Atmospheric Environment Vol. 13, pp. 1249–1255.

0004-6981/79/0901-1249 $02.00/0

J - 8

REAL TIME PREDICTION OF SO_2 CONCENTRATION IN THE VENETIAN LAGOON AREA

G. FINZI, P. ZANNETTI,* G. FRONZA and S. RINALDI
Istituto di Elettrotecnica ed Elettronica, Centro Teoria dei Sistemi C.N.R.,
Politecnico di Milano, Milan, Italy

(*First received* 12 *April* 1978 *and in final form* 12 *March* 1979)

Abstract – The paper describes a certain number of stochastic predictors of daily and hourly SO_2 concentration in the Venetian Lagoon area. More precisely, the SO_2 concentration in each inhabited subarea of the region is represented by the Dosage Area Product (DAP), namely the integral of the daily dosage over each subarea. From each stochastic model a real time predictor is derived, namely a recursive relationship allowing to forecast, at the beginning of each hour or day, the DAP levels in future periods. Moreover, a comparison is made between the performance of predictors using only recorded SO_2 data and predictors using also meteorological data such as wind directions and speeds, in order to evaluate the inclusion of such meteorological information. Possible effects of unknown fluctuations of the emission on the quality of hourly DAP forecast are also discussed.

1. INTRODUCTION

In the last few years, stochastic models such as ARIMA (Auto Regressive Integrated Moving Average) or seasonal ARIMA have been used to fit time series of pollutant concentrations (see for example Merz *et al.*, 1972; Chock *et al.*, 1975; McCollister and Wilson, 1975; Tiao *et al.*, 1975). In accordance with the techniques described by Box and Jenkins (1970), such models can be employed for supplying real time forecasts of pollutant concentrations, predicting, at intervals, future concentration levels on the basis of data recorded in previous periods.

A different and more complex type of real time predictor has also been introduced in the literature (Desalu *et al.*, 1974; Sawaragi and Ikeda, 1974; Bankoff and Hanzevack, 1975). More precisely, stochastic models derived from discretised advection–diffusion equations have been considered and the forecasts of the concentrations in all the nodes of the discretisation grid have been provided by means of Kalman filter techniques (see for instance Jazwinski, 1970). Predictors of this type exhibit a higher degree of reliability than the ARIMA ones, in particular in the forecast of episodes, though they require a much heavier computational effort.

Actually, the comparison in terms of forecast efficiency between the two approaches is not fully significant, because the amount of information used for the prediction is not the same in the two cases. In fact:

(i) no meteorological and/or emission data appear in the ARIMA models and therefore no information about meteorology and/or emission is exploited when forecasting;
(ii) in the ARIMA approach, the concentrations in the different stations are considered one by one, namely as separate stochastic processes, so that all the information supplied by the cross correlations is not taken into account.

The present paper illustrates stochastic models and predictors of daily and hourly SO_2 concentration in the Venetian Lagoon area, which has been previously investigated from different viewpoints (see for example, Runca *et al.*, 1976; Finzi *et al.*, 1977). In particular, a class of models described here represents a trade off between the ARIMA and Kalman approaches outlined above. More precisely, the simple structure of the ARIMA representation is maintained, but the characteristics (i) and (ii), lowering the forecast efficiency, are, at least partially, removed. In fact, meteorological inputs are introduced into the ARIMA models, thus turning them into the so called ARIMAX (ARIMA with eXogenous inputs, see for instance Young and Whitehead, 1977). As for (ii) instead of analysing the concentrations station by station, an overall index of pollution (Dosage Area Product or DAP) is defined for each inhabited subarea of the region under consideration (see Duckworth and Kupchanko, 1967).

2. THE VENETIAN LAGOON AREA AND THE AVAILABLE DATA

The region under consideration, the Venetian Lagoon, is shown in Fig. 1, where the three inhabited subareas of Marghera, Mestre and Venice are pointed out.

* Centro Scientifico IBM, Venice, Italy.

