



Seasonal variation of benzo(a)pyrene in the Spanish airborne PM₁₀. Multivariate linear regression model applied to estimate BaP concentrations

M.S. Callén*, J.M. López, A.M. Mastral

Instituto de Carboquímica (CSIC), Miguel Luesma Castán, 4, 50018 Zaragoza, Spain

ARTICLE INFO

Article history:

Received 3 February 2010

Received in revised form 20 April 2010

Accepted 20 April 2010

Available online 28 April 2010

Keywords:

PAH

BaP

MLR model

Internal validation

External validation

ABSTRACT

The estimation of benzo(a)pyrene (BaP) concentrations in ambient air is very important from an environmental point of view especially with the introduction of the Directive 2004/107/EC and due to the carcinogenic character of this pollutant. A sampling campaign of particulate matter less or equal than 10 microns (PM₁₀) carried out during 2008–2009 in four locations of Spain was collected to determine experimentally BaP concentrations by gas chromatography mass-spectrometry mass-spectrometry (GC-MS-MS).

Multivariate linear regression models (MLRM) were used to predict BaP air concentrations in two sampling places, taking PM₁₀ and meteorological variables as possible predictors. The model obtained with data from two sampling sites (all sites model) ($R^2 = 0.817$, PRESS/SSY = 0.183) included the significant variables like PM₁₀, temperature, solar radiation and wind speed and was internally and externally validated. The first validation was performed by cross validation and the last one by BaP concentrations from previous campaigns carried out in Zaragoza from 2001–2004. The proposed model constitutes a first approximation to estimate BaP concentrations in urban atmospheres with very good internal prediction ($Q_{CV}^2 = 0.813$, PRESS/SSY = 0.187) and with the maximal external prediction for the 2001–2002 campaign ($Q_{ext}^2 = 0.679$ and PRESS/SSY = 0.321) versus the 2001–2004 campaign ($Q_{ext}^2 = 0.551$, PRESS/SSY = 0.449).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Benzo[a]pyrene (BaP) is a complex chemical, belonging to the family of Polycyclic Aromatic Hydrocarbons (PAH), which is released widespread into the air. Most of the PAH are emitted as products of fuel thermal processes and they are released from anthropogenic activities involving the devolatilization of coal, oil, wood, diesel and gasoline [1]. Major sources of PAH are mobile sources and industrial activities although minor natural sources also include volcanoes and natural fires. PAH, once released to the atmosphere, can be transported long distances in air. This is the reason why possible adverse health and wildlife effects can take place even in places remote from the emission source.

BaP may cause cancer and genetic damage affecting the blood, the immune system, the reproductive system and the unborn child [2–5]. All these harmful effects on human health have supported the inclusion of BaP in the ninth position of the 2007 CERCLA (Comprehensive Environmental Response, Compensation, and Liability Act) [6] Priority List of Hazardous Substances, which compiles the 275 most toxic compounds. This list includes substances that are most

commonly found at facilities on the National Priorities List (NPL) which are determined to pose the most significant potential threat to human health. In the EU, BaP is on the final list of 11 substances identified as “priority hazardous substances” under the proposed Water Framework Directive [7] and it is also included in the list I of dangerous substances [8]. As an organic compound, it is listed as a candidate substance for selection, assessment and prioritisation under the Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR) and Helsinki Conventions [9]. In addition, BaP is included as indicator for the purposes of emission inventories under the United Nations Economic Commission for Europe (UN-ECE) Persistent Organic Pollutants Protocol [10]. Therefore, different legislations are led to reduce pollution to levels which minimise harmful effects on human health, paying particular attention to sensitive populations. With regard to BaP, the aim is to achieve air concentrations lower than 1.0 ng/m³ in the PM₁₀ according to Directive 2004/107/EC [11] and to establish a monitoring and information support system for the protection of the air quality. Some countries, like the United Kingdom, have even adopted more restricted air quality standards for BaP (annually mean standard for BaP of 0.25 ng/m³).

Although there is bibliography related to PAH in air [12–18], conventional sample preparation relies on tedious and time-consuming procedures so that the models development, which

* Corresponding author. Tel.: +34 976 733977; fax: +34 976 733318.
E-mail address: marisol@icb.csic.es (M.S. Callén).

allows estimating and predicting PAH concentrations, would be advantageous. Up to now, scarce bibliography has been found related to BaP and PAH predictions in the atmosphere [19,20].

The aim of this paper is to summarize the results of the monitoring of BaP concentrations in four locations of Spain with different anthropogenic activities for the warm and cold seasons. Because this pollutant is mainly associated with the particle phase, a study of the PM₁₀ concentrations was also carried out. Finally, BaP concentrations were predicted based on meteorological conditions and PM₁₀ concentrations by applying multivariate linear regression models (MLRM) to the obtained data. These estimations were validated, not only internally by cross validation but also externally with BaP concentrations obtained experimentally from previous measurements carried out in Zaragoza during 2001–2004.

2. Materials and methods

2.1. Sampling program

Four monitoring sites were chosen as representative of Aragón area in an attempt to cover all the main types of anthropogenic activities as well as biogenic sources. The first one was the city of Zaragoza ZGZ, located in the North-East of Spain, (41°39'49.38"N; 0°53'16.68"W) in which traffic pollution and industrial activities are present and previously detailed in other publication [21]. The second place was the city of Monzón (Huesca) MON (41°53'59"N; 0°10'47"E) (more than 17,000 inhabitants in 2009), the second biggest city in the Huesca province whose economy is mainly based on the industry related to food, construction, chemicals as well as service sector. The third one was Monagrega (Teruel) MNG (40°56'23"N; 0°19'15"W), a regional background station situated 7.6 km from a power station (Teruel 1050 MW) in East direction 135°. Finally, the fourth one was Torrelisa (Huesca) PIR (42°27'36"N; 0°10'48"E) localized in the Pyrenees Mountain and considered as representative of biogenic sources.

The sampling was carried out by using a GUV-15H Graseby Andersen High-Volume air sampler with volumetric flow controlled system (1.13 m³/min) provided with a PM₁₀ cut off inlet at 10 μm to capture PAH in the particle phase (PTFE-coated, glass-fibre filters, 0.6 μm pore size) during 24 h. Samples were taken in the warm season from 23-05-2008 until 03-08-2008 during consecutive days starting in ZGZ, PIR, MNG and MON and in the cold season from 13-01-2009 until 24-03-2009 in ZGZ, MNG, MON and PIR, respectively. More details regarding the sampling procedure and filter treatment have been previously published [12].

