

Polycyclic Aromatic Hydrocarbons Pollution in a Coastal Environment: the Statistical Analysis of Dependence to Estimate the Source of Pollution

Pasquale Sarnacchiaro¹, Sergi. Díez² and Paolo Montuori^{3,*}

¹Unitelma Sapienza University of Rome, V.le Regina Elena n° 295, Roma 00161, Italy

²Environmental Chemistry Department, IDAEA-CSIC, Jordi Girona, 18–26, 08034 Barcelona, Spain

³Department of Preventive Medical Sciences, University of Naples “Federico II”, Via Sergio Pansini n° 5, Naples 80131, Italy

Abstract: Polycyclic Aromatic Hydrocarbons (PAHs) are a group of carcinogenic contaminants widespread in the environment. PAHs are produced by both anthropogenic and natural processes. Difficulties exist in identifying their origins. This paper reports a practical application of Principal Component Analysis (PCA) and Principal Component Regression (PCR) to identify the pyrolytic, petrogenic and diagenetic sources of PAH pollution in the Sarno River and Estuary. Nicknamed “the most polluted river in Europe”, the Sarno River originates in south-western Italy and has a watershed of about 715 km². PCA indicated that the PAH contamination in the Sarno River and Estuary resulted from a mixed pattern. The first principal component (PC1) had significant positive loading in high molecular weight PAHs. This profile of PAH usually includes products of high temperature combustion/pyrolytic processes, reflecting the effects of traffic pyrolysis. The second principal component (PC2) had significant positive loading in two-to-four ring PAHs. So, PC2 may be considered as components from petrogenic sources. PC3 was characterized by a high loading of perylene, thought to originate from diagenetic alteration of perylenequinone pigment or some other organic matter. Therefore, this factor can be considered as natural-origin PAHs. In the PCR, the regression coefficients for components 1-3 were 66.6, 40.4 and 19.5, respectively. In this application, the PCR was a very useful statistical technique for handling the problem of multicollinearity. Results from the application of PCR have been compared with Partial Least Square (PLS) and no significant differences were reported in the prediction errors and latent variables available by PCR and PLS.

Keywords: Polycyclic aromatic hydrocarbons, Coastal environment, Contaminant transport processes, Principal component analysis, Principal Component Regression, Multicollinearity.

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are a group of lipophilic contaminants widespread in the environment. This class of compounds has been widely studied [1, 2] because of its carcinogenic and mutagenic properties [3-5].

PAHs are produced by both anthropogenic and natural processes and can be introduced into the environment through various routes. Anthropogenic inputs can originate from the incomplete combustion of organic matter (pyrolytic) and the discharge of crude oil-related material (petrogenic). The pyrogenic origin includes domestic and industrial wastewater discharges as well as atmospheric fallout of vehicle exhaust and industrial stack emission, combustion of fossil fuels, waste incineration and coke production, carbon black, coal tar pitch, asphalt and petroleum cracking [6]. Spillage of fossil fuels can be accidental or intentional (e.g. deballasting water discharges). PAHs can also originate from natural processes such as short-term diagenetic degradation of biogenic precursors (diagenesis). Direct PAH biosynthesis

by organisms such as bacteria, fungus and algae has not yet been clarified [7]. Each source (i.e. pyrolytic, petrogenic and diagenetic) gives rise to characteristic PAH patterns. In general, combustion products are dominated by relatively high molecular weight (HMW) compounds with four or more condensed aromatic rings, whereas bi- and tricyclic aromatic compounds (LMW) are more abundant in fossil fuels, which are, moreover, dominated by alkylated derivatives [8]. Diagenetic PAHs are represented by some characteristic species (e.g. perylene). Using chemical profiling and particular compound ratios, it is possible to get an idea of the process that generates these compounds [9-11]. The ratios of phenanthrene to anthracene (Phe/An) and fluoranthene to pyrene (Fl/Pyr) have been used to assess the contribution of petroleum or combustion sources of PAHs in coastal environments. A Phe/An ratio lower than 10 and a Fl/Pyr ratio higher than 1 have been shown in the majority of cases, strongly supporting the pyrogenic origin of PAHs [12]. In addition to this, when the ratio of chrysene to benzo(a)anthracene (Chr/BaA) is lower than 0.1, it is usually an indication of a pyrolytic origin, while a fluorene to fluorene + pyrene (Flu/(Flu + Pyr)) ratio lower than 0.5 is typical of a petrogenic source. However, difficulties exist in identifying these origins owing to the possible co-existence of several variables related to different sources.

*Address correspondence to this author at the Department of Preventive Medical Sciences, University of Naples “Federico II”, Via Sergio Pansini n° 5, Naples 80131, Italy; Tel: +390817463027; Fax: +39081746 3352; E-mail: pmontuor@unina.it

Currently, the interest in multivariate statistical methodologies for identifying the main sources of PAH pollution and for quantifying the incidence of each source of pollution on total pollution levels, particularly in coastal environments, is increasing [13-15]. Diagnostic ratios provide only qualitative information on the contributions of various sources to total PAH levels. Quantitative assessments can be done with several statistical approaches such as Principal Component Analysis (PCA) and Multiple Linear Regression Analysis (MLRA) [14, 15].

In many frameworks, the statistician is interested in the relationship between a set of variables, called predictors, and one response variable, both defined for the same statistical units. In regression, the objective is to explain the variation in one response variable by associating this variation with proportional variation in one or more predictor variables. When strong correlations exist between the predictor variables, the ordinary least squares estimators have large variances and, therefore, the estimates are often large and may have signs that disagree with known theoretical properties of the variables. In the statistical literature, several methods have been proposed to counter the multicollinearity problem [20], for example Ridge Regression (RR), Principal Component Regression (PCR), Partial Least Square Regression (PLS) and Principal Component Analysis onto References subspaces (PCAR) [21]. PCR shows moderate similarity with the other methods, probably because it transforms the predictor variables on the orthonormal basis of latent variables by eliminating the collinearity between predictor variables [22].

PCR has been introduced to handle the problem of multicollinearity by eliminating model instability and reducing the variances of the regression coefficients [23]. In the first step, the methodology performs a PCA on the predictor variables, and in the second step runs a regression using the principal component scores as the predictor variables with the response variable of interest. A first question in PCR is which principal component should be kept in the regression. Many statisticians argue that the decision depends only on the magnitude λ of the variance of the principal components [24]; however, there are some examples in which the principal components of small variance must be selected [25]. The main problem of this approach is that the magnitude of the eigenvalue (λ) depends on the predictor variables only and has nothing to do with the response variable; in other words, there is no general theoretical reason that the most informative linear function of the predictor variable should lie among the dominant principal components of the multivariate distribution of the predictor variables. However, when the interpretation of Principal Components is clear and useful, PCR represents a good solution to overcome the multicollinearity problem [26, 27].

