SOME PROBLEMS OF VALIDATION AND TESTING OF NUMERICAL AIR POLLUTION MODELS

by

PAUL SWITZER

and

PAOLO ZANNETTI

TECHNICAL REPORT NO. 21

SEPTEMBER 1978

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH

PREPARED UNDER SUPPORT TO SIMS FROM

ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION (ERDA)
ROCKEFELLER FOUNDATION
SLOAN FOUNDATION
ENVIRONMENTAL PROTECTION AGENCY (EPA)
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

SOME PROBLEMS OF VALIDATION AND TESTING OF NUMERICAL AIR POLLUTION MODELS

Paul Switzer
Stanford University

and

Paolo Zannetti
IBM Scientific Center
Palo Alto, California and Venice, Italy

and

Stanford University

1. Introduction.

Air pollution numerical modeling has been developed by following two main approaches: deterministic models and stochastic ones. The former consist of prediction formulas based on the physics of transport and diffusion of pollutants. In this way the knowledge of the "cause" (emission and meteorology) allows the computation of the "effect" (concentration) according to the physics of the phenomenon. This approach, after a careful calibration in the study area, can be used for short-term forecasting purposes where good short-term forecasting of meteorological and emission parameters is available. Deterministic predictors based on physical models are especially useful for control and planning problems, where temporary or permanent modifications of the emission pattern are contemplated. The most common predictors of

this type are the Gaussian plume formula, the Gifford-Hanna model and the K-model, i.e., the model based on the numerical integration of the mass-dispersion differential equation of fluid dynamics. According to our experience, e.g., see Runca et al. (1976), deterministic models and, in particular, the Gaussian model are especially suited to the fitting of long-term average concentration of values (monthly or seasonal averages). In fact, the real-time short-term forecasting for alert control using deterministic predictors shows important limitations as can be seen in Zannetti (1978).

Stochastic models are based on statistical or semi-empirical assumptions. They have been applied for the identification of the frequency distribution of concentration measures, e.g., Larsen (1969), for the analysis of trends and periodicities in meteorological and air quality time series, e.g., Trivikrama et al. (1976), and for validation and testing of semi-empirical relations among the data, e.g., Tiao et al. (1976). In particular, stochastic or mixed predictors can be successfully applied for episode control, i.e., real-time short-term forecasting of high pollution levels. These models make use of present and past available data (emission, meteorology and concentration) for the forecasting of future concentration values and, in particular, they make use of the actual concentration data measured at times close to the forecasting time. Such information is often more relevant than that obtained by the always insufficient physical understanding of the phenomenon.

An analysis of the problem of validation and testing of a general air pollution numerical model is developed in Section 2, where the differ-

ent diagnostic checks applied in the literature are discussed in order to identify a good validation procedure in this field. Finally, in Section 3, previous remarks are illustrated by developing a simple numerical model for the short-term real-time forecasting of CO hourly values in the St. Louis, Missouri area.

2. Validation of a Predictor.

All fitted models produce a time series of computed concentration $\underset{\sim}{\hat{c}}_s$ for each monitoring station s during the period of the study. These time series should be compared with the actual measured data $\underset{\sim}{c}_s$ in order to test the fitting or forecasting performance of the model acting as a predictor. Often this test is limited to the computation of the correlation coefficient r_s between the two series $\underset{\sim}{c}_s$ and $\underset{\sim}{\hat{c}}_s$. The information supplied by this correlation is certainly insufficient and represents only a measure of linear relationship between the two time series. In fact, it can be easily shown that one can have very large prediction errors and still have a very high correlation and conversely. A high r_s is associated with a good predictor if and only if the mean of the error $\underset{\sim}{\varepsilon}_s = \underset{\sim}{c}_s - \underset{\sim}{\hat{c}}_s$ is close to zero and its standard deviation is low. Otherwise a high correlation coefficient means only that there exists a linear transformation $\underset{\sim}{\hat{c}}'_s = m\underset{\sim}{\hat{c}}_s + q$ of the computed data (m and q are parameters to be estimated) which is very close to the actual data c_s.

