PERGAMON

Chemosphere 44 (2001) 125–146

CHEMOSPHERE

# Polycyclic aromatic hydrocarbons: a QSPR study

Márcia M.C. Ferreira

*UNICAMP Instituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas, SP, Brazil*

Received 9 February 2000; accepted 20 June 2000

**Abstract**

This work deals with 48 substances composed exclusively of unsubstituted six-membered fused aromatic rings. In the first step, physicochemical properties which are relevant in environmental studies such as the boiling temperature ($T_b$), the retention index (RI), *n*-octanol/water partition coefficient ($K_{OW}$) and solubility ($S$) are related with a series of electronic, geometric and topological descriptors. Among them are: electron affinity, the difference between electron affinity and ionization potential (GAP), Wiener, and connectivity indexes, volume, surface area, length-to-breadth ratio and enthalpy of formation. In a second step, these properties were incorporated into the descriptor matrix to build several quantitative structure–property relationships and to obtain prediction rules for the $\log K_{OC}$, $\log K_{OA}$, bioconcentration factor (BCF) and Henry's law constant ($H$). Finally, the photo-induced toxicity of freshwater organism *Daphinia-Magna* is modeled using the following transformed electronic descriptors: electron affinity, ionization potential and Gap. 

*Keywords:* PAHs; Boiling point; Retention index; $K_{OW}$; $K_{OC}$; $K_{OA}$; Henry's law constant; Bioconcentration factor; Phototoxicity; Chemometrics

## 1. Introduction

The polycyclic aromatic hydrocarbons (PAH) have been, for a long time, a focus of great attention by the scientific community due to their impact on public health and the environment. Some of these compounds such as benz[a]anthracene, chrysene, dibenz[a,h]anthracene and benzo[a]pyrene are mutagens and carcinogens (IARC, 1987; Jacob, 1996). Usually, the PAHs are introduced into the environment as a result of anthropogenic activities, which have increased dramatically in the last 20 years. PAHs have been detected in the atmosphere, water, soil, sediments and food (Ariese et al., 1993; Arfsten et al., 1996; Faber and Heijmans, 1996; Tolosa et al., 1996; Salau et al. 1997; Boehm et al., 1998; Franz and Eisenreich, 1998).

It is well established that the fate of PAHs in the environment is primarily controlled by their physicochemical properties. Since their aqueous solubility, volatility (e.g., Henry's law constant or air/water partition coefficient, octanol/air partition coefficient), hydrophobicity or lipophilicity (e.g., *n*-octanol/water partition coefficient) vary widely (Mackay and Callcot, 1998), the differences among their distributions in aquatic systems, atmosphere and soil are significant. On the other hand, these compounds are also quite involatile, and have relatively low vapor pressure and resistance to chemical reactions. As a result they are persistent in the environment and show a tendency to accumulate in biota, soils, sediments, and are also highly dispersed by the atmosphere.

Consequently, meaningful health and environmental assessment requires reliable data on their physicochemical properties. Unfortunately, very often these informations cannot be found in the literature, and frequently the reported data have poor precision and accuracy, specially when dealing with hydrophobic and involatile chemicals. These are mainly due to experimental difficulties from

*E-mail address:* marcia@iqm.unicamp.br (M.M.C. Ferreira).

0045-6535/01/$ - see front matter 
PII: S0045-6535(00)00275-7

procedures such as preparing, handling and analyzing the solutions when low concentrations are involved (i.e., on trace levels – ppb).

With the advent of inexpensive and rapid computation however, there has been a remarkable growth in interest in the area of quantitative structure–property relationships (QSPR), which uses multivariate methods to model relevant properties as a function of molecular structure parameters (called descriptors). A large number of descriptors have been proposed in the literature. The nature of the most commonly used descriptors (structural, topological, electronic and geometric) and the extent to which they encode structural molecular features that are related to some specific physical property is in the heart of any QSPR study. There is, in the literature, a huge amount of QSPR studies and among them, several which include the PAHs. Warne et al. (1990) used molecular descriptors and physicochemical properties to model solubility and octanol/water partition coefficient. Gerstl (1990) and Karickhoff (1981) used water solubility to estimate the organic carbon soil sorption coefficient ($K_{OC}$) for PAHs. The same authors also correlated $K_{OC}$ and *n*-octanol/water partition coefficient ($K_{OW}$). Hong et al. (1996) predicted the soil sorption coefficient by reversed phase HPLC retention time. Govers et al. (1984) and Sabljic et al. (1995) used topological index for $K_{OC}$ prediction. A good recent review on the subject is given by Gawlik et al. (1997). However, a few models for $K_{OC}$ estimation of PAHs were found using more than one descriptor at a time. Some works relate the Henry's law constant or air–water partition coefficient ($K_{AW}$) to structural molecular descriptors but usually focusing PCBs and pesticides. Bamford et al. (1999) measured the Henry's law constant ($H$) at different temperatures and relates $\log H$ with molar volume. De Maagd et al. (1998) also measured $H$ besides aqueous solubilities and $K_{OW}$. Their results were related with molar volume. The octanol/air partition coefficient ($K_{OA}$), a key property to approach the atmosphere and terrestrial organic phase, is approximately taken as to be inversely proportional to vapor pressure or the ratio $K_{OW}/K_{AW}$ as Mackey and Callcot (1998) points out and no QSPR works were found for it. The correlation of bioconcentration factor (BCF) with $K_{OW}$ is also approached by Mackay (1982). Govers et al. (1984), modeled BCF and partition coefficients, using molecular weight and molecular connectivity indices. Lu et al. (1999) also used connectivity indices to estimate BCF.

Newsted and Giesy (1987) modeled the photo-induced toxicity using a nonlinear function of the singlet and triplet excited states energies calculated from spectroscopic data. Another QSAR study was carried also by Mekenyan et al. (1994), where the authors used the HOMO–LUMO gap calculated by quantum chemistry to relate with the photo-induced toxicity.

This work deals with 48 substances composed exclusively of unsubstituted six-membered fused aromatic rings (benzenoid hydrocarbons). In a first step, physicochemical properties which are relevant in environmental studies such as the boiling temperature ($T_b$), the retention index (RI), *n*-octanol/water partition coefficient ($K_{OW}$) and solubility ($S$) are related with the descriptors (**X** matrix). In a second step, these properties were incorporated into the descriptor matrix to build several quantitative structure–property relationships and to obtain prediction rules for the $\log K_{OC}$, $\log K_{OA}$, BCF in aquatic invertebrate and Henry's law constant ($H$). Finally, the photo-induced toxicity for the freshwater organism *Daphinia-Magna* is approached by the electronic descriptors calculated for the ground states.

## 2. Data aquisition

### 2.1. Data set

The studied data set consists of 47 nonsubstituted PAHs and benzene (Fig. 1). With the availability of substantial computer power, the proposition of many different kinds of descriptors became feasible and can be easily calculated for all compounds. The series of descriptors to be mentioned were calculated, and used to correlate with the already mentioned physicochemical properties.

A large variety of electronic descriptors of whole molecules have been currently used to describe electronic features in QSPR investigations. They can give crucial direct information about the reactivity of compounds and are easily obtained through MO calculations. The simplest of such descriptors are the energies of the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals. The HOMO energy is roughly related to the ionization potential of a molecule, while LUMO energy is related to the electron affinity. The HOMO–LUMO gap, which defines the energy necessary to excite an electron from HOMO orbital to the LUMO, turned out also to be a useful descriptor. The electronic descriptors used in this work were the ionization potential (IP) which is taken as negative HOMO energy, the electron affinity (EA) which is negative LUMO energy, and the gap IP–EA energy (GAP). Some of their values were taken from Lu and Lee (1993) and the others were calculated by MOPAC 6.0 using MNDO parametrization to be consistent with values taken from the literature.

The topological descriptors are used to describe the size and shape, which are important factors for example in boiling temperature, chromatographic retention behavior and the various partition coefficients. The topological descriptors calculated in this work are the vertex-connectivity ($X_v$) (Randic, 1975), edge-connectivity ($X_e$) (Estrada, 1995) and Wiener (1947) index ($W$).

Geometric descriptors such as length, width, thickness, and molecular volume, surface area, length/breath

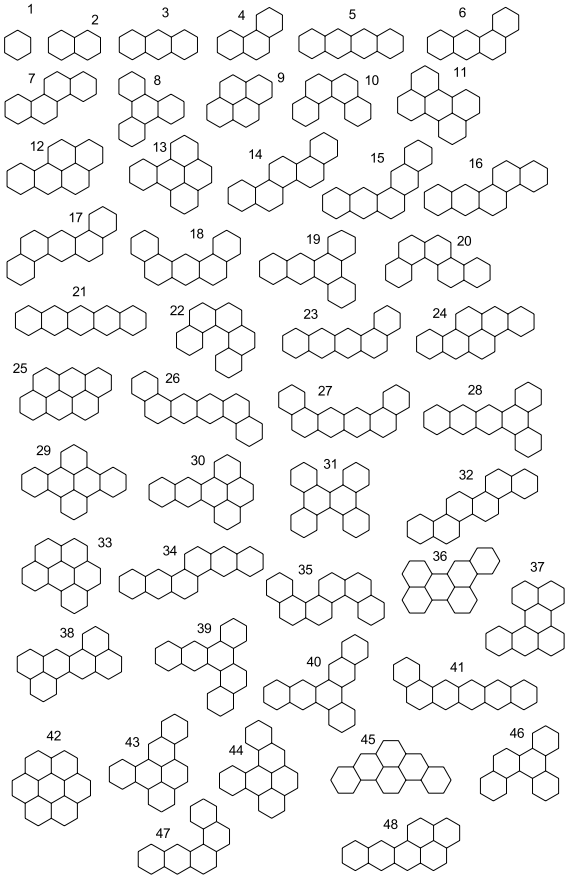



Fig. 1. 1-Benzene, 2-Naphthalene, 3-Anthracene, 4-Phenanthrene, 5-Naphthacene, 6-Benz[a]anthracene, 7-Chrysene, 8-Triphenylene, 9-Pyrene, 10-Benzo[c]phenanthrene, 11-Perylene, 12-Benzo[a]pyrene, 13-Benzo[e]pyrene, 14-Picene, 15-Pentaphene, 16-Benzo[b]chrysene, 17-Dibenz[a,h]anthracene, 18-Dibenz[a,j]anthracene, 19-Dibenz[a,c]anthracene, 20-Benzo[c]chrysene, 21-Pentacene, 22-Dibenzo[c,g]phenanthrene, 23-Benzo[a]naphthacene, 24-Dibenzo[a,h]pyrene, 25-Anthanthrene, 26-Dibenzo[a,j]naphthacene, 27-Dibenzo[a,l]naphthacene, 28-Dibenzo[a,c]naphthacene, 29-Dibenzo[e,l]pyrene, 30-Dibenzo[de,qr]naphthacene, 31-Dibenzo[g,p]chrysene, 32-Benzo[c]picene, 33-Benzo[ghi]perylene, 34-Dibenzo[b,k]chrysene, 35-Dibenzo[c,l]chrysene, 36-Benzo[b]perylene, 37-Benzo[a]perylene, 38-Dibenzo[de,mn]naphthacene, 39-Dibenz[a,n]triphenylene, 40-Benzo[h]pentaphene, 41-Benzo[a]pentacene, 42-Coronene, 43-Dibenzo[a,e]pyrene, 44-Dibenzo[a,l]pyrene, 45-Dibenzo[a,i]pyrene, 46-Benzo[g]chrysene, 47-Dibenzo[b,g]phenanthrene, 48-Naphtho[2,1,8-qra]naphthacene.

ratio, etc., encode spatial distribution of the atomic groups in the molecule. The length, width, thickness, and length/breath ratio ($L/B$) used were taken from Sanders and Wise database (http address). The $L/B$ value is defined as the ratio of the longer to the shorter side of the rectangle, which has the minimum area among all the rectangles drawn to enclose the van der Waals radii of the atoms in the molecule. The molecular volume (Vol) and surface area (SArea) were calculated by using tabulated length, thickness and width from reference above, according to the following equations:

$$\begin{aligned}\text{Vol} &= \text{length} \times \text{width} \times \text{thickness},\\ \text{SArea} &= \text{length} \times \text{width} + \text{length} \times \text{thickness}\\ &\quad + \text{width} \times \text{thickness}.\end{aligned}$$

These descriptors besides being easily calculated are convenient representations of the concepts of volume and surface area. In the literature (see for example Pearlman et al., 1984), other definitions are used and consequently yield different values for molecular volumes and surface areas.

Besides the descriptors mentioned above, the enthalpy of formation, calculated according to the Benson method (Alberty and Reif, 1988; Alberty et al., 1989), was also included as descriptor. The resulting descriptor data is reported in Table 1 and arranged in matrix **X**, where each column corresponds to a descriptor and each row represents one PAH compound.

Some physical properties of environmental interest were chosen to correlate with these descriptors. They were: experimental boiling temperatures taken from the Spectral Atlas of Polycyclic Aromatic Compounds (Karcher, 1988); reversed phase LC retention indexes of polymeric phase from Sander and Wise (1986); values of octanol/water partition coefficient and solubility ($S$) from reference (Mackay et al., 1992). All the experimental data are summarized in Table 2.

