

Frequency analysis of air quality time series for traffic related pollutants†

O. Tchepel^{*a} and C. Borrego^b

Received 10th July 2009, Accepted 14th October 2009

First published as an Advance Article on the web 18th November 2009

DOI: 10.1039/b913797a

In the present work, annual time series of traffic related pollutants (CO and PM₁₀) were considered for frequency analysis (Fourier series) with the aim to understand the underlying physical processes and the influence of emission sources on the variability of the air pollutant concentrations. Several urban traffic and suburban background air quality stations located in Porto metropolitan area (Portugal) were analysed. The results obtained for CO and PM₁₀ reveal the important contributions of short-term fluctuations (12 h and 24 h periods). However, the spectrum signals at low frequencies are significantly different between these pollutants thus stressing that temporal variations of CO and PM₁₀ are influenced by different processes. Cross-spectrum analysis of the air quality time series against wind measurements and traffic counts allowed us to identify the contribution of long-range transport over a period of about 21 days to the PM₁₀ fluctuations. Also, a correlation of over 80% between the pollution levels in the vicinity of traffic sources and suburban background levels are found for these harmonic components in the PM₁₀ spectrum, while correlations for CO is below a significant level. Thus, the spectrum and cross-spectrum analysis performed in this study reveal the distinct influence of local traffic emissions and long-range transport to CO and PM₁₀ fluctuations in the polluted urban area. The methodology shows to be a powerful tool for the analysis of the causes of air pollution.

1. Introduction

A common time series analysis based on the estimation of summary statistics, such as central tendency and spread of air pollutant concentration data, is a simple and effective tool to describe the variation in the response over time. However, more sophisticated techniques are required to better understand the main reasons for these variations. Thus, spectral analysis is used to describe time series in the frequency domain and to analyse how the variation in a time series is accounted for cyclic components at different frequencies.¹

In atmospheric sciences, spectral analysis is widely used for meteorological variables, for example, turbulence analysis. It provides information on the contribution of eddies of different

sizes to the total turbulence kinetic energy²⁻⁶ and allows the separation of synoptic and seasonal signals.⁷

However, studies on frequency analysis applied to air pollution data⁸ have emerged much later than in atmospheric physics, although frequency analysis is presently becoming an important tool in air pollution studies.⁹⁻¹³ Analysis of air quality data in the frequency domain contributes to the understanding of periodic behaviours and yields information about temporal and spatial scales of the underlying mechanisms.

There is a strong relation between temporal and spatial scales of air pollution. In this context, short-term fluctuations of the pollutant concentrations are related with local scale phenomena, including local dispersion conditions, local emissions and chemistry. By contrast, seasonal changes in the emissions and long-range transport of the pollution will contribute to the spectrum at low frequencies. Usually, air pollution time series have a broad spectrum related to periodicity of atmospheric physical processes and precursor emissions. Decomposition of the measurements into annual, seasonal, monthly, weekly, and daily components allows the study of contributions at different time scales and, therefore, from different phenomena.

^aCESAM & Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal. E-mail: oxana@ua.pt; Fax: +351 234370309

^bCESAM & Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal

† Electronic supplementary information (ESI) available: Data analysis details. See DOI: 10.1039/b913797a

Environmental impact statement

The solution to air pollution in urban areas is a complex issue and despite the continuous implementation of the abatement measures, high pollution levels are still frequently observed. Therefore, better understanding of the cause-effect relationship is required. This current work is focused on the Fourier series analysis of air quality measurements with the aim to understand underlying physical processes and the influence of emission sources on the variability of the air pollutant concentrations. The spectrum and cross-spectrum analysis performed in this study reveal the distinct influence of local traffic emissions and long-range transport to CO and PM₁₀ fluctuations in the polluted urban area. This methodology shows to be a powerful tool for the analysis of the causes of air pollution.

This paper describes the methodology for time series analysis of air quality data in the frequency domain to obtain information on variability of air pollutant concentrations, and gives an example of the application of this method to the data obtained at the National air quality monitoring network in Porto metropolitan area (Portugal). Different types of air quality stations are compared in terms of their spectrum pattern. This study is focused on traffic related pollutants and includes carbon monoxide (CO) and particulate matter with aerodynamic diameter equal to or less than $10\ \mu\text{m}$ (PM_{10}). The time series of these pollutants are analysed in the frequency domain and a contribution of different periodic fluctuations to the variance is discussed. Additionally, cross spectra of the pollutant concentrations with traffic counts and wind characteristics are analysed for a selected location searching for an interpretation of the spectrum peaks obtained for the time series.