The analysis of the time series of the errors $\underset{\sim}{\varepsilon}_s$ should not be limited to the computation of its mean and standard deviation. A careful investigation between input parameters (meteorology and emission), space-

time parameters (station position, hour of the day, season), and modeling error is recommended, as partially shown in Shieh and Shir (1976), in order both to identify the conditions under which the model does better or worse and to propose model modification. Moreover, a very important evaluation tool is the computation of the autocorrelation function $\underset{\sim}{r}_{\underset{\sim}{\varepsilon}_s}$ of $\underset{\sim}{\varepsilon}_s$. First, this function may reveal time periodicities of the error, for example a diurnal or semi-diurnal cycle. Second, the autocorrelation function of the error may be used for the computation of a low-powered but simple test of the model, the Portmanteau lack of fit, due to Box and Pierce (1970). Finally, the $\underset{\sim}{r}_{\underset{\sim}{\varepsilon}_s}$ function can be used, according to Box and Jenkins (1970), as an identification tool for more refined modeling of the error in order to improve the performance of the predictor.

If the predictor used to calculate $\underset{\sim}{\hat{c}}_s$ contains some adaptive features, i.e., the computation of some model parameters is done on a learning period of fixed length immediately preceeding the forecasting time, then, according to Zannetti et al. (1978), the performance of the model should be plotted versus the length of the learning period in order to identify its most suitable length.

The validation of an air pollution predictor might also be performed using only time periods of high measured concentration in order to identify the performance of the model and its relation to input parameters with respect to the air pollution "episodes". The study of these cases can be performed either by defining an episode as a concentration exceeding some specified standard, see Zannetti et al. (1977), or by considering an episode for a station as a concentration exceeding the station mean plus its standard deviation, according to Finzi et al. (1978).

3. Testing a Model for Short-Range CO Prediction in the St. Louis Air Basin.

For our analysis we have used meteorological and CO hourly data collected at Station 4 of the St. Louis, Missouri RAMS air pollution network (data supplied by EPA). The station is located in East St. Louis and the period of the analysis was June 1 through August 31, 1976, which provided us with data for a total of 2208 hourly time periods.

The predictor has been defined in the following way. At each hour t the meteorological and emission conditions of that hour are associated with one of N specified classes. The associated class at time t is denoted $\alpha(t)$. In our case, by analyzing wind speed and direction, temperature, day of the week (working day, Saturday and Sunday) and hour of the day, it has been possible to define $N=50$ different classes with a sufficient number (greater than 20) of hourly occurrences during the study period. The day-of-week and hours-of-day classes are chosen as surrogate emissions classes.

For any given condition class α^*, let $\underset{\sim}{T}^0_{\alpha^*} \equiv \{t_1^*, t_2^*, \ldots\}$ be the collection of hourly periods during which that condition α^* obtained. Define $\mu^0_{\alpha^*}$ and $\sigma^0_{\alpha^*}$ to be the mean and the standard deviation of the observed CO hourly concentrations at times $\underset{\sim}{T}^0_{\alpha^*}$. Similarly define $\underset{\sim}{T}^k_{\alpha^*} \equiv \{t_1^*+k, t_2^*+k, \ldots\}$ and let $\mu^k_{\alpha^*}$ and $\sigma^k_{\alpha^*}$ be the mean and standard deviation of the observed CO hourly concentrations at times $\underset{\sim}{T}^k_{\alpha^*}$. Finally, define $\rho^k_{\alpha^*}$ to be the lag k autocorrelation between the CO concentrations at times $\underset{\sim}{T}^0_{\alpha^*}$ and those at times $\underset{\sim}{T}^k_{\alpha^*}$. With these $5\times N$ parameters (in our case 250) for each k, we apply the following k-hour-ahead predictor:

$$(1) \qquad \frac{\hat{c}(t+k) - \mu^{k}_{\alpha(t)}}{\sigma^{k}_{\alpha(t)}} = \rho^{k}_{\alpha(t)} \frac{c(t) - \mu^{0}_{\alpha(t)}}{\sigma^{0}_{\alpha(t)}}$$

that allows the estimation $\hat{c}(t+k)$ of $c(t+k)$ on the basis of the observed condition class and CO concentration at time t.

In the case where $c(t)$ and $c(t+k)$ are modeled to have a joint normal distribution in each condition class, then this predictor is equivalent to the conditional mean of $c(t+k)$ given the data at time t. The predictor may also be regarded as an AR(1) model conditioned on the emission-meteorological class.

