# Final Report

# Project 14-014

# Constraining $NO_x$ Emissions Using Satellite $NO_2$ Column Measurements over the Southeast Texas

Prepared for

Air Quality Research Program (AQRP)

The University of Texas at Austin

PREPARED BY

University of Houston

4800 Calhoun Road

Houston, Texas 77204-5007

*Principal Investigator*

Yunsoo Choi

*Co-PI*

Xiangshang Li

Project Time period: 1/27/2015 – 9/30/2015

Contributors: Dr. Yunsoo Choi, Dr. Xiangshang Li, Amir Hossein Souri,

Lijun Diao, Dr. Anirban Roy, Shuai Pan

QA Requirements: Audits of Data Quality: 10% Required

September, 2015

## Contents


Executive Summary ..... 4
1. Introduction ..... 6
1.1. Background ..... 6
1.2. Objectives ..... 8
1.3. Tasks ..... 10
2. Measurements, Methodology and Model Setup ..... 11
2.1. Measurements ..... 11
Surface Data ..... 11
Aircraft ..... 12
Remote Sensing Data ..... 12
2.2. Inverse Modeling ..... 14
2.3. Model Simulation ..... 16
WRF ..... 16
CMAQ ..... 18
3. Evaluation of Meteorology Simulations ..... 23
3.1 General Meteorological Conditions ..... 23
3.2 Evaluation metrics ..... 23
3.3 Meteorological simulation results ..... 24
4. Inverse Modeling Results: $NO_x$ Emission and Model and Satellite $NO_2$ Comparison ..... 28
4.1 Emissions ..... 28
Area Source Category ..... 28
Biogenic Soil Source category ..... 30
Mobile Source Category ..... 32
Point Source Category ..... 34
Total $NO_x$ Emissions ..... 38
4.2 Simulated $NO_2$ ..... 39
Comparison to OMI Tropospheric $NO_2$ ..... 39
Comparison to CAMS ..... 42
Comparison to Aircraft ..... 45
5. Inverse Modeling: Impact on Ozone ..... 49
5.1 Surface Ozone ..... 49
Hourly Ozone Statistics ..... 49
Regional Ozone Average Time Series ..... 49

Hourly Ozone Time Series at Selected Sites ..... 50
Ozone on 09/25 – the Only High Ozone Episode ..... 53
5.2 Ozone Aloft ..... 58
6. Summary and Conclusions ..... 59
7. Future Works ..... 60
Acknowledgments ..... 60
Appendix A ..... 61
Observation Nudging ..... 61
Impact on Ozone ..... 62
Appendix B ..... 65
Audits of data quality ..... 65
Audit Personnel ..... 65
Data to be audited ..... 65
Audit Procedures ..... 65
Appendix C ..... 65
WRF Namelist ..... 65
Publications from This Study ..... 69
References ..... 70

# Executive Summary

Tropospheric ozone is produced in the planetary boundary layer through a series of non-linear reactions between oxides of nitrogen ($NO_x$) and Volatile Organic Compounds (VOCs). Reliable emissions inventories are key to accurate ozone modeling. $NO_x$ concentrations are known to respond quickly to emission changes due to the very short lifetime of $NO_x$ in the atmosphere and a linear relationship with emissions. $NO_x$ emissions from mobile and point sources have significantly decreased over time due to improvement in fuel technology and effective emissions control technologies and strategies. These reductions are well corroborated by satellite and in-situ observations. These emissions reductions imply that the National Emission Inventory (NEI) used in model simulations needs to be updated for use to represent current emissions. A robust procedure to update the emission inventory is to use additional $NO_2$ observations and a 3-D modeling framework with a realistic representation of sensitivity of $NO_x$ levels to emissions. These improvements can be achieved using the inverse modeling technique which is useful when a direct quantity (here, emissions) is not measurable.

The concept of an inverse problem involves using physical laws to back-calculate the quantity of interest ("a posteriori" emissions) from known or partially-known approximated states ("a priori" ambient measurements or satellite retrievals). Updating emission inventories using satellite measurements is a common application for which remote sensing measurements are cost-effective. Due to spatial inhomogeneity in distributions of $NO_x$ concentrations and limited number of monitoring sites, satellite remote sensing of tropospheric $NO_2$ is widely used for this purpose. The biggest advantage of using satellite data is the high spatial coverage that makes it possible to update emission inventories over large domains. The National Aeronautics and Space Administration (NASA) records this data through its Ozone Monitoring Instrument (OMI) aboard the Aura satellite; tropospheric $NO_2$ concentrations from this instrument (Level 2, V2.1) were used for this project. Several adjustments were required to remove noisy observations and re-grid them for our model simulation domain. These included corrections for cloud fraction, quality flags, solar angle and removal of a priori profile influences. Our results revealed that removal of a priori profile influences is crucial to obtaining accurate results. Disregarding this step leads to under-prediction and over-prediction of $NO_2$ in urban and rural regions respectively.

In this study, a Bayesian inversion of tropospheric OMI $NO_2$ is conducted to update four source categories of the National Emission Inventory (NEI) 2011 (area, soil, mobile and point sources), as well as the total $NO_x$ inventory. Using a "Brute Force" method, the adjusted NEI-2011 demonstrated an overall reduction in anthropogenic source categories (i.e., area, mobile and point) and an increase in biogenic soil emission of $NO_x$. The largest reduction in $NO_x$ emission for the area source category was predicted in the center of Houston and Lake Charles city. A large increase in the biogenic $NO_x$ emissions (50%) was found in most rural areas. Additionally, the results indicated that mobile $NO_x$ emissions reduced by 30-40% in urban regions. The largest decrease was found for the center of Houston. The largest decrease for point source category was found for Lake Charles and Houston Ship Channel, ranging between 40-65%. The considerable decrease was mainly due to strict emissions regulations. A new simulation was performed using the United States Environmental Protection Agency (USEPA)'s Community Multiscale Air Quality (CMAQ) model using the updated total $NO_x$ emissions. The model simulation results were evaluated using in-situ data from independent aircraft studies and the Texas Commission on Environmental Quality (TCEQ)'s Continuous Ambient Monitoring Stations (CAMS)

network. Mean error and root mean square error (RMSE) of the model-measurement comparisons were reduced with the updated $NO_x$ emissions. RMSE and bias between aircraft $NO_x$ measurements and simulated ones are 2.4 and 6.0 ppbv for the default NEI-2011 and 1.9 and 4.1 ppbv for the updated NEI-2011. The model-measurement comparisons using CAMs data indicated that, mean absolute bias and RMSE decrease by 0.8 and 1.1 ppbv from the default NEI-2011 to the inverse modeling updated case.

In order to investigate the impact of $NO_x$ emissions changes on chemical conditions (i.e., $NO_x$-saturated or –sensitive), we calculated the $HCHO/NO_2$ ratio before and after using inverse modeling. A comparison of P3-B aircraft HCHO levels to simulated values demonstrated under predicted HCHO levels likely resulting from biogenic model used. We did not find any difference in HCHO concentrations before and after using inverse modeling due to marginal indirect impacts of $NO_x$ on HCHO. After updating $NO_x$ emissions using inverse modeling, most of the urban regions became increasingly $NO_x$-sensitive (20-85% increases in the ratio) due to large decreases in anthropogenic $NO_x$ emissions. On the other hand, rural regions showed a transition from $NO_x$-sensitive toward more $NO_x$-saturated due to an increase in soil $NO_x$ emissions.

We also investigated the impact of the updated inventory on ozone predictions. Overall, the ozone changes were small both on surface and aloft, suggesting that the Houston region is $NO_x$ saturated. We found statistics (before and after updating emission inventory), for correlation (0.74, 0.76), the Index of Agreement (IOA) (0.79, 0.80), RMSE (14.6, 14.4), MAB (12.0, 11.7) and MB (9.3, 9.3). This is in agreement with previous findings suggesting that other parameters such as VOC emissions and dry depositions should be constrained as well. Small increases in ozone concentrations were observed in $NO_x$-sensitive rural areas, consistent with the increase in $NO_x$ in the new inventory. Additionally, we applied Objective Analysis (OA) to meteorological modeling using the Weather Research and Forecasting (WRF) model to constrain the modeled fields using in-situ data. The results indicated improvement of the simulated meteorology fields which provides a solid base for subsequent emission and chemistry modeling. Regarding the wind fields which are critical to pollutant transport, the IOA improved by 6% for U-wind and 11% for V-wind after using OA. For temperature, the correlation coefficient increased by about 10% while IOA rose by 8%.

# 1. Introduction

## 1.1. Background

Nitrogen oxides or $NO_x$ ($NO_x$=NO+$NO_2$) are a precursor for tropospheric ozone and particulate matter formation in the atmosphere (Seinfeld and Pandis, 2006), and a harmful species for the human respiratory system. Emissions of nitrogen oxide (NO) from anthropogenic combustion of fossil fuels (Noxon, 1978), biomass burning activity (van der Werf et al., 2006), soil microbial activity (Yienger and Levy, 1995) and lightning (Choi et al., 2009) are the major sources of $NO_2$. Our previous studies focused on showing the impact of lightning and/or anthropogenic sources on tropospheric $NO_x$ and ozone ($O_3$) over the continental US for the summers of 2005 and 2009. These have been documented in a series of publications (e.g., Choi et al., 2005, 2008a, 2008b, 2009, Choi et al., 2012; Choi, 2014). Control technologies and strategies implemented by governmental regulatory agencies have significantly reduced anthropogenic ozone precursor emissions including $NO_x$. Houston is a city known for its large petrochemical industrial facilities, located in an ozone nonattainment region. According to the EPA's National Emissions Inventory (NEI) air pollutant emissions trends data, anthropogenic sources of $NO_x$ from both stationary and mobile sources declined by 45% in the U.S. from 2000 to 2014. Additionally, anthropogenic Volatile Organic Compound (VOC) emissions were reduced by 16%. Since these reductions have not occurred evenly, monitoring and modeling $NO_x$ levels are vital to fully understanding the efficiency of efforts at reducing surface ozone.

Previous work (Choi et al., 2012; Choi, 2014; Choi and Souri, 2015; Kim et al., 2011) demonstrated that regional model simulations (e.g., CMAQ and WRF-Chem) are problematic in simulating $NO_x$ concentrations mainly due to uncertainties in emission inventories. Also, there could be uncertainties in the gas-phase chemistry module of the model. For example, the lifetime and chemical fate of alkyl nitrates for $NO_x$ simulation are still uncertain (Duncan et al., 2014). In a previous study (Choi and Souri, 2015), we studied how changes in anthropogenic emissions and have impacted (WRF-Chem) employing two modeling scenarios for September 2013. The first scenario using NEI-2005 showed a fair agreement between surface ozone from Continuous Ambient Monitoring Stations (CAMS) and model output during daytime (r = 0.63). However, the model significantly over-predicted surface $NO_2$ by a factor of ~3.6. This is due to overestimation of $NO_x$ emissions in the inventory. This finding agreed with previous studies (e.g., Anderson et al., 2014, Choi, 2014, Choi et al., 2012; Kim et al., 2011). Specifically, Anderson et al. (2014) modeled the 2011 DISCOVER-AQ field campaign in the Baltimore-Washington metropolitan area. They found that their CMAQ (Community Multiscale Air Quality) simulations using projected NEI-2005 to NEI-2011 emissions overestimated $NO_x$ emissions by 51–70% for mobile sources, which made up approximately 50–75% of total $NO_x$ emissions in the region. In the second scenario that employed NEI-2011 emissions, this overestimation was significantly mitigated, reaching a factor of only 1.7. The unchanged overestimation of surface $NO_2$ by using NEI-2011 emphasized that an adjustment should be performed for future studies using this inventory. This overestimation possibly resulted from a high bias in mobile emissions which masks the real chemical condition of cities for conducting a sensitivity analysis. It also exacerbates the quantification of relative importance of $NO_x$

emissions from other source categories such as power plants. We also compared the Ozone Monitoring Instrument (OMI) $NO_2$ and formaldehyde (HCHO) concentrations to simulated values and found that CMAQ overestimated the $NO_2$ column density while underestimating the HCHO column density. The model indicated that ozone levels reduced from the first to the second scenarios in Dallas (-3%) and Fort Worth (-0.5%) but increased surface ozone in Austin (0.7%), San Antonio (0.8%), and Houston (7%); probably resulting from underestimation of VOC contributions from petrochemical facilities and biogenic emissions (Pan et al., 2015), and overestimation of $NO_x$ in the NEI-2005 inventory that led to false low $HCHO/NO_2$ ratios. The findings of this work suggest that anthropogenic emissions should be considerably adjusted prior to making any environmental policy solely based on model results.

In addition to $NO_x$ and VOC emissions, several studies (e.g. Banta et al., 2005; Ngan et al., 2012) have emphasized on the significant influences of meteorological processes on air quality. One way of improving meteorological simulation is to use the Objective Analysis (OA) technique to assimilate observational data to reduce uncertainties in the Weather and Research Forecasting (WRF) simulation. Using a state-of-the-art meteorological model such as WRF along with OA should provide an improved meteorological simulation. OA improves meteorological analyses of coarse resolution on the model grid by incorporating information from in-situ observations. This procedure is sometimes referred to as "nudging" the WRF model nearer to the in-situ data. Without local information, simulated meteorology (especially wind fields) may deviate markedly from observational data. OA has been shown to be very effective in rectifying modeled surface level wind fields which is a parameter critical to overall model performance (e.g. Ngan et al., 2012; Li et al 2015) and air quality modeling (Czader et al., 2013). Hence, OA must be performed very carefully to preserve model integrity.

One way to improve the emission inventory is to use ambient observations to constrain emissions, i.e. to infer what must have been emitted in order to observe the measured ambient values. Inverse modeling of sources of atmospheric trace gases is a well-established technique that incorporates this inference process and has been addressed in several studies (Martin et al., 2003; Shim et al., 2006; Kemball-Cook et al., 2015). In this procedure, emissions are optimized in order to reduce differences between modeled and observed data. In our previous studies, we separated the continental US into six different geographic regions to analyze the uncertainty of anthropogenic $NO_x$ emissions modified from NEI-2005 using a 12 km grid and version 4.7.1 of the CMAQ model. Here, point source $NO_x$ emission reductions from 2005 to 2009, called as "EMIS2009" (Choi et al., 2012) were considered. Using Global Ozone Monitoring Experiment-2 (GOME-2) $NO_2$ retrievals for August 2009, we calculated a monthly-averaged $NO_2$ column. We also estimated monthly column values of $NO_2$ concentrations from the model with EMIS2009. We performed an additional simulation including remote sensing-adjusted $NO_x$ anthropogenic sources (e.g., Choi et al., 2008a; Choi et al., 2012). The GOME-2-adjusted $NO_x$ emissions (EMISGOME) were 7.8% less than EMIS2009 over the US (Choi et al., 2012). Comparisons between the observed and model-simulated $NO_x$ concentrations at the USEPA's AQS (Air Quality System) stations over the geographic regions were conducted. We found that six regions had relatively high correlation coefficients ($R>0.7$) between hourly $NO_x$ data from observations and each of two chemical transport model runs. These were the Pacific (PC), Rocky Mountain (RM), Lower Middle (LM), Upper Middle (UM), South East (SE), and North East (NE)). For each of these regions, biases of the baseline model are generally higher, but biases of the model, including emissions changes improved in terms of Normalized Mean Bias (NMB), except for the

RM region (Choi, 2014). Interestingly, among six geographical regions, noticeable changes in biases are found over LM in US, reducing NMB from +149.7% to -1.8%. Over LM, a significantly large reduction in $NO_x$ biases suggests that $NO_x$ emissions from EMIS2009 were probably too high. Zoomed-in regionalized studies were performed over Houston to highlight uncertainties of EMIS2009 emissions from the previous study (Choi, 2014). The model with EMISGOME mitigated discrepancies between simulated and observed surface $NO_x$ over Houston. Large $NO_x$ emissions reductions decreased surface $NO_x$ concentrations over Houston which mitigate the discrepancy between surface $NO_x$ of the model and in-situ data.

This study is motivated by the need to examine the possibility of using tropospheric $NO_2$ columns from the Ozone Monitoring Instrument (OMI) in order to improve the most recent national emission inventory for southeastern Texas. This work will develop the Bayesian inverse modeling which might outperform direct scaling (e.g., Kemball-Cook et al., 2015) because of incorporating the a priori values and associated uncertainties. In other words, we have some expectations of finding a reasonable optimized emissions inventory based on the a priori emission inventory. The emissions inventory also accounts for the uncertainty level of observations (which is different for different retrieval $NO_2$ columns).

Additionally, biogenic VOC emissions are temperature- (T) and solar radiation-dependent (Guenther et al., 1995). High temperatures enhance isoprene emissions leading to increased production of carbon monoxide (CO) and formaldehyde (HCHO) (e.g., Atkinson and Arey, 1998; Guenther et al., 1999; Palmer et al., 2003; Millet et al., 2006). Oxidation of short-lived biogenic VOCs (e.g., Sharkey et al., 1999; Pfister et al., 2008) affects the distribution of HCHO and CO over the US (e.g., Hudman et al., 2009; Choi et al., 2010). In section 4.2 below, we evaluate concentrations of HCHO using measurements from the DISCOVER-AQ Houston aircraft project.

