Supplementary Information

Predicting Polycyclic Aromatic Hydrocarbons using a Mass Fraction Approach in a Geostatistical Framework across North Carolina

Jeanette M. Reyes, PhD[1], Heidi Hubbard, PhD[2], Matthew A. Stiegel, PhD[3], Joachim D. Pleil, PhD[4,5], Marc L. Serre, PhD[5]*

Authors' affiliations
1 Oak Ridge Institute for Science and Education (ORISE) Research Participation Program, hosted at U.S. Environmental Protection Agency, RTP, North Carolina 27711, United States
2 ICF international, Fairfax, Virginia, USA
3 Duke University Medical Center
4 National Exposure Research Laboratory, US Environmental Protection Agency
5 Department of Environmental Sciences and Engineering, University of North Carolina – Chapel Hill

* Corresponding author
Department of Environmental Sciences and Engineering, UNC, 135 Dauer Drive, Chapel Hill, NC 27599-7431
email: marc_serre@unc.edu
telephone: 919 966 7014
fax: 919 966 7911

Projection Information

Observed PAH and PM2.5 location information were projected from decimal degrees to Lambert Conic Conformal (LCC) projection in order to deal with distances of the geostatistical estimations and predictions.

Estimation Methodology

The Bayesian Maximum Entropy (BME) posterior Probability Distribution Function (PDF) f_K describing the process x_k at an estimation point of interest $\boldsymbol{p}_k$ is given by the BME equation

$$f_K(x_k) = A^{-1} \int d\boldsymbol{x}_s \, f_S(\boldsymbol{x}_s) f_G(\boldsymbol{x}) \quad \text{(Equ. S1)}$$

where $\boldsymbol{x} = (x_k, \boldsymbol{x}_h, \boldsymbol{x}_s)$ is a realization of $\boldsymbol{X}$ at points $\boldsymbol{p} = (\boldsymbol{p}_k, \boldsymbol{p}_h, \boldsymbol{p}_s)$ and A is a normalization constant.

Our notation for the Space/Time Random Field (S/TRF) will consist of denoting a single random variable X in capital letters, its realization, x, in lower case and vectors and matrices in bold faces (e.g. $\boldsymbol{X} = [X_1, \dots, X_n]^T$ and $\boldsymbol{x} = [x_1, \dots, x_n]^T$). Let $X(\boldsymbol{p})$ be a S/TRF representing daily PAH. Typically, in the BME methodology the General Knowledge Base (G-KB) consists of a mean trend and covariance of $X(\boldsymbol{p})$ denoted as $G = \{m_X(\boldsymbol{p}), c_X(\boldsymbol{p}, \boldsymbol{p}')\}$ and the Site-specific Knowledge Base (S-KB) consists of data measured without error and data measured with error denoted as $S = \{\boldsymbol{x}_h, f_s(\boldsymbol{x}_s)\}$. In this work, $m_X(\boldsymbol{p})$ is assumed to be constant. The covariance model for the homogeneous/stationary S/TRF $X(\boldsymbol{p})$ is developed from the experimental covariance of the data fit through least squares empirical fitting.

The covariance model for the homogeneous/stationary S/TRF $X(\boldsymbol{p})$ is developed from the PAH experimental covariance. The experimental covariance value for a spatial lag r and a temporal lag τ is calculated as

$$\hat{c}_X(r, \tau) = \frac{1}{N(r,\tau)} \sum_{i=1}^{N(r,\tau)} X_{head,i} X_{tail,i} - {m_X}^2 \quad \text{(Equ. S2)}$$

where $N(r, \tau)$ is the number of pairs of values $(X_{head,i} X_{tail,i})$ separated by a spatial lag of r and time lag of τ and m_X is the mean of the $\boldsymbol{x}_h$ data. For each PAH, the 84 observed data were used to fit a single-structured space/time exponential covariance model given by the equation

$$c_X(r, \tau) = C_0 \exp\left(\frac{-3r}{a_r}\right) \exp\left(\frac{-3\tau}{a_t}\right) \quad \text{(Equ. S3)}$$

where r is the spatial distance (km), τ is the temporal distance (days), C_0 is the variance, a_r is the spatial range (km) and a_t is the temporal range (days). A one-structured exponential model was selected because it provided the best overall performance with respect to least squared fit.

Supplementary Table S1. Covariance model parameters for observed PAH data.

