Daily air pollution time series analysis of Isfahan City

[1]R. Modarres and [2*]A. Khosravi Dehkordi

[1]*Faculty of Natural Resources, Isfahan University of Technology, Isfahan, Iran*

[2]*Department of Environmental Sciences, Graduate College of Natural Resources, Isfahan University of Technology, Isfahan, Iran*

Abstract

Different time series analysis of daily air pollution of Isfahan city were performed in this study. Descriptive analysis showed different long-term variation of daily air pollution. High persistence in daily air pollution time series were identified using autocorrelation function except for SO_2 which seemed to be short memory. Standardized air pollution index (SAPI) time series were also calculated to compare fluctuation of different time series with different levels. SAPI time series indicated that NO and NO_2, CH_4 and non-CH_4 have similar time fluctuations. The effects of weather condition and vehicle accumulation in Isfahan city in cold and warm seasons are also distinguished in SAPI plots.

Key words: *Air pollution, time series, ACF, non linear dynamic, SAPI, Isfahan*

Corresponding Author, E-mail:adhkordi@cc.iut.ac.ir

Introduction

The limitless of air pollution range and sources have forced pollution managers to apply new methods in air pollution control and monitoring. Development and use of statistical and other quantitative methods in the environmental sciences have been a major communication between environmental scientists and statisticians (Herzberg and Frew, 2003). This approach is called top-down approach which starts with statistical analysis of collected air pollution data (Lee, 2002). In recent years many statistical analysis have been used to study air pollution as a common problem in urban areas. The common descriptive statistical approach used for air quality measurement and modeling is rather limited as a method to understand the behavior and variability of air quality. Different techniques have been used for air quality monitoring systems. Voigt *et al.*, 2004 applied principle component analysis in order to evaluate different air pollutants like ozone (O_3), nitrogen dioxide (NO_2) and carbon monoxide (CO) in 15 European member states. Many investigators have used probability models to explain temporal distribution of air pollutants (Bencala and Seinfeld, 1979, Yee and Chen, 1997). Time series analysis is a useful tool for better understanding of cause and effect relationship in environmental pollution (Schwartz and Marcus, 1990, Salcedo *et al.*, 1999, Kyriakidis and Journel, 2001). The principle aim of time series analysis is to describe the history of movement of a particular variable in time. Many authors have tried to detect changing behavior of

air pollution through time using different techniques (Salcedo *et al.*, 1999, Hies *et al.*, 2000, Kocak *et al.*, 2000, among others). Many others have tried to relate air pollution to human health through time series analysis (Gouveia and Fletcher, 2000, Roberts, 2003, Touloumi *et al.*, 2004). The object of this study is to examine daily time series analysis of some air pollutants in Isfahan City, in the center of Iran. The average daily air pollution concentrations (APC) of SO_2, CO, CH_4, NO_2, NO, non-CH_4 and O_3 were selected from March 2003 to March 2004.

Material and Methods
Descriptive analysis

The classical descriptive analysis is the first statistical analysis dealing with any data. Mean, standard deviation (STDEV), maximum and minimum value of selected data sets are usually calculated to have preliminary knowledge of selected variables. The calculation of coefficient of variation (cv) helps the investigator to overcome the problem of different levels and units of variables in order to compare them. Coefficient of Skewness (cs) and Kurtosis (ck) are other measures which may be used to characterize the symmetry and flatness of the probability density function of a time series, respectively (Windsor and Toumi, 2001). Because of high order, kurtosis is particularly sensitive to extremes or intermittent fluctuations and, therefore, a useful indicator of intermittency. Highly intermittent time series have a higher kurtosis. However,

descriptive analysis is of rather limited value due to the large variability associated with air quality data through time (Salcedo, *et al.*, 1999).

