This article was published in Journal of Environmental Management, 156, 52-61, 2015
http://dx.doi.org/10.1016/j.jenvman.2015.03.019

RELATIONSHIP BETWEEN LEVELS OF POLYCYCLIC AROMATIC HYDROCARBONS IN PINE NEEDLES AND SOCIO-GEOGRAPHIC PARAMETERS

Raquel Fernández-Varela[1], Nuno Ratola[2,3,*], Arminda Alves[3] and José Manuel Amigo[4]

[1]Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.
[2]Physics of the Earth, University of Murcia, Edificio CIOyN, Campus de Espinardo, 30100 Murcia, Spain
[3]LEPABE, Department of Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[4]Department of Food Science, Spectroscopy and Chemometrics, Faculty of Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

[*]Email: nrneto@um.es

Abstract

The ability of pine needles to capture polycyclic aromatic hydrocarbons (PAHs) from the surrounding air is well known. In this work the current knowledge of this affinity will be enhanced, investigating the plausible links between the concentrations of PAHs found in pine needles collected in different sites in Portugal, and several socio-geographic variables with environmental relevance. Canonical correlation analysis (CCA) has proven to be a suitable and innovative technique to look for relationships within environmental datasets. In the current work, CCA will simultaneously include chemical information (concentration of PAHs found in pine needles) and socio-geographic information associated to the sampling areas. In order to be more robust in these conclusions, *Pinus pinea* and *Pinus pinaster* species were considered separately, allowing an accurate direct comparison between them. The information concerning the different seasons and land occupation was also taken into account. Our results demonstrate how CCA can be a useful tool in environmental impact assessment, and highlight the importance of pine needles as trustful biomonitors of the influence of socio-geographic parameters on the levels of PAHs in a given area.

Keywords

Canonical correlation, environmental pollution, PAHs, *Pinus pinaster*; *Pinus pinea*

Highlights

Pine needles can be used as trustful biomonitors.

Socio-geographic parameters influence the levels of PAHs in urban areas.

Canonical correlation analysis as a helpful tool in environmental impact assessment.

1. Introduction

Although petroleum-related sources are a major cause of the environmental presence of anthropogenic PAHs, another important contribution comes from processes such as biomass burning, volcanic eruptions and diagenesis (Wang et al., 2007). Consequently, these chemicals can be ranked as petrogenic, pyrogenic or biogenic (Dahle et al., 2003). Petrogenic PAHs are related to combusted oil, coal and their by-products (da Silva and Bícego, 2010), and mainly comprised by 2 and 3-aromatic ring structures (Baumard et al., 1998), whereas in pyrogenic PAHs the "heavier" 4-6 ring molecules predominate, and their sources are primarily anthropogenic combustion processes (e.g. car exhaust, local heating facilities, industrial-related activities), forest fires, or tar use in asphalt (Bjørseth and Ramdahl, 1983; Guillon et al., 2013; Jensen et al., 2007). Biogenic PAHs are linked to biological material or early diagenetic processes, and are characterised by a strong presence of perylene (Gocht et al., 2001).

In many recent studies, vegetation has been used as a biomonitor of atmospheric contamination to identify point sources of pollution and to determine regional and global contamination patterns. Deciduous tree leaves, lichens, mosses and coniferous needles are the most used matrices (Alfani et al., 2001; Augusto et al., 2010; Desalme et al., 2013; Holoubek et al., 2000; Piccardo et al., 2005).

In particular, pine needles of several species yielded good performances in biomonitoring studies of PAHs worldwide, due to their unique morphology and waxy content, leading to a successful uptake of semi-volatile organic compounds (Amigo et al., 2011; Ratola et al., 2010a; Ratola et al., 2010b; Ratola et al., 2011; Simonich and Hites, 1995). A useful characteristic of pine needles is that it is relatively easy to establish their age. They can grow up to several years old, depending on the species and the environmental conditions. Therefore, it has been demonstrated that pine needles contain complementary monitoring information, such as a trustworthy time-integrated pollution dataset able to produce a comprehensive assessment of levels, sources and spatio-temporal patterns of PAHs (Amigo et al., 2011; Lehndorff and Schwark, 2009a; Lehndorff and Schwark, 2009b; Navarro-Ortega et al., 2012; Noth et al., 2013; Ratola et al., 2010b; Tremolada et al., 1996). There are major difficulties in obtaining accurate estimations of the origin of the PAHs, due to the complex mixtures of individual PAHs released to the environment from different sources of emission that may be localized within a specific geographical area. This makes the quantification and the linkage of the PAHs with a given source particularly challenging (Mostert et al., 2010), although some methods like molecular ratios assessment were proposed (Yunker et al., 2002).

To overcome these barriers and improve the accuracy of source apportionment studies, chemometric tools have been widely used (Gredilla et al., 2013; Mas et al., 2010). Chemometric methods in analytics is the discipline that uses mathematical and statistical methods to obtain the maximum chemical information when analysing chemical data. Methods such as principal component analysis (PCA), widely employed in several environmental areas (Rojo-Nieto et al., 2013; Terrado et al., 2006), partial least squares-discriminant analysis (PLS-DA), artificial neural networks (ANNs) (Álvarez-Guerra et al., 2010), cluster analysis (Chen et al., 2012) or positive matrix factorisation (PMF) (Jang et al., 2013) are presented in the literature as a complement to classic univariate statistics, often unveiling concealed environmental information (Álvarez-Guerra et al., 2010; Chen et al., 2012; Gredilla et al., 2013; Jang et al., 2013; Navarro et al., 2006). Canonical correlation analysis (CCA) is an exploratory method designed to study the relationship between data matrices containing different information within the same samples, which make it especially attractive in the field of environmental monitoring (Amigo et al., 2012; Galloway et al., 2002). This allows one to link the behaviour of variables of different nature that, otherwise, would be very difficult to assess (e.g. relating metal content in water with physico-chemical variables (Amigo et al. 2012).

The objective of this study, thus, is to examine the suitability of using pine needles as biomonitors of PAH levels, relating them with several socio-geographic parameters. We consider needles of different ages in different pine species and different sampling sites, and examine parameters such as population, population density, urban waste, petroleum and gas consumption, water consumption or elevation. We use CCA to establish these coordinates and identify behavioural patterns and potential sources for the presence of PAHs.

2. Material and methods

2.1. Sampling and sample analysis

The samples were collected in different locations distributed throughout mainland Portugal, covering major industrial areas, large and small cities, rural areas and also two remote mountainous sites. Since each new shoot of needles blooming each spring is easily separated in the branch from the previous ones, it was possible to separate the needles by year of exposure, which varied according to the pine species and the site. A total of 29 sampling sites (14 urban, 5 industrial, 8 rural, 2 remote) and four sampling campaigns (one per season) yielded a field dataset of 670 samples divided into two clearly differentiated blocks: 294 samples from *Pinus pinaster* (including duplicates of each sample) and 376 samples from *Pinus pinea* (with duplicates), as reported previously (Ratola et al., 2010b). This high amount of samples was obtained since it was possible to collect needles from different ages (in our case, life spans of two to four years depending on the species and the tree). The sampling strategy, maps and the methodology for sample analysis and quantification were described in detail elsewhere (Ratola et al., 2006; Ratola et al., 2009; Ratola et al., 2010a; Ratola et al., 2010b). In brief, the needles were collected from the bottom branches (1.5 to 2 m above the ground), always from the same trees in all four campaigns, and properly conditioned until analysis. This procedure consisted firstly of an extraction by sonication for 10 min with 30 mL hexane/dichloromethane (1:1),repeated twice using fresh solvent. The extracts were combined and evaporated to almost dryness (0.5 mL, approximately) in a rotary evaporator (Flawil, Switzerland) before being further purified using a clean-up procedure with polypropylene SPE alumina cartridges (5 g, 25 mL). After conditioning the cartridges with 50 mL hexane/dichloromethane (1:1), the extract was added to the column and eluted with 50 mL hexane/dichloromethane (1:1) and 50 mL dichloromethane into a pear-shaped flask, pre-concentrated in a rotary evaporator to 0.5 mL and transferred to 2 mL amber glass vials. Extracts were then blown down under nitrogen at room temperature, and reconstituted in 1 mL of hexane into the GC-MS injection vials. Chromatographic analysis of PAHs was done with a Varian CP-3800 gas chromatograph (Lake Forest, CA, USA) coupled to a Varian 4000 mass spectrometer in electron impact mode (70 eV) and a CP-8400 auto-sampler. Injection was in splitless mode (2 μL) and the GC column was a FactorFour VF-5MS (30 m x 0.25 mm ID x 0.25 μm film thickness) from Varian. Overall recoveries were between 65 and 110% and repeatabilities below 10%.