2. Data and methodology

Several steps of the frequency analysis as well as pre-processing of the air quality data applied in this study are described below.

2.1. Original data

Air quality data for CO and PM_{10} , measured at different stations during 2004 were analysed (Fig. 1). The criteria for the data selection were proximity to the Porto urban area, completeness of the measurements (data capture above 90%) and the type of air quality station. Thus, urban traffic stations with direct influence of road transport emissions and suburban background stations to characterise background pollution levels were selected. The time series for each pollutant consist of 1 year concentration values with 1 hour resolution.

Table 1 Statistics for air quality data measured in 2004

		Concentration/ $\mu\text{g m}^{-3}$				No. of exceedences ^a
		Mean	Median	SD	Max.	
CO	Matosinhos	703	594	436	5416	—
	S. Hora	548	434	405	5669	—
	Antas	726	593	508	5033	—
	Boavista	531	399	403	4288	—
	Leça do Balio	521	407	365	3457	—
	V. N. Telha	481	404	326	3175	—
PM_{10}	Matosinhos	42	35	32	249	107
	S. Hora	37	31	31	243	81
	Antas	41	35	28	226	97
	Boavista	50	40	46	641	136
	Leça do Balio	35	26	33	221	84
	V. N. Telha	36	29	28	240	73

^a Number of days with an average PM_{10} concentration above $50\ \mu\text{g m}^{-3}$.

Descriptive statistics for the air quality measurements are presented in Table 1. The CO concentrations measured at the different stations are within the legislation limits, while PM_{10} concentrations are high and 3 of the 6 stations have annual average values above the legislation limit ($40\ \mu\text{g m}^{-3}$). Also, all the analysed stations are in noncompliance with the daily average PM_{10} limit value ($50\ \mu\text{g m}^{-3}$ exceeded no more than 35 days per year). To improve the air quality, understanding the principal causes for high pollutant concentrations is important.

For the cross spectrum analysis, Antas urban traffic station located in Porto was selected concerning the data availability. The automatic vehicle counts available for two months only (November and December) were used in the study. Looking at air mass transport, wind speed and wind direction were selected from the meteorological parameters and considered in the analysis as u and v components.

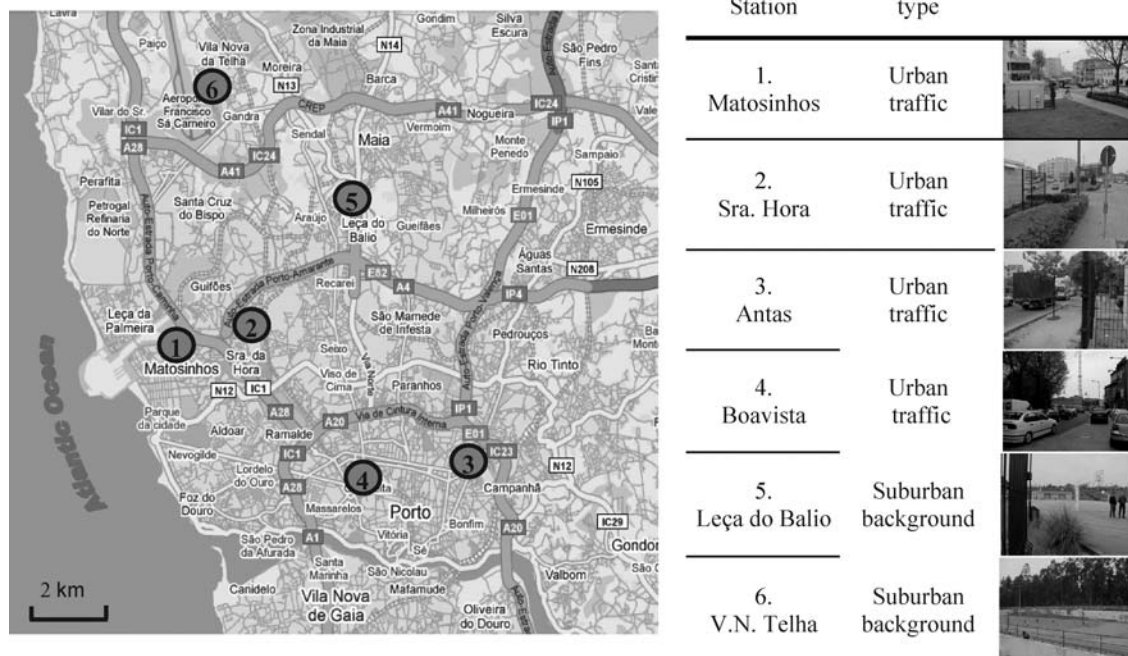


Fig. 1 Air quality monitoring stations considered for the analysis.

