

Temporal and spatial variations of polycyclic aromatic hydrocarbon concentrations around a coke oven plant.

Supplementary material.

- Appendix A: Sample Analysis Method 2
- Appendix B: explorative analysis of the preliminary data set (Jan 1995- Oct 1996) 3
- Appendix C: Statistical Analyses 7
 - Data sets
 - Annual Moving Averages
 - Data distribution (Table SC1: Statistical data of annual BaP_{eq} concentrations)
 - Regression analyses
 - Statistical Comparisons of PAH profiles and Diagnostic PAH ratios
- Appendix D: Figures 13
 - Figure SD1 Correlation between annual average coke BaP concentrations and distance from the coke oven
 - Figure SD2 Relationship between BaP_{eq} concentrations and downwind hours during sampling

APPENDIX A

Sample Analyses Method

Sample Extraction. Each high-volume glass fibre filter (Gelman Type A/E 20 \times 25 cm) was extracted by ultrasound sonication with cyclohexane, adding 3,6-dimethyl-phenanthrene and indeno (1,2,3-cd) fluoranthene as internal standards. Extraction was repeated two other times.

The extracts were pooled and concentrated using rotary evaporator, then further concentrated to about 0.2 ml using a stream of nitrogen. Concentrated extract was applied to a preconditioned SPE glass column (Chromabond CN/SiOH, by Macherey-Nagel). The PAH fraction was eluted with 8 ml of a n-pentane and dichloromethane mixture (80:20 v/v), concentrated near to dryness by a nitrogen flow and solved with 1 ml of toluene. All solvents were of Lichrosolv (Merck) grade.

Low-volume filters from site 1 or 2, of seven consecutive days, were pooled and altogether extracted twice by soxhlet with cyclo-hexane for 24 hours, then purified following the previously described method.

PAH Analysis. The qualitative/quantitative identification of PAH was performed using a Perkin Elmer (PE) Autosystem gas chromatograph (GC) equipped with an auto-sampler and a flame ionization detector (FID). Analyses of PAHs were performed on a 5% dimethyl-syloxane capillary column (length: 30 m; 0.25 mm i.d). One micro litre of the PAH solution was injected by split-less mode, with the injector at 300 °C, FID at 320°C. Programmed column temperature was: 80 °C for 1 min, a ramp 10 °C/min to 290 °C, hold at 290 °C for 20 min, then another ramp 10 °C/min to 310°C, final temperature hold 10 minutes.

The following PAHs were quantified: phenanthrene (PHE), anthracene (AN), fluoranthene (FA), pyrene (PY), benzo(a)anthracene (BaA), chrysene (CHR), benzo-fluoranthenes (BF), benzo(e)pyrene (BeP), benzo(a)pyrene (BaP), indeno(1,2,3-cd)pyrene (IP), dibenzo anthracenes (DBA), benzo(ghi)perylene (BPE). BF is the sum of benzo(b)fluoranthene, benzo(j)fluoranthene and benzo(k)fluoranthene. DBA is the sum of dibenzo(a,h)anthracene, dibenzo(a,c)anthracene and dibenzo(a,j)anthracene. All these isomers were not separated completely by the gas chromatographic method.

Quality Controls. During all the studied period, the following control charts were regularly used: PAH absolute and relative retention times, internal standard areas and heights, column efficiency (theoretical plates number).

The recovery of the analytical procedures was evaluated by the additions method; mean recoveries for three rings PAHs (PHE and AN) were 96%, recoveries for all heavier PAHs were greater than 98%.

Accuracy and precision of the analytical method were evaluated by periodical analyses of reference standard material samples (Urban Dust/Organics SRM 1649), whose PAH concentrations were certified by National Institute of Standards and Technology (NIST). Seven different analyses of this reference material, carried out with an annual frequency during the studied period, gave results in agreement with certified values: our PAH determinations ranged within 86-115% of certified values, with the exclusion of BF (76%) and DBA (180%).

Blank analyses, repeated every thirty samples, excluded the occurrence of contaminants with relative retention times corresponding 184 to PAHs and internal standards.

