Environmental Science & Technology

pubs.acs.org/est

Article

Evaluation of Bayesian Maximum Entropy Data Fusion Approaches to Estimate Styrene, Benzene, Toluene, Ethylbenzene, and Xylenes and to Inform Epidemiological Analyses in the US Gulf States

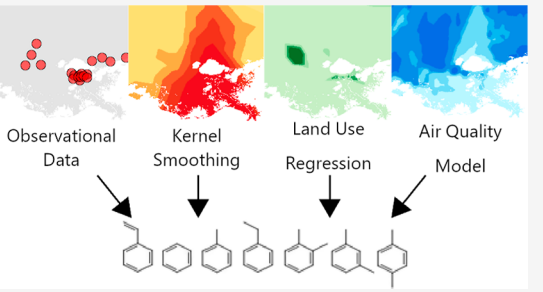
Nora A. Abbott, Lucie Semone, Richard Strott, Praful Dodda, Chi-Tsan Wang, Jaime Green, Bok Haeng Baek, Lawrence S. Engel, William Vizuete, and Marc L. Serre*

Cite This: *Environ. Sci. Technol.* 2025, 59, 45−55

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The Gulf States are home to industries emitting styrene, benzene, toluene, ethylbenzene, and xylenes (SBTEX). Presently, adverse health effects of ambient SBTEX exposure in highly polluted regions, such as the Gulf States, must be evaluated. Epidemiologists, however, are limited by inadequate estimates of ambient SBTEX. Using Bayesian Maximum Entropy, SBTEX estimation methods of varying resource intensity were evaluated, including simple kriging (least intense), incorporation of observational and emissions data trends (moderately intense), and data fusion of observed and Comprehensive Air quality Model with extensions (CAMx) data (most intense). Generally, as resource intensity increased, so did SBTEX estimation performance, where SBTEX Spearman *R* values increased by 0.48 on average from the least to most intense methods. Data fusion of observed and CAMx data was identified as the best ambient SBTEX estimation method in the Gulf States. Exposure estimates revealed that Gulf States residences within commuting distance of high industrial activity experienced 1.64 times higher 97.5th percentile daily exposures to SBTEX on average than those living in less industrialized areas, which could contribute to total occupational and ambient exposure disparities. Furthermore, ambient benzene exposure was greater than the acceptable one-in-a-million excess cancer risk threshold for 75% of estimated residence locations in the Gulf States.



KEYWORDS: *Bayesian Maximum Entropy, data fusion, styrene, BTEX, air pollution, modeled data, exposure*

1. INTRODUCTION

Styrene, benzene, toluene, ethylbenzene, and xylenes (SBTEX) are a group of volatile organic compounds (VOCs) that are primarily emitted by anthropogenic sources, such as petroleum, chemical, and manufacturing industries, vehicular exhaust, and construction and are known or suspected occupational neurotoxicants.[1−4] Recent epidemiological research identified elevated blood SBTEX levels among some Texas (TX), Louisiana (LA), Mississippi (MS), Alabama (AL), and Florida (FL) (hereafter referred to as the Gulf States) residents that, given the short biological half-lives of SBTEX, indicated recent exposures to ambient SBTEX.[3,5] However, epidemiological analyses of adverse health outcomes related to ambient SBTEX exposures are limited by the coarse spatial and temporal resolution of SBTEX monitoring data.[4,6,7]

As non-criteria pollutants,[8] SBTEX is primarily monitored for regulatory purposes near known emission sources, with monitors located sparsely throughout the United States (US).[9−11] The nature of SBTEX monitoring poses several challenges for exposure assessment, including limited historical records,[12−14] nonrepresentative observational data for general exposures,[9,15] the inability to capture small-scale variability of SBTEX,[7,16,17] and difficulty interpolating measurements over long distances due to the SBTEX's high reactivity in the atmosphere (i.e., average half-life of 11 h to 6.5 days).[1,9,17−22] Additional SBTEX monitor deployment is often prohibited by high operational costs, demanding alternative, nonobservational methods for SBTEX estimation, and to inform epidemiological studies.[9,12]

VOC estimates away from monitoring stations have been improved through alternative methods, including land use regression (LUR),[2,23−26] kriging,[9,27] low-cost sensors,[28] satellites,[29] and air quality models.[1,30] Alternative methods, however, have a variety of limitations, including a restricted spatial and temporal resolution and reliance on sparse monitoring data for LUR and kriging,[2,9,24−27] difficulty capturing seasonality for LUR,[31] quantitative inaccuracies for

Received: June 11, 2024
Revised: December 9, 2024
Accepted: December 10, 2024
Published: January 3, 2025

low-cost sensors,[28] an inability to directly measure species-specific concentrations for satellites,[29] and nonlinear performance and model bias for air quality models.[1,30,32−34]

Bayesian Maximum Entropy (BME) is an ideal technique to address SBTEX estimation method limitations because BME accounts for global trends in observational and emissions data,[32,33,35] local emissions,[32,33] and model bias,[32−34,34,36,37] and can estimate over larger spatial and temporal domains.[38] Previous applications of BME to similar air pollutants increased the Spearman R of daily ozone (O_3) by 0.04 on average compared to kriging,[33] and increased the Pearson R of nitrogen dioxide (NO_2) by 0.11 compared to an air quality model.[39] Furthermore, BME has improved NO_2 and O_3 estimated away from monitoring stations[32,33,37,40] and informed exposure assessments when observational data alone was insufficient.[37,39]

The present study uses BME data fusion to (1) evaluate the performance of SBTEX estimation methods with implications for epidemiological study design and (2) select the best methods to estimate SBTEX in the Gulf States from 2011 to 2012. Estimation methods of varying resource intensity evaluated in this work include simple kriging (least intense), kernel smoothing (KS)[32,33,41] and LUR[42,43] informed global offsets (moderately intense), and air quality model data fusion (most intense). To the authors' knowledge, this is the first study to create SBTEX estimates on a regional scale and in the Gulf States and to evaluate the performance effects of global offset selection and data fusion methods.

2. MATERIALS AND METHODS

2.1. Observational Data. Subdaily (i.e., 1, 3, or 4 h duration) styrene, benzene, toluene, ethylbenzene, *m*/*p*-xylenes, and *o*-xylene measurements from 183 monitoring stations were downloaded from the Environmental Protection Agency (EPA) Air Quality System (AQS) database for the Gulf States and Georgia (GA) in 2006 to 2017. These observation data were assigned a measurement error based on their measurement federal method detection limit (MDL). Subdaily data were averaged to produce daily estimates. To reflect the ubiquitous nature of environmental VOCs,[44] measured-zero daily values were assigned the mean and variance of log−normal distribution truncated above the federal MDL. Data from collocated monitors were averaged to account for the incompleteness of subdaily observations and measurement error, yielding unique, daily SBTEX concentrations at each monitoring site (see Supporting Information for details).

An exploratory analysis revealed that using the natural logarithm of measurements, as opposed to untransformed data, enhanced the visualization of the distribution of the data and the spatial and temporal trends (see Supporting Information, Figures S1 and S2, for details). This indicated that performing a natural logarithm transformation of the data before BME data fusion would improve the estimation performance.

2.2. Land Use Regression. To evaluate the usefulness of emissions data trends for predicting SBTEX, a LUR model was created to predict annual concentrations of each SBTEX component from 2008 to 2017 in the Gulf States. SBTEX emissions data was downloaded from the EPA's National Emissions Inventory (NEI), and traffic data was downloaded from the Federal Highway Administrations (FHWA) Highway Performance Monitoring System.

Three LUR predictor variables were created for point source and nonpoint source emissions and on-road traffic using the Sum of Exponentially Decaying Contributions (SEDC) method, which assumes pollution from a source will experience a 95% (exponential) decay over a distance, α.[42,45,46] The Find, Inform, and Test (FIT) method[45,46] was used to determine the optimal α for each SEDC predictor based on a positive and maximum Pearson and Spearman R, and p-value <0.05. No multicollinearity (i.e., variance inflation factor <10) was identified between the three SBTEX LUR predictor variables. Forward stepwise regression, with a priori positive relationship between the SEDC predictor and annual SBTEX concentration was performed using the Akaike Information Criterion and p-value <0.05.

LUR results differed across SBTEX components and were primarily predictive on a coarser scale than CAMx. Predictor data, processing, and results are described in the Supporting Information, and the final SBTEX LUR equations and model performance from 2011 to 2012 are shown in Table S1.