This study is part of a large project which has the objective of enhancing knowledge of pollution in the Sarno River and its environmental impact in the gulf of Naples. This project has attempted to assess the pollution derived from local industry, agriculture and urban impact through the determination of several groups of organic and inorganic chemicals and some indicators of microbial pollution in water and

sediments. In the present work, we report a practical application of PCA and PCR to identify the pyrolytic, petrogenic and diagenesis origin of PAH pollution in sediments of the Sarno River and Estuary.

EXPERIMENTAL

Study Area

Nicknamed “the most polluted river in Europe”, the Sarno River originates in south-western Italy and has a watershed of about 715 km². It flows through the Sarno flatland, delimited in the west by Mt. Vesuvius and in the east by the Lattari Mountains, and reaches the Sea at the Gulf of Naples (Tyrrhenian Sea), flowing through Pompei city (Fig. 1).

The Sarno watershed collects water from two important effluents, the Cavaiola and Solofrana torrents. The Sarno flatland is one of the most fertile in Italy due to the high quality of the soil, constituted by layers with volcanic and alluvial origins. The high human presence, the massive use of fertilisers and pesticides in agriculture and the development of industry represent the main causes of pollution in the Sarno River [16]. Agriculture is focused on tomato production in the San Marzano area. In terms of industry, Solofra, a city around the Solofrana River, boasts an ancient tradition in the leather industry that currently counts for about 400 productive units and 3500 workers. The pharmaceutical industry is represented principally by the Novartis Farma factory, placed exactly 200 m from the river mouth and covering an area of about 201,000 m². This factory is one of the largest facilities of Novartis Farma and is one of the most important in the world. Another source of environmental pollution can be attributed to urban agglomerations and their wastewaters [17-19].

Sampling Points and Sediment Sample Collection

An intensive sampling campaign was conducted in the spring of 2008 (Fig. 1). Surface sediment samples were collected at four locations along the Sarno river (near the source of the river, just before and after the junction with Alveo Comune and at the river mouth) and nine points in the continental shelf around the river mouth (three points were sampled 50 m from the Sarno River mouth, another three points 150 m away and, finally, another three points 500 m from the river mouth).

Surface sediment (0-20 cm) samples were collected by using a grab sampler (Van Veen Bodemhappe 2 L) and put in aluminium containers. The sediments were transported refrigerated to the laboratory and kept at -20°C before analysis.

Sediment Sample Extraction

Sediments were oven dried at 60°C and sieved at 250 μ m. Then, 5 g of sediment were spiked with the surrogate mixture (10 ng of anthracene-*d*₁₀, pyrene-*d*₁₀ and perylene-*d*₁₂) and extracted three times by sonication using 15 mL of DCM/methanol (1:1) (Carlo Erba, Milano, Italy) for 15 min. The recovered extracts were combined and dried with anhydrous sodium sulphate (Carlo Erba). Then, they were concentrated in a rotary evaporator to 0.5 mL and solvent-exchanged to hexane (Carlo Erba). Extract fractionation was

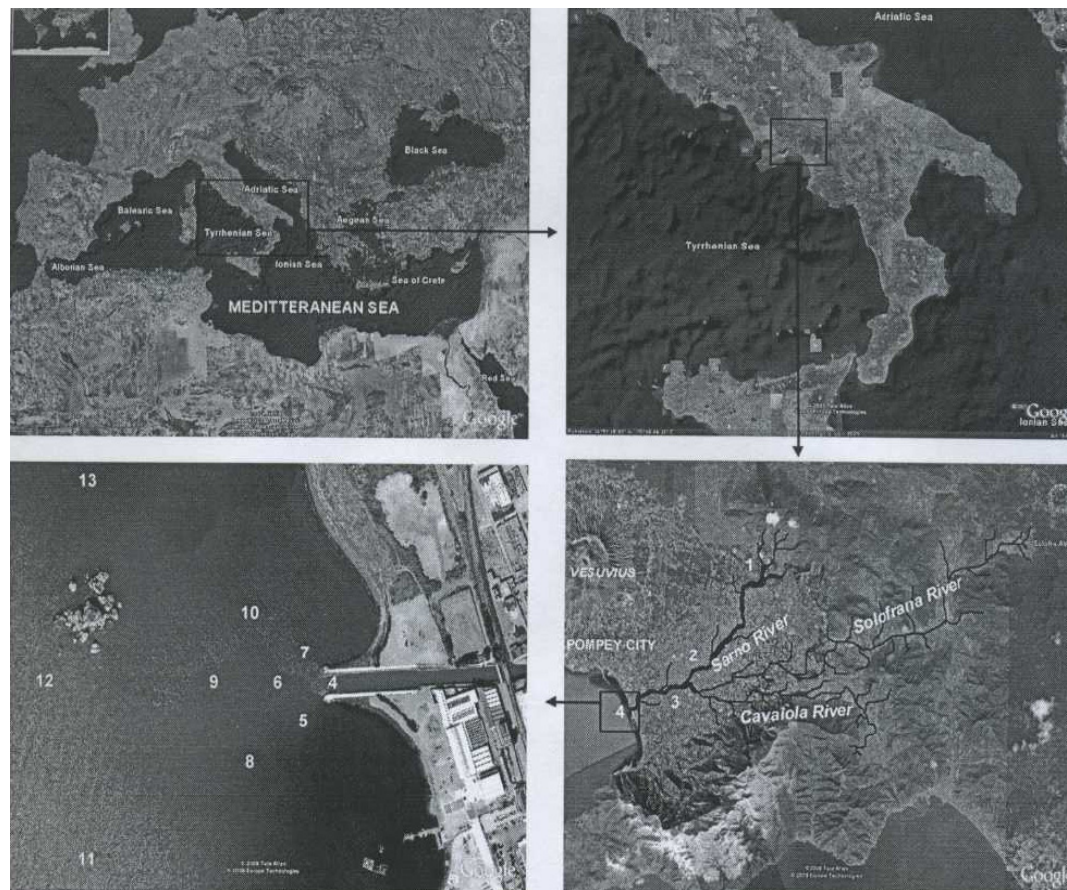


Fig. (1). Map of the study areas and sampling sites in the Sarno River and Estuary, Southern Italy (from Google Earth).

carried out by open column chromatography (3 g of neutral alumina Carlo Erba, deactivated with 3% (w/w) Milli-Q water). Three fractions were collected: fraction I with 5.5 mL of hexane (Carlo Erba), fraction II with 6 mL of hexane:ethylacetate (9:1) (Carlo Erba) and, finally, fraction III with 12 mL of ethylacetate (fraction III or F3). PAHs were eluted in fraction II, while fractions I and III contained other organic pollutants which were also determined in the samples.