APPENDIX B

Explorative analysis of the preliminary dataset (January 1995 to October 1996)

The explorative analysis investigated, separately, both the quantitative and the qualitative PAH characterization of samples. The aim was to study the variability in PAH concentrations and the differences between PAH profiles.

Methods. Principal Component Analysis (PCA) summarises the information present in a multivariable dataset by exploiting the correlations between the variables.

The calculated principal components (PCs) are independent directions of major variations of the data, explaining a percentage of its variance equal to each eigenvalue divided by the total number of variables (Wold et al. 1987). Kaiser's rule establishes that the significant components are those with eigenvalue greater than 1, meaning that they bring more information than the single original variables.

Normality of all the variables have been previously screened, those not attaining symmetry have been logarithmically transformed before applying PCA and correlation analyses.

The projection of other variables not constituting the material of PCA on the score plots of principal components can elucidate the driving forces that cause the variation (Lê et al. 2008).

SCAN 1.1 (Minitab Inc.) and FactoMineR (<http://factominer.free.fr>) software have been used.

Materials. Dataset p1 counts 149 samples (46 from site 3, 43 from site 4, 60 from site 5) described with the variables: BaP_{eq} concentration, mean daily temperature, number of leeward hours during the sampling. Objects were also characterized by other supplementary variables (mean daily pressure, relative humidity, number of hours of calm winds (speed less than 0.6 m/s), downwind distance parallel and orthogonal to SSE direction) to explore the correlations of these variables with the main components.

Dataset p2 is a subset of dataset p1, consisting only of the samples collected from site 3, described with eleven ratios between PAH concentrations and BeP concentration. Here again descriptive variables (the quantitative BaP_{eq} concentration, number of downwind hours, temperature, pressure, number of hours of calm winds) were not introduced in PCA but projected afterward to find an explication to the observed variations.

Results. The plane of the first two PCs contains 92% of the variance of PAH concentrations of dataset p1. Figure B.1 shows, with different colours, the samples from the various sites.

Objects of site 3 are placed both in the lowest and the highest concentration part of the graph, (left and right of abscissa), samples from sites 4 and 5 lay in between.

Figure SB1: PCA of PAH concentrations, biplot of scores and loading vectors for objects in the preliminary dataset.

