Supplementary Data

Source Apportionment of PAH in Hamilton Harbour Suspended Sediments: Comparison of Two Factor Analysis Methods

U. Sofowote, B. McCarry*, C. Marvin

List of Tables

List of Figures

Table S1: SumPAH values, sample names, sampling dates and seasons for Hamilton Harbour suspended sediments and creek samples. Identification (ID) numbers are used in Figure 2.

ID	Sampling Station Name	Date	Season	SumPAH (µg/g)
1	Spencer Creek	Nov-06	Fall	2.4
2	Indian Creek	Nov-06	Fall	1.8
3	Borer's Creek	Nov-06	Fall	0.7
4	Desjardin Canal	Nov-06	Fall	3.0
5	Grindstone Creek	Nov-06	Fall	1.6
6	Chedoke Creek	Nov-06	Fall	20.6
7	Red Hill Creek	Nov-06	Fall	48.5
8	9031	Aug-03	Summer	13.1
9	9031	Dec-04	Fall	13.9
10	9032	May-03	Spring	14.9
11	9032	Nov-03	Fall	25.9
12	9032	Jun-04	Summer	13.9
13	9081 (Randle Reef A)	May-05	Spring	15.1
14	9081 (Randle Reef A)	Jun-05	Summer	26.9
15	9081 (Randle Reef A)	Sep-05	Fall	26.8
16	9081 (Randle Reef A)	Aug-06	Summer	17.6
17	9083 (Randle Reef B)	May-05	Spring	36.9
18	9083 (Randle Reef B)	Jun-05	Summer	44.2
19	9083 (Randle Reef B)	Aug-06	Summer	31.1
20	9030	Nov-02	Fall	27.0
21	9030	Jul-04	Summer	6.7

22	9033	Nov-03	Fall	30.4
23	9033	Aug-03	Summer	32.5
24	9033	Dec-04	Fall	30.8
25	914 (Windermere Bridge)	May-03	Spring	71.2
26	914 (Windermere Bridge)	Feb-04	Winter	73.1

Figure S1: Principal component scores regression plot for Hamilton Harbour suspended sediments and creek sediments.

