Supporting Information for:

Estimating $PM_{2.5}$ Concentrations in the Conterminous United States Using the Random Forest Approach

Xuefei Hu[1], Jessica H. Belle[1], Xia Meng[1], Avani Wildani[2], Lance A. Waller[3], Matthew J. Strickland[4], Yang Liu[1]*

[1] Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

[2] Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

[3] Department of Biostatistics & Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA

[4] School of Community Health Sciences, University of Nevada Reno, Reno, NV 89557, USA

* Corresponding Author:
Yang Liu
Tel: +1-404-727-2131, Fax: +1-404-727-8744
E-mail: yang.liu@emory.edu

This file includes:

Pages S1-S7

Tables S1-S4

Figure S1

Table S1. Descriptive statistics for dependent and independent variables (N=80254).

	Variable	Mean	Std. Dev.	Minimum	Maximum
	$PM_{2.5}$ ($\mu g/m^3$)	9.69	6.38	0	136.8
	AOD	0.14	0.15	-0.05	2.35
NARR	Dew point temperature (K)	289.61	4.79	254.8	299.85
	Visibility (m)	19167.65	2142.66	6.46	20024.2
	2-m pressure (pa)	98490.24	4030.34	74487.3	103552
	10-m pressure (pa)	98398.94	4026.85	74422	103384
	30-m pressure (pa)	98171.44	4018.1	74225.8	103214
	30-m temperature (K)	294.93	4.09	262.99	309.6
	180-150mb temperature (K)	286.24	3.52	253.7	298.75
	Potential evaporation (kg/m^2)	0.57	0.24	-0.09	1.36
NLDAS	Downward longwave radiation flux (W/m^2)	330.98	63.41	132.46	528.01
	Downward shortwave radiation flux (W/m^2)	616.74	159.43	86.03	1104.58
	Connective available potential energy (J/kg)	243.81	643.19	0	5250.3
	Pressure (pa)	97009.34	5498.84	68102	103859
	2-m temperature (K)	294.14	8.93	256.94	317.06
	2-m relative humidity (%)	45.69	17.14	3.12	98.69
	East-west component of the wind vector at 10 m (m/sec)	1.13	2.67	-12.77	14.5
	North-south component of the wind vector at 10 m (m/sec)	0.15	3.17	-12.93	13.15
	Forest cover	0.14	0.25	0	1
	Elevation (m)	386	491.42	-32.95	3312.35
	Point emissions (tons/year)	2.25	38.26	0	1009.28
	Local road length (m)	1306.59	1353.77	0	10609.03
	Impervious surface (%)	32.36	23.33	0	90.19
	Population density (population/km^2)	1179.68	2781.98	0	50666.81

Table S2. Pearson correlation coefficients between dependent and independent variables.

Variable	$PM_{2.5}$	AOD	Dew point temperature (NARR)	Visibility (NARR)	2-m pressure (NARR)	10-m pressure (NARR)
$PM_{2.5}$	1	0.41	0.17	-0.01	0.09	0.09
AOD	0.41	1	0.33	0.01	0.09	0.09
Dew point temperature (NARR)	0.17	0.33	1	-0.02	0.58	0.58
Visibility (NARR)	-0.01	0.01	-0.02	1	0.03	0.03
2-m pressure (NARR)	0.09	0.09	0.58	0.03	1	1
10-m pressure (NARR)	0.09	0.09	0.58	0.03	1	1
30-m pressure (NARR)	0.09	0.09	0.59	0.03	1	1
30-m temperature (NARR)	0.22	0.35	0.72	0.05	0.17	0.17
180-150mb temperature (NARR)	0.23	0.3	0.76	0	0.48	0.48
Potential evaporation (NLDAS)	0.14	0.3	0.43	0.05	0	0
Downward longwave radiation flux (NLDAS)	0.29	0.5	0.67	-0.02	0.12	0.12
Downward shortwave radiation flux (NLDAS)	0.05	0.27	0.38	0.06	-0.01	-0.01
Connective available potential energy (NLDAS)	0.32	0.5	0.4	0.01	0.08	0.08
Pressure (NLDAS)	0.24	0.14	0.26	-0.02	0.51	0.51
2-m temperature (NLDAS)	0.26	0.4	0.58	0.01	0.02	0.02
2-m relative humidity (NLDAS)	0.05	0.12	0.2	-0.06	0.24	0.24
East-west component of the wind vector at 10 m (NLDAS)	-0.06	-0.02	-0.04	0.01	-0.08	-0.08
North-south component of the wind vector at 10 m (NLDAS)	0.18	0.23	0.13	-0.01	-0.02	-0.02
Forest cover	-0.18	-0.1	-0.06	0	-0.12	-0.12
Elevation	-0.25	-0.15	-0.29	0.02	-0.5	-0.5
Point Emissions	0.02	0.01	0.02	0	0.01	0.01
Local road length	0.1	0.07	-0.01	0.01	0.05	0.05
Impervious surface	0.15	0.12	-0.01	-0.01	0.05	0.05
Population density	0.08	0.08	-0.01	0.01	0.08	0.08