Time series analysis

A time series is a set of observations that are arranged chronologically. In time series analysis, the order of occurrence of the observation is crucial. If the chronological ordering of data were ignored, much of the information contained in time series would be lost. A variety of different important terminologies in time series analysis are existence such as stationarity, periodicities and trend which fall into temporal categories of air pollutant concentration (Klemm and Lange, 1999, Lee, *et al.*, 2003). Stationarity of a process can be qualitatively interpreted as a form of statistical equilibrium. Therefore, the statistical properties of the process are not a function of time. For interpretation purposes, it is often useful to plot Autocorrelation function (ACF) against lag time, K. ACF is a simple graphical method to find time relationship of an event. The sample autocorrelation coefficient is written as (Box and Jenkins,1976):

$$r_k = \frac{\sum_{t=1}^{N-k}(X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{N}(X_t - \overline{X})^2} \qquad (1)$$

k=0,1, …,N

Where N is the total length of record, K is lag time, X_t is observation at time t and $\overline{X}$ is the mean of series. All r_k lie between -1 and +1 and values significantly above zero denote correlation at the given lag. ACF can also tell us whether the observation depends on time or not. When ACF decays rapidly to zero after a few lags, it may be an indication of stationarity in the series, while a slow decay of ACF may be the indication of nonstationarity (Salas, 1993, Hipel and McLeod, 1994). Strict stationarity means that there is no systematic change in mean (no trend), no systematic change in variance, and strictly periodic variations have been removed. Quantitatively, this means that the joint distribution of X(t1), …, X(tn) is the same as the joint distribution of X(t1+t),….,X(tn+t), for all t1,…tn. This means that shifting the time origin by t has no effect on the joint distributions, which only depend on the time intervals between t1,t2,…,tn. Second-order stationarity for weak stationarity. In other words, a finite memory of the series leads to a gradual decline of the envelope of the ACF (Klemm

and Lange, 1999). Another way to investigate whether a series is time dependent or not is time series regression (Bowerman and O'Connell, 1993). The polynomial time regression between dependent variable, yt and time is written as follows:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_p t^p + \varepsilon_t \qquad (2)$$

Where β_0 to β_p are the parameters of regression equation and least square point estimates of them may be obtained by using regression techniques. For statistical inference on the significant of regression and parameters, the reader is referred to Bowerman and O'connel (1993).

Standardized time series

The above analysis will show us whether air pollutant are dependent on time or not but in air pollution time series analysis, it would be useful to find time periods of risky air pollution levels. In order to compare different air pollutants with different levels and units, we use standardized air pollution index which is written as follows:

$$SAPI = \frac{P_i - \overline{P}}{\sigma} \qquad (3)$$

Where P_i is the pollutant concentration at time i, $\overline{P}$ and σ are the mean and standard deviation of the series and SAPI is the Standardized Air Pollution Index. Standardized air pollution index is not only useful to determine risky periods of air quality characteristics but to define the risky periods as well. It is also possible to determine air pollution interaction through time using cumulative SAPI. The ASAPI will disclose the cumulative risky periods of air quality and is useful in air health monitoring.

Results
Descriptive analysis

The descriptive statistics of selected daily air pollutions are presented in Table 1. As the coefficient of variation (cv) is a measure of variation over time (Lee, 2002), the comparison of pollutants indicates that NO has the highest variation over time while CH_4 has the smallest variation. The degree of variation decrease in the order: $NO<NO_2<SO_2<TSP<CO<non\text{-}CH_4<O_3<CH_4$. This variation may be the result of variation in generating resources or weather condition. The coefficient of skewness (cs) measures the relative skewness of frequency distribution; as time series, air pollution concentration data are characterized buy strongly right-skewed frequency distribution in this study like other previous studies by Georgopoulos and Seinfeld

(1982) and Lee (2002). The degree of right-skewness decrease in order: $NO < NO_2 < SO_2 < non-CH_4 < TSP < CO < CH_4 < O_3$.

The higher ck of most of the pollutants indicates intermittency in most of air pollutants, except for CH_4, CO and O_3. Non-CH_4 is perhaps the nearest air pollutants to Normal distribution with cs=1 and ck=3. The degree of kurtosis decrease in order: $NO < NO_2 < SO_2 < non-CH_4 < TSP < CH_4 < CO < O_3$.

pollutants like NO_2 and NO, CH_4 and non-CH_4 are appealing. The positive correlation indicates synchronous time fluctuations of air pollutants. The different time correlation behavior of the pollutants is further discussed in section 3.3.