2.2. Data pre-processing

Pre-processing of the original time series is necessary prior to their analysis because of log-normal distribution of the data and missing values. In this study, pollutant concentration measurements were log-transformed for variance stabilisation. The missing data in the time series were computed by interpolation from the adjacent non-missing points. This approach assumes that there is serial correlation in the data and that each observation is to some extent related to the previous observation. However, the interpolation would introduce signals that are not in the original data. Alternatively, padding with zeroes was tested to resolve the problem of missing points. No considerable differences were obtained from these two approaches in terms of the resulting power spectra.

The overall mean is subtracted from the series prior to the analysis because the goal of spectrum analysis is to detect underlying periodicity and the overall mean is not of interest. Detrending of the data is also important as spectral analysis is intended for stationary series. The trend may be defined as low-frequency components with a wavelength longer than the record length. In this work, linear trend have been removed from the data.

2.3. Fourier analysis

To describe a variation of concentration values over time, it is assumed that measured data can be approximated by a periodic function with period T , such that:

$$F(t) = F(t + nT), \text{ where } n \text{ is an integer.}$$

Then, a time series X_t of length N is represented as a linear combination of harmonic functions with frequencies $\{f_j\}$ and amplitudes $\{A_j\}$ and $\{B_j\}$:

$$X_t = \mu + \sum_{j=1}^{[N/2]} [A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)],$$

$$t = 1, 2, \dots, N$$

where μ is a constant, $[N/2]$ is the greatest integer less than or equal to $N/2$, and the frequencies f_j are related to the sample size N by

$$f_j \equiv j/N, 1 \leq j \leq [N/2]$$

Thus, the measurement data with 1 hour resolution do not allow analysing waves with a period of less than 2 hours (Nyquist theorem).

The fast Fourier transform (FFT) algorithm was used to study the variance in the frequency-domain and to identify which frequencies are more important to the variability of the time series.

2.4. Periodogram

The results of the Fourier analysis are presented in the form of a periodogram. The periodogram value indicates the strength of the signal at the respective frequency and can be considered as

a representation of the spectral density of variance. The highest periodogram values reveal the most important cyclic components and the contribution of seasonal, weekly or diurnal fluctuations to the total variance of the pollutant concentration measurements could be quantified.

For a discrete spectrum, the area of the periodogram is defined as

$$S(f) * \Delta f = (A_j^2 + B_j^2) / 2$$

where $S(f)$ is the height of the histogram, Δf is the width of the histogram ($\Delta f = 1/N$) and N is the overall length of the series. In a linear plot of spectrum values S against frequency, the area under the curve between any pair of frequencies is proportional to the portion of variance explained by that range of frequencies,⁵ and the total area under the curve represents the variance of the series.

However, it is difficult to represent and analyse a wide range of frequencies at linear scale. Therefore, log-normal presentation of the data has been chosen in this study. Also, the smoothed spectrum is used as an alternative to the raw periodogram in order to remove noisy peaks making interpretation of the spectrum easier.

2.5. Bivariate spectral analysis

Cross spectral analysis establishes the relationship between two time series as a function of frequency and, therefore, allows us to investigate how periodicities in the two datasets are interrelated. The cross-spectral estimate is a complex function and can be presented for two observed variables x and y by:

$$S_{xy}(f) = C_{xy}(f) - iQ_{xy}(f),$$

where $C_{xy}(f)$ is the cospectrum (the real part) and $Q_{xy}(f)$ is the quadrature spectrum (the imaginary part) and they can be used as a measure of covariance between the respective frequency components in the two series.