The fitting performance of (1) during summer 1976 has been compared with those of the following models:

a. persistence model: $\hat{c}(t+k) = c(t)$;

b. AR(1) model: $c(t+k)-\mu = \rho^{k}(c(t)-\mu)$ with μ and ρ parameters to be estimated;

c. the log-antilog version of (1): $\hat{c}(t+k) = \exp \hat{c}_{log}(t+k)$;

d. model (1) applied with $N=6$ classes depending on day type and hour of the day only;

e. model (1) applied with $N=8$ classes depending on wind direction sector only.

The results obtained are summarized in the figures. For the six models described, we have plotted the prediction lag k against the root mean square of the prediction error (Figure 1), the correlation coefficient between observed and predicted values (Figure 2), and the root mean square of the prediction error for the episodes, i.e., where measured concentration values are greater than their average plus one

standard deviation (Figure 3). All these show that model (1) with N=50 produces the best performance. We would expect that the differences in performance for the various models would not be so great if the validation had been done on an independent data set which was not used for the estimation of model parameters.

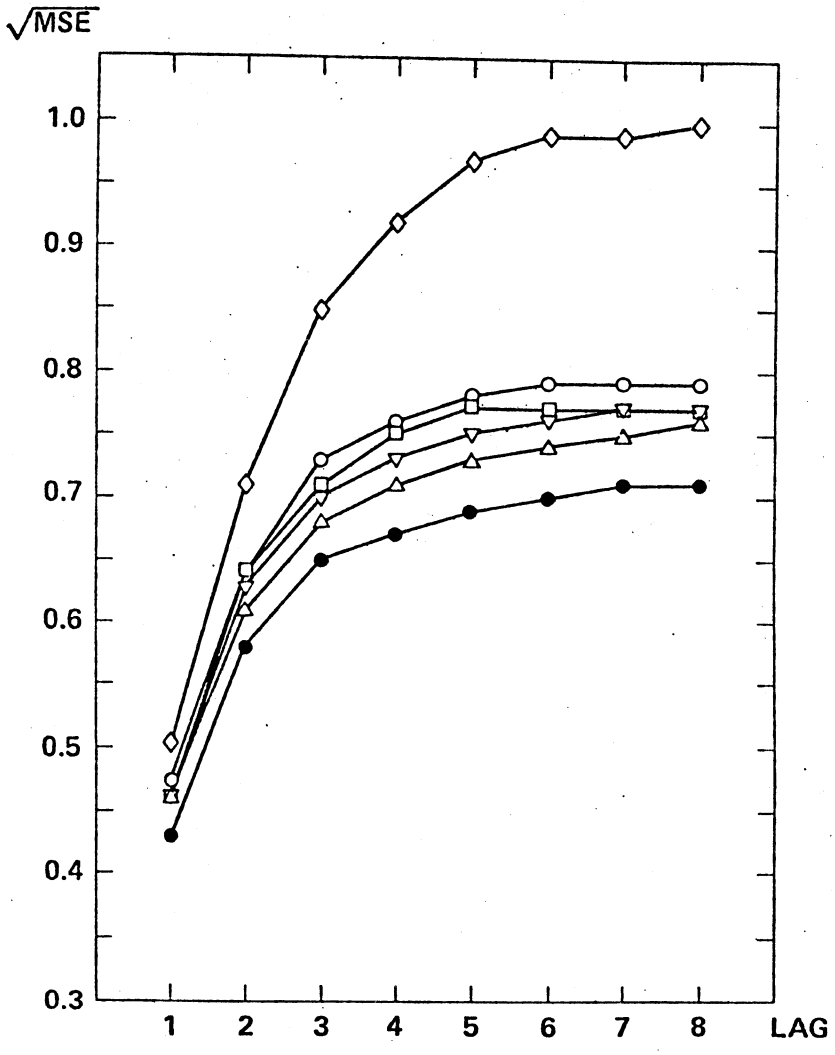

Figure 1. Root mean square of the prediction errors plotted against the forecasting lag k for six forecasting models: persistence [◇], AR(1) [o], model (1) with N=50 [•], model (1) with N=50 and log-antilog transformation [Δ], model (1) with N-6 [▽], and model (1) with N=8 [□].