2.3. Data from the Air Quality Model. This work continues prior research that developed an air quality model to improve high-resolution SBTEX estimation for health risk assessment in the Gulf States.[1] Hazardous air pollutant (HAP) reporting (except benzene) to the NEI is voluntary, so air quality models constructed using NEI emissions typically have incomplete input data.[1] In earlier work by the authors, SBTEX was modeled using Comprehensive Air quality Model with extensions (CAMx); meteorological data from the Weather Research and Forecasting Model (WRF); and styrene, toluene, ethylbenzene, and xylenes (STEX) emissions not reported to the NEI were estimated using the novel HAPs Imputation (HAPI).[1,12] Hourly SBTEX concentrations were estimated on a horizontal 4 km-by-4 km grid resolution in the Gulf States, capturing spatiotemporal variability, seasonality, chemical reactions, transport, deposition, and more complete emissions.[1] Results suggested that STEX emissions were underestimated by 2−22% in the 2012 NEI, with missing emissions primarily from oil and gas nonpoint and nonelectric power plant point sources.[1] CAMx (+HAPI) estimates were limited by emissions data accuracy, model simulation mechanisms, and the exclusion of observed data, which we seek to address through BME data fusion.[1]

Daily SBTEX CAMx estimates for 2011−2012 were calculated as the arithmetic average of the hourly CAMx (+HAPI) estimates. The CAMx output was lognormally distributed based on exploratory analyses, so a natural logarithm transformation was applied (see Figure S3 for details). CAMx zero-values were omitted because they appeared to be a boundary artifact based on exploratory analyses.

2.4. Constant Air Model Performance Correction. Daily CAMx outputs were calibrated for BME data fusion using the Constant Air Model Performance (CAMP) correction method for 2011 and 2012 separately. The CAMP method is a stationary, nonlinear, nonhomoscedastic correction for model bias performed by pairing observations and model outputs.[32−34,36,37,40]

Daily observed SBTEX log-concentrations inside and within 0.05° of the CAMx-modeled region were paired with the closest CAMx model output. Observed-modeled pairs were sorted in ascending modeled values and divided into deciles of model values. The mean (λ_1) and variance (λ_2) of observed log-concentrations within each model decile were calculated to

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

create a piecewise-linear λ_1 curve used for model correction, and a λ_2 curve providing the associated uncertainty. The end points of the curves were extrapolated using the average slope of the piecewise linear segments, and segments of the λ_1 curve with negative slopes were replaced with zero slope segments to ensure the λ_1 curve remained monotonic and positive between observed and CAMx-modeled SBTEX.[37] Lastly, modeled values were interpolated along the λ_1 and λ_2 curves to create CAMP-corrected CAMx mean, $\mathbf{ym}_s$, and variance, $\mathbf{yv}_s$, values, respectively. Detailed CAMP curves are shown in Figures S4−S8 of Supporting Information.

2.5. BME Data Fusion with Global Offsets. BME is an advanced nonlinear method of modern geostatistics[38,47] that has been used with global offsets to successfully map many pollutants;[32,35,38,47,48] however, SBTEX has not previously been evaluated. Hence, a critical goal of this work is to implement BME with global offsets for each VOC species.

Let $\boldsymbol{p} = (\boldsymbol{s},t)$ be a space/time location, where $\boldsymbol{s} = [s_1,s_2]$ is a 2D spatial location and t is time. To implement BME with various global offsets, we generalize the space-time BME data fusion framework as follows:

1. Define a spatiotemporal link function $g(.;\boldsymbol{p})$ to transform the VOC z-data observed at $\boldsymbol{p}$ into the transformed data $x = g(z;\boldsymbol{p})$,
2. Perform a BME data fusion of the transformed hard and soft x-data to estimate the distribution of the random variable X_k, representing the transformed x-value at some estimation point $\boldsymbol{p}_k$, and
3. Use the back transform $g^{-1}(.;\boldsymbol{p}_k)$ to obtain the back-transformed random variable $Z_k = g^{-1}(X_k;\boldsymbol{p}_k)$ representing VOC concentration at estimation point $\boldsymbol{p}_k$.

In step 1, we consider a link function that log-transforms the z-data and removes a global offset capturing global spatial and temporal trends in log concentrations at some resolution δ, i.e.

$$x = g(z;\boldsymbol{p}) = \log(z) - o_\delta(\boldsymbol{p}) \tag{1}$$

$$o_\delta(\boldsymbol{p}) = o_{\delta,s}(\boldsymbol{s}) + o_{\delta,t}(t) - \overline{o_{\delta,t}(t)} \tag{2}$$

where $log(.)$ is the natural log, $o_\delta(\boldsymbol{p})$ is the global offset of log-concentrations at resolution δ, $o_{\delta,s}(\boldsymbol{s})$ is its spatial component with an average $\overline{o_{\delta,s}(\boldsymbol{s})}$ equal to that of log concentrations, and $o_{\delta,t}(t) - \overline{o_{\delta,t}(t)}$ is its temporal component on average equal to zero.[33] In this work, we aim to evaluate the performance of three global offsets at the following resolutions: flat (or constant-value), medium, and fine.

The flat global offsets were calculated as the mean of the observed log-concentrations across the study domain. The temporal component, $o_{\delta,t}(t)$, was calculated via exponential KS (described in Akita et al., 2007) of spatially averaged, observed log-concentrations using a KS range of 365−730 days for the medium resolution, and 90−180 days for the fine resolution. The spatial component of the global offset, $o_{\delta,s}(\boldsymbol{s})$, was obtained using LUR (described earlier) for benzene and, for the other VOCs, via exponential KS of time-averaged, observed log-concentrations using a KS range of 2.5−7.5° for the medium resolution and 0.2−0.6° for the fine resolution. Details are described in the Supporting Information (Tables S2−S6 and Figures S9−S14). Critically, as the global offset resolution progresses from flat to fine, it becomes informed, and the link function removes z-data variability at progressively finer scales, reducing residual variability of the transformed x-data.

In step 2, we perform a BME data fusion[38,47] of hard and soft data transformed using the spatiotemporal link function (eqs 1 and 2). For a given VOC and resolution δ, we obtain the transformed hard data, $\boldsymbol{x}_h$, using the equation $\boldsymbol{x}_h = \log(\boldsymbol{z}_h) - o_\delta(\boldsymbol{p}_h)$, where $\boldsymbol{z}_h$ is the vector of observed values at hard data points $\boldsymbol{p}_h$.[33] The soft data consists of the means, $\boldsymbol{ym}_s$, and variances, $\boldsymbol{yv}_s$, of the CAMP-corrected CAMx log-concentrations at each CAMx computational points, $\boldsymbol{p}_s$. The corresponding transformed soft x-data consists of Gaussian distributions, $f_S(\boldsymbol{x}_s)$, at points, $\boldsymbol{p}_s$, with means $\boldsymbol{xm}_s = \boldsymbol{ym}_s - o_\delta(\boldsymbol{p}_s)$ and variances $\boldsymbol{xv}_s = \boldsymbol{yv}_s$.[33]

We model the residual variability and associated autocorrelation of $\boldsymbol{x}_h$ using a homogeneous/stationary space/time random field (S/TRF), $X(\boldsymbol{p})$, with a constant mean $m_X = E[X(\boldsymbol{p})]$ and a covariance between any two points $\boldsymbol{p}$ and $\boldsymbol{p}'$ expressed as

$$c_X(r,\tau) = E[X(\boldsymbol{p})X(\boldsymbol{p}')] - m_X^2 \tag{3}$$

where $E[.]$ is the expectation operator and $r = \boldsymbol{s} - \boldsymbol{s}'$ and $\tau = |t - t'|$ are the spatial and temporal lags, respectively, separating $\boldsymbol{p}$ and $\boldsymbol{p}'$.[33] The covariance was manually modeled as a three-structure exponential model described in detail in the Supporting Information (Tables S7−S11 and Figure S14). The covariance models contained long and short structures, which were driven by steady emissions over time and the short half-life (or quick decay) of SBTEX.[1,18−22] Generally, $\boldsymbol{x}_h$ derived from global offsets with finer resolution produced lower residual variability.