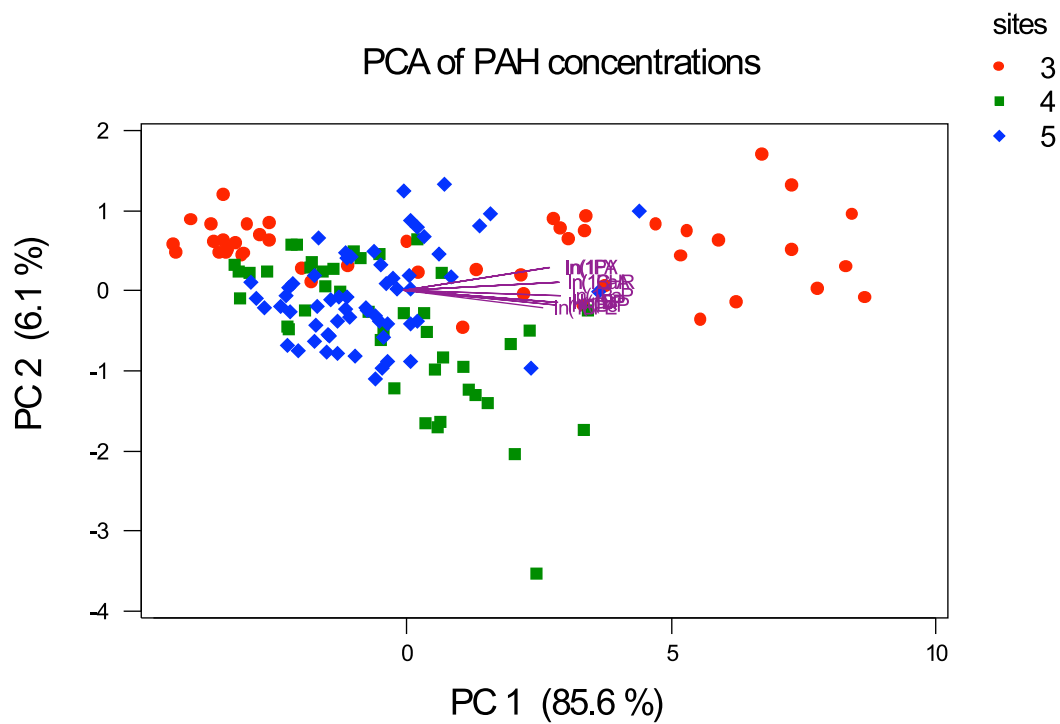
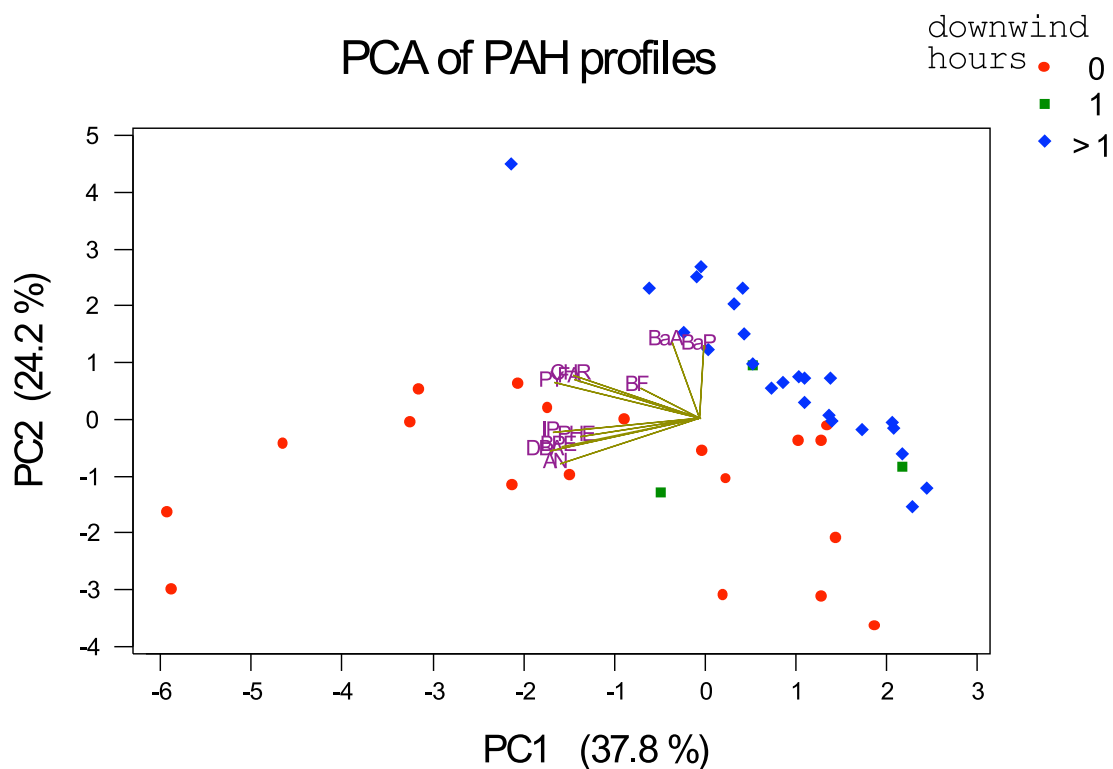


Table SB1 The supplementary variables with significant correlations ($p < 0.01$) with PCs

Supplemental variable	PC1
R ²	
pressure	
temperature	-
ln (1+downwind_hours)	0.58
ln (1+BaP _{eq})	0.94
n. hours of calm winds	-
distance	0.10

The first PC, loading almost equally on concentration of all PAHs, correlates with the number of downwind hours ($R=0.76$, $p < 0.0001$), with BaP_{eq} concentration ($R=0.97$, $p < 0.0001$), with distance of the sampling point from coke ovens ($R= - 0.31$, $p = 0.0001$). Figure 1s.1 suggest that sampler distance from and number of hours leeward to the coking plant are important factors determining the measured BaP_{eq} concentration.