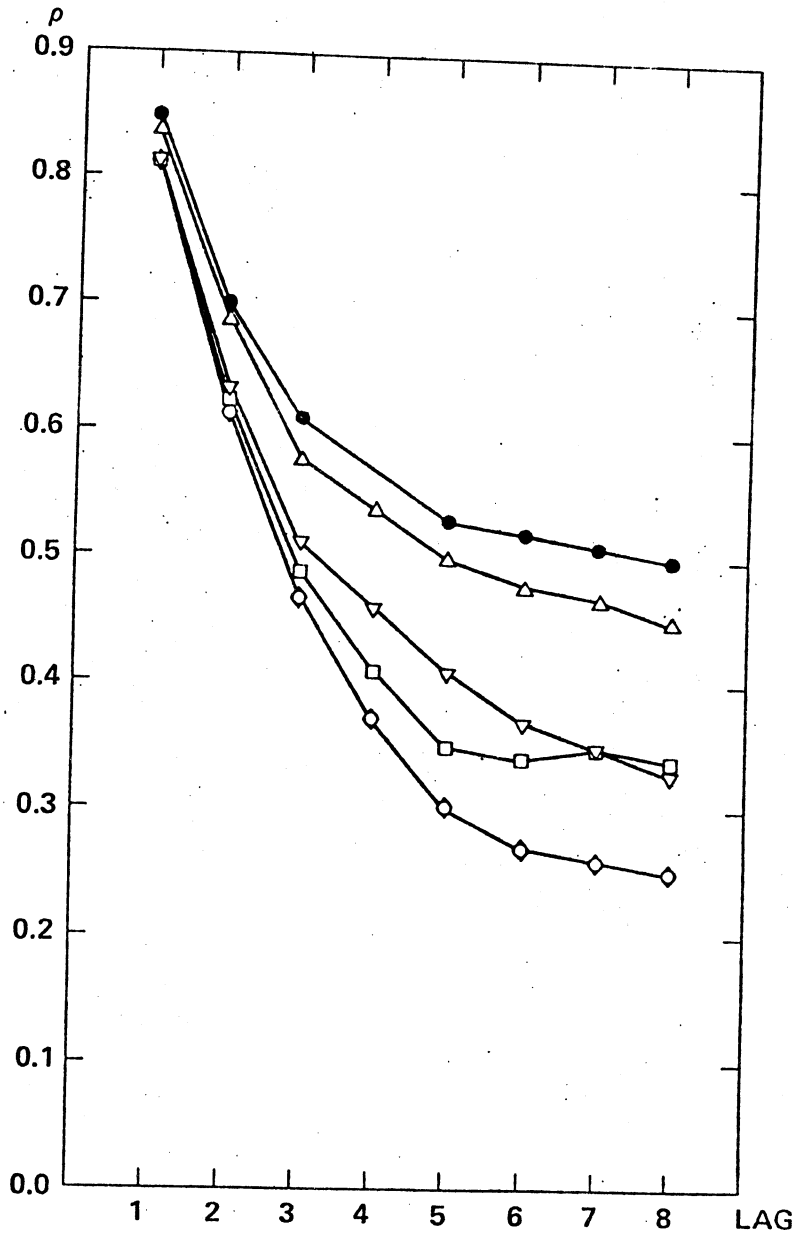

Figure 2. Correlation coefficient between observed and predicted values plotted against the forecasting lag k (same six models of Figure 1).