2.2. Extraction and analysis

The PM₁₀ concentrations were determined by gravimetric analysis and BaP concentrations were measured according to the procedure previously published by GC–MS–MS with the internal standard method [12,13]. Briefly, filters were extracted by Soxhlet after the addition of a surrogate standard (BaP-d₁₂). Extracts were then evaporated in a rotary evaporator, processed through a silica gel column and concentrated by N₂ stream until GC–MS–MS analysis. Before injection to the GC–MS–MS, p-terphenyl was added as internal standard.

2.3. Quality control and quality assurance

Analyses of four samples of an appropriate standard reference material (SRM 1649a, urban dust) provided by the National Institute of Standards and Technology (NIST) were carried out in order to check the analytical accuracy and precision. Measured values were comparable to certified values with relative errors lower than 4% for BaP.

The detection (LOD) and quantification (LOQ) limits for BaP were calculated as the concentrations equivalent to multiply by three and by ten the standard deviation of the blank filters, respectively (0.006 ng/m³; 0.020 ng/m³). The mean surrogate recovery was higher than 90% for the four sampling places therefore BaP concentrations reported in this work were not corrected for recovery efficiency.

2.4. Meteorological variables

The meteorological variables recorded for each sampling place and providing information regarding photooxidation, long-range atmospheric transport and gas-particle partitioning were: temperature (°C), relative humidity (%), solar radiation (W/m²), UV radiation (W/m²), pressure (mbar), rainfall (mm), wind speed (m/s) and season (warm season is considered from 21st March to 21st September and cold season is considered from 22nd September to 20th March). In Zaragoza, the meteorological data were provided daily by the Estación Experimental de AULA-DEI (CSIC) whereas the data for the other three places were provided hourly by the Aragon Government (DGA). In PIR and because there were not available data for all dates, data were obtained by the SIAR network [22].

2.5. Statistical tools

The SPSS Version 15.0 statistical package was used as statistical tool: (a) to measure the correlation between two variables (the considered variables were: PM₁₀ and BaP concentrations and the meteorological variables) by using Pearson correlation coefficients, (b) to test for significant differences in seasonal air concentrations (PM₁₀ and BaP) by using a parametric test (Student's *t*-test of independent samples for each sampling place) and (c) to run MLRM in order to find the best-fit model between the estimated and the experimental BaP concentrations according to PM₁₀ concentrations and meteorological variables.

2.5.1. Model development

Individual models were built for each sampling place by considering the stepwise (forward and backward) model, in which the choice of predictive variables is carried out by an automatic procedure [23–25]. The BaP concentrations were taken as dependent variable whereas the PM₁₀ and the meteorological variables were taken as independent variables. The variable *season* was additionally considered into the possible models because statistically variations were observed for BaP concentrations in two of the sampling places. Usually, this stepwise procedure takes the form of a sequence of *F*-tests (Fisher's ratio). An *F*-test is based on the null hypothesis. The test calculates the value of *F* and then compares it with the critical *F* values available in statistical tables according to freedom degrees and a confidence level. The null hypothesis is rejected if the *F* calculated from the data is greater than the critical value of the *F* distribution for a particular confidence level. In this work, only those variables showing a significant correlation with the BaP (*p* < 0.05) were considered as potential predictors for the best-fit models. Because in PIR and MNG most of the BaP concentrations were below the detection limit, both sites were excluded of the dataset.

The model acceptability was determined using several criteria [26]:

- Dataset used to obtain the model contained at least five times more components than the amount of model variables used.
- The model had got good performance. This can be expressed as coefficient of determination (*R*²) although unfortunately a high *R*² value does not guarantee that the model fits the data well. There are many statistical tools for model validation, but the primary

tool for most modeling applications is graphical residual analysis. The following assumptions about the residuals were also verified: They must be independent (and thus random), they should be of normal distribution and they must have a constant variation across the X values. Other parameters giving information over the model are the mean bias error (MBE), which indicates whether the observed concentrations are over or under estimated and the mean absolute error (MAE). The MAE and the low root mean square error (RMSE) measure residual errors and give a global idea of the difference between the observed and modeled values [27–30]:

$$\text{MBE} = \frac{\sum(\hat{y} - y)}{n} \quad (1)$$

$$\text{MAE} = \frac{\sum|\hat{y} - y|}{n} \quad (2)$$

where \hat{y} is the BaP value estimated by the model, y is the observed BaP value and n is the number of samples.

2.5.2. Validation of a model

The validation of a model is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. This is the reason why one of the most important considerations for obtaining a model is to validate that model. The validation of the model was performed by two methods: internal validation and external validation.

The internal validation was carried out by cross validation, which involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the *training set*), and validating the analysis on the other subset (called the *validation set* or *testing set*) [31,32]. Correlation coefficient of the regression between experimental and estimated data obtained by cross validation is known as cross-validated correlation coefficient or Q_{cv}^2 parameter.

The cross validation parameter, Q_{cv}^2 is mentioned in the equation:

$$Q_{cv}^2 = \frac{(\text{SD} - \text{PRESS})}{\text{SD}} \quad (3)$$

where the PRESS (predictive residual sum of squares) and SD or SSY (the sum of squared deviations of the dependent variable values from their mean) values are obtained as:

$$\text{SSY} = \text{SD} = \sum (y - y_{\text{mean}})^2 \quad (4)$$

$$\text{PRESS} = \sum (y - \hat{y})^2 \quad (5)$$

where y = the experimental BaP value, y_{mean} = the mean of the experimental BaP values, \hat{y} = the value of BaP predicted by the model and n = number of samples.

A model is considered to be significant when $Q_{cv}^2 > 0.3$. A $Q_{cv}^2 = 0.5$ is considered as good and a $Q_{cv}^2 > 0.9$ as excellent. To be a reasonable QSAR or QSPR model (quantitative structure-activity relations and quantitative structure-property relations), the PRESS/SSY ratio must be smaller than 0.4. A PRESS/SSY value < 0.1 is considered to be indicative of an excellent model [33].