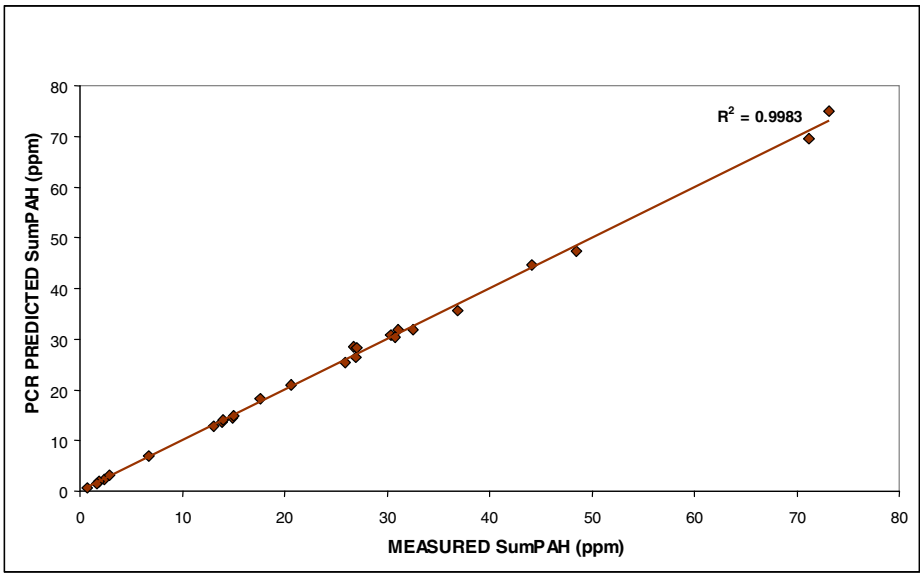


Figure S2: Positive Matrix Factorization Factor Contributions Regression Plot

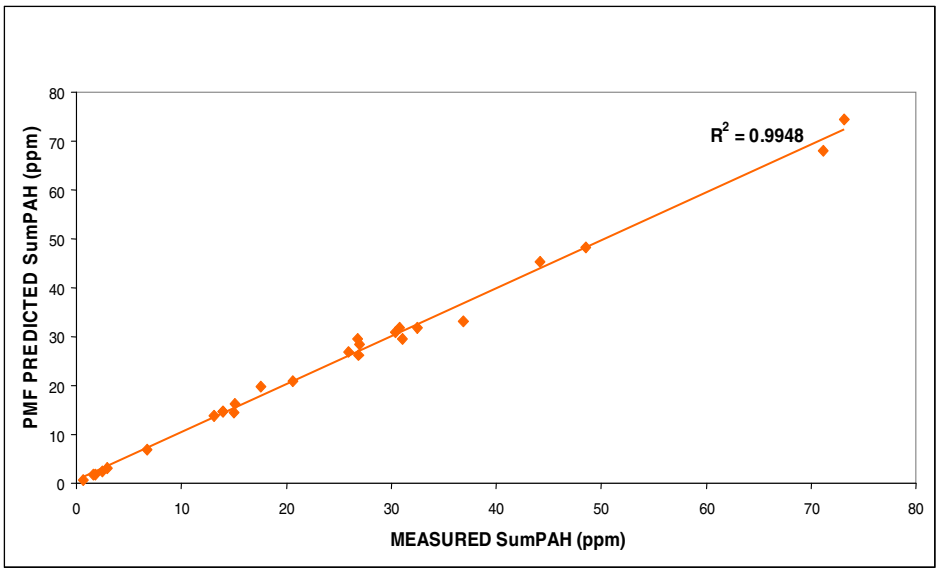


Figure S3: Plot of PMF Predicted Average vs. Measured PAH Average for each PAH

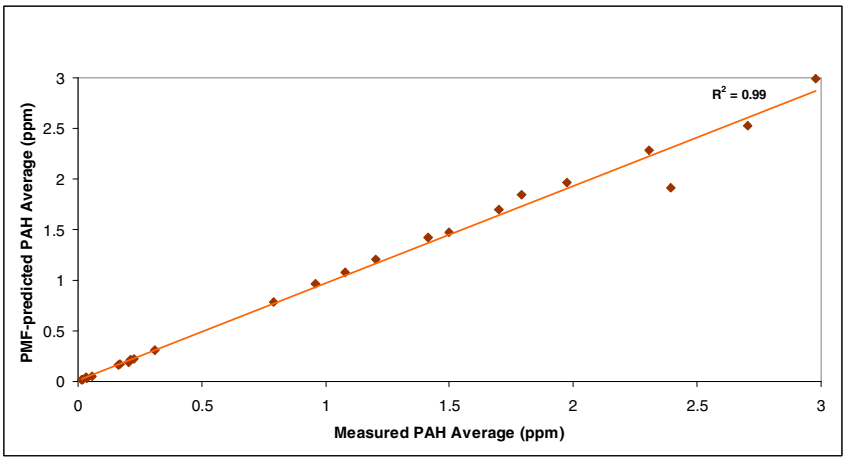


Table S2: Mean Recoveries and Standard Deviations of Deuterated PAH Recovery

	Mean Recovery (%)	Standard Deviation (%)	Ions Integrated (m/z)
Phenanthrene-d_{10}	63.6	21.5	188
Chrysene-d_{12}	97.7	2.2	240
Dibenz[a,h]anthracene-d_{14}	101	1	289, 290, 291, 292

Figure S4: PMF explained variation plots for 25 compounds in the four PMF factors identified in the Hamilton Harbour data set (order of compounds same as in Table 3).

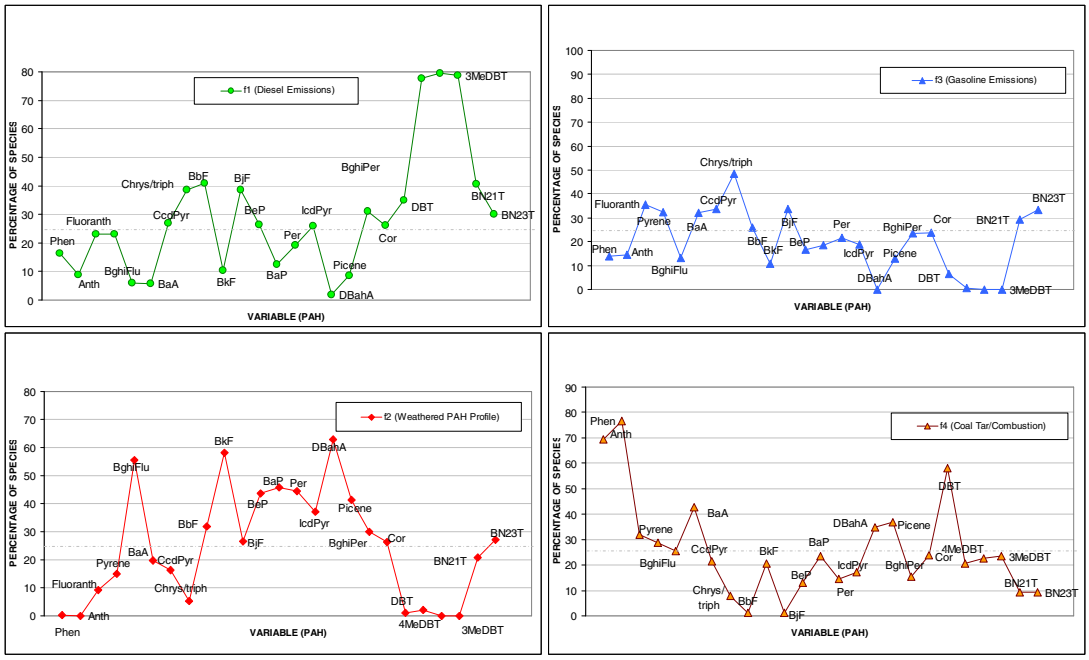


Figure S5: Comparison of the PAH profiles of (a) factor 3 (Gasoline Emissions) with SRM 1649a (NIST Urban Dust Reference Material) profile; (b) factor 4 (Coal Tar/Combustion with SRM 1597a (NIST Coal Tar Reference Material).