Structure–property relationships are derived using the descriptor data and the physical properties $T_b$, RI, $K_{OW}$ and $S$. These four properties are then incorporated into the data matrix **X** and the structure–property relationships for $K_{OC}$, $K_{OA}$, $H$ and BCF derived.

Experimental data for $K_{OC}$, $H$ and BCF were taken (unless otherwise stated) from Mackay et al. (1992). $K_{OA}$ values are reported by Mackay and Callcot (1998).

Finally, the PAHs photo-induced toxicities are modeled using the pretreated electronic descriptors. The GAP, IP and EA are transformed prior analysis using gaussian type transformations.

## 3. Methodology

The descriptor matrix $\mathbf{X} = (n, m)$ with $n = 48$ rows (i.e., number of samples) and 13 columns was autoscaled as pretreatment, i.e., was meancentered (i.e., each element of data matrix was subtracted by its mean column) and scaled to variance of one (each element divided by the standard deviation of its column) before the analysis.

The multivariate technique of partial least-squares (PLS) regression, was used for modeling (Martens and

Table 1
List of descriptors used

| | IP (eV) | EA (eV) | GAP (eV) | $W$ | $L/B$ | $X_v$ | $X_e$ | Width (Å) | Length (Å) | Thickness (Å) | Volume (Å$^3$) | SArea (Å$^2$) | $\Delta H_f$ (kJ/mol) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Benzene | 9.391 | −0.368 | 9.759 | 27 | 1.099 | 3.000 | 3.000 | 6.740 | 7.406 | 3.883 | 193.826 | 104.845 | 82.8 |
| Naphthalene | 8.574 | 0.332 | 8.242 | 109 | 1.238 | 4.966 | 5.454 | 7.428 | 9.915 | 3.884 | 286.051 | 141.009 | 150.6 |
| Anthracene | 8.049 | 0.842 | 7.207 | 279 | 1.566 | 6.933 | 7.942 | 7.439 | 11.651 | 3.882 | 336.460 | 160.779 | 218.3 |
| Phenanthrene | 8.478 | 0.481 | 7.997 | 271 | 1.463 | 6.950 | 7.926 | 8.031 | 11.752 | 3.888 | 366.951 | 171.297 | 209.1 |
| Naphthacene | 7.721 | 1.185 | 6.536 | 569 | 1.896 | 8.899 | 10.430 | 7.446 | 14.116 | 3.885 | 408.344 | 188.876 | 286.1 |
| Benz[a]anthracene | 8.114 | 0.830 | 7.284 | 557 | 1.599 | 8.916 | 10.414 | 8.717 | 13.942 | 3.887 | 472.397 | 209.608 | 276.9 |
| Chrysene | 8.261 | 0.719 | 7.542 | 545 | 1.734 | 8.933 | 10.397 | 8.039 | 13.939 | 3.992 | 447.326 | 199.792 | 267.7 |
| Triphenylene | 8.505 | 0.557 | 7.947 | 513 | 1.119 | 8.950 | 10.414 | 10.444 | 11.682 | 3.373 | 411.529 | 196.638 | 258.5 |
| Pyrene | 8.039 | 0.902 | 7.137 | 362 | 1.257 | 7.933 | 9.409 | 9.279 | 11.662 | 3.888 | 420.727 | 189.630 | 230.5 |
| Benzo[c]phenanthrene | 8.283 | 0.692 | 7.591 | 529 | 1.277 | 8.933 | 10.403 | 9.323 | 11.909 | 4.987 | 553.695 | 216.912 | 280.5 |
| Perylene | 7.815 | 1.119 | 6.695 | 654 | 1.276 | 9.933 | 11.897 | 9.247 | 11.800 | 3.913 | 426.965 | 191.472 | 279.9 |
| Benzo[a]pyrene | 7.861 | 1.088 | 6.773 | 680 | 1.493 | 9.916 | 11.897 | 9.297 | 13.882 | 3.891 | 502.176 | 219.250 | 279.9 |
| Benzo[e]pyrene | 8.108 | 0.889 | 7.219 | 652 | 1.118 | 9.933 | 11.897 | 10.520 | 11.765 | 3.887 | 481.085 | 210.390 | 289.1 |
| Picene | 8.237 | 0.751 | 7.486 | 963 | 2.005 | 10.916 | 12.868 | 8.037 | 16.112 | 3.897 | 504.631 | 223.601 | 326.3 |
| Pentaphene | 8.113 | 0.844 | 7.268 | 993 | 1.748 | 10.882 | 12.902 | 9.207 | 16.093 | 3.893 | 576.819 | 246.661 | 344.7 |
| Benzo[b]chrysene | 7.972 | 0.995 | 6.997 | 975 | 1.858 | 10.899 | 12.885 | 8.755 | 16.264 | 3.890 | 553.902 | 239.715 | 339.0 |
| Dibenz[a,h]anthracene | 8.149 | 0.829 | 7.320 | 975 | 1.822 | 10.899 | 12.885 | 8.726 | 15.898 | 3.890 | 539.644 | 234.513 | 335.5 |
| Dibenz[a,j]anthracene | 8.174 | 0.829 | 7.320 | 973 | 1.530 | 10.899 | 12.885 | 9.502 | 14.542 | 3.887 | 537.098 | 231.637 | 335.5 |
| Dibenz[a,c]anthracene | 8.174 | 0.823 | 7.351 | 911 | 1.238 | 10.916 | 12.902 | 11.247 | 13.922 | 3.889 | 608.942 | 254.463 | 326.3 |
| Benzo[c]chrysene | 8.269 | 0.747 | 7.522 | 931 | 1.519 | 10.916 | 12.874 | 9.342 | 14.193 | 5.395 | 715.328 | 259.562 | 335.5 |
| Pentacene | 7.499 | 1.429 | 6.070 | 1011 | 2.260 | 10.865 | 12.919 | 7.447 | 16.577 | 3.885 | 479.599 | 216.782 | 353.9 |
| Dibenzo[c,g]phenanthrene | 8.217 | 0.822 | 7.352 | 899 | 1.165 | 10.916 | 12.880 | 10.146 | 11.815 | 6.240 | 748.020 | 353.900 | **333.0**[a] |
| Benzo[a]naphthacene | 7.797 | 1.145 | 6.652 | 987 | 1.801 | 10.882 | 12.902 | 9.004 | 16.217 | 3.887 | 567.571 | 244.052 | 344.7 |
| Dibenzo[a,h]pyrene | 7.646 | 1.318 | 6.327 | 1142 | 1.732 | 11.899 | 14.385 | 9.287 | 16.084 | 3.895 | 581.804 | 248.192 | 347.7 |
| Anthanthrene | 7.618 | 1.330 | 6.287 | 839 | 1.345 | 10.899 | 13.397 | 9.588 | 12.898 | 3.884 | 480.319 | 211.002 | 310.5 |
| Dibenzo[a,j]naphthacene | 7.869 | 1.107 | 6.762 | 1581 | 2.010 | 12.865 | 15.373 | 8.986 | 18.065 | 3.947 | 640.725 | 269.102 | 397.6 |
| Dibenzo[a,l]naphthacene | 7.872 | 1.106 | 6.766 | 1545 | 1.790 | 12.865 | 15.373 | 9.512 | 17.029 | 3.884 | 629.130 | 265.065 | 394.0 |
| Dibenzo[a,c]naphthacene | 7.851 | 1.126 | 6.725 | 1488 | 1.410 | 12.882 | 15.390 | 11.481 | 16.185 | 3.891 | 723.026 | 293.468 | 397.6 |
| Dibenzo[e,l]pyrene | 8.168 | 0.881 | 7.288 | 1062 | 1.330 | 11.933 | 14.385 | 10.551 | 13.982 | 3.886 | 573.279 | 242.859 | 329.2 |
| Dibenzo[de,qr]naphthacene | 8.166 | 0.866 | 7.301 | 1113 | 1.288 | 11.974 | 14.385 | 11.161 | 14.375 | 4.411 | 707.698 | 273.079 | 347.7 |
| Dibenzo[g,p]chrysene | 8.006 | 0.920 | 7.086 | 1377 | 1.095 | 12.933 | 14.885 | 12.376 | 13.546 | 6.154 | 1031.689 | 327.169 | 412.4 |
| Benzo[c]picene | 8.152 | 0.868 | 7.284 | 1562 | 1.274 | 12.899 | 15.339 | 8.048 | 18.297 | 3.899 | 574.144 | 249.973 | 403.2 |
| Benzo[ghi]perylene | 7.940 | 1.071 | 6.869 | 815 | 1.124 | 10.916 | 13.380 | 10.484 | 11.779 | 3.887 | 480.010 | 210.027 | 301.3 |
| Dibenzo[b,k]chrysene | 7.818 | 1.155 | 6.663 | 1567 | 2.126 | 12.865 | 14.873 | 8.752 | 18.608 | 3.891 | 633.677 | 269.315 | 403.2 |
| Dibenzo[c,l]chrysene | 8.142 | 0.872 | 7.270 | 1461 | 1.288 | 12.899 | 15.351 | 9.211 | 15.463 | 5.965 | 849.593 | 289.610 | 421.6 |
| Benzo[b]perylene | 7.819 | 1.144 | 6.675 | 1088 | 1.396 | 11.916 | 14.385 | 10.301 | 14.375 | 4.844 | 717.284 | 267.607 | 338.5 |
| Benzo[a]perylene | 7.602 | 1.329 | 6.273 | 1071 | 1.177 | 11.916 | 14.391 | 11.591 | 13.646 | 5.469 | 865.036 | 296.192 | 351.2 |
| Dibenzo[de,mn]naphthacene | 7.497 | 1.406 | 6.091 | 1122 | 1.532 | 11.899 | 14.385 | 9.393 | 14.386 | 4.894 | 661.315 | 251.502 | 347.7 |
| Dibenz[a,n]triphenylene | 7.991 | 0.962 | 7.029 | 1410 | 1.107 | 12.899 | 15.379 | 12.541 | 13.890 | 5.432 | 946.224 | 317.768 | 406.8 |
| Benzo[h]pentaphene | 8.190 | 0.835 | 7.355 | 1440 | 1.359 | 12.882 | 15.390 | 11.703 | 15.901 | 4.850 | 902.534 | 319.969 | 408.8 |
| Benzo[a]pentacene | 7.566 | 1.390 | 6.176 | 1579 | 2.034 | 12.849 | 15.390 | 9.128 | 18.565 | 3.887 | 658.696 | 277.104 | 384.8 |

Table 1 (Continued)

| | IP (eV) | EA (eV) | GAP (eV) | $W$ | $L/B$ | $X_v$ | $X_e$ | Width (Å) | Length (Å) | Thickness (Å) | Volume (Å$^3$) | SArea (Å$^2$) | $\Delta H_f$ (kJ/mol) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Coronene | 8.077 | 1.012 | 7.065 | 1002 | 1.002 | 11.899 | 14.863 | 11.701 | 11.722 | 3.887 | 533.137 | 228.204 | 322.7 |
| Dibenzo[a,e]pyrene | 7.947 | 1.047 | 6.900 | 1082 | 1.241 | 11.916 | 14.385 | 11.207 | 13.904 | 4.194 | 653.52 | 261.138 | 338.5 |
| Dibenzo[a,l]pyrene | 7.884 | 1.090 | 6.794 | 1066 | 1.171 | 11.916 | 14.390 | 11.696 | 13.693 | 5.167 | 827.512 | 291.338 | 351.2 |
| Dibenzo[a,i]pyrene | 7.816 | 1.142 | 6.673 | 1142 | 1.732 | 11.899 | 14.385 | 9.291 | 16.089 | 3.893 | 581.937 | 248.287 | 347.7 |
| Benzo[g]chrysene | 8.157 | 0.815 | 7.341 | 899 | 1.314 | 10.993 | 12.891 | 10.488 | 13.780 | 5.323 | 769.305 | 273.703 | 329.8 |
| Dibenzo[b,g]phenanthrene | 7.972 | 0.967 | 7.005 | 942 | 1.373 | 10.899 | 12.891 | 10.130 | 13.909 | 5.229 | 736.756 | 266.598 | 348.3 |
| Naphtho[2,1,8-qra]naphthacene | 7.692 | 1.264 | 6.429 | 1165 | 1.693 | 11.882 | 14.052 | 9.587 | 16.230 | 3.890 | 605.272 | 256.025 | 356.9 |

[a] Predicted in this work by PLS using some of these descriptors.

Naes, 1989; Beebe et al., 1998). Multivariate methods are advantageous in this case since they make it possible to estimate more than one regression coefficient for each property being modeled. PLS is a projection method in which the data matrix, represented as a set of $n$ points in an $m$-dimensional space, is projected on to a $k$-dimesional hiperplane, in such a way that the coordinates of the projection are good predictors of some **y** property. In other words, in PLS modeling a "new set of $x$ variables" called latent variables (LV) are generated as linear combination of the old ones, and then used as predictors of **y**. In the PLS method, the latent variables are computed taking into account both the descriptors and property values, making use of more information when building the model. Contrary to the conventional multiple linear regression method (MLR), PLS allows the simultaneous use of strongly intercorrelated $x$-descriptors by focusing the systematic covariances in the **X** block in a few latent variables. In modeling it is essential to determine its complexity to avoid overfitting. The predictive capability of the resulting model depends on the quality of the data, (the more and better the data available, the more accurate prediction is possible), and the number $k$ of significant latent variables necessary. Cross-validation is a practical and reliable method for testing this significance. In principle, the so-called "leave-one-out" approach consists in developing a number of models with one sample omitted at the time. After developing each model, the omitted data is predicted and the differences between actual and predicted $y$ values are calculated. The sum of squares of these differences is computed and finally, the performance of the model (its predictive ability) can be given by the standard error of prediction (SEP) defined as

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n}},$$

where $y$ is the experimental value, $\hat{y}$ the predicted value and $n$ is the number of samples used for model building.