This methodology involving a classic approach such as sonication has been used successfully in many other related studies (Augusto et al., 2010; Capuano et al., 2005; Holoubek et al., 2000; Librando et al., 2002; Piccardo et al., 2005; Tian et al., 2008), similar to the performances shown by the more recent accelerated solvent extraction (ASE). Although two studies from one group show the latter with better recoveries that the former, the opposite was seen by Ratola et al (2006). These differences may be due to the use of different pine species by both groups (*P. sylvestris* and *P. pinea*, respectively) and the still existing gaps and uncertainties pointed out by two reviews (Barber et al., 2004; Desalme et al., 2013) in the understanding of the air-vegetation partition of PAHs, despite the number of different models attempted for its description. Since the intention of the current work is to make a comparison between samples, the key concern is to have all of them passing through the same handling/extraction/quantification procedure, to avoid differences in the systematic errors associated.

Six geo-social-economic parameters with an expected environmental impact on the concentrations of PAHs, namely population (number of inhabitants), population density (people/km^2), elevation (above sea level, m), urban waste produced, (t - tonnes), water consumption (m^3) and petroleum and gas consumption (t - tonnes) were obtained for the areas of concern from the databases of the Portuguese National Statistics Institute (INE, 2013). The last four are calculated as the production/consumption of the population of the areas of each sampling site. The variables were chosen considering their availability in the database, the complete information for the chosen areas and, as mentioned above, their expected relevance concerning the generation of different PAHs.

2.2. Data treatment

The PAH concentrations obtained for all the samples were collected in a matrix **X** ($M\times16$) with as many rows as samples (M) and as many columns as number of PAHs (16). In the same way, the socio-geographical parameters were collected in a second matrix **Y** ($M\times6$) with the same number of rows as **X**, but with as many columns as the socio-geographical variables considered. Both matrices were independently normalized prior to the analysis. In our case, auto-scaling was employed for both matrices.

Canonical correlation analysis (CCA) is a well-established method that calculates the linear correlation between each one of the scores ($\mathbf{T}_X$ and $\mathbf{T}_Y$) and two independent PCA models **X** and **Y** (Amigo et al., 2012; Næs and Martens, 1984; Sharma et al., 2006) as indicated in equations (1) and (2):

$$\mathbf{X} = \mathbf{T}_X\mathbf{P}^T{}_X + \mathbf{E}_X \qquad (1)$$

$$\mathbf{Y} = \mathbf{T}_Y\mathbf{P}^T{}_Y + \mathbf{E}_Y \qquad (2)$$

Once this correlation is calculated and its significance is assessed, the loadings (**P**) for both matrices (X and Y) allow us to identify which variables for each dataset are responsible for that correlation, thus providing valuable information about the inner relationships of both matrices. The residuals (E) are included as well in both models. *P. pinea* and *P. pinaster* needles were used independently. CCA was applied using the statistics toolbox under Matlab 7.12 (The Mathworks, MA, USA).

3. Results and discussion

3.1. Correlation coefficients

The diverse affinity of different pine species towards the uptake of PAHs has already been demonstrated in previous studies, such as between *P. halepensis* and *P. pinea* (Librando et al., 2002), *P. pinaster* and *P. nigra* (Piccardo et al., 2005) and *P. pinaster* and *P. pinea* (Ratola et al., 2011). Therefore, aiming for an enhanced robustness in our conclusions, both pine species should be treated separately. Moreover, the valuable information that may be obtained from two sampling sites included in this analysis where both species could be found adjacent to each other must be taken into consideration, since it allows a direct comparison between *P. pinea* and *P. pinaster*. Consequently, these samples were divided into two datasets (one for each pine species) and their individual performance showed clear differences in the evolution of the canonical coefficients. The first three correlation coefficients set between the canonical variables showed quite high correlation factors in both species, ranging from 0.81 to 0.69 and 0.67 to 0.56 for *P. pinaster* and *P. pinea* respectively (Figure 1), with significance p values below 0.001, indicating the correlation between the chemical content of PAHs and the socio-geographic variables.

a) b)

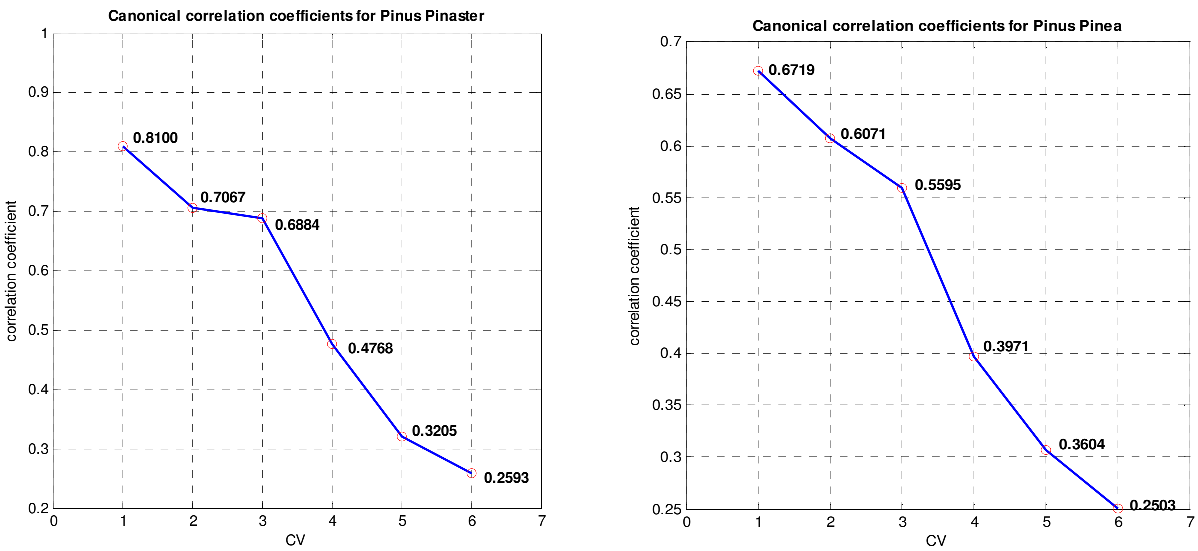


Figure 1**.** Correlations coefficients obtained for a) *Pinus pinaster* and b) *Pinus pinea*. (The p-values for the CVs are: a) 1- 1,38x10^{-111}, 2- 6,88 x10^{-66}, 3- 3,69 x10^{-37}, 4- 8,76 x10^{-12}, 5- 0,000174, 6- 0,022861; b): 1- 5,54 x10-82, 2- 1,16 x10-50, 3- 4,26 x10-28, 4- 1,06 x10-10, 5- 4,51 x10-05, 6- 0,009682).

