Spatiotemporally resolved emissions and concentrations of Styrene, Benzene, Toluene, Ethylbenzene, and Xylenes (SBTEX) in the U.S. Gulf region

Chi-Tsan Wang[1], William Vizuete[2], Lawrence S. Engel[3], Jia Xing[1], Jaime Green[2], Marc Serre[2], Richard Strott[2], Jared Bowden[4], Jung-Hun Woo[5], and Bok H. Baek[1*]

[1]Center for Spatial Information Science and Systems (CSISS), George Mason University, Fairfax, VA, U.S.

[2]Department of Environmental Sciences & Engineering, University of North Carolina, Chapel Hill, NC, U.S.

[3]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, U.S.

[4]Departement of Applied Ecology, North Carolina State University, Raleigh, NC, U.S.

[5]Civil and Environmental Engineering, College of Engineering, Konkuk University, Seoul, Republic of Korea

[*]Corresponding authors: Bok H. Baek (Email: bhbaek@gmail.com; Telephone: +1 919-308-6519)

Abstract

Styrene, benzene, toluene, ethylbenzene, and xylenes (SBTEX) are established neurotoxicants. These SBTEX are hazardous air pollutants (HAPs) and released from the petrochemical industry, combustion process, transport emission, and solvent usage sources. Although several SBTEX toxic assessment studies have been conducted, they have mainly relied on ambient measurements to estimate exposure and limiting their scope to specific locations and observational periods. To overcome these spatiotemporal limitations, an air quality modeling system over the U.S. Gulf region was created predicting the spatially and temporally enhanced SBTEX modeling concentrations from May to September 2012. Due to the incompleteness of SBTEX in the official US EPA National Emission Inventory (NEI), Hazardous Air Pollutions Imputation (HAPI) program was used to identify and estimate the missing HAPs emissions. The improved emission data was processed to generate the chemically-speciated hourly gridded emission inputs for the Comprehensive Air Quality Model with Extensions (CAMx) chemical transport model to simulate the SBTEX concentrations over the Gulf modeling region. SBTEX pollutants were modeled using

a "Reactive Tracer" feature in CAMx that accounts for their chemical and physical processes in the atmosphere. The data shows that the major SBTEX emissions in this region are contributed by mobile emission (45%), wildfire (30%), and industry (26%). Most SBTEX emissions are emitted during daytime hours (local time 14:00 -17:00), and the emission rate in the model domain is about 20 - 40 t hr^{-1}, which is about 4 times higher than that in the night-time (local time 24:00 – 4:00, about 4 – 10 t hr^{-1}). High concentrations of SBTEX (above 1 ppb) occurred near the cities close to the I-10 interstate highway (Houston, Beaumont, Lake Charles, Lafayette, Baton Rouge, New Orleans, and Mobile) and other metropolitan cities (Shreveport and Dallas). High styrene concentrations were co-located with industrial sources, which contribute the most to the styrene emissions. The HAPI program successfully estimated missing emissions of styrene from the chemical industry. The change increased total styrene emissions was increased by 22% resulting in maximum ambient concentrations increasing from 0.035 ppb to 1.75 ppb across the model domain. The predicted SBTEX concentrations with imputed emissions present good agreement with observational data, with a correlation coefficient (R) of 0.75 (0.46 to 0.77 for individual SBTEX species) and normalized mean bias (NMB) of -5.6% (-24.9% to 32.1% for individual SBTEX species), suggesting their value for supporting any SBTEX-related human health studies in the Gulf region.

Keywords: BTEX, styrene, SMOKE, Reactive Tracer, Toxicants, HAP, CAMx, Exposure

1. Introduction

Styrene, benzene, toluene, ethylbenzene, and xylenes (SBTEX) are listed as Hazardous Air Pollutants (HAPs) by the U.S. Environmental Protection Agency (EPA) (Declet-Barreto et al., 2020) and can be detected in unhealthy amounts in the ambient environment. The SBTEX is primarily from industrial emission sources and can be found in the petrochemical, construction, and manufacturing industries (Polvara et al., 2021; Declet-Barreto et al., 2020) with 98% of benzene emissions attributed to coal and petroleum sources (ATSDR, 2007a, b, 2010a, b, 2017). Exposure studies of total SBTEX at industrial sources in the Middle East, Europe, and West Asia, have shown that workers experience a cumulative yearly environmental exposure of 25 - 176 ppb (Al-Harbi et al., 2020; Rajabi et al., 2020; Christensen et al., 2018; Rahimpoor et al., 2022; Niaz et al., 2015; Moshiran et al., 2021). The inhalation reference concentration for benzene shows low-dose linearity utilizing maximum likelihood estimate E-5 risk level of benzene (1 in 100,000) range is 0.4 -1.4 ppb of air concentration for leukemia (USEPA, 2000).

Given the importance of SBTEX from industrial sources, the heavily industrialized Gulf region of the U.S. could be a significant source of exposure for the population living there. According to the Agency for Toxic Substances and Disease Registry (ATSDR) report, the petrochemical industry in the Gulf region states contributes approximately 52% (~5.3 million tons yr^{-1}) of benzene production capacity in the U.S. (ATSDR, 2007a) and ~75% (~6.2 million tons yr^{-1}) of xylenes production capacity (ATSDR, 2007b). Texas and Louisiana have significant production of styrene and ethylbenzene, with an annual production of 5.5 and 7.2 million tons yr^{-1}, respectively (SRI, 2008; ATSDR, 2010a). A recent study of SBTEX exposures in the U.S. Gulf region, conducted within the Gulf Long-term Follow-up Study (GULF Study) cohort (NIEHS, 2021), observed associations of blood concentrations and annual average air concentrations of these chemicals with neurological symptoms (Werder et al., 2019; Werder et al., 2018). The average blood BTEX concentration among the 146 tobacco smoke-unexposed participants with blood measurements in this study was 255 ng L^{-1} es (Doherty et al., 2017; Werder et al., 2018) . This value is similar to that for a representative nationwide sample assessed as part of the US National Health and Nutrition Examination Survey (NHANES) in 2005-2008 (NCHS, 2021), which measured an average of 247 ng L^{-1}. In the GuLF Study, however, the 95th percentile of BTEX concentrations was 991 ng L^{-1}, which is 23% higher than the 95th percentile for the NHANES nationwide sample of 803 ng L^{-1}. The mean blood concentration of styrene for the GuLF Study sample was 52 ng L^{-1}

(95th percentile: 882 ng L^{-1}), or twice the NHANES nationwide mean of 25 ng L^{-1} (95th percentile: 55 ng L^{-1}) (NCHS, 2021). Due to the short biological half-lives of SBTEX species, the study concluded that this high average SBTEX concentration in blood in the Gulf Region resulted from recent, presumably local emission sources.

Most ambient exposure studies of SBTEX have relied directly on local measurements from the field, or at existing ambient monitors. These measurements can then be used in statistical models to spatially predict exposures to SBTEX (Pankow et al., 2003; O'Leary and Lemke, 2014; Miller et al., 2018; Hsieh et al., 2020b). For example, Hsieh et al. (Hsieh et al., 2020a) developed the Multivariate Linear Regression (MLR) models to estimate SBTEX concentrations using correlations with other criteria air pollutants, including nitrogen oxides (NOx), carbon monoxide (CO), sulfur dioxide (SO_2), particulate matter (PM), and meteorological conditions (temperature, wind speed). The MLR model predicted a strong correlation with NOx and CO. The limitations of the statistical model are that they require measurement data, and they assume that the measurements originate from a single source in a relatively small region. The use of a dispersion model is another way to estimate ambient SBTEX concentrations when local measurements are lacking. Chen et al., 2016 (Chen et al., 2016) applied a dispersion model to predict SBTEX and other toxicant concentrations in two industrial complexes in Kaohsiung City, Taiwan. The dispersion model performed better for stationary point sources than a statistical based model and predicted up to ~78% of the ambient observation. These dispersion models, however, only account for exposures at a smaller spatial scale (USEPA, 2022, 2023) thus cannot support regional scale (e.g., state-level) application. Furthermore, these models assumed that the exposure rate to SBTEX is linear, without considering any chemical destruction and wet/dry deposition losses in the atmosphere.

An accurate SBTEX assessment in the Gulf region must address the known uncertainties associated with current statistical, biometric, and dispersion model approaches. Improved accuracy in exposure estimation is dependent on the inclusion of all industrial emission sources, must capture the temporal and spatial variability known to occur in industrial emission rates, and should include the chemical and physical decay processes of the atmosphere. These issues can be addressed using a regional-scale chemical transport model (CTM), like the Comprehensive Air Quality Model with Extension (CAMx) (RAMBOLL, 2021) coupled with an emission inventory

that provides a comprehensive account of all SBTEX sources. Besides, the reactive tracer function, which is one of the CAMx probing tools, allows the model to explicitly simulate SBTEX concentrations. Currently, SBTEX emission data can be found in EPA's National Emission Inventory (NEI), which includes data from the Toxics Release Inventory (TRI) program database (USEPA, 2021d). Unlike for benzene sources, the TRI data for the other four species (STEX) is based on voluntary reports, and as a result, the 2011 NEI has emission rate data for these air toxics only for a limited set of emission sources (USEPA, 2021a).

The following work describes the development of a new STEX emission inventory for the Gulf Coast region that includes the emission sources absent from the 2011 NEI. Missing emission rate data of STEX was provided by analyzing NEI emissions of similar industrial sources that did provide emission rates and applying their rates to the missing source. Diurnal profiles for STEX were based on the hourly profiles of other pollutants with the same type of industrial source. This study then applied the Sparse Matrix Operator Kernel Emissions (SMOKE) model system (Baek and Seppanen, 2021) to generate a CAMx-ready emission inventory. Since STEX are not included as explicit species in the chemical mechanisms used by CAMx, a reactive tracer was included to account for chemical losses. This new emission inventory was then utilized in CAMx to predict STEX concentrations over the Gulf region for 5 months in 2012.

2. Materials and Methods

In this study, the 2011 version 6 NEI was applied as the base emission inventory (USEPA, 2021b). Subsequently, the SMOKE modeling system was employed to produce hourly gridded emissions of SBTEX across the Gulf modeling region for the year 2012. These SBTEX emissions were utilized in conjunction with the CTM model and a reactive tracer function to generate the SBTEX concentration map. In the end, the USEPA Ambient Monitoring Technology Information Center (AMTIC) data were employed to evaluate the model performance in simulating SBTEX concentrations.

2.1 Emission Data Preparation

2.1.1 The HAPs Emissions in NEI

The NEI is a national database providing comprehensive annual air emission estimates for both criteria air pollutants (CAPs) (e.g., CO, NO_X, SO_2, NH_3, VOC, and $PM_{2.5}$), and HAPs (e.g., benzene, acetaldehyde, formaldehyde, xylenes, styrene, and more) from all types of emissions sources (e.g., point, nonpoint, and mobile). While CAPs emissions are reported by the state agencies is mandatory, the report of HAPs is usually voluntary. Consequently, only a limited set of HAPs have been reported to the USEPA, and their spatial coverage can vary significantly by source type (e.g., industrial, vehicles) and region (e.g., county and state) (Strum et al., 2017).

The VOC emission species generated by SMOKE from NEI have three types which are “model surrogate”, “model explicit”, and “HAPs explicit” species. The “model surrogate species”, such as XYL (xylene and other poly-alkyl aromatics), TOL (toluene and other mono-alkyl aromatics), and PAR (paraffin carbon bond), are calculated by speciation profiles in emission model platform and used to predict ozone and secondary organic aerosol (SOA) in the CTM but not for individual HAPs emission and simulation. Only five HAP emissions in NEI are “model explicit” specie: Naphthalene (NAPH), benzene (BENZ), Acetaldehyde (ALD2), Formaldehyde (FORM), and Methanol (MEOH), known as "NBAFM" to represent their individual emission (Strum et al., 2017), and are directly processed in CTM model, too. The “HAPs explicit” species emission in NEI includes hundreds of toxicants (such as styrene, xylenes, Mercury, and Acrolein). Those “HAPs

153 explicit" species cannot be directly used in the current CTM model because their explicit chemical
154 mechanisms are not developed in the current CTM chemical mechanism.

155 The "model explicit" species, benzene (B), and other "HAPs explicit" species, including styrene,
156 toluene, ethylbenzene, and xylenes (STEX) are targeted for this SBTEX human exposure study.
157 The SMOKE model system (Baek and Seppanen, 2021) assigned the annual or monthly SBTEX
158 emission inventory in NEI to hourly emission patterns by the temporal profiles based on emission
159 processes and locations by Source Category Code (SCC) and Federal Information Processing
160 Standards (FIPS) county codes. These processes are coupled with the CAPs when generating the
161 CTM-ready emission data.