Analytical Determination of PAHs

Cleaned extracts of fractions II were analysed on a GC-MS QP5050A Shimadzu (Kyoto, Japan) working in the electron impact mode at 70 eV. A SPB 20 chromatographic column (20% diphenyl 80% dimethylpolysiloxane) (60 m, 0.25 mm ID and 0.25 μm of film thickness) was used. The initial oven temperature was temperature programmed at 50°C (2 min) to 250°C at 10°C min^{-1} , holding for 30 min. The MS transfer line and ion source were kept at 250°C. Acquisition was carried out in the single ion monitoring mode (SIM) using two characteristic ions for each target analyte. Compound identification was carried out by comparing retention times with standards and using the characteristic ions and their ratio for each target analyte. Furthermore, for the higher

concentrated samples, the identification of target analytes was confirmed in full-scan mode (m/z range from 60 to 350). The concentrations were calculated from the calibration curves for the 17 PAHs, and a good chromatographic linearity ($R^2 > 0.98$) of the Standard Reference Material (Dr. Ehrenstorfer GmbH, Augsburg, Germany) was found. Triphenylamine was used as the internal standard to compensate for the sensitivity variation of the MS detector. In each sample, the concentration of the following sixteen selected PAHs monitored by the US Environmental Protection Agency (USEPA) as priority pollutants and perylene were measured: naphthalene (Nap), acenaphthylene (Acy), acenaphthalene (Ace), fluorene (Flu), phenanthrene (Phe), anthracene (An), fluoranthene (Fl), pyrene (Pyr), benz(a)anthracene (BaA), chrysene (Chr), benzo(b) fluoranthene (BbF), benzo(k)fluoranthene (BkF), benzo(a)pyrene (BaP), dibenz(a,h)anthracene (DahA), benzo(ghi)perylene (BghiP), indeno (1,2,3-cd) pyrene (InD) and perylene. Total PAH concentrations were calculated as the sum of the concentrations of the 16 PAH compounds (ΣPAHs) (Table 1).

Quality Assurance and Quality Control

The limit of detection (LOD) and limit of quantification (LOQ) were calculated as having signal-to-noise ratios of

Table 1. Description of the Sampling Sites and Concentration of PAHs in the Sediments of the Sarno River and the Continental Shelf, Southern Italy

| Sampling Location | | | Σ PAHs (ng g ⁻¹ dry wt) |
|-------------------------------|---------------------------------|--------------------------------|----------------------------------------------|
| Site Number Identification | Site Characteristics | Site Location | |
| 1 (river water) | Sarno River Source | 40°48'54.03"N 14°36'45.36"E | 5.6 |
| 2 (river water) | Upstream Alveo Comune | 40°46'42.73"N 14°34'00.48"E | 101.3 |
| 3 (river water) | After Alveo Comune | 40°46'00.34"N 14°33'10.68"E | 205.2 |
| 4 (river water) | Sarno River Mouth | 40°43'42.62"N 14°28'07.89"E | 352.4 |
| 5 (sea water) | River Mouth at 50mt South | 40°43'40.11"N 14°28'06.45"E | 446.6 |
| 6 (sea water) | River Mouth at 50mt Central | 40°43'42.46"N 14°28'05.03"E | 501.7 |
| 7 (sea water) | River Mouth at 50mt North | 40°43'45.09"N 14°28'05.17"E | 180.4 |
| 8 (sea water) | River Mouth at 150mt South | 40°43'35.68"N 14°28'02.94"E | 651.3 |
| 9 (sea water) | River Mouth at 150mt Central | 40°43'42.25"N 14°27'59.97"E | 679.2 |
| 10 (sea water) | River Mouth at 150mt North | 40°43'49.26"N 14°27'59.82"E | 161.9 |
| 11 (sea water) | River Mouth at 500mt South | 40°43'30.31"N 14°27'58.93"E | 66.5 |
| 12 (sea water) | River Mouth at 500mt Central | 40°43'42.29"N 14°27'46.41"E | 87.3 |
| 13 (sea water) | River Mouth at 500mt North | 40°43'57.85"N 14°27'48.68"E | 40.4 |

above 3 and 10, respectively, by five replicate analyses. The surrogate averaged recoveries in the sediment samples were $84.8 \pm 7.8\%$ for anthracene- d_{10} , $91.1 \pm 8.5\%$ for pyrene- d_{10} and $102 \pm 12\%$ for perylene- d_{12} . Blank assays were carried out and used for the calculation of LODs (ranged from 0.01 ng g⁻¹ to 0.15 ng g⁻¹) and LOQs (ranged from 0.03 ng g⁻¹ to 0.2 ng g⁻¹).

Statistical Analysis

Data analysis was performed with the statistical software SPSS, version 14.01 for Windows (SPSS Inc., Chicago, IL, USA) and SPAD, version 5.5 for Windows (Decisia, Pantin, France).

PCA was conducted to identify the source contributions to total PAH concentrations. PCA is a useful exploratory multidimensional data analysis technique for examining factors that contribute to data variability. Data submitted to the analysis were arranged in a matrix, where each column corresponds to one PAH component and each row represents

one sample site. The data matrix was evaluated through PCA, allowing the summarised data to be further analysed and plotted.

The statistical analysis of dependence was performed in order to determine the percentage contributions of different PAH sources to the Σ PAH for a given sediment sample. First, to analyse the dependence between the total PAH level and the seventeen selected PAHs, a preliminary study of the correlation between the predictor variables was performed (Table 2). As can be seen in Table 2, strong positive correlations were found between several pairs of predictors. When strong correlations exist between the predictor variables, it induces an inflation of the variance of the least squares estimates, increase in the magnitude of these estimates and leads to unreliable signs of the regression coefficients.