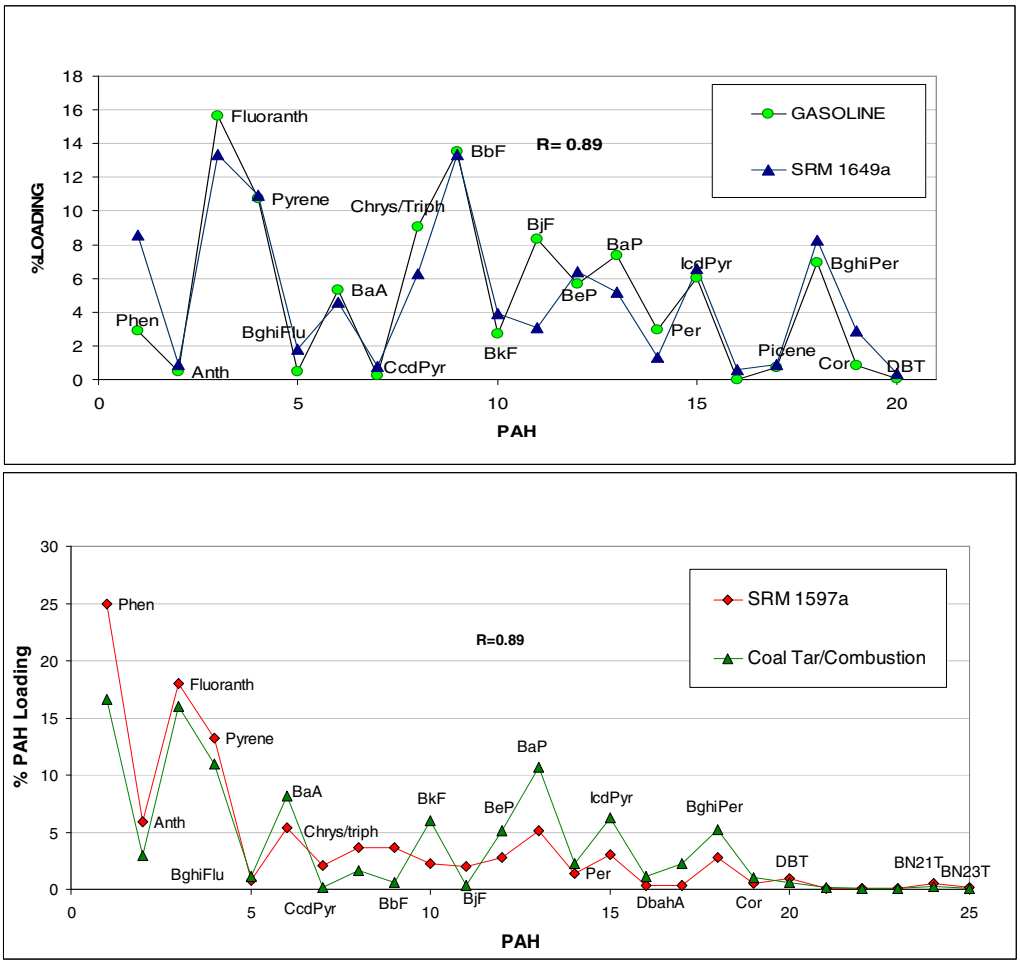


Figure S6: PAH loadings profile of unidentified factor f_2 tentatively attributed to a weathered PAH profile.

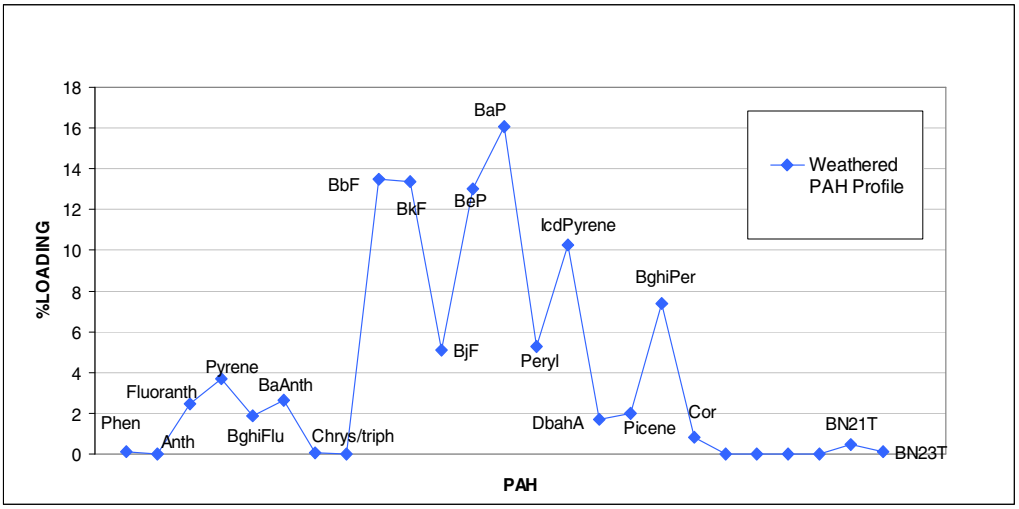
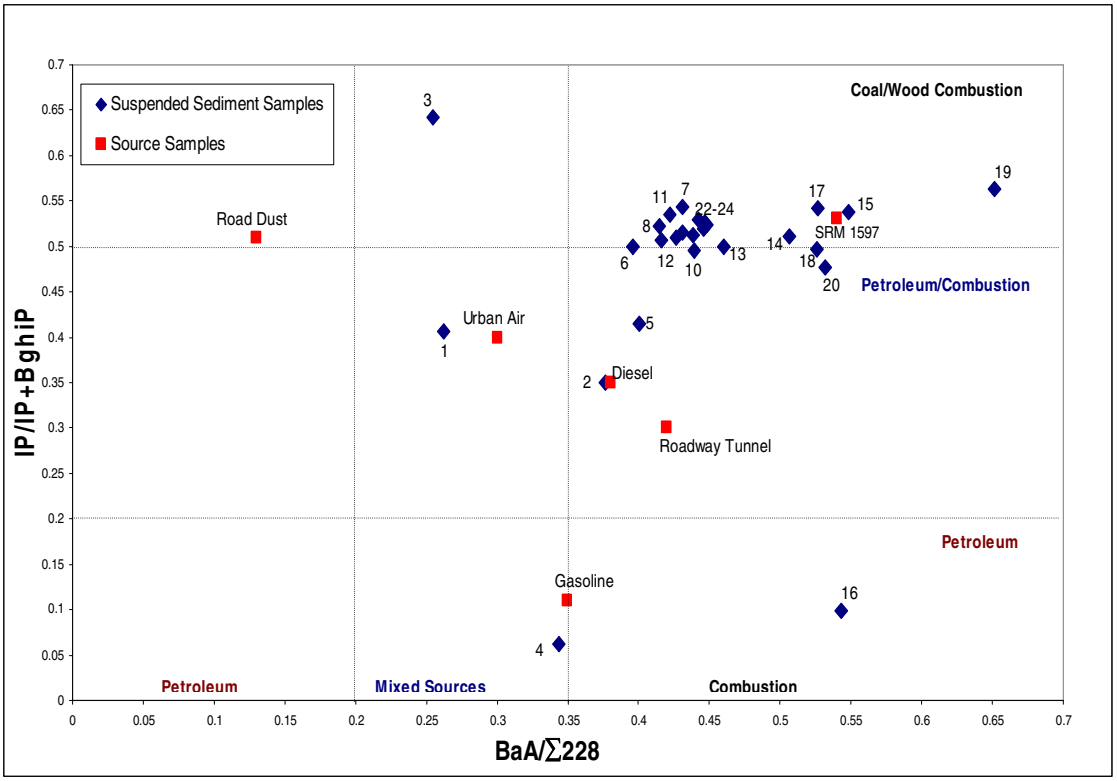


