Uncertainties in Emissions Processing: Effects of using different Emissions Processing Tools and Surrogate Data Inputs

Daewon W. Byun and Soontae Kim
Institute for Multi-dimensional Air Quality Studies (IMAQS)
University of Houston, Houston, Texas TX 77204, USA
dwbyun@math.uh.edu; kimsoon@math.uh.edu

ABSTRACT

The Houston-Galveston Area (HGA) is designated as one of the non-attainment areas due to high ground-level ozone and particulate matter concentrations. Several air quality modeling studies are actively carried out to find cost-effective measures for improving air quality in the region. One essential part of the modeling input data, the emissions inventory (EI), should be processed though emissions modeling systems like SMOKE (Sparse Matrix Operator Kernel for Emissions) and EPS2 (Emissions Preprocessing System version 2) for use in air quality modeling. These emission processing systems may present different AQM-ready emission inputs depending on the use of different cross-reference files, profiles for spatial distribution, temporal allocation methods, and chemical speciations as well as the EIs that are used. Therefore, it is worthwhile to compare one emission modeling system to another by processing the same EI.

In this study, we have characterized the emissions processing uncertainties using tools such as SMOKE and EPS2 for point and area sources, BEIS3 and GloBEIS3 for biogenic emissions, and MOBILE5 and MOBILE6 for mobile emissions with the Texas emissions inventory developed by Texas Commission on Environmental Quality for the Houston-Galveston ozone non-attainment area. SMOKE and EPS2 presented different gridded, speciated, and hourly emission rates of model species due to differences in surrogates for spatial allocation, split factors for chemical speciation and hourly activity factors for temporal allocation for each source in the emissions inventory. In addition to the anthropogenic emissions, biogenic emissions estimated with BEIS3 and GloBEIS3 in the systems utilizing different LULC and meteorological data were compared. Different inputs as well as split factors for the lumped VOC emissions of OVOC and monoterpenes presented significant difference in biogenic emissions estimations in BEIS3 and GloBEIS3. For isoprene estimated explicitly, BEIS3 showed around 10 % higher domain-wide emissions than GloBEIS3.

INTRODUCTION

The Houston-Galveston area is classified as one of the nation's non-attainment areas due to high ground-level ozone concentrations. Several air quality modeling studies are actively carried out to find cost-effective measures for improving air quality in the region. One essential part of the modeling input data, the emissions inventory (EI), should be processed though emissions modeling systems like SMOKE (Sparse Matrix Operator Kernel for Emissions) or EPS2 (Emissions Preprocessing System version 2) for use in air quality models such as CMAQ and CAMx. Along with U.S. EPA's National Emissions Inventory (NEI), the Texas Commission on Environmental Quality (TCEQ) has implemented emissions processing methods for building Texas Emissions Inventory (TEI) used for the Houston-Galveston Area (HGA) state implementation plan (SIP) modeling studies. In particular, the inventory data, which include Houston-Galveston Ship Channel point-source speciated VOC emissions, are processed through EPS2 and GloBEIS3, and EPA's MOBILE6 modified by the Texas Transportation Institute. The emissions data are used with the CAMx air quality model to assess the efficacy of the emissions control strategies in the HGA. To evaluate effects of different physical and chemical processes with a different air quality model, for example CMAQ, the TEI must be prepared in the CMAQ-ready format.

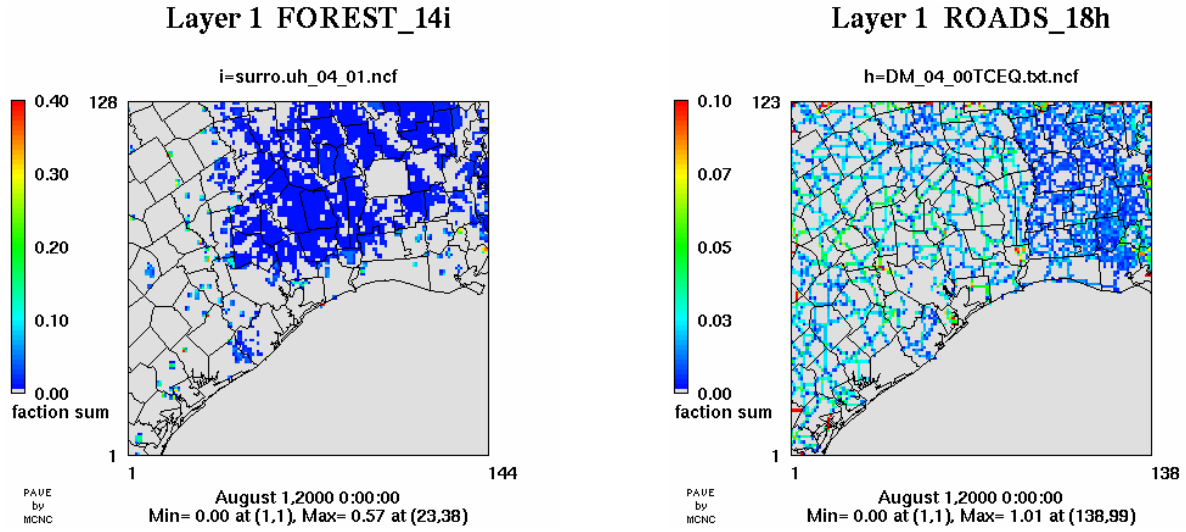
Recently, we have developed the TEIPS (Texas Emissions Inventory Preparation System for SMOKE), a computational tools that allows conversion of the TEI data in EPS2 formats into the data format accepted by the SMOKE (Kim & Byun, 2003). TEIPS is used to cross-check the Texas emission inventories by processing them with the SMOKE system and to extend chemical speciation for the other mechanism such as SAPRC99 in addition to CB-4 currently used. We have found that the EPS2 and SMOKE emission processing systems result in different AQM-ready emission inputs because of the use of different cross-reference files, profiles for spatial distribution, temporal allocation methods, and chemical speciations. In this study, we summarize the differences between the SMOKE and EPS2 results for point and area sources, BEIS3 and GloBEIS3 for biogenic emissions, and mobile emissions with the Texas emissions inventory.