Once the model was validated internally, an external validation was performed in order to determine the "realistic" predictive power of the model. In this work, the external validation was performed on an independent set of data that had undergone strict quality assurance measures, a set of 87 samples corresponding to two previous PAH samplings carried out in Zaragoza during 2001–2004 [12,13] by applying the best-fit MLR model. The predictive power of the regression model developed (training set)

was estimated on the predicted values by the external Q^2 which is defined [34]:

$$Q_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{\text{pred}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{\text{pred}} (y_i - \bar{y}_{\text{tr}})^2} \quad (6)$$

where y_{tr} is the mean value of the dependent variable for the training set, y_i and \hat{y}_i are defined above (Eq. (2)). According to the current OECD guidelines (Organization for Economic Co-operation and Development), Q_{ext}^2 for external validation should be calculated with SD referring to the training set activity mean.

3. Results and discussion

3.1. PM10 concentrations

Because PM10 is a regulated pollutant [35,36], firstly a study of the evolution of the PM10 concentrations for each sampling point was performed. A representation of the seasonal variations of PM10 for each sampling point during the two sampling periods (warm and cold seasons) is shown in Fig. 1. It is also reflected the mean PM10 concentrations for each sampling campaign as well as the daily and the annual limit values according to Directive 2008/50/EC [36].

For the warm season, the PM10 concentrations followed the decreasing order: MON > ZGZ > MNG > PIR with mean values of 32.81, 21.23, 17.18 and 10.17 $\mu\text{g}/\text{m}^3$, respectively whereas for the winter season, the order was the following: MON > ZGZ > PIR > MNG with mean values of 36.59, 30.06, 18.78 and 6.70 $\mu\text{g}/\text{m}^3$, respectively. In both seasons, the highest PM10 concentrations were obtained for MON and ZGZ in the cold period. The highest PM10 concentrations were also obtained in the cold season for PIR. Nevertheless, MNG followed a different trend in the PM10 evolution with the highest concentrations during the warm season.

In general, it has been reported in the literature that PM10 concentrations during the cold period are typical from urban profiles [37–39] affected by anthropogenic emissions joined to the high atmospheric stability and reduced mixing of air on cold season. This is the case of ZGZ and MON. Nevertheless, previous samplings carried out in Zaragoza [21,40,41] found higher PM10 concentrations in the warm season, mostly due to African episodes and resuspension processes, low precipitations and higher solar irradiation which favoured the particulate matter accumulation. This trend has also been observed in urban background stations in Madrid [42], in rural areas close to Zaragoza [43] and in some control pollution stations, which did not follow the typical trend of industrialized urban areas [39], in Cataluña.

The mean PM10 concentrations for each sampling point are shown in Table 1 for the two seasons. The exceedance percentages of the daily (50 $\mu\text{g}/\text{m}^3$) and the annual (40 $\mu\text{g}/\text{m}^3$) limit values of PM10 as well as the upper (35 $\mu\text{g}/\text{m}^3$) and the lower (25 $\mu\text{g}/\text{m}^3$) assessment thresholds of daily PM10 according to Directive 2008/50/EC [36] are also shown in Table 1. For the cold season, ZGZ and MON exceeded both, the lower and the upper assessment thresholds, whereas in the warm season, only MON exceeded the lower assessment threshold of PM10. It is noteworthy that during the 16–18 of January 2009, some rural areas close to the city celebrate the San Antón festival by burning bonfires that implied an increase in PM10 concentrations in ZGZ.

By considering the number of exceedances of the lower assessment threshold, it is expected that the concentration of 25 $\mu\text{g}/\text{m}^3$ will be exceeded in MON and ZGZ more than 35 times in a calen-

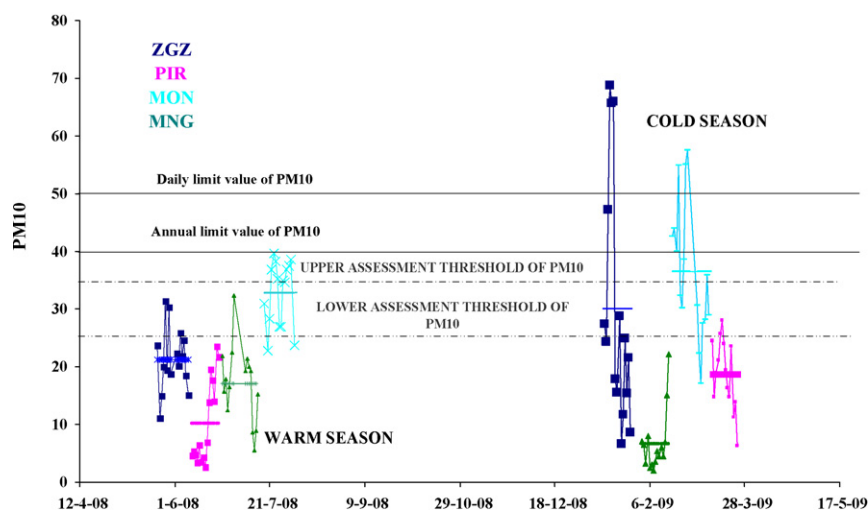


Fig. 1. Temporal evolution of PM10 ($\mu\text{g}/\text{m}^3$) and average value for each sampling point during the warm and cold seasons. The daily and the annual limit values of PM10 as well as the upper and lower assessment thresholds of PM10 are also shown.

dar year according to the Directive 2008/50/EC [36]. In MON this probability is not depending on the season while in ZGZ, these exceedances are mainly produced during the cold season. In this way, the fulfilling of the European Directives with regard to the PM10 is quite difficult in Aragón and this problem has already been reported in bibliography [21,39,43,44] in Mediterranean countries in which the particulate matter of natural origin contributes remarkably due to the influence of African episodes.

3.2. BaP concentrations

One of the main objectives of this work is to assess the BaP concentrations in different localizations in Aragón with regard to the fulfilling of the Directive 2004/107/EC [11] related to PAH in air. Independently of the sampling point, the highest BaP concentrations were obtained in the cold season. This trend was already observed in different campaigns carried out in ZGZ where the low temperatures favoured the PAH accumulation [12,13,17]. In addition, there are additional anthropogenic activities which show seasonality, one of them is the residential heating that during warm season does not contribute to PAH formation. Zaragoza was the sampling point with the highest BaP concentrations in both, warm ($\text{BaP}_{\text{mean}} = 0.089 \text{ ng}/\text{m}^3$) and cold seasons ($\text{BaP}_{\text{mean}} = 0.500 \text{ ng}/\text{m}^3$) (Fig. 2a and b), exceeding the guideline value of BaP ($1.0 \text{ ng}/\text{m}^3$) established by the Directive 2004/107/EC [11] twice in the cold season (13%) as well as the upper ($0.6 \text{ ng}/\text{m}^3$; 27%) and the lower assessment thresholds ($0.4 \text{ ng}/\text{m}^3$; 33%). In fact, the mean BaP concentration during the cold season also exceeded the lower assessment threshold of BaP. Most of these BaP exceedances were produced during San Antón (bonfire festival), obtaining the maximum BaP concentrations for those dates.