The BME data fusion of the hard data and soft distributions is performed in two stages. In the first, Maximum Entropy (ME), stage, we construct a ME probability density function (PDF) $f_G(x_k,\boldsymbol{x}_h,\boldsymbol{x}_s)$, where $\boldsymbol{x}_{map} = (x_k,\boldsymbol{x}_h,\boldsymbol{x}_s)$ is the value of $X(\boldsymbol{p})$ at $\boldsymbol{p}_{map} = (\boldsymbol{p}_k,\boldsymbol{p}_h,\boldsymbol{p}_s)$, based on general knowledge $G = \{m_X, c_X(.)\}$.[32] In the second, Bayesian (B) knowledge blending, stage, we integrate site-specific knowledge $S = \{\boldsymbol{x}_h, f_s(\boldsymbol{x}_s)\}$ and obtain the B-ME PDF under knowledge $K = \{G,S\}$

$$f_K(x_k) = A^{-1}\int d\boldsymbol{x} f_G(x_k,\boldsymbol{x}_h,\boldsymbol{x}_s) f_S(\boldsymbol{x}_s) \tag{4}$$

where A is the normalization constant.[32] Lastly, we calculate the expected value and variance of $f_K(x_k)$ to obtain the mean xm_k and variance xv_k of X_k.

In step 3, we evaluate $Z_k = g^{-1}(X_k;\boldsymbol{p}_k) = \exp(X_k + o_\delta(\boldsymbol{p}_k))$. Since Z_k represents VOC concentration, it is convenient to define $Y_k = X_k + o_\delta(\boldsymbol{p}_k)$ so that $Z_k = \exp(Y_k)$. Taking the log, we get $Y_k = \log(Z_k)$, indicating that Y_k represents log-concentrations. Since $Y_k = X_k + o_\delta(\boldsymbol{p}_k)$, then its mean and variance are $ym_k = xm_k + o_\delta(\boldsymbol{p}_k)$ and $yv_k = xv_k$. The BME PDF (eq 4) is Gaussian since the soft data is Gaussian, as a result $Y_k \sim$ Gaussian(ym_k,yv_k), and it follows that $Z_k = \exp(Y_k)$ has a median, mean, and variance given by

$$z_{median,k} = \exp(ym_k) \tag{5}$$

$$z_{mean,k} = \exp\left(ym_k + \frac{yv_k}{2}\right) \tag{6}$$

$$z_{variance,k} = [\exp(yv_k) - 1][\exp(2ym_k + yv_k)] \tag{7}$$

In this work, MATLAB and the BMElib package,[47] which provides a computational implementation of the theoretical BME framework, were used to perform SBTEX estimation. A flowchart for the BME data fusion with global offsets methodology (Figure S15) and a detailed description of how

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

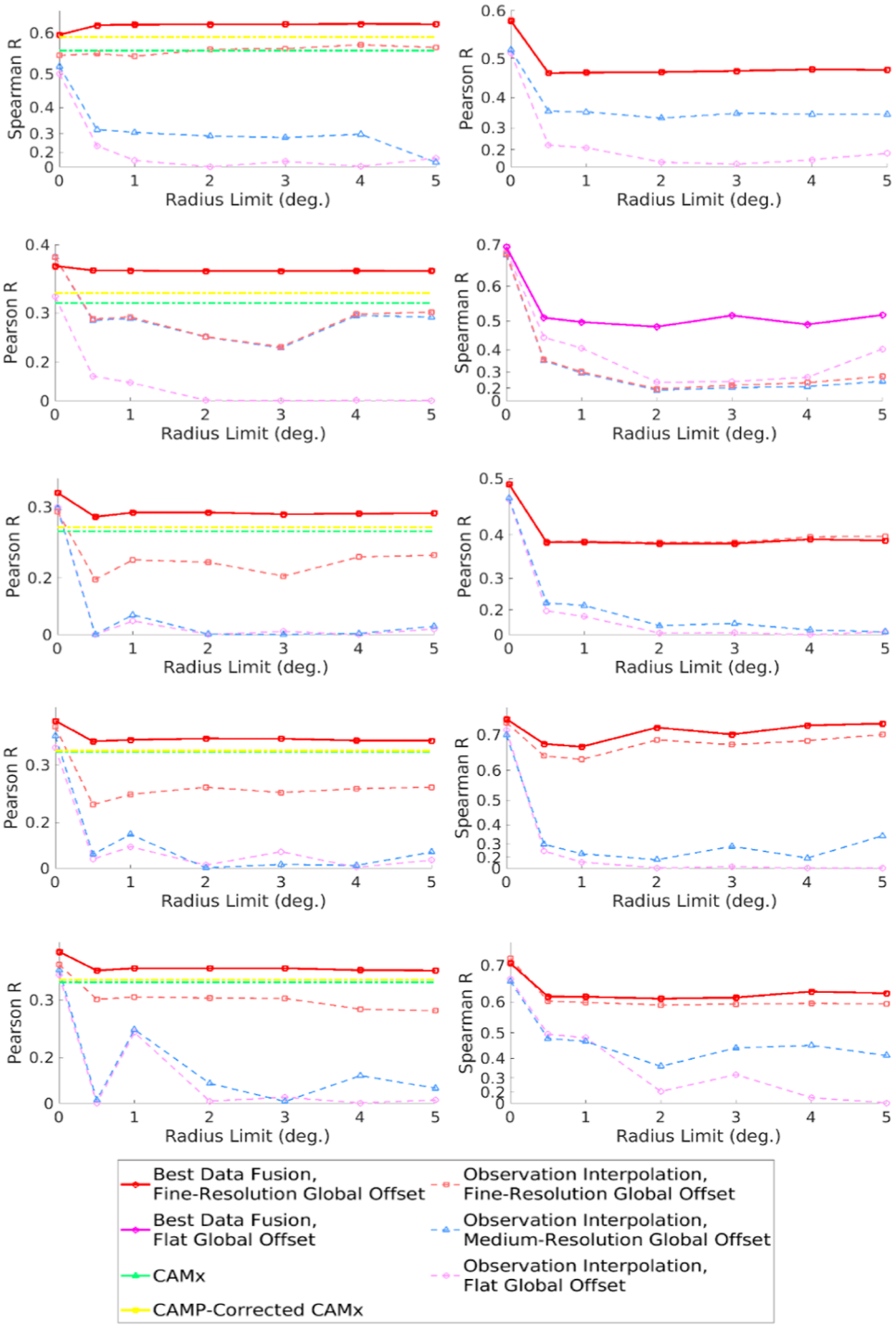


Figure 1. Radius cross-validation (RCV) curves for observation interpolation methods at three resolutions (flat, medium, and fine), CAMx, CAMP-corrected CAMx, and the best data fusion method. Note that the best data fusion method changes color to match the selected global offset resolution, where a solid magenta indicates a flat global offset for data fusion and a solid red indicates a fine-resolution global offset for data fusion. *Y*-axis scaled to the 0.5 power. Left column: RCV curves calculated within the CAMx-modeled domain. Performance statistic used for benzene, toluene, ethylbenzene, and xylenes was the Pearson *R*. Performance statistic used for styrene was the Spearman *R* because the average Pearson *R* values were less than 50% of the average Spearman *R* values. Right panel: RCV curves for method selection outside of the CAMx-modeled domain. Performance statistic used for styrene, toluene, and ethylbenzene was Pearson *R*. Performance statistic used for benzene and ethylbenzene was Spearman *R* because the Pearson *R* curves did not monotonically decrease. Performance statistic used for xylenes was the Spearman *R* because the average Pearson *R* values were less than 50% of the average Spearman *R* values.

each BME step contributes to SBTEX estimates are in the Supporting Information (Figures S16–S23).