Figure SB2: Biplot of scores and loading vectors resulting from the PCA of PAH profiles in site 3 preliminary dataset, objects are coloured following the number of downwind hours during the sampling.



The PCA on PAH profiles of samples from site 3 has three significant components, explaining 37.8%, 24.2% and 16.6% of variance respectively.

Figure SB2 shows the separation of samples collected in days with or without hours of winds blowing from the coke ovens, which exhibit a different relative abundance of PAHs relative to BeP concentration.

Table SB2

Significant correlations, Spearman's R^2 ($p < 0.01$, 44 freedom degrees), between supplementary variables and the first principal components of PAH profiles dataset (site 3, 1995-1996)

Variable	Principal Component 1	Principal Component 2	Principal Component 3
Pressure	-	-	-
temperature	0.22	-	-
ln (1+downwind_hours)	0.18	0.38	-
Ln (1+ BaP _{eq})	0.20	0.44	-
no. hours of calm winds	0.16	-	-

PAH profiles of samples collected during downwind days are richer in BaA and BaP, variables loading on the second axis.

Bibliography

Lê S., Josse J., Husson F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**(1), 1-18.

Wold S., Esbensen K, Geladi P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2**(1-3), 37-52.

APPENDIX C

Statistical analyses

Datasets

PAH concentrations measured at site 3, 4 and 5 from January 1995 until October 1996, the mean daily temperature and prevailing wind directions during these samplings, constituted the *preliminary* dataset, used to identify main sources of variability.

The 24 months between August 1997 and July 1999, after the substitution of all oven door frames, were chosen to represent the well-maintained activity of the four coke oven batteries; data collected in all the five sampling sites during this period constituted the *full activity* dataset.

The 24 months following the coke oven closure (April 2002–March 2004) represented the total absence of coke coal production, with two blast furnaces, a steel plant and rolling mills still operative inside Cornigliano industrial area. The data collected at sites 3, 4 and 5 during this period constituted the *closure* dataset.

Six subsets of samples collected at sites 3, 4, 5 (total number: 142) with common characteristics (BaP_{eq} concentrations grouped by deciles, similar meteorological conditions during samplings) were drawn out from the *full activity* dataset to test the differences or similarities of their PAH profiles (ratios of each PAH with BeP concentration).

Annual moving average

To highlight the long-term trends of PAH concentrations collected at sites 3, 4 and 5 and to smooth out short-term and seasonal fluctuations of these samples, an Annual Moving Average (AMA) was applied from July 1997 to March 2004. The AMA of BaP_{eq} concentrations was calculated as the arithmetic mean of all daily BaP_{eq} values received during the previous 12 months. In this work, AMAs are dated according to the last day of

the involved period. For each AMA, its 95% confidence interval was calculated.

Data distributions

The fitting of Gaussian or log-normal distributions of produced data was tested by means of an online Kolmogorov–Smirnov test (Kirkman, 1996). The quartile coefficient of dispersion, given by the formula $(P75-P25)/(P75+P25)$ (Bonett, 2006; Lewis and Holmes, 1953) involves a ratio between inter-quartile range and an approximate central locator, and this was used as an index of sample variability.

Table SC1, shows statistical data of annual BaP_{eq} concentrations measured in the five sampling sites around the Genoa coke oven, from 1998 to 2004: annual moving averages and standard deviation; geometric mean and geometric standard deviation; 10th, 50th, 75th, 90th percentiles; minimum and maximum concentrations; quartile coefficient of dispersion (Q_{cd}).

Variations of daily BaP_{eq} concentrations

The distribution parameters of the yearly datasets of BaP_{eq} concentrations monitored in all five sampling sites, during the two years of full activity (1998-1999) and at sites 3, 4 and 5, during the two years after the coke oven closure (2003-2004), are shown in Table SC1. All these distributions were skewed and Kolmogorov–Smirnov test refused normality in all monitored sampling sites and in both annual datasets.