Time series results

The first step in time series analysis is to draw time series plot. Time series plot can give a

Table 1: Descriptive statistics of selected air pollutants of Isfahan city in 2003

Pollutant	SO$_2$ (ppb)	O$_3$ (ppb)	TSP (μg/m^3)	non-CH$_4$ (ppm)	CH$_4$ (ppm)	CO (ppm)	NO (ppb)	NO$_2$ (ppb)
Sample size (day)	350	349	337	323	323	344	351	351
Max.	39.7	87.5	137.4	0.42	6.2	1.7	27.5	75.3
Mean	10.5	48.4	44.6	0.15	3.9	0.78	4	17.6
Min	0.41	3.46	12.4	0.05	2.30	0.23	0.69	3.15
STDEV	5.6	14.1	18.7	0.05	0.7	0.3	2.8	10.7
cv	53.2	29.1	42	33.3	18.7	37.2	70.7	61.2
cs	1.47	0.03	0.76	0.96	0.27	0.32	2.87	1.99
ck	4.24	-0.66	1.6	3.36	-0.09	-0.37	16.4	5.6
Significant autocorrelations (lags)	1-6	1-72	1-48	1-45	1-48	1-41	1-17	1-42

Perhaps the most interesting result of descriptive analysis is small value of cs and ck of O_3. It is well known that high O_3 concentration is produced only when sunlight is strong (almost during summer in Isfahan city). Therefore it is assumed that the difference between low and high concentration is large, resulting in the largest skewness and variation in all examined air pollutants (Lee, 2002). The result of this study is exactly in the opposite, where O_3 has the minimum values of skewness and variation. The correlation matrix (Table 2) also indicates the time correlation between different pollutants. High, positive correlations between chemically-similar

preliminary understating of the time behavior of the series. Fig.1. shows time series plot of selected time series air pollution concentration. This Figure shows different time behavior of air pollutants. For example, the concentration of O_3 and TSP seem to have a similar trend from the beginning of the year to the end but the maximum and minimum concentrations occur in different time. It is also obvious that SO_2 and NO have not a significant trend through time. The fluctuations of NO_2, SO_2 are more irregular at the end of the year but the fluctuation of non-CH_4 is more obvious at the beginning of the year.

Table 2: Correlation matrix of selected pollutants

Pollutants	SO$_2$	NO$_2$	NO	CO	O3	CH$_4$	nonCH$_4$	TSP
SO$_2$	1							
NO$_2$	0.30	1						
NO	0.12**	0.65*	1					
CO	-0.02	0.46*	0.45*	1				
O3	-0.17*	-0.48*	-0.20*	-0.04				
CH$_4$	-0.08	-0.16	0.03	-0.01	0.52**	1		
NonCH$_4$	-0.06	-0.31*	-0.17*	-0.23*	0.42**	0.43*	1	
TSP	0.01	-0.20	0.01	-0.01	0.28*	0.20*	0.16*	1

(*: Significant at 5%, **: Significant at 1%)

The autocorrelation functions of the selected time series also show different time stationarity of the series. The autocorrelation functions of them are presented in Fig. 2. The significant lags for all selected pollutants are also presented in Table 1. Except for SO_2, all other pollutants show nonstationarity (long serial correlation). The amplitude of autocorrelation functions do not become less pronounced for most of the series, except for SO_2. This is an indication that the memory of SO_2 is finite. Time series regression was applied to the selected series. The significant time regression was selected based on R-Square and F-statistics. Statistical inference for regression parameters was done based on null hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_0 = 0$ and all parameters were significant at $\alpha = 95\%$. The parameters of regression models are presented in Table 3 while they are graphically presented in Fig. 1. Except for SO_2 and NO, it is clear that most of the air pollutions follow a non linear trend through time. The non linear time dynamic of air pollution is probably the result of non linear behavior of pollution generating mechanism or weather fluctuations.

