CrossMark

Air pollutant exposure field modeling using air quality model-data fusion methods and comparison with satellite AOD-derived fields: application over North Carolina, USA

Ran Huang[1] · Xinxin Zhai[1] · Cesunica E. Ivey[1,2] · Mariel D. Friberg[1] · Xuefei Hu[3] ·
Yang Liu[3] · Qian Di[4] · Joel Schwartz[4] · James A. Mulholland[1] · Armistead G. Russell[1]

Abstract In order to generate air-pollutant exposure fields for health studies, a data fusion (DF) approach is developed that combines observations from ambient monitors and simulated data from the Community Multiscale Air Quality (CMAQ) model. These resulting fields capture the spatiotemporal information provided by the air quality model, as well as the finer temporal scale variations from the pollutant observations and decrease model biases. Here, the approach is applied to develop daily concentration fields for $PM_{2.5}$ total mass, five major particulate species (OC, EC, SO_4^{2-}, NO_3^-, and NH_4^+), and three gaseous pollutants (CO, NO_x, and NO_2) from 2006 to 2008 over North Carolina (USA). Several data withholding methods are then conducted to evaluate the data fusion method, and the results suggest that typical approaches may overestimate the ability of spatiotemporal estimation methods to capture pollutant concentrations in areas with limited or no monitors. The results show improvements in capturing spatial and temporal variability compared with CMAQ results. Evaluation tests for $PM_{2.5}$ led to an R^2 of 0.95 (no withholding) and 0.82 when using 10% random data withholding. If spatially based data withholding is used, the R^2 is 0.73. Comparisons of DF-developed $PM_{2.5}$ total mass concentration with the spatiotemporal fields derived from two other methods (both use satellite aerosol optical depth (AOD) data) find that, in this case, the data fusion fields have slightly less overall error, with an RMSE of 1.28 compared with 3.06 μg/m^3 (two-stage statistical model) and 2.74 (neural network-based hybrid model). Applying the Integrated Mobile Source Indicator (IMSI) method shows that the data fusion fields can be used to estimate mobile source impacts. Overall, the growing availability of chemically detailed air quality model fields and the accuracy of the DF field, suggest that this approach is better able to provide spatiotemporal pollutant fields for gaseous and speciated particulate pollutants for health and planning studies.

Introduction

Exposure to fine particulate air pollution ($PM_{2.5}$) has been associated with increased morbidity and premature mortality, suggesting that sustained reductions in pollution exposure could result in improved health and increased life expectancy (Gilboa et al. 2005; Sarnat et al. 2005; Pope et al. 2009; Matte et al. 2009; Solomon et al. 2012; Hubbell 2012). Estimating population exposure to $PM_{2.5}$ has traditionally been done by assigning measurements of a central ground monitor to people living within the region (Kanaroglou et al. 2005; Sampson et al. 2013). However, a number of studies have shown the limitations of using central ground monitor data as the exposure metric (Lefohn et al. 1987; Wade et al. 2006; Beelen et al. 2009; Kim et al. 2014; Dionisio et al. 2016). These limitations

✉ Armistead G. Russell
  Ted.russell@gatech.edu

[1] Civil and Environmental Engineering Department, Georgia Institute of Technology, Atlanta, GA, USA

[2] Department of Physics, University of Nevada Reno, Reno, NV, USA

[3] Rollins School of Public Health, Emory University, Atlanta, GA, USA

[4] Department of Environmental Health, Harvard T.H. Chan School of Public Heath, Harvard University, Boston, MA, USA

include monitoring sites in national regulatory networks that are relatively sparse across broad regions of the country (Hu et al. 2014a) and pollutant concentrations that can be impacted by local emissions, leading to local variations (Hu et al. 2014b). A variety of modeling approaches are now being used to better estimate pollutant concentration variations not captured by monitors (Marmur et al. 2005; Johnson et al. 2010; Liu et al. 2012).

One approach to develop air quality fields is using chemical transport models (CTMs) that account for local variations affected by emissions and meteorology (Godowitch et al. 2015; Kim et al. 2015; Pleim et al. 2016). The Community Multiscale Air Quality (CMAQ; Binkowski 2003; Byun and Schere 2006) model is a state-of-the-science chemical transport model (CTM) designed to follow the dynamics of air pollutants from emissions. CMAQ captures spatial and temporal variations (Friberg et al. 2016) but is subject to errors due to limitations in insufficient characterization of meteorological (Yu et al. 2012) and emission inputs (Gilliland et al. 2008; Xiao et al. 2010; Ivey et al. 2015), as well as physical and chemical processes (Carlton et al. 2008; Tang et al. 2011; Ivey et al. 2016).

The objective of this research is to use the data fusion (DF) approach to develop spatiotemporal concentration fields for $PM_{2.5}$ mass, five PM species, and three gases for the state of North Carolina to support the University of North Carolina at Chapel Hill's health analysis of coronary heart disease patients in NC (McGuinn et al. 2017). The data fusion approach is developed at a spatial resolution of 12 km that combines observations from ambient monitors and data from CMAQ to better estimate ground-level air pollutant concentration fields for improved exposure estimates (Friberg et al. 2016). Several data withholding methods, which involve the use of monitor observations, were used to evaluate the stability of the data fusion method. A comparison of total $PM_{2.5}$ mass concentration is made between the results using unadjusted CMAQ pollutant fields, the data fusion application, ordinary kriging, and two satellite aerosol optical depth (AOD) data-included methods (Hu et al. 2014a; Di et al. 2016). These were compared as a part of evaluating the performance of various $PM_{2.5}$ exposure methods. Exposure fields of five PM species and three gases were also compared between CMAQ results and data fusion method results.

Methods

Four statistical methods were used to create the spatiotemporal fields, and the results were compared with each other and evaluated against observations. The first statistical method used was the data fusion method. The data fusion method combines observations and modeled pollutant fields and was used during 2006–2008 period over