Comparing both species, in general, higher correlations between PAH values and geo-social variables were obtained for *P. pinaster*. There is no obvious explanation for this fact, but a different behaviour of these two pine species was already highlighted in previous works, namely in the uptake of the PAHs, higher in *P. pinaster* needles (mean total PAHs from 90 to 1212 ng/g (dry weight)) than in *P. pinea* (17 to 514 ng/g (dw)); and in the affinity of both species towards individual PAHs (stronger affinity of *P. pinaster* towards the lighter PAHs (2–4 aromatic rings) and of *P. pinea* to the heavier PAHs (5 and 6 rings)) (Ratola et al., 2011). Some explanations for this fact were suggested, namely their morphology - longer needles with a thicker waxy layer and, consequently higher lipid content, makes *P. pinaster* more prone to the uptake of the lighter gas-phase PAHs. On the other hand, the higher mean surface area of *P. pinea* needles conveys a stronger affinity towards the heavier particulate-phase PAHs, which settle on the needle's surface (Ratola et al., 2011). This can be an important aspect to take into account when trying to explain potential differences in CCA results in both pine species.

To explore in depth the correlations found, the canonical coefficients (as well as the scatter plot of the canonical variables) were studied separately for each species and compared. In order to make the comparison easier, the absolute values of the canonical coefficients were used, allowing for a discussion about the influence of each parameter. Before starting the individual discussion, it is important to mention that elevation does not seem to have a significant influence on the two models. Other authors have suggested an effect of elevation, e.g. a decrease in the most chemically labile PAHs in high mountain locations with respect to sea-level areas (van Drooge et al., 2010), or the increasing accumulation of PAHs in general with elevation in Tibetan mountain slopes above 2400 m (Wang et al., 2006). However our sampling coverage of this parameter is much more limited than in those studies - the most elevated sampling point is just above 1500 m, and only two are over 1000 m – precluding a direct comparison.

3.2. Canonical correlations for *Pinus pinaster*

The first three canonical coefficients (CV) for *P. pinaster* and socio-geographical variables are depicted in Figure 2. As observed, in general, population density, urban waste and petroleum consumption have a high influence in the model and also present a strong direct correlation with each other. In fact, higher population density leads to more residue production and is likely to result in a higher fuel consumption and subsequent generation of heavy PAHs. Moreover, the mentioned parameters are influencing the levels of the lighter-to-medium molecular weight PAHs (from Naph to BbF), particularly for the 3- and 4-ring PAHs. This is indicative of fuels and vehicular traffic (diesel motorisation in particular), more typical of urban conglomerates (Hwang et al, 2003; Marr et al., 1999). In Portugal, the incidence of diesel-powered automobiles is high, even in light vehicles (about 35% of the total vehicles in 2009 and 70% of the new passenger cars acquired in 2007, according to Cames and Helmers (2013)), and traffic density is intense. According to the World Bank database (The World Bank, 2014), in the 2005-2009 period Portugal had the second highest rate of vehicles per km of road (281), only surpassed by Monaco (388) and well above the UK (77), The Netherlands (64), Greece (56) or Spain (41) and this can explain the aforementioned influence. Water consumption has a slightly smaller impact, but this parameter besides reflecting an urban presence can also reveal an extended use in rural areas. In that sense, Figure 2 shows a similarity to population in general, and not only with a strong urban pressure.

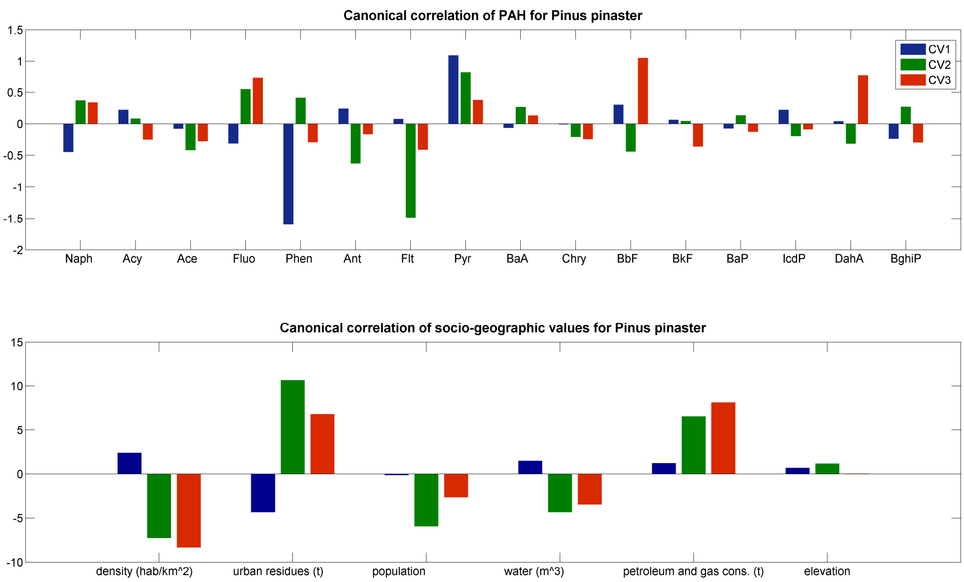


Figure 2. Canonical correlations found for the first three components (First, blue; second, green; third, red) for *P. pinaster*.

Individually, the first component (Figure 2, in blue) highlights the relationship between density, urban waste and water consumption, and the amount of three low/medium-molecular weight PAHs (Naph, Phen and Pyr). The second component (Figure 2, in green) had a quite high correlation (0.71, see Figure 1a). This component indicates that all the socio-geographical variables (except elevation for the reasons explained previously) influence the generation of PAHs, especially Flt and Pyr (associated with vehicular traffic) and, to some extent, Naph, Ace, Fluo, Phen and Ant. Moreover, the third component (Figure 2, in red) unveiled the relationship between petroleum and gas consumption, population density and urban waste, and heavier PAHs such as BbF, DahA and BghiP, and Fluo, a lighter 3-ring PAH. These high molecular weight PAHs are largely linked to important pyrogenic origins (Wang et al., 2008) such as the combustion of petroleum, wood, coal, etc., hence directly linked with petroleum and gas consumption. The PCA loading plots which represent the distribution of variables in the subspace of PAH can be seen in Ratola et al. (2010b).

To further clarify the correlations between PAH content and sociogeographic variables, scatter plots of the scores of both matrices were compared (Figures 3a-f). The scatter plots show a bivariate dataset of the scores obtained for both matrices after CCA for PAH content and socio-geographical variables, in order to identify relationships between these two data sets. The connecting lines emphasize the region occupied by the samples in a scatter plot according to the needle year, season and sampling site type. Based on the scores obtained for CV1 and CV2, high correlations can be found in the distribution of the samples according to the needle year (Figure 3a and b), season (Figure 3c and d) and sampling site type (Figure 3e and f) for both PAH content and socio-geographic variables. Concerning their age, needles exhibited an annual trend of PAH contamination related to socio-geographical variables (Figure 3a and b), explained by their generally increasing uptake capacity towards persistent atmospheric pollutants with the time of exposure to the atmosphere (Piccardo et al., 2005; Ratola et al., 2010a). Piccardo et al. (2005) found that 18-month-old pine needles showed a similar increase of Phen, Ant, Flt and Pyr concentrations compared to six-month-old needles.