This non linearity behavior was also identified by other investigators such as Kocak et al., 2000, Salcedo et al., 1999 and Hies et al., 2000. The linearity of NO and SO_2 also matches stationary condition of the series indicated in Table 1. In other words, the mean concentration of SO_2 and NO does not vary through time.

SAPI time series

For analyzing air quality data, it is very important to find the periods with adverse effect on public health (Gouvvia and Fletcher, 2000, McKee et al., 1993), even at historically low level of air pollution (Touloumi et al., 2004). In order to find these periods, standardized air pollution index was calculated and presented in Fig. 3. The figure shows that the fluctuation of air pollution differs through time. For example, NO_2 does not differ from zero for 219 (spring and summer) days but the concentration is significant after that. In other words, the adverse level of NO_2 appears mostly at the end of the year. This may occur due to increase in gas-fired home heating systems in winter. In contrast, O_3 has negative values of SAPI at the end of the year, approximately after day 219. This is the result of inverse correlation between NO_2 and O_3 (Table 2) which indicates the interaction of O_3 as a secondary pollutants and NO_2 as the primary pollutants. Because O_3 is an oxidizing chemical and its production requires the presence of sunlight, high levels of O_3 occurs in summer (Fig. 1). SO_2 is again attractive air pollution as the fluctuation of SO_2 around zero is approximately symmetric and intermittent. SO_2 highest values of SAPI's are observed between 270 to 290 days.

The main source of SO_2 emission in Isfahan city is intermittent road transport of diesel vehicles and large power stations and industrial process around the city. Other air pollutants indicate different periods of high and low SAPI. The comparison of different SAPI shows the inverse time behavior characteristics of NO_2 and non-CH_4. Whole NO_2 has negative value of SAPI at the beginning of the year to almost 220[th.] day, the negative values of Non-CH_4 begins approximately from 240[th.] day. The time behavior of two hydrocarbon pollutants, CH_4 and non-CH_4, seems to be mostly similar to each other, except for the beginning of the year when CH_4 begins to increase earlier than non-CH_4.

Table 3: The properties of Regression models of the selected air pollution time series