North Carolina. (The data fusion method was actually applied from 2002 to 2010; 2006–2008 is in the middle part of that period and could be representative of the meteorological conditions experienced over that time.) The second and third methods were a two-stage statistical model and a neural network-based hybrid model, which both use satellite aerosol optical depth (AOD) and other data to develop $PM_{2.5}$ fields separately. Reliance on AOD data led to those methods being applied just to $PM_{2.5}$ mass, not individual PM or gaseous species. The fourth method uses ordinary kriging of observations at monitoring sites and was applied to develop $PM_{2.5}$ and CO fields. Other pollutant species were monitored at very few locations, limiting the amount of information available to develop spatiotemporal exposure fields as well as conduct a more thorough evaluation.
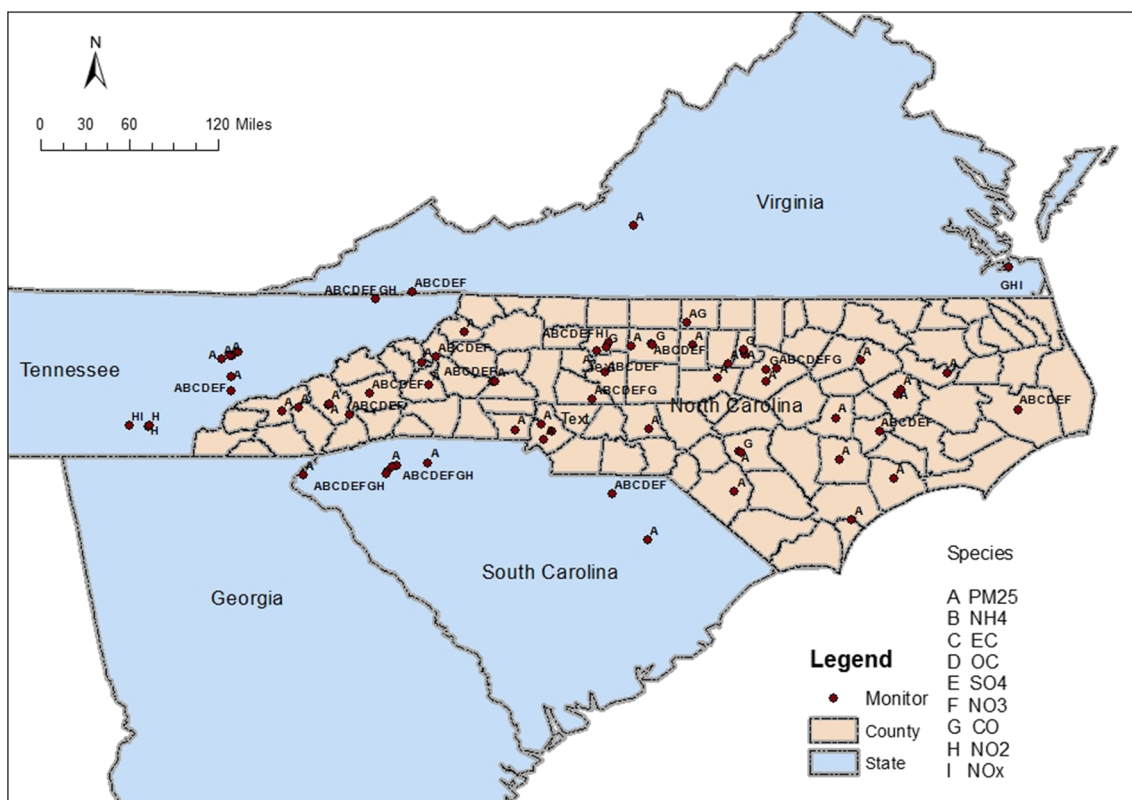
Air quality data

The observations used for data fusion come from the State and Local Air Monitoring Stations (SLAMS), Chemical Speciation Network (CSN) (Chu 2004) and Interagency Monitoring of Protected Visual Environments (IMPROVE) (Malm et al. 1994) networks. Observations from all available networks are utilized together. Pollutants include concentrations of three gases (carbon monoxide (CO), nitrogen dioxide (NO_2), and nitrogen oxide (NO_x)), $PM_{2.5}$ mass, and five $PM_{2.5}$ components (elemental carbon (EC), organic carbon (OC), ammonium (NH_4^+), nitrate (NO_3^-), and sulfate (SO_4^{2-})) (Fig. 1). Because of the limited number of monitoring sites for some species (e.g., CO, NO_2, and NO_x) in NC, we also included monitoring sites in neighboring states.

Twenty-four-hour average $PM_{2.5}$ concentrations for years 2006 to 2008 were collected from the EPA's Air Quality System Technology Transfer Network for use in the two-stage statistical model. The MODIS aerosol data (collection 5) at 550 nm wavelength were obtained from the NASA Earth Observing System Data Gateway at the Goddard Space Flight Center.

Chemical transport model simulated concentrations

Pollutant concentration fields used in this paper are developed using CMAQ model version 4.5 at 12-km resolution for the 2006–2008 period over North Carolina. A comprehensive model evaluation (Wyat Appel et al. 2008) of CMAQ version 4.5 conducted by the USEPA showed that simulated particulate nitrate and ammonium are biased high in the fall due to an overestimation of seasonal ammonia emissions (Qin et al. 2015). The EPA evaluation also found that simulated carbonaceous aerosol concentrations are biased low during the late spring and summer due to the lack of some secondary organic aerosol (SOA)

Fig. 1 Ambient air quality monitor locations used in this analysis. (Not all monitor locations have all species)

formation pathways in the model (Jathar et al. 2016; Woody et al. 2016).

Data fusion

The approach used to combine the CMAQ-derived fields with observed pollutant concentrations was described in detail in Friberg et al. (2016). The method blends observations and CMAQ results based on spatial correlation analysis between observations and CMAQ simulations and generates a new field that captures local observations, as well as spatial variability from CMAQ. A summary is provided in the Electronic supplementary material.

Data fusion results were integrated with the Integrated Mobile Source Indicator (IMSI) method (Pachon et al. 2012) to estimate the influence of mobile sources on $PM_{2.5}$. The IMSI method, which is developed for use in air quality and epidemiologic analyses, uses EC and NO_x as indicators of diesel vehicle (DV) and CO and NO_x as indicators of gasoline vehicle (GV) impacts. Here, the IMSI method, along with pollutant fields derived from the data fusion method, are used to provide spatiotemporal fields of mobile source impacts for use in source-specific, multipollutant, health analyses. The method is described in detail in the Electronic supplementary material.

Interpolation

Ordinary kriging (Cressie 1988) was applied to observed $PM_{2.5}$ and CO to develop air quality fields for comparison with the more advanced methods. $PM_{2.5}$ originates from multiple sources, both primary and secondary, whereas CO originates largely from mobile sources. $PM_{2.5}$ and CO are monitored at more sites than PM species and primary mobile source gases.

Methods utilizing satellite aerosol optical depth for $PM_{2.5}$ estimation

Two-stage statistical model

A two-stage statistical model (Hu et al. 2014a) employing satellite-retrieved aerosol optical depth (AOD) at 10 km resolution from Moderate Resolution Imaging SpectroRadiometer (MODIS) was used to develop $PM_{2.5}$ fields. The grids were restructured for comparison at 12 km resolution. The model includes a linear mixed effects module with day-specific random intercepts and slopes for AOD and meteorological fields as the first stage to account for the day-to-day variability in the $PM_{2.5}$-AOD relationship. The second stage is a geographically weighted regression model to capture spatial variation. Details of the method are found elsewhere (Hu et al. 2014a).

Table 1 Annual average concentrations from data fusion and CMAQ over the NC domain

|	$PM_{2.5}$ ($\mu g/m^3$) Monitor no.	EC ($\mu g/m^3$)	OC ($\mu g/m^3$)	NH_4^+ ($\mu g/m^3$)	NO_3^- ($\mu g/m^3$)	SO_4^{2-} ($\mu g/m^3$)	NO_2 (ppb)	NO_x (ppb)	CO (ppb)
	60	19	19	19	19	19	9	4	14
Data fusion									
2006	11.12 ± 5.09	0.47 ± 0.44	2.03 ± 1.82	1.39 ± 0.75	2.16 ± 2.91	3.91 ± 2.40	9.00 ± 4.30	8.39 ± 14.26	302.14 ± 104.53
2007	10.78 ± 5.15	0.45 ± 0.36	2.18 ± 1.96	1.37 ± 0.76	0.75 ± 1.03	3.92 ± 2.39	8.80 ± 3.93	6.70 ± 11.64	231.87 ± 79.87
2008	9.70 ± 4.69	0.31 ± 0.24	1.87 ± 1.81	1.44 ± 0.78	0.47 ± 0.69	3.31 ± 1.85	8.25 ± 3.79	5.28 ± 7.50	279.83 ± 92.97
CMAQ									
2006	8.97 ± 5.30	0.33 ± 0.39	1.20 ± 1.25	1.24 ± 0.77	0.86 ± 1.32	3.30 ± 2.07	2.61 ± 2.53	2.80 ± 2.83	163.33 ± 49.28
2007	9.09 ± 5.62	0.31 ± 0.27	1.51 ± 1.54	1.17 ± 0.73	0.96 ± 1.50	2.79 ± 1.71	3.16 ± 3.06	3.16 ± 3.54	165.48 ± 52.60
2008	6.90 ± 4.66	0.38 ± 0.33	1.19 ± 1.30	0.79 ± 0.49	0.60 ± 1.04	2.01 ± 1.14	2.81 ± 2.64	3.07 ± 3.09	153.64 ± 49.35

Neural network-based hybrid model

Di et al. (2016) applied another method that uses a neural network-based hybrid model that includes satellite-based AOD data from MODIS, absorbing aerosol index (AAI), chemical transport model (GEOS-Chem) output, land-use terms, and meteorological variables. The method has been used to estimate the national $PM_{2.5}$ fields at 1 km × 1 km resolution. Detailed description is found in a previous publication (Di et al. 2016). We extracted the results for North Carolina for 2006 to 2008.