## 4. Results and discussion

### 4.1. Boiling point ($T_b$)

The boiling point reported for 23 compounds were used in this section. The $T_b$ values (Table 2) follow the number of benzene rings: for 1–3 rings $T_b \leqslant 340$°C, for 4 rings 393°C $< T_b <$ 450°C, for 5 rings $493 < T_b < 535$°C, etc. Fig. 2 shows descriptors EA, $X_e$, $X_v$, Vol, SArea, $\log W$ and $\Delta H_f$ which are best correlated with $T_b$. It is well known that the connectivity indexes are successful descriptors for predicting $T_b$.

Table 2
List of physicochemical properties under study

| | $T_b$ (°C) | RI | log $S$ (g/m$^3$) | log $K_{OW}$ |
|---|---|---|---|---|
| Benzene | 80.1 | 1.00 | 3.25 | 2.13 |
| Naphthalene | 218.0 | 2.00 | 1.50 | 3.33 |
| Anthracene | 340.0 | 3.20 | −1.13 | 4.54 |
| Phenanthrene | 338.0 | 3.00 | 0.061 | 4.55 |
| Naphthacene | 450.0 | 4.51 | −3.24 | 5.96 |
| Benz[a]anthracene | 435.0 | 4.00 | −1.95 | 5.91 |
| Chrysene | 431.0 | 4.10 | −2.48 | 5.84 |
| Triphenylene | 429.0 | 3.70 | −1.37 | 5.45 |
| Pyrene | 393.0 | 3.58 | −0.87 | 5.14 |
| Benzo[c]phenanthrene | | 3.64 | | 5.84 |
| Perylene | 497.0 | 4.33 | −3.40 | 6.30 |
| Benzo[a]pyrene | 496.0 | 4.53 | −2.42 | 6.30 |
| Benzo[e]pyrene | 493.0 | 4.28 | −2.25 | |
| Picene | 519.0 | 5.18 | | |
| Pentaphene | | 4.67 | | |
| Benzo[b]chrysene | | 5.00 | | |
| Dibenz[a,h]anthracene | 535.0 | 4.73 | −3.22 | 6.75 |
| Dibenz[a,j]anthracene | 531.0 | 4.56 | | |
| Dibenz[a,c]anthracene | 535.0 | 4.40 | | 7.19 |
| Benzo[c]chrysene | | 4.45 | | |
| Pentacene | | | | 7.19 |
| Dibenzo[c,g]phenanthrene | | 4.07 | | |
| Benzo[a]naphthacene | | 4.99 | | 6.81 |
| Dibenzo[a,h]pyrene | 596.0 | 6.00 | | |
| Anthanthrene | 547.0 | 5.08 | | |
| Dibenzo[a,j]naphthacene | | | | |
| Dibenzo[a,l]naphthacene | | | | |
| Dibenzo[a,c]naphthacene | | | | |
| Dibenzo[e,l]pyrene | | | | |
| Dibenzo[de,qr]naphthacene | | 4.92 | | |
| Dibenzo[g,p]chrysene | | | | |
| Benzo[c]picene | | | | |
| Benzo[ghi]perylene | 542.0 | 4.76 | −3.59 | 6.87 |
| Dibenzo[b,k]chrysene | | | | |
| Dibenzo[c,l]chrysene | | | | |
| Benzo[b]perylene | | 5.04 | | |
| Benzo[a]perylene | | | | |
| Dibenzo[de,mn]naphthacene | | | | |
| Dibenz[a,n]triphenylene | | | | |
| Benzo[h]pentaphene | | | | |
| Benzo[a]pentacene | | | | |
| Coronene | 590.0 | | −3.85 | |
| Dibenzo[a,e]pyrene | 592.0 | 4.97 | | |
| Dibenzo[a,l]pyrene | 595.0 | 4.89 | | |
| Dibenzo[a,i]pyrene | 594.0 | 5.73 | | |
| Benzo[g]chrysene | | 4.27 | | |
| Dibenzo[b,g]phenanthrene | | 4.33 | | |
| Naphtho[2,1,8-qra]naphthacene | | 5.87 | | |

Several models were built using these descriptors and compared. A previous principal component analysis had shown that the electron affinity (EA) has a great contribution in PC1 and PC2, so, it was included in all considered models. The results are shown in Table 3.

The first column contains the experimental results and the following columns show the residuals (experimental–estimated value) obtained by leave-one-out cross-validation. Note that when $X_v$ is included in the model (Mod1), the prediction errors are higher for those PAHs with higher $T_b$ and the opposite happens when $X_e$ replaces $X_v$ in the model. So, both were used as descriptors (Mod3 in Table 1) in an attempt to improve the model at both extremities. Similar trends are visible

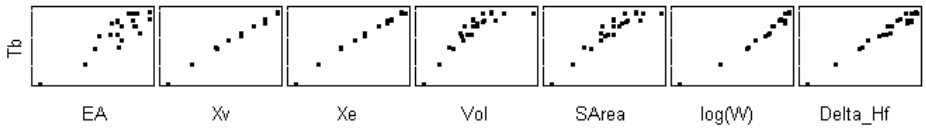



Fig. 2. Descriptors best correlated with the boiling temperature.

Table 3
Actual boiling points, cross-validated prediction errors, standard errors of prediction (SEP) and correlation coefficients ($r$)

| | Measured $T_b$ (°C) | Prediction errors | | | |
|---|---|---|---|---|---|
| | | Mod1[a] | Mod2[b] | Mod3[c] | Mod4[d] |
| Benzene | 80.10 | −19.70 | 2.47 | −6.68 | 1.59 |
| Naphthalene | 218.0 | −1.61 | −2.53 | −1.99 | −4.25 |
| Anthracene | 340.0 | −0.96 | −0.24 | −0.42 | 0.14 |
| Phenanthrene | 338.0 | 13.77 | 9.28 | 11.31 | **9.46** |
| Naphthacene | 450.0[e] | −2.49 | 2.38 | 0.21 | 2.89 |
| Benz[a]anthracene | 435.0 | −3.61 | −4.08 | −3.86 | −6.16 |
| Chrysene | 431.0 | −3.64 | −4.26 | −4.00 | −3.48 |
| Triphenylene | 429.0 | 1.75 | −0.32 | 0.51 | 2.20 |
| Pyrene | 393.0 | 2.17 | −1.94 | −0.04 | −4.11 |
| Perylene | 497.0 | 0.18 | 1.84 | 0.99 | 6.48 |
| Benzo[a]pyrene | 496.0 | 0.56 | −0.05 | 0.15 | −1.62 |
| Benzo[e]pyrene | 493.0 | 5.61 | 3.29 | 4.21 | 4.71 |
| Picene | 519.0 | −15.23 | −13.16 | −14.22 | −10.19 |
| Dibenz[a,h]anthracene | 535.0 | 1.12 | 2.85 | 1.84 | 3.25 |
| Dibenz[a,j]anthracene | 531.0 | −3.45 | −1.69 | −2.68 | −0.65 |
| Dibenz[a,c]anthracene | 535.0 | 1.75 | 5.17 | 3.37 | 1.01 |
| Dibenzo[a,h]pyrene | 596.0 | −4.58 | 1.99 | −1.21 | 0.16 |
| Anthanthrene | 547.0 | −5.73 | −8.87 | −7.62 | −5.82 |
| Benzo[ghi]perylene | 542.0 | 0.52 | −4.81 | −2.58 | −0.26 |
| Coronene | 590.0 | 5.30 | −8.55 | −2.32 | −2.97 |
| Dibenzo[a,e]pyrene | 592.0 | 3.54 | 6.78 | 4.99 | 3.85 |
| Dibenzo[a,l]pyrene | 595.0 | 5.33 | 9.50 | 7.30 | −1.33 |
| Dibenzo[a,i]pyrene | 594.0 | 1.33 | 4.22 | 2.64 | 3.60 |
| SEP | | 6.68 | 5.52 | 5.70 | 4.40 |
| $r$ | | 0.9994 | 0.9993 | 0.9995 | 0.9996 |
| Regression coefficient | EA | 0.083 | | | |
| | $X_e$ | 0.654 | | | |
| | SArea | 0.063 | | | |
| | log($W$) | 0.224 | | | |

[a] EA, $X_v$, log $W$.
[b] EA, $X_e$, log $W$.
[c] EA, $X_v$, $X_e$, log $W$.
[d] EA, $X_e$, SArea and log $W$. Four latent variables in the model.
[e] From Mackay et al. (1992) and White (1986).

(with less intensity) when the volume (Vol) and SArea are included for modeling. When the four descriptors, EA; $X_e$, SArea and log $W$, were included in the model (Mod4), the residuals show that excellent results are obtained, being higher than 2% only for phenanthrene. The regression coefficients are included in Table 4. Fig. 3 contains experimental versus estimated (by cross-validation) $T_b$ for the best validated model (Mod4).

In the following step, the validated model (Mod4) is used to predict the boiling points for those compounds where there are no experimental measurements and so, not included in the modeling step. Table 4 contains the predicted $T_b$ values. According to White (1986), the calculated $T_b$ for benzo[b]chrysene is 541°C which is in excellent agreement with the predicted value of 538°C obtained in the present work (Table 4). The predicted value for pentacene using Mod4 is 548.3°C while their calculated is 529, about 4% lower.

The $T_b$ values in the calibration set are in the range between 80°C and 596°C and several of the predicted

Table 4
Predicted values of boiling point (°C) using Mod4

| | |
|---|---|
| Benzo[c]phenanthrene | 435.84 |
| Pentaphene | 536.09 |
| Benzo[b]chrysene | 538.04 |
| benzo[c]chrysene | 532.91 |
| Pentacene | 548.29 |
| Dibenzo[c,g]phenanthrene | 533.59 |
| Benzo[a]naphthacene | 544.13 |
| Dibenzo[a,j]naphthacene | 631.93 |
| Dibenzo[a,l]naphthacene | 630.35 |
| Dibenzo[a,c]naphthacene | 635.76 |
| Dibenzo[e,l]pyrene | 579.71 |
| Dibenzo[de,qr]naphthacene | 586.78 |
| Dibenzo[g,p]chrysene | 619.87 |
| Benzo[c]picene | 619.87 |
| Dibenzo[b,k]chrysene | 619.23 |
| Dibenzo[c,l]chrysene | 625.93 |
| Benzo[b]perylene | 593.03 |
| Benzo[a]perylene | 603.70 |
| Dibenzo[de,mn]naphthacene | 598.50 |
| Dibenz[a,n]triphenylene | 633.70 |
| Benzo[h]pentaphene | 631.44 |
| Benzo[a]pentacene | 642.17 |
| Benzo[g]chrysene | 537.00 |
| Dibenzo[b,g]phenanthrene | 541.56 |
| Naphtho[2,1,8-qra]naphthacene | 587.27 |

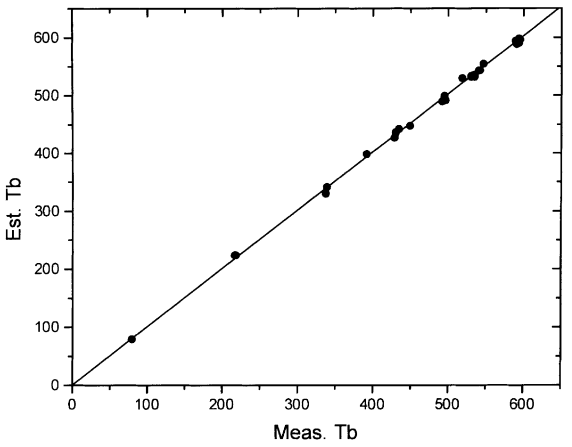



Fig. 3. Actual versus estimated (predicted by leave-one-out crossvalidation (**Mod4**)) boiling temperatures (°C).

values, are beyond this range. Are these extrapolated values reliable? To answer this question, a new model was built, considering only those 16 PAHs from Table 1 which had $T_b$ lower than 540°C in order to obtain a measure of the likely performance of the regression model and evaluate how "robust" or stable it is. This new model is well fitted since the omission of some samples did not greatly disturb the regression coefficients, which proves its stability. The predicted values

Table 5
Experimental and predicted boiling temperatures (°C) for testing model robustness

| Boiling point (°C) | Experimental | Predicted |
|---|---|---|
| Dibenzo[a,h]pyrene | 596.0 | 596.26 |
| Anthanthrene | **547.0** | **554.35** |
| Benzo[ghi]perylene | 542.0 | 545.26 |
| Coronene | 590.0 | 595.25 |
| Dibenzo[a,e]pyrene | 592.0 | 587.81 |
| Dibenzo[a,l]pyrene | 595.0 | 591.54 |
| Dibenzo[a,i]pyrene | 594.0 | 591.13 |

(which are extrapolation) for the other 9 PAHs with $T_b$ greater than 540°C (Table 5) have shown with exception of anthanthrene, a prediction error within 1% indicating that the proposed model in Table 3 is robust. Concluding, it might be expected that those predicted boiling points shown in Table 4, which are higher than 600°C, can be acceptable.