With regard to MON, this place also showed BaP concentrations exceeding the quantification limit for all samples, with higher con-

centrations during the cold season ($\text{BaP}_{\text{mean}} = 0.223 \text{ ng}/\text{m}^3$) versus the warm season ($\text{BaP}_{\text{mean}} = 0.050 \text{ ng}/\text{m}^3$) (Fig. 2c and d). Nevertheless, the guideline value of $1.0 \text{ ng}/\text{m}^3$ and the upper and lower assessment thresholds established by the Directive 2004/107/EC [11] were not exceeded.

In PIR and MNG, the BaP concentrations were almost negligible with values lower than the detection and/or quantification limits for most of the samples (PIR: $\text{BaP}_{\text{mean cold}} = 0.020 \text{ ng}/\text{m}^3$, $\text{BaP}_{\text{mean warm}} = 0.013 \text{ ng}/\text{m}^3$; MNG: $\text{BaP}_{\text{mean cold}} = 0.021 \text{ ng}/\text{m}^3$, $\text{BaP}_{\text{mean warm}} = 0.024 \text{ ng}/\text{m}^3$). These minimum concentrations are predictable by considering that PIR is localised in the Pyrenees Mountain where the anthropogenic contributions are minimum. Regarding MNG, a rural background area, minimum BaP concentrations were also detected despite the possible contribution of anthropogenic sources related to a power station and to long-range transport. For the four sampled places, ZGZ showed a higher pollution level than MON. Both localizations showed urban profiles with higher BaP concentrations during cold season in which local anthropogenic activities were responsible for most of this pollutant. This is expected by considering that the main PAH pollution sources are related to big cities involving different anthropogenic activities such as transport, industry, residential heating, etc. However, the mean BaP concentration for the two periods was $0.284 \text{ ng}/\text{m}^3$, lower than the lower assessment threshold of BaP, $0.4 \text{ ng}/\text{m}^3$.

3.3. Meteorological conditions and seasonal behaviour of PM10 and BaP

One of the factors in addition to the anthropogenic activities that contributes to PAH behaviour is the meteorological variables. A summary of the meteorological conditions as well as the mean BaP and PM10 concentrations for the

Table 1

Percentage and number of exceedances in brackets of the daily and the annual limit values, lower and upper assessment thresholds of PM10 ($\mu\text{g}/\text{m}^3$) for the four sampling sites during the cold and warm seasons. The average value of PM10 is also expressed for each season.

	ZGZ		PIR		MNG		MON	
	Cold	Warm	Cold	Warm	Cold	Warm	Cold	Warm
Daily limit value of PM10	20(3)	0	0	0	0	0	19(3)	0
Annual limit value of PM10	27(4)	0	0	0	0	0	38(6)	7(1)
Upper assessment threshold of PM10	27(4)	0	0	0	0	0	53(8)	53(8)
Lower assessment threshold of PM10	40(6)	40(6)	14(2)	0	0	1	87(13)	87(13)
Average value of PM10	30.06	21.22	18.78	10.17	6.70	17.18	36.59	32.81

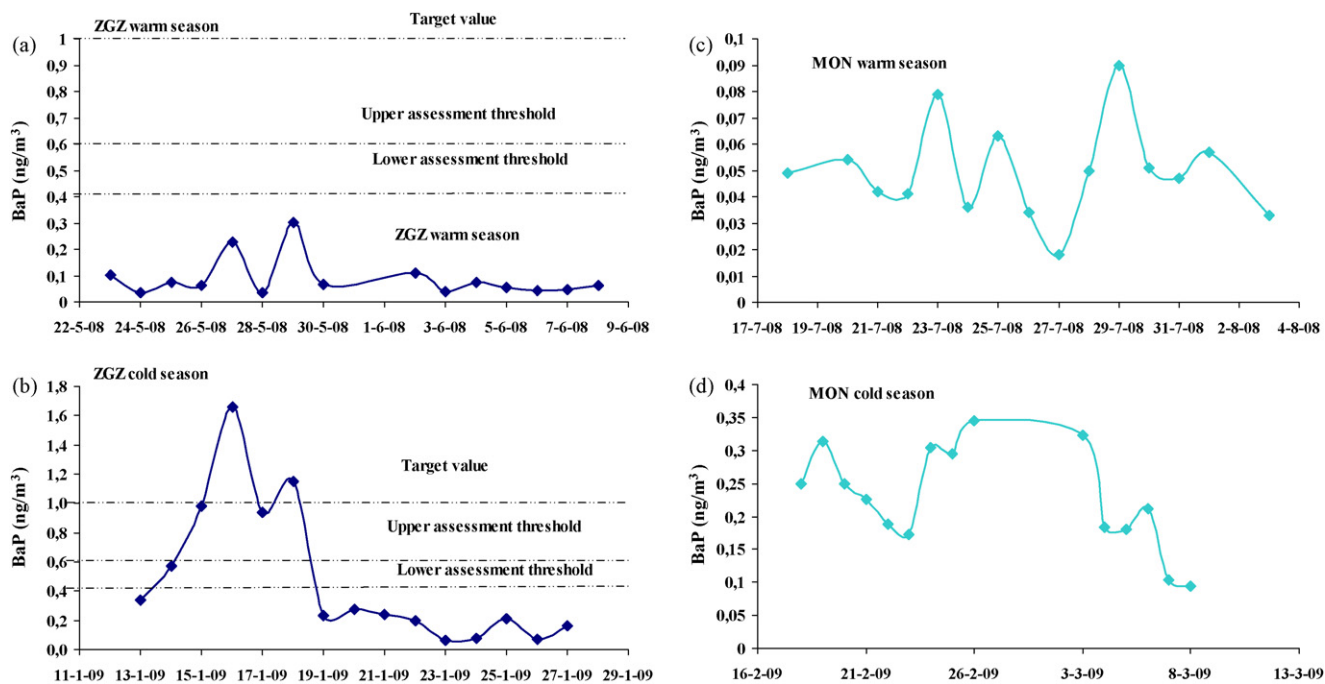


Fig. 2. Temporal evolution of BaP for ZGZ during the a) warm and b) cold seasons and for MON during the c) warm and d) cold seasons.

warm and cold period is shown in Table S1, Supplementary data.