Model evaluation methods

The performance of the data fusion method was evaluated by using three data withholding methods, as described in following subsections.

Random data withholding

Ten groups of observational data were constructed, with each group having 10% of the data randomly (not linked to specific monitors) withheld. Each group was run independently.
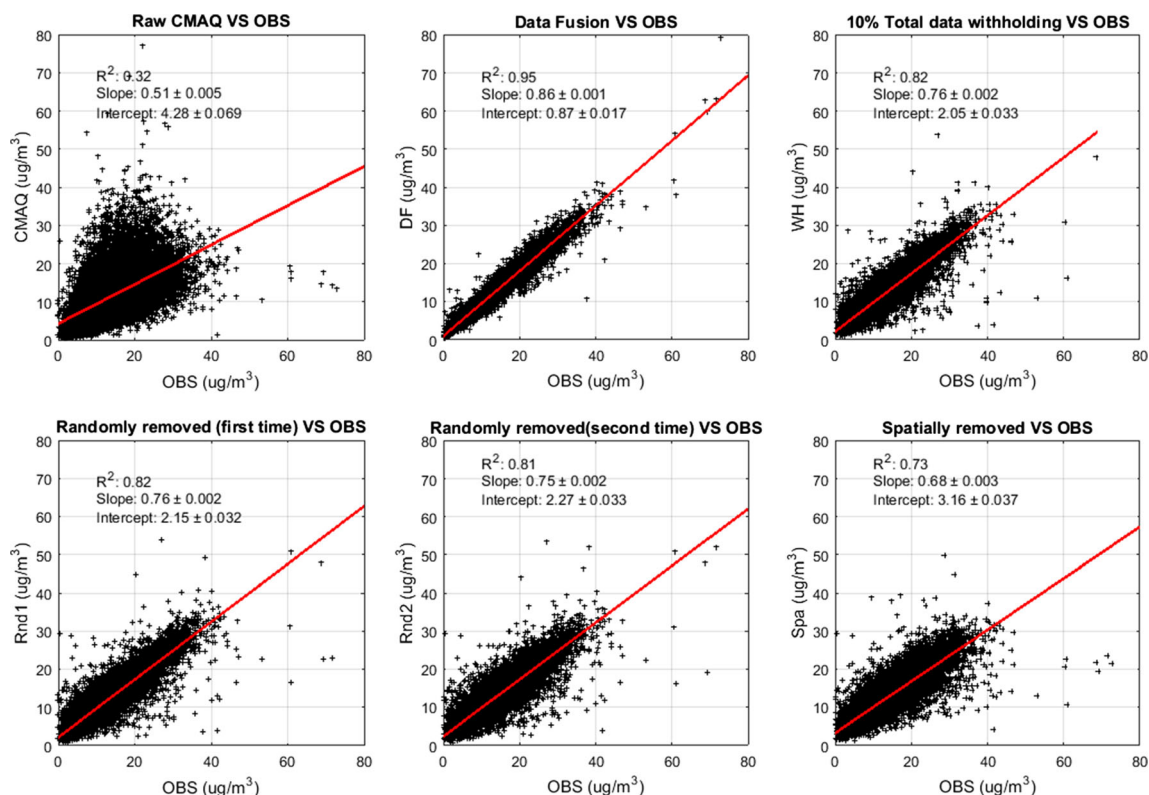
Fig. 2 Linear regression between observation (*OBS*) and simulations ($PM_{2.5}$)

Performance was assessed by comparing the simulated values to the data that were withheld for that iteration.

Randomly based monitor data withholding

Even though the random data withholding method is commonly used, it may overestimate the performance of the data fusion method. Monitor-based cross-validation may better reflect performance of the data fusion method because it is representative of areas where no monitor is located as opposed to a situation where a measurement is missing. In this case, the entire set of 60 $PM_{2.5}$ monitors were randomly split into ten subsets with six monitors in each subset. For each of ten cross-validation iterations, one subset (10% of monitors) was selected as the testing sample and the remaining nine subsets (90% of the monitors) were used to reapply the method. Estimates of the withheld monitor values were compared with the actual monitor values. This randomly based monitor data withholding was repeated twice to check the stability of this evaluation to the random choice of monitor grouping. For NO_2 and CO, leave one monitor out (LOO) was applied (i.e., in each test only one monitor data has been removed) due to the limited number of monitors available in the domain.

Spatially based monitor data withholding

Monitors may be clustered such that when one is removed there are nearby monitors that lead to the various methods being able to accurately estimate the pollutant levels for the removed monitor. This can result in an overestimation of a model's ability to provide accurate concentration estimates in a region with no monitors. Here, the entire set of monitors was spatially split into ten subsets (Fig. S1) according to their locations, and withholding was performed with the spatially based removed subsets.

Results and discussion

CMAQ

As a baseline, the unadjusted CMAQ results are evaluated over the NC domain. Annual average $PM_{2.5}$ shows that concentrations from CMAQ results (Table 1) are higher in 2007 for most species than in 2006 and 2008. For $PM_{2.5}$, the R^2 between pollutant observations and CMAQ simulations over the 3-year period is 0.32 and a root mean square error (RMSE) is 5.16 μg/m^3. Linear regression (Fig. 2; Table 2) between pollutant observations and CMAQ has a slope of 0.51. Evaluation results for other species tend to be have lower correlations (Table 2).

Data fusion

There are decreasing trends in the annual average concentration for all species from 2006 to 2008 in the data fusion results (Table 1). The annual average concentrations for each species from the DF method are higher than those from the CMAQ results. The probability density distributions of all species concentrations are log-normally distributed (Fig. S2).

Spatial plots of the annual averages for each of the nine pollutants show high concentrations in major urban centers (Fig. 3; Figs. S3a and S3b). Emission impacts are evident near the major interstates in the NO_2, NO_x, and CO fields. Concentrations at the western and eastern boundaries are much lower than the other areas because these are forest and coastal areas, respectively.

Monthly trends in North Carolina averaged over 3 years (Fig. S4) show that the concentrations of $PM_{2.5}$ and SO_4^{2-} are higher in the summer and lower in the winter in North Carolina, while NO_3^-, EC and OC are lower in the summer and higher in the winter. Concentrations of CO, NO_x, and NO_2 are higher in the winter and lower in the summer. These trends are expected based on the atmospheric formation chemistry of the secondary components (i.e., sulfate formed in summer and nitrate in winter) and the mixing height (lower in winter) due to meteorological conditions.

Mobile source impacts are estimated using the IMSI method applied to the DF fields. IMSI impacts decrease in the summer and increase in the fall (Fig. 4). The reduction of gasoline vehicle impacts is larger than the reduction of diesel vehicle impacts during the summer months. Emission-based IMSI value for gasoline ($IMSI_{GV}$) and emission-based IMSI value for diesel vehicles ($IMSI_{DV}$) are higher in 2007 than 2006 and 2008 (Fig. S5). The elevated impact areas near highways indicate that the method captures a mobile source activity and the data fusion fields are trustable (Fig. S6).