Variable	30-m pressure (NARR)	30-m temperature (NARR)	180-150mb temperature (NARR)	Potential evaporation (NLDAS)	Downward longwave radiation flux (NLDAS)	Downward shortwave radiation flux (NLDAS)
$PM_{2.5}$	0.09	0.22	0.23	0.14	0.29	0.05
AOD	0.09	0.35	0.3	0.3	0.5	0.27
Dew point temperature (NARR)	0.59	0.72	0.76	0.43	0.67	0.38
Visibility (NARR)	0.03	0.05	0	0.05	-0.02	0.06
2-m pressure (NARR)	1	0.17	0.48	0	0.12	-0.01
10-m pressure (NARR)	1	0.17	0.48	0	0.12	-0.01
30-m pressure (NARR)	1	0.17	0.49	0	0.12	-0.01
30-m temperature (NARR)	0.17	1	0.74	0.68	0.76	0.55
180-150mb temperature (NARR)	0.49	0.74	1	0.52	0.67	0.45

Potential evaporation (NLDAS)	0	0.68	0.52	1	0.72	0.79
Downward longwave radiation flux (NLDAS)	0.12	0.76	0.67	0.72	1	0.54
Downward shortwave radiation flux (NLDAS)	-0.01	0.55	0.45	0.79	0.54	1
Connective available potential energy (NLDAS)	0.08	0.4	0.35	0.37	0.57	0.28
Pressure (NLDAS)	0.51	0.1	0.21	-0.02	0.21	-0.05
2-m temperature (NLDAS)	0.02	0.81	0.7	0.85	0.9	0.67
2-m relative humidity (NLDAS)	0.24	-0.19	-0.07	-0.45	-0.02	-0.25
East-west component of the wind vector at 10 m (NLDAS)	-0.08	-0.03	-0.11	0.06	-0.04	0.04
North-south component of the wind vector at 10 m (NLDAS)	-0.02	0.17	0.18	0.18	0.25	0.07
Forest cover	-0.12	-0.02	-0.08	0.02	-0.09	0.03
Elevation	-0.5	-0.14	-0.26	-0.02	-0.25	0.02
Point Emissions	0.01	0.01	0.01	0	0	0
Local road length	0.05	-0.05	-0.01	-0.08	-0.02	-0.05
Impervious surface	0.05	-0.03	0.02	-0.08	0	-0.04
Population density	0.08	-0.05	-0.02	-0.08	-0.02	-0.03

