MODELLING AND FORECAST OF THE DOSAGE POPULATION PRODUCT IN VENICE

GIOVANNA FINZI, GIORGIO FRONZA, SERGIO RINALDI

Istituto di Elettrotecnica ed Elettronica-Centro Teoria dei Sistemi
Politecnico di Milano (Italy)

PAOLO ZANNETTI

Centro Scientifico IBM-Venezia(Italy)

KYOTO 1977

IFAC

IFAC SYMPOSIUM ON

ENVIRONMENTAL SYSTEMS PLANNING, DESIGN AND CONTROL

MODELLING AND FORECAST OF THE DOSAGE POPULATION PRODUCT IN VENICE

GIOVANNA FINZI, GIORGIO FRONZA, SERGIO RINALDI

Istituto di Elettrotecnica ed Elettronica-Centro Teoria dei Sistemi
Politecnico di Milano (Italy)

PAOLO ZANNETTI

Centro Scientifico IBM-Venezia(Italy)

Summary.-The DPP, an index of global pollution damage, is defined and a time series of such index is derived from Venice SO_2 concentration data. A certain number of DPP stochastic models is set up on the basis of this series and the forecast efficiency of the corresponding predictors is measured by three performance indexes. In particular, the improvement due to the introduction of meteorological inputs such as wind direction is pointed out.

1. INTRODUCTION

In recent years many time series of pollutant concentrations have been described by means of stochastic models such as ARIMA or seasonal ARIMA [1-4] Such models have been used in order to supply real time forecast by means of the techniques recommended in [5]. Predictors have been introduced into air pollution literature also through a different approach [6-8].

Precisely, stochastic models derived from discretized advection-diffusion equations have been considered and the forecast of the concentrations in all the cells of the grid have been provided by Kalman filter techniques. The predictors of the latter type exhibit a higher degree of reliability, though they obviously require a heavier computational effort. Actually, the comparison in terms of forecast efficiency is not fully significant, since the required information, i.e. the predictor input, is not the same in the two cases. Precisely, in [1-4] no meteorological inputs are put into evidence neither the exploitation of the available data is complete. In fact the concentrations in the monitoring stations are considered as separate time series (in practice, as independent stochastic processes) and correspondingly the forecasts are made separately,so that all the information supplied by the cross correlations is not taken into account. An alternative approach would obviously consist of setting up a multivariate stochastic model and predictor of the vector of concentrations in all the stations. However models of this kind (the multivariate ARIMA for example) are generally complex and imply the estimation of a large amount of parameters.

In a certain number of studies a description of the pollution phenomenon is required which is less detailed than the one supplied by the previously mentioned stochastic models. This is the case of the present contribution, where SO_2 pollution over an urban region (the city of Venice) is measured by a global index, representing the average damage on the individual living in the area. Specifically, such index, called *dosage population product* or DPP takes the three following factors into account:

i) the pollution level;
ii) the duration of pollution events;
iii) the number of people exposed to pollution.

The precise definition of the index as well as its evaluation from Venice SO_2 concentration and population data is shown in Section 2. The time series thus obtained allows to set up a certain number of daily DPP stochastic models and predictors (Section 3). Finally, the forecast efficiency in the various cases is measured by performance indexes such as the standard prediction error, the correlation between predicted and observed data, the standard prediction error during episodes, so that a comparison between different models can be drawn.