The highest variation of daily BaP_{eq} concentrations was found, during 1998 and 1999, at site 3, during the two years of full activity, in accordance to the highest quartile coefficients of dispersion (0.92- 0.92) and geometric standard deviations (5.91-5.78 ng/m³) found at this site. Figure SC1 shows the histograms of BaP_{eq} concentrations collected at site 3 in the full activity data set, original and log-transformed data, disclosing the bimodal characteristic of the distribution.

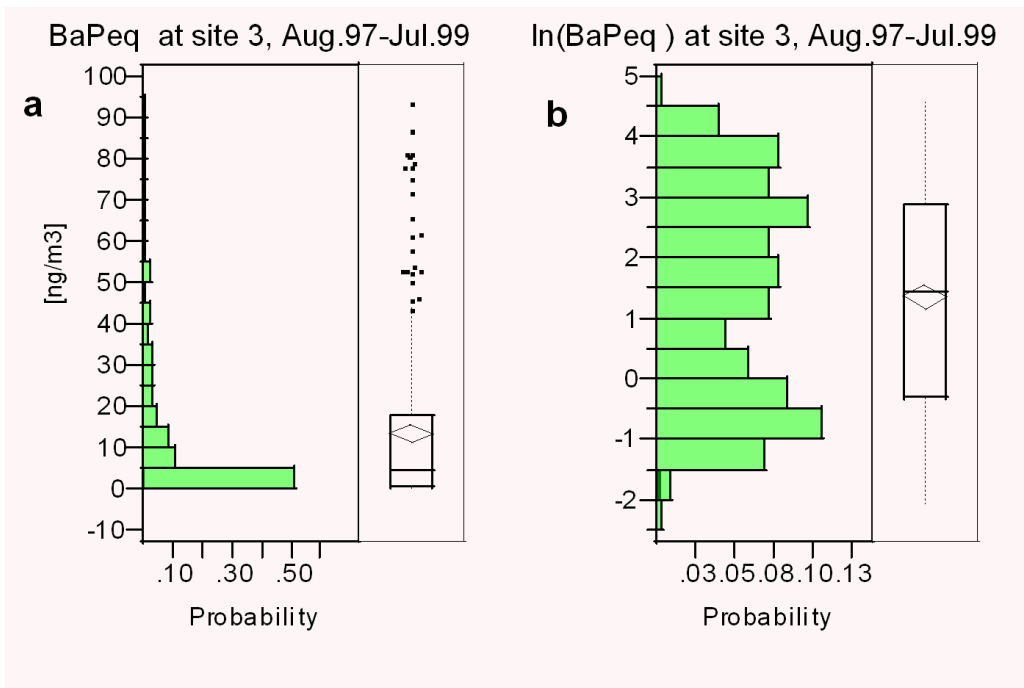
Qcd	0.4 9	0.6 5	0.5 7	0.6 6	0.9 2	0.9 2	0.4 2	0.5 7	0.7 6	0.6 1	0.4 3	0.7 3	0.2 5	0.3 4	0.4 9	0.5 3
Max	153	112	327	184	94	87	4.1	0.9	48	35	6.2	2.2	18	6	3.1	1.7
ng/m³																
min	3.8	2.1	6.1	5.9	0.2	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.6	0.6	0.1	0.1
ng/m³							6	3				9	4		2	5
P90	113	58.	217	146	42.	38.	0.5	0.5	18.	11.	0.8	1.1	3.2	4.8	1.5	1.4
ng/m³	.3	5	.4	.6	3	1	2	2	7	6	3	2			1	1
P75	80.	35.	164	98.	18	17.	0.3	0.2	7.2	6.0	0.4	0.4	2.9	3.2	0.9	0.8
ng/m³	3	7	.6	1		1	6	6			9	2			5	0
P50	61.	59.	94.	34.	5.1	3.3	0.2	0.1	2.3	3.0	0.3	0.1	2.3	2.2	0.4	0.4
ng/m³	6	2	9	6			2	6			1	4			3	8
P10	19.	3.5	16.	11.	0.4	0.4	0.1	0.0	0.5	0.9	0.1	0.0	1.2	1.0	0.2	0.1
ng/m³	3		3	3				5			4	4			5	6
95%CI	57	23	101	54	11.	9.2	0.2	0.1	4.9	4.1	0.3	0.2	1.7	2.0	0.4	0.3
ng/m³	–	–	–	–	0–	–	3–	7–	–	–	5–	5–	–	–	5–	4–
	65	28	118	66	17.	15.	0.3	0.2	7.7	5.9	0.6	0.4	3.9	3.3	1.0	0.9
					4	4	8	5			5	7			3	2
GSD	2.1	2.8	2.8	2.6	5.9	5.7	1.8	2.3	3.8	2.8	2.1	3.4	1.8	1.8	2.2	2.3
	7	2	2	1	1	8	6	8	5	9	5	3		2	1	0
GM	47.	16.	73.	39.	4.2	3.7	0.2	0.1	2.7	3.0	0.3	0.1	2.2	2.2	0.5	0.4
ng/m³	7	7	8	6			3	4			3	7			4	7
SD	38.	22.	81.	52.	20.	19.	0.4	0.2	8.9	5.9	0.8	0.5	3.0	1.5	0.7	0.5
ng/m³	5	6	6	8	4	1	1				2	2			0	1
AMA	60.	25.	109	60	14.	12.	0.3	0.2	6.3	5.0	0.5	0.3	2.8	2.6	0.7	0.6
ng/m³	7	8	.4		2	3	1	1				7			4	3
n/N	96	99.	96	99.	42	37	31	19	43	45	24	16	8	6	7	4
%		7		7												
sampled days, n	350	364	350	364	152	133	112	70	155	164	88	59	28	22	25	14
Period end date	08/ 02/ 98	08/ 01/ 99	08/ 02/ 98	08/ 01/ 99	08/ 01/ 98	08/ 01/ 99	04/ 01/ 03	04/ 01/ 04	08/ 01/ 98	08/ 01/ 99	04/ 01/ 03	04/ 01/ 04	08/ 01/ 98	08/ 01/ 99	04/ 15/ 03	04/ 15/ 04
Site	1	1	2	2	3	3	3	3	4	4	4	4	5	5	5	5