Additionally, the squared coherence is used to interpret the cross-spectrum:

$$R_{xy}(f) = \frac{C_{xy}(f)^2 + Q_{xy}(f)^2}{S_x^2(f) S_y^2(f)}$$

This function measures the square of the linear correlation between the two components of the bivariate process at frequency f and is analogous to the square of the correlation coefficient. The coherence takes values between 0 and 1, and is equal to 1 if the linear relation exists. The large coherence amplitude implies that x and y are strongly correlated at that frequency and accounts for the in-phase signal as well as to the possible lagged relationships that occur between the two series. However, if the phases of the signals are changing randomly in the selected interval of frequencies, the coherency will tend to be zero. To determine whether there is a statistically significant linear relation between the two series at frequency f , the coherence values are compared with the critical values calculated for the level of significance $\alpha = 0.01$ using F distribution, as described by Shumway and Stoffer.¹⁴

3. Results and discussion

3.1. Spectrum analysis of CO and PM₁₀ time series

The log-transformed air quality data were used for the spectrum analysis after removing the linear trend and subtracting the annual average log concentration. An example of the results for CO and PM₁₀ is presented in Fig. 2 (the results for all analysed stations can be found in the ESI†). The 95% confidence interval is presented for the smoothed spectra. The plot represents the relative contributions of different frequencies to the variance.

The spectra obtained for CO and PM₁₀ time series reveal clearly identified peaks at higher frequencies ($f \approx 0.083 \text{ h}^{-1}$ and $f \approx 0.042 \text{ h}^{-1}$) corresponding to the period of $T = 12 \text{ h}$ and $T = 24 \text{ h}$ for all monitoring stations. In some locations, the contribution of the 12 h period to the total variance is even higher than day-to-night variations. This pattern could be related to local emission changes due to the traffic peak hours and wind speed fluctuations. However, at low frequencies there is a noticeable difference between PM₁₀ and CO. The spectrum obtained for PM₁₀ reveals the significant contribution of the waves with the periods of about 4–40 days to the total variance not observed in the CO spectrum. Low frequencies (period about

6 month) are the most important spectrum components for CO at all analysed locations, while 1 week fluctuations are smoothed.

An intercomparison of the spectra for different pollutants measured at the same monitoring point could provide important information. If the emission sources for CO and PM₁₀ are the same, similar cyclic patterns can be expected. Differences in the pollutants spectra can indicate contributions from distinct emission sources or presence of chemical transformations, because the dispersion conditions are identical for both compounds when measured at the same point.

There is no doubt that CO fluctuations measured at urban traffic stations are strongly influenced by the road traffic emissions. However, PM₁₀ could also be emitted by other sources, such as soil erosion, building constructions, *etc.* Moreover, pollution transported from the other locations may have an important contribution to the observed levels. Therefore, different fluctuation patterns observed for CO and PM₁₀ may provide evidence of non-local or/and non-traffic causes for PM₁₀.

3.2. Spectrum analysis of traffic counts and wind components

Antas urban traffic station was selected to investigate how the periodicity pattern in concentration fluctuations are possibly related to road traffic and meteorological conditions.