Regarding the seasonal variation, a decreasing tendency in the mean total PAH concentrations can be observed from winter to summer months (834 to 596 ng/g (dw)) and then increasing from summer to autumn (to 651 ng/g (dw)) (Figures 3c and d and Ratola et al., 2011). This is given by the differences in the slopes of the respective seasons in the figures – the lower the slope, the higher the variation - and can be explained by the decrease of emitting sources (such as domestic heating and road traffic) in the warmer months, associated with an enhanced photochemical activity (Alfani et al., 2005; Amigo et al., 2011; Lehndorff and Schwark, 2009a).

The site type distribution according to socio-geographic variables (Figure 3e and f) showed an increasing tendency of the PAH concentrations from remote and rural to urban and industrial sites. This is more perceptible for the latter ones, which present a lower slope, and is in line with the findings reported by Ratola et al. (2010b), and other authors (Lehndorff and Schwark, 2009b; Lodovici et al., 1998; Wagrowski and Hites, 1997).

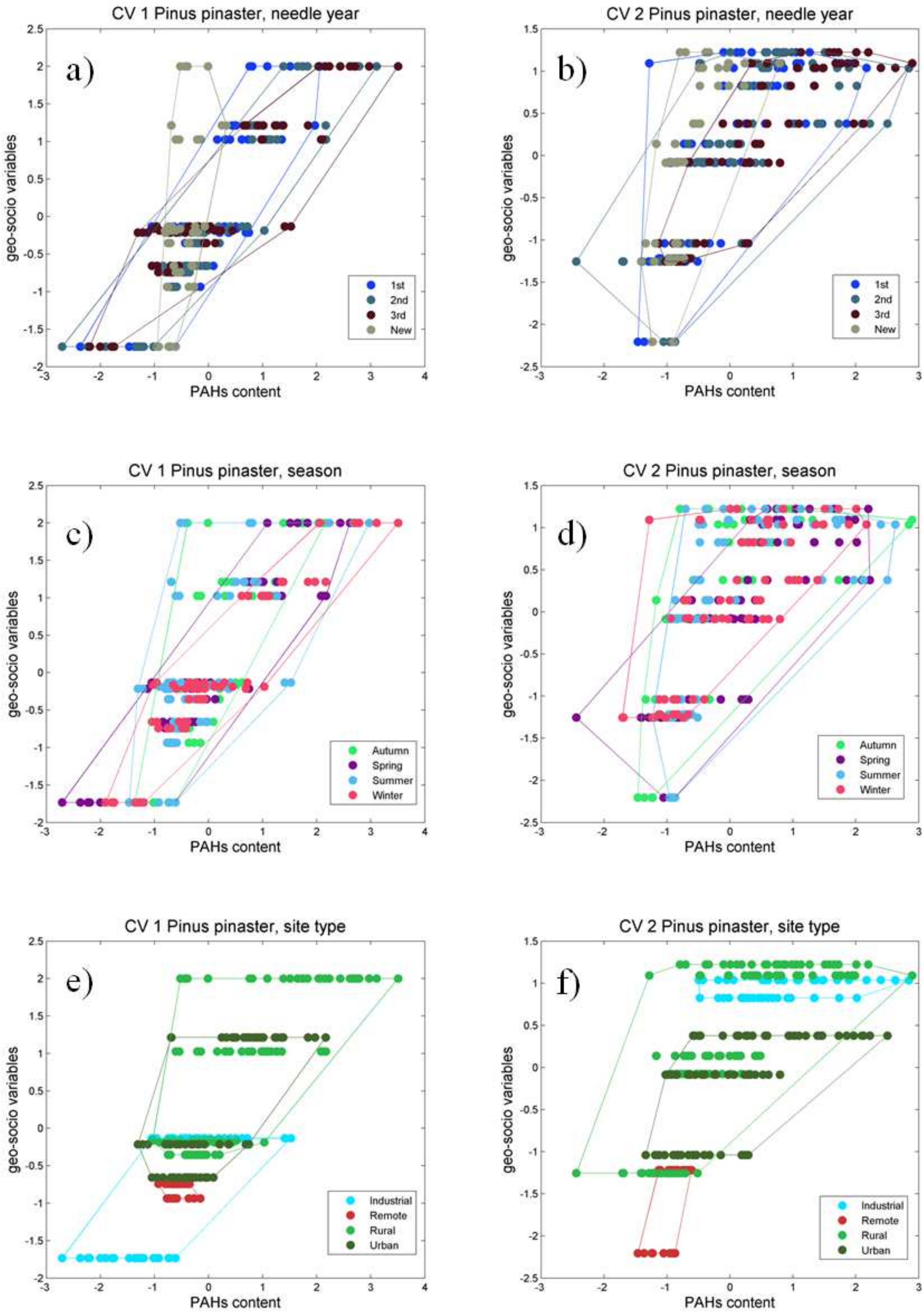


Figure 3. Scatter plots for the scores regarding a-b) needle year, c-d) season and e-f) site types for *P. pinaster*.

3.3. Canonical correlations for *Pinus pinea*

The first three canonical coefficients (CV) correlating the PAHs in *P. pinea* needles and socio-geographic variables are presented in Figure 4. Some differences can be found with respect to the *P. pinaster* results. In this case, the first component (Figure 4, in blue) shows the major influence. The link between population and urban waste, and light/medium-molecular weight PAHs, in particular Flt and Pyr, is now much clearer than with *P. pinaster* and to a lesser extent with water consumption, which was more evident in *P. pinaster* needles. Considering that the areas where *P. pinea* samples were collected had a stronger urban fingerprint (more urban sites and no remote ones), the effects of this influence can be found as deriving from higher population and, consequently, more production of waste, suggesting a stronger correlation of the pollutants to these parameters. Water consumption, also linked to rural areas is thus not such a strong influence on the correlations found.

The second component (Figure 4, in green) reflects the link between fuel consumption and urban waste, and the content of some of the lightest PAHs (Acy, Ace, Ant) and other high molecular weight PAHs (BaA, BbF, BkF). The same was verified for the third component (Figure 4, in red). It is worth mentioning the high influence on the content of Flt and Pyr (four-ring PAHs) and also 3 and 5-ring PAHs, like Fluo, BbF, BkF and BaP. This particular predominance of 4 and 5-ring PAHs in the environmental samples suggests a pyrogenic origin (Yunker et al., 2002). These results may be expected if we consider that the correlations obtained for *P. pinea* were lower than for *P. pinaster*, which were more dependent on the 5- and 6- ring PAHs.

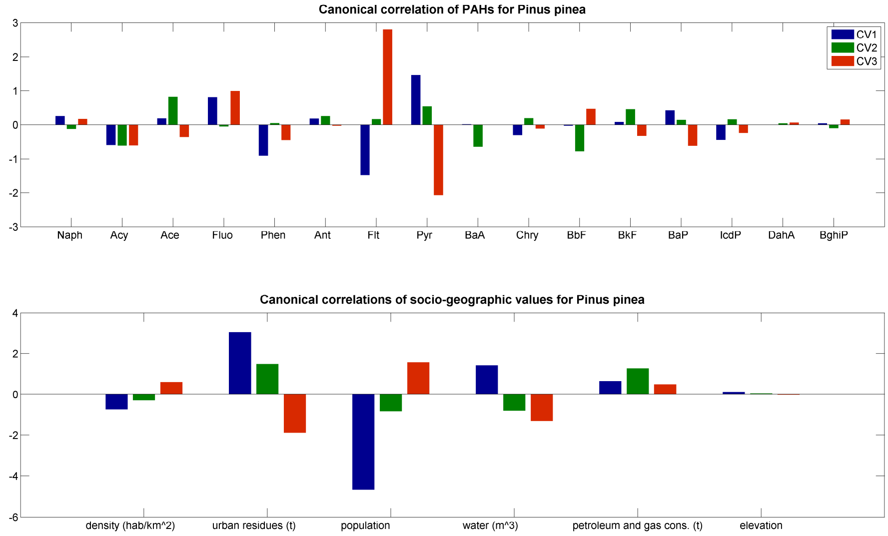


Figure 4. Canonical correlations found for the first three components (first, blue; second, green; third, red) for *P. pinea*.

