Korean J. Chem. Eng., **16**(1), 144-149 (1999)
SHORT COMMUNICATION

A STUDY ON THE PREDICTION OF OZONE FORMATION IN AIR POLLUTION

Sea Cheon Oh, Sang Hyun Sohn, Yeong-Koo Yeo† and Kun Soo Chang*

Dept. of Chem. Eng., Hanyang University, Seoul 133-791, Korea
*Automation Research Center, Dept. of Chem. Eng., POSTECH
(*Received 9 June 1998 • accepted 5 November 1998*)

Abstract – Two prediction schemes-time series analysis and parameter estimation method-were investigated to predict the formation of ozone in Seoul, Korea. Moving average method and double exponential smoothing method are applied to the time-series analysis. Three typical methods, such as extended least squares (ELS), recursive maximum likelihood (RML) and generalized least squares (GLS), were used to predict ozone formation in a real time parameter estimation. Autoregressive moving average model with external input (ARMAX) is used as the model of the parameter estimation. To test the performance of the ozone formation prediction schemes proposed in the present work, the prediction results of ozone formation were compared to the real data. From the comparison it can be seen that the prediction scheme based on the parameter estimation method gives a reasonable accuracy with limited prediction horizon.

Key words : Ozone, Prediction, Time-Series Analysis, Parameter Estimation Method

INTRODUCTION

Recently, substantial evidence has been accumulated that ozone affects the health of human beings and animals, damages vegetation and soil, deteriorates materials, affects climate, reduces visibility and solar radiation, and contributes to safety hazards. Ozone is not emitted directly by sources but formed in the atmosphere by chemical reactions. For the analysis of the ozone formation in air pollution it is therefore essential to understand the chemical processes taking place in the atmosphere. To characterize ozone formation in air pollution many efforts [Carter et al., 1979a,b; Sakamaki et al., 1982; Fan et al., 1996] have been devoted to the experimental investigation of air pollution chemistry. Roadknight et al. [1997] used an artificial neural network to model the interactions that occur between ozone pollution and climatic conditions.

It is of great interest to determine whether computer modeling can produce reasonable accuracy on the prediction of ozone formation in air pollution. Oh and Yeo [1998] modeled and simulated ozone formation from a propene-nitrogen oxide-wet air mixture by using a detailed reaction model. However, because of the large number of important reactions taking place in the atmosphere, the prediction of ozone formation based on the identification of the chemical reactions is very difficult. Furthermore, their natural fluctuations contribute to the observed variation in the frequency and the intensity of episodes in different geographic locations and at different times of the year.

Therefore, in the prediction of ozone formation, the mass conservation equation representing the combined problem of transport, source, and sink terms such as emissions, chemistry and removal at the surface have been used in most urban and regional photochemical models so far [Carmichael et al., 1986; Chang et al., 1987; Venkatram et al., 1992; Scheffe and Morris, 1993]. However, though the mass conservation equation can be used as a data analysis tool of air quality, this equation does not yet give a reasonable accuracy on the multi-period ahead prediction due to the deviations of the wind field, diffusion and chemistry by natural fluctuations. For these reasons we employed the time-series analysis and the parameter estimation method as the prediction schemes of ozone formation. The moving average method and double exponential smoothing method can be used in the time-series analysis. Extended least squares (ELS) method, recursive maximum likelihood (RML) method and generalized least squares (GLS) method can be used to predict ozone formation in real time parameter estimation. To test the performance of the ozone formation prediction schemes proposed in this work, the prediction results of ozone formation were compared to the actual measured data of ozone formation.

TIME-SERIES ANALYSIS

1. Moving Average Method

The accuracy of a prediction depends on the accuracy of the data, the stability of the data-generating process, the length of the prediction period and the prediction method used. The regression method requires that the values of the independent variables in the prediction must be known. But, if the independent variables are random variables, we can get a reasonable prediction. A class of much easier to use tracking models that does not require exogeneous independent variables is the moving average method [Elsayed and Boucher, 1994]. The moving average method tracks the changing movement of the variable of interest as a function of its prior level. When a trend occurs in the data-generating process, the appropriate descrip-