In order to evaluate possible differences between warm and cold seasons for the two pollutants studied, PM10 and BaP, Student's *t*-tests were used (Table S2, Supplementary data). A precondition for this parametric test is to assess variance homogeneity through the Levene's test. If the significance from this test is less than 0.05, then variances are significantly different and the *t*-test result corresponding to equal variances not assumed is considered. In case that the Levene's test is not significant, it is assumed that variances are equal and the *t*-test result for equal variances assumed is considered.

With regard to PM10, *t*-test showed that PM10 concentrations were statistically different (95%) for both periods for PIR and MNG indicating that the sources producing this pollutant followed a different trend in both periods. These sources were not mainly local but also from long-range transport that was also contributing to PM10 levels. On the contrary, ZGZ and MON were mainly affected by local pollution sources that did not undergo seasonal variations.

Pearson correlation coefficients were also studied in order to determine the correlation between two variables. In this case, the correlations between the meteorological variables, PM10 and BaP concentrations were studied for each sampling point (Table S3, Supplementary data). The meteorological conditions favouring the accumulation of PM10 in ZGZ were the increase of relative humidity as well as low wind speed and cold temperatures. In MON, a negative effect on the PM10 concentration was observed with strong winds and high atmospheric pressure. It was observed that in MNG, the PM10 was positively correlated at 99% level of significance with temperature, pressure and UV radiation and negatively correlated at 99% level of significance with relative humidity and wind speed. A different influence of the temperature and the relative humidity to the one found in ZGZ was obtained for MNG in such a way that high temperature conditions, high ultraviolet radiation and low relative humidity favoured the PM10 accumulation mainly due to soil resuspension episodes. In this case, typical conditions of summer and dry weather favoured the PM10 accumulation in this place. The influence of the strong winds had the same effect than the one found in ZGZ and MON diluting the PM10

concentrations. In PIR, PM10 was only negatively correlated at 95% level with rainfall.

With regard to BaP, *t*-tests demonstrated that statistically significant differences were observed for the BaP concentrations in ZGZ and MON for both seasons (Table S2, Supplementary data). These seasonal variations of pollutants can be due to several reasons. On the one hand, the anthropogenic activities related with industrial activities seem to have less variation along the year. Emissions related to traffic and in particular to the increase in domestic heating during the cold period, increase the emissions related to combustion processes and BaP. On the other hand, the atmospheric conditions are also different in winter and summer and this can favour the higher concentrations of pollutants. Therefore, during cold season, low temperatures, low solar radiation, low ozone concentration and low dispersive capacity are favourable conditions for BaP condensation and adsorption onto the particulate matter, avoiding its photochemical degradation and other chemical reactions.

Pearson correlation coefficients between the BaP concentration and the meteorological variables were studied for ZGZ and MON (Table S3, Supplementary data). For both places, there was a positive correlation statistically significant at 99% level between the BaP and the PM10 concentrations reflecting that sources producing these pollutants were the same. The relative humidity was also positively correlated with the BaP. Authors as Gustafson and Dickhut [45] also observed that during the rainy season the high relative humidity favoured the PAH concentration on the particulate matter due to PAH deposition effects. In 1996, Harrison et al. [46] reported positive correlations of PAH with the humidity, in particular, for PAH released by combustion sources such as traffic emissions.

The meteorological variables which were negatively correlated with the BaP concentrations were the temperature, the solar radiation, the UV radiation (only for MON) and the wind speed (only for ZGZ). In winter season, the number of solar radiation hours is lower which implies an increase in the PM10 and BaP concentrations. In addition, the high relative humidity, typical from foggy days and calm winds also favour the accumulation of these pollutants. Therefore, human exposure to atmospheric pollution should be avoided

Table 2Regression results for airborne BaP (ng/m³) by applying the best-MLR model to each data set (95th percentile confidence limits).

Coefficient	Variable	ZGZ	MON	All sites
A	Intercept	-5.660 (-7.129, -4.192)	110.488 (82.637, 138.338)	118.358 (82.052, 154.663)
B	Ln PM10	0.798 (0.405, 1.192)	0.697 (0.276, 1.118)	0.700 (0.433, 0.966)
C	Ln T		-19.984 (-25.124, -14.844)	-21.296 (-27.937, -14.655)
D	Ln Radiation		-0.365 (-0.733, 0.002)	-0.388 (-0.738, -0.037)
E	Ln Wind speed	-0.742 (-1.148, -0.337)		-0.227 (-0.456, -0.002)
F	Season	1.375 (0.989, 1.761)		
	N	30	31	61
	R ² (R ² adjusted)	0.794 (0.815)	0.877 (0.890)	0.796 (0.809)
	RMSE	0.505	0.300	0.451

Coefficients are significant at $p < 0.05$.

during these meteorological conditions in order to reduce harmful effects on human health.

In MON, the same meteorological conditions previously named for ZGZ also favoured the BaP accumulation. The only difference was related to the ultraviolet radiation in which the increase of such radiation influenced decreasing the pollutant concentration. PIR and MNG showed minimum BaP concentrations lower than the detection and/or quantification limits for most of the sampled dates and these correlations did not show any statistical meaning.

3.4. MLR model

Because the sampling and analysis of PAH implies different steps which consume time, it is important to develop methods capable to estimate BaP concentrations based on real measurements. One of the simplest ways to predict concentrations of pollutants is by using statistical models. In this case and with the aim of predicting BaP concentrations, a multiple linear regression (MLR) model was applied to the experimental BaP data by using SPSS software based on PM10 concentrations and meteorological variables.