162 **2.1.2 Imputation of NEI with STEX**

163 Considering that Benzene emission reporting is mandatory in the NEI and thus can be assumed to
164 have no significant missing sources, we only focused on the investigation of missing sources for
165 the STEX which is voluntary reporting. This study utilized the 2011 NEI summary reports from
166 the SMOKE modeling System (Baek and Seppanen, 2021) to identify those missing STEX
167 emission sources. The SMOKE reports provided the annual or monthly total of VOC and
168 individual HAPs emissions sorted by SCC and FIPS county codes. This study developed an R-
169 project (The R Foundation, 2021) program called "*Hazardous Air Pollutants Imputation"* (HAPI)
170 that can first read the reports from SMOKE and identify the list of inventory sources reported
171 without STEX toxics. Then it generates the imputation data for those missing STEX inventory
172 sources based on the proxy of STEX and VOC for those emission sources that share the same SCC
173 near the region (county or state).

174 Theoretically, the SCC is the reference code defining the emission process type. The same SCC
175 means they share similar emission factors with the same emission process (USEPA, 2016). The
176 profiles of HAPs for the VOC can be shared with those same SCC emission sources within the
177 surrounding regions (counties or states)(Strum et al., 2017). When there are the same SCC
178 emission sources with zero HAPs in other counties, this study performed the imputation of those
179 missing HAPs emissions based on the HAPs profiles from the matched emission source. For
180 example, the HAPs profile of styrene and toluene to the VOC emission is defined as the ratio of

styrene and toluene emissions over the VOC emission ($P_{toluene,s}$) in counties where there are styrene, toluene, and VOC emissions for that SCC (s). Then, this study will assume that those HAPs are missing when the summation of HAPs emissions are zero ($\sum_i E_{i,s,f} = 0$: i is pollutants, s is SCC code, f is FIPS county code) but VOC emission is available. Then this will apply the HAPs profile for the same SCC to the existing VOC and estimate missing styrene and toluene emissions. Therefore, this process can impute the missing HAP emissions based on the SCC-matched HAPs fractions from the surrounding counties or the same state.

The HAPI was developed based on this imputation concept. This study first separated the county and SCC level inventory data into two groups in the HAPI program: "with HAPs" and "without HAPs." For the "with HAPs" group, summations of HAPs emissions in counties and SCCs are not zero. In contrast, for the "without HAPs" group, summations of HAPs emissions in counties and SCC are zero.

In the "with HAPs" group ($\sum_i E_{i,s,f} > 0$) in Eq. (1), i is the individual HAP, such as styrene, benzene, toluene, ethylbenzene, xylenes, Acrolein, and 1,3-Butadiene; s is the SCC, and f is the county FIPS code for county. $E_{i,s,f}$ is the annual emission of pollutant i for SCC in the county. $E_{voc,s,f}$ is the CAP VOC emission for the SCC in the county. The HAPs profile ($P_{i,s}$) is a fraction of HAP-specific emission ($E_{i,s,f}$) over the summation of matched SCC and county-specific VOC emission ($E_{voc,s,f}$) from the "with HAPs" group.

This study assumed that if there is an SCC-matched "with HAPs" group HAPs profile in the inventory, they are not considered as missing HAPs emission sources. Only the emission sources for which the sum of all HAPs is zero ($\sum_i E_{i,s,f} = 0$) are considered as the “without HAPs" group.” In Eq.(2), $P_{i,s}$ is used to estimate those missing HAPs for the "without HAPs" inventory source group. The $E_{voc,s,f}$ is the CAP VOC emission in the "without HAPs" group.

When $\sum_i E_{i,s,f} > 0$, calculate individual HAPs to total VOC ratio ($P_{i,s}$):

$$P_{i,s} = \frac{\sum_f E_{i,s,f}}{\sum_f E_{voc,s,f}} \qquad \text{Eq. (1)}$$

When $\sum_i E_{i,s,f} = 0$, the HAPs emission are missing, this study applied $P_{i,s}$ and VOC emission to estimate the missing HAPs emission:

$$Em_{i,s,f} = P_{i,s} \times E_{voc,s,f} \quad \text{Eq. (2)}$$

The HAPI program then outputs the total HAPs emissions ($Em_{i,s,f}$) for the SMOKE modeling system to integrate with the CAP VOC inventory described in Section 2.1.2. Finally, the HAPI program performs the quality assurance step again to confirm that there are no missing HAPs after imputation and that the summation of HAPs emissions is not greater than the CAP VOC emission.

2.2 Model Configuration

2.2.1 Air Quality Modeling

The Comprehensive Air Quality Model with Extensions version 7.0, CAMx7.0 (RAMBOLL 2020) was implemented in this study to simulate the SBTEX concentrations in the atmosphere. The model simulation period is from April 20th to September 30th, 2012 (April 20th to April 30th are spin-up dates). The evaluated meteorological data from WRF version 3.8 over the U.S. continental region are provided by the USEPA Support Center for Regulatory Atmospheric Modeling (SCRAM) (USEPA, 2022). The WRF output data were transformed into SMOKE-ready gridded hourly meteorology through the Meteorology Chemistry Interface Processor (MCIP). The emissions sectors modulated by meteorology, such as onroad (Choi et al., 2014; Lindhjem et al., 2004) and biogenic, were estimated with the MCIP gridded hourly meteorology. The USEPA's 2012 daily total wildfire emissions (ptfire) estimated by SMARTFIRE2 (USEPA, 2015) were also incorporated (USEPA, 2021b). Additionally, the WRF meteorological data were converted to CAMx-ready meteorological data by using WRFCAMx (RAMBOLL, 2020b) for the CAMx model input. The photodissociation coefficients are calculated by the Tropospheric Ultraviolet-Visible (TUV) Radiation Model (Madronich, 1987) with Ozone Monitoring Instrument (OMI) daily data (NASA, 2021). The USEPA daily hemisphere CMAQ model results are used to calculate the boundary condition and initial condition (Hogrefe et al., 2021). The chemical mechanism is Carbon Bond 06 revision 4 (CB6r4) (Ramboll, 2020a). Figure 1 in manuscript shows that the model domain is 12 km × 12 km (blue rectangle). We also 4 km × 4 km (red rectangle) nesting simulation to enhance the model spatial resolution through the flexi-nesting method. The point source emissions are processed independently with their stack locations in the model domain

and considering the plume-raising effect by stack parameters. As a result, the model spatial allocations can be enhanced through the flexi-nesting method.

The modeling ozone evaluation results over the simulation period are shown in Table S2. The evaluation indicators followed the USEPA's model evaluation guidance (USEPA, 2006). The modeling ozone is fairly performed well over the Gulf region states (Correlation Coefficient (R) ≥ 0.55). The simulated ozone in Texas and Louisiana are close to the observation data in USEPA AQS stations (Texas: R = 0.79, NMB = 1%; Louisiana: R = 0.77, NMB = 11%). Additionally, because our model shares the same simulation period as the Texas Commission on Environmental Quality (TCEQ) 2012 Ozone SIP modeling application (TCEQ, 2016), we verified our modeling results with TCEQ's simulated OH radical-related model species, including O_3, NO_2, and formaldehyde over the Dallas and Houston region (Fig. S2). The detailed comparisons are shown in the supplementary documents and indicate that both modeling applications share a similar, good modeling performance.

2.2.2 Reactive Tracer

The overall research method scheme flowchart is shown in Fig. 2. After developing the CTM-ready emissions and CAMx model for oxidants species (OH, O_3 and NO_3), the "Reactive Tracer" (RTRAC) was used to simulate the ambient SBTEX concentration over the Gulf region. The "Reactive Tracer" is a probing tool in the CAMx modeling system to simulate explicit SBTEX concentrations. Along with the physical transport processes (diffusion and advection) and decay processes like wet and dry deposition same as the core model, there is the second-order chemical reduction rate r that is calculated using the oxidants (Ozone, OH, NO_3) concentrations [Ox], the SBTEX concentrations [Tr], and the rate constants of reactions k_{Tr+Ox} (Eq.3). In Eq.4, k is the rate constant calculated by A, B, temperature (T), and activation energy (E_a). The Master Chemical Mechanism for aromatic schemes (Bloss et al., 2005) is considered for the parameters of each specific reaction in the RTRAC process.

This study considered the initial reactions of SBTEX in the MCM mechanism version 3.3.1 (Jenkin et al., 2015). For other parameters, the National Institute of Standards and Technology (NIST) Chemistry Webbook (P.J. Linstrom and W.G. Mallard, 2018), and CAMx user guide (RAMBOLL,

263 2020b) are considered for determining the Henry's Law constant, dependence temperature, and
264 molecular weight. All parameters used in our RTRAC modeling are presented in Tables S3 and
265 S4.

266 $$r = k_{Tr+Ox}\,[Tr][Ox] \qquad \text{Eq. (3)}$$

267 $$k = A\left(\frac{T}{300}\right)^{B} \exp\left(\frac{-E_a}{T}\right) \qquad \text{Eq. (4)}$$

268 The simulated SBTEX concentration will be evaluated by comparing against with observational
269 data which will be described in the following.

270 **Ambient SBTEX Measurements**

271 The CAMx modeling evaluation was completed with the USEPA Air Quality Station (AQS) ozone
272 observational data and the Texas Commission on Environmental Quality (TCEQ) State
273 Implementation Plan (SIP) ozone modeling output data (TCEQ, 2015). The measured ambient
274 SBTEX concentrations are from the USEPA Ambient Monitoring Technology Information Center
275 (AMTIC), which is an observational network that routinely detects more than 100 air toxics in the
276 U.S. (USEPA, 2021c). It includes the federal and state monitoring stations. The 5 months (May to
277 September 2012) individual SBTEX concentrations from the AMTIC were utilized to evaluate the
278 RTRAC modeling results from CAMx.

279 A total of 46 monitoring sites measure SBTEX concentrations within our 4 km × 4 km model
280 domain, and most of them are located within Texas (42 sites), except for four sites in Louisiana.
281 The air sampling duration can be 1-hour, 3-hour, or 24-hour. There are six monitoring sites with
282 1-hour measurement data in Texas, three sites with 3-hour data in Louisiana, and the rest with 24-
283 hour data. The AMTIC sites are indicated in Fig. 1 with red stars. The USEPA conducted QA/QC
284 for the AMTIC data, which contain values that are exceptionally high due to unpredictable
285 industrial VOC release events (Couzo et al., 2012). These events are beyond the regulatory
286 emission counting; thus, the model cannot capture those unpredictable events, particularly in
287 petrochemical, oil, and gas industrial areas. Therefore, this study removed outliers (those beyond
288 twice the interquartile range (2×IQR) above Q3) for better evaluating the model performance in
289 simulating the SBTEX concentration in general.

290 The CAMx RTRAC modeling results are spatially and temporally resolved gridded hourly
291 concentrations, while the AMTIC observational data are from specific locations with time gaps.
292 Daily average and diurnal pattern analyses evaluate the predicted SBTEX concentrations. For each
293 AMTIC site, this study used the average concentration of the center grid cell and eight other
294 "surrounding" grid cells (i.e., the average of 3×3 grid cells) to compare with the observational data
295 (USEPA, 2006).

296

3. Results

3.1 SBTEX Emissions

The 2012 annual total SBTEX emissions in the model domain are shown in Table 1. The emission sectors include: agriculture fire (afgire), commercial marine vessel (cmv), non-point source (nonpt), non-road vehicle (nonroad), on-road vehicle (onroad), fire emission (ptfire), rail road (rail), residential wood combustion (rwc), non-point oil gas industry (np_oilgas), electricity power plants unit (ptegu), point source emission other than electricity generation unit (ptnonipm), and point source of oil and gas industry (pt_oilgas). The largest contributor of SBTEX emissions in the 12km×12km model domain is indicated to be from the "onroad" sector, with 89,204 t yr^{-1}, representing about 36% of the total SBTEX emissions. The "onroad" sector contributes the most of total xylenes (46%), toluene (48%), and ethylbenzene (60%) emissions, while much less to benzene (13%) and styrene (6.8%). The second largest contributor of SBTEX emissions is the "wildfire" sector (61,316 t yr^{-1}), contributing about 25% of total SBTEX. The wildfire contributes the most of total benzene (57%), 12% of total toluene and 7% of total xylenes, but no ethylbenzene and styrene due to missing explicit profiles in the 2012 wildfire emission inventory. The "nonroad" sector ranked third (35,375 t yr^{-1}), contributing about 14% of total SBTEX over our modeling region. The nonroad contributes largely to xylenes (15%), toluene (21%), and ethylbenzene (21%). Compared to other sectors, emissions from non-electricity generation unit industrial point sources (ptnonipm) contain a larger portion of styrene, 2,911 t yr^{-1}, which is 69% of total styrene emission. Our study successfully identified missing styrene emissions from the chemical industry process (see table S7), leading to a 34% increase in total styrene emissions.