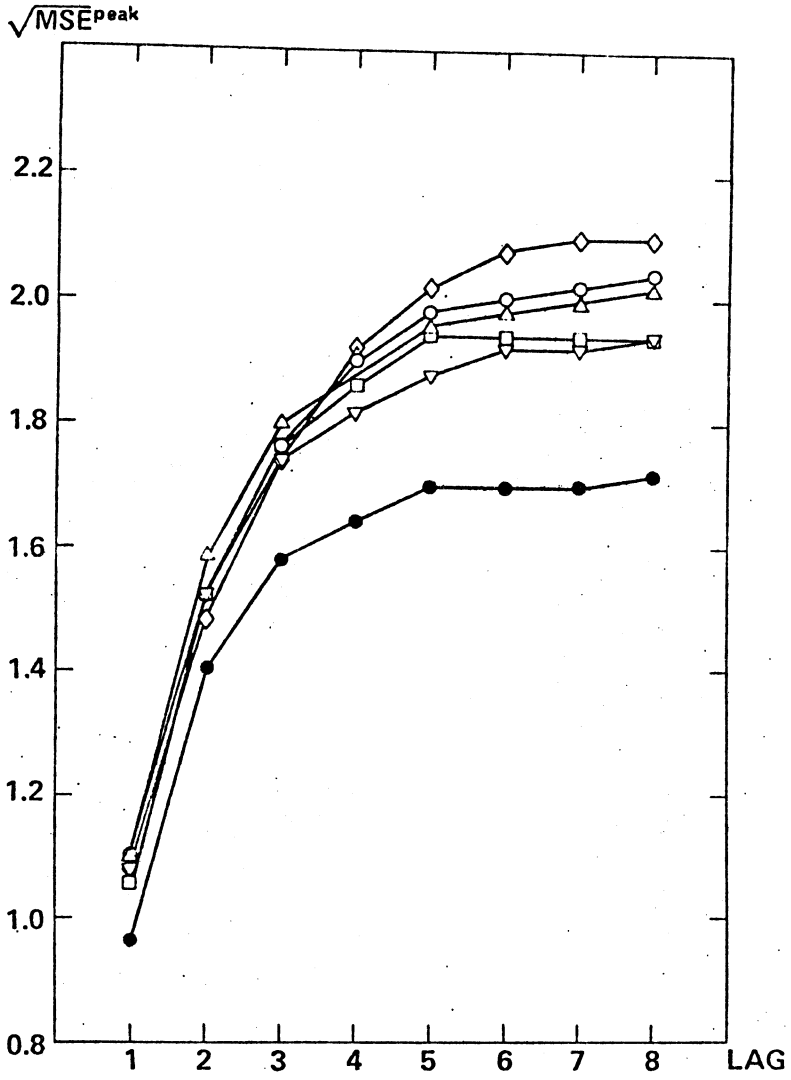

Figure 3. Root mean square of the prediction error plotted against the forecasting lag k (same six models of Figure 1) for high concentration values.

The analysis of the autocorrelation function of the error shows a well-marked daily periodicity. However, persistence or AR(1) models applied to the error time series did not seem to improve the forecasting performance of the predictor.

The plot of measured and computed concentrations shows different performance for each day. In particular all the models have difficulty handling rapid increases of concentration values (episodes) because their form is more suited to the description of damping phenomena. However, in some cases, like the three days plotted in Figure 4, the model (1) is able to handle very high peaks (1 hour ahead prediction).

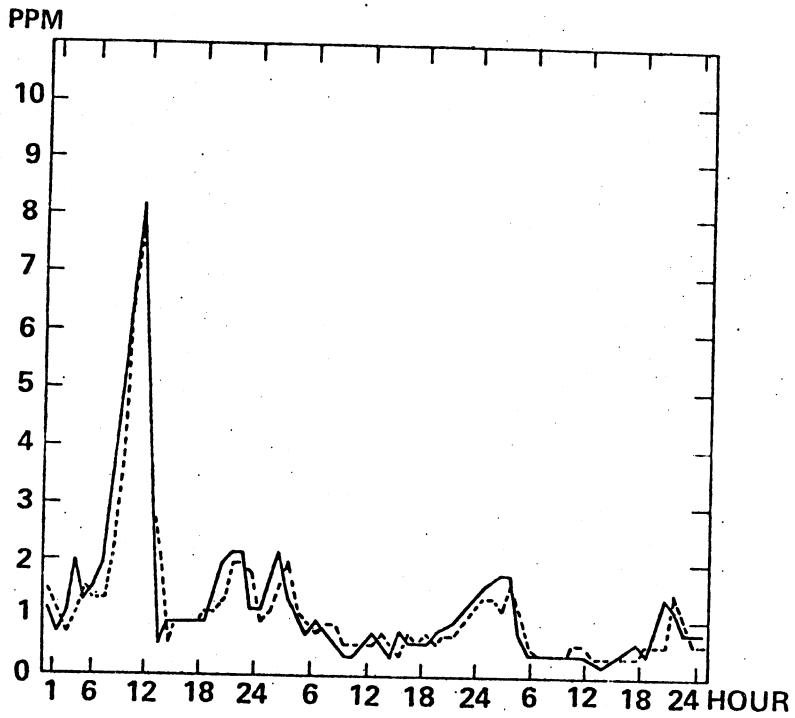

Figure 4. Measured (solid line) and forecasted (dashed line) one hour ahead CO concentrations during the three-day period June 25-27, 1976. Forecasting performed by applying model (1) with N=50.