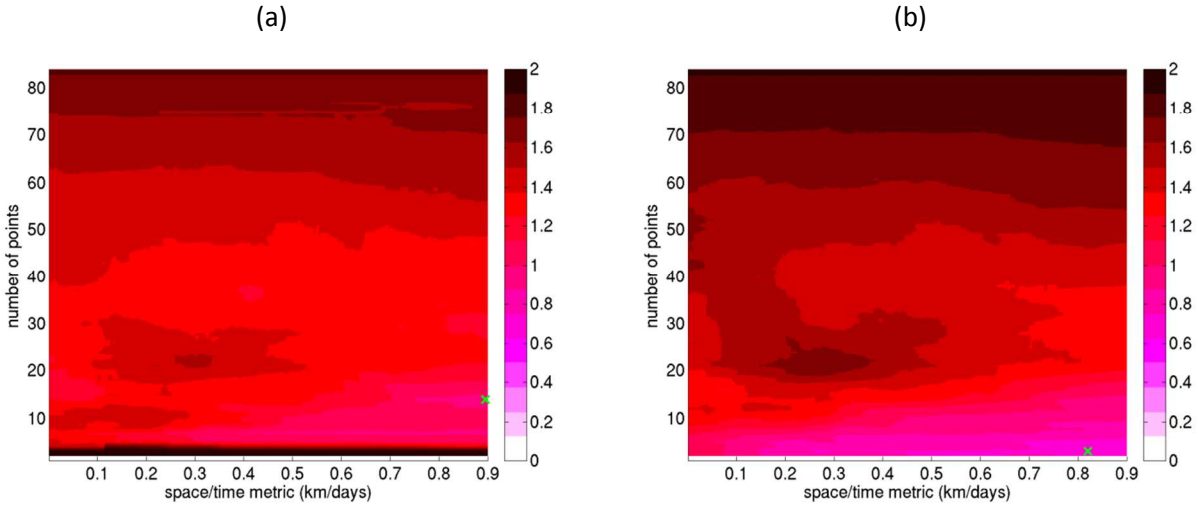
PAH	C_0 $((ng/m^3)^2)$	a_r (km)	a_t (days)
benz(a)anthracene	2.26	518	142
chrysene	2.09	521	147
benzo(b)fluoranthrene	2.38	914	109
benzo(k)fluoranthrene	2.11	362	155
benzo(e)pyrene	2.20	430	119
benzo(a)pyrene	2.13	247	146
indeno(1,2,3-c,d)pyrene	1.84	507	123
benzo(g,h,i)perylene	1.43	750	129
dibenzo(a,h)anthracene	3.01	312	123
Total PAH	1.68	356	123

Supplementary Table S2. Cokriging covariance model parameters for observed PAH and PM2.5 data. C_{PAH} is in in $(ng/m^3)^2$, C_{PM} is in $(\mu g/m^3)^2$and $C_{PAH,PM}$ is in $(ng/m^3) * (\mu g/m^3)$.

PAH	C_{PAH}	C_{PM}	$C_{PAH,PM}$	a_r (km)	a_t (days)
benz(a)anthracene	2.257	0.365	0.157	188	118
chrysene	2.091	0.365	0.135	206	119
benzo(b)fluoranthrene	2.381	0.365	-0.031	1176	78
benzo(k)fluoranthrene	2.112	0.365	0.069	223	102
benzo(e)pyrene	2.202	0.365	0.012	330	96
benzo(a)pyrene	2.131	0.365	-0.093	206	114
indeno(1,2,3-c,d)pyrene	1.838	0.365	-0.002	437	103
benzo(g,h,i)perylene	1.429	0.365	-0.054	650	101
dibenzo(a,h)anthracene	3.005	0.365	0.081	128	87
Total PAH	1.683	0.365	0.037	322	96

Supplementary Table S3. Cross validation statistics for all 9 PAHs and Total PAH.