As done previously for *P. pinaster*, scatter plots of the scores for CV1 and CV2 for both matrices were compared for *P. pinea* (Figures 5a-f). A similar high correlation in the distribution of the samples according to the needle year (Figure 5a and b), season (Figure 5c and d) and sampling site type (Figure 5e and f) was observed for PAH content and socio-geographic variables. However, some differences can be appreciated. For instance , in terms of needle year (Figure 5a and b), the pattern was not as clear as in *P. pinaster*. This could be due to the fact that *P. pinea* needles present, in general, a smaller size (shorter and thinner) (Correia et al., 2007) and, as a consequence of their slower growth, there may not be such a different entrapment rate between needlef of different ages. For site types, we observe that the correlations of PAHs with the socio-geographic parameters are higher than for *P. pinaster* (Figure 5e and f). This observation can perhaps be understood by the absense of remote sites and higher number of urban sites for *P. pinea* since socio-geographic parameters considered are strongly linked to urbanisation.

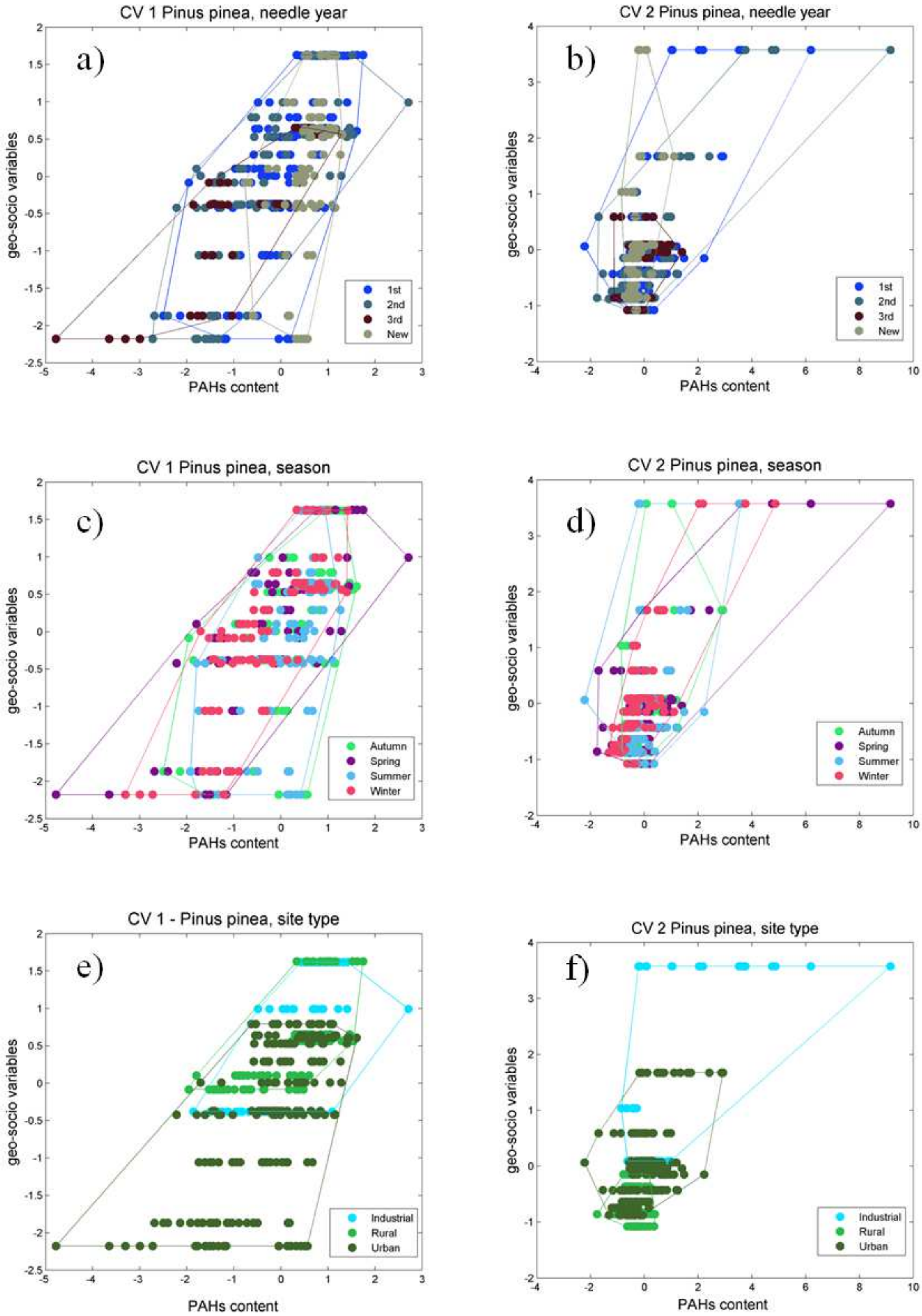
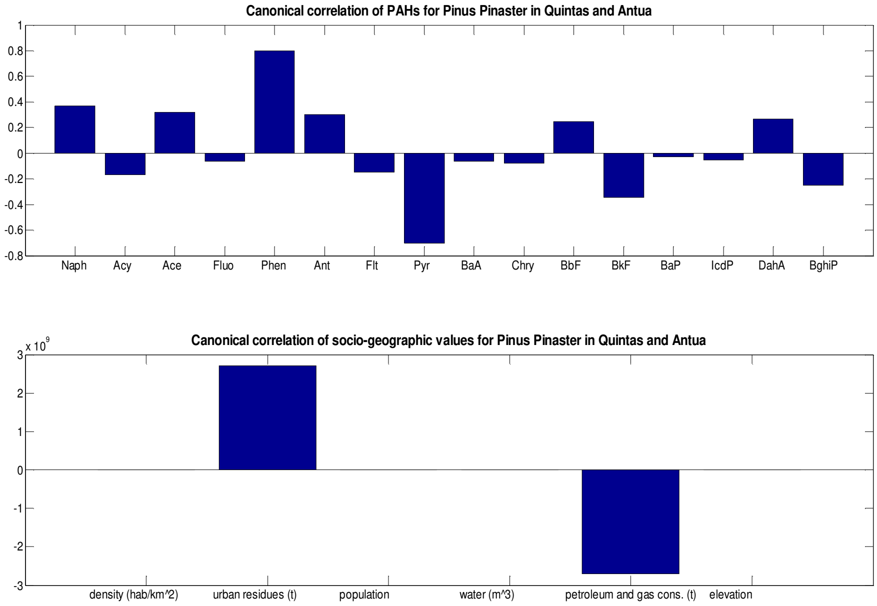


Figure 5. Scatter plots for the scores regarding a-b) needle year, c-d) season and e-f) site types for *P. pinea*.