Fig. 4 shows the linear relationship between $T_b$ and EA for linear polyacenes. Note that when comparing compounds with the same number of aromatic rings, it is found that the pericondensed, (i.e., those compounds possessing internal vertices) have lower $T_b$ than the catacondensed (i.e., those, which does not possess internal vertices). The existence of internal vertices reduces the number of external carbons and hydrogens which lead to smaller number of intermolecular interactions (lower $T_b$).

When comparing isomeric catacondensed PAHs (Fig. 4), for example isomers with four aromatic rings, ($C_{18}H_{12}$), the electron affinity decreases as the compound changes its shape, becoming more compact in the following order:

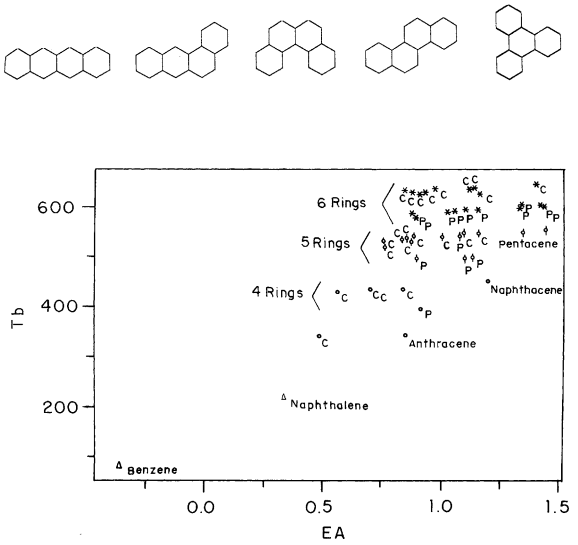



Fig. 4. Boiling temperature versus electron affinity for pericondensed (P) and catacondensed (C) PAHs.

### 4.2. Retention index (RI)

There are several studies in the literature concerning the separation of PAHs motivated by their environmental impact. The retention data used here (see Table 2) refers to the reversed-phase LC reported as the retention time index, $\log I$ (RI), by Sander and Wise (1986) for separation of 33 samples. The retention time indexes were calculated from the elution volume of the solute measured simultaneously with elution volume of standards. The standards were assigned the following values of RI: benzene 1.0; naphthalene 2.0; phenanthrene 3.0; benz[a]anthracene, 4.0 benzo[b]chrysene. 5.0; dibenzo[a,h]pyrene 6.0. Thus, a PAH with RI 4.3 elutes between benz[a]anthracene, and benzo[b]chrysene. The linear correlation among RI and some of the selected descriptors is summarized in Fig. 5, while Fig. 6 shows a plot of RI versus the $L/B$ ratio. For compounds with 4 benzene rings (Fig. 6), there is a fairly linear relationship RI versus $L/B$ ratio and the same is true for compounds with higher number of rings.

Note that with some exceptions there is also a good linear trend for catacondensed PAHs. Based on this linear trend for catacondensed PAHs, it is expected that pentacene should elute significantly later than naphthacene and picene. Those catacondensed PHAs which do not follow the linear trend are listed in Table 6, with their respective thickness. The thickness of a planar PAH is approximately 3.88 Å (see Table 1), indicating that such compounds are visibly non-planar, except the dibenzo anthracenes (two of them listed in Table 6) (Scheme 1).

PLS regression model is built and compared with values estimated by leave-one-out cross-validation in Table 7. Excluding benzene, only six of the estimated values are 5% away from their respective measured values. The regression coefficients (Table 7) were obtained using three latent variables and the following descriptors: EA, $X_v$, $W$, Vol, and $\Delta H_f$. The predicted RI for other PAHs not available in the literature are shown in Table 8. From 15 predicted RI, it can be seen that six of the PAHs should elute after dibenzo[a,h]pyrene, which seems very reasonable since all of them have six aromatic rings and are planar catacondensed compounds according to their thickness and $L/B$ ratio (see Table 1). Comparing the predicted RI for dibenzo chrysenes (see their structure representation in Scheme 2), the dibenzo[b,k]chrysene is the only planar compound among them and has the greater $L/B$ ratio, so it is expected to elute much later than dibenzo[g,p]chrysene which has the smaller $L/B$ ratio and, being the most non-planar should elute first.

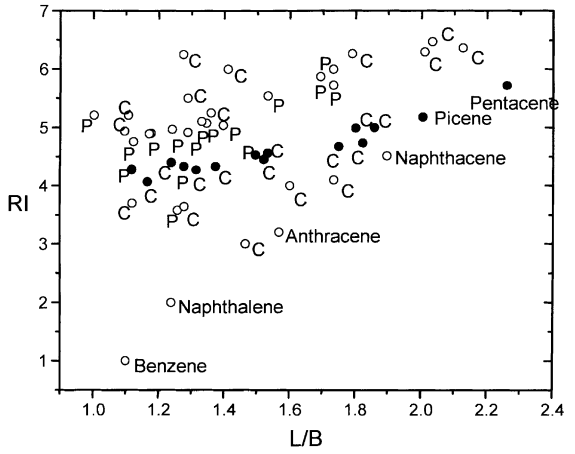



Fig. 6. LC retentions versus length-to-breadth ratio for pericondensed (P) and catacondensed (C) PAHs. • compounds with five benzene rings.

Table 6
Thickness of catacondensed PAHs which does not the follow linear trend between RI and $L/B$ ratio

| | Thickness |
|---|---|
| Triphenylene | 4.373 |
| Dibenz[a,j]anthracene | 3.887 |
| Dibenz[a,c]anthracene | 3.889 |
| Benzo[c]chrysene | 5.395 |
| Benzo[g]chrysene | 5.323 |
| Benzo[c]phenanthrene | 4.987 |
| Dibenzo[b,g]phenanthrene | 5.229 |
| Dibenzo[c,g]phenanthrene | 6.240 |

### 4.3. Water solubility (S)

The experimental data of solubility (Mackay et al., 1992) used in this section were measured at 25°C by the "shake flask method". In this technique, an excess amount of solute is added to water and after the equilibrium is achieved by shaking, chromatography or spec-

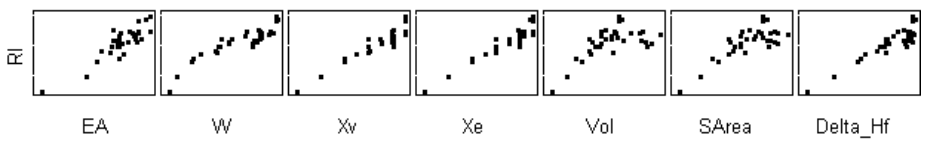
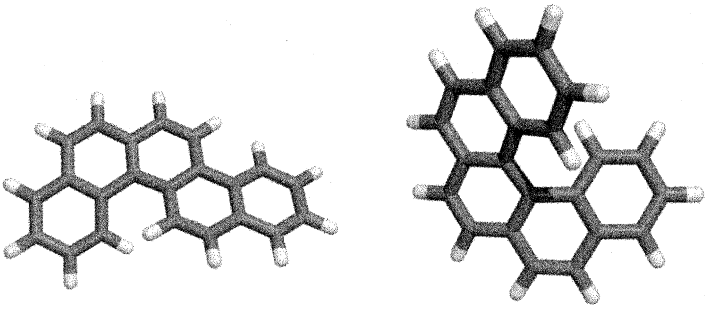



Fig. 5. Retention index versus some of the descriptors.



non-planar *Benzo[c]chrysene* *Dibenzo[c,g]phenanthrene*

Scheme 1.

troscopy is used to measure the concentration of the saturated solution. The main difficulty with the PAHs arises due to their high hydrophobicity, consequently, making their concentrations very small. In this study, picene, dibenz[a,j]anthracene and dibenz[a,c]anthracene are not included since they were not measured by the shake flask method. The linear relationship among $\log S$ and the descriptors is shown in Fig. 7. $\log S$ roughly decreases as the molecular size or weight increases leading to even more hydrophobic compounds (as the number of rings increases, the differentiation in atomic charges tends to decrease resulting in decrease of polarizability and polarity). For example, the solubility of naphthalene (1.50 g/m$^3$) > anthracene (−1.13) > naphthacene (−3.24) for carbon/hydrogen ratios 10/8, 14/10 and 18/12 respectively. The same trend is observed for phenanthrene, benz[a]anthracene and benz[a]naphthacene.

The PLS modeling is done using 14 compounds and 5 descriptors (EA, $X_e$, SArea, $\Delta H_f$ and RI) on $\log S$. The estimated $\log S$ and the residuals obtained by cross-validation are in Table 9. Basically four compounds were predicted with an error greater then 10%.

The values of $\log S$ for several PAHs including picene, dibenz[a,j]anthracene and dibenz[a,c]anthracene have been predicted and are listed in Table 10 together with the calculated values reported in the literature (Karcher, 1988; Pearlman et al., 1984; Mackay et al., 1992). Fig. 8 contains the plot of $\log S$ versus $T_b$, showing the predicted values from this work and those from literature, for the sake of comparison. From this plot, the predicted value of −3.79 g/m$^3$ for picene (represented as picene*) is in better agreement with the general trend. The calculated and predicted values for dibenz[a,j]anthracene (Db[a,j]A and Db[a,j]A*, respectively in Fig. 8) are far from each other but the predicted solubility in this work is close to the experimental value of dibenz[a,h]anthracene what would be expected. On the other hand, dibenz[a,c]anthracene* is not so well predicted.

### 4.4. n-octanol/water partition coefficient ($K_{OW}$)

The relative affinity of a compound for an aqueous or less polar medium is an important correlate due to absorption transport and partitioning phenomena. The most widely used molecular structure descriptor to encode this property, in QSPR studies, is $K_{OW}$. Compounds for which $K_{OW} > 1$ are lipophilic or hydrophobic and compounds for which $K_{OW} < 1$ are hydrophilic. By the definition, $K_{OW}$ is inversely proportional to aqueous solubility.

In this section, 22 samples are considered, most of them having reported values of $\log K_{OW}$, obtained by the shake flask method at 25°C (Mackay et al., 1992) or taken from Sangster (1989). Since a large number of measurements are available for some compounds, an average value is used in such cases (Table 11). Recently, de Maagd et al. (1998) determined $K_{OW}$ for several PAHs using the slow-stirring method. Their results are included in Table 11 also for comparison. As with $T_b$, RI and $S$, the same descriptors show a fairly good correlation with $\log K_{OW}$ (Fig. 9). $\log K_{OW}$ roughly increases with the size of the benzenoids. As mentioned before, with increasing number of rings, the differentiation of atomic charges decreases, resulting in lower polarity and so higher hydrophobicity. On the other hand, the hydrophobicity increases as the compound becomes more compact as can be seen for catacondensed isomers with four rings (from naphthacene to triphenylene).

The estimated values for a PLS model with three latent variables are in Table 11, followed by the residuals obtained by cross-validation. For this study, the same five descriptors used for the solubility were used to build the model: EA, $X_e$, SArea, $\Delta H_f$ and RI. The predicted values are in very good agreement with the experimental data. Note that only the estimated values of four compounds are higher than 3% and for dibenz[a,c]anthracene and benzo[a]naphthacene are greater than 5%. For dibenz[a,h]anthracene, the esti-

Table 7
Experimental and estimated retention indexes[a], cross-validated prediction errors, standard errors of prediction (SEP) and correlation coefficients ($r$)

| | Measured RI | Estimated RI | Residual |
|---|---|---|---|
| Benzene | 1.00 | 0.889 | **0.111** |
| Naphthalene | 2.00 | 2.160 | **−0.160** |
| Anthracene | 3.20 | 3.338 | −0.138 |
| Phenanthrene | 3.00 | 2.832 | 0.168 |
| Naphthacene | 4.51 | 4.446 | 0.064 |
| Benz[a]anthracene | 4.00 | 3.905 | 0.095 |
| Chrysene | 4.10 | 3.798 | **0.302** |
| Triphenylene | 3.70 | 3.663 | 0.037 |
| Pyrene | 3.58 | 3.469 | 0.111 |
| Benzo[c]phenanthrene | 3.64 | 3.550 | 0.090 |
| Perylene | 4.33 | 4.639 | **−0.309** |
| Benzo[a]pyrene | 4.53 | 4.412 | 0.118 |
| Benzo[e]pyrene | 4.28 | 4.289 | −0.009 |
| Picene | 5.18 | 4.770 | **0.410** |
| Pentaphene | 4.67 | 4.882 | −0.212 |
| Benzo[b]chrysene | 5.00 | 5.017 | −0.017 |
| Dibenz[a,h]anthracene | 4.73 | 4.898 | −0.168 |
| Dibenz[a,j]anthracene | 4.56 | 4.925 | **−0.365** |
| Dibenz[a,c]anthracene | 4.40 | 4.589 | −0.189 |
| Benzo[c]chrysene | 4.45 | 4.288 | 0.162 |
| Dibenzo[c,g]phenanthrene | 4.07 | 4.274 | −0.204 |
| Benzo[a]naphthacene | 4.99 | 5.175 | −0.185 |
| Dibenzo[a,h]pyrene | 6.00 | 5.630 | **0.370** |
| Anthanthrene | 5.08 | 5.232 | −0.152 |
| Dibenzo[de,qr]naphthacene | 4.92 | 4.910 | 0.010 |
| Benzo[ghi]perylene | 4.76 | 4.908 | −0.148 |
| Benzo[b]perylene | 5.04 | 5.074 | −0.034 |
| Dibenzo[a,e]pyrene | 4.97 | 5.138 | −0.168 |
| Dibenzo[a,l]pyrene | 4.89 | 4.731 | 0.159 |
| Dibenzo[a,i]pyrene | 5.73 | 5.485 | 0.245 |
| Benzo[g]chrysene | 4.27 | 4.161 | 0.109 |
| Dibenzo[b,g]phenanthrene | 4.33 | 4.563 | −0.233 |
| Naphtho[2,1,8-qra]naphthacene | 5.87 | 5.607 | 0.263 |
| SEP | | 0.196 | |
| $r$ | | 0.985 | |
| Regression coefficient | EA | 0.301 | |
| | W | 0.430 | |
| | Xv | 0.330 | |
| | Vol | −0.359 | |
| | $\Delta H_f$ | 0.258 | |

[a] PLS model using three latent variables.

mated $K_{OW}$ is 7.00, being in better agreement with some calculated values found in the literature (Mackay et al., 1992) which are 7.19. The PLS model was used to predict the $\log K_{OW}$ for other 19 PAHs (Table 12). Those compounds for which the predicted retention indexes were greater than 6.0 were not included in the prediction data set. Note that the predicted values for some compounds are very high seeming unrealistic in nature and only ten of them fall in the range of $\log K_{OW}$ used for modeling. The predicted value for coronene (Table 12) is in better agreement with the calculated value (7.64) than the experimental result (5.4) reported by Mackay et al. (1992). The other calculated values are also in good agreement with the predicted ones, being higher than 3% only for dibenz[a,j]anthracene and dibenzo[a,h]pyrene.