Model performance should be correlated with input parameters (e.g. meteorology). For example the application of (1) with N=8 (corresponding to eight wind direction sectors) suggests the analysis of prediction

errors for each wind direction separately as shown in Figure 5. In this way it is possible to identify those wind conditions associated with better and worse performance of the predictor.

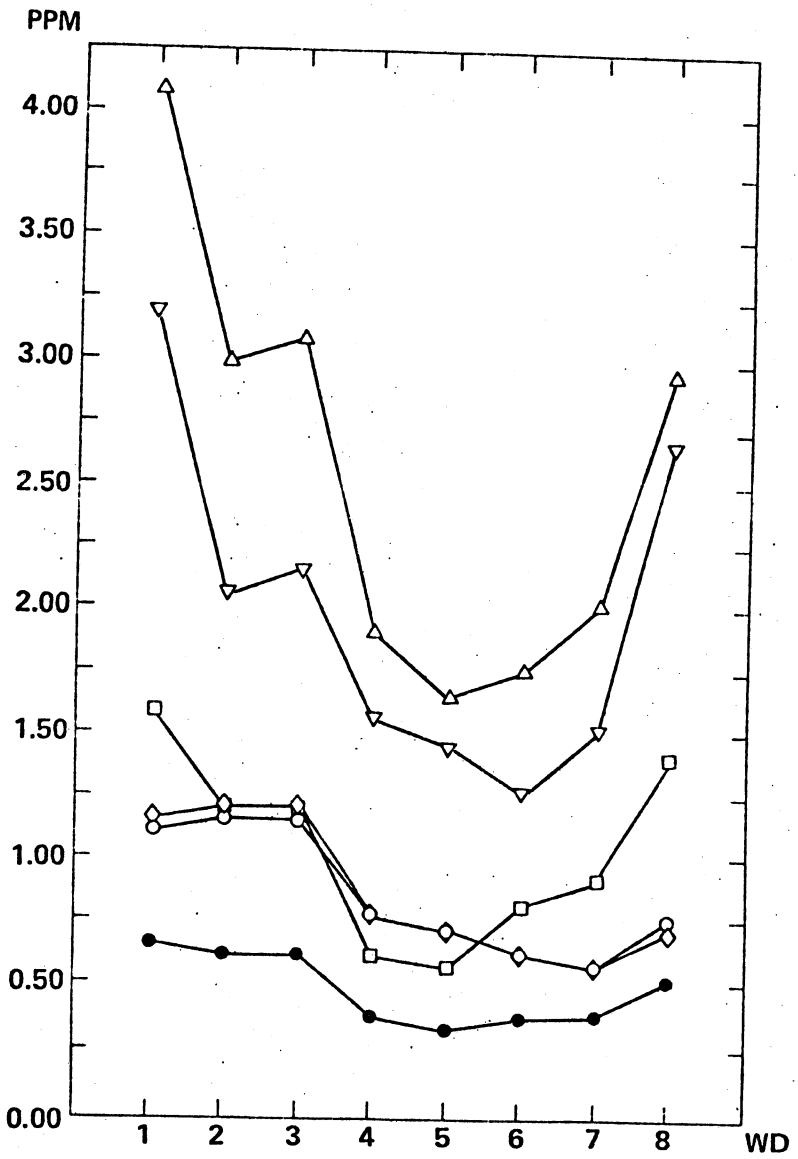

Figure 5. For each wind direction sector (1=N-NE, 2=NE-E, 3 3=E-SE, 4=SE-S, 5=S-SW, 6=SW-W, 7=W-NW, 8=NW-N) the following values are presented: average measured CO concentration [◇], average one-hour-ahead forecasted CO concentration [o], root mean square of the one-hour-ahead prediction error [•], and the same three values [Δ, ∇, and □] for the high concentration cases.

A special final test has been applied to the six predictors. Concentration values have been divided in three categories: low, medium, and high, by using dividing concentrations equal to the average ± one-half of the standard deviation. In this way, each predictor for each lag k supplies a 3×3 table of measured versus computed concentration category in which the diagonal values represent the "correctly" forecasted concentrations. In Figure 6, the percentage of "correct" predictions is plotted versus the lag k, while Figure 7 represents the same using only cases of medium or high measured concentrations. This type of test shows that the model (1) with N=50 is not always the best, although its performance remains close to the best. Of course, it must be kept in mind that much more extensive use has been made of the data in fitting model (1) than in the case for the other models.