Finally, changes in $NO_x$ and/or HCHO concentrations are known to affect the atmospheric chemical environment by changing $NO_x$/VOC ratios. This ratio in turn affects the production rate of $O_3$ (e.g., Martin et al., 2004; Duncan et al., 2010; Choi et al., 2012) and weekly cycles of surface $O_3$ as shown in our previous study (Choi et al., 2012). In section 4.2 below, we investigate how the $NO_x$/VOC ratio changes with changes in emissions. Particularly the $NO_2$/HCHO ratio derived from remotely sensed satellite retrievals (e.g., GOME-2 and/or OMI) are utilized to represent the chemical environment for the time periods of interest (e.g., Choi et al., 2012; Choi, 2014).

## 1.2. Objectives

Constrainment of $NO_x$ emissions in four different source categories (e.g., area, point, mobile and soil) is performed for urban areas of Southeast Texas during the Deriving Information on Surface Conditions from Column and Vertically Resolved Observations Relevant to Air Quality **(**DISCOVER-AQ) time period (September 1 -30, 2013) In order to perform this analysis, we estimate a posteriori emissions using an inverse method with remote sensing data that incorporates error uncertainties from modeling and remote sensing. We also examine how a posteriori $NO_x$ emissions affect modeling biases over Southeast Texas.

Our main objectives are:

1) Quantify a posteriori $NO_x$ emissions from aggregate surface a priori emissions from four source categories (point, area, mobile, and soil) by an inverse method using satellite $NO_2$ columns.
2) Evaluate model-simulated HCHO concentrations using in-situ aircraft measurements.
3) Examine how ratios of $NO_2$ (a proxy for $NO_x$) to HCHO (a proxy for VOC) vary.
4) Additionally, examine how in-situ measurement-adjusted meteorology improves model predictions.

## 1.3. Tasks

The following tasks were completed for this work:

1) Submission of Scope of Work with budget
2) Submission of Monthly and Quarterly Technical Reports
3) Implementation of a grid method and removal of a priori profile effects from OMI $NO_2$ observations
4) Implementation of a "Brute Force" method to conduct sensitivity runs in CMAQ
5) Implementation of Objective Analysis to improve simulated meteorology
6) Integration of NEI-2011 into SMOKE
7) Implementation of an inverse modeling method
8) Submission of a Draft Final Report and Final Report (this document)

# 2. Measurements, Methodology and Model Setup

## 2.1. Measurements

### *Surface Data*

Surface observational data were taken from the Continuous Ambient Monitoring Stations (CAMS) operated by the Texas Commission on Environmental Quality (TCEQ). The CAMS measurement network collects real-time meteorology and pollutant concentration data. Measured parameters differ from station to station. The station density in southeast Texas is relatively high. For example, the number of sites having meteorological, ozone and $NO_x$ measurements were 63, 52 and 30 respectively in the 4-km modeling domain during the 2013 DISCOVER-AQ campaign. The location and status of these sites measuring ozone and nitrogen oxide are shown in Figure 1. CAMS data are archived by our group using a data-spider coded in Interactive Data Language (IDL) which queries and downloads CAMS data from the TCEQ website.

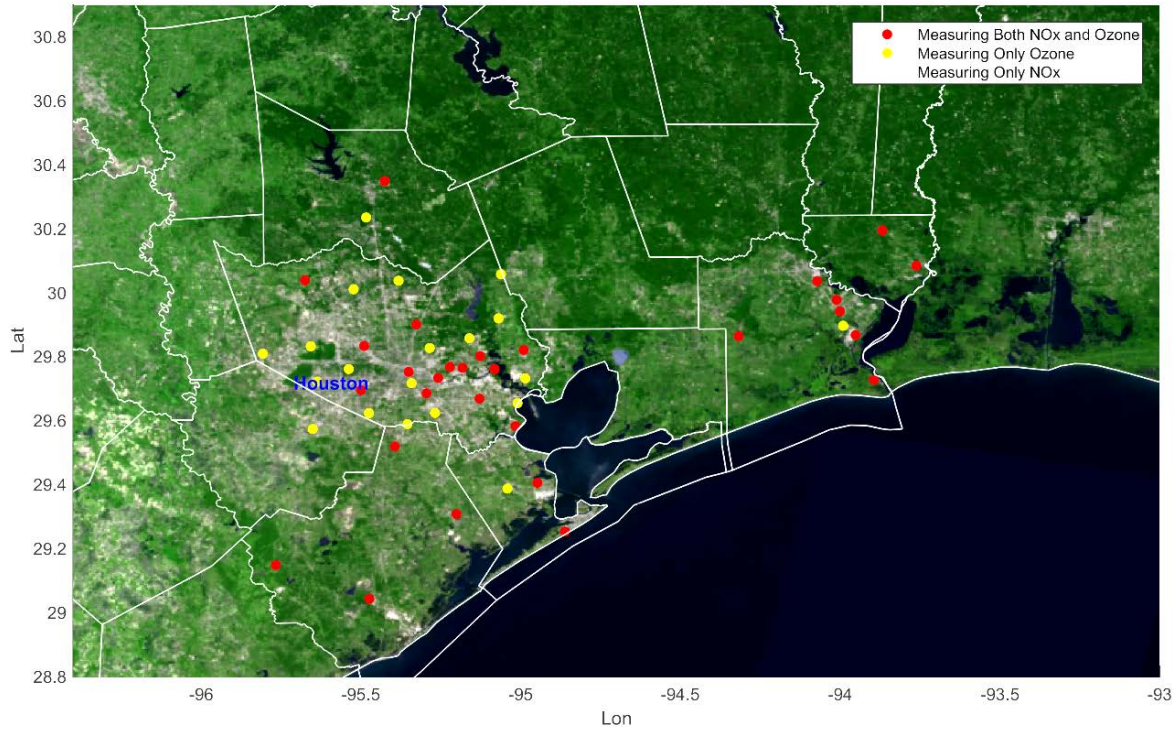


**Figure 1.** MODIS true color image showing the location and status of the CAMS ozone and $NO_x$ sites during 2013 DISCOVER-AQ in our 4-km modeling domain.

Comparison of CAMS data with model predictions is relatively straightforward. The first step is to extract model variables (NO and $NO_2$ concentrations) at the surface. The CMAQ model outputs binary Network Common Data format (NetCDF) files and we have developed in-house

code to extract data for any variable at any layer. To compare model values with observations from a site, we first use the latitude and longitude of a site to determine its cell location in the model grid, then extract the corresponding value from model output. The temporal frequency of model output and CAMS are both hourly, making direct comparison easy.

## *Aircraft*

Aircraft measurements are available online from NOAA aircraft P-3B as part of the DISCOVER-AQ campaign. The latest version of P-3B data have over 100 parameters, merged from measurements from a number of instruments on board. Data are available for 10 flight days during the DISCOVER-AQ campaign. Aircraft data can be downloaded from the DISCOVER-AQ website. Comparison of aircraft data with model output is more complicated but the idea is essentially the same: find model data matching the location and time of the aircraft measurements. We have developed an in-house code to match model and aircraft or ozonesonde data. Since aircraft data have much higher frequency (15 sec) than model output, we aggregate all aircraft data points in one grid cell during a 1-hour period to match model output.

## *Remote Sensing Data*

NASA OMI tropospheric $NO_2$ (Level 2, V2.1) data is used for this project. Level 2 data is data that has been post-processed from Level 1 (calibrated, georeferenced but unprocessed instrument data) into derived geophysical variables. Comparison of OMI data to Dutch OMI $NO_2$ (DOMINO) indicated that the NASA product was more consistent with validation studies. It has the ground footprint varying across the instrument swathe from 13×24 $km^2$ at nadir to ~40×160 $km^2$ for the edge of the orbit due to wide field of view angle and swathe width (i.e., panoramic). A detailed description of $NO_2$ retrieval algorithms can be found in Bucsela et al. (2013). Acquired spectra sensed by OMI detectors are analyzed with the Differential Optical Absorption Spectroscopy (DOAS) method in a fitting window from 405 nm to 465 nm. Calculated $NO_2$ slant column densities are then corrected for instrumental defects. This is named “destriping” due to variability of effects across the orbital track. In order to convert $NO_2$ slant column densities to vertical ones, Air Mass Factors (AMFs) which are functions of temperature, cloud cover, topography, albedo, and other factors are calculated using a pre-computed scattering-weight table from NASA’s TOMRAD forward vector radiative transfer model and monthly mean $NO_2$ profiles from the Global Model Initiative (GMI) simulation (here GEOS-CHEM in a $2.5^o \times 2.5^o$ grid). The uncertainties of the product vary from location to location and under different meteorological conditions. The overall error on the tropospheric vertical column density is <30% under clear-sky and typical polluted conditions ($>1\times10^{15}$ molecules $cm^{-2}$) (Bucsela et al., 2013).

Daily granules of tropospheric OMI $NO_2$ are available online at a NASA website. http://mirador.gsfc.nasa.gov/cgi-bin/mirador/collectionlist.pl?keyword=omno2.

Three important steps in preprocessing are:

1) Masking pixels having low quality

The common thresholds for performing the mask are: Solar Zenith Angle $0 \leq SZA \leq 85^0$, Vertical Column Density (VCD) Quality Flags=0, Root Mean Squared Error of Fit < 0.0003, Terrain Reflectivity < 30% and Cloud Fraction < 20%. Note that pixels which do not satisfy the criteria mentioned in the sections below are filtered out in this study.

2) Removing the vertical a priori profile impact from the granules to conduct an "apples-to-apples" comparison between model predictions and satellite data: A direct comparison of model output to OMI $NO_2$ requires that an a priori vertical profile of $NO_2$ in the OMI retrieval algorithm (here $2.5^o \times 2.5^o$ monthly averaged profiles from GEOS-Chem) be minimized. Although this coarse initial approximation could bias results, the main problem lies in applying an average semi-polluted profile over the large grid cell that encompasses both urban and rural regions, resulting in an underestimation of $NO_2$ vertical columns in urban regions and an overestimation in rural regions (Russell et al., 2011). Following the approach described in Duncan et al. (2014), we first use the variable called "scattering weight" provided for various pressure levels from the surface to the top of the atmosphere that is included in OMI $NO_2$ data files. We sum over all model layers the product of the scattering weight and model partial column (molecules $cm^{-2}$) in each model layer (up to the tropopause pressure provided in OMI Hierarchical Data Format (HDF) file). This sum divided by vertical column density of the model is called the air mass factor (AMF) of the model ($AMF_{model}$). Subsequently, we divide the product of VCD and AMF of satellite data from the HDF data file by $AMF_{model}$ to obtain a modified form of vertical column density of the satellite through the following equation (e.g., Martin et al., 2003; Duncan et al., 2014):

$$VCD'_{satellite} = (VCD_{satellite} \times AMF_{satellite})/AMF_{model} \quad (1)$$

Now we can directly compare $VCD'_{satellite}$ to model output. A bilinear interpolation method is used to co-register pixels between the satellite and model.

3) Gridding granules in high spatial resolution

As outlined earlier, pixels located far from nadir experience very poor spatial resolution. In order to make a smooth and uniform gridding, a recent novel method (Kuhlmann et al., 2014) is deployed. Using a parabolic spline method in this new approach, $NO_2$ maps become smoother and pixels at the extremities of scan line are more accurately reconstructed. Traditionally, oversampling was the main approach to ensure high spatial resolution for regional analysis. Here, this approach is used for the first time in the field of inverse modeling without extensive oversampling.

## 2.2. Inverse Modeling

Inverse modeling is a standard tool for combining observations of atmospheric composition with knowledge of atmospheric processes (e.g. transport, chemistry) to derive quantitative constraints on emissions to the atmosphere. The forward model chosen for our study is the CMAQ's Chemistry Transport Model (CCTM). This model solves the continuity equation to predict concentrations as a function of emissions. The inverse model then optimizes emission estimates by fitting the CTM to observed concentrations, subject to error weighting and a priori information on emissions. One approach is to define an optimal value of emissions as that which minimizes an error-weighted least squares (chi-square) scalar cost function derived from Bayes' theorem with the assumption of Gaussian errors. Generally, in this approach, we are looking to relate the probability density function (*pdf*) of observations (in this case ambient measurements or satellite retrievals) to that of the unknowns (emissions). One reason to consider the *pdf* is that there may be some a priori information about the unknowns (here original NEI-2011) which can be used to constrain the solution as a virtual measurement.

An ordinary least-squares inversion is susceptible to amplifying noise in observations due to mapping directly the unknowns to observations without having any knowledge of the unknown quantity a priori, i.e. with no constraints. The Bayesian approach is theoretically sound since we have some a priori expectation or understanding of the unknowns. Since the a priori knowledge is not zero (we have a very thorough but moderately inaccurate inventory), the first iteration can contribute to reducing estimation error. One concern is that the a priori estimate is not always reliable so that using it might lead to erroneous results. However, the level of imperfection of the a priori estimate can itself be expressed as a *pdf*. The Bayesian inversion method incorporates these effects. According to Bayes' theorem, the a posteriori *pdf* of unknowns for a measurement, *P (x|y)* (or the *pdf* of x given some knowledge about y), is given by:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{2}$$

Solving for the left hand side is the objective of this study: to update the a priori estimate *P(x)* of unknowns with knowledge from observations (y). *P (y|x)* (the *pdf* of y given x) describes the knowledge of *y* that would be estimated if the unknown is *x*. This can be obtained by the forward model and the statistical description of error in the observations. *P(y)*, the *pdf* of y, can be calculated by integrating the expression for *P (y|x) P(x)* over all *x* so that it can be regarded as a normalizing factor. By assuming that $NO_x$ emissions and tropospheric $NO_2$ concentrations are linearly correlated and all of the *pdf*s are Gaussian, we obtain the maximum probability value for *x*.

The relationship between the observation vector *y* (here OMI) and state vector *x* (here emissions) can be described as:

$$y = Kx + \varepsilon \tag{3}$$

Where the *K* matrix (Jacobian matrix) represents $NO_x$ sensitivities to the state vector defined by the CMAQ model, and ε is the error term. In order to calculate the Jacobian matrix for each source category, we employ a "Brute Force" method. In this method, sensitivity is measured based on the changes in corresponding gas concentration ($NO_2$) respect to emission changes ($NO_x$). Mathematically, the sensitivity can be expressed by:

$$S = \frac{NO_2^{+d} - NO_2^{-d}}{2d} \quad (4)$$

Where *d* is the fraction of change and $NO_2$ is the simulated tropospheric $NO_2$ column obtained from OMI. We set *d* to 100% which means two simulations: one with double $NO_x$ emissions and one without $NO_x$ emissions conducted for each source category, as well as total emissions. The Decoupled direct method (DDM) which is able to calculate local gradients of gases to emissions can also be used. Although the underlying theory of DDM is sound and has appeared to work better than the Brute-Force in some cases resulting from possible noise or numerical diffusion (e.g., Napelenok et al., 2006), our tests demonstrated that both methods generate close results in case of sensitivity of $NO_2$ concentrations to $NO_x$ emissions, The reason for not considering second-order or higher orders of the sensitivities is mainly due to the linearity assumption of the Bayesian inverse modeling (same as Kalman filter). Moreover, incorporating high order gradients in least-squares constrain, or extended Kalman filter requires iterative updating which poses a significant computational burden.

Using Bayes' theorem, we have:

$$-2lnP(x|y) = (y - Kx)^T S_o^{-1}(y - Kx) + c_1 \quad (5)$$

Where $c_1$ is a constant and $S_o$ is the observations error covariance matrix. The a priori knowledge of unknowns can be expressed by:

$$-2lnP(x) = (x - x_a)^T S_\varepsilon^{-1}(x - x_a) + c_2 \quad (6)$$

Where $x_a$ is the a priori value of *x*, $c_2$ is a constant and $S_\varepsilon$ is the associated error covariance matrix of the emissions. Uncertainties for each source category were set to 50% for area, mobile and point sources and to 300% for biogenic soil emissions. Uncertainty for OMI was based on the uncertainty provided in instrument retrieval products (Bucsela et al., 2013). After mathematical computations, the a posteriori state ($\hat{x}$) vector can be given by:

$$\hat{x} = x_a + \left(K^T S_\varepsilon^{-1} K + S_o^{-1}\right)^{-1} K^T S_\varepsilon^{-1}(y - Kx_a) \quad (7)$$

Where $x_a$ is the a priori state vector, $S_\varepsilon$ is the estimated error covariance matrix for $x_a$, and $S_o$ is the error covariance matrix for observation errors.

We applied the above method to update $NO_x$ emissions using OMI $NO_2$ retrievals with CMAQ as the forward model. An interesting point here is that all satellite "retrievals" use similar methods to obtain (retrieve) concentrations of trace gases in the atmosphere (state) from measured radiances (observed) using known spectral relationships. The forward model in this case would use measurements of trace gas concentrations (which are unavailable) to predict radiances. Since $NO_x$ emissions in NEI-2011 are divided into four source categories (area, soil, mobile and point sources), a total of 8 CMAQ sensitivity simulations are needed to update emissions. An additional CMAQ sensitivity simulation was performed to update total $NO_x$ emissions.

## 2.3. Model Simulation

The WRF and CMAQ simulation period is selected to be **September 1 -30, 2013**, which is the NASA's DISCOVER-AQ 2013 field campaign period.