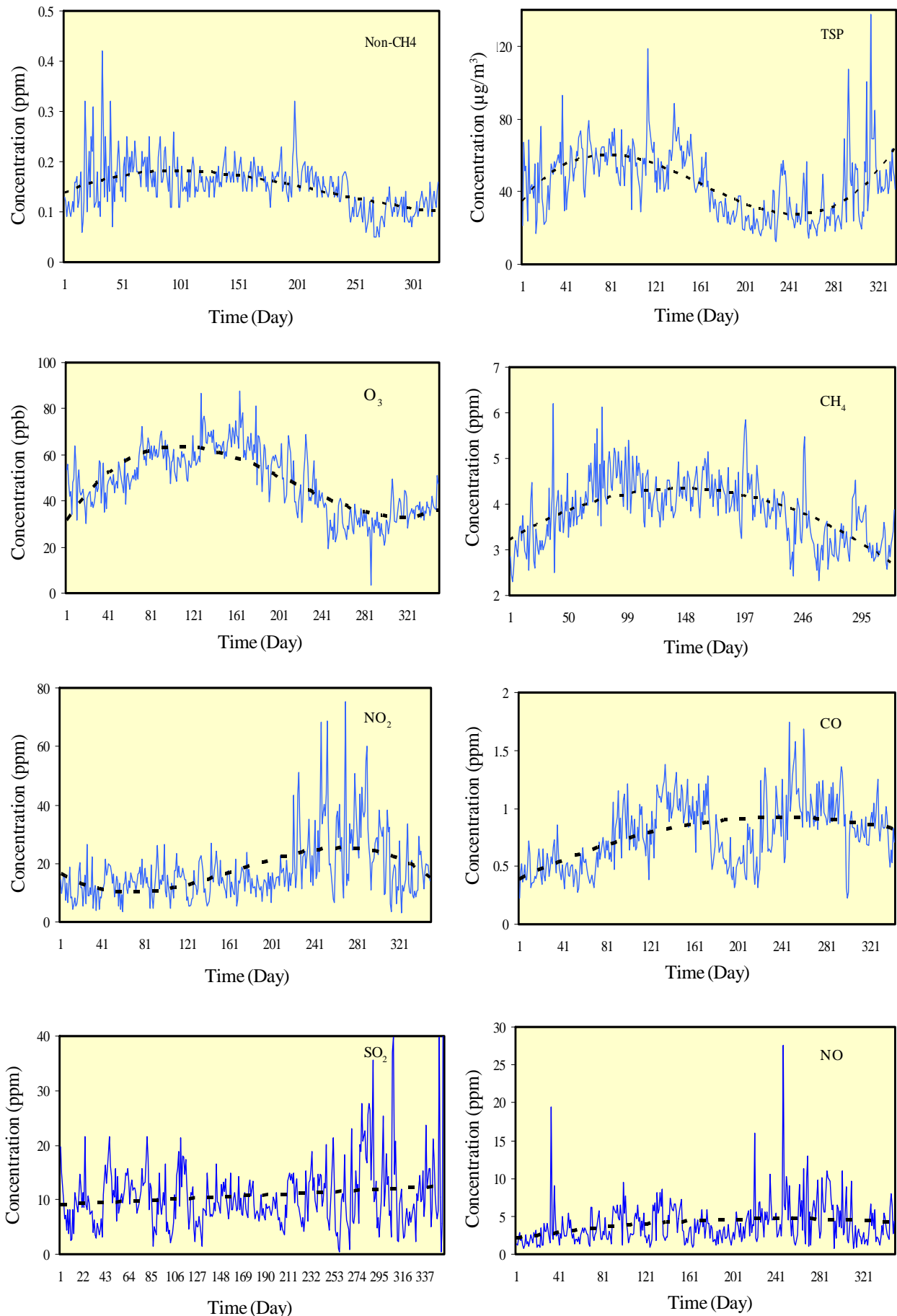| Parameters / Pollutions	β_0	β_1	β_2	β_3	F	R^2
NO_2	16.53	-0.21	0.002	4×10^{-6}	39.31	0.46
NO	1.22	0.2	-	-	38.04	0.53
CO	0.35	0.005	2×10^{-5}	1.4×10^{-8}	42.93	0.58
CH_4	2.69	0.035	-2×10^{-4}	3×10^{-7}	92.95	0.47
Non-CH_4	0.14	0.001	-7×10^{-6}	1×10^{8}	41.88	0.68
TSP	33.81	0.77	-0.006	1.3×10^{-5}	67.94	0.68
SO_2	9.2	.007	-	-	6.25	0.52
O_3	30.86	0.68	-0.004	6×10^{-6}	193.25	0.62