Figure S7: Plot of two diagnostic PAH ratios (BaA/∑228 against IP/IP+BghiP) for the Hamilton Harbour and creek suspended sediments. Six source sample ratios and zones related to source types as described are included by Yunker et al. (5). Numbers refer to sediment samples in Table S1



Explanation for Figure S7

A number of diagnostic PAH ratios (5) were calculated using the suspended sediment sample data. The most useful ratios to discriminate between sources were the BaA/∑228 and IP/IP+BghiP ratios. The plot of these data are shown in Figure S6, together with the ratio ranges suggested by Yunker et al. (5) that are indicative of the certain source types; the values for six sources cited by Yunker et al are also included on the figure. The harbour sediment samples are clustered in the upper right hand section of the plot. Five samples are very proximate to the NIST SRM 1597 (Coal Tar) reference material,

indicative of a high correlation to a coal tar source; four of these samples were collected at Randle Reef stations (#14, #15, #17, and #18). The remaining harbour samples (except #16 and #19) are found to the left of the SRM 1597 value and approximately halfway between the coal tar sample and the diesel and urban air samples. With the exception of the Chedoke Creek and Red Hill Creek samples (#6 and #7), the creek samples were located well away from the harbour sediment samples, in some cases proximate to the urban air source sample (#1), the diesel source sample (#4). This diagnostic ratio approach showed that sources of PAH to Hamilton Harbour suspended sediments are split between coal tar or coal combustion sources and a variety of urban-based combustion sources. However, this method was not amenable to determining the relative contributions to each sample from the source types. Thus, while diagnostic PAH ratios approach can be useful in elucidating qualitative relationships, a more quantitative approach was needed for this work. Thus, we turned our attention to receptor modeling techniques employing factor analyses.

Table S3: Comparison of percentage contributions of factors identified by PCA and PMF

PMF Factor	PMF Contribution	PCA Factor	PCA Contribution	Source Identification
f_2	26%	t_3	19%	Coal Tar/Coal Combustion
f_1	28%			Gasoline Emissions
f_4	24%			Diesel Emissions
		t_1	61%	Vehicular Emissions
f_3	22%			Weathered PAH Profile
		t_2	13%	Unburned Fossil Fuels
		t_4	7%	Unknown PC Source

Data Pre-treatment for PMF Analysis: One hundred and thirty one data points of the total 650 points (~20%) in the concentration matrix (see Table S4) were below their method detection limits, thus it was of utmost interest to observe how PMF processed the uncensored data set. The data points below detection limits were not replaced by the mean concentrations for any of the PAH variables or by their method detection limits. Undetected PAH were reported in the concentration matrix as zero. This is critical as no missing values are allowed in the PMF analysis the EPA PMF 1.1 program executes. The concentration matrix contains no negative values, because the PAH concentration values from the extracted method blanks which provided the background concentration were zero. No approximation of concentration values was performed prior to PMF analysis. Nine PAH variables (see Table S5) were weighted as weak due to any of the following reasons; (i) low S/N values; (ii) low R^2, and; (iii) residuals being greater than ±3 standard deviation. Cyclopenta[cd]pyrene had the lowest R^2 of 0.4 and was down-weighted for this reason. This confirmed that the use of this variable as having equal weight as strong variables as executed by the PCA was wrong, invariably leading to a highly suspect component being identified by the PCA. BghiFlu, BaP and IcdPyr were downweighted even though their S/N ratios were good because their residuals had distances greater than ±3 standard deviations. A re-run with these three variables weakened showed only IcdPyr having a noisy residual in sample #16 (see Table S5). The robust Q for 7 of 15 runs (all convergent) was 215.62. Random run #3 (robust Q = 215.62) was selected for further processing because it had the lowest true Q (216.74) and was converged in the fewest number of steps.

Table S4: PAH concentration (µg/g dry sediment extracted) and uncertainty data matrices for Hamilton Harbour suspended sediments

Concentration Matrix											
	Phen	Anth	Fluoranth	Pyrene	BghiFlu	BaA	CcdPyr	Chrys/Triph	BbF	BkF	BjF
SPENCER	0.125	0.019	0.343	0.297	0.030	0.066	0.001	0.185	0.278	0.139	0.120
INDIAN	0.103	0.013	0.323	0.155	0.020	0.072	0.001	0.118	0.214	0.115	0.081
BORER'S	0.037	0.001	0.080	0.057	0.006	0.014	0.000	0.042	0.113	0.048	0.051
DESJARDIN	0.137	0.011	0.357	0.228	0.027	0.091	0.000	0.174	0.419	0.159	0.191
GRINDSTONE	0.083	0.005	0.193	0.231	0.014	0.058	0.000	0.087	0.177	0.113	0.105
CHEDOKE	0.761	0.134	2.461	1.977	0.206	0.741	0.007	1.132	2.697	1.380	1.240
REDHILL	0.752	0.117	3.623	3.649	0.434	1.573	0.170	2.076	7.292	3.327	3.330
STN 9031 AUG 12 2003	0.716	0.064	1.383	0.926	0.122	0.510	0.013	0.721	1.636	0.951	0.824
STN 9031 DEC 02 2004	0.682	0.125	1.439	1.156	0.138	0.541	0.040	0.714	1.893	0.856	0.884
STN 9032 MAY 20 2003	1.060	0.181	1.907	1.348	0.122	0.479	0.002	0.611	1.743	0.970	0.819
STN 52 NOV 27 2003	1.656	0.237	2.462	1.499	0.245	0.892	0.038	1.106	3.762	1.560	1.641
STN 9032 JUN 16 2004	0.822	0.151	1.263	1.120	0.122	0.508	0.011	0.713	1.785	0.832	0.809
STN 9081 MAY 17 2005	1.126	0.166	1.656	0.910	0.126	0.735	0.024	0.863	1.721	0.854	0.864
STN 9081 JUN 21 2005	1.999	0.198	3.487	3.785	0.195	1.264	0.030	1.232	2.583	1.333	1.431
STN 9081 SEPT 07 05	1.917	0.376	2.991	2.162	0.188	1.560	0.028	1.285	2.712	1.485	1.486
STN 9081 AUG 12 2006	1.412	0.201	2.141	1.454	0.186	0.971	0.021	0.818	2.314	0.979	1.091
STN 9083 MAY 2005	2.925	0.657	6.085	6.863	0.268	1.715	0.046	1.544	2.754	1.352	1.400
STN 9083 JUN 21 2005	2.678	0.738	5.582	6.834	0.306	2.405	0.130	2.166	4.261	1.939	2.270
STN 9083 AUG 12 2006	2.695	0.475	4.368	3.685	0.376	1.462	0.045	0.781	2.100	2.148	0.777
STN 9030 NOV 07 2002	1.448	0.225	2.649	1.762	0.255	1.035	0.034	0.910	3.667	2.060	1.763
STN 9030 JUL 29 2004	0.406	0.051	0.742	0.663	0.062	0.266	0.006	0.358	0.785	0.350	0.362
STN 53 NOV 27 2003	1.750	0.321	2.959	2.707	0.297	1.012	0.020	1.277	4.256	1.735	1.833
STN 9033 AUG 12 2003	1.879	0.231	3.221	3.103	0.265	1.000	0.014	1.230	4.452	1.909	1.895
STN 9033 DEC 02 2004	1.835	0.254	2.680	1.969	0.247	1.012	0.049	1.252	4.347	2.014	1.986
WINDERMERE MAY 13 2003	1.178	0.180	8.694	8.079	0.636	2.396	0.016	3.071	9.549	4.377	4.440
WINDERMERE FEB 19 2004	1.113	0.170	7.181	5.640	0.962	2.598	0.081	3.561	9.903	6.006	5.074
MEAN CONC	1.204	0.204	2.703	2.395	0.225	0.961	0.032	1.078	2.978	1.500	1.414
Equation Based Uncertainty Matrix											
	Phen	Anth	Fluoranth	Pyrene	BghiFlu	BaA	CcdPyr	Chrys/Triph	BbF	BkF	BjF
MDL (e_j)	0.0196	0.0196	0.0246	0.0246	0.0246	0.0246	0.0246	0.0246	0.147	0.147	0.147
d_j (%)	21.5	21.5	21.5	21.5	2.2	2.2	2.2	2.2	2.2	2.2	2.2