PAH	statistic	kriging	cokriging	LR BME	LMF BME
benz(a)anthracene	ME	-0.210	-0.246	-0.228	-0.049
	VE	1.037	1.017	0.923	0.716
	MSE	1.069	1.065	0.964	0.710
	r^2	0.747	0.750	0.773	0.830
chrysene	ME	-0.202	-0.238	-0.200	-0.020
	VE	1.092	0.999	1.032	0.807
	MSE	1.119	1.044	1.060	0.798
	r^2	0.712	0.735	0.717	0.789
benzo(b)fluoranthrene	ME	0.015	0.029	0.157	0.018
	VE	1.215	1.329	1.724	1.019
	MSE	1.200	1.314	1.728	1.007
	r^2	0.723	0.701	0.637	0.762
benzo(k)fluoranthrene	ME	-0.219	-0.239	-0.179	-0.130
	VE	1.416	1.369	1.276	1.347
	MSE	1.447	1.410	1.292	1.348
	r^2	0.611	0.621	0.636	0.611
benzo(e)pyrene	ME	-0.137	-0.147	-0.068	0.019
	VE	0.948	0.933	0.871	0.646
	MSE	0.955	0.944	0.865	0.639
	r^2	0.776	0.779	0.781	0.848
benzo(a)pyrene	ME	-0.169	-0.174	0.069	-0.006
	VE	0.998	0.969	0.895	1.159
	MSE	1.015	0.988	0.889	1.145
	r^2	0.739	0.748	0.767	0.698
indeno(1,2,3-c,d)pyrene	ME	-0.078	-0.083	-0.017	0.020
	VE	0.842	0.831	0.896	0.686
	MSE	0.839	0.828	0.885	0.678
	r^2	0.758	0.761	0.755	0.804
benzo(g,h,i)perylene	ME	0.030	0.023	0.095	0.076
	VE	0.921	0.923	0.940	0.885
	MSE	0.911	0.912	0.938	0.880
	r^2	0.650	0.650	0.650	0.646
dibenzo(a,h)anthracene	ME	-0.205	-0.249	-0.031	-0.019
	VE	1.028	1.205	0.897	0.885
	MSE	1.058	1.253	0.888	0.874
	r^2	0.818	0.781	0.848	0.845
Total PAH	ME	-0.145	-0.137	-0.102	-0.042
	VE	0.806	0.782	0.764	0.591
	MSE	0.818	0.792	0.766	0.586
	r^2	0.747	0.752	0.744	0.821



Supplementary Figure S1. Exhaustive validation search of optimal estimation neighborhood for the (a) the Linear Regression method and (b) the log-Mass Fraction method for Total PAH displaying the MSE. The green "X" marks the lowest MSE $((ng/m^3)^2)$.

Supplementary Table S4. Mean difference in PAH near versus far from fires for kriging. 95% confidence intervals comparing the mean difference in predicted PAH near (within $100\ km$) versus far ($> 100\ km$) from fires for each of the 9 PAH and Total PAH for the **kriging** method where the control date of fires was extended 0-3 days. Units are in ng/m^3. *mean difference is statistically significant (p-value≤ 0.05), $^{\#}$mean difference > 0.

PAH	0 days	1 day	2 days	3 days
benz(a)anthracene	(-4.94E-03,-2.17E-03)*	(-5.08E-03,-2.69E-03)*	(-5.68E-03,-3.55E-03)*	(-6.38E-03,-4.37E-03)*
chrysene	(-6.67E-03,-3.53E-03)*	(-6.72E-03,-3.96E-03)*	(-7.29E-03,-4.81E-03)*	(-8.24E-03,-5.88E-03)*
benzo(b)fluoranthrene	(3.98E-03,1.11E-02)*,#	(7.08E-03,1.33E-02)*,#	(5.73E-03,1.14E-02)*,#	(4.04E-03,9.55E-03)*,#
benzo(k)fluoranthrene	(3.14E-03,6.47E-03)*,#	(4.47E-03,7.29E-03)*,#	(3.58E-03,6.09E-03)*,#	(2.76E-03,5.18E-03)*,#
benzo(e)pyrene	(-2.92E-03,2.23E-03)	(3.28E-04,4.98E-03)*,#	(-4.57E-04,3.74E-03)	(-2.14E-03,1.98E-03)
benzo(a)pyrene	(-3.83E-03,1.84E-03)	(1.48E-03,6.62E-03)*,#	(1.50E-03,6.16E-03)*,#	(4.25E-04,4.95E-03)*,#
indeno(1,2,3-c,d)pyrene	(1.87E-02,3.05E-02)*,#	(2.85E-02,3.88E-02)*,#	(2.75E-02,3.66E-02)*,#	(2.40E-02,3.29E-02)*,#
benzo(g,h,i)perylene	(3.04E-02,4.27E-02)*,#	(4.04E-02,5.09E-02)*,#	(4.17E-02,5.10E-02)*,#	(4.00E-02,4.89E-02)*,#
dibenzo(a,h)anthracene	(-2.07E-02,-1.36E-02)*	(-2.13E-02,-1.49E-02)*	(-2.27E-02,-1.68E-02)*	(-2.50E-02,-1.93E-02)*
Total PAH	(2.28E-02,6.75E-02)*,#	(5.79E-02,9.73E-02)*,#	(5.20E-02,8.74E-02)*,#	(3.69E-02,7.18E-02)*,#