Results of applying the best-MLR model to each individual place (ZGZ, MON) and to data from the two sampling sites (all sites model) are shown in Table 2 where N is the number of data points. The models were run both with and without transformation of variables (with the exception of the season and rainfall variables) by natural logarithmic transformation obtaining a higher R^2 and significance with the transformation. The variable temperature (°C) was modified to K. The PM10, the temperature, the solar radiation and the wind speed were the significant variables considered in the all sites model. When models were obtained for each individual site, ZGZ and MON, some of these variables were not significant. With regard to statistical parameters obtained for each model, it was found that R^2 coefficients were quite similar for the three models with the highest one for MON and the lowest RMSE also for MON. For the three MLR models, all the BaP concentrations estimated by each model were within the 95th percentile confidence interval.

3.5. Validation of the model

Firstly it was necessary to choose the best-fit model before doing the whole validation. In order to reach this aim, the all sites model and the MON model, which showed the best parameters of fitting, were taken as the possible best-fit models. The model obtained with all sites ($\text{Ln BaP} = 118.358 + 0.70 \cdot \text{Ln PM10} - 21.296 \cdot \text{Ln T} - 0.388 \cdot \text{Ln Radiation} - 0.227 \cdot \text{Ln Windspeed}$) (Table 2) was applied to predict the BaP concentrations for each individual place and Q_{CV}^2 were obtained.

By applying this, the Q_{CV}^2 was 0.798 for ZGZ and 0.848 for MON obtaining good cross validation (Table 3). Afterwards, the model with the highest R^2 and lowest RMSE corresponding to MON ($\text{Ln BaP} = 110.488 + 0.697 \cdot \text{Ln PM10} - 19.984 \cdot \text{Ln T} - 0.365 \cdot \text{Ln Radiation}$) was applied to ZGZ and to two sites data obtaining the following

Table 3

Cross validation parameters resulting from applying the all sites MLR model to ZGZ, MON and 13 random data and the MON MLR model to ZGZ and two sites data set.

	ZGZ	MON	2 SITES	13 DATA
N	30	31	61	13
All sites model				
Q_{CV}^2	0.798	0.848		0.813
RMSE	0.498	0.336		0.415
MAE	0.378	0.265		0.327
MBE	-0.002	0.024		0.145
PRESS/SSY	0.202	0.152		0.187
MON model				
Q_{CV}^2	0.760		0.802	
RMSE	0.565		0.449	
MAE	0.443		0.338	
MBE	-0.131		-0.071	
PRESS/SSY	0.240		0.198	

parameters: $Q_{CV}^2 = 0.760$ for ZGZ and $Q_{CV}^2 = 0.802$ for the two sites data. The application of the all sites model allowed obtaining a good correlation coefficient slightly over-predicting the BaP concentrations (MBE = 0.002) so that finally, cross validation was performed by randomly selecting a set of 13 data from two sites data set (61) and predicting the BaP concentrations obtained with the all sites MLR model. The PRESS/SSY was $0.183 < 0.4$ showing that all sites MLR model was a reasonable model (Table 3).

Once the model was internally validated, an external validation was carried out by applying the all sites MLR model to a set of 87 samples from previous PAH campaigns carried out during 2001–2004 [12,13]. The linear regression analysis between the experimental and the predicted values is shown in Fig. 3. Initially, outliers and influential points were removed from the data set by considering standardized residuals greater than three standard deviations units. A total of twelve points were rejected, most of them corresponding to summer dates in which experimental BaP concentrations were much lower than the predicted by

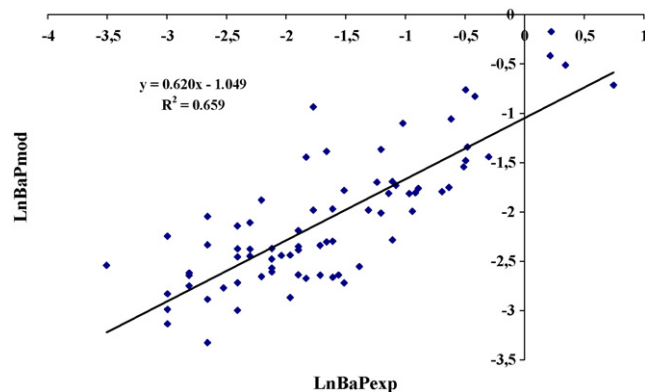


Fig. 3. Results of linear regression analysis between the experimental and predicted Ln BaP obtained with the all sites model applied to 2001–2004 sampling data set.

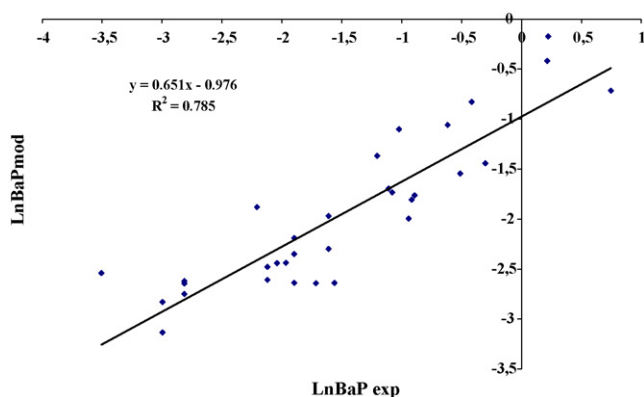


Fig. 4. Results of linear regression analysis between the experimental and predicted LnBaP obtained with the all sites model applied to 2001–2002 sampling data set.

the model so that the final set data was 75. A $Q_{\text{ext}}^2 = 0.551$ and $R^2 = 0.659$ slightly higher than 0.5 was obtained with $\text{RSME} = 0.666$ and $\text{PRESS/SSY} = 0.449$. By comparing these parameters with previous results regarding the all sites model and the cross validation (Table 3), it can be deduced that although Q_{ext}^2 is higher than 0.3 and it is a significant model, the PRESS/SSY is slightly higher than 0.4. In addition, a model with a good predictivity will show similar RMSE values for the training model, the cross validation and the external validation. Therefore, this model did not fulfil the requirements of a good external validation. However, it was able to predict more than half of the BaP concentrations.