Concentration Matrix										
	BeP	BaP	Per	IcdPyr	DBahA	Picene	BghiPer	Cor	DBT	4MeDBT
SPENCER	0.207	0.163	0.055	0.140	0.007	0.019	0.205	0.000	0.003	0.003
INDIAN	0.111	0.115	0.155	0.065	0.016	0.016	0.121	0.000	0.004	0.002
BORER'S	0.085	0.063	0.026	0.035	0.004	0.005	0.020	0.000	0.001	0.001
DESJARDIN	0.266	0.245	0.194	0.017	0.016	0.038	0.259	0.039	0.006	0.004
GRINDSTONE	0.132	0.150	0.099	0.064	0.003	0.008	0.090	0.000	0.001	0.001
CHEDOKE	1.705	1.873	0.483	1.440	0.125	0.190	1.444	0.255	0.031	0.044
REDHILL	4.657	6.250	1.918	4.065	0.303	0.710	3.416	0.404	0.027	0.019
STN 9031 AUG 12 2003	1.137	1.128	0.354	1.095	0.073	0.166	1.003	0.110	0.026	0.022
STN 9031 DEC 02 2004	1.139	1.151	0.483	1.102	0.076	0.174	1.039	0.112	0.033	0.030
STN 9032 MAY 20 2003	1.189	1.306	0.481	1.081	0.086	0.146	1.104	0.106	0.050	0.035
STN 52 NOV 27 2003	2.254	2.454	0.812	2.197	0.156	0.315	2.034	0.205	0.077	0.064
STN 9032 JUN 16 2004	1.224	1.245	0.427	1.164	0.068	0.195	1.133	0.135	0.045	0.029
STN 9081 MAY 17 2005	1.225	1.599	0.520	1.047	0.093	0.190	1.046	0.116	0.051	0.025
STN 9081 JUN 21 2005	1.718	2.318	0.684	1.814	0.178	0.337	1.740	0.236	0.085	0.037
STN 9081 SEPT 07 05	1.935	2.787	0.749	2.050	0.195	0.426	1.759	0.307	0.083	0.036
STN 9081 AUG 12 2006	1.350	1.897	0.545	0.139	0.146	0.280	1.268	0.137	0.053	0.026
STN 9083 MAY 17 2005	2.036	2.552	0.944	2.418	0.185	0.403	2.048	0.292	0.102	0.038
STN 9083 JUN 21 2005	2.635	3.763	1.171	2.734	0.236	0.611	2.766	0.370	0.130	0.048
STN 9083 AUG 12 2006	2.299	3.188	0.899	2.456	0.251	0.502	1.903	0.267	0.106	0.047
STN 9030 NOV 07 2002	2.488	2.569	0.880	1.868	0.370	0.370	2.045	0.252	0.074	0.045
STN 9030 JUL 29 2004	0.544	0.544	0.224	0.553	0.044	0.079	0.532	0.060	0.021	0.013
STN 53 NOV 27 2003	2.632	2.538	0.923	2.462	0.217	0.389	2.186	0.279	0.114	0.088
STN 9033 AUG 12 2003	2.815	2.941	1.012	2.625	0.257	0.419	2.392	0.313	0.116	0.099
STN 9033 DEC 02 2004	2.825	2.772	1.019	2.647	0.202	0.416	2.384	0.309	0.109	0.081
WINDERMERE MAY 13 2003	6.040	6.981	2.538	5.365	0.487	0.786	5.109	0.548	0.054	0.039
WINDERMERE FEB 19 2004	6.681	7.369	2.916	5.939	0.584	0.865	5.165	0.605	0.049	0.035
MEAN CONC	1.974	2.306	0.789	1.792	0.168	0.310	1.700	0.210	0.056	0.035
Equation Based Uncertainty Matrix										
	BeP	BaP	Per	IcdPyr	DBahA	Picene	BghiPer	Cor	DBT	4MeDBT
MDL (e_j)	0.147	0.147	0.147	0.0966	0.0966	0.0966	0.0966	0.0966	0.0196	0.0196
d_j (%)	2.2	2.2	2.2	1	1	1	1	1	21.5	21.5