In order to detect collinearity, the variance inflation factor (VIF) were computed. VIF resulted upper than 10 for all the predictor variables, therefore a serious collinearity problems were observed. A statistical collinearity between the

Table 2. Correlation Data Matrix for PAHs

| Correlation Matrix | Nap | Acy | Ace | Flu | Phe | An | Pyr | Chr | BaA | BbF | BkF | BaP | InD | DahA | BghiP | Fl | Perylene |
|--------------------|-------|-------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| Nap | 1,00 | | | | | | | | | | | | | | | | |
| Acy | 0,88* | 1,00 | | | | | | | | | | | | | | | |
| Ace | 0,94* | 0,77* | 1,00 | | | | | | | | | | | | | | |
| Flu | 0,85* | 0,93* | 0,73* | 1,00 | | | | | | | | | | | | | |
| Phe | 0,88* | 0,88* | 0,81* | 0,79* | 1,00 | | | | | | | | | | | | |
| An | 0,92* | 0,93* | 0,84* | 0,94* | 0,82* | 1,00 | | | | | | | | | | | |
| Pyr | 0,84* | 0,98* | 0,72* | 0,96* | 0,85* | 0,90* | 1,00 | | | | | | | | | | |
| Chr | 0,90* | 0,93* | 0,79* | 0,83* | 0,97* | 0,87* | 0,89* | 1,00 | | | | | | | | | |
| BaA | 0,30 | 0,21 | 0,23 | 0,07 | 0,05 | 0,29 | 0,10 | 0,22 | 1,00 | | | | | | | | |
| BbF | 0,27 | 0,16 | 0,18 | 0,03 | 0,00 | 0,23 | 0,06 | 0,18 | 0,99* | 1,00 | | | | | | | |
| BkF | 0,27 | 0,18 | 0,15 | 0,03 | 0,06 | 0,24 | 0,08 | 0,23 | 0,98* | 0,98* | 1,00 | | | | | | |
| BaP | 0,27 | 0,17 | 0,20 | 0,06 | -0,02 | 0,26 | 0,07 | 0,15 | 0,99* | 0,99* | 0,97* | 1,00 | | | | | |
| InD | 0,24 | 0,13 | 0,16 | 0,01 | -0,02 | 0,22 | 0,03 | 0,16 | 0,99* | 0,99* | 0,98* | 0,99* | 1,00 | | | | |
| DahA | 0,22 | 0,11 | 0,14 | -0,02 | -0,04 | 0,19 | 0,01 | 0,14 | 0,99* | 1,00* | 0,98* | 0,99* | 1,00* | 1,00 | | | |
| BghiP | 0,14 | 0,04 | 0,08 | -0,08 | -0,12 | 0,13 | -0,07 | 0,07 | 0,97* | 0,98* | 0,96* | 0,98* | 0,99* | 0,99* | 1,00 | | |
| Fl | 0,29 | 0,19 | 0,21 | 0,06 | 0,04 | 0,27 | 0,09 | 0,21 | 1,00* | 1,00* | 0,99* | 0,99* | 1,00* | 1,00* | 0,98* | 1,00 | |
| Perylene | -0,23 | -0,34 | -0,40* | -0,29 | -0,38 | -0,18 | -0,37 | -0,27 | 0,40* | 0,40* | 0,47* | 0,14 | 0,47* | 0,45* | 0,48* | 0,42* | 1,00 |

p-value <0.01

independent variables can be also detected by the proximity of the smallest eigenvalues of the correlation matrix R to 0 (Table 4).

To overcome multicollinearity problems, a large number of statistical methods are available: RR, PCR, PLS and PCAR. Given the clear popularity of PLS and PCR among chemists we have compared them. These two techniques are similar in many ways and the theoretical relationship between them has been treated extensively in the literature [28]. Both methods construct new predictor variables, known as components, as linear combinations of the original predictor variables, but they construct those components in different ways. PCR creates components to explain the observed variability in the predictor variables, without considering the response variable at all. On the other hand, PLS does take the response variable into account, and therefore often leads to models that are able to fit the response variable with fewer components. Whether or not that ultimately translates into a more parsimonious model, in terms of its practical use, depends on the context.

Historically, PCR predates PLS, with the latter appearing in the chemical literature around 1983 [28]. Since its introduction, however, PLS appears by most accounts to have become the method of choice among chemists. The main reasons are: (1) PLS should predict better because correlations with the y variable are sought in determining the scores, (2) PLS requires fewer latent variables than PCR and

should therefore be more parsimonious, (3) PLS loadings are more readily interpreted, and (4) PLS should handle nonlinearities better than PCR.

In their research P.D. Wentzell and L.V. Montoto considered many papers in order to obtain a clear picture of what kinds of studies had already been done to compare these two methods and what conclusions had been drawn. Their conclusion was "The results of the study were somewhat surprising. Given the clear popularity of PLS over PCR among chemists, we had expected to see many instances where the superior predictive performance of PLS was evident. While there were a few cases which indicated that PLS gave better results than PCR, a greater number of studies indicated no real difference in performance. Additionally, there were no theoretical studies which suggested that one method should predict better than the other" (pag.3).

Also in our research, results from the application of PCR, compared with PLS, shown no significant differences in the prediction errors and latent variables available.

PCR was preferred since the interpretation of the extracted principal components had a clear reading in relation to research, in particular the first principal component had significant positive loading in high molecular weight PAHs and this profile reflects the effects of traffic pyrolysis. The second principal component had significant positive loading in two-to-four ring PAHs, then it may be considered as components from petrogenic sources. The third component was

Table 3. Concentration Ranges and Mean Value of PAHs in Sediment Samples from Different Rivers, Estuaries and Coasts in the World

| Study Area | References | N. of PAHs Analyzed | Sediments (ng G ⁻¹) | |
|------------------------------------|-------------------------------|---------------------|---------------------------------|-------|
| | | | Range | Mean |
| Asia | | | | |
| Daliao River watershed, China | Guo <i>et al.</i> (2009) | 18 | 62 - 841 | 287 |
| Qiantang River, China | Chen <i>et al.</i> (2007) | 15 | 91 - 614 | 313 |
| Lanzhou Reach, Yellow River, China | Xu <i>et al.</i> (2007) | 16 | 464 - 2621 | 1414 |
| Tonghui River, China | Zhang <i>et al.</i> (2004) | 16 | 127 - 928 | 540 |
| Kishon River, Israel | Oren <i>et al.</i> (2006) | 22 | 59 - 298 | 153 |
| Pearl River Estuary, China | Luo <i>et al.</i> (2006) | 17 | 189 - 637 | |
| Australia/Oceania | | | | |
| Sydney Harbour, Australia | McCready (2000) | 16 | <0.1 - 380.2 | |
| North America | | | | |
| Susquehanna River, USA | Ko <i>et al.</i> (2007) | 36 | | 3000 |
| Athabasca River, Canada | Headley <i>et al.</i> (2001) | 16 | 10 - 34700 | |
| Todos Santos Bay, Mexico | Macias <i>et al.</i> (2002) | 16 | 7 - 813 | 96 |
| Mediterranean sea | | | | |
| Adriatic Sea | Guzzella <i>et al.</i> (1994) | 9 | 27 - 530 | 130 |
| Adriatic Sea | Caricchia (1993) | 12 | 18 - 580 | 200 |
| Gulf of Rijeka, Croatia | Bihari <i>et al.</i> (2006) | 10 | 213 - 695 | |
| Izmir Bay, Turkey | Darilmaz (2007) | 16 | 2 - 113 | 24 |
| Tiber River, Italy | Minissi <i>et al.</i> (1998) | 13 | 4 - 625 | |
| Adriatic Sea (Chioggia), Italy | Magi <i>et al.</i> (2002) | 16 | 24 - 501 | 333 |
| Adriatic Sea (Ancona), Italy | Magi <i>et al.</i> (2002) | 16 | 33 - 306 | 184 |
| Gulf of Trieste, Italy | Notar <i>et al.</i> (2001) | 22 | 30 - 600 | |
| <i>This study</i> | | 17 | 5.6 - 679.2 | 267.7 |