EXPERIMENTAL DATA

Texas Emissions Inventory

To perform modeling studies for the development ozone state implementation plan (SIP) for the HGA ozone nonattainment area, TCEQ has prepared Texas EI data. The data is expected to be more updated and to have more detailed speciated emissions inventory data than the national inventories, such as NEI99. Texas area and point EIs, which can be downloaded from the web site (http://www.tnrcc.state.tx.us/air/aqp/airquality_photomod.html#ei), were prepared in several categories by dividing geographically total emissions of each source type into five and six sub-emission types, respectively. The area and nonraod mobile sources include the EI subcategories such as Texas area, Texas nonroad, Louisiana emissions, off-shore shipping lanes, and elevated ship emissions for peak ozone day. Point source emissions are separately prepared for Electric Generating Utilities (EGU) and Non-EGUs (NEGUs) using the Aerometric Information Retrieval System (AIRS) Facility Subsystem work file format (TCEQ, 2002). The subcategories for the point source include Texas EGU, Texas NEGU, Louisiana EGU, Louisiana NEGU, offshore platforms, and Texas upset and additional emissions. Texas emissions inventories for EGU sources are available for hourly and peak ozone day while NEGU sources are available for peak ozone day. The on-road mobile sources for the HGA 8 counties were processed with the MOBILE6 using link-based vehicle mileage traveled data.

Figure 1. Examples of emission surrogates processed with SMOKE Tool and MIMS for forest (left) and road emissions (right), respectively.



Emission Shape Files

For spatial allocation of criteria pollutants emissions with SMOKE, GIS surrogate files were downloaded from U.S. EPA (ftp://ftp.epa.gov/EmisInventory/emiss_shp/), and then processed with SMOKETOOL or MIMS (Multi-scale Integrated Modeling System). Figure 1 provides some examples of the GIS shape files processed for SMOKE processing. These surrogate files are different from those employed for the processing of Texas emissions inventory with EPS2. For certain categories, such as the onshore and offshore, the necessary shape files are not available from EPA. Therefore, we have implemented the same shape files used for EPS2 onto SMOKE. In order to prevent the emissions from being mismatched spatially, onshore and offshore emissions were prepared in separate files and processed with TCEQ's shape file instead of the EPA's.

Internal Data for EPS2 and SMOKE

The same emissions inventory processed with different emission processing systems such as EPS2 and SMOKE may present different model-ready emissions data because they are heavily dependent on the cross-reference and profile files used for spatial and temporal allocations, and chemical speciation (Kim & Byun, 2003). To compare the results after processing the Texas emissions inventory with both EPS2 and SMOKE, the cross-reference and profile files for EPS2 were downloaded from the TCEQ web site. SMOKE used the EPA-recommended default cross-reference and profiles for chemical speciation and spatial/temporal allocation.

Biogenic Emissions data

Biogenic emissions depend heavily on the vegetation types and some canopy meteorological conditions, such as canopy temperature and photo-synthetically available radiation (PAR).

Vegetation Data

GloBEIS and BEIS3 require different land use and land cover (LU/LC) data for the estimations of biogenic emissions. For the HGA SIP modeling purpose, TCEQ is utilizing a specially compiled dataset for land use and vegetation information for the state of Texas and the surrounding states. Compared to other parts of Texas and US, the LU/LC database available for Eastern Texas is up-dated relatively recently. It is a composite land use database that includes a mapping of ground cover, vegetation species, and leaf mass densities for the state of Texas (Weidinmyer, et al., 2001). Land use and vegetation were divided into over 600 classifications at

approximately 1 km spatial resolution. Some field surveys were performed to estimate leaf biomass densities of certain tree species. Special emphasis was put in to generate more detailed urban LU/LC classifications. When no recent data were available, the USGS LU/LC database at 1-km resolution was applied to provide spatial distribution of the urban land use types. In addition to the municipal, state, and Federal government land use, land cover, and vegetation data at resolutions from 30 m to 1 km, county-based agricultural LU/LC data were incorporated as well. Although we may consider this LU/LC database has reference year of 2000, the representative years of the different data sources vary widely and in some case uncertain. This dataset can be considered as a more detailed and up-dated land type distributions than those available from EPA as BELD3 (Biogenic Emissions Land use Data), which is used in SMOKE processing. The data is available from (http://www.tnrcc.state.tx.us/air/aqp/airquality_photomod.html#ei4c).

Canopy Temperature

Temperature measurements were obtained from several different monitoring networks. Networks were chosen if they had acceptable QA procedures in place, and data were available for the time period of interest. Differences in sensor height among the temperature networks are usually not an issue during hot summer days, when vigorous mixing leads to small temperature gradients, but they might be an issue during dry, cool, still conditions when larger temperature gradients might occur near the ground. Data from the following networks were used: TCEQ network, Aerometric Information Retrieval System (AIRS) network, National Weather Service network, Texas Crop Weather Program, Conrad Blucher Institute Texas Coastal Observation Network, and National Automated Buoy Data network. Overall, data from over 100 stations were used. The statistical technique of kriging was used to interpolate temperature measurements, thus creating a temperature field for each hour of the chosen episode.