3.4. Comparison of both pine species in adjacent locations

In previous studies using this database, it was possible to use two sampling sites with adjacent trees from both species to state clearly their uptake differences. Ratola et al. (2011) reported mean total PAH uptakes of 126 to 1666 ng/g (dw) in *P. pinaster* needles and 64 to 813 ng/g (dw) in *P. pinea*. So, it was decided to embrace the same approach for a CCA set-up. Thus, 92 pine needle samples from both species (54 *P. pinaster* and 38 *P. pinea*) were collected from adjacent trees in two sites. CV1 showed that urban waste and petroleum and gas consumption had a higher influence in the model for both pine species. This result was predictable, since if the trees are adjacent, the socio-geographic parameters affect them equally. However, the PAHs showing the strongest correlations were not exactly the same: whereas for *P. pinaster*, Phen and Pyr predominated, for *P. pinea*, Ant and BaA, followed by Phen, BbF and Chry, were the prevailing ones (see Figures 6a and b), which is in line with the uptake ability and morphological differences found between species (Ratola et al., 2011). There is, however, a noteworthy observation: for *P. pinea* the individual PAHs show in general stronger correlations with the socio-geographic parameters, which somehow goes against the findings when all samples from the different sites were compared inter-species. This example shows that biomonitoring studies have to consider the sampling strategies very carefully, since the features that ensure vegetation as an excellent matrix to monitor the atmospheric contamination are probably the ones that make its interpretation complicated. The different possible responses even within the same tree can be challenging, and a very fine statistical analysis such as that performed in this work becomes even more important to overcome such difficulties.

a)



b)

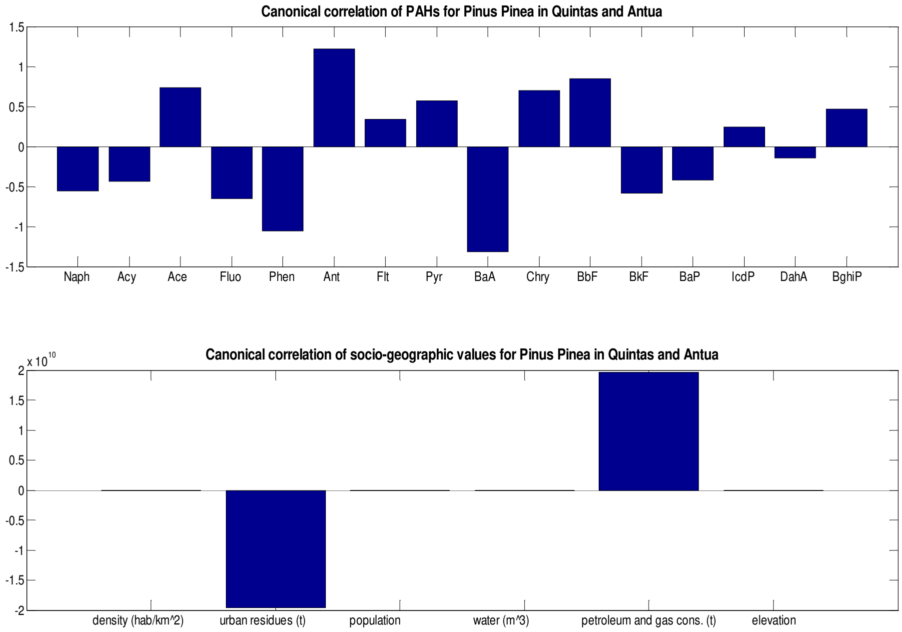


Figure 6. Canonical correlations found for the first component in two sampling sites with adjacent *Pinus pinaster* (a) and *Pinus pinea* (b) trees.

Considering each PAH individually, the most significant differences were seen for Phen (higher in *P. pinaster*), which is also the most concentrated overall, and for Flt and Pyr (higher in *P. pinea*) (Ratola et al., 2010a), closely linked to motorised traffic, probably the major source of PAH incidence in Portugal, as mentioned previously (Ratola et al., 2010a). It is worth mentioning that our study considers a higher number of urban sampling sites for *P. pinea*, prone to reflect increasing traffic contamination and therefore, the presence of those PAHs.

Taking into account the PAH incidence by number of aromatic rings (inferred from the individual PAH results found in the CCA, figures 2 and 4), significant differences between both pine species were found, especially for the 4-ring and 5-ring PAHs. In general, the most abundant were for *P. pinaster* the 4-ring and 3-ring PAHs, whereas for *P. pinea*, 4-ring PAHs prevailed, followed by 3-rings, in line with Ratola et al. (2011). These values are expected because the main PAH uptake process onto the waxy layer of the needles is gaseous absorption from the atmosphere, and most of the 3- and 4-ring PAHs occur in the vapour phase (Hwang and Wade, 2008). These low molecular weight compounds are more widely dispersed (Yang et al., 1991) and denote a potential source of long-range contamination, away from the emitting sources. The heaviest PAHs are generally removed from the air by direct (e.g., wet) deposition processes (Blasco et al., 2006), thus having a lower transportability.

Overall, all the particular features of the studied dataset highlighted the capabilities of CCA in withdrawing important links between many different parameters. The assessment of temporal and seasonal trends in biomonitoring studies using vegetation in very complicated, due to all the intrinsic aspects of the matrix that make any analytical protocol a challenge in the field of environmental chemistry. Even more when different species are included. In spite of this, CCA was able to distinguish between pine species, needles with different times of exposure to the atmospheric contaminants and land use types of the sampling areas and find singular correlations with socio-geographic indicators. Being some plant species able to “store” information for several years, the usefulness of CCA is demonstrated ever further when such a framework is planned. Obviously, in most cases, the more field data is available the better can the subsequent interpretation be performed. But large datasets are often not possible, and this is where a strong statistical tool can provide a comprehensive approach to all the potential inputs of some results to the understanding of our surrounding environment.

Conclusions

The variability in the presence of PAH content in pine needles is the result of the interaction between several environmental factors (such as the age of the needles, the season and the sampling site type). Canonical correlation analysis (CCA) allowed the identification of the most relevant features in the PAH content, showing high correlations with socio-geographic variables, such population density, production of urban waste, petroleum and gas consumption and water consumption, but not with elevation, as there was not a significant difference in the range of this parameter for the sampling sites considered. Many of the environmental datasets include a matrix of contaminants and, to some extent, correlated environmental conditions. Linking these two information sources is the main keystone to be able to describe the environmental behaviour of a specific area. CCA has widely demonstrated to be the perfect tool to study those intrinsic relationships between both sources of information. Thus, stronger assessment of environmental conditions and the influence of atmospheric conditions can be done directly and linked to the trend of the measured contaminants.

In the current study, results reinforced that *P. pinaster* and *P. pinea* needles can be used as biomonitoring tools, not only to compare the temporal and spatial distribution of airborne semi-volatile organic pollutants such as PAHs, but also to disclose important information on the influence of other related socio-geographic parameters to trace the fingerprint of individual or aromatic-ring distributed PAH behaviour. Within such a framework, CCA was ideal as a comprehensive approach to assess the relationship between PAH concentration in pine needles and socio-geographic information associated to the sample.

Acknowledgements

Raquel Fernández-Varela acknowledges the support of the European Commission under the Marie Curie Intra-European Fellowship Programme, FP7-PEOPLE-2010-IEF, 273643. Nuno Ratola and Arminda Alves wish to thank Fundação para a Ciência e a Tecnologia (FCT, Portugal) for the funding provided by the Project SAFEPINE (PTDC/AGR-CFL/102597/2008). This work has been partially funded by the European Union Seventh Framework Programme-Marie Curie COFUND (FP7/2007-2013) under UMU Incoming Mobility Programme ACTion (U-IMPACT) Grant Agreement 267143. Finally, the authors wish to thank Ramón Guitián Martínez-Piñeiro for producing the graphical abstract and Dr. Miguel Jorge for his valuable insights. Finally, we would like to thank the anonymous reviewers whose input greatly improved this manuscript.