characterized by a high loading of perylene and it can be considered as natural-origin PAHs.

Criteria for the selection of principal components in regression are: Student's tests of the significance of the components or the magnitude of the eigenvalues of the explanatory variables correlation matrix. The two procedures are not quite compatible and a third procedure based on a decision rule with respect to the partial correlation coefficient between the principal component and the explained variable has been proposed [29]. In this research we considered three components.

Moreover, in this study, the concentrations of 17 PAHs as active variables and 13 sediment samples as subjects were selected.

RESULTS AND DISCUSSION

The concentrations of total PAHs in sediment samples are illustrated in Table 1. In sediment samples the mean and the standard deviation of total PAHs concentrations were

267.7 and 233.4 ng g⁻¹, respectively. The detected concentrations ranged from 0.2 to 31.6 ng g⁻¹ with a mean value of 9.7 ng g⁻¹ for two-ring PAHs (Nap), from 0.2 to 46.3 ng g⁻¹ for three-ring PAHs (Acy, Ace, Flu, Phe, An), from 0.3 to 47.2 ng g⁻¹ for four-ring PAHs (Fl, Pyr, BaA, Chr), from 0.2 to 46.6 ng g⁻¹ for five-ring PAHs (BbF, BkF, BaP, DahA) and from 0.5 to 46.7 ng g⁻¹ for six-ring PAHs (BghiP, InD).

Compared with other polluted rivers in the world (Table 3), the concentrations of \sum PAHs in sediment samples from the Sarno River (5.6 - 352.4 ng g⁻¹) were close to those found in the Daliao River watershed (China) by Guo *et al.* (2009) and in the Qiantang River (China) by Chen *et al.* (2007) [28,29]. Values slightly higher than those found in the Sarno River were found in the Tiber River (Italy) by Minissi *et al.* (1998) and in the Tonghui River (China) by Zhang *et al.* (2004) [32, 33], while much higher concentrations were encountered in the Lanzhou Reach of the Yellow River (China) by Xu *et al.* (2007), in the Susquehanna River (USA) by Ko *et al.* (2007) and in the Athabasca River (Canada) by Headley *et al.* (2001) [15, 34, 35]. The concentrations of \sum PAHs

Table 4. Principal Component Analysis – Eigen Values Table

| Eigenvalues Table | | | |
|---------------------|------------|------------|-----------------------|
| Principal Component | Eigenvalue | Percentage | Percentage Cumulative |
| 1 | 8,6968 | 51,16% | 51,16% |
| 2 | 6,7744 | 39,85% | 91,01% |
| 3 | 0,6997 | 4,12% | 95,12% |
| 4 | 0,3997 | 2,35% | 97,47% |
| 5 | 0,3124 | 1,84% | 99,31% |
| 6 | 0,0542 | 0,32% | 99,63% |
| 7 | 0,0317 | 0,19% | 99,82% |
| 8 | 0,0218 | 0,13% | 99,95% |
| 9 | 0,0052 | 0,03% | 99,98% |
| 10 | 0,0031 | 0,02% | 99,99% |
| 11 | 0,0008 | 0,00% | 100,00% |
| 12 | 0,0002 | 0,00% | 100,00% |
| 13 | 0,0000 | 0,00% | 100,00% |
| 14 | 0,0000 | 0,00% | 100,00% |
| 15 | 0,0000 | 0,00% | 100,00% |
| 16 | 0,0000 | 0,00% | 100,00% |
| 17 | 0,0000 | 0,00% | 100,00% |

in sediment samples from the Sarno River were higher than those found in the Heman Reach of the Yellow River, China (2009) and the Kishon River, Israel (2006) [36, 37]. Σ PAHs in sediments samples from the continental shelf of the Sarno River (40.3 - 679.2 ng g⁻¹) were close to those found in the Pearl River Estuary (China) by Luo *et al.* (2006), in the Gulf of Rijeka, Croatia (Bihari *et al.*, 2006) and in the Gulf of Trieste, Italy (Notar *et al.*, 2001) [13, 14, 38]. The concentrations of Σ PAHs in sediment samples from the continental shelf of the Sarno River were higher than those found in Sydney Harbour, Australia (2000), in Todos Santos Bay, Mexico (2002), in Izmir Bay, Turkey (Darilmaz *et al.*, 2007) and in the Adriatic Sea, Italy (1993, 1943 and 2002) [2, 6, 39-42]. The concentrations of Σ PAHs in sediment samples from the continental shelf of the Sarno River were lower than those found in the Venice Lagoon by Manodori *et al.* (2006) [43].

The ratio study reflected a prevailing pattern of pyrolytic inputs of PAHs in the Sarno River and Estuary [8-12]. In fact, the results showed that the Phe/An ratio was <10 at all sites (mean 1.19; range 0.43-3.62), and Flu/Pyr was >1 at most sites (mean 1.33; range 0.74-4.01). Chr/BaA and Flu/Flu + Pyr varied within a range of 0.12 - 2.07 (mean 0.88) and 0.34 - 1.44 (mean 0.72), respectively. Thus, the LMW/HMW ratio was relatively low (<1 for most sites), suggesting a pyrolytic origin of PAHs at these sites (mean 0.90; range 0.24-3.99).