Supplementary Table S5. Mean difference in PAH near versus far from fires for cokriging. 95% confidence intervals comparing the mean difference in predicted PAH near (within $100\ km$) versus far ($> 100\ km$) from fires for each of the 9 PAH and Total PAH for the **cokriging** method where the control date of fires was extended 0-3 days. Units are in ng/m^3. *mean difference is statistically significant (p-value≤ 0.05), $^{\#}$mean difference > 0.

PAH	0 days	1 day	2 days	3 days
benz(a)anthracene	(-2.61E-03,-1.07E-05)*	(-2.92E-03,-7.93E-04)*	(-2.88E-03,-9.55E-04)*	(-3.69E-03,-1.86E-03)*
chrysene	(-3.74E-03,-8.28E-04)*	(-4.10E-03,-1.69E-03)*	(-4.05E-03,-1.88E-03)*	(-5.12E-03,-3.04E-03)*
benzo(b)fluoranthrene	(3.78E-03,1.10E-02)*,#	(6.14E-03,1.26E-02)*,#	(5.10E-03,1.09E-02)*,#	(3.32E-03,8.95E-03)*,#
benzo(k)fluoranthrene	(2.32E-03,5.01E-03)*,#	(3.72E-03,6.05E-03)*,#	(3.43E-03,5.53E-03)*,#	(2.50E-03,4.56E-03)*,#
benzo(e)pyrene	(-3.17E-03,1.71E-03)	(2.19E-04,4.63E-03)*,#	(-1.85E-04,3.80E-03)	(-1.85E-03,2.07E-03)
benzo(a)pyrene	(-6.24E-03,-8.42E-04)*	(-1.03E-03,3.88E-03)	(-9.77E-04,3.49E-03)	(-1.95E-03,2.42E-03)
indeno(1,2,3-c,d)pyrene	(1.73E-02,2.87E-02)*,#	(2.74E-02,3.74E-02)*,#	(2.68E-02,3.55E-02)*,#	(2.33E-02,3.19E-02)*,#
benzo(g,h,i)perylene	(2.54E-02,3.66E-02)*,#	(3.58E-02,4.56E-02)*,#	(3.76E-02,4.63E-02)*,#	(3.63E-02,4.47E-02)*,#
dibenzo(a,h)anthracene	(-1.71E-02,-1.07E-02)*	(-1.75E-02,-1.19E-02)*	(-1.83E-02,-1.31E-02)*	(-2.05E-02,-1.55E-02)*
Total PAH	(2.13E-02,6.57E-02)*,#	(5.66E-02,9.61E-02)*,#	(5.09E-02,8.63E-02)*,#	(3.49E-02,6.98E-02)*,#

Supplementary Table S6. Mean difference in PAH near versus far from fires for linear regression BME. 95% confidence intervals comparing the mean difference in predicted PAH near (within $100\ km$) versus far ($> 100\ km$) from fires for each of the 9 PAH and Total PAH for the **linear regression BME** method where the control date of fires was extended 0-3 days. Units are in ng/m^3. *mean difference is statistically significant (p-value≤ 0.05), $^{\#}$mean difference > 0.