Photosynthetically Active Radiation (PAR)

Photosynthesis by the plant leaves occurs within the wavelengths between 400 nm - 700 nm and this is termed photosynthetically active radiation (PAR). Therefore, biogenic emissions modeling requires input of hourly PAR data for the modeling domains. Because of the sparseness of surface radiation monitoring sites, interpolation of measurements is unlikely to yield a satisfactory field, given the heterogeneous nature of clouds. In general, for the photochemical modeling purpose, the short-wave radiation predicted by a numerical meteorological model has been used to generate PAR fields. However, because of the difficulties in simulating clouds in the mesoscale

models, the estimated PAR fields would reflect the effects of spurious clouds predicted by the model. To avoid such problem, TCEQ has generated the hourly PAR fields with input data from the GOES8 satellite using algorithms developed by Pinker et al., 2003. Cloud cover estimates from satellite imagery were fed into the radiation balance algorithm(s) to create a large-scale field of PAR. High resolution PASR fields were created from 1/16 degree solar field data. Comparisons between GOES-derived PASR fields and ground-based broadband solar radiation measurements found very high degrees of correlation. Correlations for TCEQ sites ranged from 0.94 to 0.97, with slopes ranging from 0.47 to 0.53, indicating that PASR comprised approximately 50% of broadband solar radiation (i.e., 20 nm - 2000 nm) (Estes, 2004; personal communication).

METHODS

Implementation of the Texas EI into SMOKE for TexAQS 2000 episodes

Texas area and point source emissions in AMS (AIRS Area and Mobile Subsystem) and AFS (AIRS Facility Subsystem) work file formats were modified for use in SMOKE. The header part describing file format and information required in SMOKE to recognize the input file type was added for each file, and SMOKE run scripts and some input files were modified to process the Texas EI. Texas area source EI was divided into five categories; a) Texas area, b) Texas nonroad, c) Louisiana all area, d) offshore, and e) elevated ship emissions. All of these categories were prepared in AMS format except elevated ship emissions, which were prepared in AFS work file format and were treated as point emissions. During the SMOKE run with the area EI, the default temporal and gridding cross-reference and profile files were used except for offshore emissions. Area on-/off-shore emissions were prepared in separate files because SMOKE does not have surrogates for offshore emissions. To resolve this problem, emissions shape files for EPS2 were also implemented onto SMOKE to process the emissions of which surrogates are not available from SMOKE.

While other emissions sources such as area, biogenic and mobile sources are spatially distributed with surrogates, point sources can be spatially allocated with their own location data. The location of one point source in AFS format can be presented as either in UTM distance (km) or in LAT-LON (hereafter, LL) decimal degrees (U.S. EPA, 1992). Although SMOKE can also read the location data of a point source in any coordinates (MCNC, 2002), we found some problems using Texas emissions inventory (TEI) in SMOKE without converting because TCEQ uses LCP (Lambert Conformal Conic Projection) coordinates instead of UTM and LL coordinates. Therefore, we have

replaced the TCEQ's LCP parameters with the LL coordinates. We have performed additional format conversion of the TEI hourly point emissions to process them with SMOKE, which can import hourly emissions in AFS format. However, if we follow standard SMOKE processing steps, the temporal variation pattern becomes different from the real data after processing because SMOKE just imports hourly emissions and adds them up to save daily emission rates, but does not keep the emission patterns. Therefore, we have re-processed the hourly emissions data with the day-specific temporal profiles, while keeping the AFS format. The basic concept of this work is first to prepare a separate hourly emissions file on each day, and second, to create the day-specific hourly temporal cross-reference and profile files, and, finally, to process these inputs with SMOKE.

Table 1. HGA-specific surrogates for EPS2 and the availability from SMOKE.

Surrogates	Remarks	Surrogates	Remarks
Population	A, B	Gas stations	B
Urban population	A, B	Dry cleaner	B
Rural population	A, B	Restaurants	B
Commercial airports	-	Residential area	A, B
General airports	A, B	Forest	A, B
Military airports	B	Agriculture	B
County yards	A, B	Commercial & industrial	B
Water	A, B	Commercial & residential	-
Ships	B	Oil and gas wells (Inland)	-
Harbors	A, B	Offshore oil and gas wells	-
Canal	-	Offshore	-
Railroad	A, B	Shipping lanes	-
Auto body shops	-	Platforms	
Marine coating facilities	-		

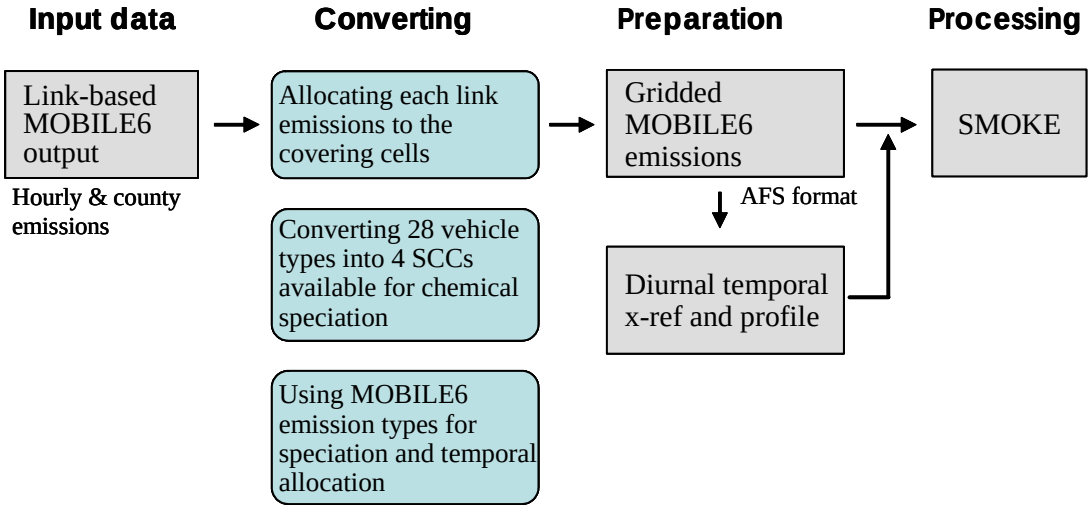
Remarks: A and B represent the surrogates are included in U.S. EPA's old and new surrogating system, respectively, using 15 and 64 surrogates.