†To whom correspondence should be addressed.
E-mail : ykyeo@email.hanyang.ac.kr

tion is as

$$y_t = a_{0,t} + a_{1,t}\tau + \varepsilon_t \tag{1}$$

where $a_{0,t}$, $a_{1,t}$, τ, and ε_t are the levels assumed to be changing with the trend at time t, the slope at time t, the number of periods into the future for which the prediction is to apply, and the random variable that represents the random noise in the data, respectively. The model of the process (1) can be written as

$$y_t^* = a_{0,t}^* + a_{1,t}^*\tau \tag{2}$$

where the superscript $*$ denotes the values related to the model. In Eq. (2) $a_{0,t}^*$ and $a_{1,t}^*$ are parameters to be adjusted, which can be obtained from the simple moving average and the double moving average. The observed value y_t can be written by using the simple moving average as

$$y_t = MA_t^{[1]} + \left(\frac{n-1}{2}\right)a_{1,t}^* \tag{3}$$

where $MA_t^{[1]}$ represents the simple moving average taken at time t and n is the number of observations in $MA_t^{[1]}$. The double moving average is defined by the form

$$MA_2^{[2]} = \frac{1}{n}\sum_{i=t-n+1}^{t} MA_i^{[1]} \tag{4}$$

The double moving average is a moving average of the simple moving average. The double moving average lags the simple moving average by the same amount as the simple moving average lags the actual level of the data. This observation leads to the following estimation for the level $a_{0,t}$ as

$$a_{0,t}^* = MA_t^{[1]} + (MA_t^{[1]} - MA_t^{[2]}) = 2MA_t^{[1]} - MA_t^{[2]} \tag{5}$$

Also, since $y_t = a_{0,t}^*$, the parameter $a_{1,t}^*$ can be obtained from Eqs. (3) and (5) as

$$a_{1,t}^* = \left(\frac{2}{n-1}\right)(MA_t^{[1]} - MA_t^{[2]}) \tag{6}$$

In this work the prediction of demand is computed by using Eqs. (2), (5) and (6).

2. Double Exponential Smoothing Method

Exponential smoothing is a mathematical technique that utilizes the same principle of an *N*-period moving average and does not require historical data over a long period of time. Exponential smoothing is especially useful for short-term prediction [Elsayed and Boucher, 1994]. The model of the system (1) can be written by using simple exponential smoothing method of the form

$$\begin{aligned} y_t^* &= \alpha y_t + (1-\alpha)y_{t-1}^* \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} \\ &\quad + \ldots + \alpha(1-\alpha)^{t-1}y_1 + (1-\alpha)^t y_0^* \end{aligned} \tag{7}$$

where α is the weight parameter on the actual data. Defining β as $\beta = 1-\alpha$, we have

$$\begin{aligned} y_t^* &= \alpha y_t + \alpha\beta y_{t-1} + \alpha\beta^2 y_{t-2} + \ldots + \alpha\beta^{t-1}y_1 + \beta^t y_0^* \\ &= \alpha\sum_{i=0}^{t-1}\beta^i y_{t-i} + \beta^t y_0^* \end{aligned} \tag{8}$$

The expectation of y_t^* is given by

$$\begin{aligned} E[y_t^*] &= \alpha\sum_{i=0}^{t-1}\beta^i E[y_{t-i}] + \beta^t y_0^* \\ &= \alpha\sum_{i=0}^{t-1}\beta^i [a_{0,t} + a_{1,t}(t-i)] + \beta^t y_0^* \end{aligned} \tag{9}$$

As time periods increase ($t\rightarrow\infty$), $\beta^t \rightarrow 0$ and

$$\begin{aligned} E[y_t^*] &= a_{0,t} + a_{1,t}t - \frac{\beta}{\alpha}a_{1,t} \\ &= E[y_t] - \frac{\beta}{\alpha}a_{1,t} \end{aligned} \tag{10}$$

Double exponential smoothing can be obtained by substituting y_t^* for y_t in Eq. (7) as the exponential smoothing of a simple exponentially smoothed series such as

$$y_t^{*[2]} = \alpha y_t^* + \beta y_{t-1}^{*[2]} \tag{11}$$

where $y_t^{*[2]}$ is the double exponential smoothing prediction value. The expectation of $y_t^{*[2]}$ can be written as

$$E[y_t^{*[2]}] = E[y_t^*] - \frac{\beta}{\alpha}a_{1,t} \tag{12}$$

Therefore,

$$a_{1,t} = \frac{\alpha}{\beta}(E[y_t^*] - E[y_t^{*[2]}]) \tag{13}$$

From Eqs. (10) and (12) we have

$$E[y_t] = 2E[y_t^*] - E[y_t^{*[2]}] = a_{0,t} \tag{14}$$

The estimates of $a_{0,t}$ and $a_{1,t}$ can be obtained from Eqs. (13) and (14) respectively, as

$$a_{0,t}^* = 2y_t^* - y_t^{*[2]} \tag{15}$$

$$a_{1,t}^* = \frac{\alpha}{\beta}(y_t^* - y_t^{*[2]}) \tag{16}$$

In this work, the prediction using double exponential smooth method is computed by Eqs. (2), (15) and (16).