2.6. Model Validation. SBTEX estimation methods were evaluated away from monitoring stations using radius cross-validation (RCV), which consists for a given radius, of removing observations at each station one at a time and estimating these observations using only data from stations that are at least one radius distance away from the station in

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

Table 1. CAMx, Observational Interpolation Using Global Offsets, and Selected Data Fusion Method Spearman *R* Performance to Estimate Observed SBTEX Inside of the CAMx-Modeled Domain[a]

species name	CAMx Spearman *R*	average RCV Spearman *R*				LOOCV Spearman *R*
		observational interpolation			selected data fusion method	selected data fusion method
		flat global offset	medium-resolution global offset	fine-resolution global offset		
styrene	0.559	0.132	0.303	0.559	0.616	0.757
benzene	0.628	0.259	0.462	0.474	0.651	0.772
toluene	0.611	0.148	0.186	0.499	0.647	0.749
ethylbenzene	0.554	0.102	0.190	0.510	0.598	0.735
xylenes	0.617	0.101	0.269	0.572	0.643	0.752

[a]Observational interpolation and selected data fusion methods were evaluated using the average radius cross-validation (RCV). Selected data fusion methods were further evaluated using leave-one-out cross-validation (LOOCV). Performance was measured using median SBTEX estimates.

consideration.[37,41] Performance for a radius is evaluated using the Pearson or Spearman correlation R between observed values and estimated values. For each method, we obtained the full RCV curve and displayed the Pearson R if its curve decreases monotonically and is on average at least half the Spearman R; otherwise, we displayed the more robust Spearman R. 'Best (SBTEX) estimation methods' were selected based on the highest *R*-value at a cross-validation radius of 3°. For estimation methods that use only observations, the *R*-values are expected to drop substantially over small cross-validation radii and then slowly decay over larger distances.[37,41] More detail on RCV is available in the Supporting Information (Figure S24).

We used RCV to compare the performance of only observational data (i.e., observational interpolation) using (1) flat (i.e., simple kriging; least intense), (2) medium-resolution (moderately intense), and (3) fine-resolution (moderately intense) global offset, and observational data, and CAMx predictions (i.e., BME data fusion) with a global offset (most intense). We evaluated BME data fusion (i.e., we used both observations and CAMx output) using the best of three global offset subtypes. Method intensities were based on data and computational and temporal demands. Details on global offsets and their subtypes (Tables S2−S6, S13 and Figures S9−S13) and method intensities (Table S12) are described in the Supporting Information.

When calculating the *R*-values and determining the best estimation methods, we divided these data into two geographical domains: within the CAMx 4 km modeled domain[1] and the outside CAMx domain. For the inside CAMx domain, we calculated the *R*-value between observed and CAMx-predicted values and displayed them as a flat line. For the outside domain, we could not calculate an R-value for CAMx since CAMx did not predict outside of the CAMx domain and selected data fusion as the best estimation method if it did not diminish observation interpolation performance.

Composite methods to estimate SBTEX concentration in the Gulf States were created from the best estimation methods (1) within and (2) outside the CAMx-modeled domain. Best estimation and composite methods were also evaluated using the Pearson and Spearman R values from averaged RCV curves and leave-one-out cross-validation (LOOCV), which measured estimation performance at a monitoring station location when historical measurement data at that location was available. Full validation and evaluation results are available in the Supporting Information (Tables S14−S16 and Figures S25−S39).

2.7. GuLF Study Exposure Analysis. All residences for GuLF Study cohort members within the CAMx-modeling domain were gathered (n = 33,894). On average, there were 1.62 residence locations per participant enrolled because some participants moved during the study period. Using the composite SBTEX estimation methods, daily SBTEX from 2011 to 2012 (731 days) was estimated at selected points within the CAMx-modeled domain (30 × 15 regular grid points, monitoring station locations, and Voronoi points). Daily SBTEX estimates at residence locations were approximated by interpolating grid estimates to residence locations.

A 2003 survey from the US Department of Transportation found that over 50% of commuters traveled 10 miles or less to work.[49] Based on this assumed commuting radius, at each residence, the number of NEI-reported point sources (preparation described in Section 3.2 and Supporting Information) active between 2011 and 2012 and within 0.15° of each residence location was calculated. Residence locations that had a greater number of point sources than the 75th percentile number of point sources were said to have "high source density", corresponding to a larger emitting industry presence. Residence locations with an equal or fewer number of point sources were said to have "low source density" or low industry presence.

The 2.5th, 50th, and 97.5th percentile daily exposures were calculated for SBTEX for all residence locations, high source density residence locations, and low source density residence locations within the GuLF Study cohort.

3. RESULTS AND DISCUSSION

3.1. Within the CAMx-Modeled Domain. Interpolation of SBTEX observations using global offsets performed in the expected order within the CAMx-modeled domain. The flat global offset had the worst performance, the medium-resolution offset performed slightly better, and the fine-resolution offset performed best (Figure 1). Cross-validation R-values for interpolation of SBTEX observations using a flat global offset decreased quickly as the radius limit increased to 0.5° (Figure 1), demonstrating that observational interpolation using a flat global offset (i.e., simple kriging) can only reliably be used to estimate SBTEX well when close to monitoring stations. Improvements from flat to fine resolution were surprisingly noticeable, however, with SBTEX global offsets created using a fine-resolution global offset maintained consistent performances over larger radius limits (Figure 1), suggesting that historical SBTEX measurements and emissions were representative of 2011 and 2012 observations. Compared to observational interpolation with a flat global offset, medium- and fine-resolution global offsets performed better, demonstrating that moderately intense methods are a useful tool for

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

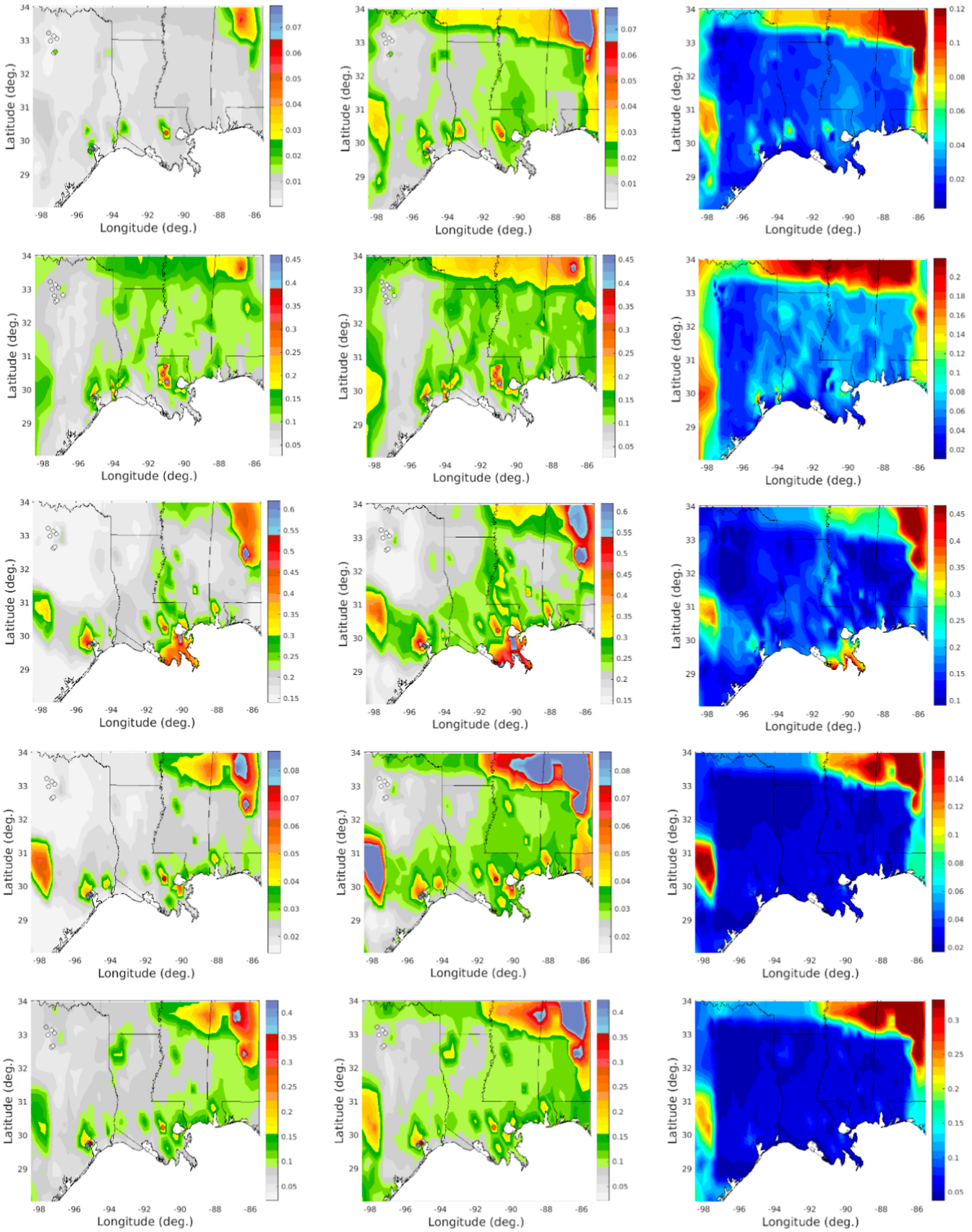


Figure 2. SBTEX estimates and estimation standard deviation (SD) near the CAMx-modeled region on August 4, 2012. From top to bottom, row 1: styrene; row 2: benzene; row 3: toluene; row 4: ethylbenzene; and row 5: xylenes. Left column: median concentration estimates (part per billion, ppb); center column: mean concentration estimates (ppb); and right column: estimation SD (ppb).