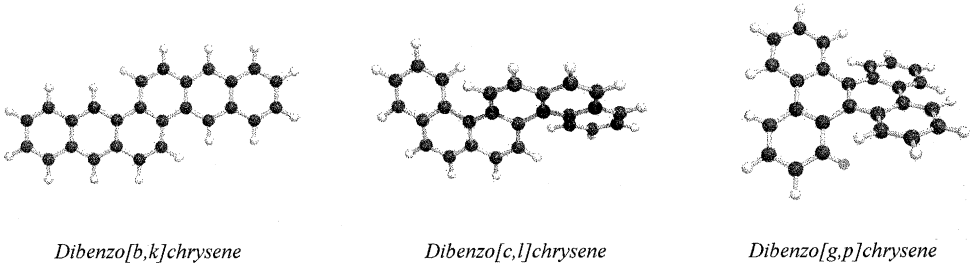
This ends the first part of this work, where some physical properties of interest were modeled by few descriptors, which could be easily calculated, showing that these properties can be well predicted even when there is a high degree of correlation among descriptors.

The second part deals with the modeling of some physicochemical factors of high interest in environmental assessment.

Table 8
Predicted retention indexes from PLS model, the number of aromatic rings and planarity[a]

| | | Predicted RI | # Rings | Planar |
|---|---|---|---|---|
| Pentacene | C[b] | 5.719 | 5 | +[c] |
| Dibenzo[a,j]naphthacene | C | 6.303 | 6 | + |
| Dibenzo[a,l]naphthacene | C | 6.266 | 6 | + |
| Dibenzo[a,c]naphthacene | C | 6.000 | 6 | + |
| Dibenzo[e,l]pyrene | P | 5.105 | 6 | + |
| Dibenzo[g,p]chrysene | C | 4.942 | 6 | − |
| Benzo[c]picene | C | 6.245 | 6 | + |
| Dibenzo[b,k]chrysene | C | 6.370 | 6 | + |
| Dibenzo[c,l]chrysene | C | 5.500 | 6 | − |
| Benzo[a]perylene | P | 4.903 | 6 | − |
| Dibenzo[de,mn]naphthacene | P | 5.537 | 6 | − |
| Dibenz[a,n]triphenylene | C | 5.212 | 6 | − |
| Benzo[h]pentaphene | C | 5.248 | 6 | − |
| Benzo[a]pentacene | C | 6.467 | 6 | + |
| Coronene | P | 5.214 | 7 | + |

[a] Descriptors used for modeling: EA, $X_v$, $W$, Vol and $\Delta H_f$.
[b] C: catacondensed; P: pericondensed.
[c] +: Planar; −: non-Planar.




Scheme 2.

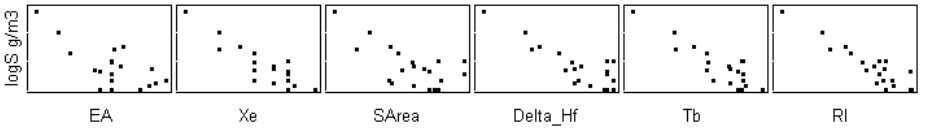



Fig. 7. Plots of log $S$ versus descriptors, showing their correlation.

### 4.5. Octanol/air partition coefficient ($K_{OA}$)

$K_{OA}$ is thought to be a key descriptor for the partitioning of semivolatile compounds between the atmosphere and terrestrial organic phases

$$K_{OA} = \frac{C_O}{C_A},$$

where $C_O/C_A$ is the ratio of the solute concentration in octanol, to the concentration in air phases. Unfortunately, there are only two experimental values measured directly (Harner and Bidleman, 1998) (Table 13). However, it is also possible to approximate $K_{OA}$ as being the ratio of octanol–water and air–water partition coefficients ($K_{OW}/K_{AW}$) or inversely proportional to the vapor pressure, where $K_{OW}$ represents the partitioning of chemicals between octanol saturated with water and water saturated with octanol, whereas $K_{AW}$ represents the partitioning between air and pure water. These are approximated values and there may be errors especially due to error propagation caused by variability in literature values of $K_{OW}$ and $K_{AW}$. Experimental results (Mackay and Callcot, 1998) at 25°C are correlated with the physicochemical descriptors, including log $S$ and log $K_{OW}$. A good regression model was obtained using EA, $W$, $X_e$, $T_b$ and log $S$ (Table 13) as descriptors. Only two estimated values of log $K_{OA}$ differ by 5% from their respective experimental results.

Table 9
Actual and estimated[a] solubility values (PLS, three latent variables)

| | Measured log $S$ (g/m$^3$) | Estimated log $S$ (g/m$^3$) | Residual |
|---|---|---|---|
| Naphthalene | 1.50 | 1.434 | 0.066 |
| Anthracene | −1.13 | −1.066 | −0.064 |
| Phenanthrene | 0.061 | −0.205 | **0.266** |
| Naphthacene | −3.24 | −3.292 | 0.052 |
| Benz[a]anthracene | −1.95 | −1.741 | −0.209 |
| Chrysene | −2.48 | −1.800 | **−0.680** |
| Triphenylene | −1.37 | −1.416 | 0.046 |
| Pyrene | −0.87 | −1.029 | **0.159** |
| Perylene | −3.40 | −3.057 | −0.343 |
| Benzo[a]pyrene | −2.42 | −2.359 | −0.061 |
| Benzo[e]pyrene | −2.25 | −2.577 | **0.327** |
| Dibenz[a,h]anthracene | −3.22 | −3.214 | −0.006 |
| Benzo[ghi]perylene | −3.59 | −3.588 | −0.002 |
| Coronene | −3.85 | −4.172 | 0.322 |
| SEP | | 0.260 | |
| $r$ | | 0.991 | |
| Regression coefficient | EA | 0.090 | |
| | $X_e$ | −0.278 | |
| | RI | −0.518 | |
| | SArea | 0.600 | |
| | $\Delta H_f$ | −0.669 | |

[a] Estimated by leave-one-out crossvalidation.

Table 10
Predicted values of log $S$[a]

| | log $S$ | Literature |
|---|---|---|
| Benzene | 3.34 | 3.25[b] |
| Benzo[c]phenanthrene | −1.16 | |
| Picene | −3.79 | −2.6[c] |
| Pentaphene | −2.93 | |
| Benzo[b]chrysene | −3.45 | |
| Dibenz[a,j]anthracene | −3.16 | −1.92[c] |
| Dibenz[a,c]anthracene | −2.00 | −1.64[d], −2.8[b] |
| benzo[c]chrysene | −2.02 | |
| Dibenzo[c,g]phenanthrene | −1.74 | |
| Benzo[a]naphthacene | −3.49 | |
| Dibenzo[e,l]pyrene | −3.43 | |
| Dibenzo[de,qr]naphthacene | −2.56 | |
| Dibenzo[g,p]chrysene | −2.09 | |
| Benzo[b]perylene | −2.82 | |
| Benzo[a]perylene | −2.01 | |
| dibenzo[a,n]triphenylene | −2.67 | |
| Benzo[h]pentaphene | −2.61 | |
| Dibenzo[a,e]pyrene | −2.94 | |
| Dibenzo[a,l]pyrene | −2.08 | |
| Benzo[g]chrysene | −1.26 | |
| Dibenzo[b,g]phenanthrene | −2.04 | |

[a] log $S$ (g/m$^3$).
[b] From Mackay et al. (1992).
[c] From Pearlman et al. (1984).
[d] From Karcher (1988, p. 21).

The resulting model was used to predict the log $K_{OA}$ for some unknown PAHs and the results are in Table 14. The predicted values for dibenzo anthracenes (13.38 and 13.19) have the same order of magnitude as the experimental data for dibenz[a,h]anthracene (13.91). Naphthacene and triphenylene predicted results that are close to the experimental values for benz[a]anthracene and chrysene, which seems very reasonable since they are all catacondendesed compounds having four aromatic rings.

### 4.6. Soil sorption ($K_{OC}$)

For most of the non-ionic organic chemicals, the content of organic-carbon is the determining factor of soil adsorption making the organic-carbon partition coefficient $K_{OC}$, the most used parameter to indicate soil mobility in soil–water systems. Compounds with higher values of log $K_{OC}$ tend to be less mobile than those with lower values. Several coefficients were available for some PAHs in the literature and so, an average value including experimental data for sediments were taken. The experimental data in this section was also taken from Mackay et al. (1992) except for benzo[ghi]perylene (Luers and Hulscher, 1996) which was measured at 20°C. According to the authors, the influence of temperature is very small. Fig. 10 shows good correlation of log $K_{OC}$ with several calculated descriptors, $X_v$, $T_b$, log $S$, log $K_{OW}$, etc. It is well known that mobility depends on water solubility ($S$) and hydrophobicity ($K_{OW}$) (Gawlik et al., 1997). $K_{OC}$ increases with molecular size which is expected since bigger molecules are extremely hydrophobic meaning that their concentration in water remains small and accumulation in sediments is significant.

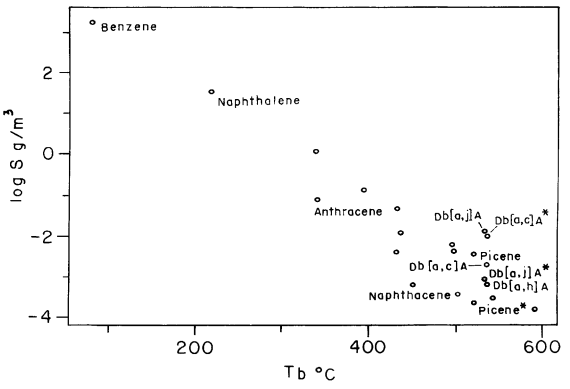



Fig. 8. Plot of log $S$ versus boiling temperature.

Table 15 contains the experimental and estimated results obtained by PLS and the regression coefficients. Only the estimated values for naphthacene, benz[a]anthracene, pyrene differ by more than 2% from the respective experimental coefficients.

The resulting model (Table 15) was used to predict log $K_{OC}$ for some PAHs and the results are in Table 16. The predicted partition coefficient for chrysene and coronene are in good agreement with those reported from literature. On the other hand, the predicted value for triphenylene and benzo[e]pyrene in this work (5.20 and 6.28), differs greatly from the literature data reported by Mackay et al. (1992). However, being very similar to the experimental value for benzo[a]pyrene (6.46), it is an indication that the present result is more reasonable.

Table 11
Actual and estimated values of log $K_{OW}$, and their residuals (PLS model using three latent variables)

| | Measured log $K_{OW}$[a] | Estimated | Residuals | Exp. Literature[b] |
|---|---|---|---|---|
| Benzene | 2.13 | 2.063 | 0.067 | |
| Naphthalene | 3.33 | 3.359 | −0.029 | 3.33 |
| Anthracene | 4.54 | 4.593 | −0.054 | 4.68 |
| Phenanthrene | 4.55 | 4.552 | −0.002 | 4.57 |
| Naphthacene | 5.96[c] | 5.857 | 0.103 | |
| Benz[a]anthracene | 5.91[c] | 5.806 | 0.104 | 5.91 |
| Chrysene | 5.84 | 5.758 | 0.082 | 5.81 |
| Triphenylene | 5.45 | 5.668 | **−0.218** | |
| Pyrene | 5.14 | 5.244 | −0.104 | |
| Benzo[c]phenanthrene | 5.84[c] | 5.750 | 0.090 | |
| Perylene | 6.30 | 6.114 | 0.186 | |
| Benzo[a]pyrene | 6.30 | 6.388 | −0.088 | 6.13 |
| Dibenz[a,h]anthracene | 6.75[c] | 6.998 | **0.248** | |
| Dibenz[a,c]anthracene | 7.19[c] | 6.724 | **0.466** | |
| Pentacene | 7.19[c] | 7.145 | 0.045 | |
| Benzo[a]naphthacene | 6.81 | 7.136 | **−0.326** | |
| Benzo[ghi]perylene | 6.87 | 6.760 | 0.110 | 6.22 |
| SEP | | 0.179 | | |
| $r$ | | 0.996 | | |
| Regression coefficient | EA | −0.0148 | | |
| | $X_e$ | 0.509 | | |
| | RI | 0.224 | | |
| | SArea | 0.120 | | |
| | $\Delta H_f$ | 0.173 | | |

[a] Mackay (1982).
[b] De Maagd et al. (1998).
[c] Sangster (1989).