The individual and total SBTEX annual emission spatial plots in 12km×12km model domain are presented in Fig.3. The grid cell with the highest SBTEX emissions is found in Houston city near the ship channel (1059 t yr^{-1}), which is about 35 times higher than average emission (28 t yr^{-1}) across the domain, followed by one in San Antonio in Texas (1022 t yr^{-1}) and one near Sabine Lake in Louisiana (1022 t yr^{-1}). In Fig.3 (b), the missing sources of SBTEX emissions in the NEI are mostly located in Texas and Louisiana, particularly for the grid cells in Lake Charles (increased by 373 t yr^{-1}, +282%), Baton Rouge (167 t yr^{-1}, +31%) in Louisiana; Belton (61 t yr^{-1}, +21%), Fort Worth (50 t yr^{-1}, +85%), Dallas (44 t yr^{-1}, +52%) in Texas, and some rural areas in Texas.

These missing sources of SBTEX are mostly from the np_oilgas and ptnonipm emission sectors (detailed in Supplementary document 3.1 and 3.2). Although the total SBTEX emission increased by only 2% based on the domain average (Table 1), the localized impacts for certain areas can be up to 60% of the total SBTEX emissions.

The SBTEX emissions exhibit strong diurnal variations across a day, as presented in Fig. 4a. The daytime hourly emission (up to 77 t hr^{-1}) is about 4.3 times higher than the night-time emission rate, mainly due to the larger emissions from on-road and off-road mobile sources (half of total emissions) during the daytime. The diurnal variations in the chemical composition of total SBTEX also suggested the increased percentage of toluene and xylenes (indicating the transport sources) during the morning (L.T. 6:00 – 10:00) and evening (L.T. 19:00) rush hour. The inclusion of the missing sources will slightly reduce the emission variation across a day, as most of the missing sources come from industrial manufacturing and oil processes (detailed in SI) whose diurnal profiles are much flatter (about only 20% increase during the daytime) compared to the total emission (see Fig. 4b), with much smaller differences between day (0.86 t hr^{-1}) and night (0.69 t hr^{-1}). The chemical composition of missing emission sources were relatively constant throughout the day with about 50% comprised of xylenes, 30% toluene, and styrene was 10-15%. The relative amount of missing styrene was higher than that found in total emissions.

3.2 Model performance

CAMx simulations predicting SBTEX concentrations were completed using two sets of emissions: the National Emission Inventory (Base), and the emission scenario adjusted in this study (Adj). The differences between the two scenarios can be regarded as the impacts of the missing emission sources in the original NEI, suggesting the importance of the completeness of emissions.

The simulated concentrations were compared with the observations to evaluate the accuracy of SBTEX emission and concentration estimated in this study. The normalized mean bias (NMB, %) and correlation coefficient (R) of both Base and Adj cases were compared in Table 2. Overall, the CAMx model can capture the pollution level and spatiotemporal variation of all SBTEX species. More specifically, the model reproduced the daily variation of SBTEX concentrations, with R of 0.65 (0.54-0.65 for individual SBTEX) for all daily observational records (N=2,717), as well as

their spatial distribution across observational sites (N=46, averages of the whole simulation period), with R of 0.75 (0.46 to 0.77 for individual SBTEX species), and NMB of -5.6% (-24.9% to 32.1% for individual SBTEX species).

The inclusion of emissions can slightly improve the overall model performance with decreased NMBs for toluene (+3.5%), xylenes (+5.7%), ethylbenzene (+3.8%), and total SBTEX (+3.2%). The NMB for styrene is increased from 17.4% to 32.1%, while R is increased by 0.01, suggesting better correlations with the new-estimated emission data, while uncertainties associated with emission factors or other parameters lead to the overestimation of SBTEX. Fig. 5 shows the spatial distribution of average concentration simulated in the Adj case, overlapping the average observational data for total SBTEX (5a) and individual species (5b to 5f). The observational data (diamond shapes) shows a high concentration in industrial or city sites, and a lower concentration at rural sites. The model results showed a continual concentration gradient pattern from cities to a rural area with 4 km × 4 km resolution and the results are close to the observational data in Houston, Dallas, Beaumont, and Baton Rouge.

We further classified the observation sites into four groups, including "Airport", "Industry", "Rural", and "Urban" based on their geographical locations (Table S8). For total SBTEX (Fig. 6a), the correlation coefficient (R) is 0.75 (R-square is 0.56) across all locations, and the black solid line is the regression line for all sites (N=46). The red dots indicated that the industrial sites have a higher concentration in both model and observational results, and the cities (blue diamonds) showed that their concentrations are slightly overestimated and lower than industrial sites. The Airport (black squares) and Rural (green triangles) have lower SBTEX concentrations than City and Industry, and Rural is the lowest group. Fig. 6b to 6f are similar plots for explicit benzene, toluene, xylenes, ethylbenzene, and styrene. The R ranges from 0.46 to 0.77. The benzene (R is 0.68), toluene (R is 0.46), and styrene (R is 0.64) are overestimated, but xylenes (R is 0.77), ethylbenzene (R is 0.77) are close to observational data. Although toluene has the lowest R (0.46), which is caused by two industry sites that largely underestimate in Houston (Site ID: 482011015) and Nederland (Site ID: 482450014), in case we remove those two industrial sites data, the R for toluene in Fig. 6c will become 0.7 (Fig. S7). This phenomenon is probably caused by the missing toluene industrial sources near those two sites. The inclusion of missing emission sources definitively improved the model performance (Table 2), especially in Rural (+5.4%) and Airport

groups (+6.8%) which suffered the most due to the missing industrial sources. The NMBs for xylenes are also reduced across all emission groups (Industry: +3%, Urban: +12%, Airport: +20%, and Rural: +13%).

Because only few sites have hourly data, this study compared the diurnal variation of SBTEX concentrations for Houston industry area (using data from only three monitoring sites) in Fig. S8. The hourly data shows that benzene, ethylbenzene, and styrene are overestimated (NMB for benzene is 69%, ethylbenzene is 36%, and styrene is 27%) during nighttime hours (LT 21:00 to 6:00). Toluene is underestimated at nighttime (NMB is -45%), whereas xylenes closely aligns with the observed data (-16%) range. On the other hand, all species experience underestimation during daytime hours (NMB for benzene is -25%, Toluene is -65%, Xylenes is -51%, ethylbenzene is -46%, and Styrene is -82%) (from LT 10:00 to 17:00). Such results indicate that the hourly emission rate may overestimate during nighttime but underestimate during the daytime in Houston industry area.

3.3 SBTEX Concentrations Patterns

3.3.1 Spatial Distribution

Figure 7a presents the spatial distribution of SBTEX concentration during the model period (May 1st to Sep 30th) in the Adj scenario. The highest SBTEX concentration (3.07 ppb) occurs near Lake Charles, followed by Baton Rouge (2.06 ppb), Houston ship channel (2.04 ppb), Shreveport (1.69 ppb), Beaumont (1.59 ppb). The spatial distribution patterns of individual SBTEX compounds exhibit similarities due to shared emission sources, except for styrene. Styrene primarily originates from ptnonipm, while other species predominantly arise from vehicle emissions and wildfires. benzene (max: 1.06 ppb), toluene (max: 1.01 ppb), and ethylbenzene (max: 0.16 ppb) reach their highest concentrations in Houston, reflecting their significant emissions. Further, xylenes (0.78 ppb) originate from sources in Shreveport. Remarkably elevated concentrations of styrene (reaching 1.97 ppb) are conspicuously identified proximal to Lake Charles, a locale characterized by an abundant emission of styrene from non-Electricity Generating Unit point sources, which have been absent in the original NEI records.

This study further investigated the influence of missing emission sources in the original NEI on the SBTEX concentrations by taking the differences between Adj and Base scenarios. The majority of missing emissions are associated with the np_oilgas and ptnonipm sectors, with increased contributions geographically concentrated in Texas and Louisiana (Fig. 7b). In particular, the largest impact on SBTEX concentration is shown near Lake Charles by up to 1.82 ppb (+68%), which is mostly related to the increase of styrene concentration (by 1.75 ppb, +5315%). This increase is due to the NEI missing one large point source (364.12 t yr^{-1}) in the ptnonipm sector near Lake Charles. The inclusion of missing emission sources also led to the increase of styrene concentrations in other cities, such as Baton Rouge (0.07 ppb, +389%), LA, and Houston, TX (0.03 ppb, +62%). Baton Rouge, LA also suffers the largest increase of toluene concentrations by 0.44 ppb (+92%) due to the inclusion of missing emissions, followed by Beaumont (0.07 ppb, +50%), and Carthage (0.048 ppb, +66%) in TX. Fort Worth, TX exhibits the most increase of xylenes concentrations by 0.07 ppb (+95%), followed by Center (0.06 ppb, +273%), Teague (0.06 ppb, +340%), and Beaumont (0.036 ppb, +70%) in TX. The largest increase of ethylbenzene concentration occurred at Longview (0.01 ppb, +85%), followed by Beaumont (0.009 ppb, +40%) and Houston (0.006 ppb, +9%) in TX.

3.3.2 The Diurnal Variation

In general, the diurnal variations of SBTEX concentrations are primarily influenced by various factors (such as ventilation, emissions, diffusion, deposition, and chemical reactions). These variations typically manifest with lower concentrations during the daytime compared to nighttime due to increased ventilation, diffusion, and chemical loss, even though emissions are about 4 times higher during the daytime, as presented earlier (Fig. 4). Diurnal meteorological and emission patterns suggest more sensitivity of the concentrations to the emissions during nighttime than daytime, implying that implementing emission controls to reduce the concentrations at night would be most effective. The variation of emission sources might also modulate the diurnal pattern in concentrations. To demonstrate that, here we selected two industrial locations and one city location with high SBTEX concentrations to compare the diurnal variation of concentrations.

The first one is Channelview city (Latitude: 29.8, Longitude: -95.12), located at the Houston ship channel industrial area on the eastern side of downtown Houston. Driven by both emission

temporal profiles and meteorological conditions, the peak SBTEX concentration (about 12 ppb) in Channelview city occurs at LT 23:00 to 1:00, contributed mostly by benzene (56%) which indicates the industrial sources, with a small amount of toluene (19%), xylenes (13%), styrene (4.8%), ethylbenzene (7%) (Fig. 8a). The second case, Bayland Park (Latitude: 29.69, Longitude: -95.49), located nearby at the western side of Houston, presents the same level of peak SBTEX concentration (about 12 ppb) (Fig.9a) as Channelview city. In contrast to Channelview, the peak concentration of Bayland Park occurs at traffic rush hour (LT 7:00 to 8:00), contributed mostly by toluene (53%) and xylenes (23%) (indicating the mobile vehicle sources) rather than benzene (18%). Meanwhile, the adjusted industry emission sources, as presented in table S5, play a significant role in driving the peak concentration (0.4 ppb) in Channelview city (Fig. 8b), yet exhibit a reduced impact on Bayland Park (Fig. 9b), where is far from the industry area.

A similar pattern is also shown in Baton Rouge, Louisiana (Latitude: 30.46, Longitude: -91.17), located near downtown Baton Rouge (affected by onroad sources), and also close to the industry area (~ 1 mile from the north). Like Houston industry area, the daytime SBTEX concentration is much lower (<3 ppb) than night-time, and the peak SBTEX concentration (about 9.4 ppb) occurs at LT 22:00 (Fig. 10). Because Baton Rouge is impacted by both traffic and industrial sources, emissions differ from Houston in that both benzene (35% – 40%) and toluene (35% - 40%) become the major portion of SBTEX (Fig. 10a). The missing emission sources (Fig. 10b) will further enhance the peak concentration by 2 ppb at LT 5:00 - 8:00, with the largest chemical contribution from toluene (about 70 - 85%), followed by styrene (about 7 - 20%) associated with the industrial sources.

4. Conclusion and Discussion

To address the urgent need for health assessment of SBTEX exposures in the Gulf region, this study developed high spatiotemporally resolved emissions and concentrations of individual SBTEX. The HAPI program was developed and implemented to identify and gap-fill the missing SBTEX inventory for the SMOKE emissions modeling system. Then the state-of-the-science chemical transport modeling system, CAMx, was applied to generate the high temporal and spatial resolution predictions of explicit SBTEX concentrations based on the improved SBTEX emission inventory and a "Reactive Tracer" (RTRAC) feature. The modeled average SBTEX concentrations

exhibit good agreement with observational data (R is 0.75 and NMB is improved in Adj case to -5.6% for total SBTEX), suggesting that the emissions and concentrations estimates developed in this study can be used to support well the SBTEX-related human health studies in the Gulf region.