A new external validation was also carried out by considering only the data corresponding to the sampling 2001–2002 ($N = 32$). The linear regression analysis between the experimental and the predicted values by the all sites model is shown in Fig. 4. In this case, it was obtained that $Q_{\text{ext}}^2 = 0.679$, $R^2 = 0.785$ with $\text{PRESS/SSY} = 0.321$ and $\text{RMSE} = 0.667$. The correlation coefficient slightly increased with a reduction in the PRESS/SSY fulfilling the requirements for a good external validation. Because most of the outliers corresponded to warm season in which, experimental BaP concentrations differed remarkably from the predicted concentrations, it could be deduced that the influence of the high temperature during summer in both sampling places favoured the volatilization and photodegradation of BaP. This difficulty in predicting summer time concentrations was recently reported by Akyüz and Çabuk [19].

As mentioned in the introduction, there is not a lot of bibliography regarding estimation of BaP or PAH concentrations by using MLR models [19,20]. In this way, this MLR model constitutes a first step in developing BaP predictions in urban locations with predominance of anthropogenic sources using PM10 concentrations as well as meteorological variables. In our case, the robustness of the model was carried out by internal and external validation by considering different statistical indicators used to provide a general indication between the observed and the predicted data. More research should be led in order to improve the method and to validate it with data proceeding of other urban locations.

4. Conclusions

The evolution of the PM10 and BaP concentrations in four locations of North-East of Spain (Aragón) during two sampling periods: cold and warm seasons, has been studied. Results allowed deducing that the highest PM10 concentrations were obtained in urban areas with predominance of anthropogenic sources corresponding to MON followed by ZGZ, independently of the season. High relative humidity, low winds and cold temperature, typical conditions of winter season, are the meteorological conditions favouring the

PM10 accumulation in ZGZ and MON. MNG followed a different trend showing maximum PM10 concentrations in the warm period.

The highest BaP concentrations were obtained in ZGZ followed by MON for both sampling periods, reaching the maximum concentrations during the cold season. Both locations corresponded to urban locations with predominance of anthropogenic sources: traffic and industry at which the domestic heating was added during the cold period. A seasonal behaviour of BaP was obtained in ZGZ and MON and the meteorological conditions favouring the BaP concentrations in both places were the ones obtained for the PM10. The guideline value of BaP established by Directive 2004/107/EC [11] of 1.0 ng/m^3 was only exceeded in ZGZ during the cold season for the 13% of the dates.

Regarding the model, a statistical model based on MLR was studied to predict PAH concentrations in urban areas based on data corresponding to two sampling sites in Aragón. This proposed model had got good stability, robustness and predictivity when verified by internal and external validation, in particular when applied to data from 2001–2002 sampling campaign. For the external validation, those dates corresponding to warm season in which, low experimental concentrations of BaP were obtained, seemed not to fit very well to predicted values by the model developed. High temperatures, favouring the volatility and photodegradation of PAH, could explain the difficulty in predicting summer BaP concentrations. Nevertheless, it could be concluded that this method could be used as initial estimative tool to reduce the number of studies in air pollution samplings.

Acknowledgements

Authors would like to thank Aula Dei-CSIC (R. Gracia) and the Government of Aragon (DGA) for providing the meteorological data and the CSIC for the JAE post-doct contract of J.M.L.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhazmat.2010.04.085.

References

- [1] A.M. Mastral, M.S. Callén, A review on polycyclic aromatic hydrocarbon (PAH) emissions from energy generation, *Environ. Sci. Technol.* 34 (2000) 3051–3057.
- [2] S.K. Pohjola, M. Lappi, M. Honkanen, L. Rantanen, K. Savela, DNA binding of polycyclic aromatic hydrocarbons in a human bronchial epithelial cell line treated with diesel and gasoline particulate extracts and benzo(a)pyrene, *Mutagenesis* 18 (2003) 429–438.
- [3] W.J. Deng, P.K.K. Louie, W.K. Liu, X.H. Bi, J.M. Fu, M.H. Wong, Atmospheric levels and cytotoxicity of PAHs and heavy metals in TSP and PM2.5 at an electronic waste recycling site in southeast China, *Atmos. Environ.* 40 (2006) 6945–6955.
- [4] IARC (International Agency for Research on Cancer), Benzo(a)pyrene, in: IARC Monographs on the evaluation of carcinogenic risk of the chemical to man Certain Polycyclic Aromatic Hydrocarbons and Heterocyclic Compounds, vol. 3, World Health Organization, Lyon, France, 1973, 91–136.
- [5] IARC (International Agency for Research on Cancer), Benzo(a)pyrene, in: IARC Monographs on the evaluation of carcinogenic risks to humans. Polynuclear Aromatic Compounds. Part 1 Chemical, Environmental and Experimental Data, vol. 32, World Health Organization, Lyon, France, 1983, 33–224.
- [6] CERCLA Priority List of hazardous substances. Available at: <http://www.atsdr.cdc.gov/cercla/> (accessed November 18, 2009).
- [7] Water Framework Directive 2000/60/EC. Available at: <http://www.euwfd.com/> (accessed November 25, 2009).
- [8] Directive 76/464/EEC of 4 May 1976 on pollution caused by certain dangerous substances discharged into the aquatic environment of the Community.
- [9] Helsinki Convention, Convention on the protection of the marine environment of the Baltic Sea area, 1992.
- [10] UN-ECE, Protocol to the 1979 convention on long range transboundary air pollution on persistent organic pollutants and executive body decision 1998/2 on information to be submitted and the procedure for adding substances to annexes I II or III to the protocol on persistent organic pollutants, United Nations, New York and Geneva, 1998 (ECE/EB.AIR/60).
- [11] Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and PAH in ambient air.