References

Alfani, A., De Nicola, F., Maisto, G., Prati, M.V., 2005. Long-term PAH accumulation after bud break in *Quercus ilex* L. leaves in a polluted environment. Atmos. Environ. 39, 307-314.

Alfani, A., Maisto, G., Prati, M.V., Baldantoni, D., 2001. Leaves of *Quercus ilex* L. as biomonitors of PAHs in air of Naples. Atmos. Environ. 35, 3553–3559.

Álvarez-Guerra, M., Ballabio, D., Amigo, J.M., Bro, R., Viguri, J.R., 2010. Development of models for predicting toxicity from sediment chemistry by partial least squares-discriminant analysis and counter-propagation artificial neural networks. Environ. Pollut. 158, 607-614.

Amigo, J.M., Gredilla, A., Fernández-Ortiz de Vallejuelo, S., de Diego, A., Madariaga, J.M., 2012. Study of parameters affecting the behaviour of trace elements in a polluted estuary. Canonical correlation analysis as a tool in environmental impact assessment. Chemom. Intell. Lab. Syst. 119, 1-10.

Amigo, J.M., Ratola, N., Alves, A., 2011. Study of geographical trends of polycyclic aromatic hydrocarbons using pine needles. Atmos. Environ. 45, 5988-5996.

Augusto, S., Máguas, C., Matos, J., Pereira, M.J., Soares, A., Branquinho, C., 2010. Lichens as an integrating tool for monitoring PAH atmospheric deposition: a comparison with soil, air and pine needles. Environ. Pollut. 158, 483-489.

Barber, J.L., Thomas, G.O., Kerstiens, G., Jones, K.C., 2004. Current issues and uncertainties in the measurement and modelling of air–vegetation exchange and within-plant processing of POPs. Environ. Pollut. 128, 99–138.

Baumard, P., Budzinski, H., Garrigues, P., 1998. Polycyclic aromatic hydrocarbons in sediments and mussels of the western Mediterranean Sea. Environ. Toxicol. Chem. 17, 765-776.

Bjørseth. A., Ramdahl, T., 1983. *Handbook of polycyclic aromatic hydrocarbons*. Marcel Dekker Eds., New York, USA.

Blasco, M., Domeño, C., Nerín, C., 2006. Use of lichens as pollution biomonitors in remote areas: comparison of PAHs extracted from lichens and atmospheric particles sampled in and around the Somport tunnel (Pyrenees). Environ. Sci. Technol. 40, 6384-6391.

Cames, M., Helmers, E., 2013. Critical evaluation of the European diesel car boom - global comparison, environmental effects and various national strategies. Environ. Sci. Europe 25, 22 p.

Capuano, F., Cavalchi, B., Martinelli, G., Pecchini, G., Renna, E., Scaroni, I., Bertacchi, M., Bigliardi, G., 2005. Environmental prospection for PCDD/PCDF, PAH, PCB and heavy metals around the incinerator power plant of Reggio Emilia town (Northern Italy) and surrounding main roads. Chemosphere 58, 1563-1569

Chen, C.-W., Chen, C.-F., Dong, C.-D., Tu, Y.-T., 2012. Composition and source apportionment of PAHs in sediments at river mouths and channel in Kaohsiung Harbor, Taiwan. J. Environ. Monit. 14, 105-115.

Correia, A.V., Oliveira, Â.C., Fabião, A., 2007. Biologia e ecologia do pinheiro-bravo. In: *Árvores e Florestas de Portugal, vol. 4 – Pinhais e eucaliptais, a floresta cultivada*. Eds. Público, Comunicação Social, SA and Fundação Luso-Americana para o Desenvolvimento, Lisbon, Portugal, pp. 17-34.

da Silva, D.A.M., Bícego, M.C., 2010. Polycyclic aromatic hydrocarbons and petroleum biomarkers in São Sebastião Channel, Brazil: Assessment of petroleum contamination. Mar. Environ. Res. 69, 277-286.

Dahle, S., Savinov, V.M., Matishov, G.G., Evenset, A., Næs, K., 2003. Polycyclic aromatic hydrocarbons (PAHs) in bottom sediments of the Kara Sea shelf, Gulf of Ob and Yenisei Bay. Sci. Total Environ. 306, 57-71.

Desalme, D., Binet, P., Chiapusio, G., 2013. Challenges in tracing the fate and effects of atmospheric polycyclic aromatic hydrocarbon deposition in vascular plants. Environ. Sci. Technol. 47, 3967-3981.

Galloway, T.S., Sanger, R.C., Smith, K.L., Fillmann, G., Readman, J.W., Ford, T.E., Depledge, M.H., 2002. Rapid assessment of marine pollution using multiple biomarkers and chemical immunoassays. Environ. Sci. Technol. 36, 2219-2226.

Gocht, T., Moldenhauer, K.M., Püttmann, W., 2001. Historical record of polycyclic aromatic hydrocarbons (PAH) and heavy metals in floodplain sediments from the Rhine River (Hessisches Ried, Germany). Appl. Geochem. 16, 1707-1721.

Gredilla, A., Fdez-Ortiz de Vallejuelo, S., Amigo, J.M., de Diego, A., Madariaga, J.M., 2013. Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries. Trends Anal. Chem. 46, 59-69.

Guillon, A., Le Ménach, K., Flaud, P.-M., Marchand, N., Budzinski, H., Villenave, E., 2013. Chemical characterization and stable carbon isotopic composition of particulate polycyclic aromatic hydrocarbons issued from combustion of 10 Mediterranean woods. Atmos. Chem. Phys. 13, 2703-2719.

Holoubek, I., Kořínek, P., Šeda, Z., Schneiderová, E., Holoubková, I., Pacl, A., Tříska,, J., Cudlín, P., Čáslavský, J., 2000. The use of mosses and pine needles to detect persistent organic pollutants at local and regional scales. Environ. Pollut. 109, 283–292

Hubert, A., Wenzel, K.-D., Engelwald, W., Schuurmann, G., 2001. Accelerated solvent extraction - more efficient extraction of POPs and PAHs from real contaminated plant and soil samples. Rev. Anal. Chem. 20, 101-144.

Hwang, H., Terry, L.W., Sericano, J.L., 2003. Concentrations and source characterization of polycyclic aromatic hydrocarbons in pine needles from Korea, Mexico, and United States. Atmos. Environ. 37, 2259–2267.

Hwang, H.H., Wade, T.L., 2008. Aerial distribution, temperature-dependent seasonal variation, and sources of polycyclic aromatic hydrocarbons in pine needles from the Houston metropolitan area, Texas, USA. J. Environ. Sci. Health Part A 43, 1243–1251.

INE (Instituto Nacional de Estatística – Statistics Portugal) website, http://www.ine.pt, last assessed in October 2013.

Jang, E., Alam, M.S., Harrison, R.M., 2013. Source apportionment of polycyclic aromatic hydrocarbons in urban air using positive matrix factorization and spatial distribution analysis. Atmos. Environ. 79, 271-285.

Jensen, H., Reimann, C., Finne, T.E., Ottesen, R.T., Arnoldussen, A., 2007. PAH-concentrations and compositions in the top 2 cm of forest soils along a 120 km long transect through agricultural areas, forests and the city of Oslo, Norway. Environ. Pollut. 145, 829-838.