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

SBTEX estimation in monitoring station-scarce when most resource-intensive methods are not possible.

CAMx generally estimated SBTEX better than least- and moderate-intensity estimation methods within the CAMx-modeled domain (Figure 1). CAMP correction of CAMx improved the SBTEX estimation performance within the CAMx-modeled domain (Figure 1). Based on Figure 1, a fine-resolution global offset + CAMx data fusion was selected as the best estimation method within the CAMx-modeled domain for all SBTEX. Surprisingly, the data fusion approach improved upon both observational interpolation and CAMx, indicating that it is more than just the best of the two. The data fusion curve for styrene demonstrated improving performance as the radius limit increased from 0° to a radius limit of 0.5°, which may be attributed to the high variability in styrene measurements (Figure 1).

Compared to CAMx performance alone, the selected estimation method within the CAMx-modeled domain yielded 6.1−15.5% increases in average RCV Pearson R values of BTEX, but a 10.6% decrease for styrene and 3.8−10.1% increases in Spearman R for SBTEX (Tables 1, S14). SBTEX Spearman R values increased by 0.48 on average from the least to most intense methods (Table 1). The LOOCV performance using selected estimation methods was 32.2 to 51.8% higher for Pearson *R* and 21.9 to 31.5% higher for Spearman *R* than the respective CAMx Pearson or Spearman *R* values for all SBTEX (Tables 1, S14). Validation results suggest that the most resource-intensive estimation methods always improved rank estimation of SBTEX exposures within the CAMx-modeled domain.

Compared to previous BME estimation, this work (1) outperformed previous BME rank performance improvements compared to kriging (e.g., average Spearman *R* increase of 0.04 (over RCV radii of 0−108 km) for daily O_3 versus 0.39−0.54 for SBTEX),[33] (2) mostly outperformed previous BME linear performance improvements compared to an air quality model (e.g., Pearson *R* increase of 0.11 for NO_2 versus 0.10−0.14 for SBTEX),[39] and (3) similarly improved estimation away from monitoring stations.[32,33,37,40] This work's improved success can be attributed to the CAMP correction of CAMx data and the incorporation of observational and emissions trends in global offsets.

3.2. Outside of the CAMx-Modeled Domain. Outside of the CAMx-modeled domain, interpolation of STEX observations using global offsets performed as expected, where the performance from worst to best was a flat global offset, a medium-resolution global offset, and a fine-resolution global offset (Figure 1). Again, observation interpolation was best at small radius limits and quickly decreased as the radius increased (Figure 1). The interpolation of the benzene observations did not behave as expected when the flat global offset performed the best (Figure 1). The flat global offset may have performed better for benzene estimation than the more spatially resolved global offsets because the flat offset had the largest spatial long-range covariance structure among the tested methods, meaning that it could pull information from monitoring stations further away. These results demonstrated that, outside of the CAMx-modeled domain, moderately intense STEX estimation methods improved estimation performance compared to the least resource-intense methods. In the case of benzene, the least resource-intense methods performed better than moderately intense methods, which demonstrated the trade-off between higher-resolution global offsets revealing less autocorrelation in datasets.

The best estimation methods outside the CAMx-modeled domain were a fine-resolution global offset + CAMx data fusion to predict STEX and a flat global offset + CAMx data fusion to predict benzene (Figure 1). Data fusion with CAMx improved the SBTEX estimation performance for locations outside the CAMx-modeled domain (Figure 1). This was likely due to low validation R values from only observational data and long-range spatial covariance structures that allowed CAMx information to be pulled when estimation locations were far away. For each of the best-selected SBTEX methods outside the CAMx-modeled domain, data fusion performance appeared to plateau (or slightly increase) at large RCV radii (Figure 1). It was not possible to confirm whether this plateau was real or was a slow decrease in the performance because of the limited size of the study domain. RCV results of SBTEX outside the CAMx-modeled domain suggest that the most intense methods, utilizing information from nearby CAMx, performed the best for ambient SBTEX estimation (Figure 1).

3.3. Estimate Mapping Using Composite Methods. Figure 2 shows SBTEX median and mean estimations and error standard deviation (SD) on the randomly selected date, August 4, 2012. Toluene, ethylbenzene, and xylene estimates showed similar spatial distribution, which is consistent with previous findings of high Pearson correlation and comparable industry patterns between these pollutants[50,51] (Figure 2). Benzene estimates demonstrated similar areas of elevated concentrations to toluene, ethylbenzene, and xylenes, but showed more widespread distribution, potentially from high wildfire and on-road vehicle emissions contributions in 2012[1] (Figure 2). Conversely, styrene estimates exhibited more localized high concentrations likely from point source emissions and its short ambient lifetime[1,18] (Figure 2). Additional log-, mean-, and median-concentration spatial estimates (Figures S40−S45) and a detailed comparison of CAMx and BME estimates (Figure S46) are available in the Supporting Information.

Importantly, all SBTEX estimates demonstrated high concentrations along the Interstate 10 corridor (between Houston, TX, and Baton Rouge, LA) and near the Mississippi River corridor (between Baton Rouge and New Orleans, LA). These corridors are known for high vehicular traffic and petrochemical industry activity[1,52,53] and the Mississippi corridor, colloquially referred to as "Cancer Alley", has long been the subject of environmental justice concerns within lower-income and Black communities.[52−54] The ability of BME to estimate SBTEX at user-specified space/time locations provides epidemiologists with a novel framework to determine, at a high resolution and accounting for seasonality (from observational and air quality model information), the most at-risk populations and exposure inequities.

SBTEX estimation SD within the CAMx-modeled domain was mostly lower than outside the CAMx-modeled domain because estimates within the CAMx-modeled domain had more information sources to pull from, leading to less uncertainty and more detail in the median and mean SBTEX estimates (Figure 2). For median SBTEX estimates, the transition from within to outside the CAMx-modeled domains was gradual over a short distance and minimal, which would not greatly affect the rank order of participants' exposure for an epidemiological study that included participants residing outside (and nearby) the CAMx-modeled domain (Figure

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

Table 2. Select Daily Percentile (Median) SBTEX Exposures at Low Source Density and High Source Density Residence Locations within the CAMx-Modeled Domain from 2011−2012[a]

species name	75th percentile point source count within commuting distance	source density group	percentile daily exposure (ppb)		
			2.5th	50th	97.5th
styrene	15	low (n = 26453)	2.2×10^{-3}	5.7×10^{-3}	1.0×10^{-2}
		high (n = 7441)	2.6×10^{-3}	6.3×10^{-3}	2.4×10^{-2}
benzene	19	low (n = 25423)	7.0×10^{-2}	1.3×10^{-1}	2.7×10^{-1}
		high (n = 8471)	8.2×10^{-2}	1.6×10^{-1}	3.9×10^{-1}
toluene	23	low (n = 25866)	1.4×10^{-1}	2.3×10^{-1}	4.5×10^{-1}
		high (n = 8028)	1.4×10^{-1}	2.9×10^{-1}	7.7×10^{-1}
ethylbenzene	20	low (n = 25547)	2.0×10^{-2}	3.0×10^{-2}	5.0×10^{-2}
		high (n = 8347)	2.0×10^{-2}	3.6×10^{-2}	8.8×10^{-2}
xylenes	8	low (n = 27499)	5.8×10^{-2}	1.0×10^{-1}	2.1×10^{-1}
		high (n = 6395)	5.8×10^{-2}	1.1×10^{-1}	2.1×10^{-1}

[a]Low source density residence locations were defined as residence locations having less than or equal to the 75th percentile number of point source emitters within commuting distance (0.15°).