### *WRF*

*Domain setup*

The WRF domains have sizes of 161×145 for 12-km domain, and 97×79 for 4-km domain. WRF domains are shown in Figure 2 as red and blue boxes. The geographic parameters for the domains are listed in Table 1.

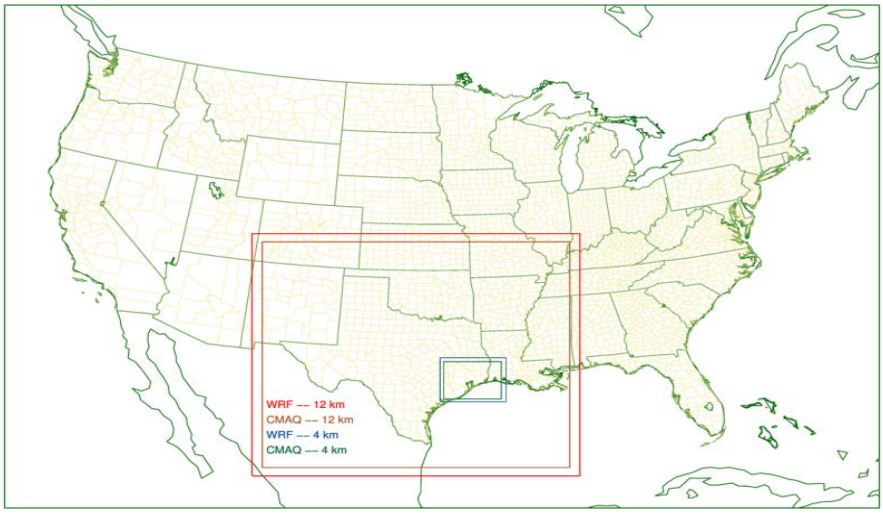


**Figure 2.** WRF (thick lines) and CMAQ (thin lines) used for the UH Air Quality Forecasting (AQF) System. There are two domains: the 12-km Texas domain and the 4-km Houston-Galveston-Brazoria (HGB) domain.

**Table 1.** Projection Parameters and Domain Origin

| | |
|---|---|
| First True Latitude (Alpha) | 33°N |
| Second True Latitude (Beta) | 45°N |
| Central Longitude (Gamma) | -97°W |
| Projection Origin | (31.55113°N, -98.13650 °W) |
| 12km WRF domain Lower Left Corner [x, y] (m) | [-1007980, -1739860] |
| 4km WRF domain Lower Left Corner [x, y] (m) | [56020, -1251860] |

Both WRF and CMAQ share the same vertical structure since no layer collapsing has been employed in the Meteorology Chemistry Interface Processor (MCIP). The vertical structure is listed in Table 2.

**Table 2.** Vertical layer structures of WRF and CMAQ used for the modeling

| Layer | Height Above Ground Level (AGL, m) | Layer | AGL(m) |
|---|---|---|---|
| 1 | 32.4 | 15 | 1517.8 |
| 2 | 81.2 | 16 | 1751.4 |
| 3 | 163.1 | 17 | 1990 |
| 4 | 245.9 | 18 | 2233.9 |
| 5 | 329.5 | 19 | 2534.7 |
| 6 | 413.7 | 20 | 3164.8 |
| 7 | 498.4 | 21 | 4193.1 |
| 8 | 583.8 | 22 | 5415.3 |
| 9 | 669.7 | 23 | 6964.2 |
| 10 | 756.2 | 24 | 9083.3 |
| 11 | 887.2 | 25 | 11444.6 |
| 12 | 1019.6 | 26 | 14549.2 |
| 13 | 1153.4 | 27 | 16540.7 |
| 14 | 1288.8 | | |

*Input analysis data*

We have evaluated existing analysis datasets and decided to use NCEP's (National Centers for Environmental Prediction) NARR (North American Regional Reanalysis) as input. The NARR data are based on an NCEP Eta 221 regional North American grid (Lambert Conformal) NCEP Eta 221 regional North American grid (Lambert Conformal) at 29 pressure levels. Its horizontal resolution is 32-km and the frequency is 3-hourly. An alternative to NARR is the Eta-North American Mesoscale Model (NAM) analysis data. However, data frequency is reduced from every three hours to every six hours starting in 2013. Our validation tests showed it is not as good as NARR for WRF input, probably because of lower temporal resolution.

*Major WRF configurations*

Implemented WRF options are shown in Table 3 below. The first approximation and boundary conditions will be from NCEP NARR analyses. Grid nudging is turned on with the same NARR analysis data.

**Table 3.** WRF physics options

| | |
|---|---|
| WRF Version | V3.6.1 |
| Microphysics | Lin et al. Scheme |
| Long-wave Radiation | Rapid Radiative Transfer Model –Global Climate Applications (RRTMG) |
| Short-wave Radiation | New Goddard scheme |
| Surface Layer Option | Monin-Obukhov with Compressible Boundary layer (CB) viscous sublayer scheme |
| Land-Surface Option | Unified Noah Land Surface Model |
| Urban Physics | None |
| Boundary Layer | Yonsei University (YSU) |
| Cumulus Cloud Option | Kain-Fritsch |
| Four Dimensional Data Assimilation (FDDA) | Grid and 1-hr observation-nudging |

*Nudging (grid, surface and observation nudging)*

Obs-nudging is regarded as a low-cost and effective method for improving meteorological model performance but it requires additional observational data. In this study, we acquire input observation data and generating files in little_r format using codes developed in-house at the University of Houston (UH). Observational data come from the Meteorological Assimilation Data Ingest System (MADIS) and TCEQ CAMS network. MADIS is a NOAA program that collects, integrates, quality-controls, and distributes observations from NOAA and non-NOAA organizations. Four MADIS datasets used for obs-nudging are NOAA Profiler Network (NPN), Cooperative Agency Profilers (CAP), Aviation Routine Weather Report (METAR) and NOAA Radiosonde (RAOB). Most of the observation data are available at hourly frequency. Therefore, we implemented hourly observation nudging despite input analysis data being at 3-hourly frequency. Key settings for hourly obs-nudging are “intf4d” in “namelist.oa”, and “auxinput11_interval”, “sgfdda_interval_m” in “namelist.input”.

## *CMAQ*

*Emission processing*

Emission modeling was performed with the Sparse Matrix Operator Kernel Emissions (SMOKE) model. The 2008 National Emission Inventory (NEI) generated by the Environmental Protection Agency (EPA) (with the adjustment of mobile and point source changes using EPA national emission trends from 2008 to 2013, Czader et al., 2015; Pan et al., 2015; Li et al., 2015) was used to estimate hourly emission rates from anthropogenic sources for the continental U.S. (for 12-km). The 2011 NEI was used for Southeast Texas domains (for 4-km). Emissions from natural sources were estimated with BEIS3 (Biogenic Emission Inventory System version 3). Mobile emissions were processed with MOVES. Various surface $NO_x$ emissions were prepared for inverse modeling. Again, in this project, we used the latest NEI-2011 emissions “as is” for

4km domain, that is, without adjusting for possible emission changes. A brief summary of the emissions data used in this emissions modeling platform follows:

- 2011 platform v6.1 represents all platform source categories (see Table 4) other than onroad mobile sources
- For onroad mobile source emissions, the latest 2011 platform v6.2 based on the latest Motor Vehicle Emissions Simulator (MOVES) 2014 was used.

Table 4 lists source categories that we used to represent the year 2011 air pollutant emissions for this emission modeling system and notes on data preparation.

**Table 4**. Platform source categories for the NEI 2011 platform

| Platform Source category | Abbreviation | Description |
|---|---|---|
| Electricity Generating Units (EGU) non-peaking units | Ptegu | 2011 NEI point source EGUs determined to operate as non-peaking units. Hourly 2011 CEMS are not used since the year of modeling is 2014. |
| EGU peaking units | ptegu_pk | Same as ptegu source category but limited to EGUs that are determined to operate as peaking units. Hourly 2011 CEMS are not used since the year of modeling is 2014. |
| Point source oil and gas | pt_oilgas | 2011NEIv1 point sources with oil and gas production emissions processes. Annual resolution. |
| Remaining non-EGU point | Ptnonipm | All 2011NEIv1 point source records not matched to the ptegu, ptegu_pk, and pt_oilgas source categories, except for offshore point sources that are in the "othpt" source category. Includes all aircraft emissions and some rail yard emissions. Annual resolution. |
| Agricultural | Ag | $NH_3$ emissions from 2011NEIv1 nonpoint livestock and fertilizer application, county and annual resolution. |
| Area fugitive dust | Afdust | Particulate matter less than size 10 and 5 microns respectively ($PM_{10}$ and $PM_{2.5}$) from fugitive dust sources from the 2011NEIv1 nonpoint inventory including building construction, road construction, and agricultural dust, and road dust. County and annual resolution. |
| Nonpoint oil and gas | np_oilgas | 2011NEIv1 nonpoint sources from oil and gas-related processes. County and annual resolution. |
| Residential Wood Combustion | Rwc | This is a new source category in 2011NEIv1. NEI nonpoint sources with Residential Wood Combustion (RWC) processes. County and annual resolution. |

| | | |
|---|---|---|
| Class 1 & 2 CMV and locomotives | c1c2rail | Locomotives and primarily category 1 (C1) and category 2 (C2) commercial marine vessel (CMV) emissions sources from the 2011NEIv1 nonpoint inventory. Midwestern states' CMV emissions, including Class 3 sources, are from a separate year 2010 emissions inventory. County and annual resolution. |
| Commercial marine | c3marine | Category 3 (C3) CMV emissions projected to 2011 from year 2002 values. These emissions are not from the 2011NEIv1, but rather were developed for the rule called "Control of Emissions from New Marine Compression-Ignition Engines at or Above 30 Liters per Cylinder", usually described as the Emissions Control Area- International Maritime Organization (ECA-IMO) study. Emissions Control Area-International Maritime Organization (ECA-IMO) study. (EPA-420-F-10-041, August 2010). U.S. states-only emissions (zero in Midwest); see othpt source category for all non-U.S. emissions. Treated as point sources to reflect shipping lanes, annual resolution. |
| Remaining nonpoint | nonpt | 2011NEIv1 nonpoint sources not otherwise removed from modeling or included in other platform source categories; county and annual resolution. |
| Nonroad | nonroad | 2011NEIv1 nonroad equipment emissions developed with the National Mobile Inventory Model (NMIM) using NONROAD2008 version NR08a. NMIM was used for all states except California and Texas, which submitted their own emissions to the 2011NEIv1. County and monthly resolution. |
| Onroad RatePerDistance | Rateperdistance | EPA ran MOVES2014 for 2011 in emissions factor mode. The MOVES lookup tables include on-network (RatePerHour) to represent exhaust and most evaporative emissions during running, tirewear, and brake wear modes. These data include the reference county and reference fuel month assignments that EPA used for the MOVES simulation. |
| Onroad RatePerHour | Rateperhour | EPA ran MOVES2014 for 2011 in emissions factor mode. The MOVES lookup tables include off-network (RatePerDistance) that represents emissions from extended idling and Auxiliary Power Unit (APU) operation. These data include the reference county and reference fuel month assignments that EPA used for the MOVES simulation. |

| | | |
|---|---|---|
| Onroad RatePerVehicle | Ratepervehicle | EPA ran MOVES2014 for 2011 in emissions factor mode. The MOVES lookup tables include off-network starts/stops (RatePerVehicle) that represents emissions from start exhaust and most evaporative emissions that occurs off-network. These data include the reference county and reference fuel month assignments that EPA used for the MOVES simulation. |
| Onroad RatePerProfile | Rateperprofile | EPA ran MOVES2014 for 2011 in emissions factor mode. The MOVES lookup tables include off-network (RatePerProfile) that represents emissions from evaporative fuel vapor venting. These data include the reference county and reference fuel month assignments that EPA used for the MOVES simulation. |
| Non-US Point | Othpt | Point sources from Canada's 2006 inventory and Mexico's Phase III 2012 inventory, annual resolution. Mexico's inventory is year 2012 and grown from year 1999 (ERG, 2009; Wolf, 2009). Also includes all non-U.S. C3 CMV and U.S. offshore oil production, which are unchanged from the 2008 NEI point source annual emissions. |
| Non-US nonpoint and nonroad | Other | Annual year 2006 Canada (province resolution) and year 2012 (grown from 1999) Mexico Phase III (municipio resolution) nonpoint and nonroad mobile inventories. |
| Non-US onroad | Othon | Year 2006 Canada (province resolution) and year 2012 (grown from 1999) Mexico Phase III (municipio resolution) onroad mobile inventories, annual resolution. |
| Biogenic | Biog | No updates made (Stay constant) |

As a QA/QC check, we provide area source category comparison of the two inventories below. Figure 3 shows area emissions from NEI2008 and NEI-2011, as well as differences in percentage. The reduction is evident – the NEI2008 large area emission in the DFW area (upper left corner) can hardly be seen in the NEI-2011. Intensity also decreased in the Houston area.

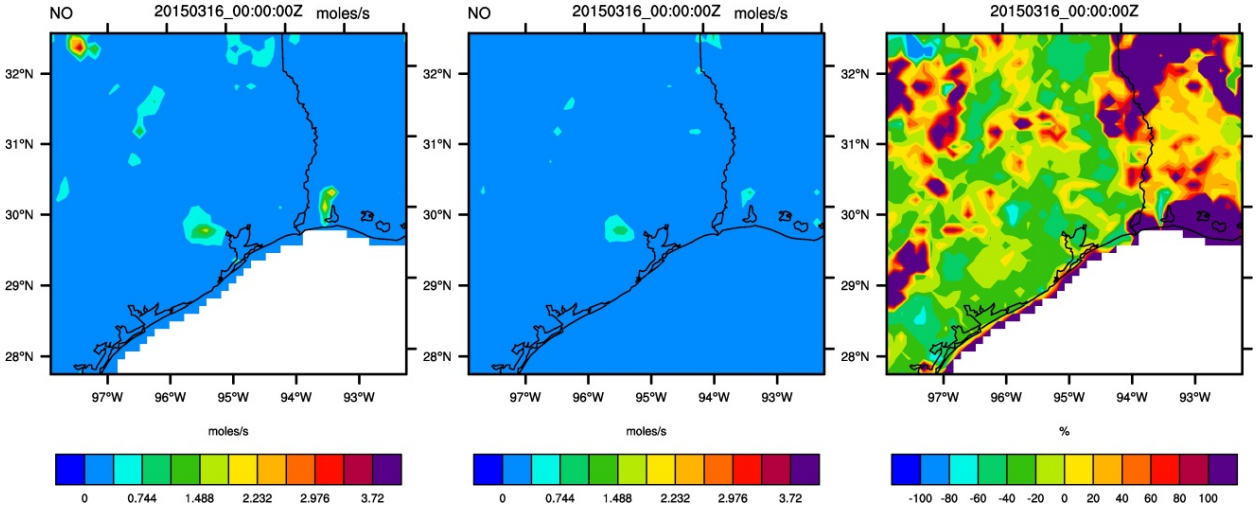

**Figure 3**. Area emissions. left: NEI2008, center: NEI-2011, and right: difference in percent.

*Generating meteorological input using MCIP*

Meteorological input for CMAQ was processed using UH-modified MCIP over WRF output. UH-modified MCIP corrected a few bugs and had minor enhancements over default MCIP.

*Major CMAQ configurations*

Major CMAQ configurations are shown in Table 5. All of these options have been tested by the UH modeling group.

**Table 5.** Major CMAQ options

| | |
|---|---|
| CMAQ version | V5.0.1 |
| Chemical Mechanism | cb05tucl_ae5_aq: CB05 gas-phase mechanism with active chlorine chemistry, updated toluene mechanism, fifth-generation CMAQ aerosol mechanism with sea salt, aqueous/cloud chemistry |
| Lightning $NO_x$ emission | Included by using inline code |
| Horizontal advection | Yamartino (YAMO (hyamo) |
| Vertical advection | WRF omega formula (vwrf) |
| Horizontal mixing/diffusion | Multiscale (multiscale) |
| Vertical mixing/diffusion | Asymmetric Convective Model version 2 (acm2) |
| Chemistry solver | Euler Backward Iterative (EBI) optimized for the Carbon Bond-05 mechanism (ebi_cb05tucl) |
| Aerosol | Aerosol Module version 5 (AERO) 5 for sea salt and thermodynamics (aero5) |
| Cloud Option | ACM cloud processor for AERO5 (cloud_acm_ae5) |
| Initial and Boundary Conditions (IC/BC) source | Default static profiles |

# 3. Evaluation of Meteorology Simulations

## 3.1 General Meteorological Conditions

The weather during the September 2013 simulation period was relatively dry with mostly southerly, easterly or southeasterly winds. From 09/05 to 09/19, there was a lack of influence of strong synoptic weather systems. Shifting wind patterns were observed during the period: light northeasterly in the early morning gradually turned clockwise to southeasterly in the afternoon and evening hours. In this period, winds shifted from south east to near east and there were more clouds after 09/10. The only cold front arrived early 09/21. Figure 4 shows daily regional average temperatures and periods marked with temperature drops. The "1Hr-OA" case clearly tracked observations better than the "No-OA" case. Although not very significant to photochemistry, temperature drop is usually a good proxy for the critical factors affecting ozone production or transport such as cloudiness, wind and precipitation.