Table SC1

SD= standard deviation, MA= moving average, GM = geometric mean, GSD= geometric SD, 95%CI= margin error of the mean, P10= 10th percentile, P50= median, P75= 75th percentile, Qcd=quartile coefficient of dispersion = (P75 – P25) / (P75 + P25)

Figure SC1 a) Histogram and box-plots of BaP_{eq} distribution at site 3, 285 daily samples collected during the coke oven full activity period. b) The distribution of the same data after logarithmic transformation.

Rhomb in the box points to mean value, with height equal to standard deviation of the mean, central segment in the box points at the median.



After the coke plant's closure, during the two following years (2003-2004), daily BaP_{eq} concentrations were less variable. Site 3 displays now the lowest coefficient of dispersion, coherently with its position farthest from the residuary PAH source (vehicular traffic), followed by the dispersions, comparable, of sites 4 and 5, leant on the same street.

Regression analyses

Regression analyses, significant at $p \leq 0.01$, were performed with original or log-transformed variables, according to their fitting of the normal distribution, using JMP 5.0.1 (SAS Institute Inc.) and Stata 10 (StataCorp LP 2007) statistical software.

Regression analyses were applied to daily BaP_{eq} concentrations and to downwind coke oven hours during samplings carried out at sites 3 and 4, for two years (1997- 1999) belonging to the “full activity” period.

The log-normal fixed effects model was applied to BaP and BaP_{eq} measures to estimate the effect of downwind hour number, on data variability.

Regression analyses were also applied to average BaP_{eq} concentrations measured at sites 2, 3 and 4 during two periods, twelve months long each (August 1997- July 1998 and August 1998-July 1999) regressed with distances of these sites from the coke oven centre.