Variable	Connective available potential energy (NLDAS)	Pressure (NLDAS)	2-m temperature (NLDAS)	2-m relative humidity (NLDAS)	East-west component of the wind vector at 10 m (NLDAS)	North-south component of the wind vector at 10 m (NLDAS)
$PM_{2.5}$	0.32	0.24	0.26	0.05	-0.06	0.18
AOD	0.5	0.14	0.4	0.12	-0.02	0.23
Dew point temperature (NARR)	0.4	0.26	0.58	0.2	-0.04	0.13
Visibility (NARR)	0.01	-0.02	0.01	-0.06	0.01	-0.01
2-m pressure (NARR)	0.08	0.51	0.02	0.24	-0.08	-0.02
10-m pressure (NARR)	0.08	0.51	0.02	0.24	-0.08	-0.02
30-m pressure (NARR)	0.08	0.51	0.02	0.24	-0.08	-0.02
30-m temperature (NARR)	0.4	0.1	0.81	-0.19	-0.03	0.17
180-150mb temperature (NARR)	0.35	0.21	0.7	-0.07	-0.11	0.18
Potential evaporation (NLDAS)	0.37	-0.02	0.85	-0.45	0.06	0.18
Downward longwave radiation flux (NLDAS)	0.57	0.21	0.9	-0.02	-0.04	0.25
Downward shortwave radiation flux (NLDAS)	0.28	-0.05	0.67	-0.25	0.04	0.07
Connective available potential energy (NLDAS)	1	0.11	0.45	0.17	0.05	0.26
Pressure (NLDAS)	0.11	1	0.04	0.44	-0.11	-0.12
2-m temperature (NLDAS)	0.45	0.04	1	-0.33	-0.04	0.25
2-m relative humidity (NLDAS)	0.17	0.44	-0.33	1	0.01	-0.05
East-west component of the wind vector at 10 m	0.05	-0.11	-0.04	0.01	1	0.01

(NLDAS)						
North-south component of the wind vector at 10 m (NLDAS)	0.26	-0.12	0.25	-0.05	0.01	1
Forest cover	-0.03	-0.34	-0.04	-0.13	0.01	0.03
Elevation	-0.13	-0.98	-0.09	-0.42	0.08	0.1
Point Emissions	0.02	0.02	-0.01	0.03	0.01	0
Local road length	-0.02	0.14	-0.05	0.12	0.04	-0.05
Impervious surface	-0.03	0.21	-0.02	0.09	0.01	-0.05
Population density	-0.02	0.14	-0.05	0.08	0.03	-0.04

Variable	Forest cover	Elevation	Point emissions	Local road length	Impervious surface	Population density
$PM_{2.5}$	-0.18	-0.25	0.02	0.1	0.15	0.08
AOD	-0.1	-0.15	0.01	0.07	0.12	0.08
Dew point temperature (NARR)	-0.06	-0.29	0.02	-0.01	-0.01	-0.01
Visibility (NARR)	0	0.02	0	0.01	-0.01	0.01
2-m pressure (NARR)	-0.12	-0.5	0.01	0.05	0.05	0.08
10-m pressure (NARR)	-0.12	-0.5	0.01	0.05	0.05	0.08
30-m pressure (NARR)	-0.12	-0.5	0.01	0.05	0.05	0.08
30-m temperature (NARR)	-0.02	-0.14	0.01	-0.05	-0.03	-0.05
180-150mb temperature (NARR)	-0.08	-0.26	0.01	-0.01	0.02	-0.02
Potential evaporation (NLDAS)	0.02	-0.02	0	-0.08	-0.08	-0.08
Downward longwave radiation flux (NLDAS)	-0.09	-0.25	0	-0.02	0	-0.02
Downward shortwave radiation flux (NLDAS)	0.03	0.02	0	-0.05	-0.04	-0.03
Connective available potential energy (NLDAS)	-0.03	-0.13	0.02	-0.02	-0.03	-0.02
Pressure (NLDAS)	-0.34	-0.98	0.02	0.14	0.21	0.14
2-m temperature (NLDAS)	-0.04	-0.09	-0.01	-0.05	-0.02	-0.05
2-m relative humidity (NLDAS)	-0.13	-0.42	0.03	0.12	0.09	0.08
East-west component of the wind vector at 10 m (NLDAS)	0.01	0.08	0.01	0.04	0.01	0.03
North-south component of the wind vector at 10 m (NLDAS)	0.03	0.1	0	-0.05	-0.05	-0.04
Forest cover	1	0.39	-0.01	-0.34	-0.61	-0.21
Elevation	0.39	1	-0.02	-0.15	-0.23	-0.15
Point Emissions	-0.01	-0.02	1	-0.02	0.02	-0.02
Local road length	-0.34	-0.15	-0.02	1	0.61	0.37
Impervious surface	-0.61	-0.23	0.02	0.61	1	0.45
Population density	-0.21	-0.15	-0.02	0.37	0.45	1

Table S3. Cross validation results for models without convolutional layers for nearby $PM_{2.5}$ measurements and land use terms.