This study found that the "onroad" sector contributes the most to total xylenes (46%), toluene (48%), and ethylbenzene (60%) emissions, while the styrene emissions are mostly contributed by non-EGU point sources (ptnonipm, 69%) but were substantially underestimated in the original NEI data, resulting in 34% underestimation of total styrene emissions. The highest SBTEX concentration (3.07 ppb) occurs near Lake Charles, followed by Baton Rouge (2.06 ppb), Houston ship channel (2.04 ppb), Shreveport (1.69 ppb), Beaumont (1.59 ppb), corresponding to a large amount of SBTEX emissions in these cities.

The 5-month average SBTEX modeled concentrations are close to the average measurement data (R of total SBTEX is 0.74, benzene is 0.68, toluene is 0.45, xylenes is 0.77, ethylbenzene is 0.77, and styrene is 0.64). These spatiotemporally fine modeled air SBTEX concentrations can be used for conducting epidemiologic analyses or in risk assessment. The diurnal variation of SBTEX concentrations that is opposite to its emissions pattern indicates that the concentration is more sensitive to emission at night than daytime. The high SBTEX concentration during nighttime affects individuals who engage in more nighttime activities or reside in houses lacking isolation of outdoor air. Therefore, the HAPs emission control policy should also focus on night-time emissions. Further, the hourly SBTEX data can be used in epidemiologic analyses to investigate effects of acute exposures and short-term changes in those exposures.

This study acknowledges the considerable uncertainties in this approach, including the accuracy of emission data, the meteorological condition data, and oxidants concentrations (OH radical, O_3, and NO_3) simulation in the CB6 mechanism. There are limited observational data to verify the model performance. This study is mainly based on the bottom-up NEI dataset, thus the uncertainties in the original NEI emissions and SMOKE process influenced on this study. For example, despite our implementation of imputation for the HAPs annual data, the emission activity within hourly, daily, and monthly temporal profiles as well as parameters (e.g., emission rates, and compositions) also remains unchanged. The emergency emissions from unreported flaring (such as final treatment equipment) or leakage events that have not been considered in the original NEI,

also not included in this study. Further, the concentrations of oxidants are simulated in the CAMx model with the CB6r4 mechanism; this mechanism is designed to simulate ozone and PM. Therefore, the model species OH radical, NO_3, and O_3 may differ from the actual concentrations. These oxidant concentrations affect the chemical decay rate, especially in big metropolitan cities with higher NOx emissions. Nevertheless, the high spatiotemporally resolved emissions and concentrations of individual SBTEX developed in this study, with acceptable performance, can be a good reference dataset to support SBTEX-related human health studies in the Gulf region. Besides, this approach can be extended to other chemical compounds to estimate their concentrations. The USEPA provides emission data for approximately one hundred HAPs in the NEI for certain emission sources. Those emission data can also be processed to derive HAPs concentrations. The dataset provided in this study will facilitate epidemiologic studies of SBTEX exposures in relation to a range of health outcomes in the Gulf region and can be extended to provide similar health research opportunities elsewhere.

Code availability:

1. The source code of the CAMx7.00 model and model preprocess tools (O3map, tuv4.8, wrfcamx, camq2camx) can be downloaded on the Environ website: http://www.camx.com (RAMBOLL, 2021)
2. Python 2.7 is used to treat the model output and can be downloaded on anaconda python website: https://www.anaconda.com/distribution/ (Anaconda, 2020)
3. R project for statistical computing can be downloaded at https://www.r-project.org (The R Foundation, 2021)
4. HAPI program code can be downloaded on GitHub: https://github.com/tatawang/HAPI (Wang and Baek, 2023)

Data availability:

1. The result of this study, including SBTEX emission, concentration data and evaluation code in this study can be downloaded at: https://zenodo.org/record/8303346, DOI: 10.5281/zenodo.8303346 (Wang et al., 2023)
2. Beside the 2019 May to September, we also provided the whole 2012 SBTEX hourly concentration data of Adj case in NetCDF format and "comma-separated values (csv)".
3. The 2011 NEI emission model platform (EMP) and SMOKE model system can be downloaded on the EPA ftp website: https://www.epa.gov/air-emissions-modeling/2011-version-6-air-emissions-modeling-platforms (USEPA, 2021b)
4. The meteorological data can be found on CMAS Data Warehouse website: https://dataverse.unc.edu/dataverse/cmascenter (UNC-IE, 2021)
5. The AMTIC data can be found at: https://www.epa.gov/amtic/amtic-ambient-monitoring-archive-haps (USEPA, 2021c)

Author contribution

CTW and BHB are the lead researchers in this study and are responsible for research design, producing data, experiments, results analysis, and manuscript writing. WV and JX are co-head researchers and guided the research design, assessed model results, and contributed to writing

the manuscript. JG, MS, RS, LE, JB, and JHW helped to collect and verify data and write the manuscript.

Competing interests

The Authors declare that they have no conflict of interest.

Acknowledgments

We want to thank the National Institute of Environmental Health Sciences (NIEHS) support this study, the research project is "Neurological Effects of Environmental styrene and BTEX Exposure in a Gulf of Mexico Cohort" (Grant No. R01ES031127). We appreciate other grant support by NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle, and Climate (AC4) program and Climate Observations and Monitoring (COM) program, Grant No. NA21OAR4310225 (GMU) / NA21OAR310226 (UMD), and the Fine Particle Research Initiative in East Asia Considering National Differences (FRIEND) project through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2020M3G1A1114621). This work was also supported by Korea Environment Industry & Technology Institute (KEITI) through Project for developing an observation-based GHG emissions geospatial information map, funded by Korea Ministry of Environment (MOE) (RS-2023-00232066). We also thank the Texas Commission on Environmental Quality (TCEQ), South Coast Air Quality Management District (AQMD), Ramboll CAMx team, and the University of North Carolina at Chapel Hill (UNC-CH) for their invaluable assistance.

Reference

Al-Harbi, M., Alhajri, I., AlAwadhi, A., and Whalen, J. K.: Health symptoms associated with occupational exposure of gasoline station workers to BTEX compounds, Atmospheric Environment, 241, 117847, 2020.

Anaconda: Anaconda python: https://www.anaconda.com/products/individual, last access: May, 1st, 2020.

ATSDR: Toxicological Profile for benzene, https://www.atsdr.cdc.gov/toxprofiles/tp3.pdf, 2007a.

ATSDR: Toxicological Profile for xylene, https://www.atsdr.cdc.gov/toxprofiles/tp71.pdf, 2007b.

ATSDR: Toxicological Profile for ethylbenzene, https://www.atsdr.cdc.gov/toxprofiles/tp110.pdf, 2010a.

ATSDR: Toxicological Profile for styrene, https://www.atsdr.cdc.gov/toxprofiles/tp53.pdf, 2010b.

ATSDR: Toxicological Profile for toluene, https://www.atsdr.cdc.gov/toxprofiles/tp56.pdf, 2017.

Baek, B. H. and Seppanen, C.: SMOKE v4.8.1 Public Release (January 29, 2021) (Version SMOKEv481_Jan2021): http://doi.org/10.5281/zenodo.4480334 last 2021.

Bloss, C., Wagner, V., Jenkin, M. E., Volkamer, R., Bloss, W. J., Lee, J. D., Heard, D. E., Wirtz, K., Martin-Reviejo, M., Rea, G., Wenger, J. C., and Pilling, M. J.: Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons, Atmos. Chem. Phys., 5, 641-664, 10.5194/acp-5-641-2005, 2005.

Chen, W.-H., Chen, Z.-B., Yuan, C.-S., Hung, C.-H., and Ning, S.-K.: Investigating the differences between receptor and dispersion modeling for concentration prediction and health risk assessment of volatile organic compounds from petrochemical industrial complexes, Journal of environmental management, 166, 440-449, 2016.

Choi, K.-C., Lee, J.-J., Bae, C. H., Kim, C.-H., Kim, S., Chang, L.-S., Ban, S.-J., Lee, S.-J., Kim, J., and Woo, J.-H.: Assessment of transboundary ozone contribution toward South Korea using

multiple source–receptor modeling techniques, Atmospheric Environment, 92, 118-129, https://doi.org/10.1016/j.atmosenv.2014.03.055, 2014.

Christensen, M. S., Vestergaard, J. M., d'Amore, F., Gørløv, J. S., Toft, G., Ramlau-Hansen, C. H., Stokholm, Z. A., Iversen, I. B., Nissen, M. S., and Kolstad, H. A.: styrene exposure and risk of lymphohematopoietic malignancies in 73,036 reinforced plastics workers, Epidemiology, 29, 342-351, 2018.

Couzo, E., Olatosi, A., Jeffries, H. E., and Vizuete, W.: Assessment of a regulatory model's performance relative to large spatial heterogeneity in observed ozone in Houston, Texas, Journal of the Air & Waste Management Association, 62, 696-706, 10.1080/10962247.2012.667050, 2012.

Declet-Barreto, J., Goldman, G. T., Desikan, A., Berman, E., Goldman, J., Johnson, C., Montenegro, L., and Rosenberg, A. A.: Hazardous air pollutant emissions implications under 2018 guidance on US Clean Air Act requirements for major sources, Journal of the Air & Waste Management Association, 70, 481-490, 2020.

Doherty, B. T., Kwok, R. K., Curry, M. D., Ekenga, C., Chambers, D., Sandler, D. P., and Engel, L. S.: Associations between blood BTEXS concentrations and hematologic parameters among adult residents of the U.S. Gulf States, Environ Res, 156, 579-587, 10.1016/j.envres.2017.03.048, 2017.

Esri, D., HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), Swisstopo, MapmyIndia, and the GIS User Community: World Topographic Map, 2013.

Hogrefe, C., Gilliam, R., Mathur, R., Henderson, B., Sarwar, G., Appel, K. W., Pouliot, G., Willison, J., Miller2, R., Vukovich1, J., Eyth, A., Talgo, K., Allen, C., and Foley, K.: CMAQv5.3.2 ozone simulations over the Northern Hemisphere: model performance and sensitivity to model configuration

Office of Research and Development, 2021.

Hsieh, M.-T., Peng, C.-Y., Chung, W.-Y., Lai, C.-H., Huang, S.-K., and Lee, C.-L.: Simulating the spatiotemporal distribution of BTEX with an hourly grid-scale model, Chemosphere, 246, 125722, 2020a.

Hsieh, M. T., Peng, C. Y., Chung, W. Y., Lai, C. H., Huang, S. K., and Lee, C. L.: Simulating the spatiotemporal distribution of BTEX with an hourly grid-scale model, Chemosphere, 246, 10.1016/j.chemosphere.2019.125722, 2020b.

Jenkin, M. E., Young, J. C., and Rickard, A. R.: The MCM v3.3.1 degradation scheme for isoprene, Atmospheric Chemistry and Physics, 15, 11433-11459, 10.5194/acp-15-11433-2015, 2015.

Lindhjem, C., Chan, L., Pollack, A., Corporation, E. I., Way, R., and Kite, C.: Applying Humidity and Temperature Corrections to On and Off-Road Mobile Source Emissions, 13th International Emission Inventory Conference, Clearwater, FL2004.

Miller, L., Xu, X. H., Wheeler, A., Zhang, T. C., Hamadani, M., and Ejaz, U.: Evaluation of missing value methods for predicting ambient BTEX concentrations in two neighbouring cities in Southwestern Ontario Canada, Atmospheric Environment, 181, 126-134, 10.1016/j.atmosenv.2018.02.042, 2018.

Moshiran, V. A., Karimi, A., Golbabaei, F., Yarandi, M. S., Sajedian, A. A., and Koozekonan, A. G.: Quantitative and semiquantitative health risk assessment of occupational exposure to styrene in a petrochemical industry, Safety and health at work, 12, 396-402, 2021.

NASA: OZONE & AIR QUALITY: https://ozoneaq.gsfc.nasa.gov, last 2021.

NCHS: National Center for Health Statistics (NCHS) National Health and Nutrition Examination Survey Data.: https://www.cdc.gov/nchs/nhanes/index.htm, last 2021.

Niaz, K., Bahadar, H., Maqbool, F., and Abdollahi, M.: A review of environmental and occupational exposure to xylene and its health concerns, EXCLI journal, 14, 1167, 2015.

NIEHS: GuLF STUDY: https://gulfstudy.nih.gov/en/index.html, last access: 16, November, 2021.

O'Leary, B. F. and Lemke, L. D.: Modeling spatiotemporal variability of intra-urban air pollutants in Detroit: A pragmatic approach, Atmospheric Environment, 94, 417-427, 10.1016/j.atmosenv.2014.05.010, 2014.

P.J. Linstrom and W.G. Mallard, E.: NIST Chemistry WebBook, NIST Standard Reference Database Number 69: https://webbook.nist.gov/chemistry/, last 2018.