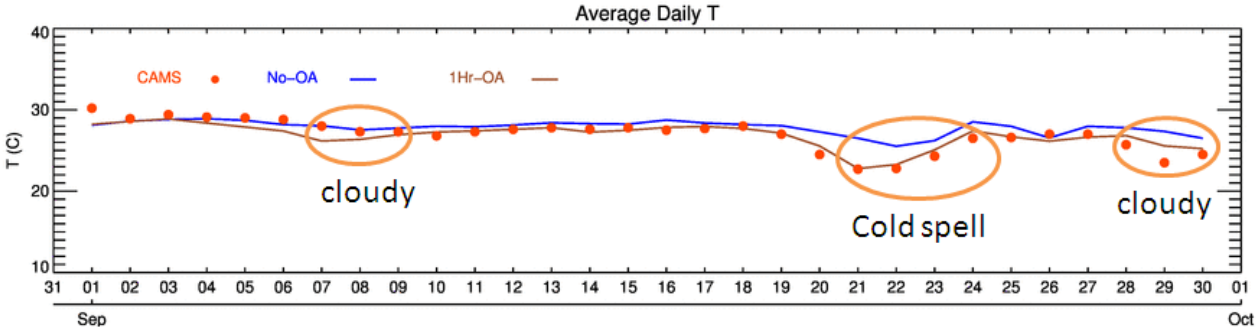


**Figure 4**. Regional daily temperature averaged over all available (typically around 1,200) hourly CAMS observations

Light rain events occurred on 09/02, 09/10, 09/16, 09/19 to 09/21 and 09/28 to 09/30. The 09/20 and 09/21 events consisted of widespread light to medium showers. Besides above-mentioned dates, there were a few other days with sporadic drizzles.

A majority of the days between 09/01 and 09/20 were mostly sunny to mostly cloudy. The periods from 09/08 to 09/10 and 09/18 to 09/20 had more clouds than other days. The period from 09/21 to 09/30 was influenced by the passage of a cold front. The days between 09/22 and 09/24 were sunny and cool. Then the surface wind direction reversed in mid-09/25 and brought clouds from 09/26 to 09/30.

## 3.2 Evaluation metrics

To evaluate performance of WRF simulations we used statistics commonly used by the modeling community, listed below. Observational CAMS data were used to validate model results.

*Correlation (r) between model values and observed values*

$$r = \frac{\sum_{t=1}^{n}[(x_t - \bar{x})(y_t - \bar{y})]}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 * \sum_{t=1}^{n}(y_t - \bar{y})^2}} \quad (8)$$

n = number of data points, x = observed values, y = model values, values with an over bar indicate the mean.

*Index of Agreement (IOA) between model values and observed values*

$$IOA = 1 - \frac{\sum_{t=1}^{n} {e_t}^2}{\sum_{t=1}^{n}(|y_t - \bar{x}| + |x_t - \bar{x}|)^2} \quad (9)$$

n = number of data points, $e_t = y_t - x_t$, x = observed values, y = model values, values with an over bar indicate the mean.

*Root Mean Square Error (RMSE)*

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} {e_t}^2} \quad (10)$$

n = number of data points, $e_t = y_t - x_t$, x = observed values, y = model values

*Mean Absolute Error (MAE)*

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (11)$$

n = number of data points, $e_t = y_t - x_t$, x = observed values, y = model values

*Mean Bias (MB)*

$$MB = \frac{1}{n}\sum_{t=1}^{n} e_t \quad (12)$$

n = number of data points, $e_t = y_t - x_t$, x = observed values, y = model values

## 3.3 Meteorological simulation results

We performed three sets of WRF simulations with different objective analysis settings: 1) no observation nudging is performed (No-OA); 2) observation nudging input is updated every 3 hours; and 3) observation nudging input is updated every 1 hour. The different OA frequencies were set in the namelist file of WRF while input files of different frequencies were generated by the OBSGRID program.

Although not very significant to photochemistry, temperature change is usually a good proxy for the critical factors affecting ozone production or transport such as cloudiness, wind, and precipitation. Table 6 lists statistics of hourly surface temperature and shows that the OA cases are clearly better than the no-OA case, with the correlation coefficient increasing by about 10% and IOA improving by 7%-8%. Biases for all cases are low, indicating excellent energy budget in the model. On the other hand, the two OA cases have similar statistics with "1Hr-OA" slightly superior.

**Table 6.** Statistics of hourly surface temperature

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| No-OA | 41058 | 0.83 | 0.89 | 2.0 | 1.5 | 0.9 | 27.4 | 28.3 | 3.1 | 2.8 |
| 3Hr-OA | 41058 | 0.93 | 0.96 | 1.2 | 0.9 | 0 | 27.4 | 27.4 | 3.1 | 3.1 |
| 1Hr-OA | 41058 | 0.94 | 0.97 | 1 | 0.8 | 0 | 27.4 | 27.4 | 3.1 | 3.1 |

N – data points; Corr – Correlation; IOA – Index of Agreement; RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MB – Mean Bias; O – Observation; M - Model; O_M – Observed Mean; M_M – Model Mean; SD – Standard Deviation; Units for RMSE/MAE/MB/O_M/M_M/O_SD/M_SD: degree C

Compared to temperature, hourly winds at local scale are difficult to predict by meteorological models such as WRF. Performance of the model is also greatly influenced by the quality of input analysis data. Running the model at a finer resolution can provide more local meteorological details than at coarser resolution. A fine resolution run usually does not alter average winds inherited from the coarse resolution run unless objective analysis is performed. Therefore, when large-scale winds from the input analysis differ from observations, there is little chance that the model can predict winds correctly. One remedy is to add objective analysis (observation nudging) during the WRF run.

Statistics for U-wind are displayed in Table 7 while those for V-wind are shown in Table 8. An alternative approach for calculating wind statistics is to compute statistics for wind speed and wind direction. The drawback for this approach is that wind direction is measured in degrees. For example, a direction of 5 degrees is actually close to the direction of 355 degrees but statistically, 5 and 355 are far apart.

The statistics indicate that model winds generally have lower correlation with observations than model temperature. Hence, evaluating wind statistics requires extra caution as correlation can be misleading for days with light winds. Specifically, low correlation on a day with light winds does not necessarily mean poor model performance since wind direction on a light wind day can vary from hour to hour somewhat randomly at a given site and it is usually hard for the model to capture these local-scale changes.

Statistics for U and V wind components are listed in Table 7 and 8 respectively. The statistics are based on CAMS data. Results are quite consistent across variables. Overall, winds are reasonably

well simulated after OA was performed. OA cases have significantly better IOA than the No-OA case, while differences between the 1Hr-OA and 3Hr-OA are quite small. A further examination of the numbers shows OA cases have smaller wind biases (lower winds) than No-OA case. The IOA improved by 6% for U-wind and 11% for V-wind.

**Table 7.** Statistics of hourly surface U wind

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| No-OA | 43246 | 0.76 | 0.84 | 1.4 | 1.1 | -0.6 | -1.3 | -1.9 | 1.6 | 1.9 |
| 3Hr-OA | 43246 | 0.79 | 0.88 | 1.1 | 0.8 | -0.3 | -1.3 | -1.6 | 1.6 | 1.6 |
| 1Hr-OA | 43246 | 0.81 | 0.89 | 1.0 | 0.8 | -0.3 | -1.3 | -1.6 | 1.6 | 1.6 |

**Table 8.** Statistics of hourly surface V wind

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| No-OA | 43246 | 0.76 | 0.8 | 2.1 | 1.7 | 1.2 | 0.4 | 1.7 | 2 | 2.6 |
| 3Hr-OA | 43246 | 0.77 | 0.88 | 1.3 | 1 | -0.1 | 0.4 | 0.4 | 2 | 2 |
| 1Hr-OA | 43246 | 0.8 | 0.89 | 1.2 | 0.9 | -0.1 | 0.4 | 0.4 | 2 | 2 |

N – data points; Corr – Correlation; IOA – Index of Agreement; RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MB – Mean Bias; O – Observation; M - Model; O_M – Observed Mean; M_M – Model Mean; SD – Standard Deviation; Units for RMSE/MAE/MB/O_M/M_M/O_SD/M_SD: m/s

The ability of OA to correct the wind field can be seen from Figure 5: (a) No-OA (b) 1Hr-OA. At 09:00 CST of 09/25, observed winds (orange) at most Houston area sites are from the northwest. The No-OA case shows northerly winds while the 1Hr-OA case has northwesterly winds. Therefore, the 1Hr-OA case matched much better with observations.

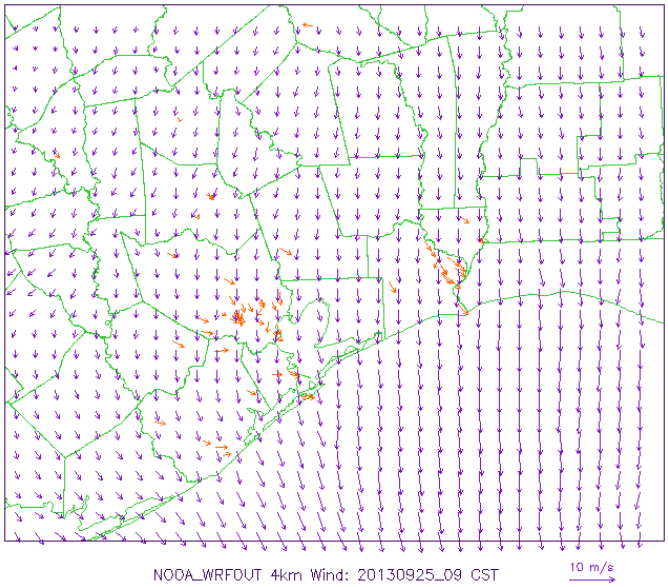


**Figure 5(a).** Winds of No-OA case at 09/25_09 CST, orange arrows show observations. The model has strictly northerly winds in Houston.

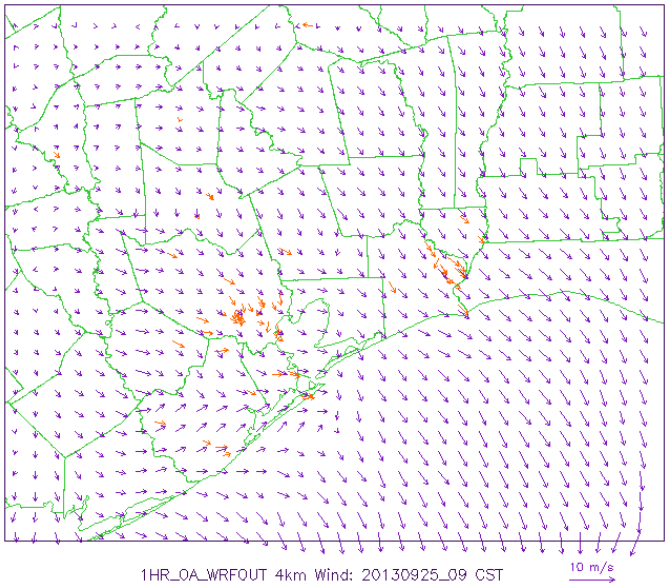


**Figure 5(b).** Winds of 1Hr-OA case at 09/25_09 CST, orange arrows show observations. The model has mostly northwesterly winds in Houston.

# 4. Inverse Modeling Results: $NO_x$ Emission and Model and Satellite $NO_2$ Comparison

As outlined in Section 2.2, a forward chemical transport model links the emissions to tropospheric $NO_2$ concentrations. For this study we used the USEPA's Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006) to simulate air quality on a 4km gridded domain during September 2013. As mentioned before, lateral boundary conditions were derived from NARR. The chemical boundary conditions for the 4-km CMAQ run were obtained from the University of Houston air quality forecasting system using a parent domain which had a grid resolution of 12 km. Our 4 km simulation model domain covering southeast Texas consisted of 84 and 66 grid cells in the x and y directions respectively with 27 vertical layers. As indicated previously, the emissions inventory for simulation was NEI-2011 processed with the Sparse Matrix Operator Kernel Emissions (SMOKE) and the MOVES (Motor Vehicle Emission Simulator).

In order to set up the inverse modeling framework, some variables including uncertainties of emissions and observations should be defined. Since each pixel in an OMI satellite image has a different uncertainty depending on meteorological conditions (e.g., air mass factor), a matrix of uncertainty is needed. Here we used a pre-defined uncertainty of tropospheric $NO_2$ provided in OMI granules. By following the error propagation rule, the uncertainty of each pixel based on all available granules was calculated. Since inverse modeling is sensitive to the uncertainty of observations, assuming fixed uncertainty might lead to a false change in emission inventory in areas where OMI observations have large uncertainty. Estimating error of emissions is not a straightforward task. We assumed a 50% uncertainty level for anthropogenic sources (Shim et al., 2005) and 300% for biogenic soil sources (Hudson et al., 2010).

## 4.1 Emissions

Based on the method explained in Chapter 2, we updated $NO_x$ emissions in NEI-2011 and called the updated version NEI-2011n. The old and updated $NO_x$ emissions by source categories are shown in Figures 6 through 10. The upper column plots are for old (priori, or NEI-2011) $NO_x$ emissions while the lower column shows the updated (posteriori, or NEI-2011n) $NO_x$ emissions.

### *Area Source Category*

$NO_x$ emission in area source category is shown in Figure 6. We can see that in the updated version, $NO_x$ emissions decreased significantly in both Houston and the Lake Charles region. There are only small changes in the rural area and Beaumont region. The largest decrease appears to be the two red grid cells in the middle of Harris County.

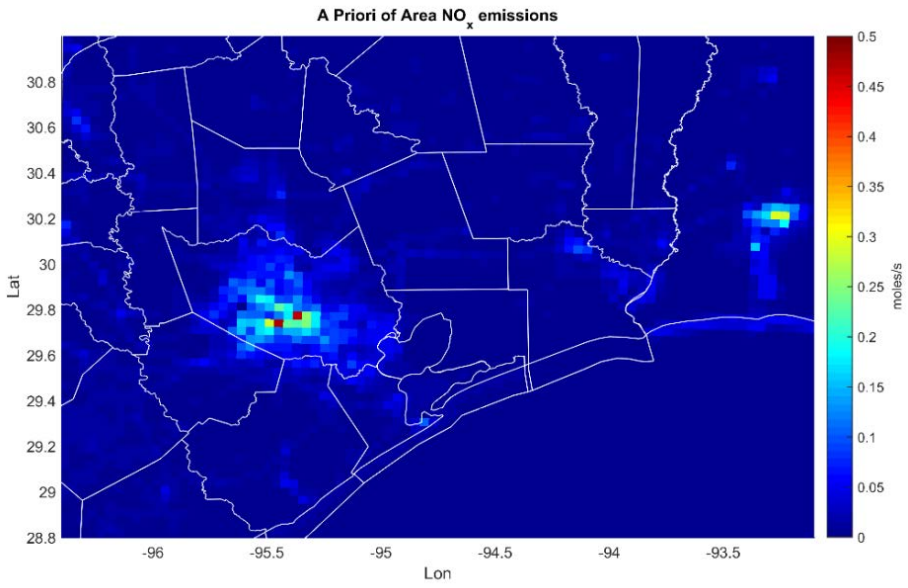


**Figure 6(a).** Area source $NO_x$ emissions estimates in southeast Texas, NEI-2011 (Old)

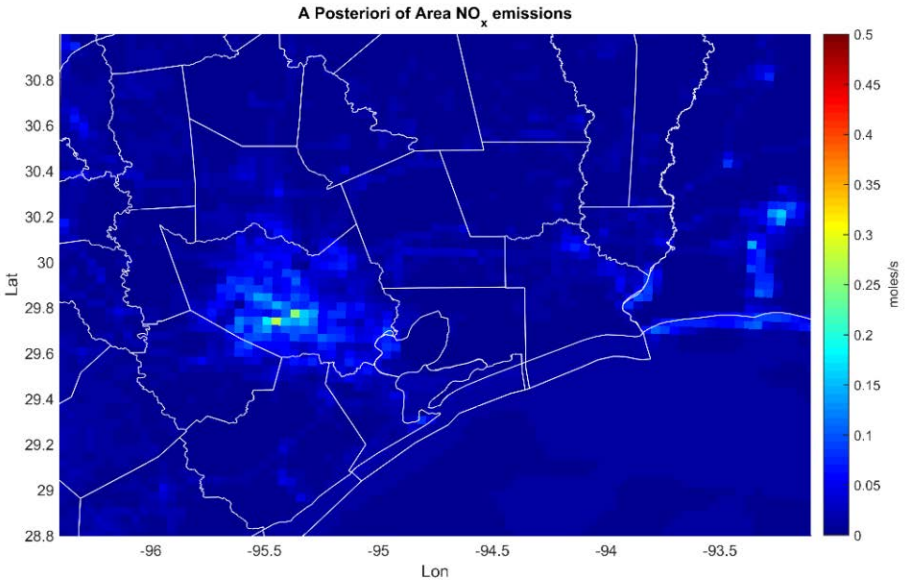


**Figure 6(b).** Area source category $NO_x$ emissions estimates in southeast Texas, NEI-2011n (Updated)

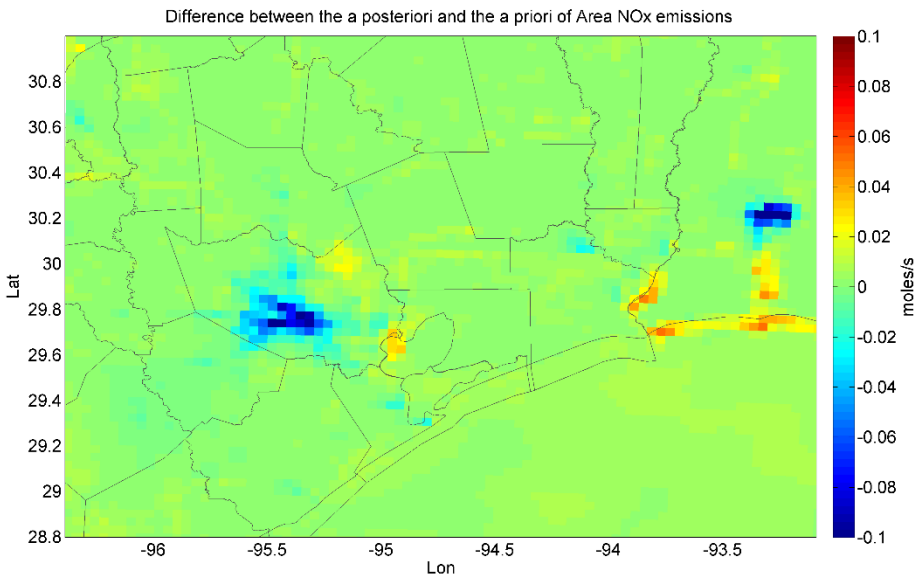


**Figure 6(c).** The difference between NEI-2011n and NEI-2011 for area source category.

## *Biogenic Soil Source category*

Figure 7 depicts $NO_x$ emissions from the biogenic soil source category. The original inventory has higher $NO_x$ emissions in the western and coastal regions of the modeling domain and lower emissions in urban Houston, the Piney Woods of eastern Texas and areas adjacent Louisiana. The updated emission shows enhanced biogenic soil $NO_x$ emissions in western and coastal regions and a rather small change in the Piney Woods region. This can be well explained by the tendency of earlier versions of the BEIS (Biogenic Emission Inventory System) model to underestimate biogenic emissions, since we used BEIS3 for biogenic soil processing.