Temporal correlations between IMSI impacts and $PM_{2.5}$ concentrations indicate that highly populated and busy traffic areas have lower temporal correlations than other areas (Fig. S7). The correlations between $PM_{2.5}$ and EC, CO, and NO_x are low in rural areas (Fig. S8). The low temporal correlation between $PM_{2.5}$ and the primary pollutants is because much of the $PM_{2.5}$ in the area is secondary (Gertler et al. 2000; Gertler 2005). The annual average spatial correlations between IMSI impacts and $PM_{2.5}$ concentrations are 0.72 (2006), 0.71 (2007), and 0.78 (2008).

Ten percent random data withholding (Fig. S9) led to a R^2 of 0.82 (Fig. 2) for $PM_{2.5}$, 0.24 (Fig. S10) for CO and 0.78 (Fig. S11) for NO_2. Reapplying the method led to very similar correlations (e.g., for $PM_{2.5}$, the R^2 was 0.81). Spatial 10% monitor withholding cross-validation (only applied to $PM_{2.5}$ due to the lack of monitors) led to a lower R^2 of 0.73 (Fig. 2). The LOO results for CO and NO_2 also have lower R^2 values than the random data withholding, with a decrease

Table 2 Method performance evaluation (CMAQ, DF, and DF-WH) for PM$_{2.5}$ and PM$_{2.5}$ species (EC, OC, NH$_4^+$, NO$_3^-$, and SO$_4^{2-}$) and mobile source-related gases NO$_2$, NO$_x$, and CO, 24-h average values

OBS versus simulations

Species (no. of monitors)	PM$_{25}$ (60) (µg/m³) CMAQ	DF	DF-WH	EC (19) (µg/m³) CMAQ	DF	DF-WH	OC (19) (µg/m³) CMAQ	DF	DF-WH	NH$_4$ (19) (µg/m³) CMAQ	DF	DF-WH	NO$_3$ (19) (µg/m³) CMAQ	DF
Bias (NME)	0.36	0.10	0.16	0.45	0.30	0.39	0.51	0.24	0.36	0.37	0.12	0.23	0.96	1.11
RMSE	5.16	1.28	2.48	0.51	0.35	0.36	2.03	1.06	1.27	0.74	0.25	0.42	1.15	1.59
R^2	0.32	0.95	0.82	0.28	0.62	0.32	0.26	0.71	0.45	0.34	0.92	0.67	0.41	0.52
Linear regression ($y = \alpha x + \beta$)														
Slope (α)	0.507 ± 0.011	0.864 ± 0.003	0.76 ± 0.002	0.651 ± 0.004	0.811 ± 0.003	0.44 ± 0.01	0.562 ± 0.002	0.892 ± 0.002	0.59 ± 0.01	0.851 ± 0.003	1.029 ± 0.001	0.72 ± 0.01	1.306 ± 0.008	1.881 ± 0.01
Intercept (β)	5.827 ± 0.163	0.835 ± 0.038	2.05 ± 0.03	0.023 ± 0.001	0.017 ± 0	0.31 ± 0.009	0.038 ± 0.004	0.042 ± 0.003	1.05 ± 0.04	0.019 ± 0.002	0.005 ± 0	0.51 ± 0.02	0.013 ± 0.003	0.008 ± 0.004

OBS versus simulations

Species	NO$_3$ (19) (µg/m³) DF-WH	SO$_4$ (19) (µg/m³) CMAQ	DF	DF-WH	NO$_2$ (9) (ppb) CMAQ	DF	DF-WH	NO$_x$ (4) (ppb) CMAQ	DF	DF-WH	CO (14) (ppb) CMAQ	DF	DF-WH
Bias (NME)	1.25	0.34	0.07		0.51	0.15	0.19	0.52	2.03	0.27	0.52	0.39	0.47
RMSE	1.49	1.86	0.42		7.14	2.44	0.98	12.29	49.44	3.16	268.76	231.16	178.04
R^2	0.40	0.60	0.97		0.18	0.81	0.82	0.26	0.76	0.78	0.08	0.26	0.24
Linear regression ($y = \alpha x + \beta$)													
Slope (α)	1.62 ± 0.03	0.699 ± 0.002	0.9996 ± 0		0.573 ± 0.008	0.878 ± 0.003	0.81 ± 0.006	0.496 ± 0.008	3.11 ± 0.022	0.61 ± 0.007	0.368 ± 0.003	0.82 ± 0.005	0.25 ± 0.008
Intercept (β)	0.25 ± 0.03	0.05 ± 0.004	0.0084 ± 0.001		0.837 ± 0.082	0.483 ± 0.082	0.76 ± 0.03	3.203 ± 0.132	−2.025 ± 0.378	3.54 ± 0.08	63.445 ± 1.087	91.417 ± 1.91	362.75 ± 3.49

*DF-WH* 10% random data withholding, *NME* normalized mean error
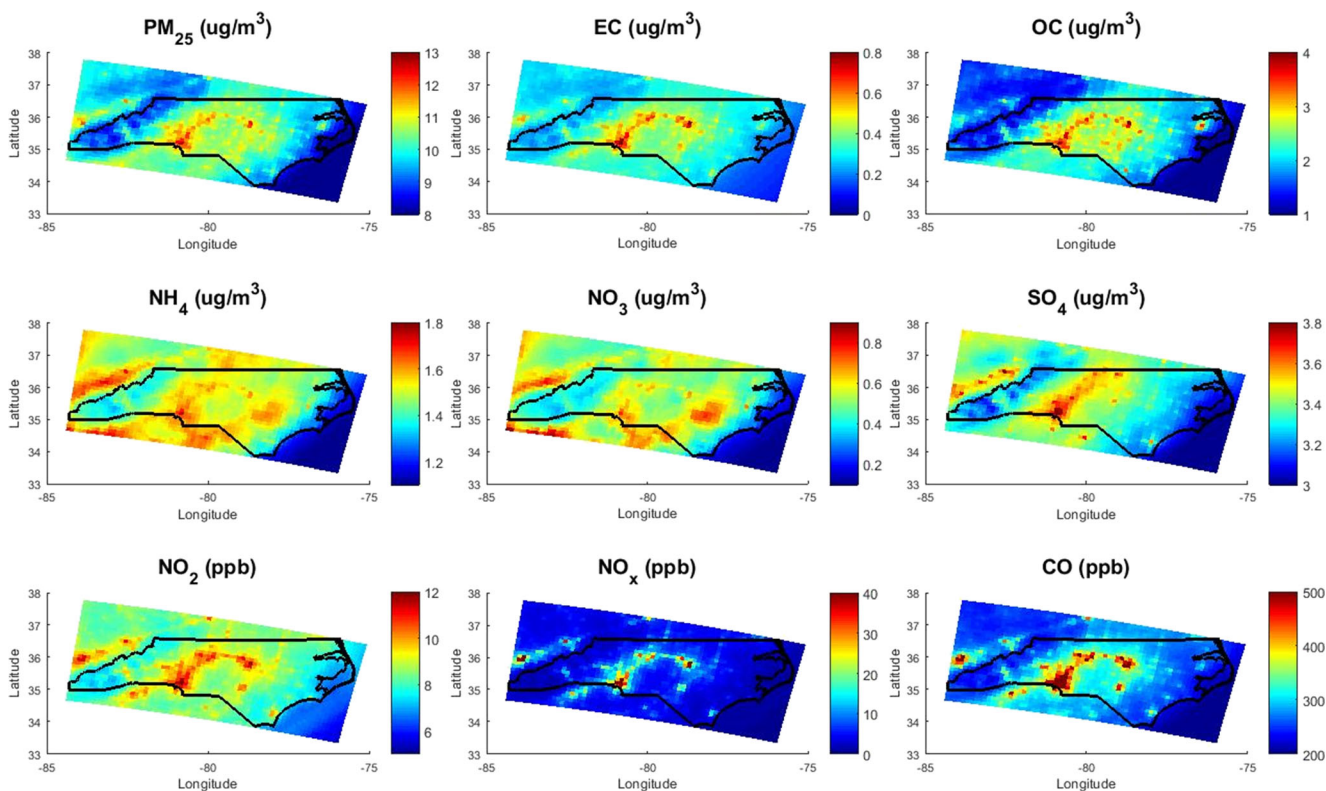
Fig. 3 Annual average spatial distribution fields from data fusion, 2008