PARAMETER ESTIMATION METHOD

The parameter estimation method used in the present work is primarily intended for process identification in an adaptive control system. Many efforts have been devoted to the development of the effective parameter estimation [Frick and Valavi, 1978; Kubrusly, 1981; Zhang, 1983; Wang et al., 1987]. Samson [1983] analyzed the identification methods for the discrete-time system subject to bounded disturbances. Oh and Yeo [1996] and Choi et al. [1988] have used the Autoregressive moving average (ARMA) model and ARMAX model in on-line identification. However, the ARMA model cannot represent the effect of the disturbances on the system output. Therefore, in this work the ARMAX model, ARMA model with external input, is used as the model of the parameter estimation. The multi-input single-output process for ozone formation can be described by ARMAX representation of the form

$$y(t+1)=\sum_{i=1}^{n}a_i y(t+1-i)+\sum_{j=1}^{m}\sum_{i=1}^{n}b_{j,i}u_j(t+1-d-i)+\sum_{i=1}^{n}c_i e(t+1-i)+e(t+1) \tag{17}$$

where n and m denote the model order and the number of inputs. The model of the process (17) can be written as

$$\begin{aligned}y^*(t+1)&=\sum_{i=1}^{n}a_i^* y(t+1-i)+\sum_{j=1}^{m}\sum_{i=1}^{n}b_{j,i}^*u_j(t+1-d-i)+\sum_{i=1}^{n}c_i^* e(t+1-i)\\&=\theta^*(t+1)^T\phi(t)\end{aligned} \tag{18}$$

where θ^* and ϕ are the model parameter vector and the process data vector, respectively, and d is the known time delay, but we do not need exact knowledge of the process structure. In order to estimate the model parameter vector θ^*, we use a recursive parameter estimation algorithm [Landau, 1990] of the form

$$\theta^*(t+1)=\theta^*(k)+F(t)\phi(t)\eta(t+1) \tag{19}$$

where

$$\theta^*(t)^T=[a_1^{0*}(t)\ldots a_n^*(t),\ b_{11}^*(t)\ldots b_{1n}^*(t),\ldots,\ b_{m1}^*(t)\ldots b_{mn}^*(t),\ c_1^*(t)\ldots c_n^*(t)]$$

and the prediction error η is defined by

$$\eta(t+1)=y(t+1)-y^*(t+1)=\frac{y(t+1)-\theta^*(t)^T\phi(t)}{1+\phi(t)^T F(t)\phi(t)} \tag{20}$$

The adaptation gain F in Eq. (19) is calculated as follows:

$$F(t+1)=\frac{1}{\lambda_1}(t)\left[F(t)-\frac{F(t)\phi(t)\phi(t)^T F(t)}{\frac{\lambda_1(t)}{\lambda_2(t)}+\phi(t)^T F(t)\phi(t)}\right] \tag{21}$$

where

$$0<\lambda_1(t)\le 1$$

$$0\le\lambda_2(t)<2$$

$$F(0)=\begin{bmatrix}GI & & 0\\ & \ddots & \\ 0 & & GI\end{bmatrix}$$

$$GI>0$$

The process data vector ϕ for the ELS method is represented as

$$\phi(t)^T=[y(t)\ldots y(t-n+1),\ u_1(t)\ldots u_1(t-n+1),\ \ldots,\ u_m(t)\ldots u_m(t-n+1),\ \eta(t)\ldots\eta(t-n+1)] \tag{22}$$

For RML method, ϕ can be written by

$$\phi(t)^T=\frac{1}{C^*(t,q^{-1})}[y(t)\ldots y(t-n+1),\ u_1(t)\ldots u_1(t-n+1),\ \ldots,\ u_m(t)\ldots u_m(t-n+1),\ \eta(t)\ldots\eta(t-n+1)] \tag{23}$$