Pankow, J. F., Luo, W. T., Bender, D. A., Isabelle, L. M., Hollingsworth, J. S., Chen, C., Asher, W. E., and Zogorski, J. S.: Concentrations and co-occurrence correlations of 88 volatile organic compounds (VOCs) in the ambient air of 13 semi-rural to urban locations in the United States, Atmospheric Environment, 37, 5023-5046, 10.1016/j.atmosenv.2003.08.006, 2003.

Polvara, E., Roveda, L., Invernizzi, M., Capelli, L., and Sironi, S.: Estimation of Emission Factors for Hazardous Air Pollutants from Petroleum Refineries, Atmosphere, 12, 1531, 2021.

Rahimpoor, R., Sarvi, F., Rahimnejad, S., and Ebrahimi, S. M.: Occupational exposure to BTEX and styrene in West Asian countries: a brief review of current state and limits, Archives of Industrial Hygiene and Toxicology, 73, 107-118, 2022.

Rajabi, H., Mosleh, M. H., Mandal, P., Lea-Langton, A., and Sedighi, M.: Emissions of volatile organic compounds from crude oil processing–Global emission inventory and environmental release, Science of The Total Environment, 727, 138654, 2020.

Ramboll: Speciation Tool User's Guide: https://www.cmascenter.org/speciation_tool/documentation/5.0/Ramboll_sptool_users_guide_V5.pdf, last 2020a.

RAMBOLL: CAMx7.00 User's Guide: https://camx-wp.azurewebsites.net/Files/CAMxUsersGuide_v7.00.pdf, last 2020b.

RAMBOLL: Comprehensive Air Quality Model with Extensions, CAMx: https://www.camx.com, last access: Nov 29, 2021.

SRI: Directory of Chemical Producers: United States of America, https://books.google.com/books?id=Nx0vAQAAIAAJ, 2008.

Strum, M., Kosusko, M., Shah, T., and Ramboll: SPECIATE and using the Speciation Tool to prepare VOC and PM chemical speciation profiles for air quality modeling, https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=532731&Lab=NRMRL, 2017.

TCEQ: 2015 Ozone NAAQS Transport SIP Modeling (2012 Episode): https://www.tceq.texas.gov/airquality/airmod/data/gn/gn2012, last access: Nov 18, 2015.

TCEQ: TCEQ 2012 Modeling Platform Technical Support Document, https://wayback.archive-it.org/414/20210529054550/https://www.tceq.texas.gov/assets/public/implementation/air/am/modeling/gn/doc/TCEQ_2012_Modeling_Platform_TSD.pdf, 2016.

The R Foundation: The R Project for Statistical Computing: https://www.r-project.org, last access: Jan, 25th, 2022, 2021.

UNC-IE: CMAS data warehouse: https://dataverse.unc.edu/dataverse/cmascenter, last access: Jan, 25th, 2023, 2021.

USEPA: Integrated Risk Information System (IRIS) benzene Integrated Risk Information System: https://iris.epa.gov/ChemicalLanding/&substance_nmbr=276, last access: Dec 12, 2000.

USEPA: Guidance on the Use of models and other analyses for demonstrating attainment of air quality goals for ozone, PM2.5 and Regional Haze, https://permanent.fdlp.gov/LPS74329/draft_final-pm-O3-RH.pdf, 2006.

USEPA: 2014 Fire NEI Workshop Emissions Processing-SmartFire Details, https://www.epa.gov/sites/default/files/2015-09/documents/emissions_processing_sf2.pdf, 2015.

USEPA: Source Classification Codes (SCCs): https://sor-scc-api.epa.gov/sccwebservices/sccsearch/, last access: Dec 5, 2016.

USEPA: National Emissions Inventory (NEI) 2011 Version 6 Air Emissions Modeling Platforrm (EMP): https://www.epa.gov/air-emissions-modeling/emissions-modeling-platforms, last access: JAN, 27th, 2022, 2021a.

USEPA: 2011 Version 6 Air Emissions Modeling Platforms: https://www.epa.gov/air-emissions-modeling/2011-version-6-air-emissions-modeling-platforms, last access: April, 04, 2022, 2021b.

USEPA: Ambient Monitoring Technology Information Center (AMTIC): https://www.epa.gov/amtic, last access: Nov, 18, 2021, 2021c.

USEPA: Toxics Release Inventory (TRI) program: https://www.epa.gov/toxics-release-inventory-tri-program, last access: Nov 16, 2021d.

USEPA: User's Guide for the AMS/EPA Regulatory Model (AERMOD): https://gaftp.epa.gov/Air/aqmg/SCRAM/models/preferred/aermod/aermod_userguide.pdf, last access: August 15, 2023, 2022.

USEPA: Air Quality Dispersion Modeling: https://www.epa.gov/scram/air-quality-dispersion-modeling#:~:text=Dispersion%20modeling%20uses%20mathematical%20formulations,at%20selected%20downwind%20receptor%20locations., last access: August 15, 2023, 2023.

Wang, C.-T. and Baek, B. H.: Hazardous Air Pollutants Imputation (HAPI) (v1.0.1): https://doi.org/10.5281/zenodo.7987106, last access: May 30, 2023, 2023.

Wang, C.-T., Baek, B. H., Vizuete, W., and Engel, L. S.: The styrene, benzene, toluene, ethylbenzene and xylenes (SBTEX) hourly gridding modeled emission and concentration in the U.S. Gulf region (1), Zenodo [dataset], 10.5281/zenodo.7967541, 2023.

Werder, E. J., Engel, L. S., Blair, A., Kwok, R. K., McGrath, J. A., and Sandler, D. P.: Blood BTEX levels and neurologic symptoms in Gulf states residents, Environmental research, 175, 100-107, 2019.

Werder, E. J., Engel, L. S., Richardson, D. B., Emch, M. E., Gerr, F. E., Kwok, R. K., and Sandler, D. P.: Environmental styrene exposure and neurologic symptoms in US Gulf coast residents, Environment international, 121, 480-490, 2018.

Tables

Table 1. The Annual emission rates (metric tons yr^{-1}) of styrene, benzene, toluene, ethylbenzene, and xylenes (SBTEX) in 2012 including the increases resulting from this work. The percent increase from the 2012 National Emission Inventory is given in parentheses. The bold font indicates the emission sector with the maximum SBTEX rates.

Emission Sectors	**benzene tons yr^{-1}**	**toluene tons yr^{-1}**	**xylenes tons yr^{-1}**	**ethylbenzene tons yr^{-1}**	**styrene tons yr^{-1}**	**Total tons yr^{-1}**	**Sectoral share of total**
agriculture fire (agfire)	1,128	745	0	0	0	1,873	0.76%
commercial marine vehicle (cmv)	103	16	24	10	11	164	0.07%
non-point source (nonpt)	3,070	16,932	5,156	1,188	777	27,123	11%
non-road vehicle (nonroad)	4,752	13,506	14,265	2,682	171	35,376	14%
on-road vehicle (onroad)	10,495	**43,657**	**27,271**	**7,472**	309	**89,204**	36%
wild fire (ptfire)	**46,052**	10,909	4,355	0	0	61,316	25%
Rail (rail)	10	14	20	8	9	61	0.02%
residential wood combustion (rwc)	395	92	26	0	0	513	0.21%
non-point oil gas industry (np_oilgas)	5,421	2,694 (+69%)	4,683 (+51%)	455 (+100%)	2 (+100%)	13,255 (+28%)	5.4%
electricity power plants unit (ptegu)	277	131 (+2%)	60 (+3%)	35 (+3%)	7 (0%)	510 (+1%)	0.21%
point source emission other than electricity generation unit (ptnonipm)	7,305	2,608 (+17%)	2,644 (+12%)	667 (+12%)	**2,911** (+34%)	16,135 (+10%)	5.9%
point source emission of oil and gas industry (pt_oilgas)	510	314 (+25%)	209 (+24%)	36 (+24%)	2 (+100%)	1071 (+11%)	0.43%
Total	79,518	90,080 (+2%)	58,713 (+3%)	12,553 (+3%)	4,199 (+22%)	246,601 (+2%)	100%

Table 2. Normalized Mean Bias (NMB, %) and Correlation Coefficient (R) comparison of average observational data and model result during the model simulation period, May 1st, 2012 to Sep 30th, 2012 for the 2012 National Emission Inventory (Base), and the emission scenario adjusted in this study (Adj). Bold font indicates the model improvement, and gray color font indicates poorer model performance. Also shown is the count (N) of available daily average data across all sites.

	Group	N	benzene	toluene		xylenes		ethylbenzene		styrene		SBTEX	
				Base	Adj	Base	Adj	Base	Adj	Base	Adj	Base	Adj
R (daily average comparison for all sites)	All	2717	0.54	0.57	0.57	0.58	0.56	0.56	**0.56**	0.55	**0.57**	0.65	0.65
R (5 month average comparison for all sites)	All	46	0.68	0.46	0.46	0.79	0.77	0.76	**0.77**	0.63	**0.64**	0.75	0.75
NMB (%) (average comparison for all sites)	All	46	12.53	-10.2	**-6.7**	-30.6	**-24.9**	-25.2	**-21.4**	17.4	32.1	-8.8	**-5.6**
NMB (%) (daily average comparison for all sites)	Rural	508	-22.3	-10.6	**-5.4**	-33.2	**-19.8**	-26.8	**-23.0**	-63.9	**-54.8**	-19.3	**-13.9**
	Airport	95	-41.0	-4.5	**0.6**	-18.4	**1.5**	-26.0	**-19.5**	34.5	42.3	-11.8	**-5.0**
	Urban	272	61.7	82.9	87.5	-20.9	**-8.8**	17.0	19.9	-50.6	**-39.9**	32.6	39.3
	Industry	1842	88.0	-6.6	**-2.2**	-26.5	**-23.5**	-9.4	**-4.9**	54.6	76.1	15.5	19.0

Figures:

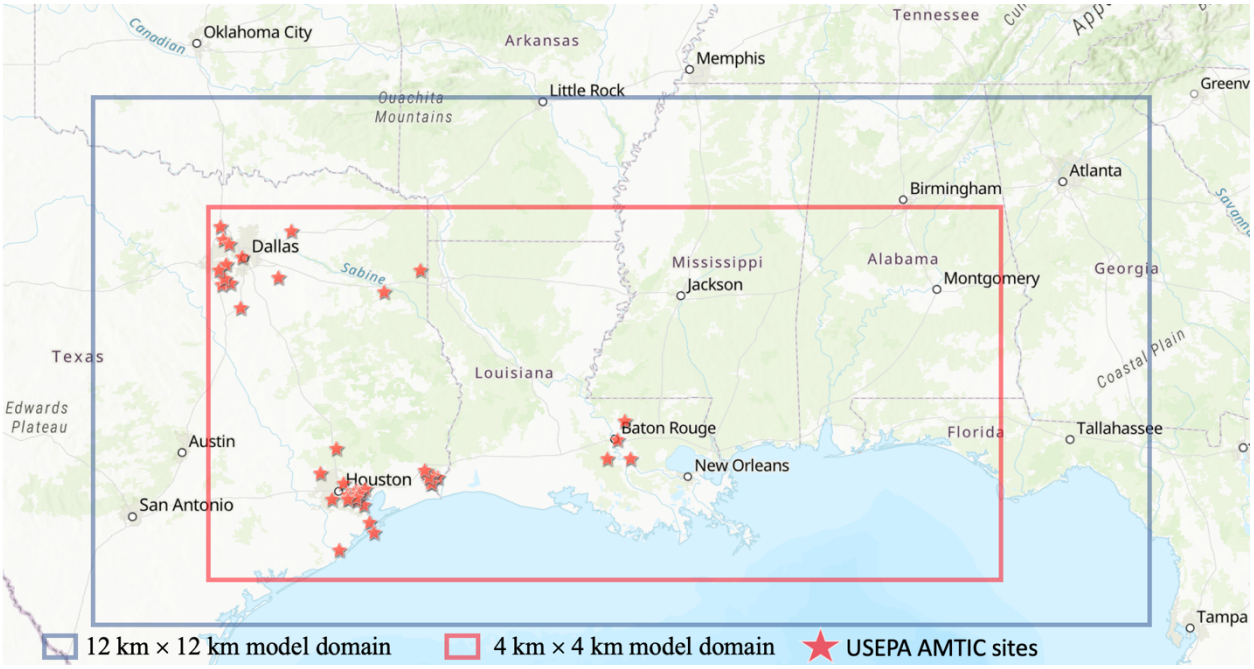


Figure 1. The modeling domains with the outer 12 × 12 km resolution domains (blue rectangle) and inner 4 × 4 km resolution domain (red rectangle). The red stars are the USEPA Ambient Monitoring Technology Information Center observational (AMTIC) sites for Hazardous Air Pollutants (HAPs). There are 4 sites are in Louisiana, and 42 sites in Texas. Generated with ArcGIS map (Esri, 2013).