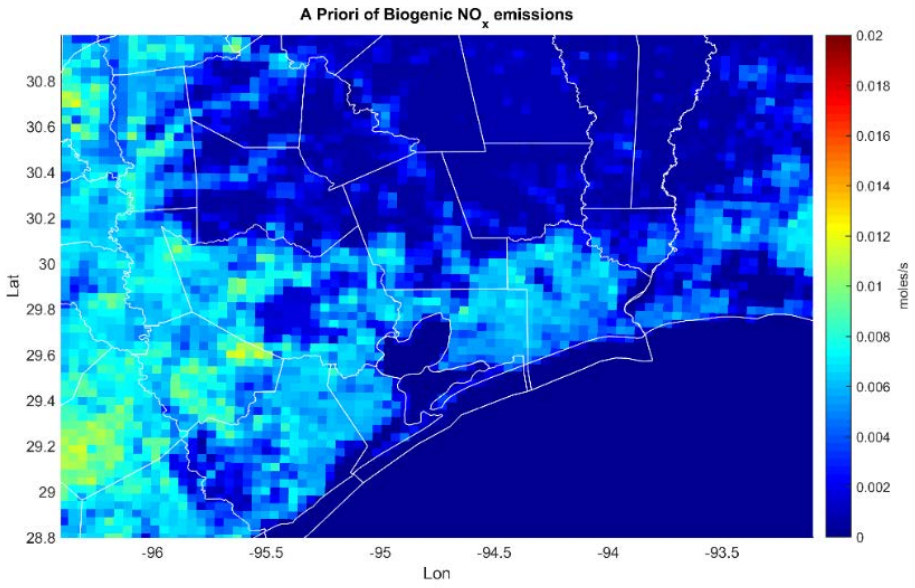


**Figure 7(a).** Biogenic soil $NO_x$ emissions estimates in southeast Texas, NEI-2011 (Old)

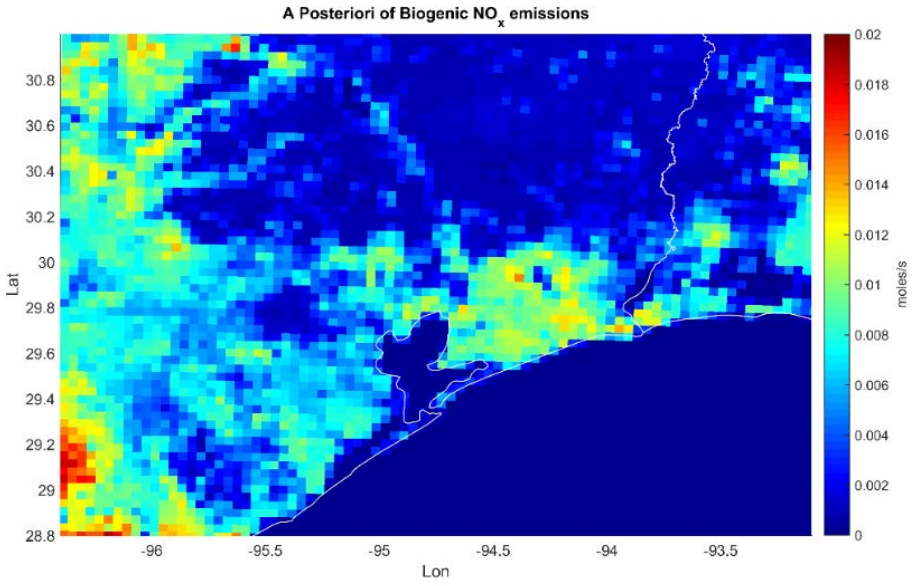


**Figure 7(b).** Biogenic soil $NO_x$ emissions estimates in southeast Texas, NEI-2011n (Updated)

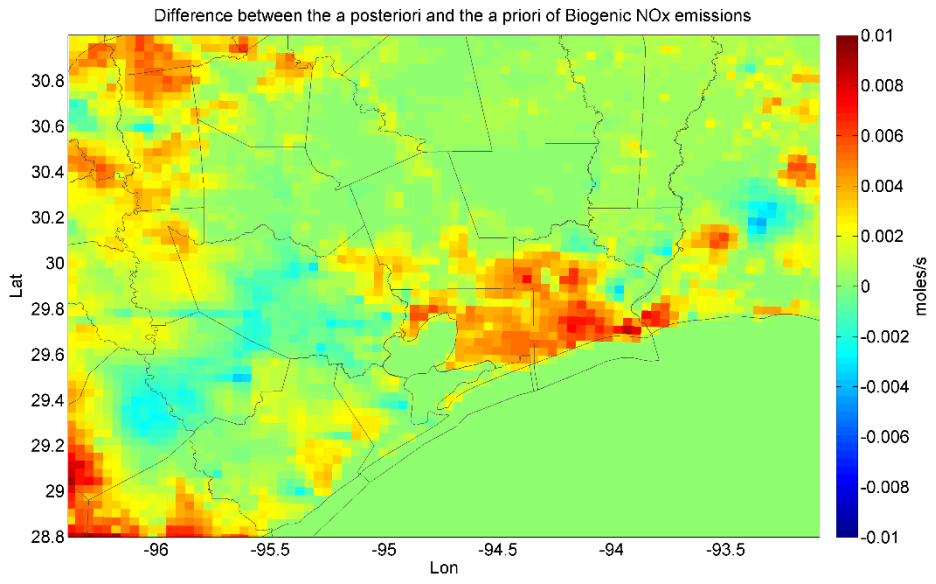


**Figure 7(c).** The difference between NEI-2011n and NEI-2011 for area source category.

## *Mobile Source Category*

Figure 8 shows estimates of $NO_x$ emissions from the mobile source category which makes up approximately 50-75% of total $NO_x$ emissions in Houston. As stated earlier, because of improvements in fuel technology and emissions control strategies and technology, a decreasing trend in mobile $NO_x$ emissions is expected. Therefore, a priori mobile $NO_x$ emissions in 2011 should be considerably reduced to match those of 2013. The a posteriori mobile $NO_x$ emissions have decreased substantially to a factor of 2 in center of Houston. It emphasized that an adjustment should be performed for future studies when using NEI-2011 for model simulation. This overestimation masks the real chemical condition of the cities for conducting any sensitivity analysis, and also makes it hard to quantify the relative importance of $NO_x$ emissions from other source categories such as power plants.

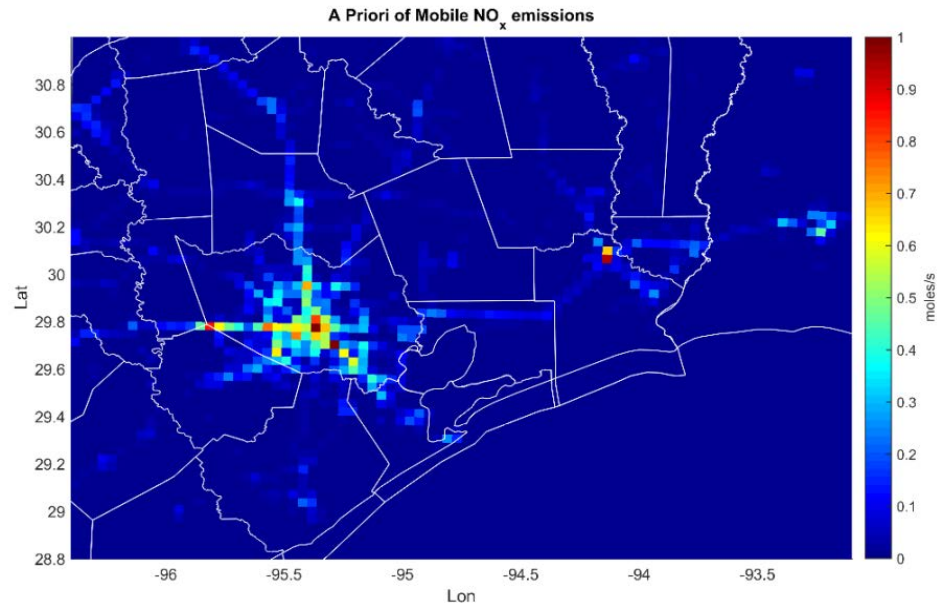


**Figure 8(a).** Mobile source category $NO_x$ emissions estimates in southeast Texas, NEI-2011 (Old)

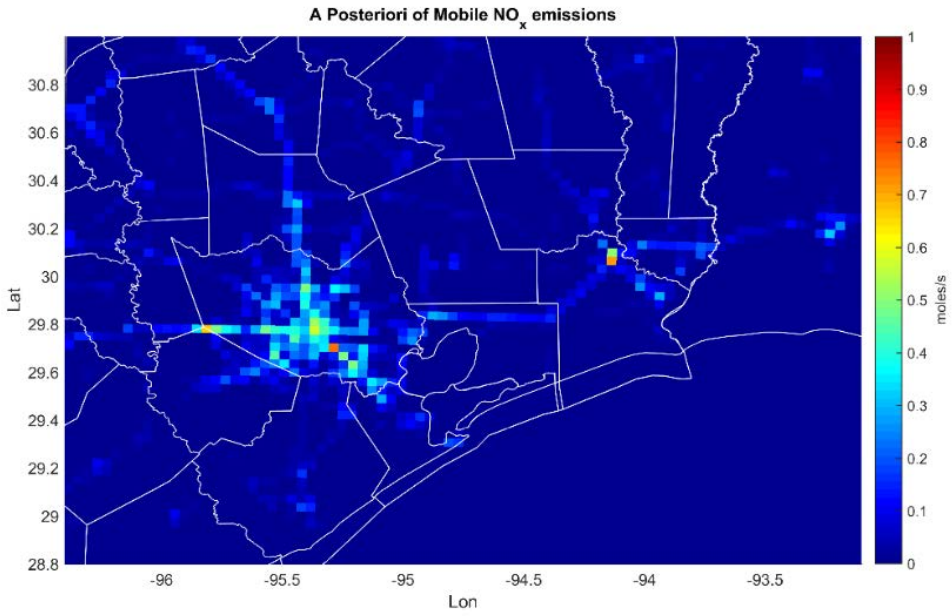


**Figure 8(b).** Mobile source category $NO_x$ emissions estimates in southeast Texas, NEI-2011n (Updated)

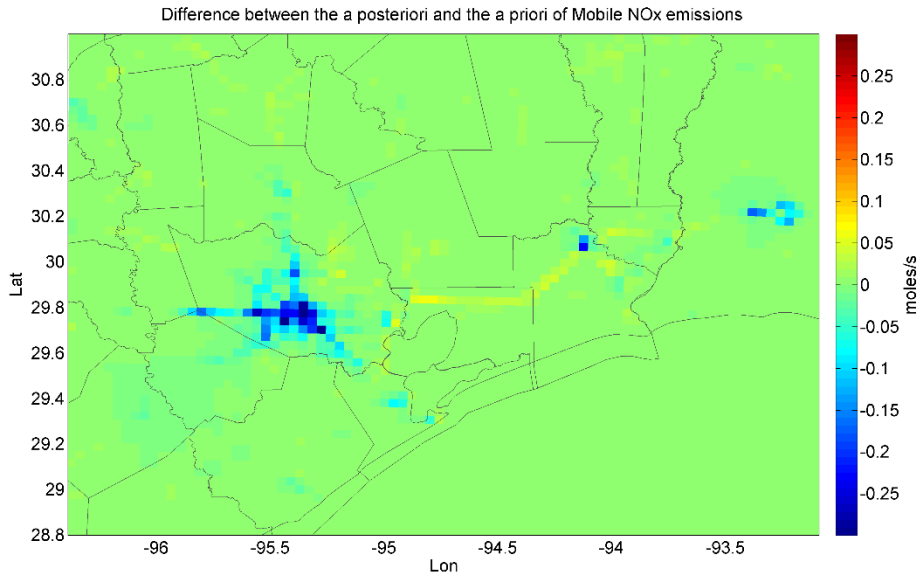


**Figure 8(c).** The difference between NEI-2011n and NEI-2011 for mobile source category.

## *Point Source Category*

Figure 9 represents the point source $NO_x$ emissions. Point sources are more concentrated near the Houston Ship Channel. Posterior emissions revealed that besides mobile $NO_x$ emissions, point source was also over-predicted in the 2013 simulation. This is likely due to strict regulations which reduced point source emissions.