Concentration Matrix							
	2MeDBT	3MeDBT	BN21T	BN23T		SumPAH	SumPAH Deviate
SPENCER	0.000	0.000	0.026	0.001		2.433	-1.152
INDIAN	0.001	0.001	0.016	0.002		1.838	-1.184
BORER'S	0.000	0.000	0.004	0.000		0.694	-1.244
DESJARDIN	0.001	0.001	0.025	0.002		2.908	-1.127
GRINDSTONE	0.000	0.000	0.006	0.001		1.623	-1.195
CHEDOKE	0.008	0.023	0.208	0.029		20.594	-0.193
REDHILL	0.008	0.010	0.318	0.076		48.524	1.281
STN 9031 AUG 12 2003	0.007	0.010	0.085	0.012		13.095	-0.589
STN 9031 DEC 02 2004	0.009	0.015	0.089	0.017		13.937	-0.545
STN 9032 MAY 20 2003	0.013	0.016	0.098	0.016		14.968	-0.490
STN 52 NOV 27 2003	0.022	0.031	0.169	0.039		25.924	0.088
STN 9032 JUN 16 2004	0.012	0.015	0.096	0.019		13.942	-0.545
STN 9081 MAY 17 2005	0.010	0.015	0.077	0.015		15.073	-0.485
STN 9081 JUN 21 2005	0.017	0.020	0.154	0.035		26.911	0.140
STN 9081 SEPT 07 05	0.016	0.020	0.166	0.048		26.765	0.132
STN 9081 AUG 12 2006	0.013	0.015	0.113	0.023		17.594	-0.352
STN 9083 MAY 17 2005	0.015	0.018	0.193	0.033		36.886	0.667
STN 9083 JUN 21 2005	0.021	0.028	0.270	0.066		44.159	1.051
STN 9083 AUG 12 2006	0.020	0.026	0.163	0.032		31.068	0.359
STN 9030 NOV 07 2002	0.015	0.019	0.181	0.034		27.018	0.146
STN 9030 JUL 29 2004	0.006	0.007	0.049	0.009		6.736	-0.925
STN 53 NOV 27 2003	0.037	0.048	0.271	0.050		30.404	0.324
STN 9033 AUG 12 2003	0.039	0.052	0.215	0.047		32.541	0.437
STN 9033 DEC 02 2004	0.036	0.048	0.226	0.046		30.765	0.343
WINDERMERE MAY 13 2003	0.014	0.018	0.481	0.116		71.192	2.478
WINDERMERE FEB 19 2004	0.013	0.017	0.479	0.126		73.130	2.580
MEAN CONC	0.014	0.018	0.161	0.034	MEAN	24.259	
					STDEVP	18.943	
Equation Based Uncertainty Matrix							
	2MeDBT	3MeDBT	BN21T	BN23T			
MDL (e_j)	0.0196	0.0196	0.0246	0.0246			
d_j (%)	21.5	21.5	2.2	2.2			

Table S5: PMF diagnostics file for Hamilton Harbour suspended sediments

____________________ANALYSIS START________________________							
Time of run (approx.): 12-Apr-2008 17:15:50							
Concentrations File: C:\Documents and Settings\Yemi\My Documents\YEMIResData\HH SPREADSHEETS\HH PAC SPREADSHEETS\HH_conc.csv							
Uncertainty File: C:\Documents and Settings\Yemi\My Documents\YEMIResData\HH SPREADSHEETS\HH PAC SPREADSHEETS\HH_un_ eqn.csv							
Number of random starting points: 15							
Number of factors: 4							
Seed: Used random seed.							
c3 Modeling Constant(Percent): 0.00							
Species included:							
Strong - Phen,Anth,Fluoranth,Pyrene,						Weak (down-weighted) - BghiFlu,CcdPyr,BaP,IcdPyr,	
			BaA,Chrys/triph,BbF,BkF,				DBahA,4MeDBT,2MeDBT,3MeDBT,
			BjF,BeP,Per,Picene,				BN23T,
			BghiPer,Cor,DBT,BN21T,				
Species not included	Bad (not included) -		No "Bad" variables				
~~							
Q Values for random-start runs							
~~							
Random R	Q(Robust)	Q(True)	Converged	# Steps			
1	215.62	216.75	Yes	551			
2	215.63	216.75	Yes	794			
3	**215.62**	**216.74**	**Yes**	**622**			
4	215.63	216.76	Yes	747			
5	215.62	216.74	Yes	637			
6	215.62	216.75	Yes	365			
7	215.65	216.78	Yes	540			
8	215.66	216.79	Yes	581			
9	215.62	216.75	Yes	500			
10	215.65	216.78	Yes	964			
11	215.63	216.75	Yes	662			
12	215.62	216.75	Yes	639			
13	215.63	216.76	Yes	484			
14	215.63	216.75	Yes	468			
15	215.62	216.75	Yes	580			

_____ START BASE FACTOR ANALYSIS of Random Run # 3_____								
Regression diagnostics of run# 3								
Species	Intercept	Slope	RMSE	r^2	S/N			
Phen	-0.06	1.05	0.2	0.95	2.32			
Anth	0.03	0.79	0.03	0.95	2.15			
Fluoranth	0.39	0.79	0.52	0.92	2.32			
Pyrene	0.55	0.56	0.58	0.84	2.32			
BghiFlu	0.01	0.94	0.05	0.95	5.19			
BaA	0.01	1	0.02	1	16.25			
CcdPyr	0.02	0.47	0.02	**0.4**	**0.62**			
Chrys/Tripl	0	1	0.03	1	17.61			
BbF	0.04	0.99	0.16	1	11.04			
BkF	-0.02	1	0.17	0.99	5.49			
BjF	0.03	0.99	0.1	0.99	5.15			
BeP	-0.02	1.01	0.06	1	7.16			
BaP	0	0.99	0.28	0.98	8.58			
Per	0	1	0.07	0.99	3.06			
IcdPyr	0.14	0.94	0.28	0.97	**9.88**			
DBahA	0.01	0.96	0.03	0.95	**0.76**			
Picene	0	0.99	0.03	0.98	1.53			
BghiPer	0	1	0.07	1	9.84			
Cor	0.01	0.96	0.03	0.98	1.05			
DBT	0	0.95	0.01	0.95	1.19			
4MeDBT	0	0.88	0.01	0.88	**0.78**			
2MeDBT	0	0.87	0	0.86	**0.24**			
3MeDBT	0	0.89	0.01	0.85	**0.34**			
BN21T	0.01	0.97	0.02	0.98	3.53			
BN23T	0	0.94	0	0.98	**0.62**			
**********								