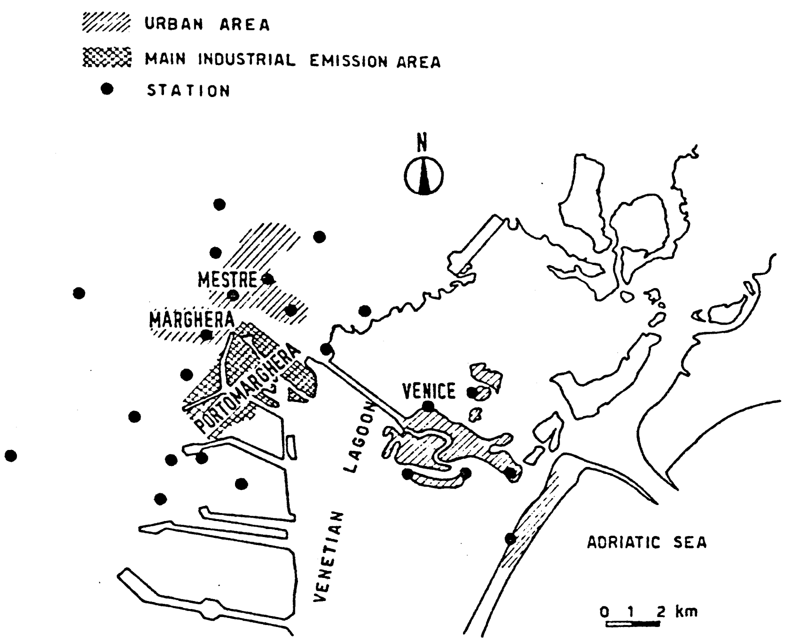

Fig. 1. The Venetian Lagoon region, the three subareas and the monitoring network.

Data used for the present analysis have been provided by the Governmental Health Department network installed in the Venetian area by Tecneco. Details on this network, such as the criteria followed for the determination of the measurement sites, the kind of instruments, the expected error of individual hourly measurements, the correlation between measures at neighboring sites, can be found in Tecneco (1973). At present the network (Fig. 1) consists of 24 stations for measuring ground level SO_2 concentration and one meteorological station located in the historical center 15 m above the ground.

The meteorological station and ten SO_2 stations have been operating continuously since February 1973, and the others since 1974. A small computer scans the sensors at each station every minute and computes hourly average values, as well as the highest daily 30 min average. Moreover, any time that the 30 min legal standard of 0.30 ppm is exceeded, a warning signal is transmitted to the local authority. The data recorded, on an hourly basis, at the meteorological station are wind speed and direction, pressure, relative humidity, temperature, rainfall, fog and cloudiness.

All the emissions of SO_2 are lumped in the industrial zone of Porto Marghera, at least in the "summer" considered in the present analysis (the period 1 May–30 September 1974). There is no reliable information on the emission rates of the sources; however, the nature of the industries suggests that the total daily emission can reasonably be considered as constant. The assumption is less reasonable at the hourly level: in fact, the performances of the hourly forecasts suggest that the lack of information about hourly fluctuations of the emissions is the main limitation of the use of the ARIMAX predictors described in this paper.

3. DEFINITION AND EVALUATION OF THE DOSAGE AREA PRODUCT

The analysis has been carried out with respect to an index which can be considered as representative of the overall pollution in each inhabited subarea.

The Dosage Area Product over a region $\mathscr{R}$ in the k-th period of time is defined as

$$\mathrm{DAP}(k) = \frac{1}{A}\int_{A}\int_{kT}^{(k+1)T} c(x, y, t)\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y, \quad (1)$$

where T = time interval, $c(x, y, t)$ = concentration at time t at point (x, y), A = area of $\mathscr{R}$.

If $D(x, y, k)$ denotes the dosage at point (x, y) in the k-th period of time, Equation (1) can also be written in the form

$$\mathrm{DAP}(k) = \frac{1}{A}\int_{A} D(x, y, k)\,\mathrm{d}x\,\mathrm{d}y. \quad (2)$$

For each k, an estimation of the integral in Equation (2) must be carried out on the basis of the dosages $\{D_i(k)\}_{i=1}^{N}$ recorded in the N monitoring stations. Such estimation can be made by means of the polygons method (see, for instance, Gray, 1974), which simply consists of the following procedure.

(i) Divide $\mathscr{R}$ in N subregions $\mathscr{R}_i$. The subregion $\mathscr{R}_i$ is defined as the locus of points in $\mathscr{R}$ for which the i-th station is the nearest station. (Hence the boundary of $\mathscr{R}_i$ in the interior of $\mathscr{R}$ turns out to have a polygonal form.)