Fig. 1: Time series plots of selected air pollutions (solid line) and fitted regression curves (dashed lines)
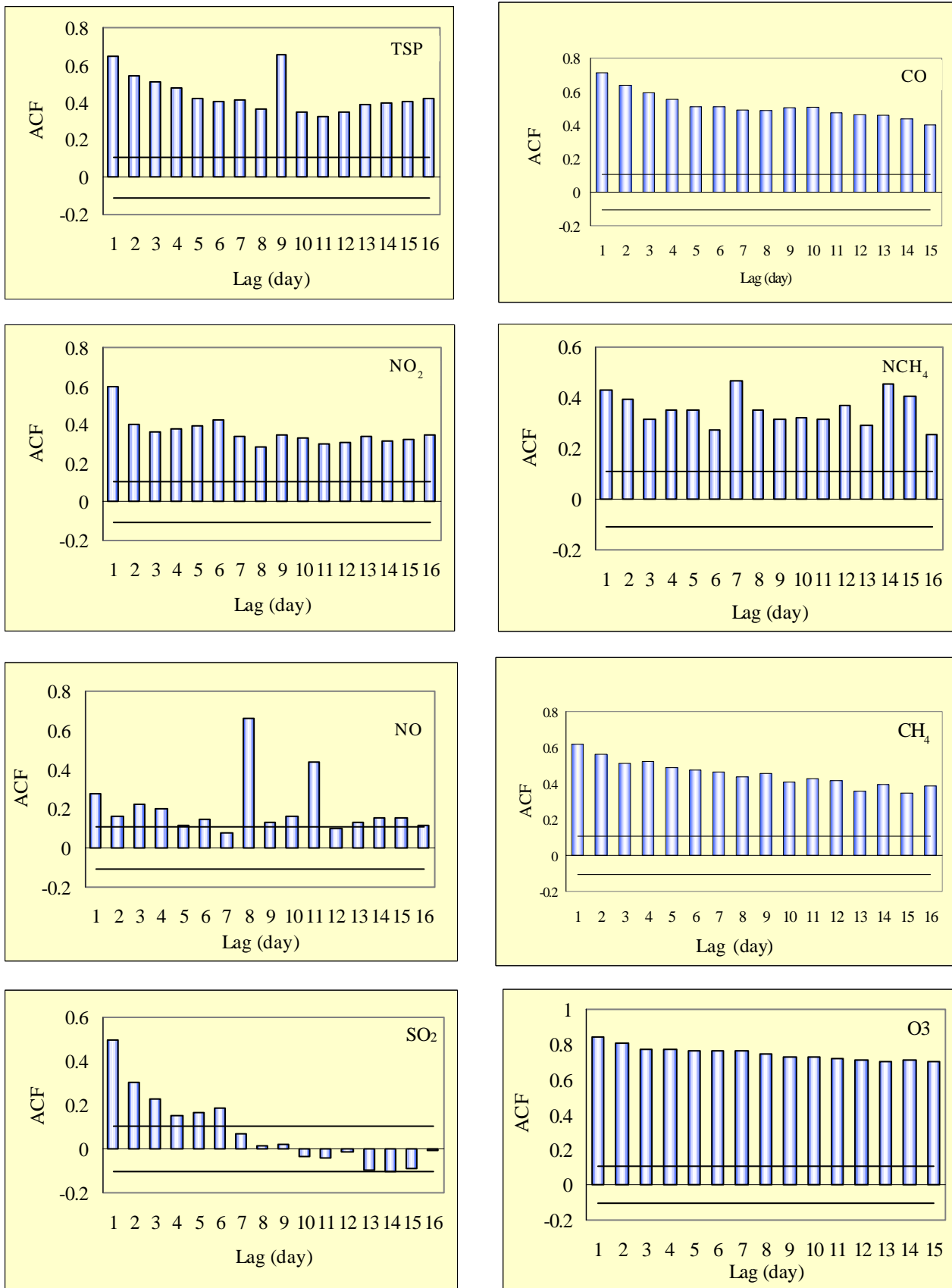
Fig. 2: Autocorrelation Function of selected pollutants, showing different stationarity behaviors
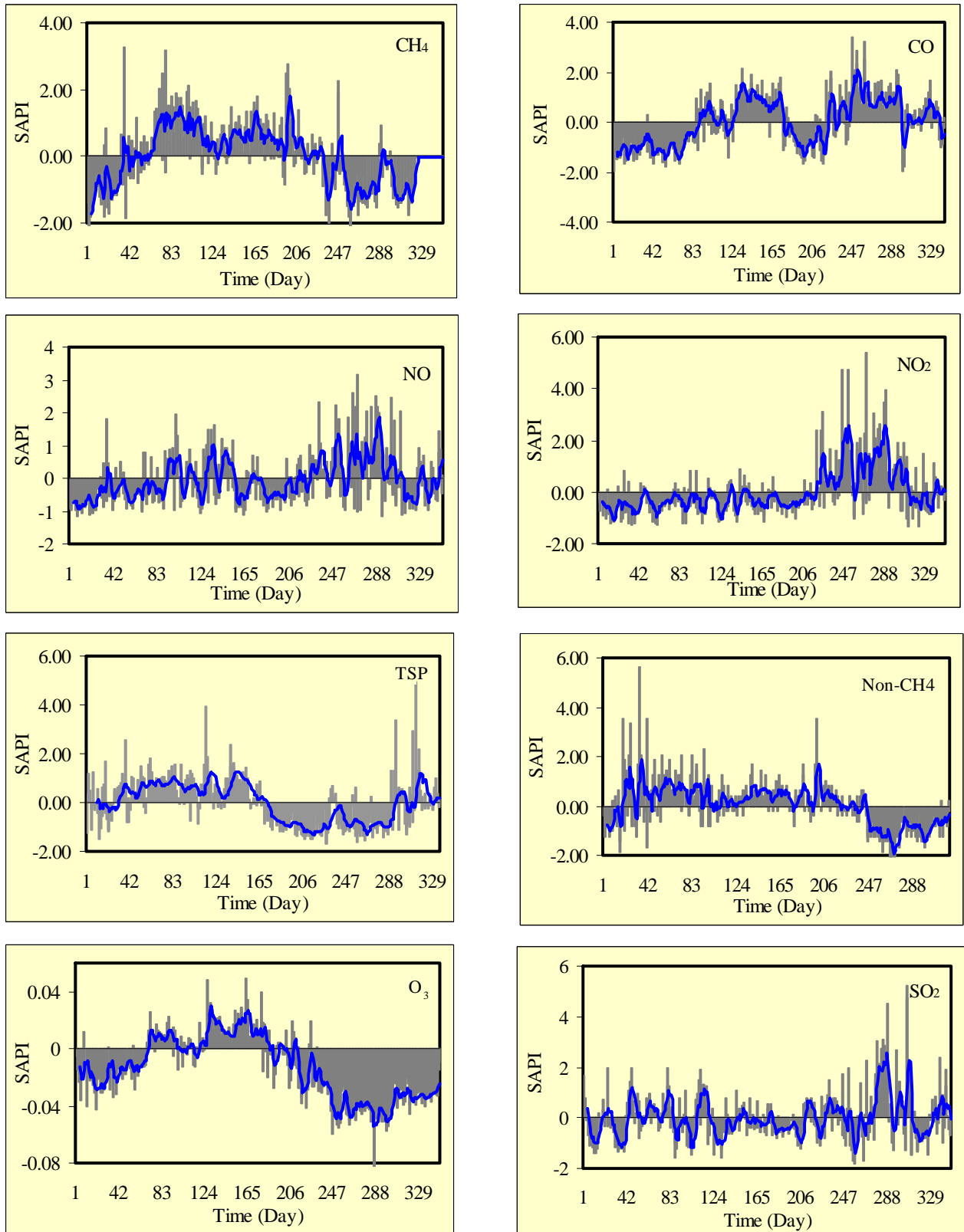
Fig. 3: Time series plot of standardized air pollution index (bars) and smoothed moving average line (solid line)