Observations (residuals) beyond 3 Std. Dev.								
Species	Obs. (residuals)							
lcdPyrene	16(-4.3)							
**********								
Species (residuals) beyond 3 Std. Dev.								
Obs.	Species (residuals)							
16	lcdPyrene (-4.3)							
_____ END BASE FACTOR ANALYSIS of Random Run # 3_____								

Figure S8: PCA-MLR source contributions plot for Hamilton Harbour suspended sediments and creek sediments (µg/g).

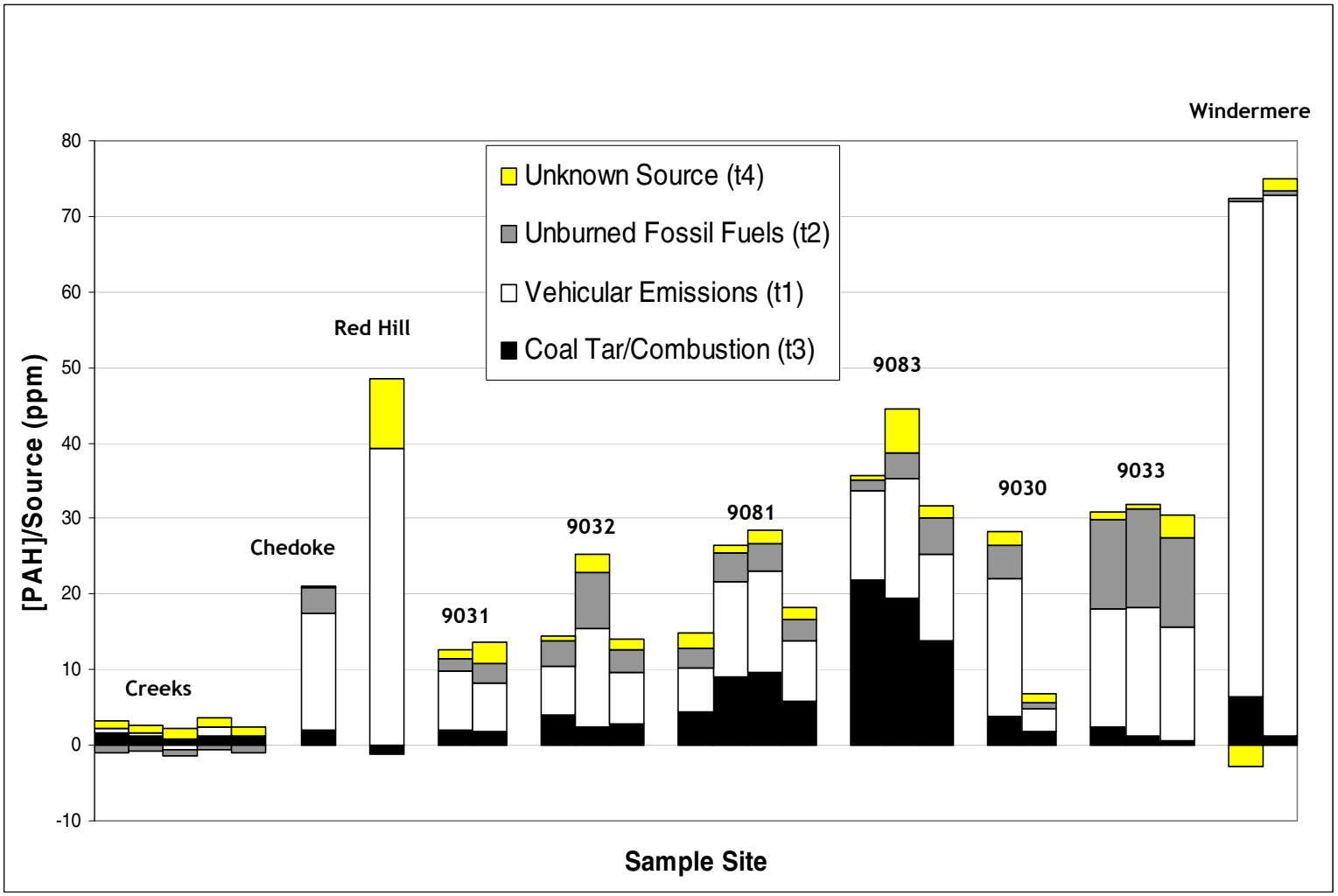
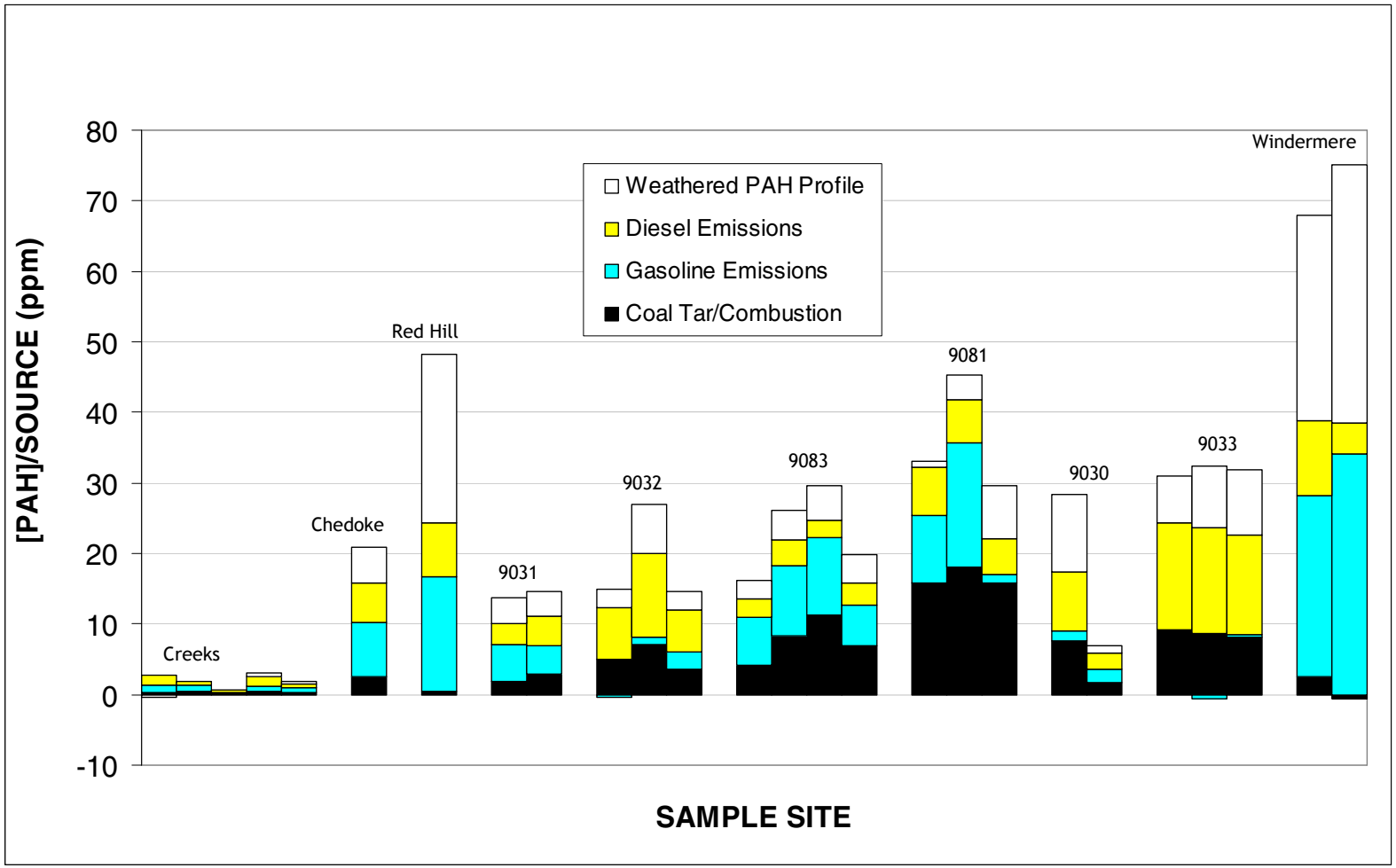