(ii) Compute the integral in Equation (2) by the approximation $D(x, y, k) \simeq D_i(k)$ for all $(x, y) \in \mathscr{R}_i$. Thus, if A_i denotes the area of $\mathscr{R}_i$, the DAP turns out to be a convex combination of the measurements in the monitoring stations:

$$\mathrm{DAP}(k) = \frac{1}{A} \sum_{i=1}^{N} A_i D_i(k). \quad (3)$$

Through Equation (3), daily and hourly DAP time series for the three inhabited subareas have been computed in correspondence with summer 1974. Alternative methods for evaluating the DAP from the available concentration measurements, such as least squares surface interpolating techniques (see, for instance, Anderson, 1970; Finzi *et al.*, 1978) have also been used without obtaining significant differences in the statistical characteristics of the DAP series.

The DAP series evaluated through the procedure described above represent, together with the meteorological records, the data set for the stochastic models and predictors described below. Thus, for brevity, such series, though resulting from an approximate evaluation and not from direct measurement, will henceforth be referred to as "the observed DAP series".

Global pollution indicators different from the DAP could also be used to analyze the Venetian Lagoon area. In particular, the Dosage Population Product (DPP), introduced by Finzi *et al.* (1977), is particularly attractive, since it is much more clearly related with the health damages caused to the population. Nevertheless, in the present case the DAP and the DPP are proportional one to each other since the population is almost uniformly distributed in the three sub-areas. For this reason, reference is made in the following only to a single global indicator, namely the DAP.

4. STOCHASTIC MODELS AND PREDICTORS OF DAILY DAP

First, the three daily DAP time series, observed in summer 1974 in the three inhabited sub-areas, have been considered as (partial) realizations of stationary stochastic processes and, in accordance with the techniques recommended by Box and Jenkins (1970), different ARMA (=stationary ARIMA) models have been calibrated in order to fit the time series. The results of such analysis can be summarized as follows.

(a) The distribution of each process $\{\mathrm{DAP}(k)\}$ has turned out to be lognormal [see also Larsen (1969)] at more than 20% level of significance through the Kolmogorov–Smirnov test. Thus the process of logarithms or logprocess $\{y(k)\}$, where $y(k) = \ln \mathrm{DAP}(k)$, has been considered as normal.

(b) In the ARMA class, the best model of each logprocess has turned out to be AR(1) (Auto-Regressive of order 1), namely

$$y(k) - \mu = \phi(y(k-1) - \mu) + \varepsilon(k), \quad (4)$$

where μ = logprocess mean; ϕ = model parameter; $\{\varepsilon(k)\}$ = zero mean purely random process (white noise).

(c) The overall statistical test on the validity of the model (diagnostic check), precisely the cumulative periodogram test on the residuals, has shown that $\{\varepsilon(k)\}$ is actually white noise at more than 25% level of significance. Hence model (4) is not only the best in the ARMA class but is also acceptable from the view point of statistical inference.

(d) If $\hat{y}$ denotes forecast values, the one day ahead DAP predictor, derived from the stochastic model (4), is simply given by

$$\hat{y}(k) = \mu + \phi(y(k-1) - \mu), \quad (5a)$$

$$\hat{\mathrm{DAP}}(k) = \exp \hat{y}(k), \quad (5b)$$

where $y(k-1)$, the logarithm of the DAP which occurred on the $(k-1)$-th day, is an available datum at the instant of the prediction [namely at the end of the $(k-1)$-th day].

The performance of the three DAP predictors (5) has been evaluated in three ways:

(i) the correlation Θ between predicted and observed DAP;

(ii) the standard deviation S of the forecast error (=observed DAP − predicted DAP);

(iii) the mean square error S_h of the forecast error during "highly polluted days" (= days characterized by DAP $\geqslant$ mean of the DAP process + standard deviation of the DAP process).

The results of computer simulation of Equations (5) for summer 1974 have turned out rather unsatisfactory, in particular for the Marghera and Mestre subareas.

Even in the best case, corresponding to the Venice sub-area (the least polluted one) the performance has been rather poor, as shown in Table 1 (first row) and in the predicted vs observed DAP plane of Fig. 2.