The increasing of hydrocarbon and carbon monoxide air pollutants in summer is mostly due to high vehicle concentration in Isfahan city as one of the main tourist-attracting city of Iran. Another important hydrocarbon resource is petrochemical industry in the north west of the city. The role of hydrocarbons in O_3 formation can be seen in Table 2 and Fig. 3. The higher values of CO in winter are also the result of increase in home heating systems with fossil fuels. The other two related pollutants, NO and NO_2 also show the same behavior. Although NO_2 has negative values (low risky values) for the 220 day of the year but NO has some positive (high risky values) values between 90[th.] and 165[th.] days and then between 220[th.] and 300[th.] days. The effect of wind blowing in fall and winter is the main reason of TSP decreasing in cold seasons of Isfahan.

Discussion and Conclusion

Daily air pollution time series analysis of Isfahan city was performed in this study and showed different temporal behavior of different air pollutants. While some pollutants like NO_2 and SO_2 show simple temporal fluctuation through the year, other pollutants show high fluctuation and have mostly non linear trend through the year using time series regression. High coefficient of variation and kurtosis in most of the observed series also indicated non linearity variation of air pollutants concentration through time. Standardized time series analysis which let us compare different pollutants with different levels and units, indicates different health adverse periods from the beginning of the year to the end. This different time behavior is not only the reason of correlation of different pollutants with each other but the seasonal variation on increasing or decreasing air pollutants as well. It was also shown that most daily air pollution time series have high persistence of air pollution conditions through time. This persistence is not only dangerous for public health but also makes air pollution management and control very difficult, except for NO_2 and SO_2. The influence of weather conditions like rainfall, air moisture and wind velocity-direction on air pollution temporal dynamics is also very important in air pollution management and control which will be ongoing author's task to investigate.