In Eq. (23) the estimated polynomial $C^*(t, q^{-1})$ at instant t is written as

$$C^*(t,q^{-1})=1+c_1^*(t)q^{-1} \tag{24}$$

which corresponds to the filtering of the component of $\phi(t)$ for the ELS method by $1/C^*(t, q^{-1})$. ϕ for the GLS method is given by representation of the form

$$\phi(t)^T=[y(t)\ldots y(t-n+1),\ u_1(t)\ldots u_1(t-n+1),\ldots,\ u_m(t)\ldots u_m(t-n+1),\ \alpha(t)\ldots\alpha(t-n+1)] \tag{25}$$

The quantity $\alpha(t)$ is estimated as follows:

$$\begin{aligned}\alpha(t)=&[(1-a_1^*(t)q^{-1}-\ldots-a_n^*(t)q^{-n})y(t)]-[b_{11}^*(t)u_1(t-d-1)\\&+\ldots+b_{1n}^*u_1(t-d-n)+\ldots+b_{m1}^*u_m(t-d-1)\\&+\ldots+b_{mn}^*u_m(t-d-n)]\end{aligned} \tag{26}$$

RESULTS AND DISCUSSION

Data used to test the performance of the ozone formation prediction schemes proposed in this work were provided by the Korean Environmental Office and were collected in the form of hourly readings of pollutants for May 11-14, 1996. The air pollutants used as parameters were ozone, nitrogen dioxide, sulphur dioxide, total hydrocarbon, methane, carbon monoxide, and nitrogen monoxide. Fig. 1 shows the results of one hour ahead prediction of the ozone formation using time-series analysis. In Fig. 1 the periods of moving average and the weight parameter α of double exponential smoothing method are 24 hours and 0.3, respectively. From Fig. 1 we can see that the time-series analysis used in this work gives a reasonable trend of ozone formation, but it lags the actual level of the data. Such a lag produces a serious prediction error, which is the main reason why the time-series analysis uses a specified past data in nature.

Fig. 2 shows the results of one and two hours ahead predictions of ozone formation based on the parameter estimation method. In order to predict the ozone formation by using the parameter estimation method, the prediction of the other pol-

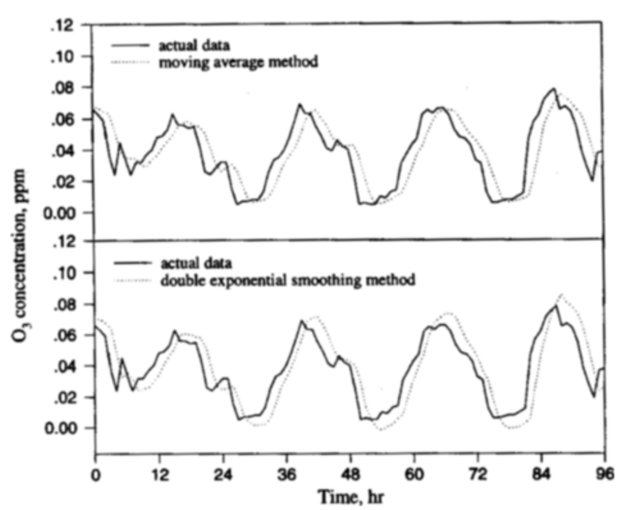

Fig. 1. One hour ahead prediction results of O_3 formation using the time-series analysis.

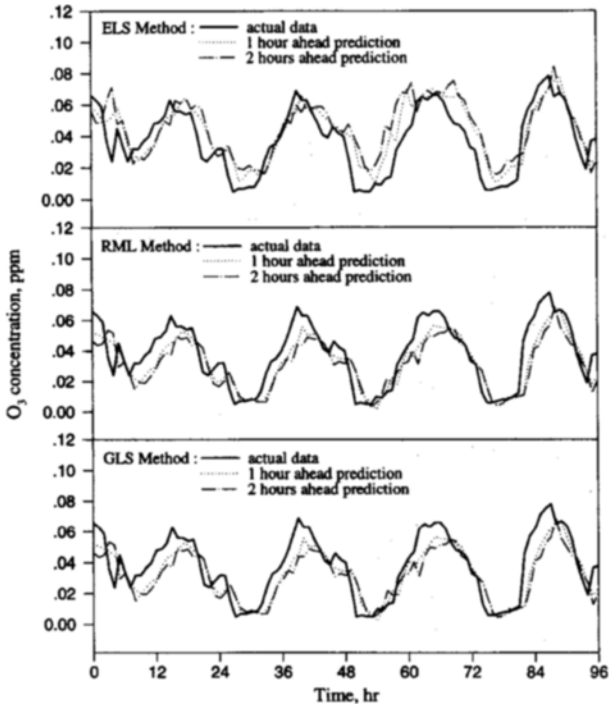

Fig. 2. One and two hours ahead prediction results of O_3 formation using the parameter estimation method.