Statistical comparison of PAH profiles and diagnostic PAH ratios

To identify PAH profile of coke oven fugitive emissions, BaP_{eq} concentrations measured at site 3, were grouped by deciles. Two main deciles groups were identified, defined as low and high deciles. In low deciles group were included samples whose BaP_{eq} concentrations were within 15^o and 25^o percentiles ($p_{15} < \text{BaP}_{eq} < P_{25}$). Samples with BaP_{eq} concentrations higher than 90^o percentiles ($\text{BaP}_{eq} > P_{90}$) were included in high deciles group.

Samples belonging to low and high deciles groups were prevalently collected during upwind and downwind conditions.

Samples collected at site 5, during the same sampling periods of low and high deciles samples collected at site 3, were selected and accordingly grouped.

The student's *t* test with Bonferroni correction (Armitage et al., 2002) was applied to

compare all grouped PAH profiles.

Samples collected at site 3, 4 and 5, during the *full activity* were used to calculate three PAH diagnostic ratios [$FA/(FA+PY)$; $BaA/(BaA+CHR)$; $IP/(IP+BPE)$] in six sub groups: low and high BaP_{eq} deciles at site 3 and contemporary samples collected at site 4 and 5. Mean diagnostic ratios of each sub group were compared with reference diagnostic PAH ratios (Yunker et al., 2002) found in emissions produced by the combustion of gasoline, diesel oil, hard coal.

All results were considered significantly different if the experiment-wise *p* value was less than 0.05.

Bibliography

Armitage, P., G. Berry, and J.N.S. Matthews. 2002. Statistical methods in medical research, 4th edition, Blackwell Publishing Ltd.

Bonett, D.G. 2006. Confidence interval for a coefficient of quartile variation. Computational Statistics & Data Analysis. 50(11): 2953-2957.

Kirkman, T.W. 1996. Statistics to use. http://www.physics.csbsju.edu/stats/KS-test.n.plot_form.html (accessed February 8, 2010).

Holmes R.L.A.. 1953. Statistics for professional students. London Pitman. 2nd edition

Yunker M.B., R.W. Macdonald, R. Vingarzan, R.H. Mitchell, D. Goyette, and S. Sylvestre. 2002. PAHs in the Fraser River Basin: a Critical Appraisal of PAH Ratios as Indicators of PAH Source and Composition. Organic Geochemistry. 33(4):489-515.

APPENDIX D

Figures and legends

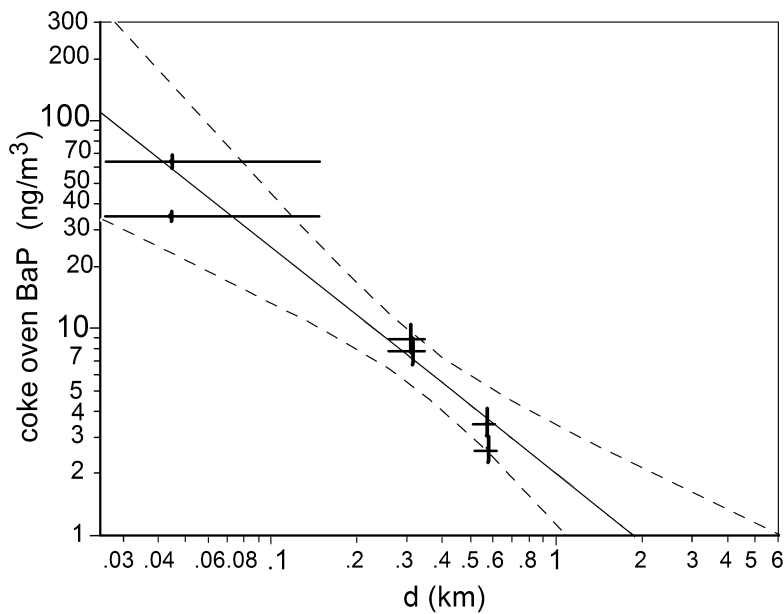


Figure SD1. Correlation between annual average *coke* BaP concentrations and distance from the coke oven.