GIS Shape Files and Emission Surrogate Data

The GIS emission shape files from the U.S. EPA (ftp://ftp.epa.gov/EmisInventory/emiss_shp/) were processed with the SMOKE Tool to prepare 15 surrogates used for spatial allocation in SMOKE. Similarly, EPS2 uses 27 surrogates developed by TCEQ (Funk et al., 2002). Table 1

lists the HGA-specific surrogates and indicates the availability from two sets of U.S. EPA's surrogate data. The updated system employs 64 surrogates and covers more categories than those in TCEQ compared with the old surrogates used in the work. However there are still other unique surrogates prepared for the HGA area in the TCEQ's data. These unique surrogates, especially for offshore emissions, are implemented in SMOKE.

Figure 2. Preparation steps for implementation of link-based MOBILE6 emissions into SMOKE.



Treatment of Mobile Emissions

TCEQ is currently using the link-based MOBILE6 emissions from TTI for the HGA 8-county area using VMT (Vehicle Mileage Traveled) data. While the current version of SMOKE (1.4 Beta) relies on county-based mobile emissions or MOBILE5, TTI has utilized the link-based VMT data for the MOBILE6 emissions processing. The data used include more detailed vehicle and emission types for the criteria pollutants emitted from on-road vehicles. Since both EPS2 and SMOKE are not able to directly process the link-based emissions inventory, pretreatments of the data are needed. Instead of TCEQ's SAS code for LBASE in EPS2, a FORTRAN code for data converting from MOBILE6 output to LBASE input format was used. To process the link-based MOBILE6 emissions in SMOKE, we have followed three major steps, the same as for other emissions components; chemical speciation, temporal allocation and spatial distribution with surrogates. Chemical speciation can be done with EPA's cross-reference and profile files prepared for the MOBILE6 species, and temporal allocation can be completed by creating daily temporal allocation files based on the FIPS code, SCC and emission types as if we did it for hourly point

emissions. The last part is to allocate the emissions spatially. SMOKE uses surrogate files for gridding, however it is not applicable to the link-based emissions. Therefore, a new method is capable of spatially allocating the link-based emissions without using the gridded surrogates as in the original SMOKE mobile processing. In the new approach, as shown in Figure 2, each link is gridded prior to processing in SMOKE using a module 'GLINE' in EPS2, and then is assigned to a fake point source located on the center of each cell. This method works the same way as the temporal allocation of point source emissions, which does not use the surrogates but uses the location of each point source.

RESULTS

After implementing into SMOKE, the Texas EI was processed both with SMOKE and EPS2 during the period of TexAQS 2000 to compare their results. The emissions modeling domain for this comparison was set up using TCEQ's model domain which covers southeastern Texas and some parts of Louisiana and the Gulf of Mexico as shown in Figure 1.

Effects of Different Surrogate Shape Files

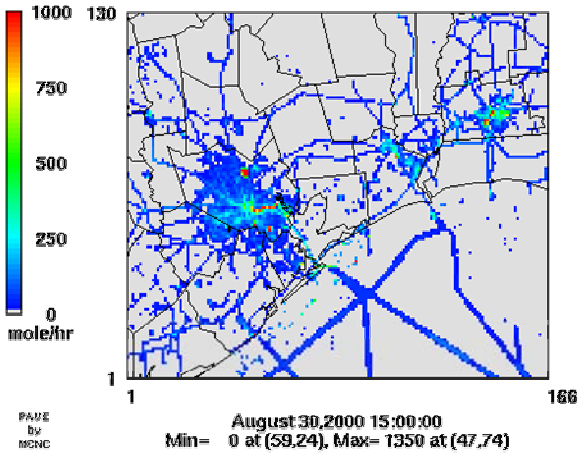
Figure 3 compares spatial distributions of Texas area source emissions between EPS2 and SMOKE. We have employed the spatial allocation factors prepared by SMOKETOOL with the EPA's emission shape files for processing the inland emissions in SMOKE. However, we have used TCEQ's surrogate files for the onshore and offshore emissions after modifying the file format.

Effects of Different Speciation Methods.

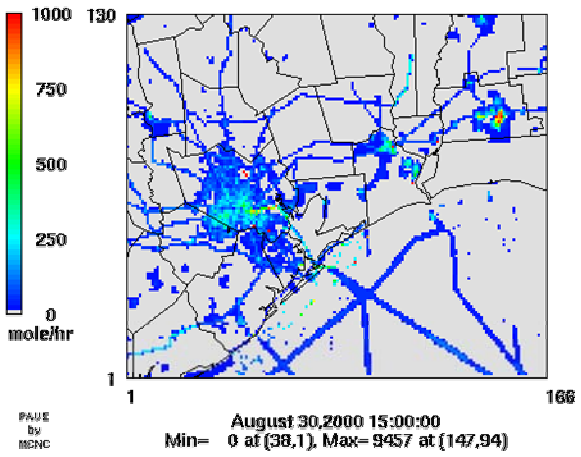
The chemical split factors are applied to speciate the lumped VOC and NO_x emissions into individual model species. SMOKE assigns a speciation profile prepared by U.S. EPA based on the SCC (Source Classification Code) of the source. By using U.S. EPA- and TCEQ-default split factors for the CB-IV species, different model species are estimated. To compare domain total emission rates after chemical speciation for the CB-IV mechanism, we present the ratio of EPS2 to SMOKE processed emission rate for each species in Figure 4. From the ratios for CO and NO_x we can deduce the implications of the other processing step differences, for example in the spatial allocation. SMOKE shows higher emission rates relative to EPS2 for all species except for TOL and XYL (see Figure 4a).

Figure 3. Spatial distributions of area and nonroad mobile NO emissions processed by SMOKE (a & c) and EPS2 (b & d). TCEQ's surrogates were used for onshore and offshore emissions in SMOKE.