In addition to pyrolytic and petrogenic sources, perylene is also produced by in situ degradation of biogenic precursors

[12, 44]. Indeed, perylene is probably the most important diagenetic PAH encountered in sedimentary environments and, thus, a high abundance of perylene relative to other PAHs can indicate an important natural origin of the compound. The highest concentrations of perylene were observed at location 3 (after Alveo Comune) and between the River and the Estuary. The clearly seaward-decreasing trends of concentrations of perylene suggest that perylene originated mainly from terrestrial inputs. Perylene has been frequently associated with inputs from rivers and estuaries [12, 45]. These authors have suggested that concentrations of perylene which are higher than 10% of the total penta-aromatic isomers indicate a probable diagenetic input, whereas those in which perylene comprises less than 10% indicate a probable pyrolytic origin of the compound. In the present study, perylene was studied in all sediment samples, and concentrations of this compound ranged from 0.12 to 30.3 ng g⁻¹. The concentrations of perylene relative to the penta-aromatic isomers indicated that the relative abundance of some of them were lower than 10%, indicating a pyrolytic origin of the compounds that could reach the river from different sources in this area.

PCA, as mentioned earlier, was used to separate PAHs that had similar sources and modes of input. The majority of the variance (95%) of the scaled data was explained by three eigenvectors-principal components (Table 4).

The first principal component (PC1) explained 51.2% of the total variance, and the second (PC2) and the third (PC3) represented 39.9% and 4.1% of the total variance, respec-

Table 5. Principal Component Analysis – Variable Coordinates

| Variable Coordinates | Axes 1 | Axes 2 | Axes 3 |
|----------------------|--------|--------|--------|
| Nap | 0,64 | 0,72 | 0,00 |
| Acy | 0,56 | 0,80 | -0,05 |
| Ace | 0,53 | 0,70 | 0,24 |
| Flu | 0,43 | 0,83 | -0,19 |
| Phe | 0,41 | 0,85 | -0,01 |
| An | 0,62 | 0,73 | -0,15 |
| Fl | 0,61 | 0,77 | 0,05 |
| Pyr | 0,46 | 0,84 | -0,10 |
| Chr | 0,57 | 0,76 | -0,07 |
| BaA | 0,92 | -0,39 | 0,06 |
| BbF | 0,90 | -0,43 | 0,09 |
| BkF | 0,90 | -0,41 | -0,02 |
| BaP | 0,90 | -0,42 | 0,07 |
| InD | 0,89 | -0,45 | 0,02 |
| DahA | 0,88 | -0,47 | 0,06 |
| BghiP | 0,84 | -0,54 | 0,05 |
| Perylene | 0,24 | -0,60 | -0,74 |

tively. The coordinates for variables show the loadings for the individual PAHs (Table 5).

Along the axis PC1, all the compounds were found on the positive-coordinate side and none on the negative-coordinate side (Fig. 2), therefore it represents a “size component”. Nevertheless, the variables which had the highest influence in building this axes are BaA, BbF, BkF, BaP, InD, DahA and BghiP. These compounds gave strong correlations with the total PAH concentration ($r > 0.7$). Moreover, it had significant positive loading in high molecular weight PAHs. This profile of PAHs usually implies products of high temperature combustion/pyrolytic processes, reflecting the effects of traffic pyrolysis. PC2 had significant positive loading in two-to-four ring PAHs (Ace, Acy, Flu, An, Fl, Pyr, Phe, Chr and BkF). So, PC2 may be considered a component of petrogenic sources. This also confirms the results obtained from the ratios of specific PAH compounds in that the PAH contamination in the sediments resulted mainly from a mixed pattern of pyrolytic and petrogenic sources.

PC3 was characterised by a high loading of perylene (Fig. 3), thought to originate from diagenetic alteration of perylenequinone pigment or some other organic matter. Therefore, this factor can be considered as natural-origin PAHs.

The PCR results are presented in Table 6. The regression coefficients for Principal Components 1–3 were 66.6, 40.4 and 19.5, respectively. Thus, a mean percentage contribution of 52.6% was from the combustion of coal and wood, 31.9% from oil-derived sources, 25% from vehicle emissions, and

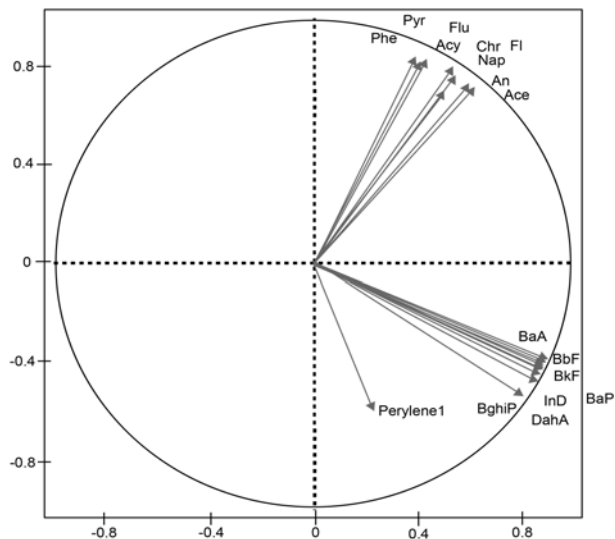


Fig. (2). Principal Component Analysis – Variables Plot (Axes 1-2).

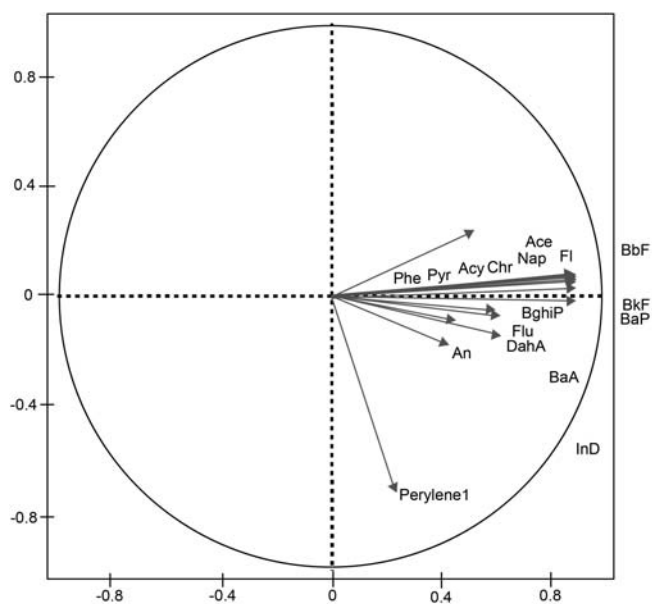


Fig. (3). Principal Component Analysis – Variables Plot (Axes 1-3).

15.4% from natural sources. Compared with the analysis by diagnostic ratios, the results of PCA and PCR are consistent with those of diagnostic ratios. This suggests that both diagnostic ratios and PCA and PCR are appropriate for PAH source identification.