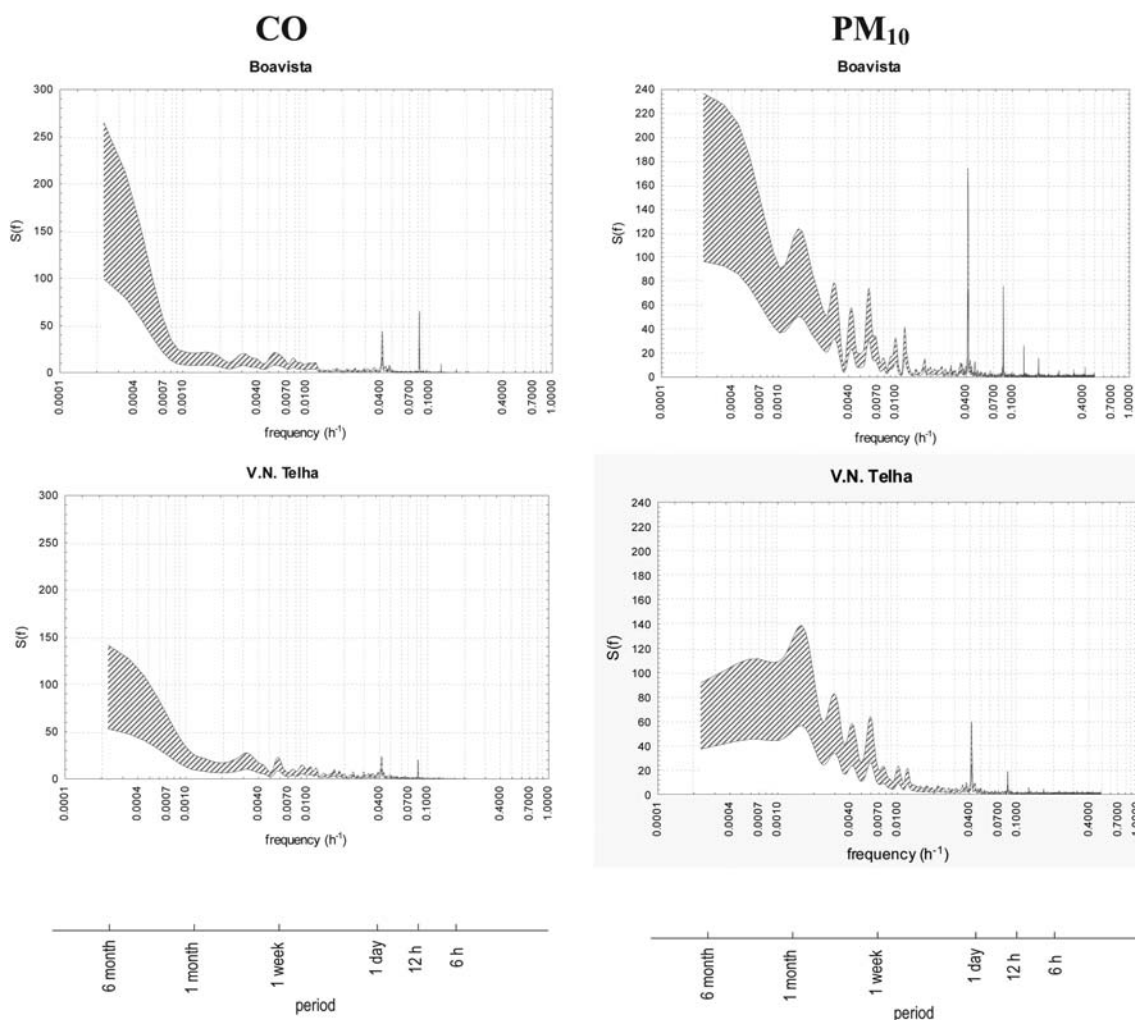


Fig. 2 Power spectrum density of CO and PM₁₀ at urban traffic station (Boavista) and suburban background station (V. N. Telha).

Firstly, the data for hourly vehicle counts (available for November and December) and wind characteristics (available for one year) were analysed. The traffic flows are strongly affected by daily variations (periods with 24 h) and peaks for 8 h, 12 h and 1 week periods are clearly identified in the resulting spectrum. Although the pollutant concentrations in the air are not directly related with the number of vehicles but with their emissions (which are also influenced by vehicle technology and driving pattern), vehicle flow is still a crucial parameter to

analyse possible correlation between the road traffic and the pollutants.

Spectra for absolute values of u and v wind components (see the plot in the ESI†) allow the analysis of main fluctuations in latitudinal and longitudinal wind components to be carried out. Taking into account the location of stations (near the coast) and the orientation of the dominant winds in Portugal, v component is primary related with synoptic scale and long transport, while u component is also affected by breeze circulations, which depend