2). As expected, mean SBTEX estimates were higher, and the CAMx-availability transition was more distinct than median SBTEX estimates due to the inclusion of variance in mean calculations but not in median calculations (Figure 2). Based on these findings, median estimates alone or mean estimates accounting for estimation SD (*e.g.,* using estimation inverse SD-weighted linear regression) are recommended for use in epidemiological analyses and to limit differential exposure misclassification.

3.4. GuLF Study Exposure Estimation. Using composite methods, we estimated daily SBTEX exposures at participant residence locations to evaluate potential exposure inequities in the Gulf States. Table 2 shows each pollutant's 2.5th, 50th, and 97.5th percentile of daily exposure. For styrene in the low source density group, the 2.5th and 97.5th percentile of exposure were 0.0022 and 0.01 ppb, indicating a noticeable contrast between the least and most exposed study participants. There was a similar ratio of least to most exposed for the other subgroup and components of BTEX (Table 2). This contrast is a potential asset for future epidemiological analyses of ambient SBTEX-related health effects.

The possibility that participants may work at facilities that emit SBTEX was also investigated based on the presence of polluting industries within typical commuting distance of residences. The 75th percentile number of point sources within commuting distance of residence locations was 15, 19, 23, 20, and 8 for SBTEX compounds, respectively. As seen in Table 2, those who lived within typical commuting distance of high styrene source density had a 97.5th percentile daily exposure of 0.024 ppb, compared to those who lived within typical commuting distance of low source density, who had a 97.5th percentile exposure of 0.010 ppb. This means there was a 2.4-fold increase in 97.5th percentile exposures between residence locations within commuting distance of low versus high source density. Likewise, the ratios of 97.5th percentile daily exposures between low versus high source density areas were 1.4, 1.7, 1.8, and 1.0 for BTEX, respectively (Table 2). It is possible this relationship was smaller for xylenes because the 75th percentile of xylenes point sources within commuting distance (n = 8) was half, or less, than the number of point sources for other SBTEX compounds (n = 15−23).

Since individuals who live near a given polluting industry are more likely than those who live further away to work in that industry,[55−58] it is ideal for epidemiological studies to account, when possible, for both workplace and environmental exposures.[59] Furthermore, previous investigations have found that nonwhite and Hispanic households more frequently lived near emitters[57,58,60−62] and were exposed to higher air pollution levels.[63−66] If the situation for SBTEX-emitting industries is like that for other polluting industries, then total SBTEX exposure inequities in the Gulf States may be magnified as a function of race, ethnicity, and occupation. Additional details on cancer and noncancer health effects and the benefits of BME estimation are included in the Supporting Information.

4. FUTURE WORK

Future work may explore computational efficiency improvements that would increase the number of BME methods that could be investigated (*e.g.,* treating observational data as soft based on detection limits, applying different CAMP correction characteristics, using a non-natural logarithm VOC transformation) and potentially improving results. Additionally, future research should evaluate this methodology in other regions of the US to determine its generalizability for broader epidemiological work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c05094.

Observational data processing and exploratory analysis; LUR development; air quality model processing and CAMP correction; BME framework description; BME spatial estimate development; BME validation description; within and outside CAMx-modeled domain and composite Gulf States validation results; log-concentration and composite Gulf States spatial estimates; composite method and CAMx output estimate comparison; and GuLF Study health risks and exposure discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marc L. Serre − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*; Email: marc_serre@unc.edu

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

Authors

Nora A. Abbott − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0002-5425-8352

Lucie Semone − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

Richard Strott − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

Praful Dodda − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

Chi-Tsan Wang − *Center for Spatial Information Science and Systems (CSISS), George Mason University, Fairfax, Virginia 22030, United States*

Jaime Green − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

Bok Haeng Baek − *Center for Spatial Information Science and Systems (CSISS), George Mason University, Fairfax, Virginia 22030, United States;* orcid.org/0000-0003-1054-6325

Lawrence S. Engel − *Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

William Vizuete − *Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, North Carolina 27599, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c05094

Author Contributions

The manuscript was written through the contributions of all authors. Serre, Baek, Engel, and Vizuete conceived and designed the analysis; Wang, Green, and Dodda contributed data or analysis tools; Abbott, Simone, and Strott performed the analysis; and Abbott wrote the paper. All authors critically reviewed the paper and have approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the National Institute of Environmental Health Sciences R01 ES031127 and by the National Institute for Occupational Safety and Health T42OH008673. This work was also supported by the Korea Environment Industry & Technology Institute Project funded by the Korea Ministry of Environment RS-2023-00232066.

■ ABBREVIATIONS

SBTEX	styrene, benzene, toluene, ethylbenzene, and xylenes
CAMx	Comprehensive Air quality Model with extension
VOC	volatile organic compound
US	United States
LUR	land use regression
BME	Bayesian Maximum Entropy
KS	kernel smoothing
EPA	Environmental Protection Agency
AQS	Air Quality System
TX	Texas
LA	Louisiana
MS	Mississippi
AL	Alabama
GA	Georgia
FL	Florida
MDL	method detection limit
NEI	National Emissions Inventory
FHWA	Federal Highway Administration
AADT	annual average daily traffic
SEDC	Sum of Exponentially Decaying Contributions
FIT	Find, Inform, and Test
HAP	hazardous air pollutant
WRF	Weather Research and Forecasting Model
HAPI	Hazardous Air Pollutants Imputation
STEX	styrene, toluene, ethylbenzene, and xylenes
CAMP	Constant Air Model Performance
S/TRF	space/time random field
RCV	radius cross-validation
LOOCV	leave-one-out cross-validation
SD	standard deviation
ppb	parts per billion
IUR	inhalation unit risk
O_3	ozone
NO_2	nitrogen dioxide.

■ REFERENCES

(1) Wang, C.-T.; Baek, B. H.; Vizuete, W.; Engel, L. S.; Xing, J.; Green, J.; Serre, M.; Strott, R.; Bowden, J.; Woo, J.-H. Spatiotemporally Resolved Emissions and Concentrations of Styrene, Benzene, Toluene, Ethylbenzene, and Xylenes (SBTEX) in the US Gulf Region. *Earth Syst. Sci. Data* **2023**, *15* (11), 5261−5279.

(2) Kheirbek, I.; Johnson, S.; Ross, Z.; Pezeshki, G.; Ito, K.; Eisl, H.; Matte, T. Spatial Variability in Levels of Benzene, Formaldehyde, and Total Benzene, Toluene, Ethylbenzene and Xylenes in New York City: A Land-Use Regression Study. *Environ. Health* **2012**, *11* (1), 51.

(3) Werder, E. J.; Engel, L. S.; Blair, A.; Kwok, R. K.; McGrath, J. A.; Sandler, D. P. Blood BTEX Levels and Neurologic Symptoms in Gulf States Residents. *Environ. Res.* **2019**, *175*, 100−107.

(4) Werder, E. J.; Sandler, D. P.; Richardson, D. B.; Emch, M. E.; Kwok, R. K.; Engel, L. S. Determinants of Environmental Styrene Exposure in Gulf Coast Residents. *Expo. Sci. Environ. Epidemiol.* **2019**, *29* (6), 831−841.

(5) Ashley, D. L.; Bonin, M. A.; Cardinali, F. L.; McCraw, J. M.; Wooten, J. V. Measurement of Volatile Organic Compounds in Human Blood. *Environ. Health Perspect.* **1996**, *104*, 871−877.

(6) Werder, E. J.; Engel, L. S.; Richardson, D. B.; Emch, M. E.; Gerr, F. E.; Kwok, R. K.; Sandler, D. P. Environmental Styrene Exposure and Neurologic Symptoms in U.S. Gulf Coast Residents. *Environ. Int.* **2018**, *121*, 480−490.

(7) Lemke, L. D.; Lamerato, L. E.; Xu, X.; Booza, J. C.; Reiners, J. J.; Raymond III, D. M.; Villeneuve, P. J.; Lavigne, E.; Larkin, D.; Krouse, H. J. Geospatial Relationships of Air Pollution and Acute Asthma Events across the Detroit−Windsor International Border: Study Design and Preliminary Results. *J. Expo. Sci. Environ. Epidemiol.* **2014**, *24* (4), 346−357.

(8) Saxena, P.; Sonwani, S. *Criteria Air Pollutants and Their Impact on Environmental Health*; Springer: Singapore, 2019.