These results obtained by different molecular ratios, PCA and PCR were correlated with the specific pollution conditions in the Sarno flatland. The Sarno flatland is a heavily industrialised area, with many heavily polluting factories. Industrial wastes enriched with combustion-derived PAHs are discharged directly into the Sarno River. The emission of atmospheric particles from factories and from the entire Naples urban area also cause serious air pollution, and particulate-associated PAHs may be transported and deposited

Table 6. Principal Component Regression – Coefficients

| Source | Estimate | Standard Error | T-Statistic | p-value |
|-----------------------|----------|----------------|-------------|---------|
| Comp.1 – (Pyrolytic) | 66,631 | 2,20512 | 30,2164 | 0,0000 |
| Comp.2 – (Petrogenic) | 40,3759 | 2,49963 | 16,1527 | 0,0000 |
| Comp.3 – (Diagenetic) | 19,5241 | 7,78436 | 2,5081 | 0,0334 |
| Constant | 266,884 | 6,50584 | 41,0222 | 0,0000 |

into the river. Additionally, besides these inputs, some other sources such as the roads on both sides of the river and along the coast, urban runoff containing street dust and municipal wastewater resulted in the pattern of pyrolytic origins of PAH contamination in the area.

5. CONCLUSIONS

This study was the first to document a comprehensive analysis of PAHs in the Sarno River and Estuary, and has provided very useful information for the evaluation of trace PAHs and probable sources of PAHs. Source analysis revealed that PAHs mainly came from combustion processes. PCA indicated that PAH contamination in the Sarno River and Estuary resulted from a mixed pattern of pyrolytic and petrogenic sources; however, a predominant pattern of pyrolytic inputs of PAHs was found. Particularly, the total variance in the composition of PAHs was 51.2%, 39.9% and 4.1% from pyrolytic, petrogenic and diagenic sources, respectively. The ratio study reflected a prevailing pattern of pyrolytic inputs of PAHs in sediments from the Sarno River and Estuary. In this study, PCR was a very good statistical technique for handling the problem of multicollinearity by eliminating model instability and reducing the variances of the regression coefficients. The PCR coefficients for the pyrolytic, petrogenic and diagenic components were 66.6, 40.4 and 19.5, respectively. Thus, the mean percentage contribution to total PAH levels were 52.6% from the combustion of coal and wood, 31.9% from oil-derived sources and 15.43% from natural sources.

REFERENCES

- [1] Tolosa, I.; Bayona J.M.; Albaigés J. Aliphatic and polycyclic aromatic hydrocarbons and sulfur/oxygen derivatives in northwestern Mediterranean sediments: spatial and temporal variability, fluxes and budgets. *Environ. Sci. Technol.*, **1995**, *29*, 2519-2527.
- [2] Caricchia, A.M.; Chiavarini, S.; Cremisini, C.; Marrini, F.; Morabito, R. PAHs, PCBs and DDE in the northern adriatic sea. *Mar. Pollut. Bull.*, **1993**, *26*, 581-583.
- [3] Lehr, R.E.; Jerima, D.M. Metabolic activations of polycyclic hydrocarbons. *Arch. Environ. Contam. Toxicol.*, **1997**, *39*, 1-6.
- [4] Yan, L.S. Study of carcinogenic mechanisms for aromatic hydrocarbons – extended bay region theory and its quantitative model. *Carcinogenesis*, **1985**, *6*, 1-6.
- [5] White, K.L. An overview of immunotoxicology and polycyclic aromatic hydrocarbons. *Environ. Carcinog. Rev.*, **1986**, *2*, 163-202.
- [6] McCreedy, S.; Slee, G. F.; Birch, G.F.; Taylor, S.E. The distribution of polycyclic aromatic hydrocarbons in superficial sediments of Sydney Harbor, Australia. *Mar. Pollut. Bull.*, **2000**, *40*, 999-1006.
- [7] Parlanti, E. *Utilisation des hydrocarbures comme traceurs d'origine de la matière organique sédimentaire en milieu marin. Etude du Golfe de Lyon et du Golfe de Gascogne (Programme Ecomerge)*. PhD thesis Nr 495, University Bordeaux I : Bordeaux, France, **1990**.
- [8] Prah, F.G.; Carpenter, R. Polycyclic aromatic hydrocarbon associations in Washington coastal sediment. *Geochim. Cosmochim. Ac.*, **1983**, *47*, 1013-1023.
- [9] Budzinski, H.; Jones, I.; Bellock, J.; Piérard, C.; Garrigues, P. Evaluation of sediment contamination by polycyclic aromatic hydrocarbons in the Gironde estuary. *Mar. Chem.*, **1997**, *58*, 85-97.
- [10] Yunker, M.B.; Macdonald, R.W.; Vingarzan, R.; Mitchell, R.H.; Goyette, D.; Sylvestre, S. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. *Org. Geochem.*, **2002**, *33*, 489-515.
- [11] De Luca, G.; Furesi, A.; Leardi, R.; Micer, G.; Panzaneli, A.; Piu, P.C.; Sanna, G. Polycyclic aromatic hydrocarbons assessment in the sediments of the porto torres harbor (Northern Sardinia, Italy). *Mar. Chem.*, **2004**, *86*, 15-32.
- [12] Baumard, P.; Budzinski, H.; Michcon, Q.; Garrigues, P.; Burgeot, T.; and Bellocq, J. Origin and bioavailability of PAHs in the Mediterranean sea from mussel and sediment records. *Estuar. Coast Shelf. Sci.*, **1998**, *47*, 77-90.
- [13] Luo, X.J.; Chen, S.J.; Mai, B.X.; Yang, Q.S.; Sheng, G.Y.; Fu, J.M. Polycyclic aromatic hydrocarbons in suspended particulate matter and sediments from the pearl river estuary and adjacent coastal areas, China. *Environ. Pollut.*, **2006**, *139*, 9-20.
- [14] Bihari, N.; Fafandel, M.; Hamer, B.; Kralj-Bilen, B. PAH content, toxicity and genotoxicity of coastal marine sediments from the rovinj area, northern adriatic, croatia. *Sci. Total. Environ.*, **2007**, *366*, 602-11.
- [15] Xu, J.; Yu, Y.; Wang, P.; Guo, W.; Dai, S.; Sun, H. Polycyclic aromatic hydrocarbons in the surface sediments from yellow river, china. *Chemosphere*, **2007**, *67*, 1408-14.
- [16] Arienzo, M.; Adamo, P.; Bianco, M.R.; Violante, P. Impact of land use and urban runoff on the contamination of the sarno river basin in southwestern italy. *Water Air Soil Pollut.*, **2001**, *131*, 349-366.
- [17] De Pippo, T.; Donadio, C.; Guida, M.; Petrosino, C. The case of sarno river (southern italy): effects of geomorphology on the environmental impacts. *Environ. Sci. Pollut. Res. Int.*, **2006**, *13*, 184-91. Review.
- [18] Legambiente. Sarno, l'inquinamento preoccupa. Sondaggio di Legambiente sul disinquinamento del fiume. Available from: <http://www.legambienteonline.it/news2001/sarno3.htm>
- [19] Novartis Pharma. Salute, Sicurezza e Ambiente. Available from: <http://www.novartis.it/downloads/media/salute2002.pdf>
- [20] Frank, I.E.; Friedman, J.H. A statistical view of chemometrics regression tools. *Technometrics*, **1993**, *35*, 109-148.
- [21] Norliza, A.; Maizah, H.A.; Robiah, A. A comparative study on some methods for handling multicollinearity problems. *Matematika*, **2006**, *22*, 109-119
- [22] Massy W. F. Principal components regression in exploratory statistical research. *J. Am. Stat. Assoc.*, **1965**, *60*, 234-246.
- [23] D'Ambra, A.; Sarnacchiaro, P. Some data reduction methods to analyze the dependence with highly collinear variables: a simulation study. *Asian J. Math. Stat.*, **2010**, *3*(2), 69-81.
- [24] Filzmoser P.; Croux C. In: *Classification, Clustering and Data Analysis*. Jajuga K.; Sokolowski A.; Bock H; ed. Springer, **2002**, 227-234.
- [25] Hocking R.R. *Methods and Applications of Linear Models: Regression on and the Analysis of Variance*, John Wiley & Sons: USA, **1996**.
- [26] Joliffe I.T. A note on the use of principal components in regression. *Appl. Stat.*, **1982**, *31*, 300-303.