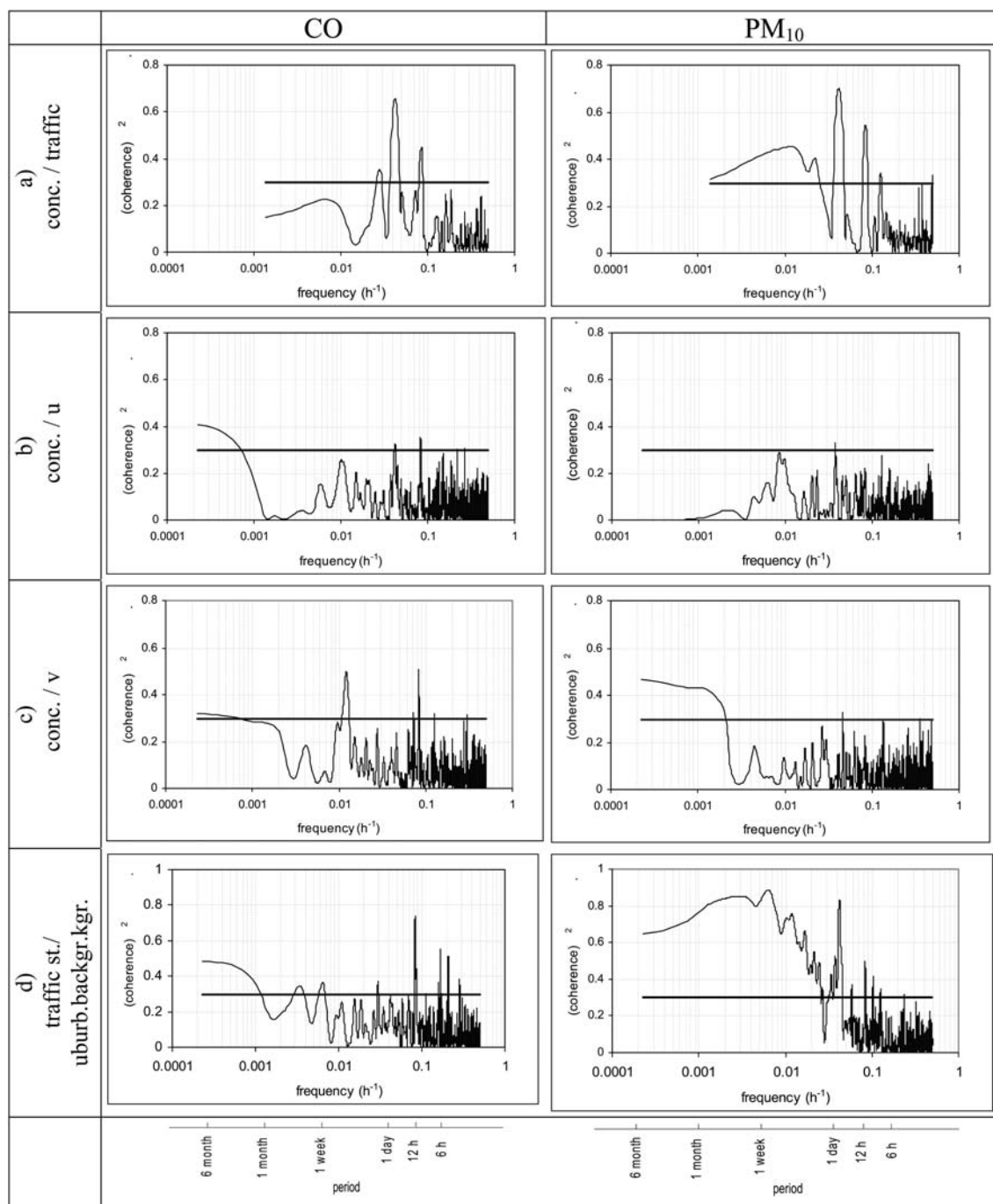


Fig. 3 Results of the cross-spectrum analysis: squared coherency spectrum and 0.01 significance level for pollutant concentrations *versus* (a) traffic counts, (b) u -wind component, (c) v -wind component and (d) concentration measurements at urban traffic station *versus* suburban background station.

on the orientation of the ocean coast. The most important spectrum peak obtained for the u component is attributed to day/night fluctuations while the v component is mostly contains harmonic waves with the period of about 21–30 days.

3.3. Cross-spectrum analysis of concentrations versus traffic counts

The cross-spectrum analysis was performed considering co-spectrum, quadrature spectrum and squared coherence. The last parameter is one of the most useful in the cross-spectrum analysis¹⁵ that provides linear correlation between two time series.

The cross spectrum analysis confirms the existence of a strong correlation (the squared coherence is about 0.7) between the road traffic and pollutant concentration measurements for fluctuations with a 24 h period (Fig. 3(a)). Additionally, both the co-spectrum and the quadrature spectrum show an important contribution of this peak to the covariance (co-spectrum and quadrature spectra are included in the ESI†). Fluctuations with 12 h, 48 h and 1 week periods are present in the coherence spectrum of the both pollutants. However, the correlation between the CO and traffic counts for 1 week waves are not distinguished from the noise (considering 0.01 significance level), while for PM₁₀ the correlation with 1 week traffic fluctuations is about 45%.