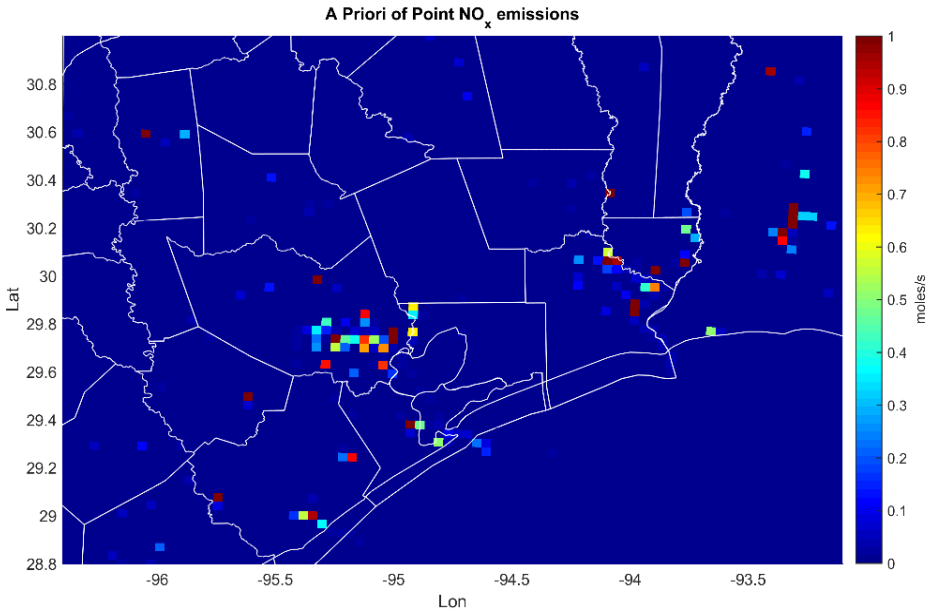


**Figure 9(a).** Point source category $NO_x$ emissions estimates in southeast Texas, NEI-2011 (Old)

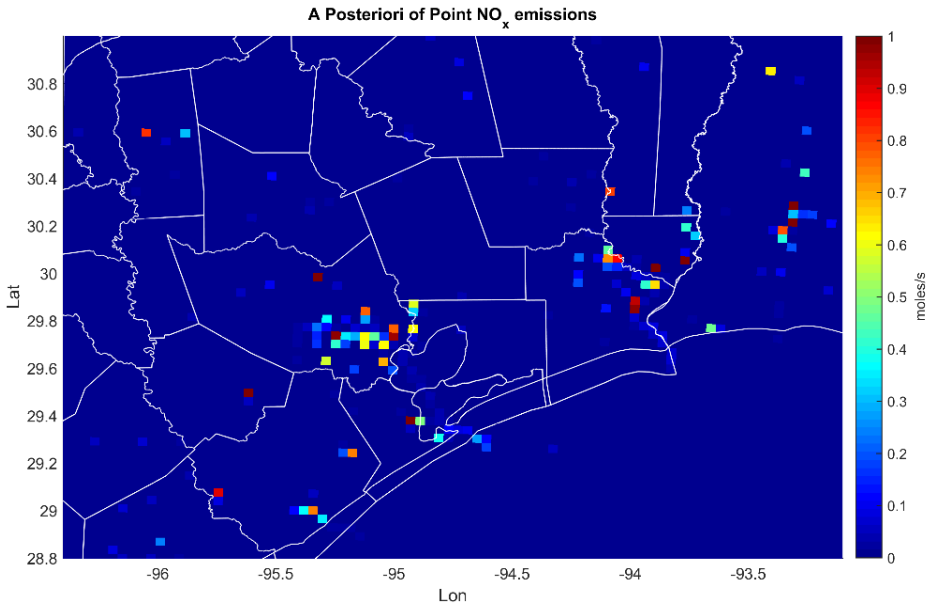


**Figure 9(b).** Point source category $NO_x$ emissions estimates in southeast Texas, NEI-2011n (Updated)

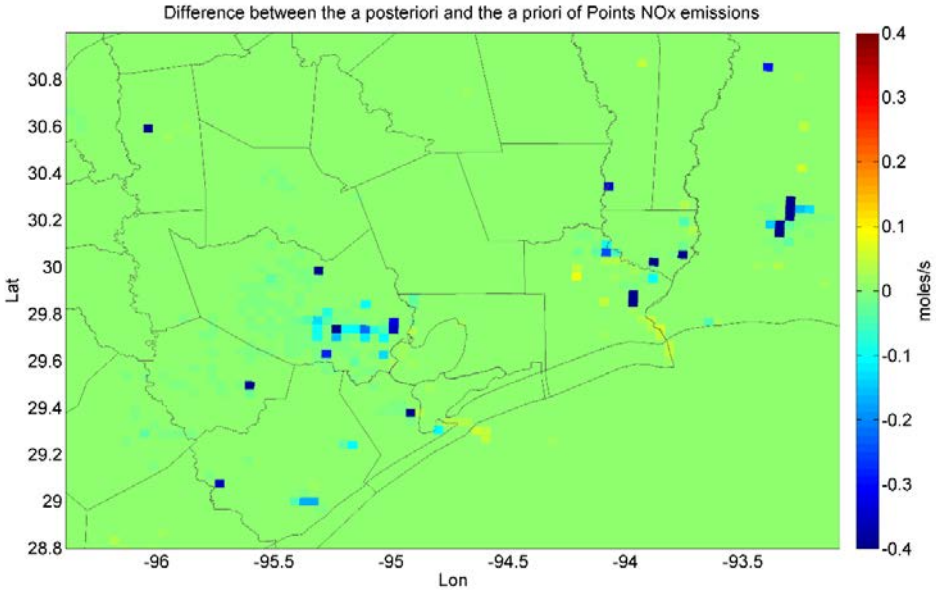


**Figure 9(c).** The difference between NEI-2011n and NEI-2011 for point source category.

In order to validate the accuracy of the a posteriori point sources, we used TCEQ-2013 point sources which more likely reflect the real emission rates in Sep-2013. Figure 10 (a) shows $NO_x$ point source emissions in units of ton per year which has been made based on the a priori of NEI-2011. Figure 10 (b) evidently shows that the a posteriori considerably reduced after performing inverse modeling. Interestingly, Figure 10 (c), which is based on TCEQ 2013 $NO_x$

emission, matches better with Figure 10 (b) results, providing convincing evidence that the method used is efficient to update emission inventory based on OMI satellite data.

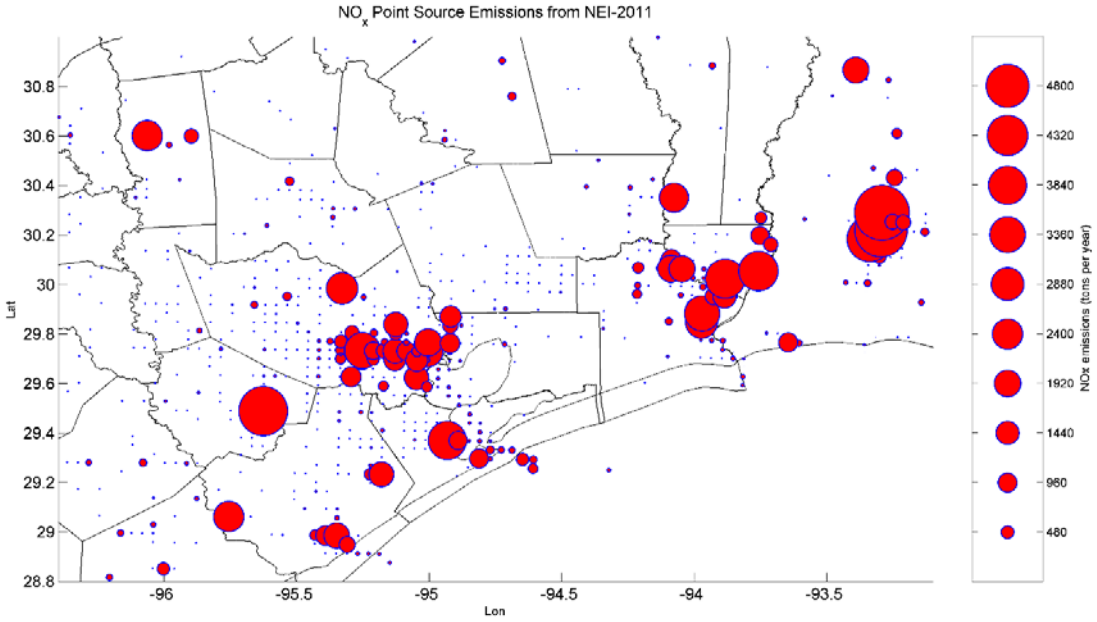


**Figure 10(a):** Point source category $NO_x$ emissions estimates in southeast Texas, NEI-2011. The unit is tons per year.

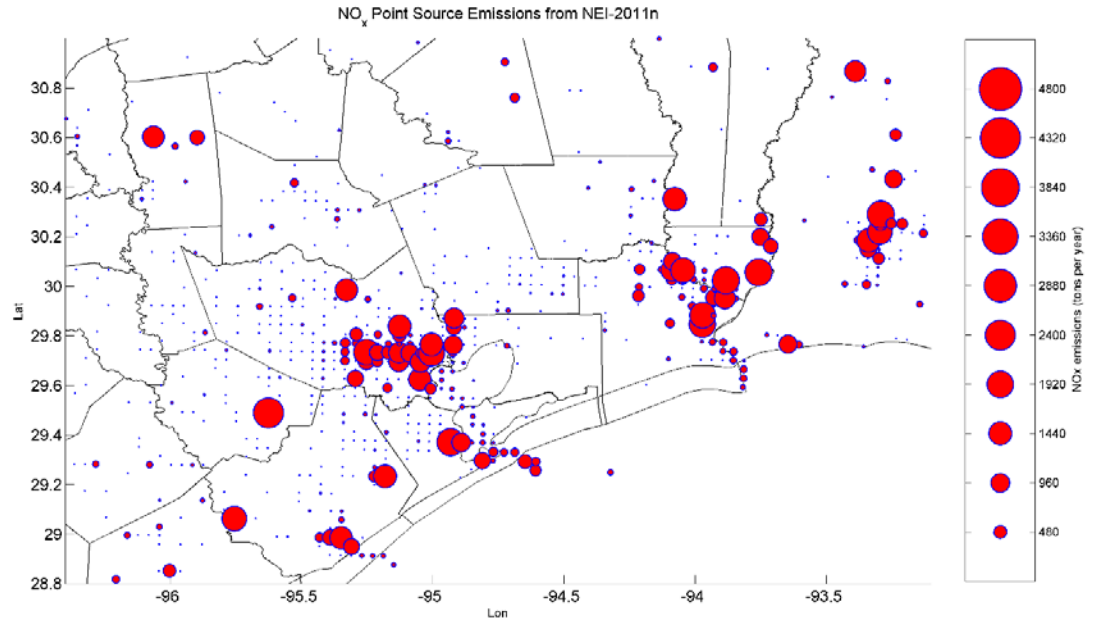


**Figure 10(b):** Point source category $NO_x$ emissions estimates in southeast Texas, NEI-2011n. The unit is tons per year.

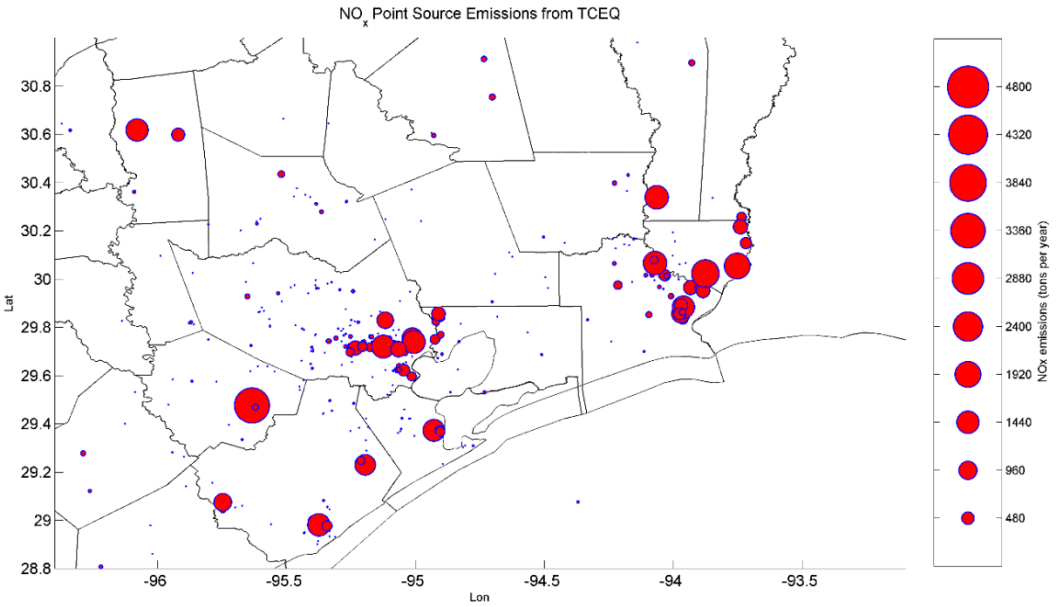


**Figure 10(c):** Point source category $NO_x$ emissions estimates in southeast Texas, from the Texas Commission on Environmental Quality (TCEQ) 2013 Point Source Emissions Inventory (PSEI). The unit is tons per year.

## *Total $NO_x$ Emissions*

Total $NO_x$ emissions before and after inverse modeling is shown in Figure 11. The $NO_x$ emissions decrease in urban regions and background biogenic and area emission increase in rural areas. This could be attributed to large uncertainty of biogenic emissions estimation, or changes in land use/land cover, all of which should be studied in future work.

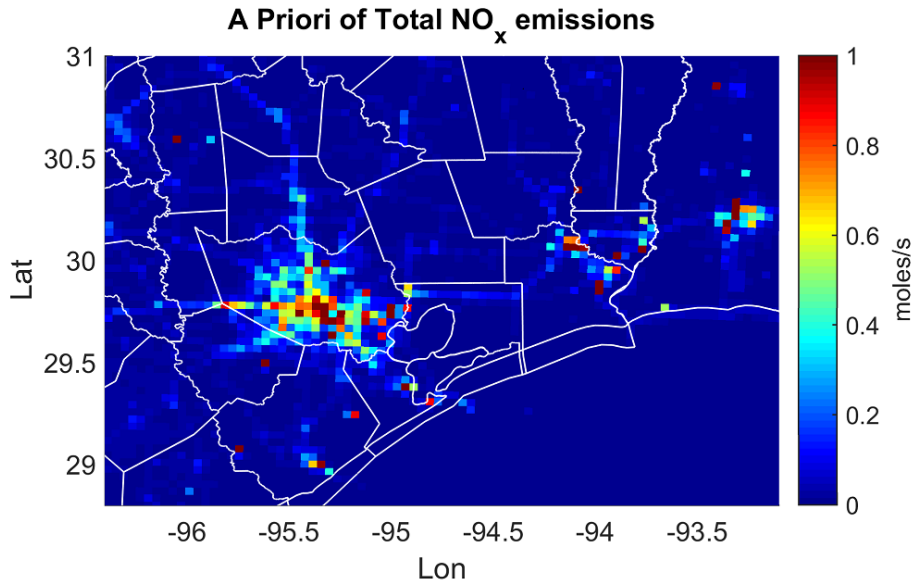


**Figure 11(a).** Total (all source categories) $NO_x$ emissions estimates in southeast Texas, NEI-2011 (Old)

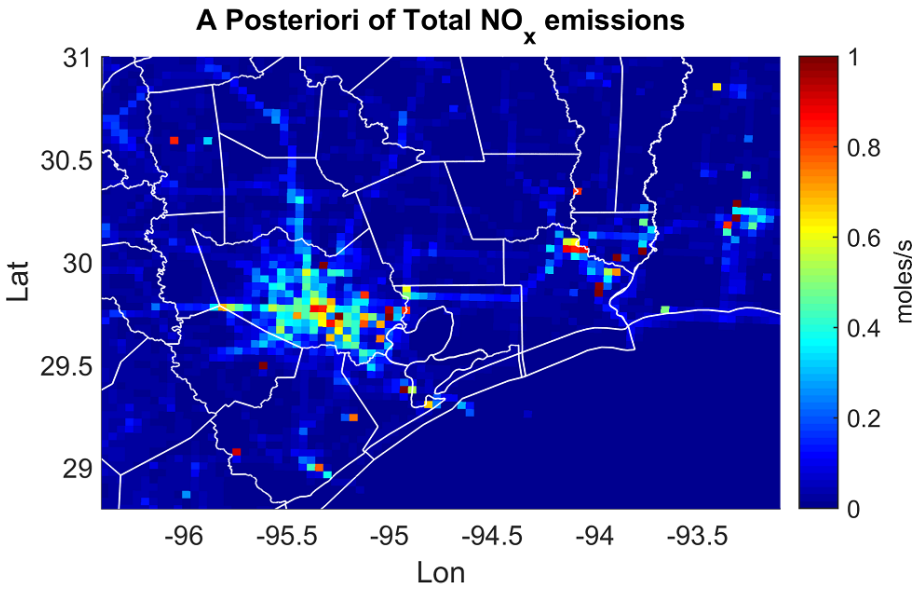


**Figure 11(b).** Total (all source categories) $NO_x$ emissions estimates in southeast Texas, NEI-2011n (Updated)

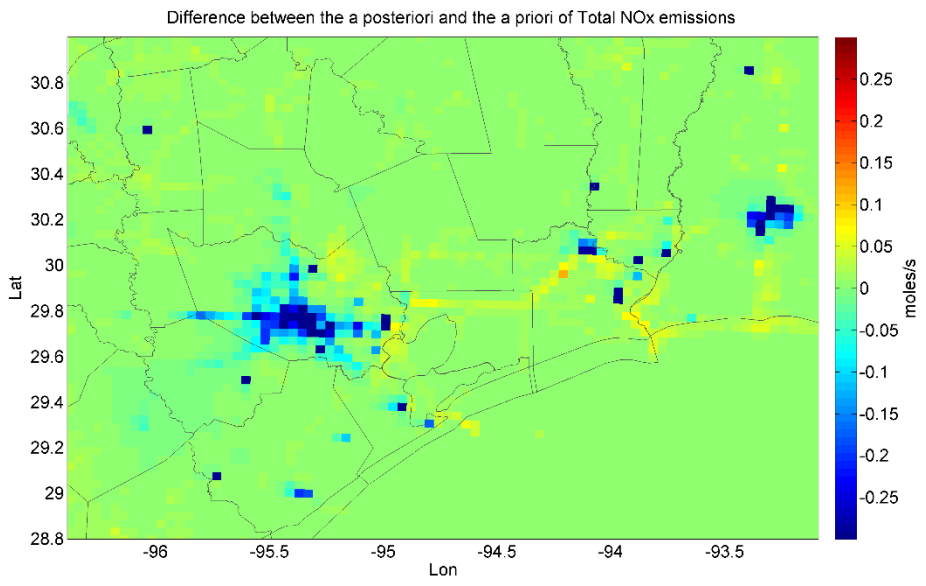


**Figure 11(c).** The difference between NEI-2011n and NEI-2011 for all source categories.

## 4.2 Simulated $NO_2$

To evaluate the updated $NO_x$ emissions, we performed CMAQ simulations to compare model $NO_2$ against OMI satellite, aircraft and surface CAMS observations.

### *Comparison to OMI Tropospheric $NO_2$*

As indicated previously, we used daily granules of NASA OMI $NO_2$ to compare model performance and to update emissions. Due to uncertainty of the product in cloudy conditions and the retrieval method, aforementioned tests have been carried out to provide reliable information (i.e., high signal/noise ratio). Since OMI acquires its imagery in different overpasses and its spatial resolution becomes coarser from nadir to the extremities of scan lines, we re-grid the granules based on the method of Kulman et al. (2014) to provide a map of $NO_2$ in the same grid as our model domain (4-km resolution). While a Level 3 product is provided by NASA, its spatial resolution is too coarse ($0.25^o \times 0.25^o$) for inverse modeling at a regional scale.