from 0.24 to 0.10 for CO and from 0.78 to 0.52 for NO_2. Although there is a small difference in $PM_{2.5}$ RMSE results of approximately 1.20 μg/m³ between the 10% random data withholding results and the original DF data sets (Fig. S12; Tables 3, 4, and 5), both of these values are much smaller than the CMAQ RMSE results of 5.16 μg/m³. Spatial distributions of the maximum root-mean-squared deviation (mRMSD: The maximum daily root-mean-squared deviation value throughout the whole year.) for $PM_{2.5}$ show that the largest mRMSD are lower than 2, except in northeastern NC in 2008 (Fig. S13a and S13b). The RMSD of spatially removed groupings (Fig. S14) is similar to randomly removed groupings (Fig. S13a and S13b) for $PM_{2.5}$, except for the northeast area of North Carolina in 2008 because of the limited monitors in this area (Fig. 1). NO_2 results are similar, with RMSE decreasing from 7.1 ppb (CMAQ) to 2.4 ppb (data fusion) (Table 5). For CO, RMSE decreases from 269 ppb (CMAQ) to 231 ppb (data fusion) (Table 4).

RMSEs of LOO results for NO_2 and CO also show larger increases compared with 10% random data withholding results (Tables 4 and 5). All monitor-based withholding cross-validation for $PM_{2.5}$, CO, and NO_2 have larger RMSE and smaller R^2 than 10% random data withholding results.

The spatial 10% monitor withholding leads to a lower R^2 and higher RMSE for $PM_{2.5}$ as compared with random 10% monitor withholding (Table 3) with RMSE increases from 2.48 μg/m³ (random) to 2.81 μg/m³ (spatial). When removing values in spatially similar groupings, kriging results are minimally impacted by distant observations. As a result, the CMAQ simulations are more heavily weighted and the performance of the withheld data fusion results worsens. The LOO test for NO_2 and CO shows the influence of the distribution and quantity of the monitoring sites. CO monitors are located mainly in urban areas, while NO_2 monitors are distributed more widely. There are fewer monitors for both NO_2 and CO than for $PM_{2.5}$.
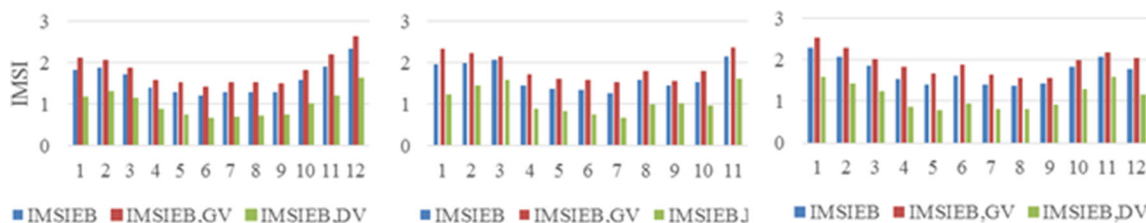
Fig. 4 Monthly trends of $IMSI_{EB}$, $IMSI_{EB, GV}$, and $IMSI_{EB, DV}$ from 2006 to 2008 (unitless)

Table 3 Performance evaluation for observation (OBS) and simulations (PM$_{2.5}$) using data withholding approaches, 24-h average values

	Mean (μg/m^3)	Median (μg/m^3)	RMSE	NME	R^2
Observations	12.7	11.5	0	0	1
CMAQ	10.9	9.6	5.16	0.38	0.32
Data fusion	11.8	10.8	1.28	0.10	0.95
DF-10% random data withholding	11.8	10.8	2.48	0.16	0.82
DF-random 10% monitors withholding (first test)	12.0	11.2	2.37	0.16	0.82
DF-random 10% monitors withholding (second test)	12.3	11.4	2.49	0.17	0.81
DF-spatial 10% monitors withholding	12.3	11.5	2.81	0.19	0.73
Interpolation (ordinary kriging)	12.7	11.5	0.67	0.02	0.99
Ordinary kriging: random 10% monitors withholding (first test)	12.8	11.6	2.64	0.13	0.83
Ordinary kriging: random 10% monitors withholding (second test)	12.7	11.6	2.74	0.14	0.81
Ordinary kriging: spatial 10% monitors withholding	12.6	11.5	3.23	0.19	0.71
Two-stage statistical model (no withholding)	12.8	11.7	3.06	0.15	0.81
Neural network-based hybrid model (no withholding)	12.2	11.0	2.74	0.15	0.82

NME normalized mean error

Ordinary kriging interpolation

Annual average PM$_{2.5}$ and CO spatial plots from kriging are shown in the Electronic supplementary material (Fig. S15). Linear regression (Figs. S16a and S16b) between ordinary kriging and observations has the highest R^2 and slope among all the methods. RMSEs are also very small, which are 0.67 μg/m^3 and 24 ppb, separately. Such performance is expected when using the same data in the application because of the ordinary kriging method's mechanism, so monitor-based data withholding was performed for evaluation.

The performance using monitor-based withholding for ordinary kriging is similar to data fusion results. R^2 for monitor-based withholding is larger than 0.70. Results for CO are worse than the total data interpolation; R^2 decreases from 0.99 (ordinary kriging) to 0.13 (ordinary kriging LOO) (Fig. S16b).

Methods using satellite-retrieved AOD for PM$_{2.5}$

Two-stage statistical model

The R^2 between observation and two-stage statistical model results is 0.81 (Table 3) lower than data fusion results (0.95, Table 2). The RMSE of two-stage statistical model (3.06 μg/m^3) is better than CMAQ data RMSE of 5.16 μg/m^3 when comparing simulated results with observations. A tenfold cross-validation (random data withholding) shows that the 3-year averaged R^2 is 0.78 and the averaged RMSE is 3.06 from 2006 to 2008.

Neural network-based hybrid model

The linear regression between neural network-based hybrid model results and pollutant observations has an R^2 of 0.82 (Fig. S17). The annual average spatial distribution fields

Table 4 Performance evaluation for observation (OBS) and simulations (CO), 24-h average values

	Mean (ppb)	Median (ppb)	RMSE	NME	R^2
OBS	388	342	0	0	1
CMAQ	242	221	269	0.52	0.08
Data fusion	461	421	231	0.38	0.26
DF-10% random data withholding	464	431	178	0.47	0.24
DF-leave 1-monitor out	461	426	260	0.48	0.10
Interpolation (ordinary kriging)	391	346	24	0.05	0.99
Ordinary kriging: leave 1-monitor out	394	355	164	0.45	0.13

NME normalized mean error

Table 5 Performance evaluation for observation (OBS) and simulations (NO_2), 24-h average values

	Mean (ppb)	Median (ppb)	RMSE	NME	R^2
OBS	11.0	10.1	0.0	0	1
CMAQ	7.4	6.0	7.1	0.51	0.18
Data fusion	10.3	9.2	2.4	0.15	0.81
DF-10% random data withholding	10.1	9.3	3.2	0.27	0.78
DF-leave 1-monitor out	10.3	9.3	3.8	0.27	0.52

NME normalized mean error

(Fig. S18) show a decreasing trend for $PM_{2.5}$ concentration from 2006 to 2008. The fields show that the method is also good at capturing the spatial information that urban areas have a high $PM_{2.5}$ concentration and rural areas have a lower concentration.