Figure S9: Contributions of four PMF factors (in µg/g) to the SumPAH (µg/g) of each suspended sediment sample collected in Hamilton Harbour and the creeks.



Principal Component Analysis of Hamilton Harbour Data

The PCA model in this context was run thus:
For a data matrix X_{nxm}
where n=number of rows indicating observations/samples
m = number of columns indicating species/elements

First standardize the x_{ij} matrix elements such that a new z matrix is produced in which

$$z_{ij} = (x_{ij} - \bar{x}_j)/\sigma_j \quad (1)$$

where Z_{ij} is the standard deviate of element x_{ij}
x_{ij} is the concentration value of the jth element in the ith sample
$\bar{x}_j$ is the mean concentration of the jth element over all observations
σ_j is the standard deviation of the concentrations of the jth element

The PCA is run such that

$$z_{ij} = \sum_{k=1}^{p} P_{ik} \bullet W_{kj} \quad (2)$$

where; k = 1…., p the pollutional source components
i = 1…., n the number of observations/samples
j = 1…., m the number of species/elements
W_{kj} is the coefficient matrix of the components (transpose of loadings matrix)
P_{ik} is the kth component's value for observation/sample i (scores)

Thurston and Spengler report that inversion of equation (2) yields

$$P_{ik} = \sum_{j=1}^{m} z_{ij} \bullet \frac{W_{jk}}{\lambda_k} \quad (3)$$

λ_k is an eigenvalue associated with the principal component P_k and it is derived from eigenvalue analysis of R_{jxj} the correlation matrix, i.e., find an eigenvector matrix U_{jxj} such that R_{jxj} is diagonalized

$$U_{jxj}^{-1} R_{jxj} U_{jxj} = \Lambda_{jxj} \quad (4)$$

since PCA assumes that components are not correlated, the eigenvector matrix must satisfy the square orthonormal rule i.e.,

$$U_{jxj}^{-1} U_{jxj} = U_{jxj} U_{jxj}^{-1} = I \quad (5)$$

for orthogonality of the components
The diagonal elements of Λ_{jxj} are the eigenvalues which are ordered in descending magnitude since each preceding eigenvector is chosen such that its eigenvalue from the diagonalization of the correlation matrix explains more variance in the data set than the succeeding one.

The elements in each column of the U_{jxj} matrix (i.e., eigenvectors) by definition are equal to $\frac{W_{jk}}{\lambda_k}$

Substituting the eigenvectors into equation (3) gives P_{ik} (scores)

A key feature of PCA is data dimension reduction. This is possible because the eigenvalues are ordered in descending magnitude, thus the number of pollutional components p is usually less than the number of species m, i.e., p<m

Multilinear regression MLR was done such that the scores P_{ik} were regressed against the standard deviate elements of SumPAH column vector i.e.,

$$z_{i_y} = \sum_{k=1}^{p} B_k P_{ik}$$

where B_k are the MLR coefficients for each column k of P_{ik} over all i.

More details on the PCA model can be found in:

1. Hopke P. K.; Receptor Modeling in Environmental Chemistry. Wiley-Interscience: NY, **1985**
2. Thurston G. D., Spengler J. D.; A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmospheric Environment*, Vol 19, No. 1, pp 9-25, **1985**