The S and S_h values can be usefully compared with the DAP mean (= 12.5 ppb × day) and standard deviation (=5.4 ppb × day). The preceding result is clearly due to the structure of model (4) where the physical causes affecting the phenomenon (meteorology and emission) do not appear explicitly but contribute to the noise term $\varepsilon(k)$. Since emission

Table 1. Performance of daily DAP predictors for the Venice sub-area

DAP predictor	Θ	S [ppb × day]	S_h [ppb × day]
AR(1)	0.39	5	12
ARX(1)	0.61	4	9

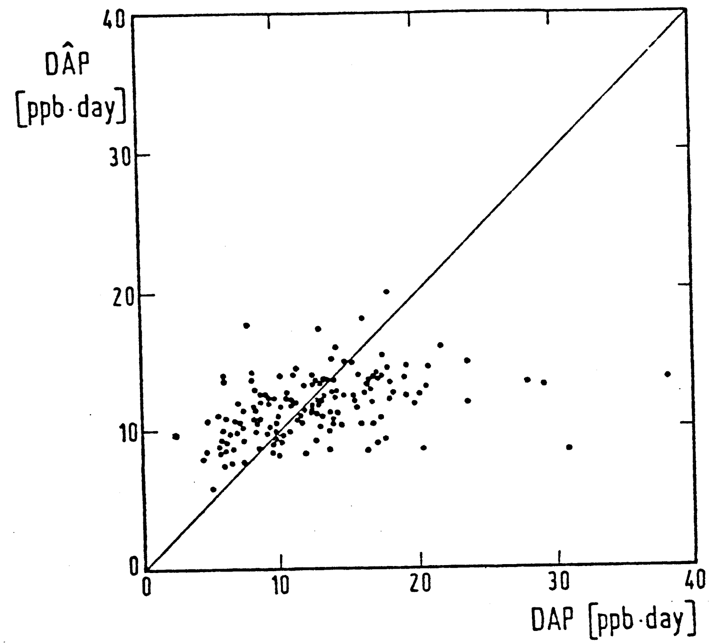

Fig. 2. Daily predicted (DÂP) vs observed DAP for the Venice sub-area [AR(1) predictor].

measurements are not available and, anyway, fluctuations from day to day are certainly not relevant, attention has been paid to qualify the role of the meteorological factors, in particular wind direction which is important in the Venetian case (see Zannetti *et al.*, 1977). In fact each sub-area is strongly asymmetric with respect to the sources, and the distribution of hourly wind direction is characterized by significant variations from day to day. Correspondingly, the three observed DAP series exhibit an "irregular" behaviour which cannot be accounted for by a simple autoregressive representation. Thus, Equation (4) has been replaced by the following ARIMAX model (specifically ARX(1):

$$y(k) - \mu = \alpha[y(k-1) - \mu] + \psi_1[u_1(k) - \mu_1] + \psi_2[u_2(k) - \mu_2] + \varepsilon(k), \quad (6)$$

where $u_1(k)$ = logarithm of the average wind speed in the k-th day; μ_1 = mean of $\{u_1(k)\}$; $u_2(k)$ = logarithm of $p(k)/24$, where $p(k)$ is the number of hours in the k-th day characterized by wind blowing from the sources towards the subarea; μ_2 = mean of $\{u_2(k)\}$; α, ψ_1, ψ_2 = model parameters.

The predictor derived from model (6) is

$$\hat{y}(k) = \mu + \alpha[y(k-1) - \mu] + \psi_1[u_1(k) - \mu_1] + \psi_2[u_2(k) - \mu_2], \quad (7a)$$

$$\hat{\mathrm{DAP}}(k) = \exp \hat{y}(k), \quad (7b)$$

namely the forecast at the end of the $(k-1)$-th day requires a previous forecast of $u_1(k)$ and $u_2(k)$ which appear as inputs in Equation (7a). In the absence of a wind predictor, it is however possible to point out an upper and a lower bound of the performance of the DAP predictor (7).

The upper bound corresponds to a perfect wind predictor [$u_1(k)$ and $u_2(k)$ are equal to their actual values in Equation (7a)] while the lower bound corresponds to a wind predictor based on the persistence assumption, [namely $u_1(k) = u_1(k-1)$ and $u_2(k) = u_2(k-1)$ in Equation (7a)].

The performance of the ARX(1) predictor (7) has given an average improvement between 10% (lower bound) and 30% (upper bound) with respect to the AR(1) predictor (5).

For instance, in the case of Venice, the predictor performs as shown in Fig. 3. From the comparison with Fig. 2 a certain reduction resulting from the peak smoothing effect can be appreciated. More specifically, the values of the three forecast quality indexes are shown in the second row of Table 1.