3.4. Cross-spectrum analysis of concentrations versus wind components

The results of the cross spectrum analysis for the pollutant concentration versus wind components presented in Fig. 3(b), (c) show that squared coherence is significantly different for CO and PM₁₀ (co-spectra and quadrature spectra are presented in the ESI†).

The highest correlation (50%) between the CO concentration and the wind velocity is observed for the frequencies of $f \approx 0.083 \text{ h}^{-1}$ and $f \approx 0.01 \text{ h}^{-1}$ corresponding to the wave periods of about 12 h and 4 days, respectively, and is related to the v component. Oppositely, low frequencies are more important in u component achieving the coherence of 0.4.

The analysis of PM₁₀ versus u wind component shows that co- and quadrature spectra fluctuate around zero. Few peaks, identified in the covariance within the frequency range of $0.001 < f < 0.01$, are not significant in the coherence spectrum. This means that the phase relationship in the two time series does not remain stable resulting in a low coherence value.

A large contribution of low frequencies ($f < 0.002 \text{ h}^{-1}$) is clearly seen in the plots for the v component. For these frequencies, the correlation between PM₁₀ concentrations and wind velocity is about 35–45%. Furthermore, a positive covariance found between the two time series for the periods of about 21 days means that the increase in the concentration levels is related with the increase of wind speed and may indicate existence of a long-range transport of the pollution. Oppositely, negative covariance will indicate a presence of higher pollutant concentrations under low wind conditions (for example in anti-cyclone conditions), or a decrease of the concentrations related with an increase of wind speed due to the transport of clean mass

of air to the monitoring location. The presence of this signal in the quadrature spectrum reveal that the fluctuations of the two variables are lagged in time.

3.5. Cross-spectrum analysis of urban versus background measurements

The cross spectrum analysis for urban traffic (Antas) and suburban background station (V. N. Telha) have been carried out to determine whether the periodicities found at these two stations are correlated to each other. As can be seen in Fig. 3(d), the correlation between the two stations for PM₁₀ is very important in a wide range of frequencies with the squared coherence of >0.8 . Additionally, the co-spectrum indicates that a major contribution to the covariance of PM₁₀ observed at Antas and V. N. Telha is related with the cycles of about 28 days ($f \approx 0.0015 \text{ h}^{-1}$). This signal was previously identified in the latitudinal (v) wind component (the figure is included in the ESI†) and is related with the synoptic scale and long transport. By contrast, for CO time series no correlation could be found at this frequency between the urban traffic and suburban background stations (Fig. 3(d)) and the highest correlation is attributed to the 12 h fluctuations. These results emphasize a different nature of the processes that contribute to CO and PM₁₀ concentration variations. A similar cross spectrum between Antas and Leça do Balio monitoring points was obtained.

4. Conclusions

In the current work, the methodology of spectral analysis applied to the air quality data is discussed and the time series of CO and PM₁₀ related to traffic emission sources in the Porto metropolitan area are analysed. The results show an important contribution of short-term fluctuations (12 h and 24 h periods) to the total variance for both pollutants. At this time scale, the contribution of the local traffic source is confirmed by the cross-spectrum analysis which indicates a 45–70% correlation between the concentration and the traffic flux variations. The meteorological conditions are also important at this scale. The highest correlation with the wind v component was obtained for CO 12 h peak (about 50%). However, for the CO concentrations a negative covariance with the v wind component was found meaning the CO concentration decrease in the presence of a strong wind.

Completely different spectrum patterns are observed for the two pollutants in the frequency range of about $0.01\text{--}0.001 \text{ h}^{-1}$ corresponding to the wave periods of approximately 4–40 days. For CO, an unexpected result was found for the 1–2 week periods, considering that in the urban areas the CO level is directly influenced by the road traffic emissions. The peaks corresponding to these periods are smoothed in the CO spectrum and their correlation with the traffic flow is not distinguished from the noise. Contrary to that, the PM₁₀ spectrum reveals an important contribution of these cycles to the total variance resulting in the cross-correlation of the PM₁₀ peaks with those in the traffic spectrum of about 45%.

A long-range transport of PM₁₀ pollution was attributed to the fluctuations with the periods of about 21 days, identified by a positive covariance of the pollutant concentration versus v wind

component in the quadrature spectrum, being the correlation between these two parameters of about 35%. Moreover, long waves with the similar periods were found to give the most important contribution to the covariance spectrum between the urban traffic measurements and suburban background levels (with correlation above 80%) thus indicating the same underlying processes for the PM₁₀ variations at distinct monitoring points.