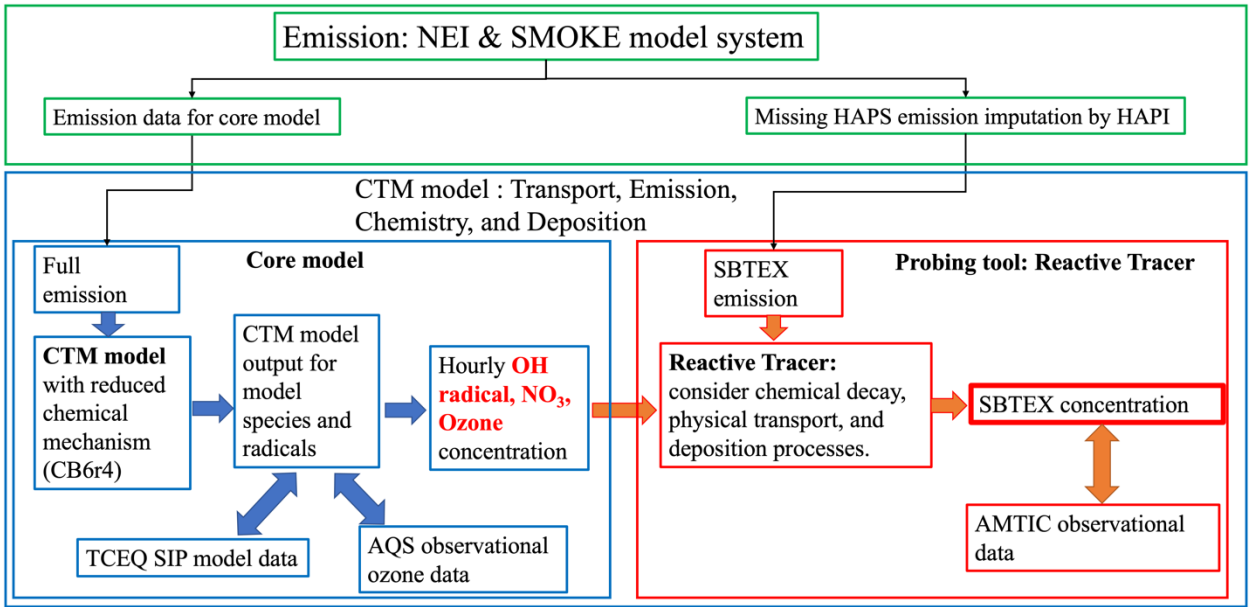
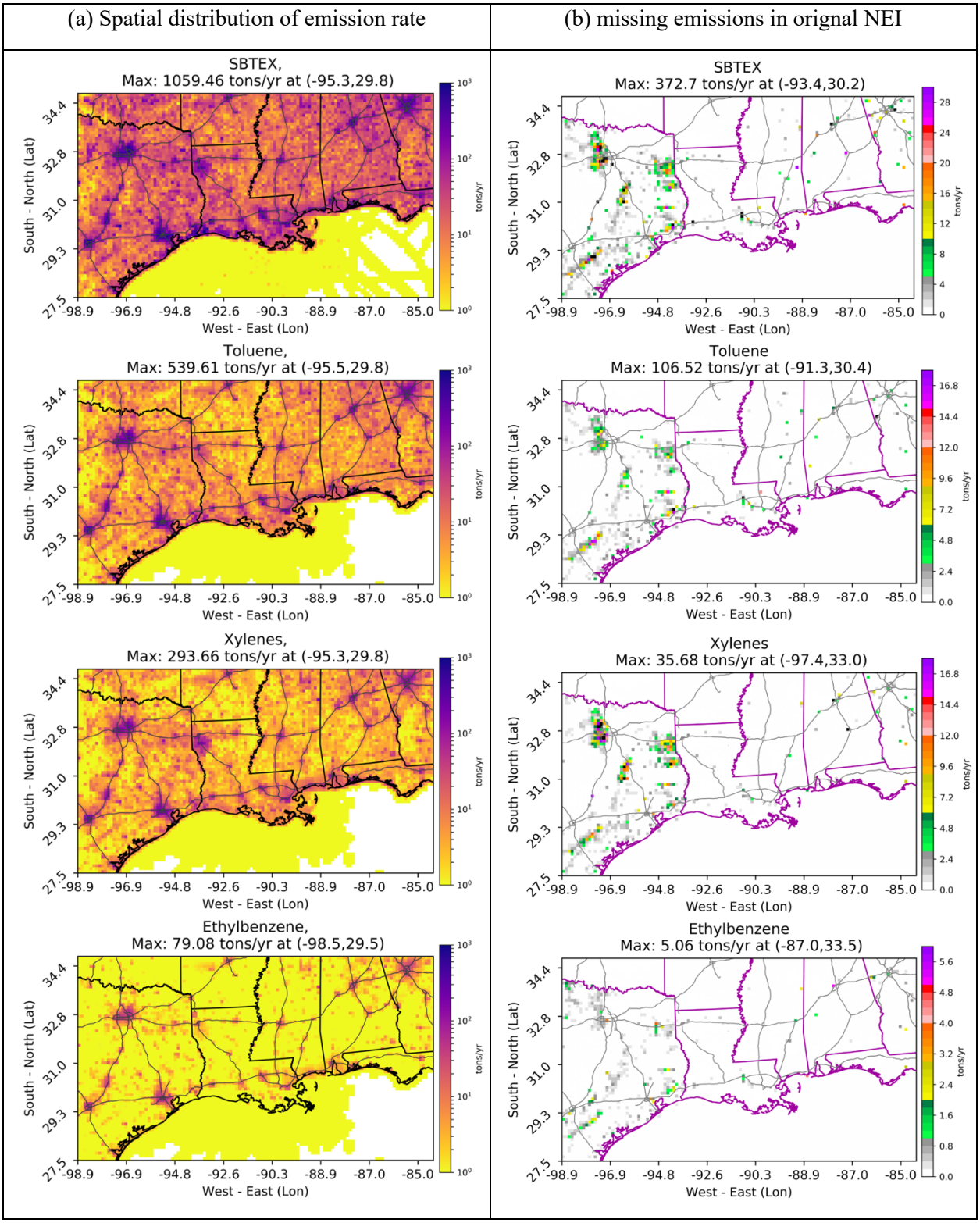


Figure 2. Toxic air quality modeling system schematic: The green rectangles are emission processes; the blue rectangles are the base CTM model process for estimating the concentration of oxidants; the red rectangles are the Reactive Tracer process for estimating individual SBTEX concentration.

755


(a) Spatial distribution of emission rate
(b) missing emissions in orignal NEI
SBTEX,
Max: 1059.46 tons/yr at (-95.3,29.8)
SBTEX
Max: 372.7 tons/yr at (-93.4,30.2)
Toluene,
Max: 539.61 tons/yr at (-95.5,29.8)
Toluene
Max: 106.52 tons/yr at (-91.3,30.4)
Xylenes,
Max: 293.66 tons/yr at (-95.3,29.8)
Xylenes
Max: 35.68 tons/yr at (-97.4,33.0)
Ethylbenzene,
Max: 79.08 tons/yr at (-98.5,29.5)
Ethylbenzene
Max: 5.06 tons/yr at (-87.0,33.5)
South - North (Lat)
West - East (Lon)
tons/yr

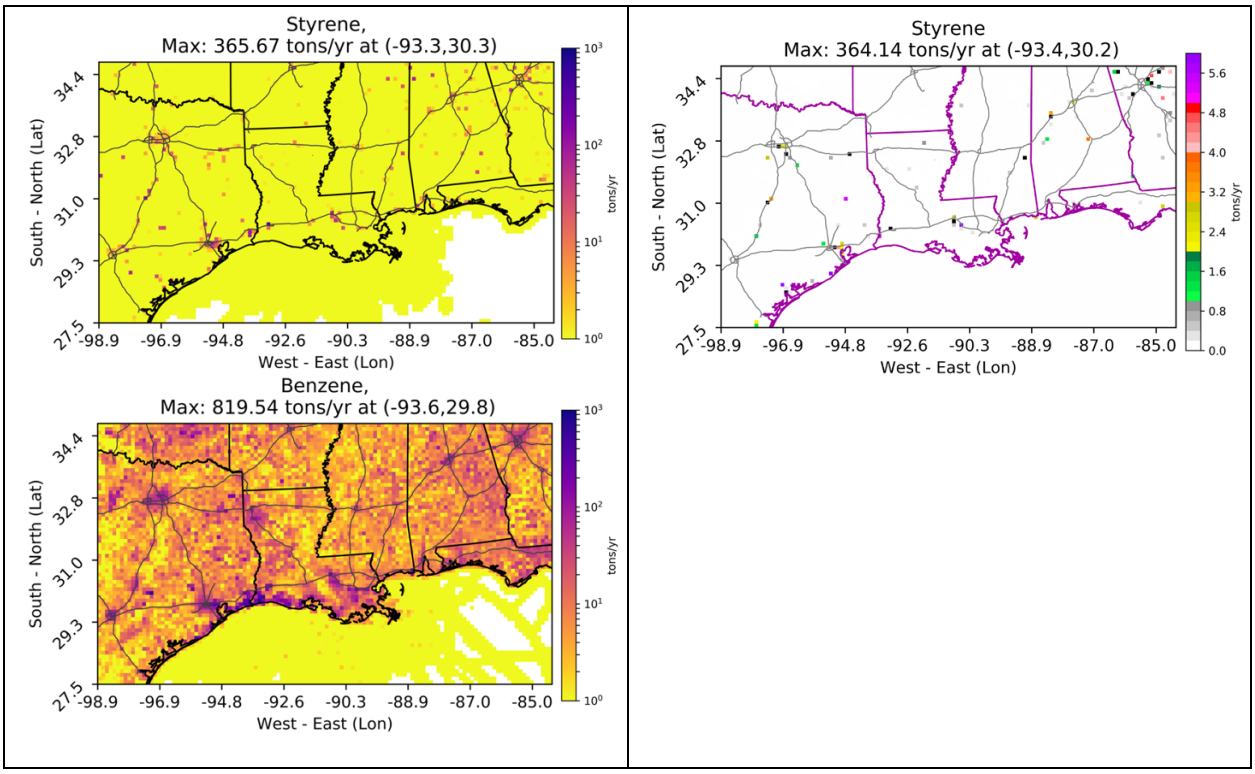
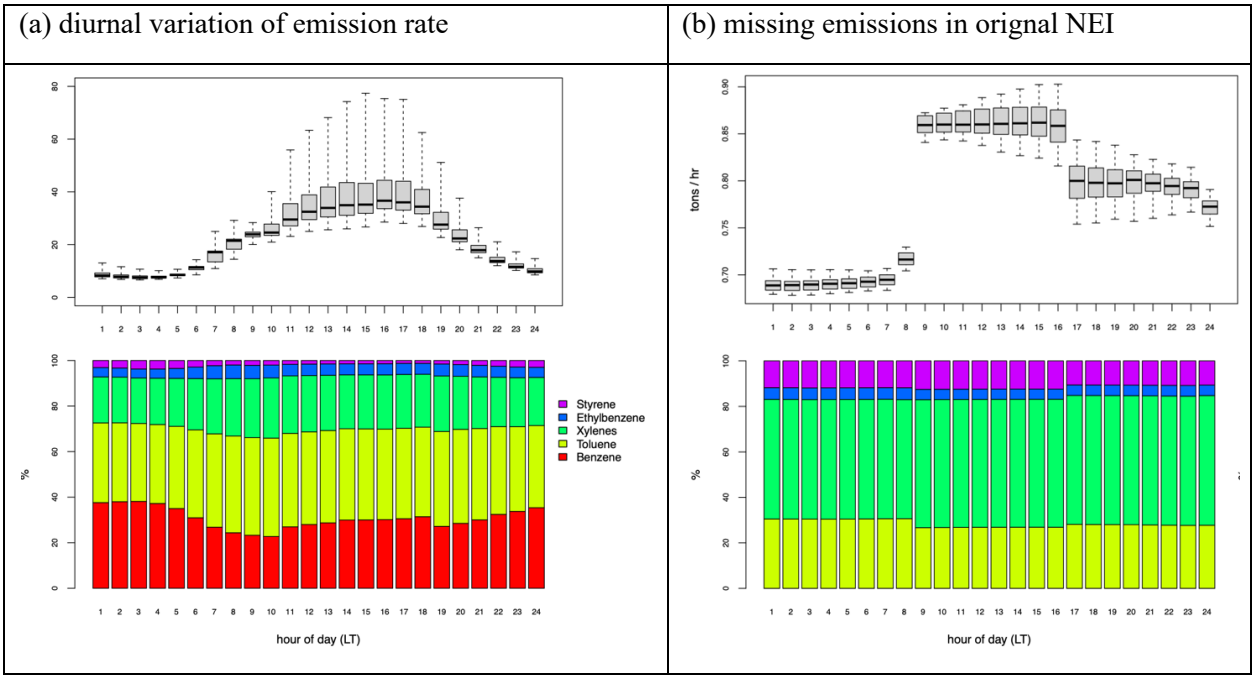