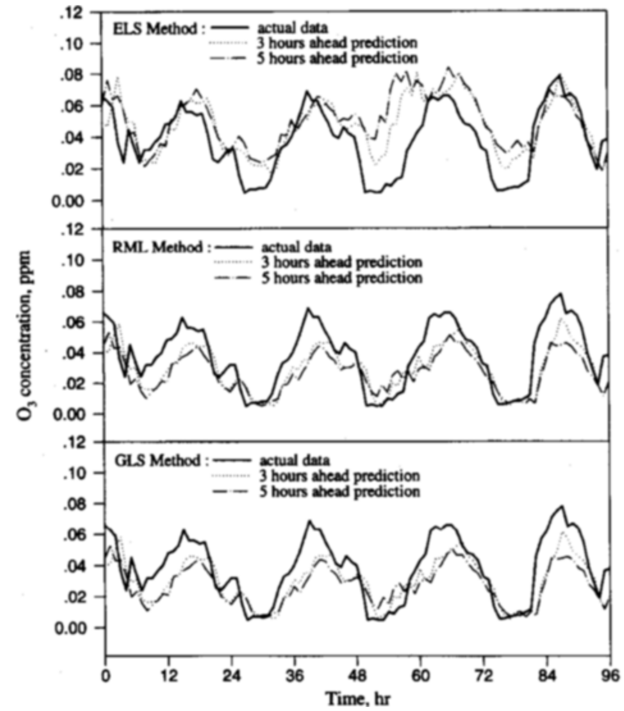

Fig. 3. Three and five hours ahead prediction results of O_3 formation using the parameter estimation method.

lutants causing the ozone formation is required unlike the time-series analysis. While ozone is formed in the atmosphere by chemical reaction, the other pollutants are mainly generated by many sources. Therefore, ozone formation was predicted by the parameter estimation method with the ARMAX model which can be applied to the identification of the chemical reactions effectively, while the other pollutants causing the ozone formation were predicted by time-series analysis and were put into the form of a set of process data vector to predict ozone formation using the ARMAX model. A high initial gain, GI=1000, was chosen because an initial parameter estimation was not available, and the initial values of the model parameters were set to 0. In Eq. (21), $\lambda_1(t)$ and $\lambda_2(t)$ were set to 0.99 and 1.5 by trials. One hour was chosen as the model time delay and the second-order model was used. From Fig. 2 we can see that the parameter estimation method gives a reasonable trend of ozone formation and the lagged errors are reduced as compared with the prediction results based on the time-series analysis. Also it was found that the ELS method results in positive prediction errors, while the prediction errors by RML and GLS methods are negative.

Three and five hours ahead prediction results are shown in Fig. 3. As can be seen, the prediction errors of ozone formation increase with increasing length of prediction periods. This is due to the accumulation of the prediction errors of the other pollutants predicted by the time-series analysis. Fig. 4 shows the results of prediction of ozone formation for the change of model time delay. From Fig. 4 we can see that the one hour model time delay gives an accurate prediction result as compared with the other model time delay. Fig. 5 shows the prediction results using the second-order model compar-

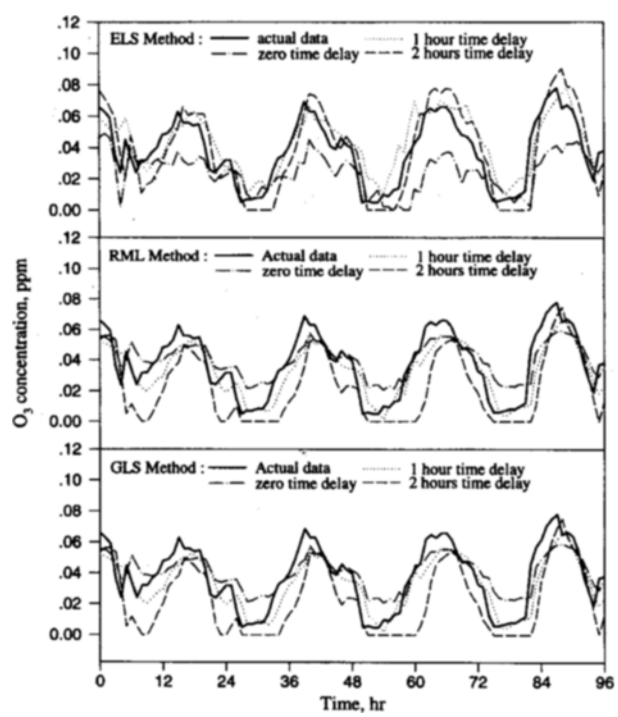

Fig. 4. One hour ahead prediction results of O_3 formation using the parameter estimation method for the change of model time delay.