Comparison between CMAQ and data fusion for all species

Correlations between 10% random data withholding results and observations are higher than CMAQ and observations (Figs. S9 and S12; Table 2). R^2 values for $PM_{2.5}$, EC, OC, NH_4^+, NO_3^-, SO_4^{2-}, NO_2, NO_x, and CO between observations and data fusion simulations increase compared with the correlations between observations and CMAQ simulations. RMSEs decrease and R^2 increases for all the species except NO_3^- and NO_x. The R^2 between observation and 10% random

data withholding for $PM_{2.5}$ is 0.82. SO_4^{2-} also performs very well with a R^2 value of 0.82. R^2 value between daily CMAQ and data fusion results for each grid over the whole year for 2008 show that the highest values correspond to the grids that are nearest to monitors for all pollutants (Fig. 5). R^2 values decrease as the distance to monitors increase, which indicates that the accuracy of this method increases with the number of monitors used because of the high dependency on the number and locations of monitors to perform the kriging step in the data fusion method.

Comparison between data fusion and two-stage statistical model

The relationship between data fusion and two-stage statistical model results for $PM_{2.5}$ simulations during 2006 to 2008 are calculated using Deming regression (Deming 1943) to equally
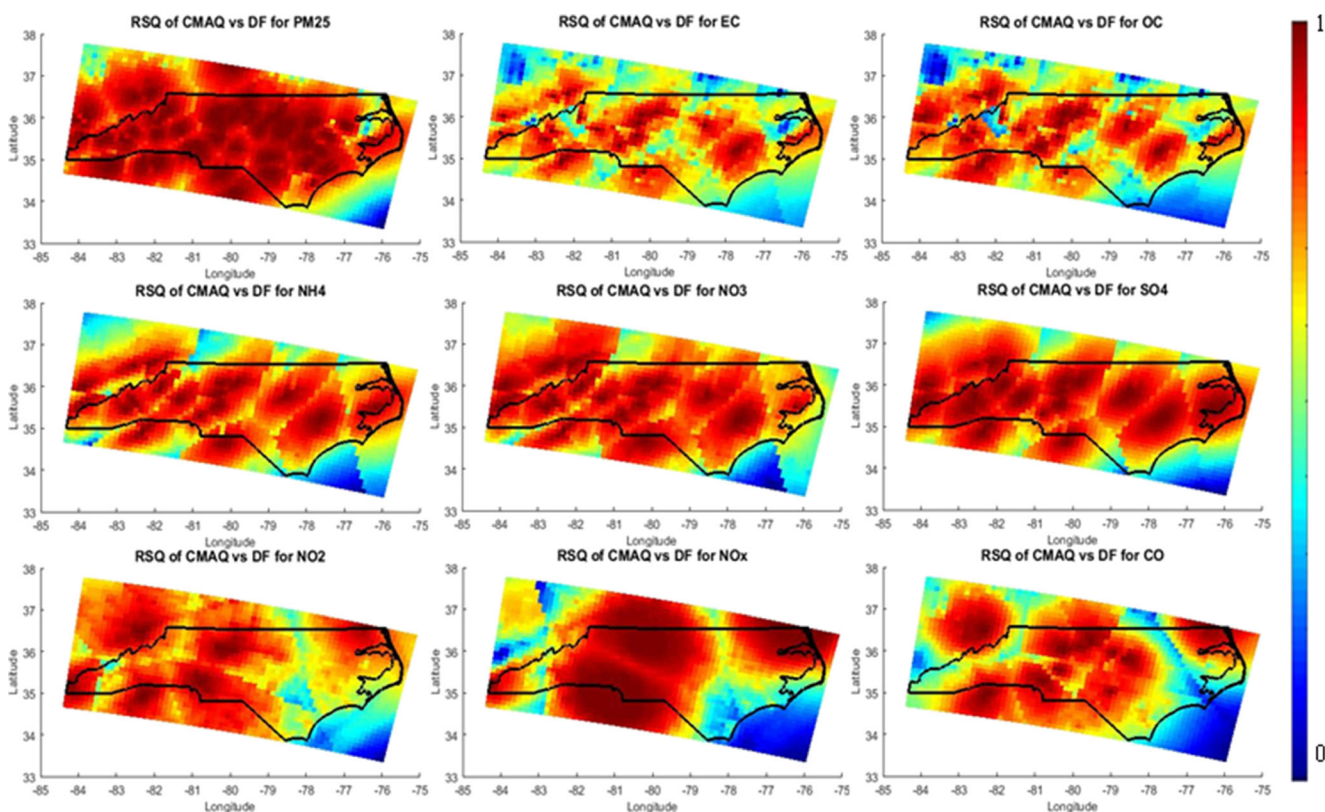
Fig. 5 R^2 values of each grid for 2008
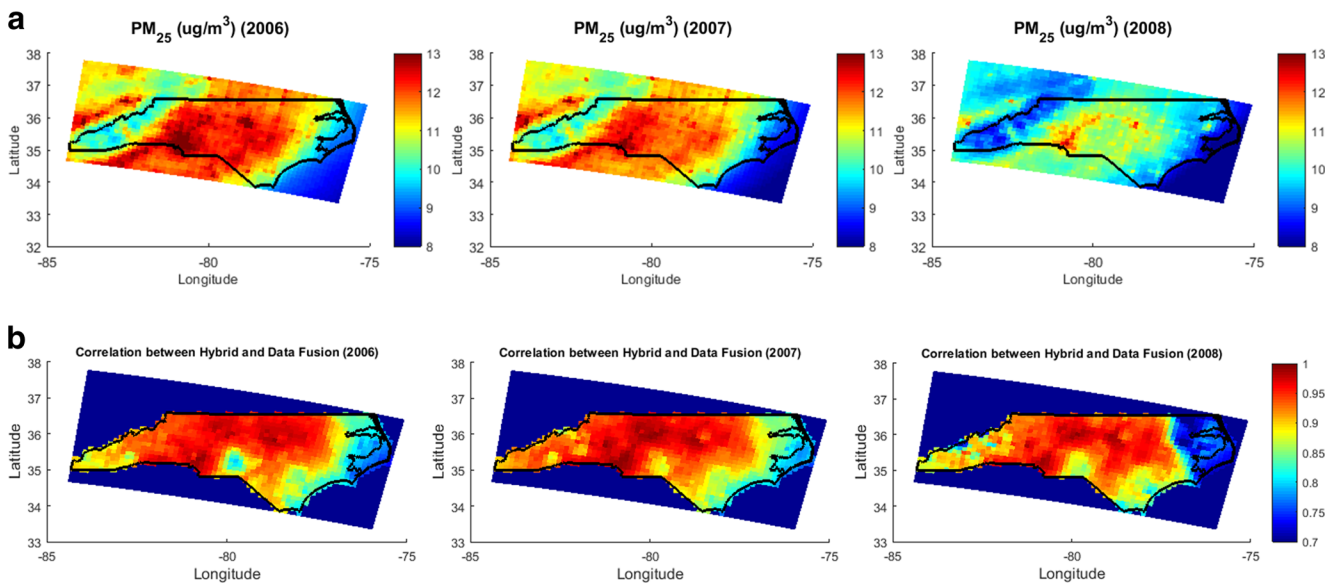
Fig. 6 **a** Temporal correlations (*R*) between data fusion and two-stage statistical model from 2006 to 2008. **b** Temporal correlations (*R*) between data fusion and Harvard's hybrid method from 2006 to 2008