PAH	0 days	1 day	2 days	3 days
benz(a)anthracene	(-1.26E-03,1.07E-03)	(-1.32E-03,6.12E-04)	(-1.31E-03,4.29E-04)	(-1.88E-03,-2.00E-04)*
chrysene	(-9.40E-04,1.67E-03)	(-6.83E-04,1.50E-03)	(-7.60E-04,1.19E-03)	(-1.49E-03,4.07E-04)
benzo(b)fluoranthrene	(1.09E-02,2.23E-02)*,#	(1.32E-02,2.25E-02)*,#	(1.38E-02,2.30E-02)*,#	(1.35E-02,2.23E-02)*,#
benzo(k)fluoranthrene	(5.27E-03,7.94E-03)*,#	(6.79E-03,9.03E-03)*,#	(6.54E-03,8.55E-03)*,#	(5.78E-03,7.76E-03)*,#
benzo(e)pyrene	(5.22E-03,9.80E-03)*,#	(8.47E-03,1.24E-02)*,#	(8.57E-03,1.21E-02)*,#	(7.08E-03,1.06E-02)*,#
benzo(a)pyrene	(2.23E-03,1.37E-02)*,#	(6.98E-03,1.60E-02)*,#	(7.17E-03,1.49E-02)*,#	(6.14E-03,1.35E-02)*,#
indeno(1,2,3-c,d)pyrene	(2.04E-02,3.24E-02)*,#	(3.09E-02,4.15E-02)*,#	(3.03E-02,3.95E-02)*,#	(2.68E-02,3.58E-02)*,#
benzo(g,h,i)perylene	(2.72E-02,4.06E-02)*,#	(3.99E-02,5.29E-02)*,#	(4.09E-02,5.21E-02)*,#	(3.86E-02,4.92E-02)*,#
dibenzo(a,h)anthracene	(-4.80E-03,2.16E-03)	(-6.11E-03,9.01E-05)	(-5.74E-03,-1.02E-04)*	(-7.44E-03,-1.95E-03)*
Total PAH	(6.29E-02,1.03E-01)*,#	(9.71E-02,1.32E-01)*,#	(9.91E-02,1.30E-01)*,#	(8.48E-02,1.16E-01)*,#

Supplementary Table S7. Mean difference in PAH near versus far from fires for log-mass fraction BME. 95% confidence intervals comparing the mean difference in predicted PAH near (within $100\ km$) versus far ($> 100\ km$) from fires for each of the 9 PAH and Total PAH for the **log-mass fraction BME** method where the control date of fires was extended 0-3 days. Units are in ng/m^3. *mean difference is statistically significant (p-value≤ 0.05), #mean difference > 0.

PAH	0 days	1 day	2 days	3 days
benz(a)anthracene	(1.57E-03,4.38E-03)*,#	(1.01E-03,3.34E-03)*,#	(1.44E-03,3.53E-03)*,#	(8.80E-04,2.88E-03)*,#
chrysene	(2.07E-03,5.39E-03)*,#	(1.42E-03,4.25E-03)*,#	(2.24E-03,4.80E-03)*,#	(1.51E-03,3.99E-03)*,#
benzo(b)fluoranthrene	(2.36E-02,3.02E-02)*,#	(2.52E-02,3.13E-02)*,#	(2.66E-02,3.21E-02)*,#	(2.52E-02,3.05E-02)*,#
benzo(k)fluoranthrene	(7.80E-03,1.08E-02)*,#	(9.34E-03,1.19E-02)*,#	(9.18E-03,1.14E-02)*,#	(8.27E-03,1.04E-02)*,#
benzo(e)pyrene	(1.83E-02,2.49E-02)*,#	(1.97E-02,2.55E-02)*,#	(2.06E-02,2.58E-02)*,#	(1.88E-02,2.39E-02)*,#
benzo(a)pyrene	(5.14E-03,1.02E-02)*,#	(9.91E-03,1.46E-02)*,#	(1.16E-02,1.59E-02)*,#	(1.05E-02,1.47E-02)*,#
indeno(1,2,3-c,d)pyrene	(4.79E-02,6.11E-02)*,#	(5.35E-02,6.51E-02)*,#	(5.41E-02,6.42E-02)*,#	(4.99E-02,5.97E-02)*,#
benzo(g,h,i)perylene	(3.13E-02,4.06E-02)*,#	(3.96E-02,4.80E-02)*,#	(4.09E-02,4.83E-02)*,#	(3.88E-02,4.59E-02)*,#
dibenzo(a,h)anthracene	(1.90E-03,8.77E-03)*,#	(1.90E-03,7.98E-03)*,#	(2.88E-03,8.40E-03)*,#	(1.32E-03,6.70E-03)*,#
Total PAH	(1.72E-01,2.30E-01)*,#	(1.91E-01,2.42E-01)*,#	(1.96E-01,2.41E-01)*,#	(1.79E-01,2.23E-01)*,#