TIME SERIES ANALYSIS OF VENICE AIR QUALITY DATA

Paolo Zannetti - IBM Scientific Center, Dorsoduro 3228, Venice, Italy

and

Giovanna Finzi, Giorgio Fronza, Sergio Rinaldi - Istituto di Elettrotecnica ed Elettronica, Centro Teoria dei Sistemi CNR, Politecnico di Milano, Milan Italy

KYOTO 1977

IFAC

IFAC SYMPOSIUM ON

ENVIRONMENTAL SYSTEMS PLANNING, DESIGN AND CONTROL

have been evaluated (K_j = number of days of the j-th season) in the period (February 1973 February 1975) corresponding to the recorded data. The corresponding parameters $^*m^l_{i,j}$ and $^*\sigma^l_{i,j}$ of the normal process

$\left\{ {}^*x^l_{i,j}(k) \right\}_k$, where $^*x^l_{i,j}(k) = \ln x^l_{i,j}(k)$,

have been computed. A typical behaviour of the above defined parameters versus the hour index i is shown in Fig.3. The daily cycle of the standard deviation is clearly removed for the process $\left\{ {}^*x^l_{i,j}(k) \right\}_k$ for all seasons and all stations.

The correlation structure of the process has been characterized by evaluating the autocorrelation functions

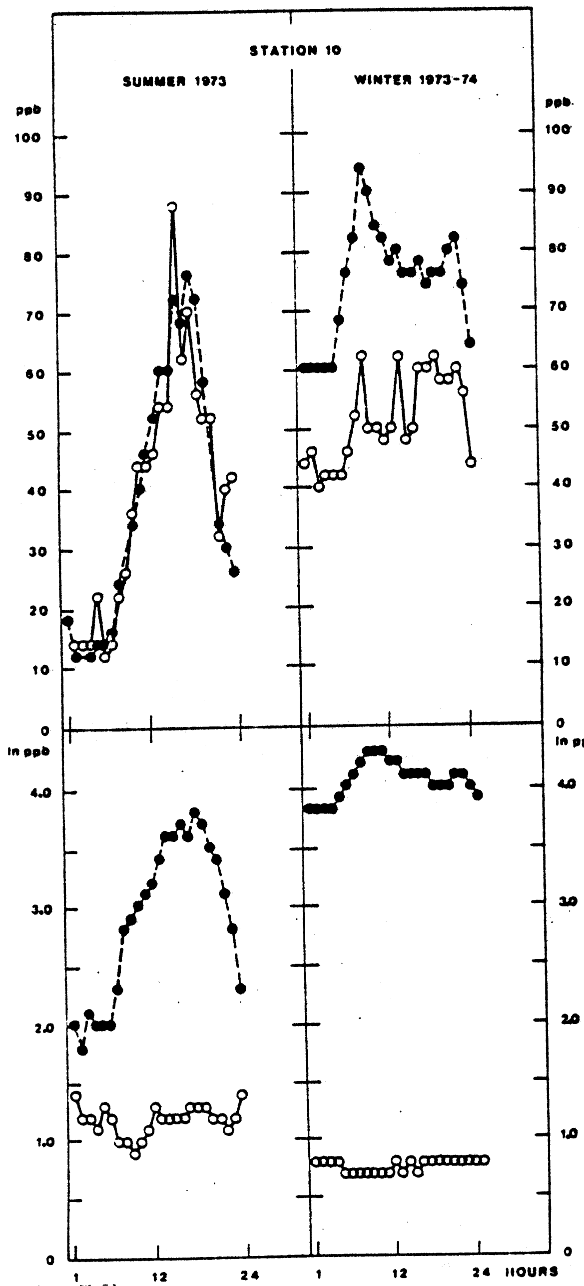

FIG.3. Daily cycle of the seasonal average (dots) and the standard deviation (circles) for the SO2 data recorded at Station 10. Upper side refers to data expressed in ppb, and the lower side to the logarithm of the data. Left side refers to the summer 1973 and the right side to the winter 1973-74.

$$\varrho^l_{i,j}(\tau) \cong \frac{\sum_{k=1}^{K'_{i,j}(\tau)} \left[x^l_{i,j}(k)-m^l_{i,j}\right]\left[x^l_{i+\tau,j}(k)-m^l_{i+\tau,j}\right]}{\sigma^l_{i,j}\,\sigma^l_{i+\tau,j}\,K_j}$$

where $K'_{i,j}(\tau) = K_j$ in the case when $i+\tau \leq 24$, and where $K'_{i,j}(\tau) = K'_{i,j}(\tau-24) - 1$, $x^l_{i+\tau,j}(k) = x^l_{i+\tau-24,j}(k+1)$, $m^l_{i+\tau,j} = m^l_{i+\tau-24,j}$ and $\sigma^l_{i+\tau,j} = \sigma^l_{i+\tau-24,j}$ in the case when $i + \tau > 24$. The lag-one correlations $\varrho^l_{i,j}(1)$ are particularly significant for describing the effect of the daily cycle of the physical causes on the entire correlation structure. In summer there are stronger variations of $\varrho^l_{i,j}(1)$ versus i, generally within the range 0.5-0.9 (0.7-0.9 for the process of the logarithms), while in winter the lag-one index has less pronounced variations within the higher range 0.7-0.9 (0.7-1 for the logarithms) because of the "background" in the pollution level due to domestic heating.