CMAQ provides $NO_2$ concentrations in mixing ratio units for each layer. One must integrate all concentrations from surface to the tropopause height in order to retrieve simulated tropospheric $NO_2$. For this purpose we made use of the following equation:

$$Tropospheric\ NO_2 = \sum_{k=1}^{n} 2.687 \times 10^{16} \times (-7.88 \times 10^5) \times (P_k - P_{k-1}) \times NO_2^k \qquad (13)$$

Where $P$ is pressure in hPa (hectoPascals) and $NO_2^k$ is mixing ratio (ppt) in the $k^{th}$ layer.

Resultant tropospheric $NO_2$ concentrations from CMAQ are depicted in Figure 12(a). The plot shows the average of all CMAQ $NO_2$ values by using original NEI-2011 in the region in September 2013 at ~1:30 p.m. CST (one per day at roughly the same time that OMI captures its imagery). It clearly shows high $NO_2$ levels in the center of the city which is caused predominantly by mobile and point sources.

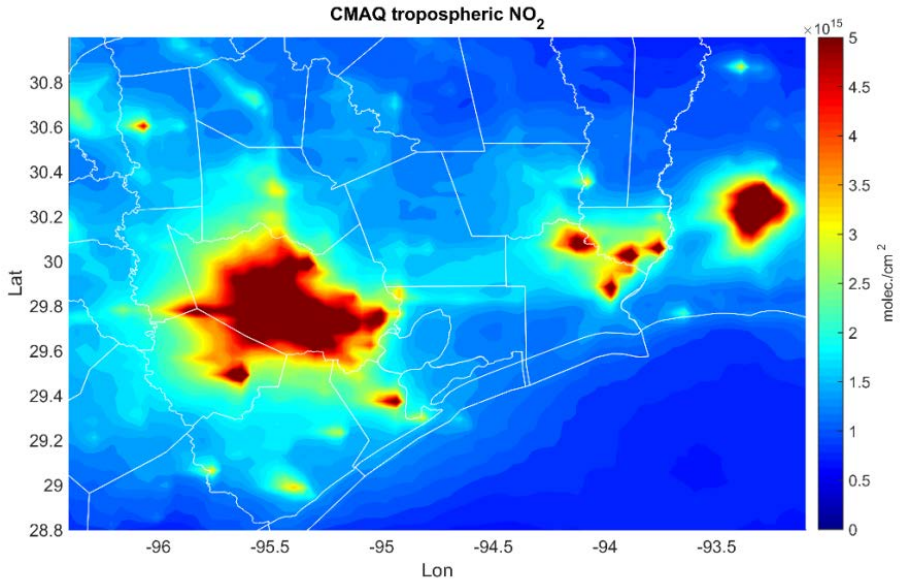


**Figure 12(a)**. Simulated tropospheric $NO_2$ from CMAQ in September of 2013 at 1:30 p.m. CST.

Average OMI tropospheric $NO_2$ in September 2013 is shown in Figure 12(b). Note that the influences of initial profile used for $NO_2$ retrieval have been removed in this result. The result in Figure 12(b) demonstrates higher and lower $NO_2$ values in non-urban and urban regions respectively compared to CMAQ output. This might be expected based on extensive evaluation of photochemistry in this region (e.g., Choi 2014 and references therein). However this large difference between observations and simulation is not only attributed to emission uncertainty but also to exclusion of influences of the a priori profile used in OMI $NO_2$ retrieval. After minimizing the mentioned influences, the model output and OMI observation become closer. We removed this a priori profile based on the method of Duncan et al. (2014) to obtain adjusted OMI tropospheric $NO_2$ depicted in Figure 12(c).

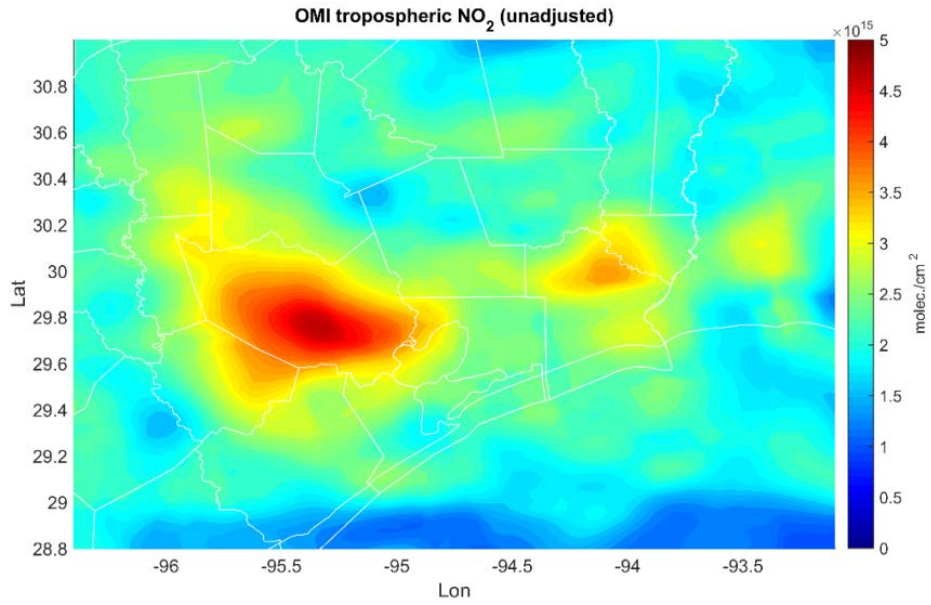


**Figure 12(b)**. Average OMI tropospheric $NO_2$ in September of 2013. Noisy pixels have been screened out but the effect of the a priori profile remains.

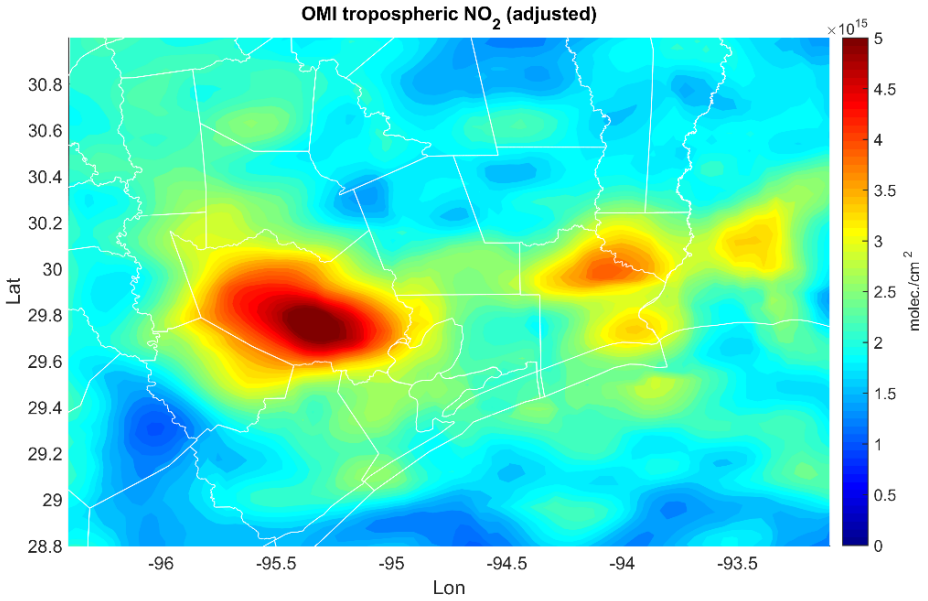


**Figure 12(c)**. Average adjusted OMI tropospheric $NO_2$ in September of 2013. Noisy pixels have been screened out and the effect of the a priori guess is mitigated.

It is evident that adjusted OMI output differs from the unadjusted one. After adjusting OMI tropospheric $NO_2$, background $NO_2$ in rural regions decreased. This result is especially noteworthy when contrasted with CMAQ model output and our assumption of small contributions of biogenic soil emissions in these regions. Figure 12(c) presents a very detailed OMI map for Houston. We believe this figure communicates a crucial message to the air quality

community: removing the influence of the a priori profile undeniably has considerable impacts (both in magnitude and distribution) on OMI tropospheric $NO_2$.

In order to validate simulated tropospheric $NO_2$ and OMI, we calculated a 2-D correlation analysis, bias, RMSE and the ratio of model to OMI. We found good agreement in correlation (~0.7) between the observation and model results. On the other hand, bias and RMSE were found to be ~$7\times10^{14}$ molecules $cm^{-2}$ and $9\times10^{14}$ molecules $cm^{-2}$ respectively, indicating over-prediction by the model. The map of the ratio of the model to OMI is depicted in Figure 12(d). It reveals in Houston, CMAQ over-prediction of $NO_2$ levels might be caused by overestimation of emissions by NEI-2011. Under-prediction in non-urban regions also can be seen.

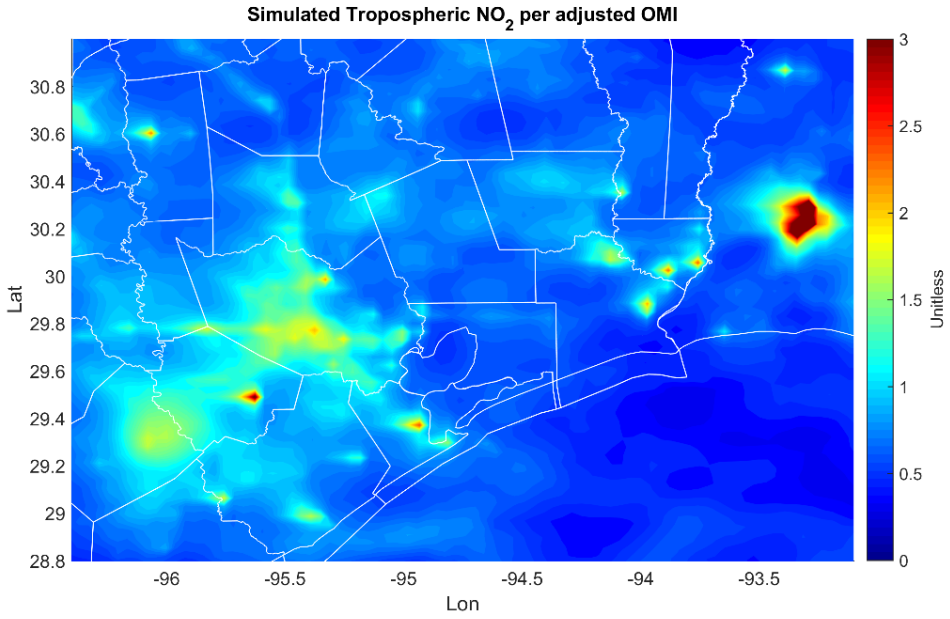


**Figure 12(d)**. Ratio of simulated tropospheric $NO_2$ to OMI output for September 2013 (both data have been co-registered in time and space).

## *Comparison to CAMS*

### *Average Morning Surface $NO_2$*

Typically $NO_2$ concentrations reach a maximum in morning time hence in Figure 13 we plot the spatial pattern for simulated $NO_2$ averaged over 06-12 local time. High $NO_2$ hotspots are seen in the urban regions: Houston, Beaumont, and Lake Charles. Comparison of CAMS sites and model simulation show a correlation coefficient of 0.9, mean absolute bias of 3.3 ppbv and RMSE of 4.1 ppbv. The comparison is made with hourly data averaged over all observation sites. Positive $NO_2$ biases are likely the result of high $NO_2$ emissions in the original inventory. Using updated emissions, the correlation coefficient did not improve but mean absolute bias (2.7 ppbv) and RMSE (3.1 ppbv) decreased. The largest bias reduction is seen at the center of Houston. Bias in rural area was small in NEI-2011 and remained so in NEI-2011n.

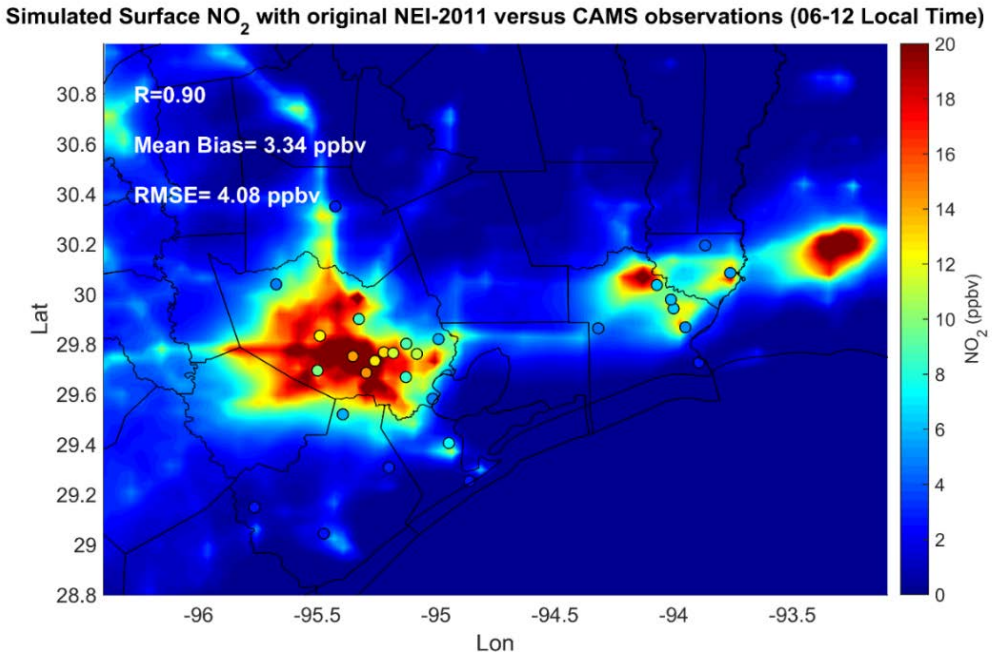


**Figure 13(a)**. Simulated morning surface $NO_2$ and CAMS observations (small circles), NEI-2011.

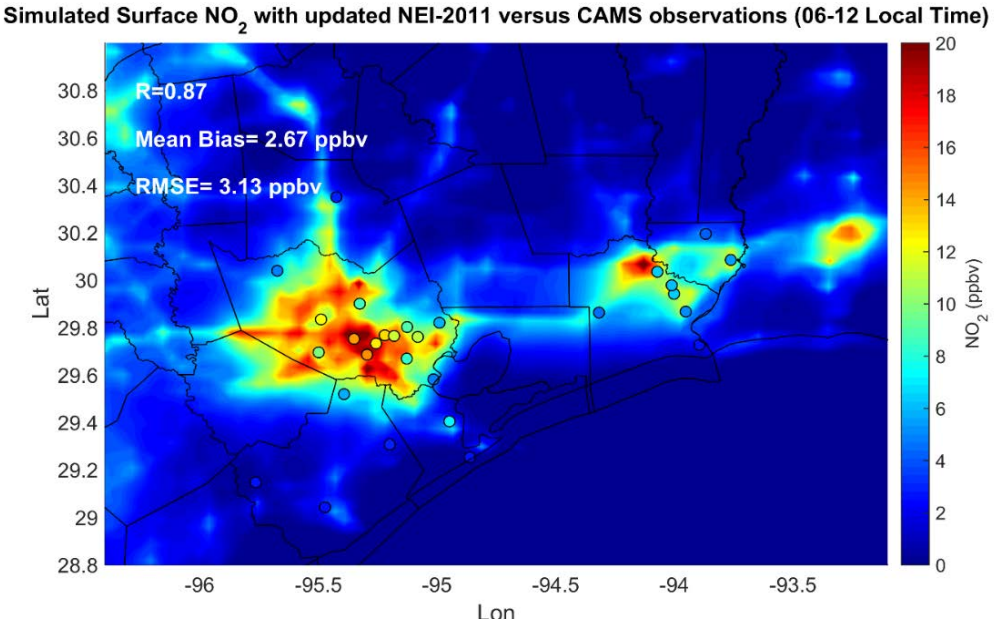


**Figure 13(b)**. Simulated morning surface $NO_2$ and CAMS observations (small circles), NEI-2011n.

*Hourly Surface $NO_2$ statistics*

Hourly surface $NO_2$ statistics are shown in Table 9.

**Table 9.** Statistics of hourly surface $NO_2$

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| NEI-2011 | 19804 | 0.65 | 0.74 | 6.9 | 4.5 | 2.6 | 5.6 | 8.2 | 5.7 | 8.4 |
| NEI-2011n | 19804 | 0.61 | 0.75 | 6.0 | 4.0 | 1.6 | 5.6 | 7.3 | 5.7 | 7.1 |

N – data points; Corr – Correlation; IOA – Index of Agreement; RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MB – Mean Bias; O – Observation; M - Model; O_M – Observed Mean; M_M – Model Mean; SD – Standard Deviation; Units for RMSE/MAE/MB/O_M/M_M/O_SD/M_SD: ppb

It should be noted that the correlation in Figure 13 was computed using average morning $NO_2$ for September while Table 9 was based on individual hourly observations. Therefore the correlation in Table 9 might be lower. Overall, updated emissions improved surface $NO_2$ statistics with IOA increasing by 0.01. A decrease in correlation is offset by better matches in model mean (dropping from 8.2 to 7.3 compared, observed 5.6) and standard deviation (dropping from 8.4 to 7.1, observed 5.7). This is expected as $NO_2$ emissions decreased in the updated inventory.