It must be noticed that correlations of order 0.6–0.7 between observed and predicted daily DAP values seem to be the upper bound of the performance of the daily predictors (see also Finzi *et al.*, 1978).

In other terms, this limitation seems inherent to the time interval, the day, which is rather long compared with the dynamics of the phenomenon, rather than to the region under consideration and to the structure of the ARIMAX models. Recall also that a non stochastic representation such as the complex and detailed advection–diffusion daily model described by Shir and Shieh (1973) has given an average fitting performance of the order mentioned above. In cases characterized by extremely "regular" meteorology the performance has been slightly better (0.70–0.75 as shown by Finzi *et al.*, 1977, for the city of Milan). On the contrary, the slightly below average result in the case of Venice can easily be explained by the much more complex features of the pollution phenomenon (Zannetti *et al.*, 1977).

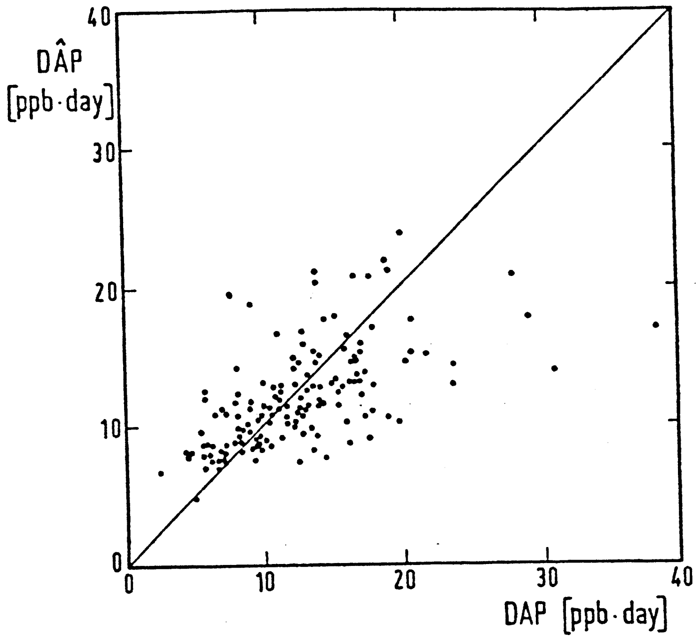

Fig. 3. Daily predicted (DÂP) vs observed DAP for the Venice subarea [ARX(1) predictor with actual wind as input].

Significant improvements have not been obtained by introducing stability categories into the ARX model. Zannetti *et al.* (1977) have already observed that the effect of the stability factor on SO_2 concentration is difficult to specify in Venice.

5. STOCHASTIC MODELS AND PREDICTORS OF HOURLY DAP

For brevity, the forecast performances of the different DAP predictors, illustrated in this section, will be shown only for Marghera (the most polluted subarea) and Venice (the least polluted subarea) since each predictor performs almost as well in all subareas.

Table 2. Performance of one hour ahead predictors for Marghera (a) and Venice (b) sub-areas

(a)

DAP predictor	Θ	S [ppb × hour]	S_h [ppb × hour]
AR(1)	0.73	21	44
ARX(1) lower bound	0.74	21	43
ARX(1) upper bound	0.74	20	43
Persistence	0.73	22	44
CSAR(1)	0.77	20	42

(b)

DAP predictor	Θ	S [ppb × hour]	S_h [ppb × hour]
AR(1)	0.78	7	20
ARX(1) lower bound	0.79	6	19
ARX(1) upper bound	0.81	6	18
Persistence	0.78	7	20
CSAR(1)	0.83	5	17

First, Equation (4) and the derived predictor (5) have been used in order to model and forecast hourly DAP. In particular, the AR(1) model (4) is again the best in the ARMA class. For 1974, the forecast performance of (5) is shown in Table 2 (first row); the performance indexes can be usefully compared with the estimated hourly DAP means and standard deviations (m = 32 ppb × hour and σ = 30 ppb × hour for Marghera, and m = 12 ppb × hour and σ = 11 ppb × hour for Venice).