ed to those from the first-order model. From these results it can be seen that the second-order model provides better prediction results than the first-order model. The prediction results of ozone formation using the arithmetic and geometric mean

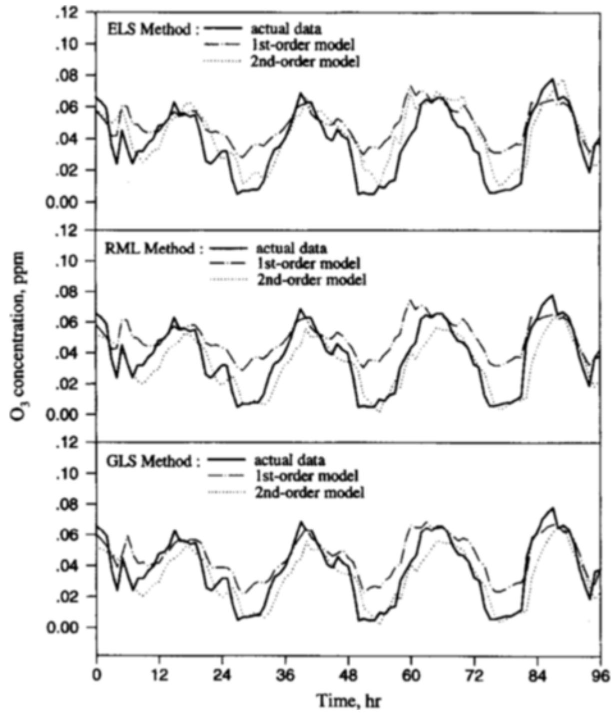

Fig. 5. One hour ahead prediction results of O_3 formation using the parameter estimation method for the change of model order.

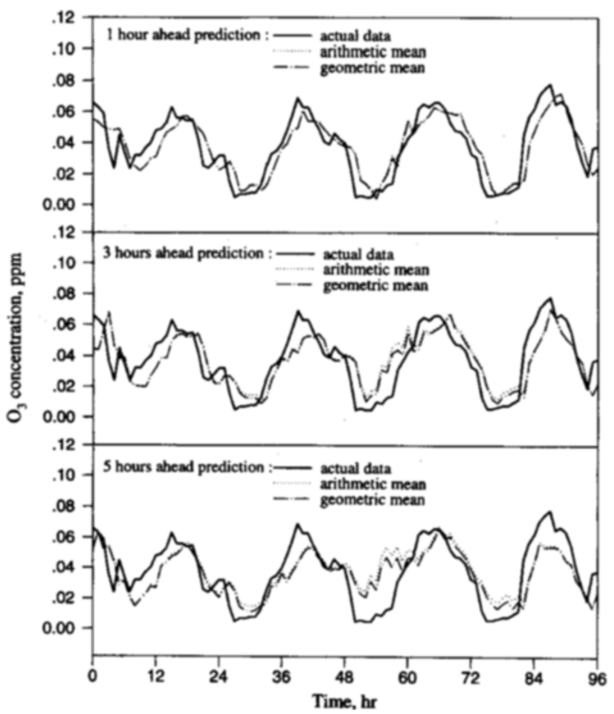

Fig. 6. Prediction results of O_3 formation using arithmetic and geometric mean of ELS and RML method.

of the ELS and RML method are shown in Fig. 6. As shown in Fig. 6, the prediction errors of ozone formation increase as the length of the prediction periods increases. It was found that the prediction scheme of the ozone formation based on the parameter estimation method gives a reasonable accuracy on the limited period ahead prediction.