Model	R^2	RMSPE ($\mu g/m^3$)	MPE ($\mu g/m^3$)	Slope
Without the convolutional layer for nearby $PM_{2.5}$ measurements	0.74	3.26	2.16	1.12
Without convolutional layers for land use terms	0.78	2.94	1.85	1.03

Table S4. Cross validation R^2 for models using our convolutional layers and surfaces generated using universal kriging.

Model	Universal kriging	Our convolutional layers
All predictor variables included	0.76	0.80
Convolutional layers for nearby $PM_{2.5}$ measurements and land use terms only	0.53	0.71
The convolutional layer for nearby $PM_{2.5}$ measurements only	0.17	0.56

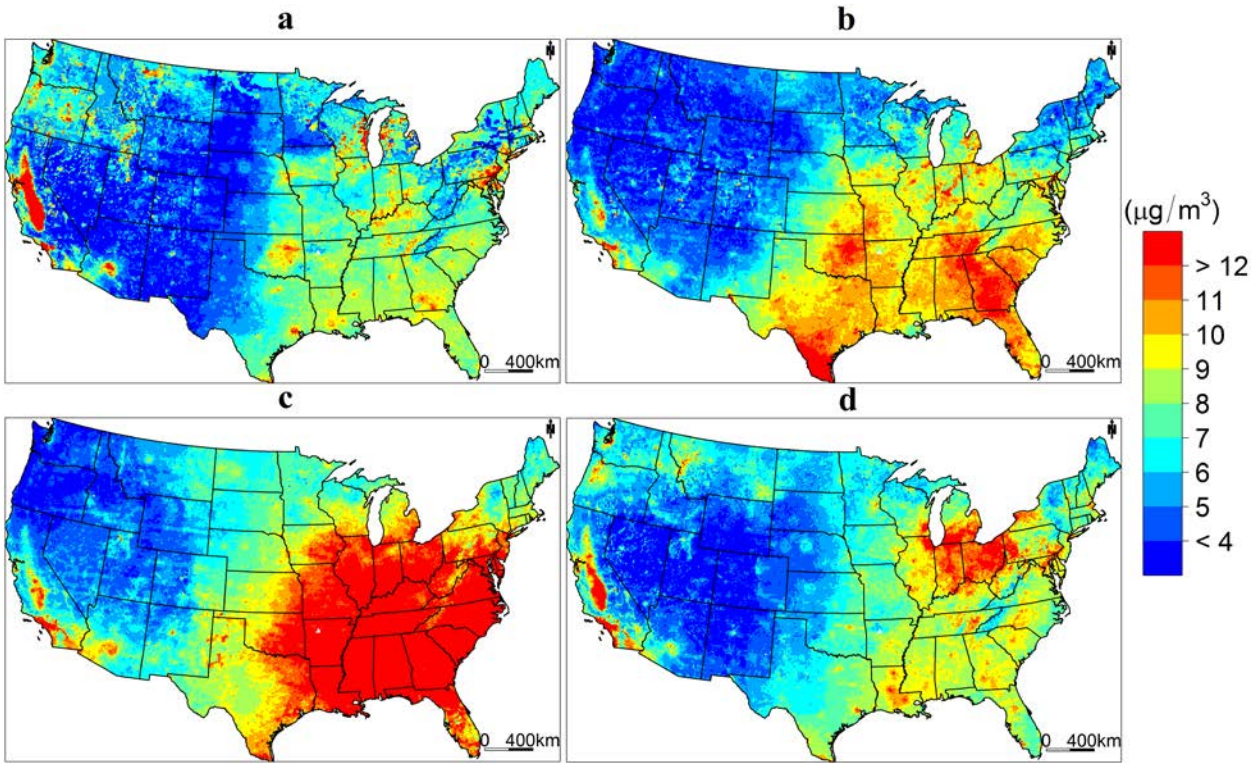

Figure S1. Seasonal mean predictions. (a) DJF (December, January, and February); (b) MAM (March, April, and May); (c) JJA (June, July, and August); (d) SON (September, October, and November).