Fig. 9. Plots of log $K_{OW}$ versus descriptors, showing their correlation.

Table 12
Predicted and calculated values of $\log K_{OW}$ by PLS model and regression equation

| | $\log K_{OW}$ | Calculated |
|---|---|---|
| Benzo[e]pyrene | 6.30 | 6.44[a] |
| Picene | 6.97 | 7.11[b] |
| Pentaphene | 7.00 | 6.84 |
| Benzo[b]chrysene | 7.03 | 7.11[b] |
| Dibenz[a,j]anthracene | 6.87 | 7.11[a] |
| benzo[c]chrysene | 6.97 | 6.84 |
| Dibenzo[c,g]phenanthrene | 6.84 | 6.84 |
| Dibenzo[a,h]pyrene | 7.72 | 7.30 |
| Anthanthrene | 6.94 | 7.04[b] |
| Dibenzo[e,l]pyrene | 7.41 | 7.30 |
| Dibenzo[de,qr]naphthacene | 7.56 | 7.30 |
| Benzo[b]perylene | 7.53 | |
| Benzo[a]perylene | 7.64 | |
| Coronene | 7.47 | 7.64[a], 5.4[c] |
| Dibenzo[a,e]pyrene | 7.48 | |
| Dibenzo[a,l]pyrene | 7.64 | |
| Dibenzo[a,i]pyrene | 7.66 | |
| Benzo[g]chrysene | 6.96 | 6.84 |
| Dibenzo[b,g]phenanthrene | 7.00 | 6.84 |

[a] From Mackay et al. (1992).
[b] Sangster (1989).
[c] Experimental value.

### 4.7. Henry's law constant (H)

The Henry's law constant is essentially an air–water partitioning coefficient differing only by a scale factor when the temperature is kept constant since $K_{AW} = H/RT$, where $R$ is the gas constant. $K_{AW}$ is important for controlling the cycling of compounds across air–water interface and it can be determined by measurement of solute concentrations in both phases. In 1981, Mackay and Shiu (1981) has reviewed the methods of determining, estimating and expressing these coefficients. Henry's Law coefficient can also be expressed as the ratio ($P/C$) where $P$ is partial vapor pressure in the gas or air and $C$ the solute concentration in water.

Two sets of data listed in Table 17 were initially considered. One of them calculated by the ratio $P/C$ at 25°C (Mackay and Callcot, 1998) and the other measured experimentally by de Maagd et al. (1998) and Bamford et al. (1999). De Maagd et al. determined $H$ directly by using the gas-purge method, where concentrations in the water phase and gas phase were measured during a dynamic air/water exchange process at 20°C. Bamford et al. also measured $H$ directly, between 4°C and 31°C, but the method of choice was the gas stripping. The data used in this work was recalculated for 20°C since the temperature dependence can be expressed by the van't Hoff equation. Fig. 11 shows the plots of experimental and calculated $\log H$ versus descriptors,

Table 14
Predicted $\log K_{OA}$ coefficients using PLS model

| | $\log K_{OA}$ |
|---|---|
| Naphthacene | 10.176 |
| Triphenylene | 10.355 |
| Benzo[e]pyrene | 11.481 |
| Picene | 14.089 |
| Dibenz[a,j]anthracene | 13.380 |
| Dibenz[a,c]anthracene | 13.189 |
| Benzo[ghi]perylene | 13.072 |

Table 13
Actual, estimated values of $\log K_{OA}$ by PLS modeling (two latent variables) and regression coefficients

| | Measured $\log K_{OA}$[a] | Estimated | Residual | $\log K_{OA}$ literature |
|---|---|---|---|---|
| Naphthalene | 5.13 | 5.014 | 0.116 | |
| Anthracene | 7.34 | 7.445 | −0.105 | |
| Phenanthrene | 7.45 | 7.655 | −0.205 | 7.57[b] |
| Benzo[a]anthracene | 10.80 | 10.175 | **0.625** | |
| Chrysene | 10.44 | 10.588 | −0.148 | |
| Pyrene | 8.43 | 8.295 | 0.135 | 8.80[b] |
| Perylene | 11.70 | 11.404 | 0.296 | |
| Benzo[a]pyrene | 10.71 | 11.545 | **−0.835** | |
| Dibenz[a,h]anthracene | 13.91 | 13.839 | 0.071 | |
| SEP | | 0.378 | | |
| $r$ | | 0.994 | | |
| Regression coeff. | EA | −0.190 | | |
| | $W$ | 0.412 | | |
| | $X_e$ | 0.267 | | |
| | $T_b$ | 0.251 | | |
| | $\log S$ | −0.2201 | | |

[a] Experimental values from Mackay and Callcot (1998).
[b] Harner and Bidleman (1998), directly measured.

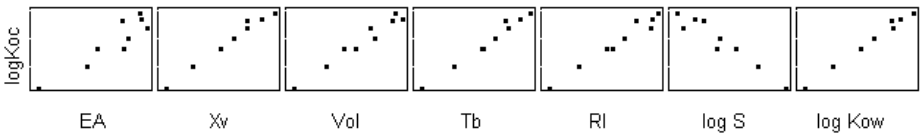



Fig. 10. Plots of log $K_{OC}$ versus descriptors showing their correlation.

Table 15
Experimental and PLS estimated values of log $K_{OC}$[a]

| | Measured[b] log $K_{OC}$ | Estimated | Residual |
|---|---|---|---|
| Benzene | 1.57 | 1.550 | 0.020 |
| Naphthalene | 3.04 | 3.011 | 0.029 |
| Anthracene | 4.37 | 4.356 | 0.014 |
| Phenanthrene | 4.36 | 4.307 | 0.053 |
| Naphthacene | 5.81 | 6.023 | **−0.213** |
| Benz[a]anthracene | 6.30 | 5.901 | **0.399** |
| Pyrene | 5.03 | 5.227 | **−0.197** |
| Benzo[a]pyrene | 6.46 | 6.593 | −0.133 |
| Benzo[ghi]perylene | 6.80[c] | 6.693 | 0.107 |
| SEP | | 0.175 | |
| *r* | | 0.9987 | |
| Regression coefficient | $X_v$ | −0.0685 | |
| | Vol | 0.598 | |
| | $T_b$ | −0.240 | |
| | log *S* | −0.396 | |
| | log $K_{OW}$ | 0.338 | |

[a] Three latent variables. Dibenz[a,h]anthracene excluded as outlier.
[b] From Mackay et al. (1992).
[c] From Luers and Hulscher (1996).

Table 16
Predicted soil sorption coefficients and reported values from literature

| | log $K_{OC}$ | Literature[a] |
|---|---|---|
| Chrysene | 6.05 | 6.27, 6.90 |
| Triphenylene | 5.20 | 6.90 |
| Perylene | 6.07 | |
| Benzo[e]pyrene | 6.28 | 7.20 |
| Picene | 7.13 | |
| Dibenz[a,h]anthracene | 7.19 | 6.31, 6.52 |
| Dibenz[a,j]anthracene | 7.20 | |
| Dibenz[a,c]anthracene | 7.69 | |
| Coronene | 7.38 | 7.80 |

[a] From Mackay et al. (1992).

where it can be noted that the experimental values are better correlated with them and so will be used throughout the analysis. There is a linear decrease of log *H* with $T_b$ and log $K_{OW}$ (Fig. 11) and increase with log *S* which means that as the molecular size increases, the concentration in water becomes smaller (higher hydrophobicity) and also the volatility causing their concentrations in both phases to be very small. That is the main difficulty in measuring the PAHs Henry's law constants.

The estimated results and residuals obtained by PLS regression are listed in Table 18. Benz[a]anthracene was considered an outlier and so, excluded during the modeling step. Although the correlation coefficient is very high, the residuals are in the range of 10%.

The model was used to predict the Henry's law constant of 9 PAHs at 20°C (see Table 19). The predicted Henry's law constant for perylene is 0.095 Pa $m^3$/mol, which is much lower than the value calculated reported by Mackay et al. (1992) (0.44. Pa $m^3$/mol calculated *P*/*C*) even at 25°C. The predicted results for benz[a]anthracene, naphthacene and triphenylene (0.18, 0.24 and 0.32 Pa $m^3$/mol, respectively) agree with the experimental data of its isomer chrysene (0.26 Pa $m^3$/mol).

### 4.8. Bioconcentration factor (BCF)

While the uptake of PAHs (bioconcentration) in terrestrial invertebrates comes from food or epidermal contact with interstitial water, in fishes and aquatic invertebrates it comes from the environment (Van-Brummelen et al., 1998).

The bioconcentration factor is derived from measured concentration of a chemical present in an aquatic organism and the ambient environment as

$$\mathrm{BCF} = \frac{C_{org}}{C_W},$$

where $C_{org}$ is the concentration in target organism (μg/kg) and $C_W$ is the concentration in (pure) water (μg/l) or calculated from rate constants.

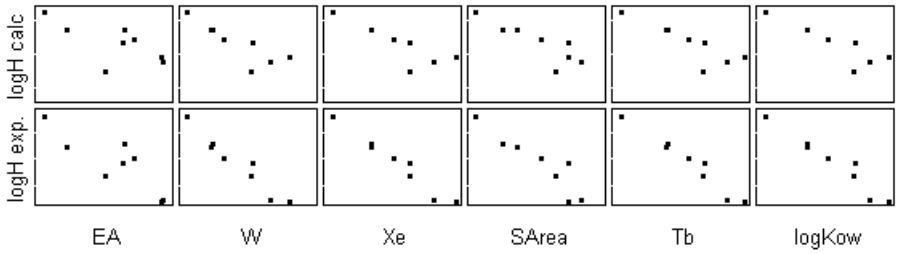
A summary of BCFs for some PAHs in the crustacean *Daphnia-Magna* are presented in Table 20. The correlation among BCF and some descriptors can be visualized in Fig. 12. Note that BCF in this aquatic invertebrate shows a high correlation with log $K_{OW}$, and log *S*. With an increase in the hidrophobicity (inversely related to water solubility), a higher mobility of PAHs to the target organism is expected.

An interesting linear trend is observed (Fig. 13) among BCF and ($K_{OC}$), what is not surprising since both of them are directly related to the organic carbon content.

Table 17
Experimental Henry's Law constants and calculated values[a] from literature Pa $m^3$/mol

| | $H$[b] ($P/C$) (25°C) | log $H$ | $H$ exp[c] (20°C) | log $H$ exp (20°C) |
|---|---|---|---|---|
| Naphthalene | 43.01 | 1.63 | 45.01 | 1.65 |
| Anthracene | 3.96 | 0.60 | 4.02[d] | 0.60 |
| Phenanthrene | 3.24 | 0.51 | 2.9, 3.05[d] | 0.47 |
| Benz[a]anthracene | 0.58 | −0.24 | 0.76[d] | −0.12 |
| Chrysene | 0.0122 | −1.91 | 0.26[d] | −0.59 |
| Pyrene | 0.92 | −0.036 | 1.25[d] | 0.097 |
| Perylene | 0.003 | −2.52 | | |
| Benzo[a]pyrene | 0.046 | −1.34 | 0.034 | −1.47 |
| Dibenz[a,h]anthracene | 0.00017 | −3.70 | | |
| Benzo[ghi]perylene | 0.075[e] | −1.12 | 0.027 | −1.57 |

[a] Calculated as a ratio of vapor pressure and concentration ($P/C$).
[b] From Mackay and Callcot (1998).
[c] From de Maagd et al. (1998).
[d] From Bamford et al. (1999).
[e] Data from Shiu and Mackay (1997).




* Calculated as a ratio of vapor pressure and concentration (25 °C). †(20 °C).

Fig. 11. Plots of log $H$ calculated (calculated as a ratio of vapor pressure and concentration (25°C)) and log $H$ experimental (20°C) versus the descriptors.

Table 18
Results from PLS modeling[a] for Henry's Law constants

| | | Measured log $H$ | Estimated | Residual |
|---|---|---|---|---|
| Naphthalene | | 1.65 | 1.679 | −0.029 |
| Anthracene | | 0.60 | 0.623 | −0.023 |
| Phenanthrene | | 0.47 | 0.557 | −0.087 |
| Chrysene | | −0.59 | −0.632 | 0.042 |
| Pyrene | | 0.097 | −0.089 | **0.186** |
| Benzo[a]pyrene | | −1.47 | −1.192 | **−0.278** |
| Benzo[ghi]perylene | | −1.57 | −1.769 | **0.199** |
| SEP | | | 0.152 | |
| $r$ | | | 0.995 | |
| Regression coeff. | $W$ | | −0.201 | |
| | $X_e$ | | −0.202 | |
| | SArea | | −0.199 | |
| | $T_b$ | | −0.202 | |
| | log $K_{OW}$ | | −0.202 | |

[a] One latent variable; (20°C). Benz[a]anthracene has been excluded as an outlier.