References

Bencala K. E. and Seinfeld J. H., (1979). On frequency distribution of air pollutant concentrations. Atmos. Environ., *10*, 941-950.

Bowerman B. L. and O'Connel R. T., (1993). Forecasting and Time Series, an Applied Approach, Duxbury, Pasific Grove, 726.

Box G. E. P. and Jenkins G. M., (1976). Time Series Analysis, Forecasting and Control. Revised Edition, Holden – Day, San Francisco, California, 575.

Georgopoulos P. G. and Seinfeld J. H., (1982). Statisticall distribution of air pollutant concentrations, Environ. Sci. Technol., 401A-416A.

Gouviea N. and Fletcher T., (2000). Time series analysis of air pollution and mortality: Effects by cause, age and socioeconomic status. J. Epidemiol. Commun. H., *54*, 750-755.

Herzberg A. M. and Frew L., (2003). Can public policy be influenced? Environmetrics, *14*, 1-10.

Hies T., Treffeisen R., Sebald L. and Reimer E., (2003). Spectral analysis of air pollutants. Part 1: elemental carbon time series. Atmospheric Environment, *34*, 3495-3502.

Hipel K. W. and McLeod A. E., (1994). Time series modeling of water resources and environmental systems, Elsevier, Amsterdam, 1013.

Klemm O. and Lange H., (1999). Trends of air pollution in the Fichtelgebrige Mountains, Bavaria. Environ. Sci. &Pollut. Res, *6*, 193-199.

Kocak K., Saylan L. and Sen O., (2000). Nonlinear time series prediction of O3 concentration in Istanbul. Atmospheric Environment, *34*, 1267-1271.

Kyriakidis P. C. A. G. Journel, (2001). Stochastic modeling of atmospheric pollution: a spatial time series framework. Part II: application to monitoring monthly sulfate deposition over Europe. Atmos. Environ., *35*, 2339-2348.

Lee C. K., (2002). Multifractal characteristics in air pollutant concentration time series. Water Air Soil Poll., *135,* 389-409.

Lee C. K., Ho D. S., Yu C., Wang C., Hsiao Y., (2003). Simple multifractal cascade model for air pollutant concentration (APC) time series. Environmetrics, *14* (2), 255-269.

McKee D. J., (1993). Health effects associated with ozone and nitrogen dioxide exposure. Water Air Soil Poll., *67*, 11-35.

Roberts S., (2003). Combining data from multiple monitors in air pollution mortality time series studies. Atmos. Environ., *37*, 3317–3322.

Salas J. D., (1993). Analysis and modeling of hydrologic time series, In: D. R. Maidment (Ed.) Handbook of Hydrology, McGraw Hill, New York.

Salcedo R. L. R., Alvim Ferraz M., Alves C. and Martins F., (1999). Time series analysis of air pollution data. Atmos. Environ., 33, 2361-2372.

Schwartz J. and Marcus A., (1990). Mortality and air pollution in London: a time series analysis. Am. J. Epidem., 131, 85-194.

Touloumi G., Atkinson R. and Terte A. L., (2004). Analysis of health outcome time series data in epidemiological studies. Environmetrics, 15, 101-117.

Voigt K., Welzl G. and Bruggemann R., (2004). Data analysis of environmental air pollutant monitoring systems in Europe. Environmetrics, 15, 577-596.

Windsor H. L. and Toumi R., (2001). Scaling and persistence of UK pollution. Atmos. Environ., 35, 4545-4556.

Yee E., and Chen R., (1997) A simple model for the probability density functions of concentration fluctuations in atmospheric plumes. Atmos. Environ., 31:, 991-1002.