Two annual (August 1997–July 1998 and August 1998–July 1999) BaP concentrations (ng/m³), estimated to be through coke oven emissions only, correlate, according to eq. 3), with the distance d (km) of sampling sites 2, 3 and 4, from the centre of the coke plant. Error bars along the abscissa represent the range of each sampling site distances from the nearest to the furthest coke battery; those along the ordinate show the 95% confidence interval of the annual BaP concentrations.

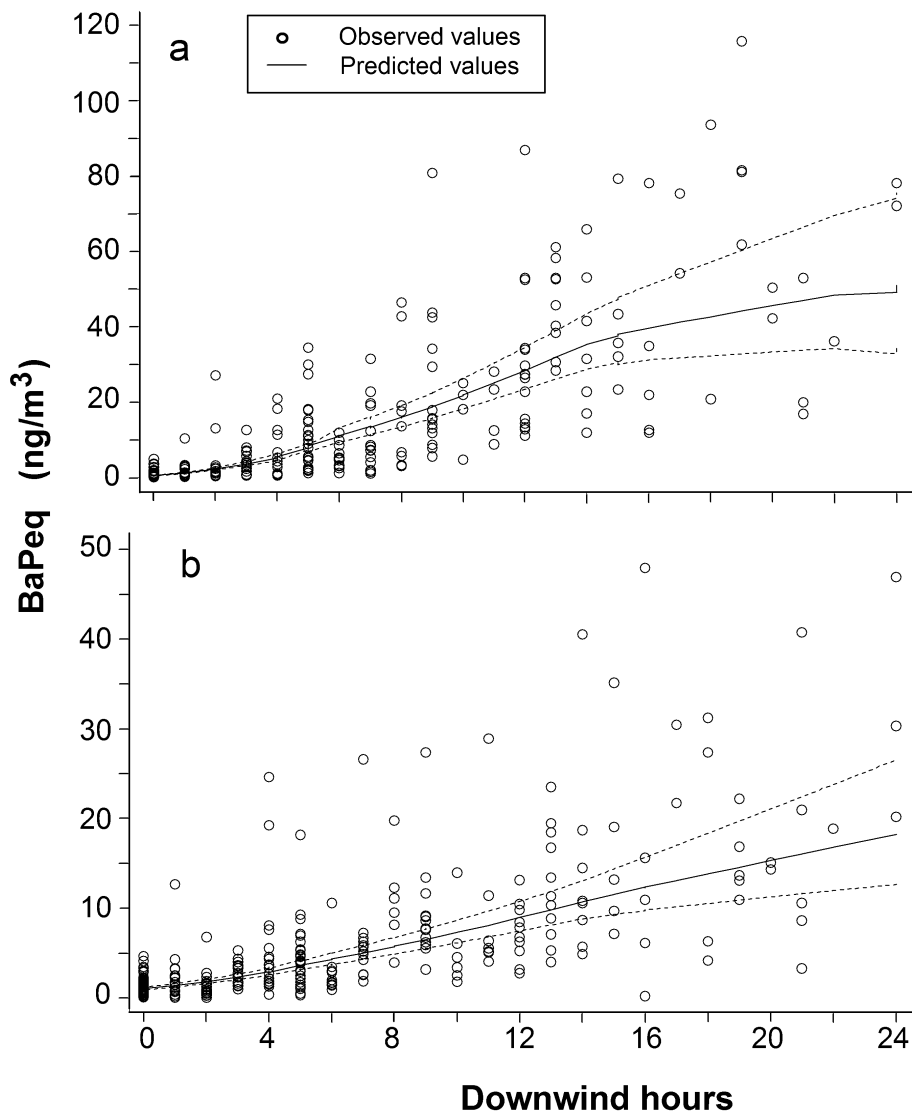


Figure SD2 Relationship between BaPeq concentrations and number of downwind hours during samplings.

BaPeq concentrations at sites 3 and 4 (parts **a** and **b** respectively) correlate with hours downwind the coke oven, during sampling, according to eq. 1) and eq. 2). Continuous lines represent the best fit, dotted lines the 95% confidence interval