- [27] Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate Analysis. Probability and Mathematical Statistics*. Academic Press: London, **1979**.
- [28] Wentzell, P.D.; Vega Montoto, L.; Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemom. Intell. Lab. Syst.*, **2003**, *65*, 257-279
- [29] Foucart, T.; A decision rule for discarding principal components in regression. *J. Stat. Plan. Inference*, **2000**, *89*, 187-195
- [30] Guo, W.; He, M.; Yang, Z.; Lin, C.; Quan, X.; Men, B. Distribution, partitioning and sources of polycyclic aromatic hydrocarbons in daliiao river water system in dry season, China. *J. Hazard. Mater.*, **2009**, *164*, 1379-85.
- [31] Chen, Y.; Zhu, L.; Zhou, R. Characterization and distribution of polycyclic aromatic hydrocarbon in surface water and sediment from qiantang river, china. *J. Hazard. Mater.*, **2007**, *6*, 148-55.
- [32] Minissi, S.; Caccese, D.; Passafiume, F.; Grella, A.; Eleonora, C.; Rizzoni, M. Mutagenicity (micronucleus test in Vicia faba root tips), polycyclic aromatic hydrocarbons and heavy metal content of sediments collected in tiber river and its tributaries within the urban area of rome. *Mutat. Res.*, **1998**, *420*, 77-84.
- [33] Zhang, Z.; Huang, J.; Yu, G.; Hong, H. Occurrence of PAHs, PCBs and organochlorine pesticides in the tonghui river of beijing, china. *Environ. Pollut.*, **2004**, *130*, 249-61.
- [34] Ko, F.C.; Baker, J.; Fang, M.D.; Lee, C.L. Composition and distribution of polycyclic aromatic hydrocarbons in the surface sediments from the susquehanna river. *Chemosphere*, **2007**, *66*, 277-85.
- [35] Headley, J.C.; Akre, C.; Conly, F.M.; Peru, K.M.; Dickson, L.C. Preliminary characterization and source assessment of PAHs in tributary sediments of the athabasca river, canada. *Environ. Forensics*, **2001**, *2*, 335-345.
- [36] Sun, J.H.; Wang, G.L.; Chai, Y.; Zhang, G.; Li, J.; Feng, J. Distribution of polycyclic aromatic hydrocarbons (PAHs) in henan reach of the yellow river, middle china. *Ecotoxicol. Environ. Saf.*, **2009**, *72*, 1614-24.
- [37] Oren, A.; Aizenshtat, Z.; Chefetz, B. Persistent organic pollutants and sedimentary organic matter properties: a case study in the kishon river, israel. *Environ. Pollut.*, **2006**, *141*, 265-74.
- [38] Notar, M.; Leskovsek, H.; Faganeli, J. Composition, distribution and sources of polycyclic aromatic hydrocarbons in sediments of gulf of trieste, northern adriatic sea. *Mar. Pollut. Bull.*, **2001**, *42*, 36-44.
- [39] Macias-Zamora, J.V.; Mendoza-Vega, E.; Villaescusa-Celaya, J.A. PAHs composition of surface marine sediments: a comparison to potential local sources in Todos Santos Bay, BC, Mexico. *Chemosphere*, **2001**, *46*, 459-468.
- [40] Darilmaz, E.; Kucuksezgin, F. Distribution and origin of hydrocarbons in surficial sediments from the eastern aegean sea (izmir bay). *Mar. Pollut. Bull.*, **2007**, *54*, 1824-30.
- [41] Guzzella, L.; De Paolis, A. Polycyclic Aromatic Hydrocarbons in sediments of the adriatic sea. *Mar. Pollut. Bull.*, **1994**, *28*, 159-165.
- [42] Magi, E.; Bianco, R.; Ianni, C.; Di Carro, M. Distribution of polycyclic aromatic hydrocarbons in the sediments of the adriatic sea. *Environ. Pollut.*, **2002**, *119*, 91-98.
- [43] Manodori, L.; Gambaro, A.; Piazza, R.; Ferrari, S.; Stortini, A.M.; Moret, I.; Capodaglio, G. PCBs and PAHs in sea-surface microlayer and sub-surface water samples of the venice lagoon (italy). *Mar. Pollut. Bull.*, **2005**, *52*, 184-92.
- [44] Venkatesan, M.I. Occurrence possible sources of perylene in marine sediments- a review. *Mar. Chem.*, **1988**, *25*, 1-27.
- [45] LaFlamme, R.E.; Hites, R.A. The global distribution of polycyclic aromatic hydrocarbons in recent sediments. *Geochim. Cosmochim. Acta*, **1978**, *42*, 289-303.