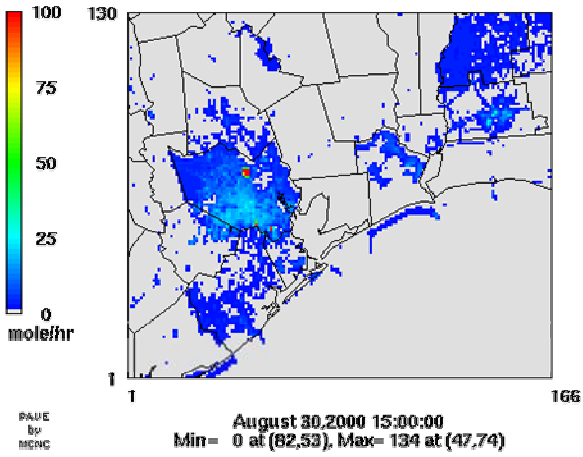
(a) Area NO from SMOKE



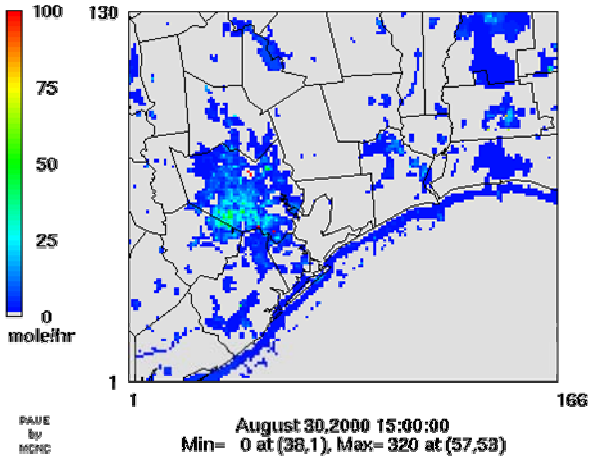
(b) Area NO from EPS2



(c) Nonroad NO from SMOKE



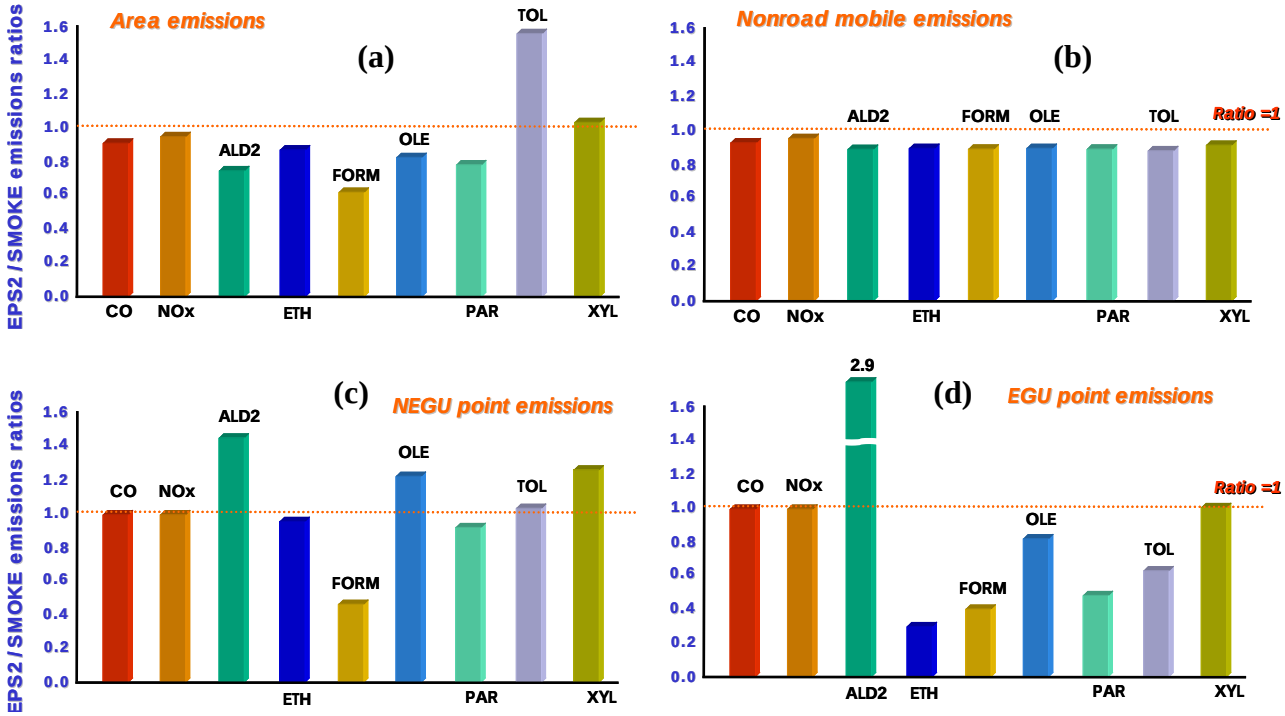
(d) Nonroad NO from EPS2



Nonroad mobile emissions (see Figure 4b) show an almost uniform ratio of EPS2 to SMOKE emissions rates (about 0.9) after chemical speciation. Figure 4b indicates that SMOKE generates around 10% more emissions than EPS2 for the same species. In Figure 4c, EPS2 shows relatively higher emissions for some CB-IV species such as ALD2, OLE, and XYL than SMOKE for NEGU point emissions. The ratios of EPS2 and SMOKE for these species vary from source to source according to the split factors applied to the source. Differences in chemical speciation become larger for EGU point emissions (see Figure 4d). EPS2 assigns the lumped VOC emissions into ALD2 much more than SMOKE and then show relatively low emission rates for other species.

In the cases of CO and NO_x emissions for point sources, surrogate data for spatial allocation are not needed and therefore uncertainties that originated from using different surrogates are not present. Thus, CO and NOx emission rates for point sources become almost the same (see Figures 4c-d).

Figure 4. Chemical speciation results for (a) Area, (b) Nonroad mobile, (c) NEGU point and (d) EGU point source emissions.