Low frequencies (approximately, $f < 0.001 \text{ h}^{-1}$, period > 40 days) contribute significantly to the variance of both pollutants analysed in the present work, and a correlation of these signals with meteorological parameters was detected. However, if the CO fluctuations are better correlated with the u wind component, the PM₁₀ peaks are related with v component. Therefore, concentration variations are associated with different wind directions in the case of CO and PM₁₀ and this fact indicates that the pollutants are transported from distinct locations and originated by different sources.

The results obtained in this study show that the frequency analysis of air quality time series is a powerful technique that can provide important information about the nature of the processes behind the measurements. The methodology implemented in this work could be applied to determine representative background concentrations of air pollutants by removing short-term fluctuations associated with influence of local emission sources.¹⁶ Also, cross spectral analysis could be valuable in air pollution modelling to examine sources of the model uncertainties. The findings from the current work contributes to the understanding of the cause-effect relationship in the pollutant concentration variations that required for air quality management.

Acknowledgements

The methodology presented in this work is partly developed within the European project Air4EU - Air Quality Assessment for Europe: from local to continental scale (SSPI-CT-2003-503596) and the Portuguese project PAREXPO - Particulate matter in ambient air and human exposure (POCI/AMB/57393/2004).

References

- 1 D. Percival and A. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, 1998.
- 2 I. Hoven, Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour, *J. Meteorology*, 1957, **14**, 160–164.
- 3 N. K. Vinnichenko, The kinetic energy spectrum in the free atmosphere – 1 s to 5 years, *Tellus*, 1970, **22**, 158–166.
- 4 T. J. Lyons, Mesoscale wind spectra, *Q. J. R. Meteorol. Soc.*, 1975, **101**, 901–910.
- 5 R. B. Stull, *An introduction to boundary layer meteorology*, Kluwer Academic Publishers, 1988.
- 6 D. H. Lenschow and B. B. Stankov, Length Scales in the Convective Boundary Layer, *J. Atmos. Sci.*, 1986, **43**, 1198–1209.
- 7 R. Eskridge, J. Y. Ku, S. T. Rao, P. S. Porter and I. G. Zurbenko, Separating Different Time Scales of motion in Time Series of Meteorological variables, *Bull. Am. Meteorol. Soc.*, 1997, **78**, 1473–1483.
- 8 S. T. Rao, P. J. Samson and A. R. Peddada, Spectral analysis approach to the dynamics of air pollutants, *Atmos. Environ.*, 1976, **10**, 375–379.
- 9 T. Hies, R. Treffeisen, L. Sebald and E. Reimer, Spectral analysis of air pollutants. Part 1: elemental carbon time series, *Atmos. Environ.*, 2000, **34**, 3495–3502.
- 10 L. Sebald, R. Treffeisen, E. Reimer and T. Hies, Spectral analysis of air pollutants. Part 2: ozone time series, *Atmos. Environ.*, 2000, **34**, 3503–3509.
- 11 L. C. Marr and R. A. Harley, Spectral analysis of weekday–weekend differences in ambient ozone, nitrogen oxide, and non-methane hydrocarbon time series in California, *Atmos. Environ.*, 2002, **36**, 2327–2335.
- 12 C. Hogrefe, P. S. Porter, E. Gego, A. Gilliland, R. Gilliam, J. Swall, J. Irwin and S. T. Rao, Temporal features in observed and simulated meteorology and air quality over the Eastern United States, *Atmos. Environ.*, 2006, **40**, 5041–5055.
- 13 Y. Choi, C. Ho, D. Chenb, Y. Noha and C. Song, Spectral analysis of weekly variation in PM₁₀ mass concentration and meteorological conditions over China, *Atmos. Environ.*, 2008, **42**, 655–666.
- 14 R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, New York, Springer, Verlag, 2000.
- 15 C. Chatfield, *Time Series Analysis: An Introduction*, London, Chapman & Hall, 2004.
- 16 O. Tchepel, A. M. Costa, H. Martins, J. Ferreira, A. Monteiro, A. I. Miranda and C. Borrego, Determination of Background Concentrations for Air Quality Models Using Spectral Analysis of Monitoring Data, *Atmos. Environ.*, 2009, DOI: 10.1016/j.atmosenv.2009.08.038.