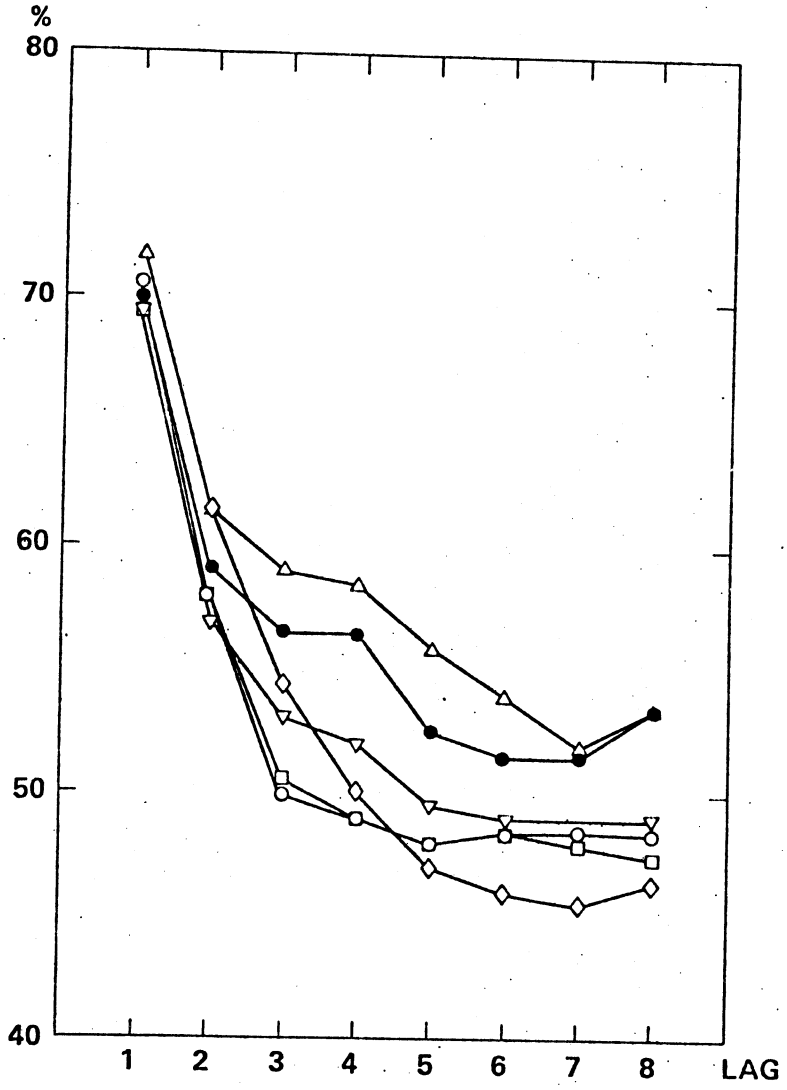

Figure 6. Percentage of "correct" prediction versus the forecasting lag (same six models of Figure 1).