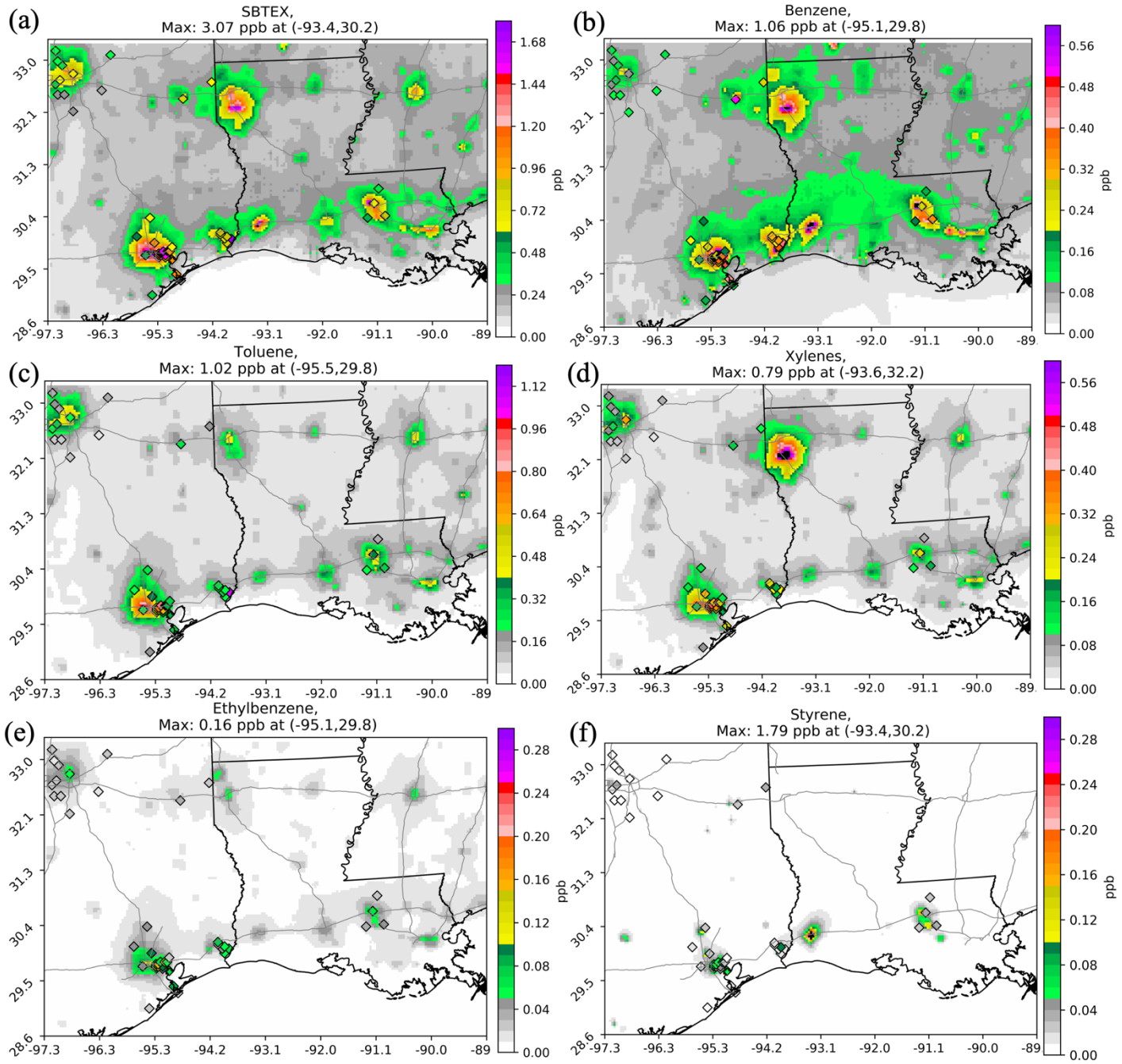
Figure 3. Spatial distribution of 2012 annual total SBTEX emission rates (ton/yr) of the modified emission inventory used in this work (a), and the location and amount of emissions that were added to the NEI (b).

759



760

761 **Figure 4**. Diurnal emission pattern of sum styrene, benzene, toluene, ethylbenzene, and xylenes
762 (SBTEX) (domain total, tons hr^{-1}) (upper panel) and the average relative composition of five
763 species (lower panel).



764

765 **Figure 5** (a) The average concentration in Adj scenario overlapped the average observational
766 measurement data (Diamond shape) during the model simulation period (May 1st, 2012 to Sep
767 30th, 2012) for (a) Total SBTEX, (b) benzene, (c) toluene, (d) xylenes, (e) ethylbenzene, (f)
768 styrene.

769

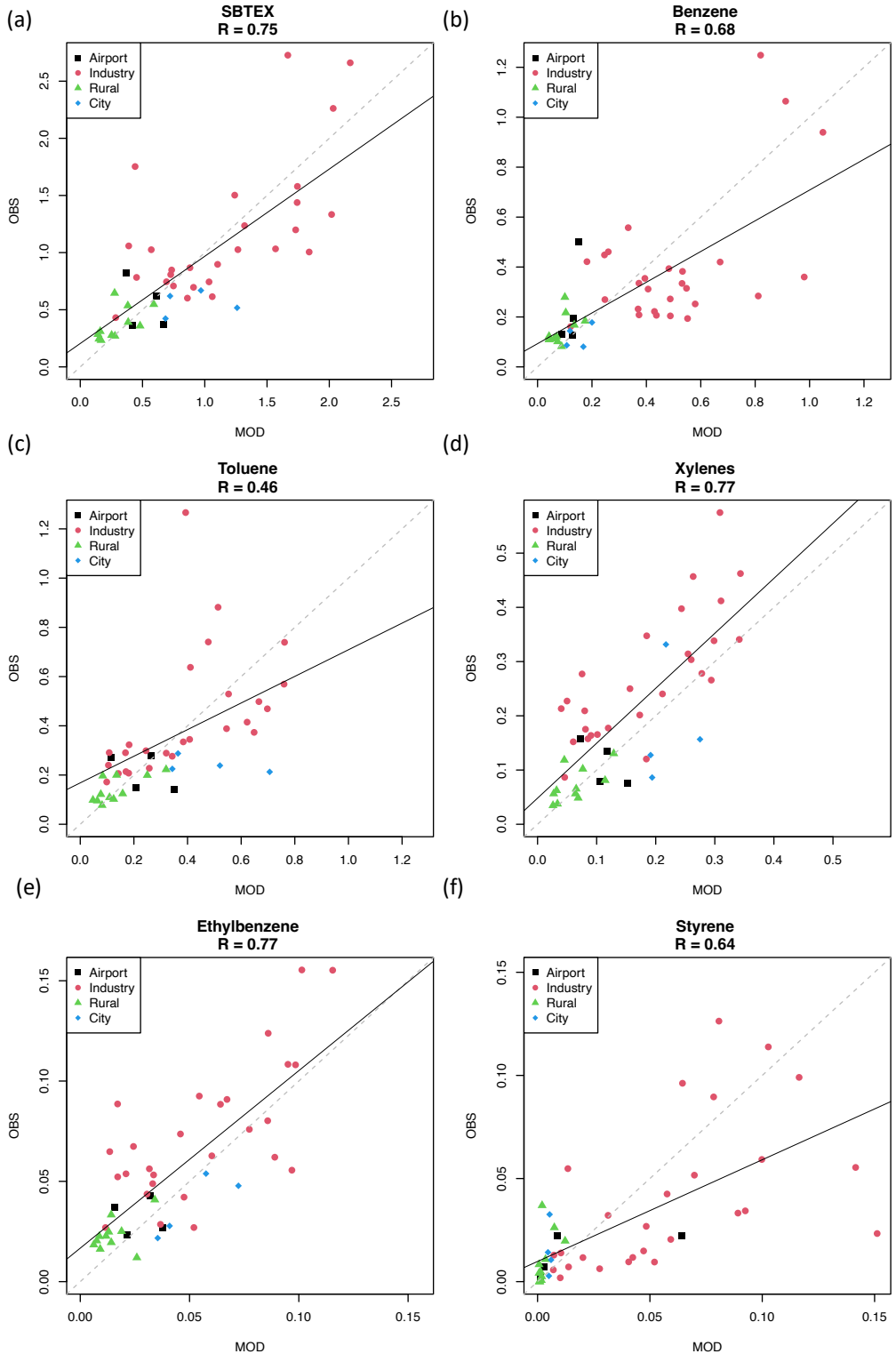
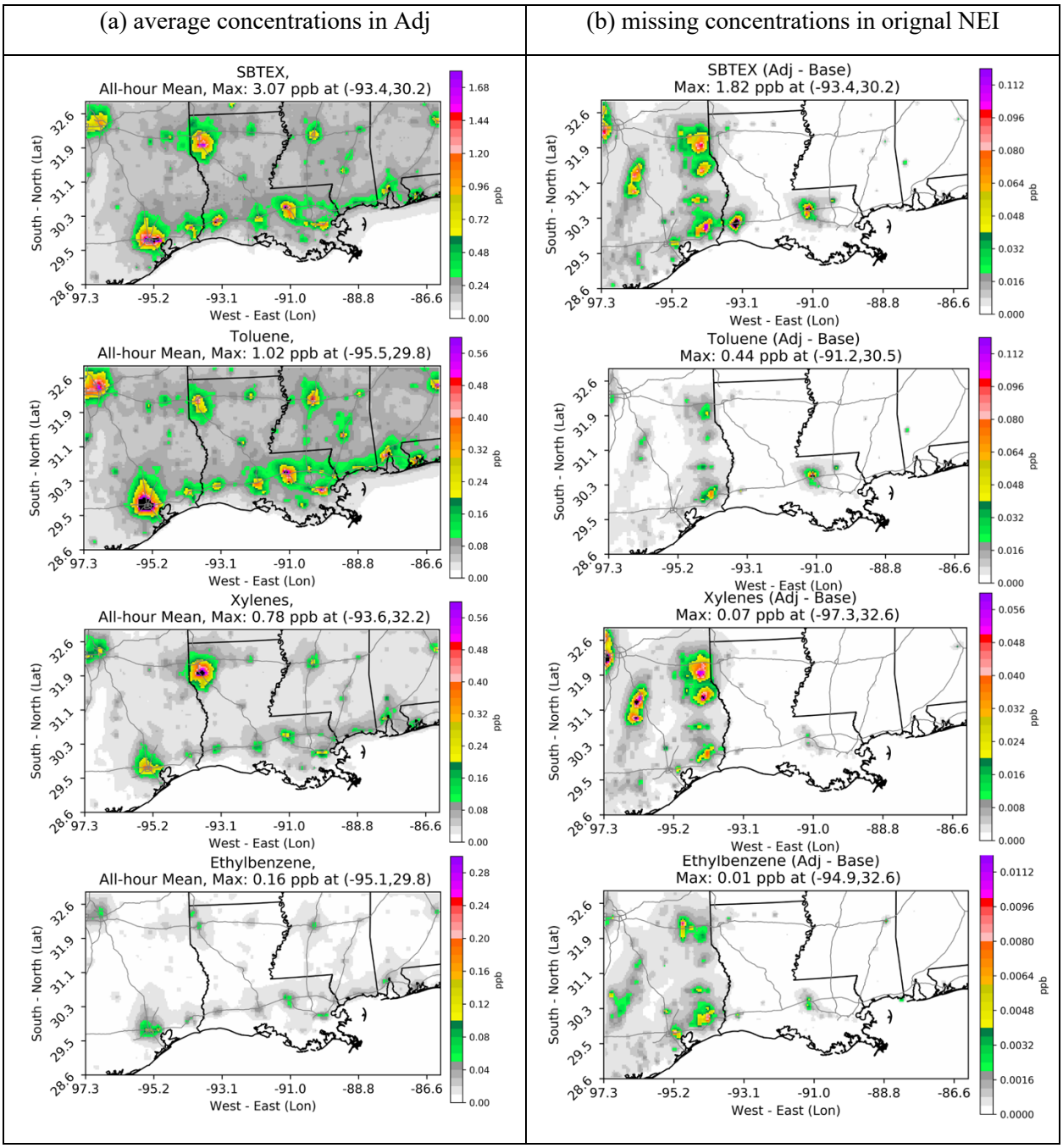
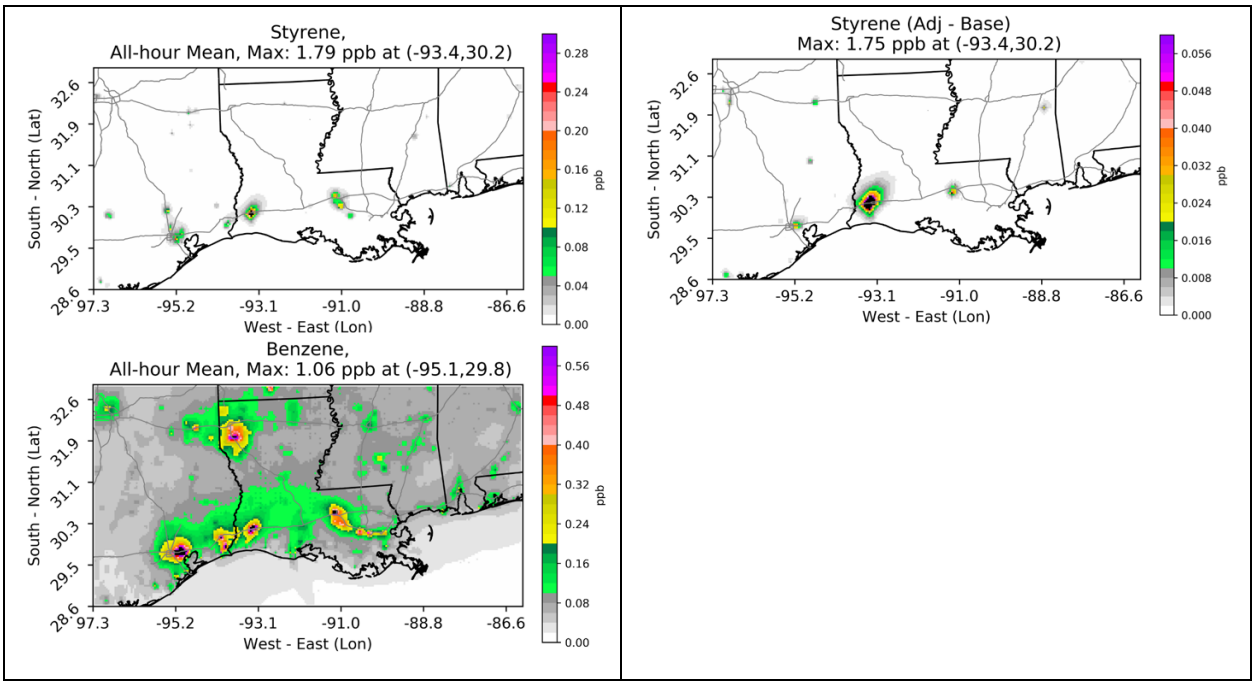