(9) Whitworth, K. W.; Symanski, E.; Lai, D.; Coker, A. L. Kriged and Modeled Ambient Air Levels of Benzene in an Urban Environment: An Exposure Assessment Study. *Environ. Health* **2011**, *10* (1), 21.

(10) Mukerjee, S.; Smith, L.; Neas, L.; Norris, G. Evaluation of Land Use Regression Models for Nitrogen Dioxide and Benzene in Four US Cities. *Sci. World J.* **2012**, *2012*, 1−8.

(11) Wong, D. W.; Yuan, L.; Perlin, S. A. Comparison of Spatial Interpolation Methods for the Estimation of Air Quality Data. *J. Expo. Sci. Environ. Epidemiol.* **2004**, *14* (5), 404−415.

(12) United States Government Accountability Office. *AIR POLLUTION: Opportunities to Better Sustain and Modernize the National Air Quality Monitoring System; GAO-21−38*; United States Government Accountability Office: Washington, DC, 2020. https://www.gao.gov/assets/gao-21-38.pdf (accessed 2024−02−28).
(13) EPA. Air Toxics Ambient Monitoring. https://www.epa.gov/amtic/air-toxics-ambient-monitoring (accessed April 25, 2024).
(14) EPA. AMTIC—Ambient Monitoring Archive for HAPs. https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps (accessed April 25, 2024).
(15) Miller, L.; Xu, X.; Luginaah, I. Spatial Variability of Volatile Organic Compound Concentrations in Sarnia, Ontario, Canada. *J. Toxicol. Environ. Health, Part A* **2009**, *72* (9), 610−624.
(16) Miller, L.; Lemke, L. D.; Xu, X.; Molaroni, S. M.; You, H.; Wheeler, A. J.; Booza, J.; Grgicak-Mannion, A.; Krajenta, R.; Graniero, P. Intra-Urban Correlation and Spatial Variability of Air Toxics across an International Airshed in Detroit, Michigan (USA) and Windsor, Ontario (Canada). *Atmos. Environ.* **2010**, *44* (9), 1162−1174.
(17) Cocheo, C.; Sacco, P.; Ballesta, P. P.; Donato, E.; Garcia, S.; Gerboles, M.; Gombert, D.; McManus, B.; Patier, R. F.; Roth, C.; De Saeger, E.; Wright, E. Evaluation of the Best Compromise between the Urban Air Quality Monitoring Resolution by Diffusive Sampling and Resource Requirements. *J. Environ. Monit.* **2008**, *10* (8), 941.
(18) Agency for Toxic Substances and Disease Registry. Toxicological Profile for Styrene, 2010. https://www.atsdr.cdc.gov/toxprofiles/tp53.pdf (accessed Jan 17, 2024).
(19) Agency for Toxic Substances and Disease Registry. Toxicological Profile for Benzene, 2007. https://www.atsdr.cdc.gov/toxprofiles/tp3.pdf (accessed Jan 17, 2024).
(20) Agency for Toxic Substances and Disease Registry. Toxicological Profile for Toluene, 2017. https://www.atsdr.cdc.gov/toxprofiles/tp56.pdf (accessed Jan 17, 2024).
(21) Agency for Toxic Substances and Disease Registry. Toxicological Profile for Ethylbenzene, 2010. https://www.atsdr.cdc.gov/toxprofiles/tp110.pdf (accessed Jan 17, 2024).
(22) Agency for Toxic Substances and Disease Registry. Toxicological Profile for Xylene, 2007. https://www.atsdr.cdc.gov/toxprofiles/tp71.pdf (accessed Jan 17, 2024).
(23) Azmi, W. N. F. W.; Pillai, T. R.; Latif, M. T.; Koshy, S.; Shaharudin, R. Application of Land Use Regression Model to Assess Outdoor Air Pollution Exposure: A Review. *Environ. Adv.* **2023**, *11*, 100353.
(24) Su, J. G.; Jerrett, M.; Beckerman, B.; Verma, D.; Arain, M. A.; Kanaroglou, P.; Stieb, D.; Finkelstein, M.; Brook, J. A Land Use Regression Model for Predicting Ambient Volatile Organic Compound Concentrations in Toronto, Canada. *Atmos. Environ.* **2010**, *44* (29), 3529−3537.
(25) Mohammadi, A.; Ghassoun, Y.; Löwner, M.-O.; Behmanesh, M.; Faraji, M.; Nemati, S.; Toolabi, A.; Abdolahnejad, A.; Panahi, H.; Heydari, H.; Miri, M. Spatial Analysis and Risk Assessment of Urban BTEX Compounds in Urmia, Iran. *Chemosphere* **2020**, *246*, 125769.
(26) Wheeler, A. J.; Smith-Doiron, M.; Xu, X.; Gilbert, N. L.; Brook, J. R. Intra-Urban Variability of Air Pollution in Windsor, Ontario—Measurement and Modeling for Human Exposure Assessment. *Environ. Res.* **2008**, *106* (1), 7−16.
(27) Hsu, C.-Y.; Zeng, Y.-T.; Chen, Y.-C.; Chen, M.-J.; Lung, S.-C. C.; Wu, C.-D. Kriging-Based Land-Use Regression Models That Use Machine Learning Algorithms to Estimate the Monthly BTEX Concentration. *Int. J. Environ. Res. Publ. Health* **2020**, *17* (19), 6956.
(28) Demanega, I.; Mujan, I.; Singer, B. C.; Anđelković, A. S.; Babich, F.; Licina, D. Performance Assessment of Low-Cost Environmental Monitors and Single Sensors under Variable Indoor Air Quality and Thermal Conditions. *Build. Environ.* **2021**, *187*, 107415.
(29) Curci, G.; Palmer, P. I.; Kurosu, T. P.; Chance, K.; Visconti, G. Estimating European Volatile Organic Compound Emissions Using Satellite Observations of Formaldehyde from the Ozone Monitoring Instrument. *Atmos. Chem. Phys.* **2010**, *10* (23), 11501−11517.
(30) EPA. AirToxScreen Overview. https://www.epa.gov/AirToxScreen/airtoxscreen-overview (accessed Feb 29, 2024).
(31) Smith, L. A.; Mukerjee, S.; Chung, K. C.; Afghani, J. Spatial Analysis and Land Use Regression of VOCs and NO2 in Dallas, Texas during Two Seasons. *J. Environ. Monit.* **2011**, *13* (4), 999.
(32) de Nazelle, A.; Arunachalam, S.; Serre, M. L. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina. *Environ. Sci. Technol.* **2010**, *44* (15), 5707−5713.
(33) Xu, Y.; Serre, M. L.; Reyes, J.; Vizuete, W. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: A National Application. *Environ. Sci. Technol.* **2016**, *50* (8), 4393−4400.
(34) Reyes, J. M.; Xu, Y.; Vizuete, W.; Serre, M. L. Regionalized PM2.5 Community Multiscale Air Quality Model Performance Evaluation across a Continuous Spatiotemporal Domain. *Atmos. Environ.* **2017**, *148*, 258−265.
(35) Akita, Y.; Carter, G.; Serre, M. L. Spatiotemporal Nonattainment Assessment of Surface Water Tetrachloroethylene in New Jersey. *J. Environ. Qual.* **2007**, *36* (2), 508−520.
(36) Akita, Y.; Baldasano, J. M.; Beelen, R.; Cirach, M.; de Hoogh, K.; Hoek, G.; Nieuwenhuijsen, M.; Serre, M. L.; de Nazelle, A. Large Scale Air Pollution Estimation Method Combining Land Use Regression and Chemical Transport Modeling in a Geostatistical Framework. *Environ. Sci. Technol.* **2014**, *48* (8), 4452−4459.
(37) Becker, J. S.; DeLang, M. N.; Chang, K.-L.; Serre, M. L.; Cooper, O. R.; Wang, H.; Schultz, M. G.; Schröder, S.; Lu, X.; Zhang, L.; Deushi, M.; Josse, B.; Keller, C. A.; Lamarque, J.-F.; Lin, M.; Liu, J.; Marécal, V.; Strode, S. A.; Sudo, K.; Tilmes, S.; Zhang, L.; Brauer, M.; West, J. J. Using Regionalized Air Quality Model Performance and Bayesian Maximum Entropy Data Fusion to Map Global Surface Ozone Concentration. *Elementa* **2023**, *11* (1), 00025.
(38) Serre, M. L.; Christakos, G. Modern Geostatistics: Computational BME Analysis in the Light of Uncertain Physical Knowledge − the Equus Beds Study. *Stoch. Environ. Res. Risk Assess.* **1999**, *13*, 1−26.
(39) Valencia, A.; Serre, M.; Arunachalam, S. A Hyperlocal Hybrid Data Fusion Near-Road PM2.5 and NO2 Annual Risk and Environmental Justice Assessment across the United States. *PLoS One* **2023**, *18* (6), No. e0286406.
(40) Xu, Y.; Serre, M. L.; Reyes, J. M.; Vizuete, W. Impact of Temporal Upscaling and Chemical Transport Model Horizontal Resolution on Reducing Ozone Exposure Misclassification. *Atmos. Environ.* **2017**, *166*, 374−382.
(41) Cleland, S. E.; West, J. J.; Jia, Y.; Reid, S.; Raffuse, S.; O'Neill, S.; Serre, M. L. Estimating Wildfire Smoke Concentrations during the October 2017 California Fires through BME Space/Time Data Fusion of Observed, Modeled, and Satellite-Derived PM $_{2.5}$. *Environ. Sci. Technol.* **2020**, *54* (21), 13439−13447.
(42) Messier, K. P.; Akita, Y.; Serre, M. L. Integrating Address Geocoding, Land Use Regression, and Spatiotemporal Geostatistical Estimation for Groundwater Tetrachloroethylene. *Environ. Sci. Technol.* **2012**, *46* (5), 2772−2780.
(43) Reyes, J. M.; Serre, M. L. An LUR/BME Framework to Estimate PM $_{2.5}$ Explained by on Road Mobile and Stationary Sources. *Environ. Sci. Technol.* **2014**, *48* (3), 1736−1744.
(44) Li, A. J.; Pal, V. K.; Kannan, K. A Review of Environmental Occurrence, Toxicity, Biotransformation and Biomonitoring of Volatile Organic Compounds. *Environ. Chem. Ecotoxicol.* **2021**, *3*, 91−116.
(45) Wiesner-Friedman, C.; Beattie, R. E.; Stewart, J. R.; Hristova, K. R.; Serre, M. L. Microbial Find, Inform, and Test Model for Identifying Spatially Distributed Contamination Sources: Framework Foundation and Demonstration of Ruminant *Bacteroides* Abundance in River Sediments. *Environ. Sci. Technol.* **2021**, *55* (15), 10451−10461.
(46) Messier, K. P.; Kane, E.; Bolich, R.; Serre, M. L. Nitrate Variability in Groundwater of North Carolina Using Monitoring and Private Well Data Models. *Environ. Sci. Technol.* **2014**, *48* (18), 10804−10812.