2. DEFINITION AND EVALUATION OF THE DPP

This paper considers SO_2 pollution in the Venice lagoon area, due to the emissions sources, in the industrial area of Porto Marghera, in the northwest of the city. The measurement network is shown in Fig. 1. Only the summer season has been examined, so that

TABLE 1. EVALUATED DPP SERIES [ppb . day]

k=1-31	15	29	8	6	20	15	8	6	6	14	9	17	13	14	16	13	13	15	16	28	13	9	7	6	16	13	18	11	11	11	7
k=32-62	18	17	12	23	14	9	7	13	13	17	38	18	16	12	12	12	20	19	21	14	14	13	11	11	12	10	14	19	18	10	8
k=63-93	9	13	13	11	10	8	10	11	8	10	7	6	6	12	8	13	10	7	17	6	7	17	8	7	10	16	12	17	19	11	14
k=94-124	15	21	24	22	17	6	6	10	10	5	7	7	10	14	20	15	16	14	15	14	10	13	9	9	8	8	3	5	4	6	5
k=125-153	6	8	11	6	10	11	12	11	13	7	10	17	17	16	13	11	14	14	9	14	18	8	10	6	31	16	18	17	17	-	-

time series analysis [5] has been performed. The process has been considered as stationary, because the variable under consideration was the daily DPP in summer, so that neither the daily nor the seasonal cycles, due to astronomical and emission factors, had to be taken into account. Moreover, the emission data did not exhibit a weekly cycle.

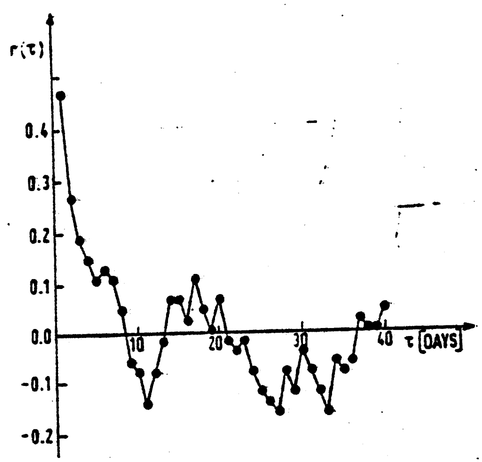

FIG. 2. Sample autocorrelation function of the DPP process.

First, the distribution of the stationary process has been determined; according to the Kolmogorov-Smirnov test, it turned out to be *lognormal* at more than 20% significance level. Furthermore, in view of the form of the sample autocorrelation function (see, Fig. 2), the normal process of DPP logarithms has been assumed to be *autoregressive-moving average* (ARMA). More precisely, AR(1), AR(2) and ARMA (1,1) models have been tested and the AR(1) model turned out to be as satisfactory as the others. The one day ahead predicted DPP is shown in Fig.3 versus the observed one while the quality of the forecast has been measured by the three following indexes.

a) The standard *forecast error*

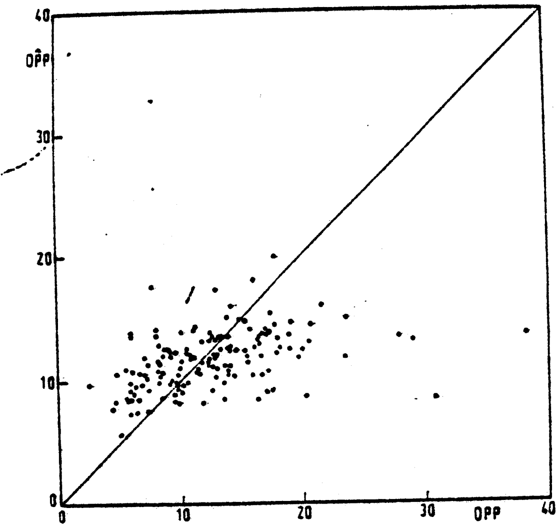

FIG. 3. One day ahead predicted DPP versus observed DPP (Model I)

$$S = \left\{ \frac{\sum_k |\widehat{DPP}(k) - \widetilde{DPP}(k)|^2}{153} \right\}^{\frac{1}{2}}.$$

where $\widehat{DPP}(k)$ is the forecast supplied by the AR(1) predictor.