In addition, the following ARX(1) model [quite similar to (6)] has been considered

$$z(k)-v = \theta(z(k-1)-v)+\psi_1(u_1(k)-v_1) + \psi_2(u_2(k)-v_2)+\varepsilon(k), \quad (8)$$

where $z(k)$ = logarithm of the k-th hour DAP in the subarea; v = mean of $\{z(k)\}$; $u_1(k)$ = logarithm of the average wind speed blowing towards the subarea during the last H (H = given number) hours (precisely, the H hours before the end of hour k); v_1 = mean of $\{u_1(k)\}$; $u_2(k)$ = logarithm of the percent of hours characterized by wind blowing towards the subarea during the last H hours; v_2 = mean of $\{u_2(k)\}$; θ, ψ_1, ψ_2 = model parameters; $\{\varepsilon(k)\}$ = noise process.

The model calibration gave the following values for H,

$$H = \begin{cases} 1 \text{ for Marghera} \\ 1 \text{ for Mestre} \\ 2 \text{ for Venice.} \end{cases}$$

The predictor derived from Equation (8) is obtained by setting the noise term to zero in Equation (8) and its performance is shown in the second and third rows of Table 2.

From the analysis of the first three rows of Table 2, one can conclude that the two predictors AR(1) and ARX(1) perform equally well, even in the case of perfect prediction of wind direction and speed. However, the quality of one hour ahead forecasts is not very significant, because of the shortness of the forecast interval. As a matter of fact, also the most trivial predictor based on the persistence assumption (next hour DAP = this hour DAP) exhibits a performance only slightly worse than the preceding ones, as shown in the fourth row of Table 2.

In conclusion, the forecast of the SO_2 concentration one hour in advance is not useful in assessing the validity of the models. On the contrary, as the forecast step increases, the performance of stochastic models becomes increasingly superior to the persistence predictor. For instance, Table 3 (first four rows) exhibits the performances of all the above mentioned predictors in the case of forecasts for four hours ahead.

Table 3. Performance of four hours ahead predictors for Marghera (a) and Venice (b) sub-areas

(a)

DAP predictor	Θ	S [ppb × hour]	S_h [ppb × hour]
AR(1)	0.29	29	65
ARX(1) lower bound	0.35	28	63
ARX(1) upper bound	0.45	27	58
Persistence	0.29	36	68
CSAR(1)	0.50	25	53

(b)

DAP predictor	Θ	S [ppb × hour]	S_h [ppb × hour]
AR(1)	0.33	10	31
ARX(1) lower bound	0.45	9	27
ARX(1) upper bound	0.58	8	24
Persistence	0.33	12	33
CSAR(1)	0.60	7	23

6. DETECTION OF EMISSION CYCLES

In the case of Marghera, the performance shown in Table 3 is not very satisfactory, even when the best ARX(1) predictor is used. Thus, an attempt has been made to ascertain whether or not the relative lack of forecast quality could be ascribed to the lack of information about hourly fluctuations of the emissions. Seasonal ARIMA models (see Box and Jenkins, 1970) have given no useful information from this view point. On the contrary, indications were given by another type of "cyclic" model, namely by CSAR(1) (Cyclo Stationary Auto Regressive of order 1, see, for instance, Thomas and Fiering, 1962):

$$\begin{aligned} z(24i+j) - v^j \\ &= \phi^j(z(24i+j-1) - v^{j-1}) + \varepsilon(24i+j) \\ &(j = 1,2,\ldots,24; \quad i = 0,1,\ldots), \end{aligned} \tag{9}$$

where $z(24i+j)$ = logarithm of the average DAP in the j-th hour of the i-th day; v^j = hourly mean of $\{z(24i+j)\}$; $\{\phi^j\}_{j=1}^{24}$ = model parameters; $\{\varepsilon(24i+j)\}$ = noise term.

Equation (9) points out that CSAR models are simply AR models with periodically varying (every 24 steps, in the present case) parameters. The one hour ahead and four hours ahead performance of the predictor derived from Equation (9) is shown in the fifth rows of Tables 2 and 3 respectively. Although no exogenous input appears in Equation (9), the forecast quality is comparable and often higher than the upper bound of ARX(1).

In general it would not be possible *a priori* to distinguish whether the good performance of the periodic model (9) is due to the presence of an emission cycle or to the existence of very regular daily cycles of the meteorological variables. Nevertheless, as already pointed out, wind direction and speed do not exhibit, in the case studied, a regular daily periodicity, so that the existence of a non-negligible emission cycle seems the more reasonable explanation of the result.

7. CONCLUDING REMARKS

Simple stochastic models and predictors of the DAP (Dosage Area Product), an overall pollution index, estimated on the Venetian Lagoon inhabited subareas, have been described in this paper.