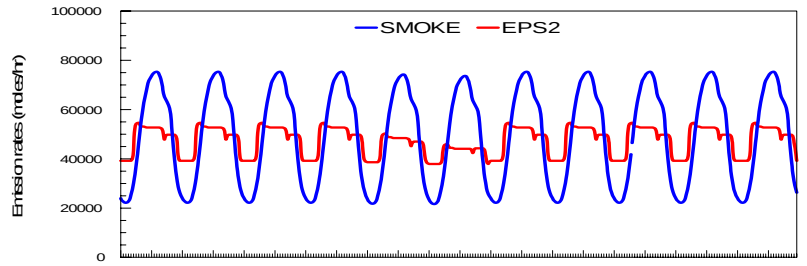


Effects of Different Temporal Allocation Factors.

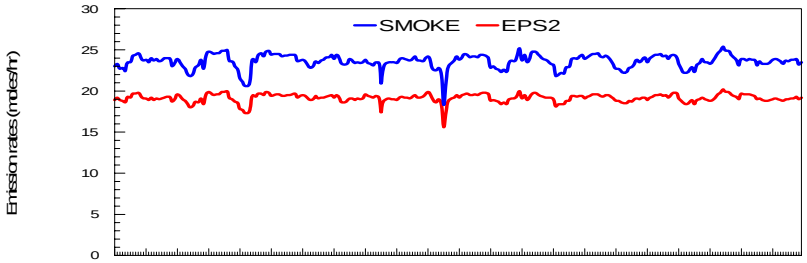
To obtain hourly emissions, the peak ozone day and annual average emissions for area/nonroad and point sources are allocated with the monthly, weekly, and weekday/weekend temporal profiles. SMOKE presents more diurnal fluctuations for the nonroad mobile emissions compared to EPS2 as shown in Figure 5a. In particular, SMOKE shows over 30% higher NO emission rates in daytime for nonroad mobile emissions. Figure 5b compares olefin emissions for the EGU point sources. SMOKE shows around 20% higher emission rates, but the variation patterns are quite similar. Usually, EPS2 does not present diurnal variations for the NEGU point emissions. However, SMOKE applies the diurnal variations to the emissions (Figure 5c).

Figure 5. Examples of temporal variations for (a) nonroad NO, (b) EGU point olefin, and (c) NEGU point NO emissions after SMOKE and EPS2.

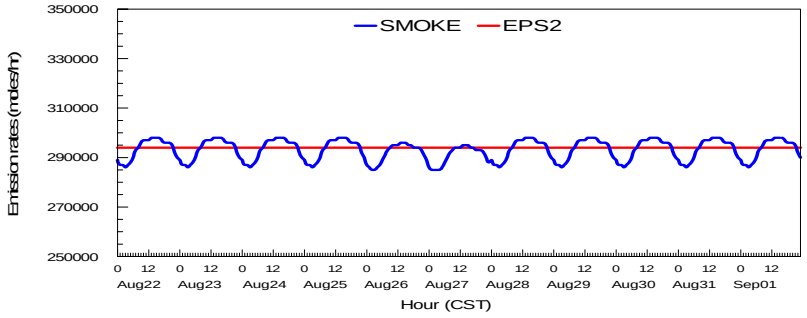
(a) Nonroad mobile NO emissions



(b) EGU point olefin emissions



(c) NEGU point NO emissions

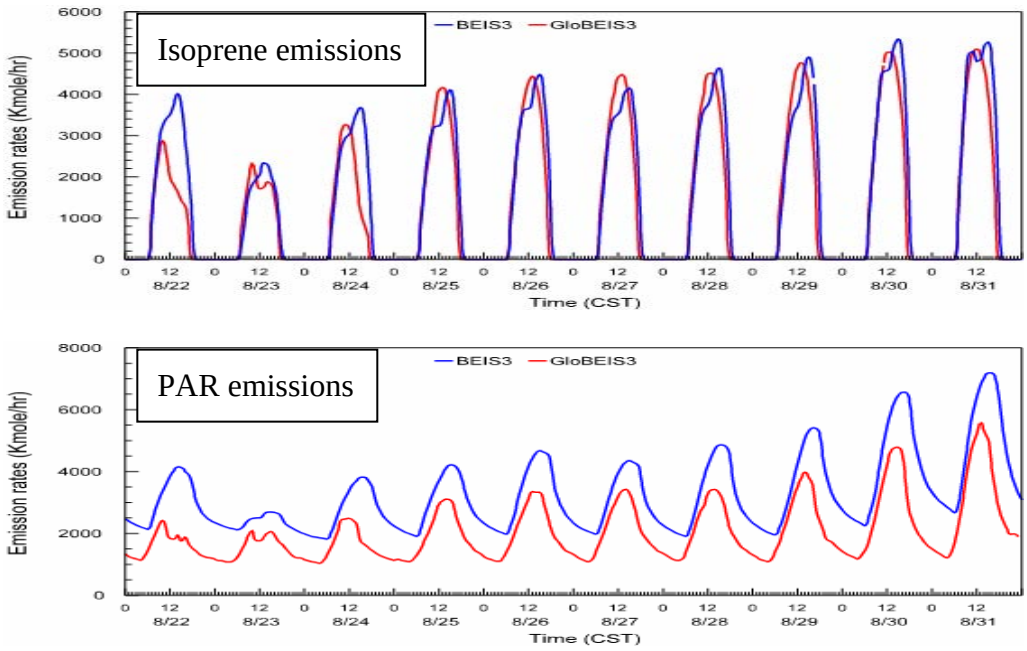


Biogenic emission estimates with BEIS3 and GloBEIS3

Figure 6 shows a comparison of ISOP and PAR emissions between GloBEIS3 used by TCEQ and BEIS3 in SMOKE. Land use data used for the GloBEIS3 was from TCEQ (Funk et al., 2002; TCEQ, 2002) and the same for the BEIS3 was prepared with SMOKETOOL using BELD3. In the case of meteorological data, SMOKE used MCIP output from MM5 and GloBEIS3 used radiation estimated by satellite data analysis and observed temperature. While ISOP spatial distributions from both cases show similar emissions rates, maximum PAR emissions rates from BEIS3 are different from GloBEIS3 substantially. More details will be examined by comparing

each estimation process and emissions factor of each species in GloBEIS3 and BEIS3 afterwards and the influences will be verified through air quality model simulations during the second year of the project.