CONCLUSION

In the present study the two prediction schemes of ozone formation based on a time-series analysis and the parameter estimation method are tested. While the time-series analysis used in this work gives a resonable trend of the ozone formation, it lags the actual level of the data. Further studies on the application of the time-series analysis are required to improve the effectiveness for the prediction of ozone formation. The parameter estimation methods used in this work give a reasonable accuracy on the limited period ahead prediction of ozone formation and the lagged errors are reduced as compared with the prediction results using the time-series analysis. However the prediction errors of ozone formation increase with an increase of the length of periods into the future. The development of an effective prediction scheme on the multi-period ahead prediction is a topic for future research.

ACKNOWLEDGEMENT

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Automation Research Center at POSTECH.

REFERENCES

Carmichael, G. R., Peters, L. K. and Kitada, T., "A Second Generation Model for Regional-Scale Transport/Chemistry/Deposition," *Atmos. Environ.*, **20**, 173 (1986).

Carter, W. P. L., Lloyd, A. C., Sprung, J. L and Pitts, J. N. Jr., "An Experimental Investigation of Chamber-Dependent Radical Sources," *Int. J. Chem. Kinet.*, **11**, 45 (1979a).

Carter, W. P. L., Winer, A. M., Darnal, K. R. and Pitts, J. N. Jr., "Smog Chamber Studies of Temperature Effects in Photochemical Smog," *Environ. Sci. Technol.*, **13**, 1094 (1979b).

Chang, J. S., Brost, R. A., Isaksen, I., Madronich, S., Middleton, P., Stockwell, W. R. and Walcek, C. J., "A Three-Dimensional Eulerian Acid Deposition Model: Physical Concepts and Formulation.," *J. Geophys. Res.*, **920**, 14681 (1987).

Choi, T. W., Lee, I. O., Lee, K. S. and Kim, K. B., "Real-Time Prediction of Hot Metal Temperature in Blast Furnace Using an ARMAX Model," *HWAHAK KONGHAK*, **23**, 237 (1988).

Elsayed, E. A. and Boucher, T. O., "Analysis and Control of Production Systems," Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1994).

Fan, Z., Kamens, R. M., Zhang, J. and Hu, J., "Ozone-Nitrogen Dioxide-NPAH Heterogeneous Soot Particle Reactions and Modeling NPAH in the Atmosphere," *Environ. Sci. Technol.*, **30**, 2821 (1996).

Frick, P. A. and Valvavi, A. S., "Estimation and Identification of Bilinear Systems," *Atomat. Contr. Theory and Applications*, **6**, 1 (1978).

Kubrusly, C. S., "Identification of Discrete-Time Stochastic Bilinear Systems," *Int. J. Control*, **33**, 291 (1981).

Landau, I. D., "System Identification and Control Design," Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1994).

Oh, S. C. and Yeo, Y. K., "Modeling and Simulation of Ozone Formation from a Propene-Nitrogen Oxide-Wet Air Mixture in a Smog-Chamber," *Korean J. Chem. Eng.*, **15**, 20 (1998).

Oh, S. C. and Yeo, Y. K., "On-Line Identification of Interacting Two-Tank System," *Korean J. Chem. Eng.*, **13**, 422 (1996).

Roadknight, C. M., Balls, G. R., Mills, G. E. and Palmer-Brown, D., "Modeling Complex Environmental Data," *IEEE Transactions on neural networks*, **8**, 852 (1997).

Sakamaki, F., Okuda, M., Akimoto, H. and Yamazaki, H., "Computer Modeling Study of Photochemical Ozone Formation in the Propene-Nitrogen Oxides-Dry Air System. Generalized Maximum Ozone Isopleth," *Environ. Sci. Technol.*, **16**, 45 (1982).

Samson, C., "Stability Analysis of Adaptively Controlled Systems Subject to Bounded Disturbances," *Atomatica*, **19**, 81 (1983).

Scheffe, R. D. and Morris, R. E., "A Review of the Development and Application of the Urban Airshed Model," *Atmos. Environ.*, **27B**, 23 (1993).

Venkatram, A., Karamchandani, P., Kuntasal, G., Misra, P. K. and Davies, D. L., "The Development of the Acid Deposition and Oxidant Model (ADOM)," *Environ. Pollut.*, **75**, 189 (1992).

Wang, M. L., Chang, R. Y. and Yang, S. Y., "Analysis and Parameter Estimation of Bilinear Systems via Generalized Orthogonal Polynomials," *Int. J. Control*, **16**, 641 (1985).

Zhang, Y. G., "Identification of a Class of Stochastic Bilinear Systems," *Int. J. Control*, **38**, 991 (1983).