weight the two inputs because both data are estimated values from models (Fig. 6a). The grid-by-grid correlations over most of the domain have a value close to 1; however, the correlations in boundary areas are lower. Both the data fusion and two-stage statistical model capture the urban area PM$_{2.5}$ concentrations. Fewer monitors are located in the forested areas of NC, so the results from the two methods are not as strongly correlated. CMAQ secondary organic carbon formation is typical biased low in forested areas (Van Donkelaar et al. 2007; Zhang et al. 2007; Baek et al. 2011), which may contribute to low correlations with the two-stage statistical model. The two-stage statistical model can overestimate concentrations in the coastal areas of eastern NC (Fig. S19) because of the high relative humidity in the area, which leads to a bias in estimated PM$_{2.5}$ from satellite-retrieved AOD (Liu et al. 2005; Hu et al. 2013). The retrieval quality of the MODIS product is sensitive to vegetation cover and has difficulty distinguishing between the mixed land and water pixels, a limitation that might also contribute to the overestimation of the two-stage model along the coast. Lacking AOD data could be another limitation of these AOD data-included methods because of the satellite pattern and cloud cover days.

Comparison between data fusion and hybrid model

Another comparison is made between the data fusion and Di et al.'s (2016) method. Temporal Deming regression (Fig. 6b) shows the higher correlation in urban areas and lower correlation in the eastern and western boundaries and mid-south areas. This is similar to the comparison of data fusion and the two-stage statistical model results except in the mid-south area, which is a national forest. The difference in annual average concentration in coastal areas (Figs. S18, S19, and

S20) illustrates that the neural network-based hybrid model could provide a more accurate spatial information because of the use of AAI and CTM outputs to improve accuracy.

Conclusion

Application of the data fusion method for primary and secondary pollutants over North Carolina demonstrates that the method provides accurate concentration fields, especially for PM$_{2.5}$ total mass, OC, SO$_4^{2-}$, NH$_4^+$, and NO$_2$, capturing the spatial and temporal variations in both gaseous and speciated particulate matter concentrations. Capturing these variations is critical for improved estimation of exposures for health studies. Cross-validation with 10% random data withholding indicates that the DF results have little bias. CMAQ-modeled, non-data fused concentration fields were subject to higher temporally and spatially varying bias and error and lower correlations. These results demonstrate that the data fusion approach, as opposed to using CTM fields directly, should be used to provide spatiotemporal exposure fields for health studies that use daily air quality metrics. Using the DF method-derived fields to estimate mobile source impacts using the IMSI method also found that the results could be used in health studies.

This study also investigated the use of random data withholding versus withholding monitors randomly and based upon spatial clustering. Findings show that the data fusion method does provide accurate fields, but random data withholding may overestimate the ability of such methods to provide accurate concentration estimates in areas lacking monitors. The number and the distribution of monitoring sites affect the accuracy of the data fusion method. The more widely the

monitors are distributed, the more stable the data fusion method results. Observation availability is an important factor in the application and evaluation of the method according to some pollutants' performances such as CO, NO_2, and NO_x have very few monitors. Moreover, CO monitors are mainly located in urban areas. However, this research and previous studies demonstrate the benefits of the method versus the use of air quality model fields directly.

Spatiotemporal $PM_{2.5}$ fields derived using the CTM-based data fusion method are compared well with similar fields derived using AOD and another chemical transport model. These and prior results suggest that the data fusion method provides a promising approach to develop exposure fields for health analysis across both urban and regional scales. A major advantage of CTM-based data fusion methods (which could potentially include the hybrid approach) over methods relying mostly on AOD to provide spatial variations is that it provides speciated $PM_{2.5}$ and gaseous pollutant fields.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

References

Baek J, Hu Y, Odman MT, Russell AG (2011) Modeling secondary organic aerosol in CMAQ using multigenerational oxidation of semivolatile organic compounds. J Geophys Res Atmos 116:D22204. https://doi.org/10.1029/2011JD015911

Beelen R, Hoek G, Pebesma E et al (2009) Mapping of background air pollution at a fine spatial scale across the European Union. Sci Total Environ 407:1852–1867. https://doi.org/10.1016/j.scitotenv.2008.11.048

Binkowski FS (2003) Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description J Geophys Res 108:4183. https://doi.org/10.1029/2001JD001409

Byun D, Schere KL (2006) Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl Mech Rev 59:51. https://doi.org/10.1115/1.2128636

Carlton AG, Turpin BJ, Altieri KE et al (2008) CMAQ model performance enhanced when in-cloud secondary organic aerosol is included: comparisons of organic carbon predictions with measurements. Environ Sci Technol 42:8798–8802. https://doi.org/10.1021/es801192n

Chu S-H (2004) PM2.5 episodes as observed in the speciation trends network. Atmos Environ 38:5237–5246. https://doi.org/10.1016/j.atmosenv.2004.01.055

Cressie N (1988) Spatial prediction and ordinary kriging. Math Geol 20:405–421. https://doi.org/10.1007/BF00892986

Deming WE (1943) Statistical adjustment of data

Di Q, Kloog I, Koutrakis P et al (2016) Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ Sci Technol 50:4712–4721. https://doi.org/10.1021/acs.est.5b06121

Dionisio KL, Baxter LK, Burke J, Özkaynak H (2016) The importance of the exposure metric in air pollution epidemiology studies: when does it matter, and why? Air Qual Atmos Heal 9:495–502. https://doi.org/10.1007/s11869-015-0356-1

Friberg MD, Zhai X, Holmes HA et al (2016) Method for fusing observational data and chemical transport model simulations to estimate spatiotemporally resolved ambient air pollution. Environ Sci Technol 50:3695–3705. https://doi.org/10.1021/acs.est.5b05134

Gertler AW (2005) Diesel vs. gasoline emissions: does PM from diesel or gasoline vehicles dominate in the US? Atmos Environ 39:2349–2355. https://doi.org/10.1016/j.atmosenv.2004.05.065

Gertler AW, Gillies JA, Pierson WR (2000) An assessment of the mobile source contribution to PM10 and PM2.5 in the United States. Water Air Soil Pollut 123:203–214. https://doi.org/10.1023/A:1005263220659

Gilboa SM, Mendola P, Olshan AF et al (2005) Relation between ambient air quality and selected birth defects, seven county study, Texas, 1997–2000. Am J Epidemiol 162:238–252. https://doi.org/10.1093/aje/kwi189

Gilliland AB, Hogrefe C, Pinder RW et al (2008) Dynamic evaluation of regional air quality models: assessing changes in O3 stemming from changes in emissions and meteorology. Atmos Environ 42:5110–5123. https://doi.org/10.1016/j.atmosenv.2008.02.018

Godowitch JM, Gilliam RC, Roselle SJ (2015) Investigating the impact on modeled ozone concentrations using meteorological fields from WRF with an updated four-dimensional data assimilation approach. Atmos Pollut Res 6:305–311. https://doi.org/10.5094/APR.2015.034

Hu X, Waller LA, Al-Hamdan MZ et al (2013) Estimating ground-level PM(2.5) concentrations in the southeastern U.S. using geographically weighted regression. Environ Res 121:1–10. https://doi.org/10.1016/j.envres.2012.11.003