Figure 6 Comparisons of ISOP (top) and PAR (bottom) emissions processed by GloBEIS3 and BEIS3 in SMOKE for the CB-4 chemistry mechanism.



Mobile Emissions Processing

Many of NO emissions were processed from hourly-resolved and link-based MOBILE6 outputs for HGA and both systems show relatively similar daily and diurnal cycles. Small differences in hourly NO emissions appear due to different temporal allocation factors applied to mobile emissions outside HGA 8 counties for which annually averaged daily emissions are used. However, chemical speciation presents different domain-wide VOC emission rates as shown in Figure 7b.

CONCLUSIONS

Texas emissions inventory used for the Houston-Galveston Area SIP modeling studies by TCEQ were processed with two emissions preparation systems, SMOKE and EPS2. This preliminary

analysis (e.g., Figure 8) presents significant impact of the using differences in the model-ready emissions. We have identified that some of these differences are originated from the base databases utilized in the systems. Many of these components can be harmonized and then the model-ready emissions used in different air quality modeling systems can be made much closer than is now. In the future these emissions data will be used to perform cross comparisons of SMOKE and EPS2 with CAMx and CMAQ.

Figure 7. Comparisons of mobile emissions processed by EPS2 and SMOKE; (a) NO emissions, (b) OLE emissions, and (c) PAR emissions for the CB-4 chemistry mechanism.

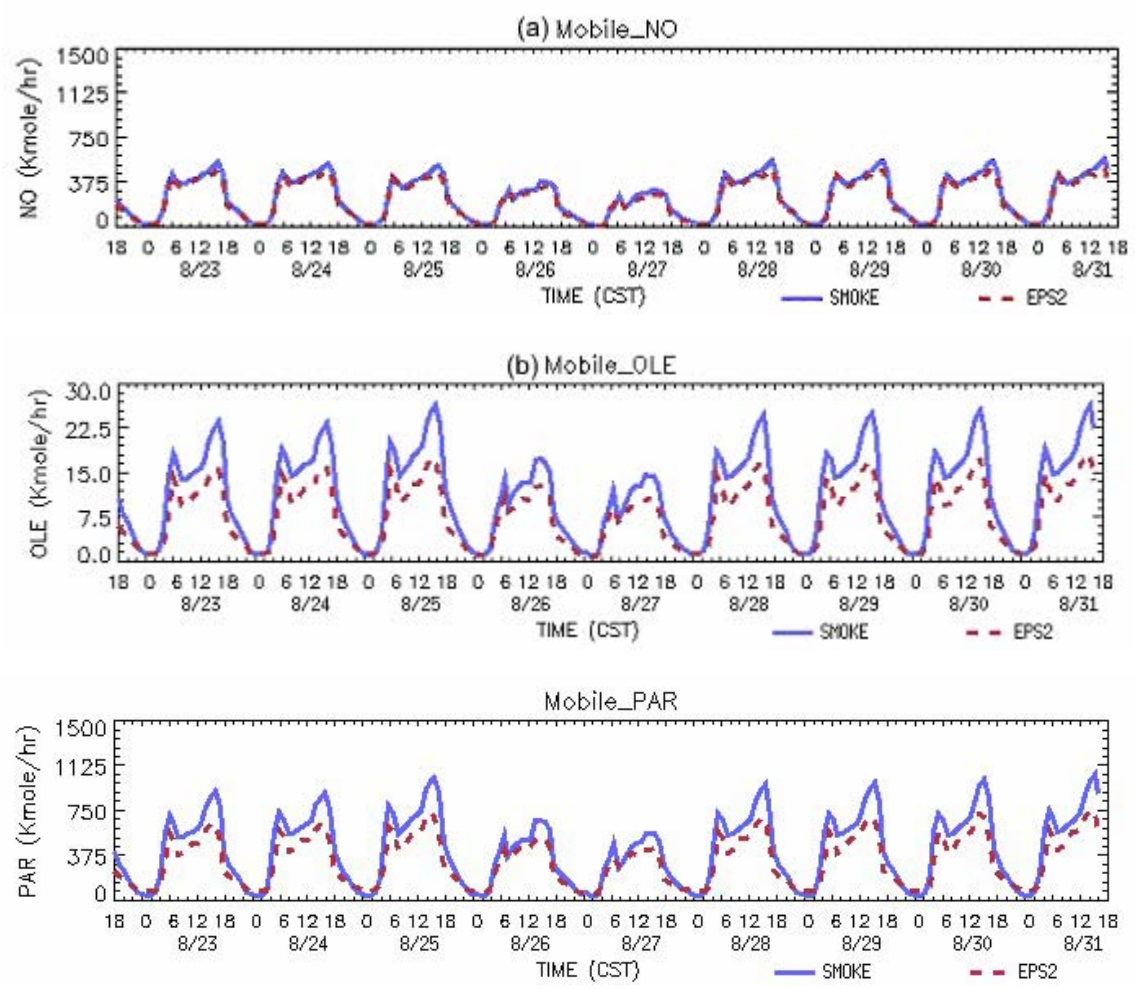
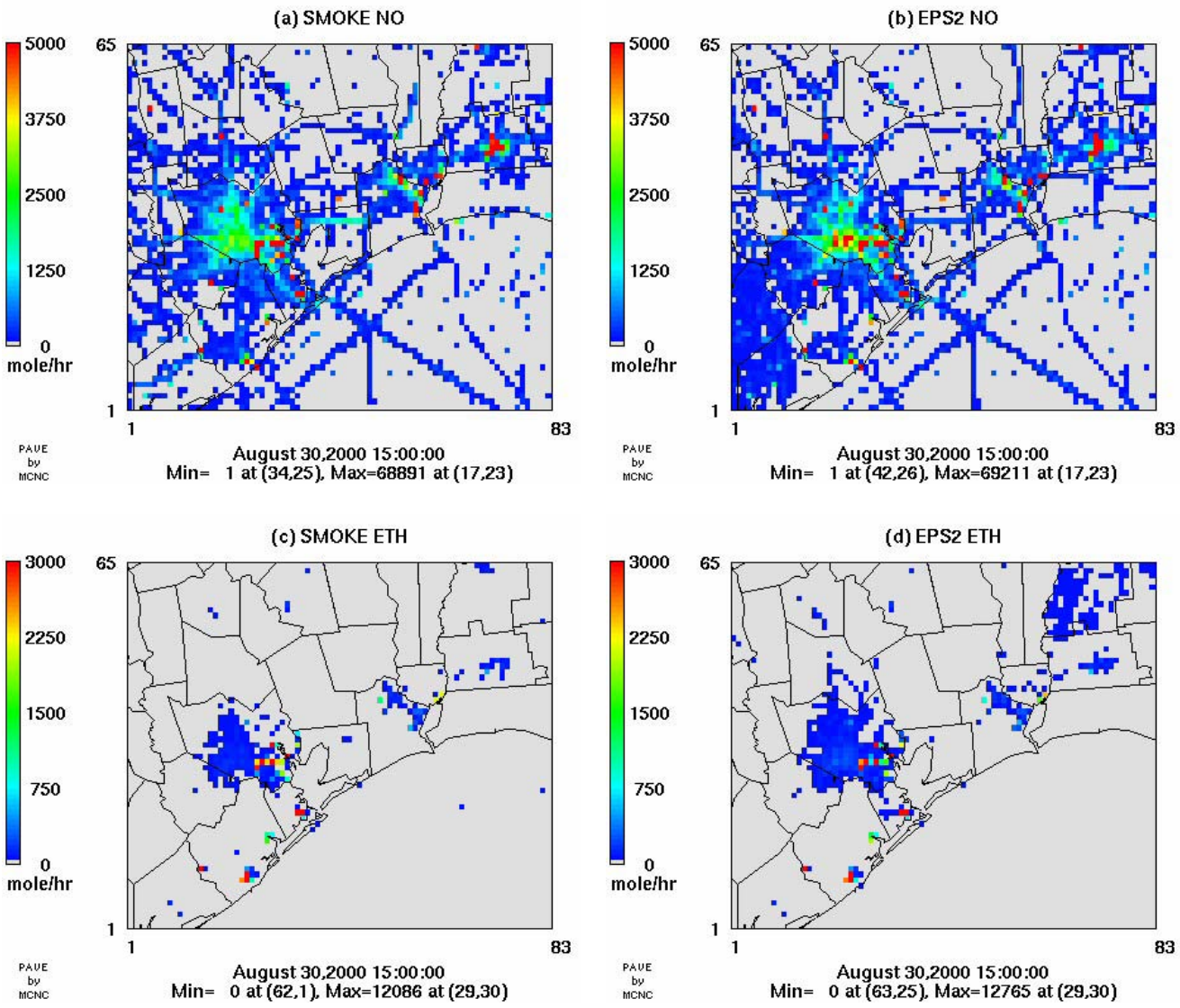