- [12] M.S. Callén, M.T. de la Cruz, J.M. López, R. Murillo, M.V. Navarro, A.M. Mastral, Long-range atmospheric transport and local pollution sources on PAH concentrations in a South European urban area, fulfilling of the European Directive, *Water Air Soil Pollut.* 190 (2008) 1–4.
- [13] M.S. Callén, M.T. de la Cruz, J.M. López, R. Murillo, M.V. Navarro, A.M. Mastral, Some inferences on the mechanism of atmospheric gas/particle partitioning of polycyclic aromatic hydrocarbons (PAH) at Zaragoza (Spain), *Chemosphere* 73 (2008) 1357–1365.
- [14] M. Rehwagen, A. Müller, L. Massolo, O. Herbarth, A. Ronco, Polycyclic aromatic hydrocarbons associated with particles in ambient air from urban and industrial areas, *Sci. Total Environ.* 348 (2005) 199–210.
- [15] M. Possanzini, V. Di Palo, P. Gigliucci, M. Concetta, T. Sciano, A. Cecinato, Determination of phase-distributed PAH in Rome ambient air by denuder/GC-MS, *Atmos. Environ.* 38 (2004) 1727–1734.
- [16] K. Prevedouros, E. Brorström-Lundén, C.J. Halsall, K.C. Jones, R.G.M. Lee, A.J. Sweetman, Seasonal and long-term trends in atmospheric PAH concentrations: evidence and implications, *Environ. Pollut.* 128 (2004) 17–27.
- [17] A.M. Mastral, J.M. López, M.S. Callén, T. García, R. Murillo, M.V. Navarro, Spatial and temporal PAH concentrations in Zaragoza, Spain, *Sci. Total Environ.* 307 (2003) 111–124.
- [18] W.-J. Lee, Y.-F. Wang, T.-C. Lin, Y.-Y. Chen, W.-C. Lin, C.-C. Ku, J.-T. Cheng, PAH characteristics in the ambient air of traffic-source, *Sci. Total Environ.* 159 (1995) 185–200.
- [19] M. Akyüz, H. Çabuk, Meteorological variations of PM_{2.5}/PM₁₀ concentrations and particle-associated PAH in the atmospheric environment of Zonguldak, Turkey, *J. Hazard. Mater.* 170 (2009) 13–21.
- [20] A.B. Lobscheid, T.E. McKone, D.A. Vallero, Exploring relationships between outdoor air particulate associated polycyclic aromatic hydrocarbon and PM_{2.5}: a case study of benzo(a)pyrene in California metropolitan regions, *Atmos. Environ.* 41 (2007) 5659–5672.
- [21] J.M. López, M.S. Callén, R. Murillo, T. García, M.V. Navarro, M.T. de la Cruz, A.M. Mastral, Levels of selected metals in ambient air PM₁₀ in an urban site of Zaragoza (Spain), *Environ. Res.* 99 (2005) 58–67.
- [22] SIAR network, Servicio de Información Agroclimática para el Regadío. Available at: <http://www.mapa.es/siar> (accessed November 3, 2009).
- [23] R.R. Hocking, The analysis and selection of variables in linear regression, *lometrics* 32 (1976).
- [24] N. Draper, H. Smith, *Applied Regression Analysis*, second ed., John Wiley & Sons, New York, 1981.
- [25] SAS Institute Inc., *SAS/STAT User's Guide*, Version 6, vol. 2, fourth ed., SAS Institute Inc., Cary, NC, 1989.
- [26] L. Ericsson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs, *Environ. Health Perspect.* 111 (2003) 1361–1375.
- [27] S.M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, John Wiley & Sons, USA, 1987.
- [28] A. Chaloulakou, M. Saisana, N. Spyrellis, Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, *Sci. Total Environ.* 313 (2003) 1–13.
- [29] M.W. Gardner, S.R. Dorling, Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmos. Environ.* 34 (2000) 21–34.
- [30] S.I.V. Sousa, F.G. Martins, M.C. Pereira, M.C.M. Alvim-Ferraz, Prediction of ozone concentrations in Oporto city with statistical approaches, *Chemosphere* 64 (2006) 1141–1149.
- [31] I. Kahn, D. Fara, M. Karelson, U. Maran, P.L. Andersson, QSPR treatment of the soil sorption coefficients of organic pollutants, *J. Chem. Inform. Model.* 45 (2005) 94–105.
- [32] P. Gramatica, E. Giani, E. Papa, Statistical external validation and consensus modeling: A QSPR case study for K_{oc} prediction, *J. Mol. Graph. Model.* 25 (2007) 755–766.
- [33] S. Wold, Validation of QSAR's, *Quant. Struct. -Act. Relat.* 10 (1991) 191–193.
- [34] L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, D.M. Sheehan, QSAR models using a large diverse set of estrogens, *J. Chem. Inform. Comput. Sci.* 41 (2001) 186–195.
- [35] Directive 1999/30/EC of 22 April 1999 relating to limit values for sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air. (1999), OJL 163, 29.6.1999, p.41.
- [36] Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.
- [37] M. Viana, Levels, composition and origin of the atmospheric particulate matter in the North and East of the Iberian Peninsula and Canary Islands (in Spanish), Doctoral Thesis, Universitat de Barcelona, 2003.
- [38] E. Manoli, D. Voutsas, C. Samara, Chemical characterization and source identification/apportionment of fine and coarse air particles in Thessaloniki, Greece, *Atmos. Environ.* 36 (2002) 949–961.
- [39] B. Artiñano, X. Querol, P. Salvador, S. Rodríguez, A. Alastuey, Assessment of air-borne particulate matter in Spain in response to the new EU-directive, *Atmos. Environ.* 35 (2001) S43–S53.
- [40] J.M. López, M.S. Callén, T. García, R. Murillo, M.V. Navarro, A.M. Mastral, Spatial and temporal evolution of atmospheric PAH concentration in Zaragoza, Spain, *Sci. Total Environ.* 307 (2003) 111–124.
- [41] M.T. de la Cruz, Study of the organic and inorganic components of the PM₁₀ in Zaragoza city. Application of receptor models (in Spanish), Doctoral Thesis, Universidad de Zaragoza, 2008.
- [42] P. Salvador, B. Artiñano, D.G. Alonso, X. Querol, A. Alastuey, Identification and characterisation of sources of PM₁₀ in Madrid (Spain) by statistical methods, *Atmos. Environ.* 38 (2004) 435–447.
- [43] S. Rodríguez, X. Querol, A. Alastuey, E. Mantilla, Origin of high summer PM₁₀ and TSP concentrations at rural sites in Eastern Spain, *Atmos. Environ.* 36 (2002) 3101–3112.
- [44] R. Tsiouridou, D. Voutsas, Th. Kouimtzi, Ionic composition of PM₁₀ in the area of Thessaloniki, Greece, *Chemosphere* 52 (2003) 883–891.
- [45] K.E. Gustafson, R.M. Dickhut, Particle/gas concentrations and distributions of PAHs in the atmosphere of southern Chesapeake Bay, *Environ. Sci. Technol.* 31 (1997) 140–147.
- [46] R.M. Harrison, D.J.T. Smith, L. Luhana, Source apportionment of atmospheric polycyclic aromatic hydrocarbons collected from an urban location in Birmingham, UK, *Environ. Sci. Technol.* 30 (1996) 825–832.