Hu X, Waller LA, Lyapustin A et al (2014a) Estimating ground-level PM2.5 concentrations in the southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sens Environ 140:220–232. https://doi.org/10.1016/j.rse.2013.08.032

Hu Y, Balachandran S, Pachon JE et al (2014b) Fine particulate matter source apportionment using a hybrid chemical transport and receptor model approach. Atmos Chem Phys 14:5415–5431. https://doi.org/10.5194/acp-14-5415-2014

Hubbell B (2012) Understanding urban exposure environments: new research directions for informing implementation of U.S. air quality standards. Air Qual Atmos Heal 5:259–267. https://doi.org/10.1007/s11869-011-0153-4

Ivey CE, Holmes HA, Hu Y et al (2016) A method for quantifying bias in modeled concentrations and source impacts for secondary particulate matter. Front Environ Sci Eng 10:14. https://doi.org/10.1007/s11783-016-0866-6

Ivey CE, Holmes HA, Hu YT et al (2015) Development of PM2.5 source impact spatial fields using a hybrid source apportionment air quality model. Geosci Model Dev 8:2153–2165. https://doi.org/10.5194/gmd-8-2153-2015

Jathar SH, Cappa CD, Wexler AS et al (2016) Simulating secondary organic aerosol in a regional air quality model using the statistical oxidation model—part 1: assessing the influence of constrained

multi-generational ageing. Atmos Chem Phys 16:2309–2322. https://doi.org/10.5194/acp-16-2309-2016

Johnson M, Isakov V, Touma JS et al (2010) Evaluation of land-use regression models used to predict air quality concentrations in an urban area. Atmos Environ 44:3660–3668. https://doi.org/10.1016/j.atmosenv.2010.06.041

Kanaroglou PS, Jerrett M, Morrison J et al (2005) Establishing an air pollution monitoring network for intra-urban population exposure assessment: a location-allocation approach. Atmos Environ 39:2399–2409. https://doi.org/10.1016/j.atmosenv.2004.06.049

Kim S-Y, Yi S-J, Eum YS et al (2014) Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities. Environ Health Toxicol 29:e2014012. https://doi.org/10.5620/eht.e2014012

Kim Y-M, Zhou Y, Gao Y et al (2015) Spatially resolved estimation of ozone-related mortality in the United States under two representative concentration pathways (RCPs) and their uncertainty. Clim Chang 128:71–84. https://doi.org/10.1007/s10584-014-1290-1

Lefohn AS, Knudsen HP, Logan JA et al (1987) An evaluation of the kriging method to predict 7-h seasonal mean ozone concentrations for estimating crop losses. JAPCA 37:595–602. https://doi.org/10.1080/08940630.1987.10466247

Liu Y, Koutrakis P, Kahn R et al (2012) Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 2—a case study. J Air Waste Manage Assoc 57:1360–1369

Liu Y, Sarnat JA, Kilaru V et al (2005) Estimating ground-level $PM_{2.5}$ in the eastern United States using satellite remote sensing. Environ Sci Technol 39:3269–3278. https://doi.org/10.1021/es049352m

Malm WC, Sisler JF, Huffman D et al (1994) Spatial and seasonal trends in particle concentration and optical extinction in the United States. J Geophys Res 99:1347. https://doi.org/10.1029/93JD02916

Marmur A, Unal A, Mulholland JA, Russell AG (2005) Optimization-based source apportionment of $PM_{2.5}$ incorporating gas-to-particle ratios. Environ Sci Technol 39:3245–3254. https://doi.org/10.1021/es0490121

Matte TD, Cohen A, Dimmick F et al (2009) Summary of the workshop on methodologies for environmental public health tracking of air pollution effects. Air Qual Atmos Health 2:177–184. https://doi.org/10.1007/s11869-009-0059-6

McGuinn LA, Ward-Caviness C, Neas LM et al (2017) Fine particulate matter and cardiovascular disease: comparison of assessment methods for long-term exposure. Environ Res 159:16–23. https://doi.org/10.1016/j.envres.2017.07.041

Pachon JE, Balachandran S, Hu Y et al (2012) Development of outcome-based, multipollutant mobile source indicators. J Air Waste Manage Assoc 62:431–442. https://doi.org/10.1080/10473289.2012.656218

Pleim J, Gilliam R, Appel W, Ran L (2016) Recent advances in modeling of the atmospheric boundary layer and land surface in the coupled WRF-CMAQ model. Springer International Publishing, pp 391–396

Pope CA, Ezzati M, Dockery DW (2009) Fine-particulate air pollution and life expectancy in the United States. N Engl J Med 360:376–386. https://doi.org/10.1056/NEJMsa0805646

Qin M, Wang X, Hu Y et al (2015) Formation of particulate sulfate and nitrate over the Pearl River Delta in the fall: diagnostic analysis using the Community Multiscale Air Quality model. Atmos Environ 112:81–89. https://doi.org/10.1016/j.atmosenv.2015.04.027

Sampson PD, Richards M, Szpiro AA et al (2013) A regionalized national universal kriging model using partial least squares regression for estimating annual $PM_{2.5}$ concentrations in epidemiology. Atmos Environ (1994) 75:383–392. https://doi.org/10.1016/j.atmosenv.2013.04.015

Sarnat SE, Coull BA, Schwartz J et al (2005) Factors affecting the association between ambient concentrations and personal exposures to particles and gases. Environ Health Perspect 114:649–654. https://doi.org/10.1289/ehp.8422

Solomon PA, Costantini M, Grahame TJ et al (2012) Air pollution and health: bridging the gap from sources to health outcomes: conference summary. Air Qual Atmos Heal 5:9–62. https://doi.org/10.1007/S11869-011-0161-4

Tang W, Cohan DS, Morris GA et al (2011) Influence of vertical mixing uncertainties on ozone simulation in CMAQ. Atmos Environ 45:2898–2909. https://doi.org/10.1016/j.atmosenv.2011.01.057

Van Donkelaar A, Martin RV, Park RJ et al (2007) Model evidence for a significant source of secondary organic aerosol from isoprene. Atmos Environ 41:1267–1274. https://doi.org/10.1016/j.atmosenv.2006.09.051

Wade KS, Mulholland JA, Marmur A et al (2006) Effects of instrument precision and spatial variability on the assessment of the temporal variation of ambient air pollution in Atlanta, Georgia. J Air Waste Manage Assoc 56:876–888. https://doi.org/10.1080/10473289.2006.10464499

Woody MC, Baker KR, Hayes PL et al (2016) Understanding sources of organic aerosol during CalNex-2010 using the CMAQ-VBS. Atmos Chem Phys 16:4081–4100. https://doi.org/10.5194/acp-16-4081-2016

Wyat Appel K, Bhave PV, Gilliland AB et al (2008) Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance; part II—particulate matter. Atmos Environ 42:6057–6066. https://doi.org/10.1016/j.atmosenv.2008.03.036

Xiao X, Cohan DS, Byun DW, Ngan F (2010) Highly nonlinear ozone formation in the Houston region and implications for emission controls. J Geophys Res 115:D23309. https://doi.org/10.1029/2010JD014435

Yu S, Mathur R, Pleim J et al (2012) Comparative evaluation of the impact of WRF/NMM and WRF/ARW meteorology on CMAQ simulations for $PM_{2.5}$ and its related precursors during the 2006 TexAQS/GoMACCS study. Atmos Chem Phys 12:4091–4106. https://doi.org/10.5194/acp-12-4091-2012

Zhang Y, Huang J-P, Henze DK, Seinfeld JH (2007) Role of isoprene in secondary organic aerosol formation on a regional scale. J Geophys Res 112:D20207. https://doi.org/10.1029/2007JD008675