In conclusion, the existence of a daily cycle affecting both the distributions and the correlation structure of the seasonal hourly SO_2 concentration process cannot be neglected. Moreover it would seem better to apply to the process of the logarithms of the concentrations, which exhibits stronger correlations and no daily variation of the standard deviation.

Finally, the screening study has been concluded by analysing the meteorological records and pointing out their relations to SO_2 data. Precisely, auto-spectra of concentrations, wind speed, temperature and pressure (Fig.4) as well as the amplitude, phase and coherence parameters of the corresponding cross-spectra have been evaluated. An objective of such spectral analysis has been to ascertain the existence of concentration cycles to be ascribed to the synoptic weather periodicities of 3.5 days. Such cycles have been pointed out in the contributions [13,14] concerning situations in North America. However, the existence of the 3.5 days cycles has not been proved for the Venetian area, probably because of the different meteorology and particularly because of the reduced influence of wind speed, with respect to wind direction, on the pollution phenomenon. In conclusion, the cross spectral analysis does not increase significantly the general information on the phenomenon. Concentration series versus meteorological series show higher level of coherence in correspondence with diurnal and semi-diurnal oscillation, but there is no definite oscillation of period greater than one day. However, almost all the stations show a certain increase of coherence, between SO_2 and meteorological data, in correspondence with periods in the range 2.5-3.5 days.

Table 1. Values of σ_ε^2 for different autoregressive (AR) and/or moving average (MA) models during the summer 1973

Station number	Variance σ_x^2 of the data	σ_ε^2 AR(1)	σ_ε^2 AR(2)	σ_ε^2 MA(1)	σ_ε^2 MA(2)	σ_ε^2 ARMA(1,1)	σ_ε^2 Seasonal model
2	2026	1029	1029	1033	-	1029	1186
10	2378	1254	1251	--	-	1251	1886

With respect to the variance of the noise, the cyclostationary model exhibits a substantial improvement if compared with the AR(1) (Table 2).

Table 2. Values of σ_ε^2 for the AR(1) and AR(1)CS models in the summer 1974 and winter 1974-75

Period	Station number	Variance σ_x^2 of the data	AR (1) σ_ε^2	AR(1)CS σ_ε^2
Summer 1974	2	1231	537	477
	10	1451	658	581
	16	292	133	121
	29	1071	617	542
Winter 1974-75	2	2583	870	779
	10	2413	650	580
	16	4023	905	792
	29	5610	1806	1659

The process of the logarithms of the hourly concentrations (which may be regarded as normal) has been considered. As a matter of fact, the above models (AR(1),ARIMA,seasonal ARIMA and AR(1)CS) have been tested also on such process and again the AR(1)CS proved to be the most satisfactory.

3.2 Real time forecast

The parameters of the above models have been estimated by means of concentration data in 1973, while the predictors corresponding to such models have been tested in correspondence with the data in 1974. There is a certain variation between the two years, characterized by a rather different meteorology, and this explain why the AR(1) predictor proved to be more efficient than the AR(1)CS (see Table 3, columns 4 and 6, where the variance of the prediction error is reported). In fact, the latter predictor has many more parameters and therefore "keeps more memory" of the data used for parameters estimation. Result suggests the application of adaptive AR(1) and AR(1)CS models,namely those characterized by parameters estimated on a learning past period of given length.

Table 3. Values of σ_ε^2 of AR(1) and AR(1)CS using last year parameters and using adaptive parameters

Period	Station number	Variance σ_x^2 of the data	AR(1) σ_ε^2 model: last year	AR(1) σ_ε^2 model: Adaptive	AR(1)CS σ_ε^2 model: last year	AR(1)CS σ_ε^2 model: Adaptive
Summer 1974	2	1231	541	545(13)	542	615(20)
	10	1451	662	695(5)	674	750(27)
	16	292	143	127(5)	180	146(30)
	29	1071	656	560(13)	681	640(25)
Winter 1974-75	2	2583	887	920(7)	927	1320(10)
	10	2413	650	650(5)	664	720(23)
	16	4023	915	775(31)	908	710(40)
	29	5610	1814	1630(25)	1807	1490(40)

The performances of the corresponding predictors are reported in Table 3, columns 5 and 7, where the parentheses denote the optimal length of the learning period. Furthermore,the adaptive models can reasonably be tested to represent the entire process, without distinguishing from season to season (at least for sufficiently short learning periods). The forecast performance in this case has been shown in Fig.5 in correspondence with various lengths of the learning period. The minima represent the optimal tradeoff between the need of having a learning period long enough for a reliable parameter estimation and the requirement of "forgetting" data corresponding to a completely different meteorological situation.

All the above mentioned predictors have been applied also to the process of the logarithms, and the forecast data have subsequently been antitransformed in order to obtain the prediction of the variable of interest. However, in terms of such variables, the prediction performance is worse, showing a first clear indication of disadvantages in the logarithm transformation in analysing and modelling air quality data. In summary, the results indicate the general validity both the followed approach, and the stochastic models proposed.