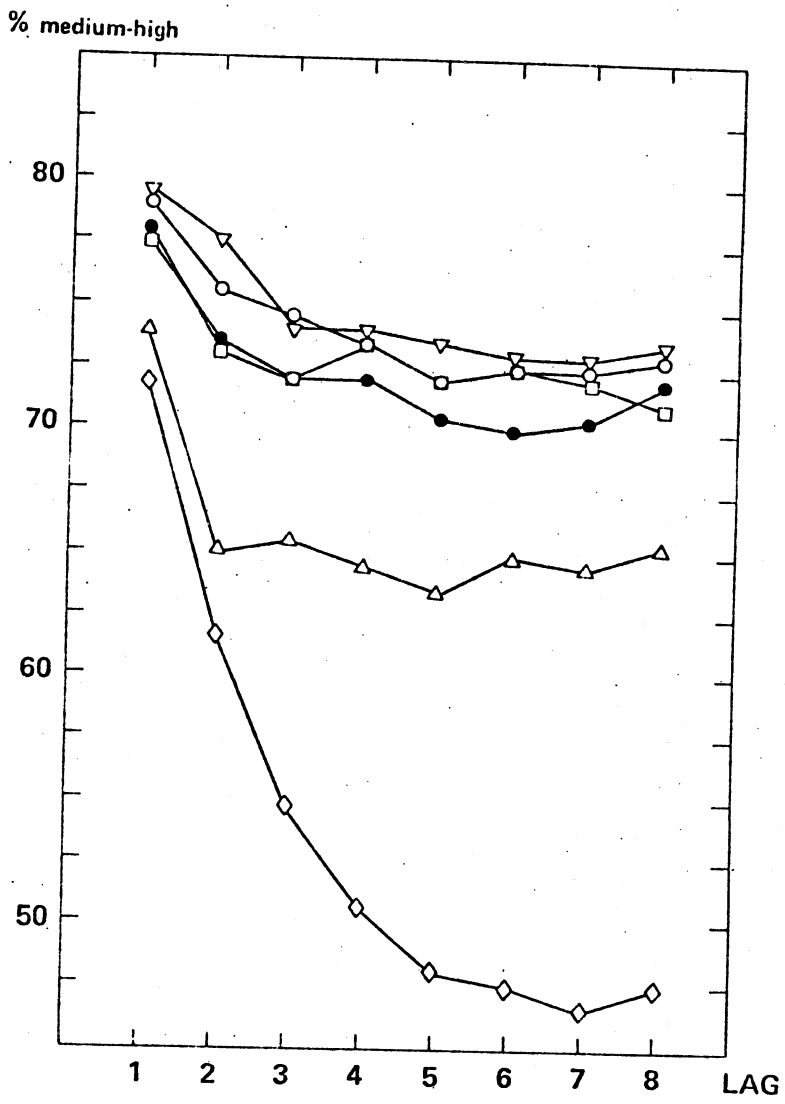

Figure 7. Percentage of "correct" prediction versus the forecasting lag for cases of medium or high measured concentrations (same six models of Figure 1).

REFERENCES

[1] Box, G.E.P., and G. M. Jenkins (1970). Time Series Analysis Forecasting and Control. San Francisco, Holden-Day.

[2] Box, G.E.P., and D. A. Pierce (1970). Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models. J.A.S.A., 65, 1509-1526.

[3] Finzi, G., G. Fronza, S. Rinaldi, and P. Zannetti (in press). Modeling and Forecasting of the Dosage Population Product in Venice. Presented at the IFAC Symposium on Environmental Systems Planning, Design and Control, August 1-5, 1977, Kyoto, Japan.

[4] Larsen, L. I. (1969). A New Mathematical Model of Air Pollution Concentration, Average Time and Frequency. J. Air Poll. Control Assoc., 19, 24-30.

[5] Runca, E., P. Melli, and P. Zannetti (1976). Computation of Long-Term Average SO_2 Concentration in the Venetian Area. Appl. Math. Modelling, 1, 9-15.

[6] Shieh, L. J., and C. C. Shir (1976). Analysis of Input Parameters and Results of Urban Air Pollution Computation. Presented at the Third Symposium on Atmospheric Turbulence, Diffusion and Air Quality, October 26-29, 1976, Raleigh, NC.

[7] Tiao, G. C., M. S. Phadke, and G.E.P. Box (1976). Some Empirical Models for the Los Angeles Photochemical Smog Data. J. Air Poll. Control Assoc., 26, 485-490.

[8] Trivikrama Rao, S., P. J. Samson, and A. R. Pedadda (1976). Spectral Analysis Approach to the Dynamics of Air Pollutants, Atmos. Environ., 10, 375-379.

[9] Zannetti, P., P. Melli, and E. Runca (1977). Meteorological Factors Affecting SO_2 Pollution Levels in Venice. Atmos. Environ., 11, 605-616.

[10] Zannetti, P. (1978). Short-Term Real-Time Control of Air Pollution Episodes in Venice. Presented at the 71st annual meeting of the APCA, June 25-30, 1978, Houston, TX.

[11] Zannetti, P., G. Finzi, G. Fronza, and S. Rinaldi (in press). Time Series Analysis of Venice Air Quality Data. Presented at the IFAC Symposium on Environmental Systems Planning, Design and Control, August 1-5, 1977, Kyoto, Japan.