Table 19
Predicted Henry's Law constants[a]

| | log $H$ | $H$ |
|---|---|---|
| Benz[a]anthracene | −0.74 | 0.18 |
| Perylene | −1.02 | 0.095 |
| Naphthacene | −0.62 | 0.24 |
| Triphenylene | −0.49 | 0.32 |
| Benzo[e]pyrene | −1.17 | 0.068 |
| Picene | −1.87 | 0.013 |
| Dibenz[a,h]anthracene | −1.93 | 0.012 |
| Dibenz[a,j]anthracene | −1.92 | 0.012 |
| Dibenz[a,c]anthracene | −2.13 | 0.007 |

[a] (20°C). Pa $m^3$/mol.

A good regression model was obtained using $X_v$, Vol, log $K_{OW}$ and log $S$, with a correlation coefficient of 0.933. The results are in Table 20 and visualized in Fig. 14. The residuals obtained by cross-validation for phenanthrene and triphenylene are greater than 10%.

Table 20
Experimental and estimated[a] values of bioconcentration factor of PAHs in aquatic invertebrate *Daphnia-Magna*

| | Measured log BCF | Estimated log BCF | Residual |
|---|---|---|---|
| Naphthalene | 2.12[b] | 2.213 | −0.093 |
| Anthracene | 2.95 | 2.943 | 0.007 |
| Phenanthrene | 2.51 | 2.973 | **−0.463** |
| Benz[a]anthracene | 4.00 | 3.716 | 0.284 |
| Chrysene | 3.785 | 3.733 | 0.052 |
| Triphenylene | 3.96 | 3.428 | **0.532** |
| Pyrene | 3.43 | 3.279 | 0.151 |
| Perylene | 3.86 | 3.989 | −0.129 |
| Benzo[a]pyrene | 3.82 | 4.057 | −0.237 |
| Dibenz[a,h]anthracene | 4.00 | 4.478 | −0.478 |
| Benzo[ghi]perylene | 4.45 | 4.245 | 0.205 |
| SEP | | 0.295 | |
| *r* | | 0.933 | |
| Regression coefficient | $X_v$ | 0.246 | |
| | Vol | 0.225 | |
| | $\log K_{OW}$ | 0.246 | |
| | $\log S$ | −0.238 | |

[a] PLS modeling using leaving-one-out crossvalidation. One latent variable.
[b] From *Daphnia-Pulex*. Mackay et al. (1992).

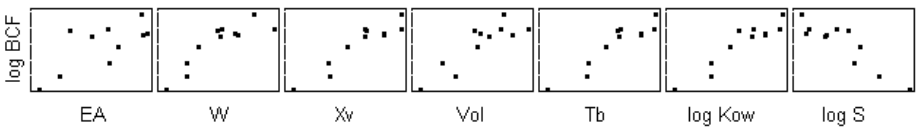
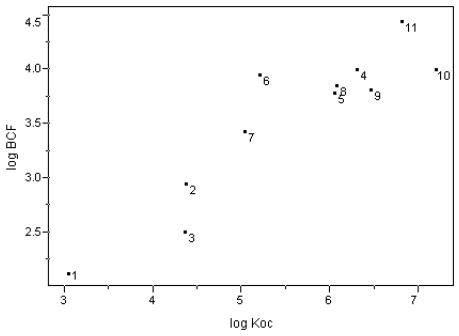



Fig. 12. Scatter plots of bioconcentration factor (log BCF) versus descriptors.




1-Naphthalene, 2-Anthracene, 3-Phenanthrene, 4-Benz[a]anthracene, 5-Chrysene, 6-Triphenylene, 7-Pyrene, 8-Perylene, 9-Benzo[a]pyrene, 10-Dibenz[a,h]anthracene, 11-Benzo[ghi]perylene.

Fig. 13. Plot log/log of bioconcentration factor versus $K_{OC}$.

Concluding this section, it is shown that by using a small set of descriptors, reliable structure–property relationships could be obtained for four different physical properties of PAHs which are of great interest in environmental sciences ($K_{OC}$, $K_{OA}$, $H$ and BCF). All the models obtained showed to have high correlation coefficients and in average small leave-one out cross-validated residuals. Prediction was done for several compounds for these properties. The next section will focus in the phototoxicity and its relation with the electronic descriptors (IP, EA and GAP).

### 4.9. Photo-induced toxicity

This term is used for the phenomenon of increasing the toxicity of certain PAHs when exposed to UV light. The formation of free radicals and subsequent damage of a variety of macromolecules is probably the cause of increase of phototoxicity. Anthracene, pyrene, benzo[a]pyrene, benz[a,h]anthracene and benzo[ghi]perylene are among the most phototoxic compounds, whereas phenanthrene and triphenylene are not phototoxic. Anthracene is bound to DNA through covalent bonds after exposure to UV-A radiation (Sinha and Chignell, 1983).

In an attempt to model photo-induced toxicity, Newsted and Giesy (1987) proposed a nonlinear model using the lowest singlet and triplet excited states energies obtained from molecular spectroscopy. Mekenyan et al. (1994) refined the previous model, considering the photo-induced toxicity as being a result of competing effects:

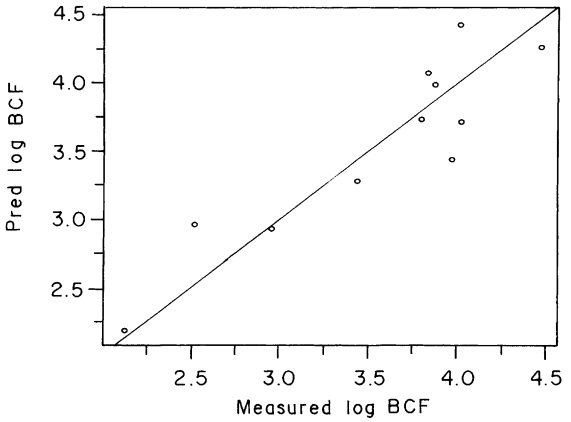
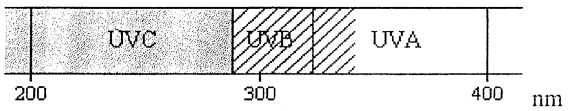



Fig. 14. Predicted versus measured log bioconcentration factor.

light absorption, molecular stability, irradiation energy and irradiation intensity.

As pointed out by Mekenyan et al., the electronic descriptor GAP could be useful to approach the observed phototoxicity since it is related with the energy necessary to excite the molecule from HOMO to LUMO. With increasing GAP, the light absorption shifts to a shorter wavelength and the potential phototoxicity should increase. The compounds stability, which can be also associated with the GAP through the hardness defined as GAP/2, also affects the phototoxicity. Increasing the GAP should enhance the stability of aromatic compounds. On the other hand, with decreasing wavelength, the intensity in sunlight slowly decreases to zero according to the chart




by which solar radiation of wavelengths below 290 nm should not reach the earth's surface because of absorption by ozone formed in the stratosphere. The UVB, UVA, visible (400–700 nm) and infrared (700–3000 nm) solar radiation reach the earth's surface, corresponding, respectively, to 0.1%, 4.9%, 39.0% and 56.0% of the total solar radiant energy (Mandronich et al., 1995). Therefore, it is expected that a nonlinear function of electronic descriptors be useful to model the phototoxicity.

In the present work, the electronic descriptors IP, EA and GAP were transformed according to

| | |
|---|---|
| IP | $\exp\text{-}(IP\text{-}8.06)^2$ |
| EA | $\exp\text{-}(EA\text{-}.91)^2$ |
| GAP | $\exp\text{-}(GAP\text{-}7.2)^2$ |

and used to model the phototoxicity.

Table 21 contains the photo-induced toxicity (log 1/LT50) for *Daphinia-Magna* (Newsted and Giesy, 1987) and some absorption wavelengths with their respective molar absorptivities. Fig. 15 shows the correlation between the transformed descriptors and phototoxicity.

PLS model was built using the experimental data and two latent variables. The estimated results (Table 21) are within 5% error compared to experimental data. Phototoxicity for all other PAHs considered in this work were predicted (results are in Table 22). From the predicted values, it can be concluded that pentacene, dib-

Table 21
Measured, and estimated photoinduced toxicity of PAHs to invertebrate *Daphnia-Magna* (two latent variables) and absorption wavelengths (nm) with the corresponding molar absorption coefficients ($\varepsilon$)[a]

| | Measured Photot.[b] | Estimated | Residual | Absorbed $\lambda$ (nm) and $\varepsilon$ |
|---|---|---|---|---|
| Anthracene | −2.46 | −2.38 | −0.08 | 356 (8.5), 374 (8.5) |
| Phenanthrene | −3.20 | −3.18 | −0.02 | 292 (14.8) |
| Naphthacene | −2.89 | −2.88 | −0.01 | 294 (29.5), 442 (10.4), 471 (14.0) |
| Benz[a]anthracene | −2.52 | −2.39 | −0.13 | 288 (83), 300 (10), 341 (71) |
| Chrysene | −2.54 | −2.54 | −0.00 | 282, 294, 307, 320 (11.6–12.6) |
| Triphenylene | −3.20 | −3.21 | 0.01 | 273 (18.8), 284 (17.6) |
| Pyrene | −2.31 | −2.38 | 0.07 | 306 (11.8), 319 (31), 335 (55) |
| Perylene | −2.67 | −2.64 | −0.03 | 342 (12.2), 386 (13.4), 408 (29), 435 (39.5) |
| Benzo[a]pyrene | −2.43 | −2.55 | 0.12 | 297 (63), 365 (25), 385 (29.5) |
| Dibenz[a,h]anthracene | −2.44 | −2.41 | −0.03 | 297 (148), 321 (19.4), 333 (17.2), 350 (15.8) |
| Benzo[ghi]perylene | −2.36 | −2.46 | 0.10 | – |
| SEP | | 0.096 | | |
| $r$ | | 0.973 | | |
| Regression coefficient | $\exp\text{-}(IP-8.06)^2$ | −0.678 | | |
| | $\exp\text{-}(EA-0.91)^2$ | −0.043 | | |
| | $\exp\text{-}(GAP-7.2)^2$ | −0.258 | | |

[a] Molar absorptivity $\varepsilon$ in parenthesis: l/cm mol.
[b] Newsted and Giesy (1987).

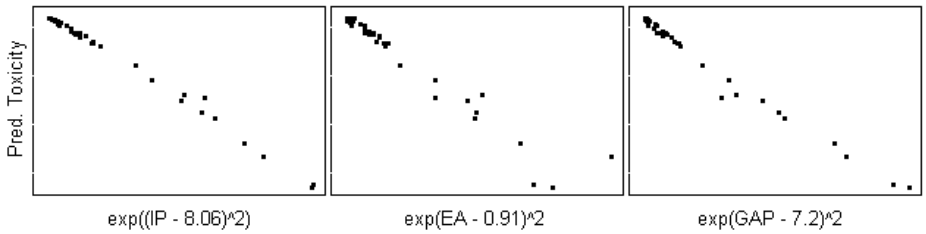



Fig. 15. Plots of photo-induced toxicity (log 1/LT50) versus transformed electronic descriptors showing their correlation.

enzo[a,h]pyrene, anthanthrene, benzo[a]perylene, dibenzo[de,mn]naphthacene and benzo[a]pentacene should not exhibit phototoxicity. Fig. 16 shows the plot of GAP versus measured and predicted phototoxicity, where it can be seen that PAHs in the upper left side should be those showing a very high phototoxicity. Overall, the addition of extra rings decreases the phototoxicity. This can be visualized comparing the isomers dibenz[a,h]anthracene,

Table 22
Predicted photoinduced toxicity of PAHs to invertebrate *Daphnia-Magna*[a]

| | Predicted Photot. | Absorbed $\lambda$ (nm) and $\varepsilon$[b] |
|---|---|---|
| **Naphthalene** | **<−3.3** | 275 (36.4), 286 (3.9) |
| Benzo[c]phenanthrene | −2.58 | 281 (83), 302 (11), 313 (10.2) |
| Benzo[e]pyrene | −2.39 | 290 (57), 307 (13.2), 318 (24.5), 333 (39) |
| Picene | −2.50 | |
| Pentaphene | −2.39 | |
| Benzo[b]chrysene | −2.42 | |
| Dibenz[a,j]anthracene | −2.42 | 286 (75), 297 (129), 322 (16.6), 337 (14.6) |
| Dibenz[a,c]anthracene | −2.43 | 287 (133) |
| Benzo[c]chrysene | −2.54 | |
| **Pentacene** | **<−3.3** | 290 (56), 302 (148), 330 (8.9) |
| Dibenzo[c,g]phenanthrene | −2.46 | |
| Benzo[a]naphthacene | −2.68 | 290 (84), 303 (125), 316 (113), 421 (10), 448 (11) |
| **Dibenzo[a,h]pyrene** | **−3.23** | |
| **Anthanthrene** | **<−3.3** | |
| Dibenzo[a,j]naphthacene | −2.55 | |
| Dibenzo[a,l]naphthacene | −2.54 | |
| Dibenzo[a,c]naphthacene | −2.58 | |
| Dibenzo[e,l]pyrene | −2.42 | |
| Dibenzo[de,qr]naphthacene | −2.42 | |
| Dibenzo[g,p]chrysene | −2.39 | 300 (67), 337 (15.4), 351 (14.8) |
| Benzo[c]picene | −2.41 | |
| Benzo[ghi]perylene | −2.46 | |
| Dibenzo[b,k]chrysene | −2.65 | |
| Dibenzo[c,l]chrysene | −2.40 | |
| Benzo[b]perylene | −2.64 | |
| **Benzo[a]perylene** | **<−3.3** | |
| **Dibenzo[de,mn]naphthacene** | **<−3.3** | |
| Dibenzo[a,n]triphenylene | −2.40 | |
| Benzo[h]pentaphene | −2.44 | |
| **Benzo[a]pentacene** | **<−3.3** | |
| Coronene | −2.39 | 302 (278), 333 (17.4), 345 (12) |
| Dibenzo[a,e]pyrene | −2.45 | |
| Dibenzo[a,l]pyrene | −2.52 | 290 (37.5), 317 (18.8), 331 (26.5), 402 (12.6) |
| Dibenzo[a,i]pyrene | −2.65 | |
| Benzo[g]chrysene | −2.42 | 286 (68), 309 (11), 320 (11.8), 333 (8.6) |
| Dibenzo[b,g]phenanthrene | −2.41 | |
| Naphtho[2,1,8 qra]naphthacene | −3.02 | |

[a] Highlighted PAHs are not expected to be phototoxic.
[b] Molar absorptivity $\varepsilon$ in parenthesis: I/cm mol.