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55

(47) Christakos, G.; Bogaert, P.; Serrre, M. L. *Temporal GIS: Advanced Functions for Field-Based Applications*; Springer: New York, NY, 2001.
(48) Christakos, G. A Bayesian/Maximum-Entropy View to the Spatial Estimation Problem. *Math. Geol.* **1990**, *22*, 763−777.
(49) Bureau of Transportation. From Home to Work, the Average Commute Is 26.4 minutes, 2003. https://www.nrc.gov/docs/ML1006/ML100621425.pdf (accessed April 25, 2024).
(50) Smith, L. A.; Stock, T. H.; Chung, K. C.; Mukerjee, S.; Liao, X. L.; Stallings, C.; Afshar, M. Spatial Analysis of Volatile Organic Compounds from a Community-Based Air Toxics Monitoring Network in Deer Park, Texas, USA. *Environ. Monit. Assess.* **2007**, *128* (1−3), 369−379.
(51) Tecer, L. H.; Tagil, S.; Ulukaya, O.; Ficici, M. Spatial Distribution of BTEX and Inorganic Pollutants During Summer Season in Yalova, Turkey. *Ecol. Chem. Eng. S* **2017**, *24* (4), 565−581.
(52) Caiazzo, F.; Ashok, A.; Waitz, I. A.; Yim, S. H. L.; Barrett, S. R. H. Air Pollution and Early Deaths in the United States. Part I: Quantifying the Impact of Major Sectors in 2005. *Atmos. Environ.* **2013**, *79*, 198−208.
(53) Robinson, E. S.; Tehrani, M. W.; Yassine, A.; Agarwal, S.; Nault, B. A.; Gigot, C.; Chiger, A. A.; Lupolt, S. N.; Daube, C.; Avery, A. M.; Claflin, M. S.; Stark, H.; Lunny, E. M.; Roscioli, J. R.; Herndon, S. C.; Skog, K.; Bent, J.; Koehler, K.; Rule, A. M.; Burke, T.; Yacovitch, T. I.; Nachman, K.; DeCarlo, P. F. Ethylene Oxide in Southeastern Louisiana's Petrochemical Corridor: High Spatial Resolution Mobile Monitoring during HAP-MAP. *Environ. Sci. Technol.* **2024**, *58* (25), 11084−11095.
(54) Blodgett, A. D. An Analysis of Pollution and Community Advocacy in 'Cancer Alley': Setting an Example for the Environmental Justice Movement in St James Parish, Louisiana. *Local Environ.* **2006**, *11* (6), 647−661.
(55) Benedetti, M.; Lavarone, I.; Comba, P. Cancer Risk Associated with Residential Proximity to Industrial Sites: A Review. *Arch. Environ. Health* **2001**, *56* (4), 342−349.
(56) Brender, J. D.; Maantay, J. A.; Chakraborty, J. Residential Proximity to Environmental Hazards and Adverse Health Outcomes. *Am. J. Public Health* **2011**, *101* (S1), S37−S52.
(57) Brender, J. D.; Suarez, L.; Langlois, P. H.; Steck, M.; Zhan, F. B.; Moody, K. Are Maternal Occupation and Residential Proximity to Industrial Sources of Pollution Related? *J. Occup. Environ. Med.* **2008**, *50* (7), 834−839.
(58) Bullard, R. D.; Mohai, P.; Saha, R.; Wright, B. TOXIC WASTES AND RACE AT TWENTY: WHY RACE STILL MATTERS AFTER ALL OF THESE YEARS. *Environ. Law* **2008**, *38*, 371.
(59) Jerrett, M. Do Socioeconomic Characteristics Modify the Short Term Association between Air Pollution and Mortality? Evidence from a Zonal Time Series in Hamilton, Canada. *J. Epidemiol. Community Health* **2004**, *58* (1), 31−40.
(60) Mohai, P.; Saha, R. Which Came First, People or Pollution? Assessing the Disparate Siting and Post-Siting Demographic Change Hypotheses of Environmental Injustice. *Environ. Res. Lett.* **2015**, *10* (11), 115008.
(61) Johnston, J. E.; Werder, E.; Sebastian, D. Wastewater Disposal Wells, Fracking, and Environmental Injustice in Southern Texas. *Am. J. Public Health* **2016**, *106* (3), 550−556.
(62) Johnston, J. E.; Chau, K.; Franklin, M.; Cushing, L. Environmental Justice Dimensions of Oil and Gas Flaring in South Texas: Disproportionate Exposure among Hispanic Communities. *Environ. Sci. Technol.* **2020**, *54* (10), 6289−6298.
(63) Werder, E. J.; Sandler, D. P.; Richardson, D. B.; Emch, M. E.; Kwok, R. K.; Gerr, F. E.; Engel, L. S. Environmental Styrene Exposure and Sensory and Motor Function in Gulf Coast Residents. *Environ. Health Perspect.* **2019**, *127* (4), 047006.
(64) Ash, M.; Boyce, J. K. Racial Disparities in Pollution Exposure and Employment at US Industrial Facilities. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (42), 10636−10641.
(65) Tessum, C. W.; Paolella, D. A.; Chambliss, S. E.; Apte, J. S.; Hill, J. D.; Marshall, J. D. $PM_{2.5}$ Polluters Disproportionately and Systemically Affect People of Color in the United States. *Sci. Adv.* **2021**, *7* (18), No. eabf4491.
(66) Mikati, I.; Benson, A. F.; Luben, T. J.; Sacks, J. D.; Richmond-Bryant, J. Disparities in Distribution of Particulate Matter Emission Sources by Race and Poverty Status. *Am. J. Public Health* **2018**, *108* (4), 480−485.

https://doi.org/10.1021/acs.est.4c05094
Environ. Sci. Technol. 2025, 59, 45−55