Figure 8. Model-ready NO and ETH emissions estimates obtained with SMOKE (a&c) and EPS2 (b&d) systems.



REFERENCES

Funk, T.H., Stiefer, P.S., and Chinkin, L.R. (2001) Development of Gridded Spatial Allocation Factors for The State of Texas, Final Report 3, STI-900570-2114-FR3, Prepared for Texas Natural Resource Conservation Commission, August 31, 2001. (ftp://www.tnrcc.state.tx.us/pub/OEPAA/TAD/Modeling/HGAQSE/Contract_Reports/EI/)

Kim, S., and Byun, D.W., 2003: Prototyping the Texas emissions inventory preparation system for the SMOKE system, 12th International Emission Inventory Conference, San Diego, April 29 - May 1, 2003. (http://www.epa.gov/ttn/chief/conference/ei12/poster/kim.pdf)

MCNC (2002) "User's manual for Sparse Matrix Operator Kernel Emissions Modeling System, Version 1.4Beta". Online available: http://www.emc.mcnc.org/products/smoke/SMOKEDOCS.html.

TCEQ (2002) "ATTACHMENT 3: Emissions Inventory Development and Modeling for the August 25-September 1, 2000 Episode", Technical Analysis Division, November, 2002. Online available: http://www.tnrcc.state.tx.us/air/aqp/airquality_photomod.html.

U.S. EPA (1992) "Users guide for the Urban Airshed Model, Volume IV: User's Manual for the Emissions Preprocessor System 2.0, Part A: core FORTRAN system".

Wiedinmyer, A., Guenther, A., Estes, M., Strange, I.W., Yarwood, G., Allen, D.T., 2001: A Land Use Database and Examples of Biogenic Isoprene Emission Estimates for the State of Texas, USA, Atmos. Environ., 35, 6465-6477.

KEY WORDS

Emission processing
GIS surrogate files
Air quality modeling
Ozone in the Houston-Galveston area

ACKNOWLEDGEMENTS

The research is funded in part by the grant from the Texas Air Research Center under the project entitled, "Development of Texas Emissions Inventory Preparation System for SMOKE," and the United States Environmental Protection Agency through Grants R-82906801 and R-83037701 to the University of Houston. The document has not been subjected to the either required peer and policy reviews from TARC and EPA and therefore does not necessarily reflect the views of the funding organizations and no official endorsement should be inferred.