Lehndorff, E., Schwark, L., 2009a. Biomonitoring airborne parent and alkylated three-ring PAHs in the Greater Cologne Conurbation I: temporal accumulation patterns. Environ. Pollut. 157, 1323-1331.

Lehndorff, E., Schwark, L., 2009b. Biomonitoring airborne parent and alkylated three-ring PAHs in the Greater Cologne Conurbation II: regional distribution patterns. Environ. Pollut. 157, 1706-1713.

Librando, V., Perrini, G., Tomasello, M., 2002. Biomonitoring of atmospheric PAHs by evergreen plants: correlations and applicability. Polycycl. Aromat. Comp. 22, 549–559.

Lodovici, M., Akpan, V., Casalini, C., Zappa, C., Dolara, P., 1998. Polycyclic aromatic hydrocarbons in *Laurus nobilis* leaves as a measure of air pollution in urban and rural sites of Tuscany. Chemosphere 36, 1703–1712.

Marr, L.C., Kirchstetter, T.W., Harley, R.A., Miguel, A.H., Hering, S.V., Hammond, S.K., 1999. Characterization of polycyclic aromatic hydrocarbons in motor vehicle fuels and exhaust emissions. Environ. Sci. Technol. 33, 3091-3099.

Mas, S., de Juan, A., Tauler, R., Olivieri, A.C., Escandar, G.M., 2010. Application of chemometric methods to environmental analysis of organic pollutants: a review. Talanta 80, 1052-1067.

Mostert, M.M.R., Ayoko, G.A., Kokot, S., 2010. Application of chemometrics to analysis of soil pollutants. Trends Anal. Chem. 29, 430-445.

Næs, T., Martens, H., 1984. Multivariate calibration. II. Chemometric methods. Trends Anal. Chem. 3, 266–271.

Navarro, A., Tauler, R., Lacorte, S., Barceló, D., 2006. Chemometrical investigation of the presence and distribution of organochlorine and polyaromatic compounds in sediments of the Ebro River Basin. Anal. Bioanal Chem. 385, 1020-1030.

Navarro-Ortega, A., Ratola, N., Hildebrandt, A., Alves, A., Lacorte, S., Barceló, D., 2012. Environmental distribution of PAHs in pine needles, soils and sediment matrices. Environ. Sci. Pollut. Res. 19, 677-688.

Noth., E.M., Hammond, S.K., Biging, G.S., Tager, I.B., 2013. Mapping and modeling airborne urban phenanthrene distribution using vegetation biomonitoring. Atmos. Environ. 77, 518-524

Piccardo, M.T., Pala, M., Bonaccurso, B., Stella, A., Redaelli, A., Paola, G., Valério, F., 2005. *Pinus nigra* and *Pinus pinaster* needles as passive samplers of polycyclic aromatic hydrocarbons. Environ. Pollut. 133, 293–301.

Ratola, N., Lacorte, S., Alves, A., Barceló, D., 2006. Analysis of polycyclic aromatic hydrocarbons in pine needles by gas chromatography mass spectrometry: Comparison of different extraction and clean-up procedures. J Chromatogr A 1114, 198-204.

Ratola, N., Amigo, J.M., Alves, A., 2010a. Levels and sources of PAHs in selected sites from Portugal: biomonitoring with *Pinus pinea* and *Pinus pinaster* needles. Arch. Environ. Contam. Toxicol. 58, 631–647.

Ratola, N., Amigo, J.M., Alves, A., 2010b. Comprehensive assessment of pine needles as bioindicators of PAHs using multivariate analysis. The importance of temporal trends. Chemosphere 81, 1517-1525.

Ratola, N., Amigo, J.M., Oliveira, M.S.N., Araújo, R., Silva, J.A., Alves, A., 2011. Differences between *Pinus pinea* and *Pinus pinaster* as bioindicators of polycyclic aromatic hydrocarbons. Environ. Exp. Bot. 72, 339-347.

Ratola, N., Lacorte, S., Barceló, D., Alves, A., 2009. Microwave-assisted extraction and ultrasonic extraction to determine polycyclic aromatic hydrocarbons in needles and bark of *Pinus pinaster* Ait. and *Pinus pinea* L. by GC-MS. Talanta 77, 1120-1128.

Rojo-Nieto, E., Sales, D., Perales, J.A., 2013. Sources, transport and fate of PAHs in sediments and superficial water of a chronically polluted semi-enclosed body of seawater: Linking of compartments. Environ. Sci. Process. Impact. 15, 986-995.

Sharma, S.K., Kruger, U., Irwin, G.W., 2006. Deflation based nonlinear canonical correlation analysis. Chemom. Intell. Lab. Syst. 83, 34-43.

Simonich, S.L., Hites, R.A., 1995. Organic pollutant accumulation in vegetation. Environ. Sci. Technol. 29, 2095-2103.

Terrado, M., Barceló, D., Tauler, R., 2006. Identification and distribution of contamination sources in the Ebro river basin by chemometrics modelling coupled to geographical information systems. Talanta 70, 691-704.

The World Bank website, http://data.worldbank.org, last assessed in November 2014.

Tian, X., Liu, J., Zhou, G., Peng, P., Wang, X., Wang, C., 2008. Estimation of the annual scavenged amount of polycyclic aromatic hydrocarbons by forests in the Pearl River Delta of Southern China. Environ. Pollut. 156, 306-315.

Tremolada, P., Burnett, V., Calamari, D., Jones, K.C., 1996. Spatial distribution of PAHs in the U.K. atmosphere using pine needles. Environ. Sci. Technol. 30, 3570-3577.

Van Drooge, B.L., Fernández, P., Grimalt, J.O., Stuchlík, E., Torres García, C.J., Cuevas, E., 2010. Atmospheric polycyclic aromatic hydrocarbons in remote European and Atlantic sites located above the boundary mixing layer. Environ. Sci. Pollut. Res. 17, 1207-1216.

Wagrowski, D.M., Hites, R.A., 1997. Polycyclic aromatic hydrocarbon accumulation in urban, suburban and rural vegetation. Environ. Sci. Technol. 31, 279–282.

Wang, X.P., Yao, T.D., Cong, Z.Y., Yan, X.L., Kang, S.C., Zhang, Y., 2006. Gradient distribution of persistent organic contaminants along northern slope of central-Himalayas, China. Sci. Total Environ. 372, 193-202.

Wang, Z., Li, K., Lambert, P., Yang, C., 2007. Identification, characterization and quantitation of pyrogenic PAHs and other organic compounds in tire fire products. J. Chromatogr. A 1139, 14-26.

Wang, Z., Yang, C., Brown, C., Hollebone, B., Landriaul, M., 2008. A case study distinguishing pyrogenic and petrogenic hydrocarbons. International Oil Spill Conference Proceedings: Vol. 2008, No. 1, pp. 311-320.

Wenzel, K.-D., Hubert, A., Manz, M., Weissflog, L., Engewald, W., Schuurmann, G., 1998. Accelerated solvent extraction of semivolatile organic compounds from biomonitoring samples of pine needles and mosses. Anal. Chem. 70, 4827-4835.

Yang, S.Y.N., Connell, D.W., Hawker, D.W., Kayal, S.I., 1991. Polycyclic aromatic hydrocarbons in air, soil and vegetation in the vicinity of an urban roadway. Sci. Total Environ. 102, 229-240.

Yunker, M.B., Macdonald, R.W., Vingarzan, R., Mitchell, R.H., Goyette D., Sylvestre, S., 2002. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. Org. Geochem. 33, 489-515.