The main conclusion is that information on wind speed and direction gives a satisfactory improvement of one day ahead forecast with respect to purely stochastic representations such as ARMA models. More precisely, the usefulness of Auto Regressive predictors with exogenous inputs (ARX models) is anyway really appreciable if a good wind forecast is available (for instance, if a reliable wind model is also developed).

As for subdaily forecast, such improvement is reduced probably because the unknown fluctuations of the emissions are relatively important. In this case, the forecast quality is improved introducing a CSAR model (Cyclo Stationary Auto Regressive Model) with periodically varying parameters.

REFERENCES

Anderson P. M. (1970) The uses and limitations of trend surface analysis in studies of urban air pollution. *Atmospheric Environment* 4, 129–147.

Bankoff S. G. and Hanzevack E. L. (1975) The adaptive-filtering transport model for prediction and control of pollutant concentration in an urban airshed. *Atmospheric Environment* **9**, 793–808.

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

Chock D. P., Terrel T. R. and Levitt S. B. (1975) Time series analysis of Riverside, California air quality data. *Atmospheric Environment* **9**, 978–989.

Desalu A. A., Gould L. A. and Schweppe F. C. (1974) Dynamic estimation of air pollution. IEEE *Trans. on Autom. Contr.* Ac-19, 904–910.

Duckworth S. and Kupchanko E. (1967) Air analysis: the standard dosage area product. *J. Air Pollut. Control Ass.* **17**, 379–383.

Finzi G., Fronza G. and Rinaldi S. (1978) Stochastic modelling and forecast of the dosage area product. *Atmospheric Environment* **12**, 831–838.

Finzi G., Fronza G., Rinaldi S. and Zannetti P. (1977) Modelling and forecast of the dosage population product in Venice. *Proc. IFAC Symp. on Environmental Systems Planning, Design and Control*, Kyoto, p. 403 (Edited by H. Akashi). Pergamon Press, Oxford.

Finzi G., Fronza G. and Spirito A. (1977) Univariate stochastic models and real time predictors of daily SO_2 pollution in Milan. *Proc. NATO-CCMS 8th Int. Technical Meeting on Air Pollution Modelling*, Louvain La Neuve, Belgium.

Gray D. M. (Ed.) (1974) Handbook on the Principles of Hydrology. Water Information Center, Inc.

Jazwinsky A. H. (1970) *Stochastic Processes and Filtering Theory.* Academic Press, New York.

Larsen R. I. (1969). A new mathematical model of air pollutant concentration, averaging time and frequency. *J. Air Pollut. Control Ass.* **19**, 24–30.

McCollister G. M. and Wilson K. R. (1975) Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutions. *Atmospheric Environment* **9**, 417–423.

Merz P. H., Painter L. J. and Ryason P. R. (1972) Aerometric data analysis – time series analysis and forecast and an atmospheric smog diagram. *Atmospheric Environment* **6**, 319–342.

Runca E., Melli P. and Zannetti P. (1976) Computation of long-term average SO_2 concentration in the Venetian area. *Appl. Math. Mod.* **1**, 9–15.

Sawaragi Y. and Ikeda S. (Eds) (1974) *Proc. Symp. on Modelling for the Control and Prediction of Air Pollution.* Kyoto, Japan.

Shir C. C. and Shieh L. J. (1973) A generalized urban air pollution model and its application to the study of SO_2 distribution in the St. Louis Metropolitan area. Technical Report RJ 1227; IBM San José Research Center, California.

Tecneco (1973) Sistema di rilevamento dell'inquinamento atmosferico nella zona di Venezia. Technical Report PROT/AGA N. 60, Tecneco, Fano (Ps), Italy.

Thomas H. A. and Fiering M. B. (1962) Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation, in *Design of Water Resources Systems* (Edited by Maass *et al.*). Harvard University Press, Cambridge.

Tiao G. C., Box G. E. P. and Hamming W. J. (1975) A statistical analysis of the Los Angeles ambient carbon monoxide data 1955–1972. *J. Air Pollut. Control Ass.* **25**, 1129–1136.

Young P. and Whitehead P. (1977) A recursive approach to time series analysis for multivariable systems. *Int. J. Control* **25**, 457–482.

Zannetti P., Melli P. and Runca E. (1977) Meteorological factors affecting SO_2 pollution levels in Venice. *Atmospheric Environment* **11**, 605–616.