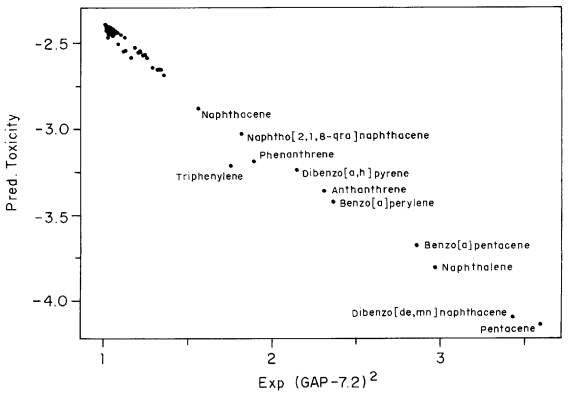



Fig. 16. Predicted photo-induced toxicity versus transformed GAP.

dibenz[a,j]anthracene and dibenz[a,c]anthracene which have similar phototoxicity 2.44, 2.42, and 2.42, respectively, to their 6 rings analogues dibenzo[a,j]naphthacene, dibenzo[a,l]naphthacene, dibenzo[a,c]naphthacene (2.55, 2.54 and 2.58, respectively).

## 5. Conclusions

The physicochemical properties of PAH compounds are a direct function of their size, topology and electronic distribution. The information encoded in one type of variable, for example molecular connectivity, can describe a contribution to a given physical property which cannot be totally expressed by the volume and so on. The use of a single descriptor can capture only a part of the property of interest, or of some occurring process, which is in many cases far from satisfactory. The use of multivariate regression instead, is a great improvement by correlating physical properties with molecular parameters. Methods based on principal component analysis, such as PLS, are advantageous by dealing well with strongly correlated descriptors, besides allowing them to estimate more than one coefficient for each property. Good regression models were built for $T_b$, RI, solubility and $K_{OW}$. Structure–property relationships were successfully obtained for four different properties, $K_{OC}$, $K_{OA}$, Henry's law constant and BCF, which are of direct application in environmental sciences. The photo-induced toxicity of the studied PAHs was also modeled using transformed electronic descriptors IP, EA and GAP. In all cases, properties were predicted for those PAHs not found in the literature.

## Acknowledgements

The author kindly thanks Dr. Rudolf Kiralj for doing some calculations and discussions and also Lucicleide Ribeiro Cirino for her great help and encouragement.

## References

Alberty, R.A., Chung, M.B., Reif, A.K., 1989. Standard chemical thermodynamic properties of polycyclic aromatic hydrocarbons and their isomer groups. II. Pyrene series, naphthopyrene series, and coronene series. J. Phys. Chem. Ref. Data 18, 77–108.

Alberty, R.A., Reif, A.K., 1988. Standard chemical thermodynamic properties of polycyclic aromatic hydrocarbons and their isomer groups. I. Benzene series. J. Phys. Chem. Ref. Data 17, 241–253.

Arfsten, D.P., Schaeffer, D.J., Mulveny, D.C., 1996. The effects of near ultraviolet radiation on the toxic effects of polycyclic aromatic hydrocarbons in animals and plants: a review. Ecotoxicol. Environ. Safety 33, 1–24.

Ariese, F., Kok, S.J., Verkaik, M., Gooijer, C., Velthorst, N.H., Hofstraat, J.W., 1993. Synchronous fluorescence spectrometry of fish bile – a rapid screening method for the biomonitoring of PAH exposure. Aquat. Toxicol. 26, 273–286.

Bamford, H.A., Poster, D.L., Baker, J.E., 1999. Temperature dependence of Henry's law constants of thirteen polycyclic aromatic hydrocarbons between 4°C and 31°C. Environ. Toxicol. Chem. 18, 1905–1912.

Beebe, K.R., Pell, R.J., Seasholtz, M.B., 1998. Chemometrics: A Practical Guide. Wiley, New York.

Boehm, P.D., Page, D.S., Gilfillan, E.S., Bence, A.E., Burns, W.A., Mankiewicz, P.J., 1998. Study of the fates and effects of the Exxon Valdez oil spill on benthic sediments in two bays in Prince William Sound, Alaska. 1. Study design, chemistry, and source fingerprinting. Environ. Sci. Technol. 32, 567–576.

De Maagd, P.G.J., TenHulscher, D.T.E.M., VandenHeuvel, H., Opperhuizen, A., Sijm, D.T.H.M., 1998. Physicochemical properties of polycyclic aromatic hydrocarbons: aqueous solubilities, *n*-octanol/water partition coefficients, and Henry's law constants. Environ. Toxicol. Chem. 17, 251–257.

Estrada, E., 1995. Edge adjacency relationships and a novel topological index related to molecular volume. J. Chem. Inf. Comp. Sci. 35, 31–33.

Faber, J.H., Heijmans, J.S.M., 1996. Polycyclic aromatic hydrocarbons in soil detritivores. In: Van Straalen, N.M., Krivolutskii, D.A. (Eds.), Bioindicator Systems for Soil Pollution. Kluwer Academic Publishers, Dordrecht, MA, pp. 31–43.

Franz, T.P., Eisenreich, S., 1998. Snow scavenging of polychlorinated biphenyls and polycyclic aromatic hydrocarbons in Minnesota. J. Environ. Sci. Technol. 32, 1771–1778.

Gawlik, B.M., Sotiriou, N., Feicht, E.A., Schulte-Hostede, S., Kettrup, A., 1997. Alternatives for the determination of the soil adsorption coefficient, $K_{OC}$, of non-ionic organic compounds – a review. Chemosphere 34, 2525–2551.

Gerstl, Z., 1990. J. Contam. Hydrol. 6, 357–375.

Govers, H., Ruepert, C., Aiking, H., 1984. Quantitative structure–activity relationships for polycyclic aromatic hydrocarbons: correlation between molecular connectivity, physico-chemical properties, bioconcentration and toxicity in daphinia pulex. Chemosphere 13, 227–236.

Harner, T., Bidleman, T.F., 1998. Measurement of octanol-air partition coefficients for polyclyclic aromatic hydrocarbons and polychlorinated naphthalenes. J. Chem. Eng. Data 43, 40–46.

Hong, H., Wang, L.S., Han, S.K., Zou, G.W., 1996. Prediction adsorption coefficients ($K_{OC}$) for aromatic compounds by HPLC retention factors ($k'$). Chemosphere 32, 343–351.

IARC, 1987. Monographs on the Evaluation of the Carcinogenic Risk of Chemical to Humans, vol. 32, Suppl. 7. International Agency for Research on Cancer, Lion.

Jacob, J., 1996. The significance of polycyclic aromatic hydrocarbons as environmental carcinogens. Pure Appl. Chem. 68, 301–308.

Karcher, W., 1988. Spectral Atlas of Polycyclic Aromatic Compounds, vol. 2. Kluwer Academic Publishers, Dordrecht, MA, p. 16.

Karickhoff, S.W., 1981. Semiempirical estimation of sorption of hydrophobic pollutants on natural sediments and soils. Chemosphere 10, 833–846.

Lu, Y.-J., Lee, S.-L., 1993. Semi-empirical calculations of the nonlinear optical properties of polycyclic aromatic compounds. Chem. Phys. 179, 431–444.

Lu, X., Tao, S., Cao, J., Dawson, R.W., 1999. Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices. Chemosphere 39, 987–999.

Luers, F., Hulscher, Th.E.M., 1996. Temperature effect on the partitioning of polycyclic aromatic hydrocarbons between natural organic carbon and water. Chemosphere 33, 643–657.

Mackay, D., 1982. Correlation of bioconcentration factors. Environ. Sci. Technol. 16, 274–278.

Mackay, D., Shiu, W.-Y., Ma, K.C., 1992. Illustrated Handbook of Physical–Chemical Properties and Environmental Fate of Organic Compounds, vol. 2. Lewis/CRC, Boca Raton.

Mackay, D., Shiu, W.-Y., 1981. Critical review of Henry's law constants for chemicals of environmental interest. J. Phys. Chem. Ref. Data 10, 1175–1199.

Mackay, D., Callcot, D., 1998. Partitioning and physical properties of PAHs. In: Neilson, A.H. (Ed.), The Handbook of Environmental Chemistry, vol. 3, Part J. PAHs and Related Compounds. Springer, Berlin, pp. 325–346.

Mandronich, S., Mckenzie, R.L., Caldwell, M.M., Björn, L.O., 1995. Changes in ultraviolet-radiation reaching the earth's surface. Ambio 24, 143–152.

Martens, H., Naes, T., 1989. Multivariate Calibration. Wiley, New York.

Mekenyan, O.G., Ankley, G.T., Veith, G.D., Call, D.J., 1994. QSARs for photoinduced toxicity. 1. Acute lethality of polycyclic aromatic hydrocarbons to *Daphnia-Magna*. Chemosphere 28, 567–582.

Newsted, J.L., Giesy, J.P., 1987. Predictive models for photoinduced acute toxicity of polycyclic aromatic-hydrocarbons to *Daphnia-Magna*, Strauss (*cladocera*, *crustacea*). Environ. Toxicol. Chem. 6, 445–461.

Pearlman, R.S., Yalkowsky, S.H., Banerjee, S., 1984. Water solubilities of polynuclear aromatic and heteroaromatic compounds. J. Phys. Chem. Ref. Data 13, 555–562.

Randic, M., 1975. Characterization of molecular branching. J. Am. Chem. Soc. 97, 6609–6615.

Sabljic, A., Gutsen, H., Verhaar, H., Hermens, J., 1995. QSAR modeling of soil sorption-improvements and systematics of log $K_{OC}$ vs log $K_{OW}$ correlations. Chemosphere 31, 4489–4515.

Salau, J.S.I, Tauler, R., Bayona, J.M., Tolosa, I., 1997. Input characterization of sedimentary organic contaminants and molecular markers in the Northwestern Mediterranean Sea by exploratory data analysis. Environ. Sci. Technol. 31, 3482–3490.

Sander, L.C., Wise, S.A., Polycyclic Aromatic Hydrocarbon Structure Index. NIST Special Publication 922. http://ois.nist.gov/PAH.

Sander, L.C., Wise, S.A., 1986. Investigations of selectivity in RPLC of polycyclic aromatic hydrocarbons. Adv. Chromatogr. 25, 139–218.

Sangster, J., 1989. Octanol–water partition coefficients of simple organic compounds. J. Phys. Chem. Ref. Data 18, 1111–1229.

Shiu, W.-Y., Mackay, D., 1997. Henry's law constants of selected aromatic hydrocarbons, alcohols and ketones. J. Chem. Eng. Data 42, 27–30.

Sinha, B.K., Chignell, C.F., 1983. Binding of anthracene to cellular macromolecules in presence of light. Photochem. Photobiol. 37, 33–37.

Tolosa, I., Bayona, J.M., Albaiges, J., 1996. Aliphatic and polycyclic aromatic hydrocarbons and sulfur/oxygen derivatives in northwestern Mediterranean sediments: spatial and temporal variability, fluxes, and budgets. Environ. Sci. Technol. 30, 2495–2503.

Van Brummelen, T.C., Van Hattum, B., Crommentuijn, T., Kalf, D.F., 1998. Biovailability and ecotoxicity of PAHs. In: Neilson A.H. (Ed.), The Handbook of Environmental Chemistry, vol. 3, Part J. PAHs and Related Compounds. Springer, Berlin, pp. 203–263.

Warne, M.St.J., Connell, D.W., Hawker, D.W., Schuurmann, G., 1990. Chemosphere 21, 877–888.

White, C.M., 1986. Prediction of the boiling point, heat of vaporization and vapor pressure at various temperatures for polyclyclic aromatic hydrocarbons. J. Chem. Eng. Data 31, 198–203.

Wiener, H., 1947. Structural determination of paraffin boiling points. J. Am. Chem. Soc. 69, 17–20.