b) The *cross correlation* between predicted and observed DPP

$$R = \frac{\sum_k (\widehat{DPP}(k) - \hat{\mu})(\widetilde{DPP}(k) - \tilde{\mu})}{153\, \tilde{\sigma}\, \hat{\sigma}}$$

where $(\tilde{\mu}, \tilde{\sigma})$ and $(\hat{\mu}, \hat{\sigma})$ denote the estimated mean and standard deviation of the DPP and the predicted DPP, respectively.

c) The standard *forecast error during episodes*

$$S_e = \left\{ \frac{\sum_{k \in K_e} [\widehat{DPP}(k) - \widetilde{DPP}(k)]^2}{17} \right\}^{\frac{1}{2}}$$

where K_e is the set of the 17"episode days", defined as the days characterized by $DPP(k) \geq \tilde{\mu} + \tilde{\sigma}$.

The values of S, R and S_e obtained in correspondence with the AR(1) predictor are reported in Table 2 (first row)

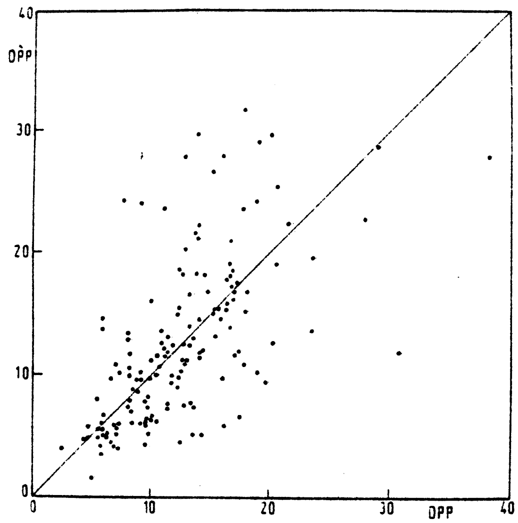

FIG. 4. One day ahead predicted DPP versus observed DPP (Model III)

wind direction, and better results might have been supplied if reliable atmospheric stability data had been available. The relatively low efficiency of the predictors can be mainly ascribed to the time unit selected. In fact the day is a rather long interval if compared with the dynamics of an extremely variant pollution phenomenon, characterized by sea-land-lagoon interactions [9]. However, it wouldn't have been reasonable to consider subdaily intervals when defining a global measure of pollution damages on the population since data on morbidity or hospital recoveries are in general available on a daily basis.

REFERENCES

[1] P.H. MERZ, L.Y. PAINTER and P.R. RYASON: Aerometric data analysis. Time series analysis and forecast and an atmospheric smog diagram. *Atmosph. Envir.* 6, 319-342 (1972).

[2] G.M. McCOLLISTER and K.R. WILSON: Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmosph. Envir.* 9, 417-423 (1975).

[3] D.P. CHOCK, T.R. TERRELL and S.B. LEVITT: Time-series analysis of Riverside, California air quality data. *Atmosph. Envir.* 9, 978-989 (1975).

[4] G.C. TIAO, G.E.P. BOX and HAMMING: A statistical analysis of monoxide data 1955-1972. JACPA 25, 1129-36, (1975).

[5] G.E.P. BOX and G.M. JENKINS : *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco (1970).

[6] A.A. DESALU, L.A. GOULD and F.C. SCHWEPPE: Dynamic estimation of air pollution. IEEE *Trans. on Autom. Contr.* AC-19, 904-910 (1974).

[7] Y. SAWARAGI and S. IKEDA (eds.): Proceedings of the Symposium on: *Modelling for the Control and Prediction of Air Pollution.* Kyoto (Japan) (1974).

[8] S.G. BANKOFF and E.L. HANZEVACK: The adaptive-filtering transport model for prediction and control of pollutant concentration in an urban airshed. *Atmosph. Envir.*, 9, 793-808 (1975).

[9] P. ZANNETTI, P. MELLI and E. RUNCA: Meteorological factors affecting SO_2-pollution level in Venice, *Atmosph. Envir.* (in press).