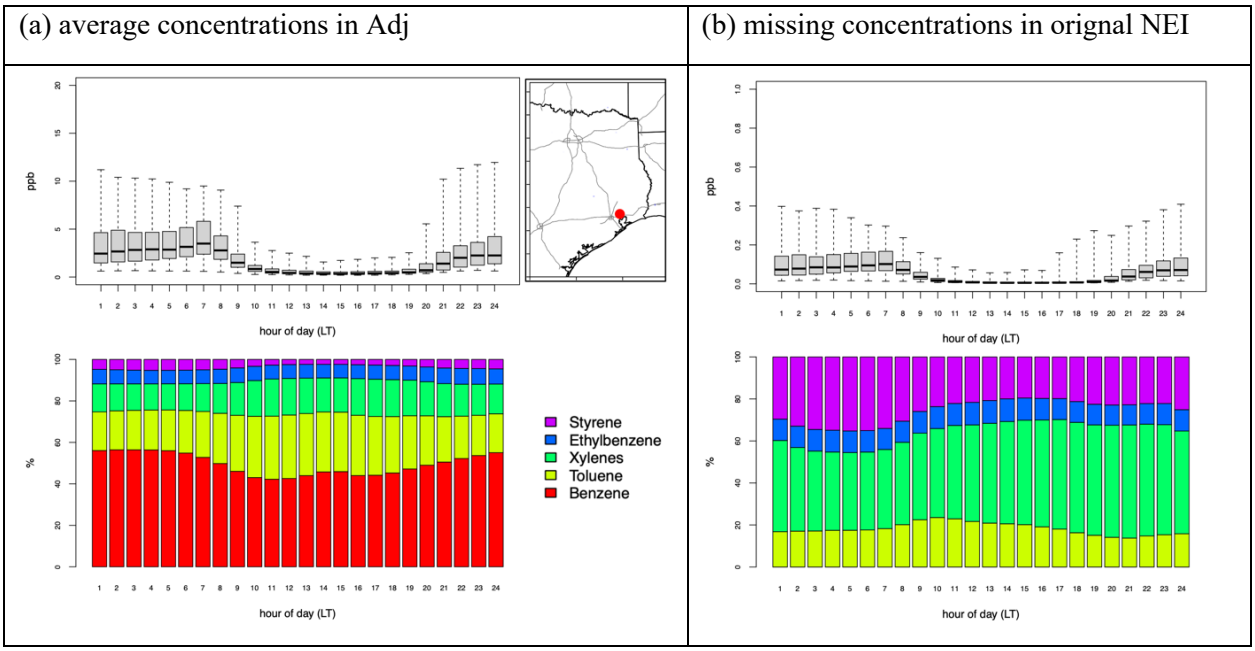
Figure 6. The average SBTEX concentration (ppb) comparison between model (MOD) Adj case and observational (OBS) data during the model simulation period (May 1st, 2012 to Sep 30th, 2012) for (a) total SBTEX, (b) benzene, (c) toluene, (d) xylenes, (e) ethylbenzene, and (f) styrene


(a) average concentrations in Adj
(b) missing concentrations in orignal NEI
SBTEX,
All-hour Mean, Max: 3.07 ppb at (-93.4,30.2)
Toluene,
All-hour Mean, Max: 1.02 ppb at (-95.5,29.8)
Xylenes,
All-hour Mean, Max: 0.78 ppb at (-93.6,32.2)
Ethylbenzene,
All-hour Mean, Max: 0.16 ppb at (-95.1,29.8)
SBTEX (Adj - Base)
Max: 1.82 ppb at (-93.4,30.2)
Toluene (Adj - Base)
Max: 0.44 ppb at (-91.2,30.5)
Xylenes (Adj - Base)
Max: 0.07 ppb at (-97.3,32.6)
Ethylbenzene (Adj - Base)
Max: 0.01 ppb at (-94.9,32.6)
South - North (Lat)
West - East (Lon)
ppb



774 **Figure 7.** The average concentration of SBTEX during the model simulation period (May 1st,
775 2012 to Sep 30th, 2012) in Adj scenario. The black color indicates the concentration is higher
776 than max color scale bar.

777



778

779 **Figure 8.** Diurnal pattern (upper panel) and relative composition (lower panel) of SBTEX
780 concentrations from May 1st to September 30th in Houston Ship Channel industry area,
781 Channelview city (red dot location)

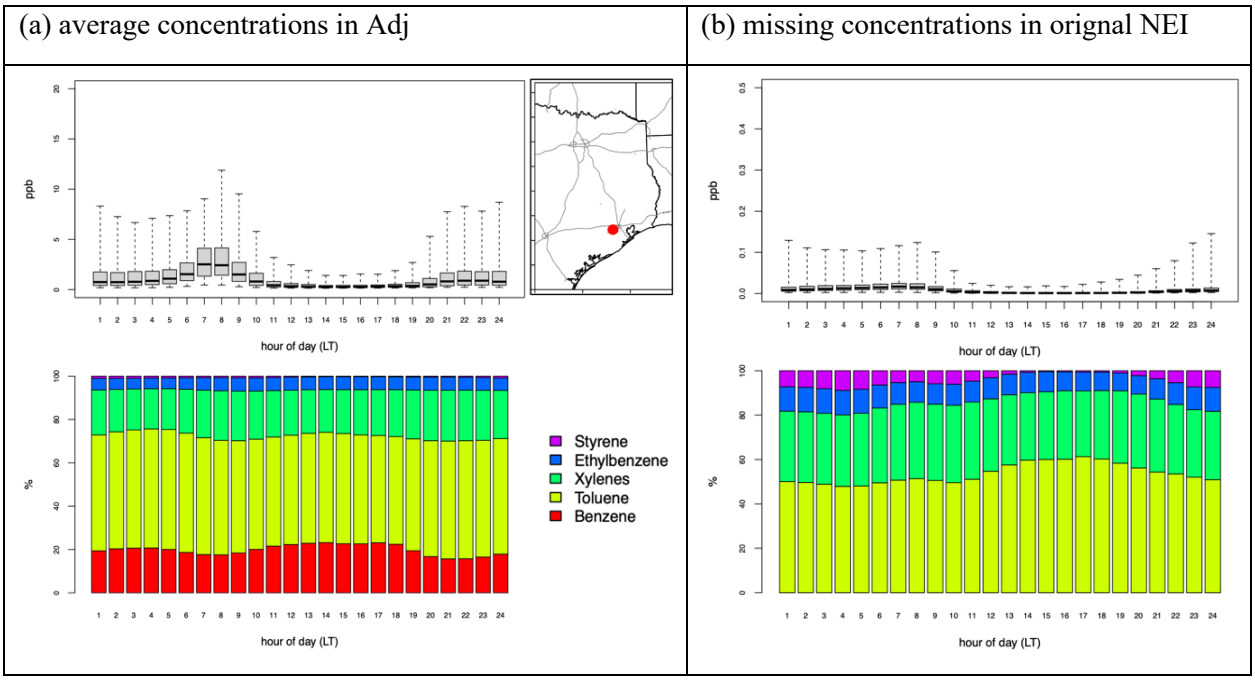
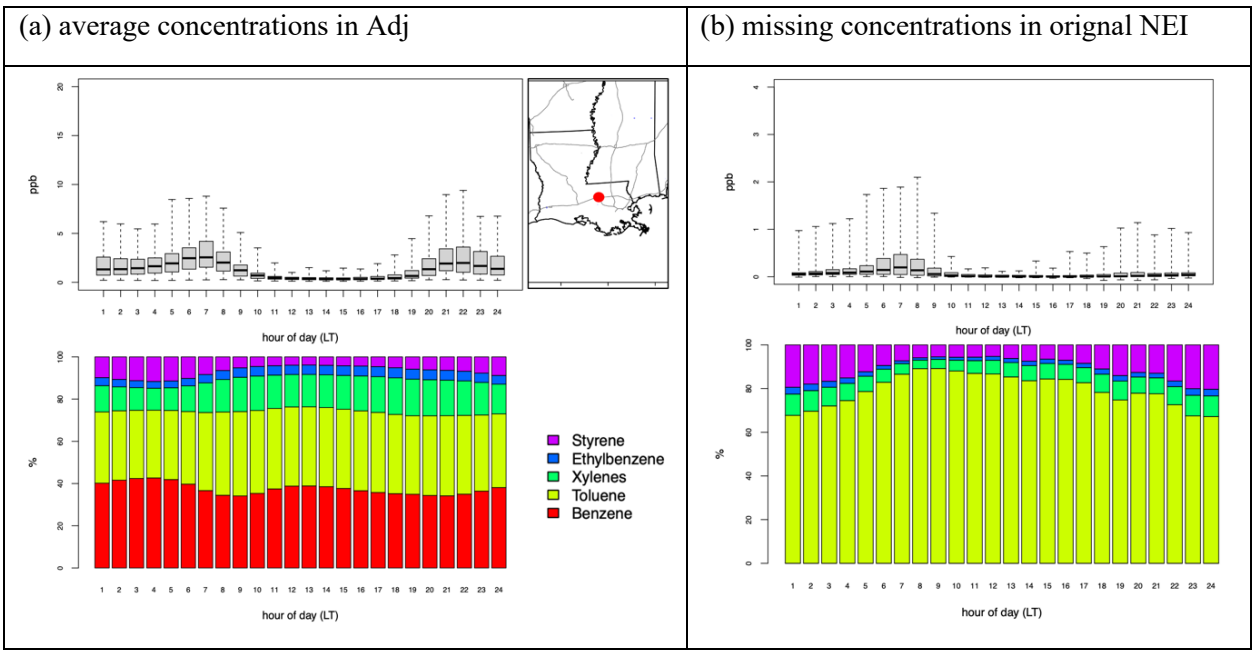


Figure 9. Diurnal pattern (upper panel) and relative composition (lower panel) of SBTEX concentrations from May 1st to September 30th in Houston residential area near Bayland Park (red dot location).

788



789

790 **Figure 10.** Diurnal pattern (upper panel) and relative composition (lower panel) of SBTEX
791 concentrations from May 1st to September 30th in Baton Rouge city (red dot location).

792

793