*Regional Average Surface $NO_2$ Time Series in Metro Houston*

In order to study temporal variations in measured $NO_2$ levels and our model simulation, we compare the time series of observed and simulated $NO_2$ levels in the Houston metro area. $NO_2$ data were hourly regional averages using observations from more than 20 CAMS sites. Figure 14 shows consistent over-prediction of daily $NO_2$ peaks using original NEI-2011 on most days. With adjusted NEI-2011n emissions, the new simulation shows that mean bias and RMSE of regional average surface $NO_2$ decreased by 24% and 15% respectively.

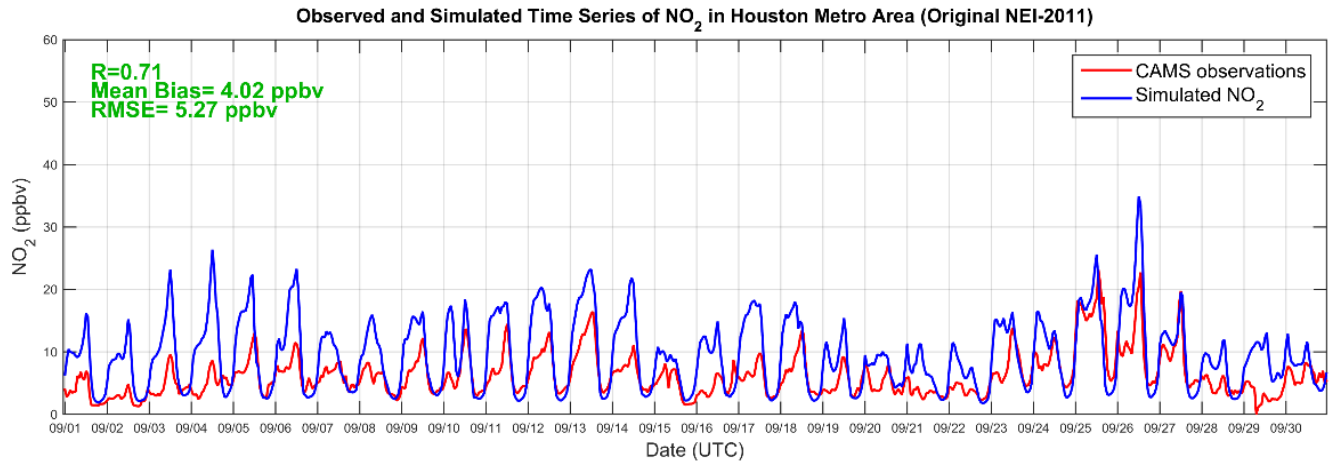


**Figure 14(a).** Time series of simulated and observed surface $NO_2$ levels -- NEI-2011

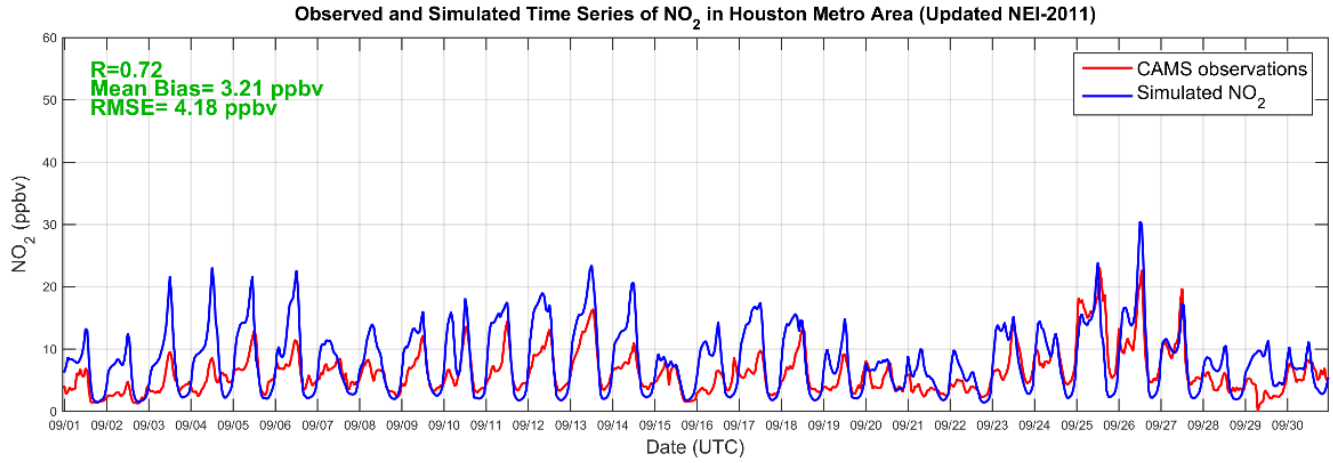


**Figure 14(b)**. Time series of simulated and observed surface $NO_2$ levels -- NEI-2011n

## *Comparison to Aircraft*

### *$NO_2$ Vertical Biases for NEI 2011 and NEI 2011n*

Figure 15 plots the $NO_2$ vertical biases against aircraft measurements for the original and updated emissions. All 10 days with sufficient observations were plotted. The results are mixed – depending on the individual day and the height level.

When the original simulation (NEI-2011) showed a positive bias, the new simulation reduced biases. On the other hand, when the old case had a negative bias, the new case made the bias worse. This is in agreement with the fact that overall $NO_2$ emissions in NEI-2011n are lower than NEI-2011. Above ~1.5 km, $NO_2$ concentrations fall below 1 ppb for both model and observation. There is usually minimal difference between the two cases.

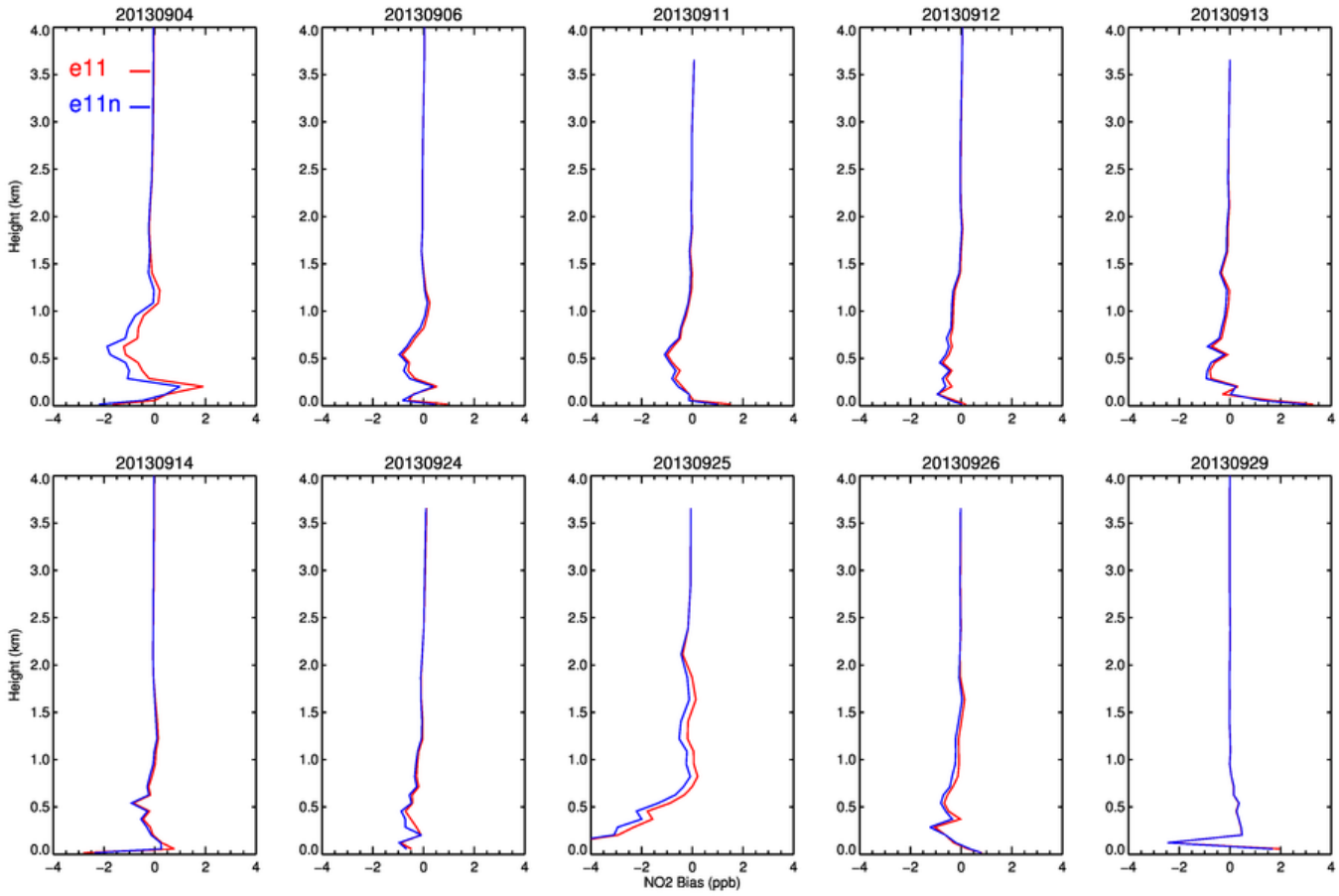


**Figure 15**. $NO_2$ vertical biases for NEI-2011 (red) and NEI-2011n (blue). Measurements are aggregated over model grid cells

## *HCHO Time Series*

Figure 16 shows HCHO time series as well as aircraft measurements for NEI-2011 and the updated NEI-2011n emission inventory.

The two figures (Figure 16a and 16b) have minor differences. Overall, the model has less oscillations and slightly lower values. There is a high observed peak on 09/25, corresponding to observed peaks in $O_3$ and $NO_2$. Similar to the situation in $O_3$ and NO, the model did not capture the observed high values. We are still investigating the cause for model misses on 09/25 and current results suggest that a problem in model winds played a role while the emission errors (either from inventory or unreported upset releases, or both) could also be at fault.

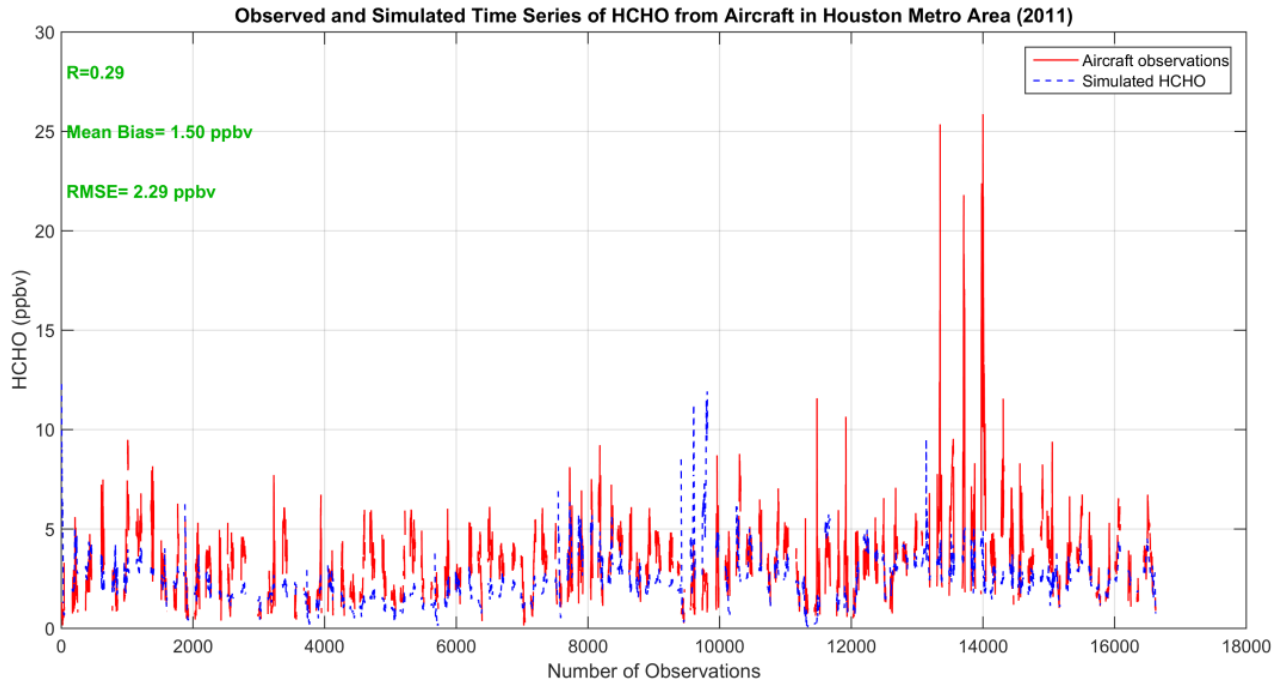


**Figure 16(a)**. Model and observed (aircraft) HCHO using NEI-2011

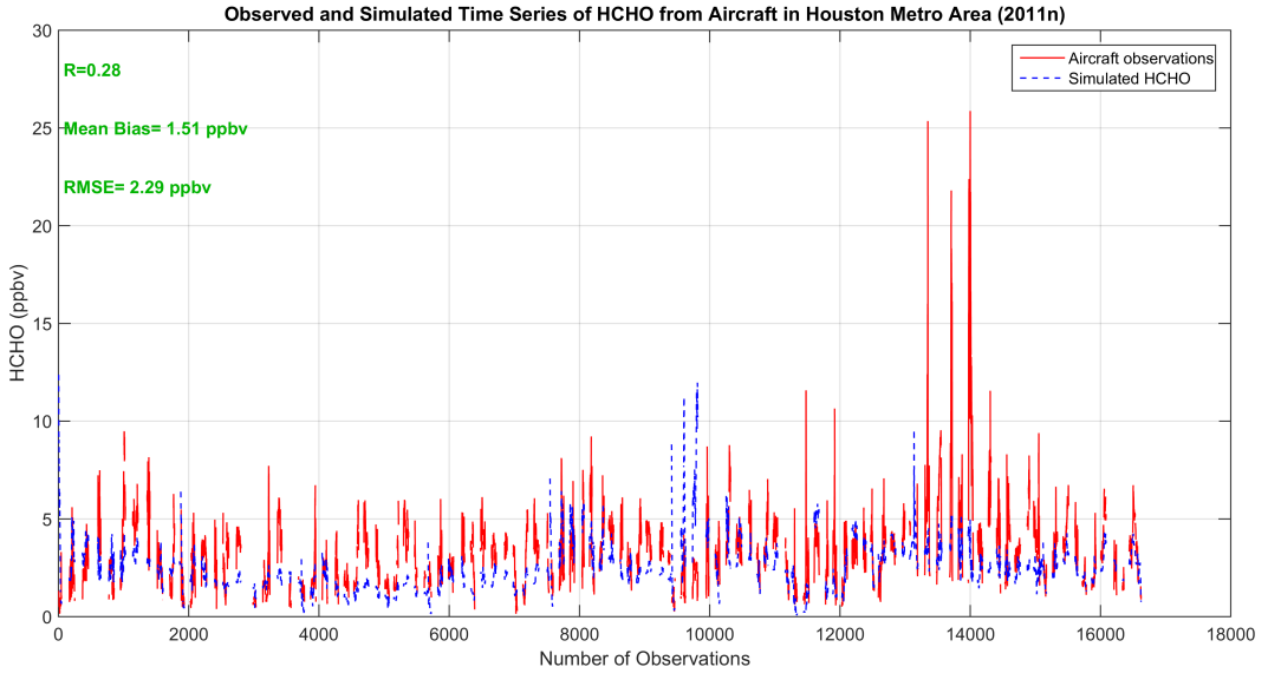


**Figure 16(b)**. Model and observed (aircraft) HCHO using NEI-2011n

*Impact of Updated Emission Inventory on $HCHO/NO_2$ Ratio*

Because of non-linearity of ozone production relative to its precursors, individual species or species ratios can provide evidence on the ozone-$NO_x$-VOC relationship (Sillman, 1995). Among these "indicator species", the $HCHO/NO_2$ ratio has been considered a robust proxy for assessing the impact of VOC and $NO_x$ on ozone production (Martin et al., 2003; Choi et al., 2012). Due to large uncertainties of the emission inventories (both biogenic and anthropogenic), the magnitude of the ratio suggests that the chemical conditions of the region are incorrectly depicted. This inaccuracy not only causes ozone production to deviate considerably from its actual rate, but also hinders conducting a reliable sensitivity analysis of ozone to its precursors (e.g., Choi and Souri, 2015). Hence one of the advantages of inverse modeling is to capture more realistic chemical conditions for a given region. Figure 17 demonstrates the ratio using original NEI-2011, updated NEI-2011n and the difference (i.e., NEI-2011n in respect to NEI-2011). Due to the fact that both anthropogenic VOC emissions (a source of HCHO) and biogenic emissions were constant, only $NO_x$ emissions changes have impacted the ratio before and after adjustment. Figure 17 shows that in urban regions, more correctly specified chemical conditions produce a more $NO_x$-sensitive regime using the adjusted emission inventory. On the other hand, in remote regions (i.e., rural regions), an opposite trend is seen.

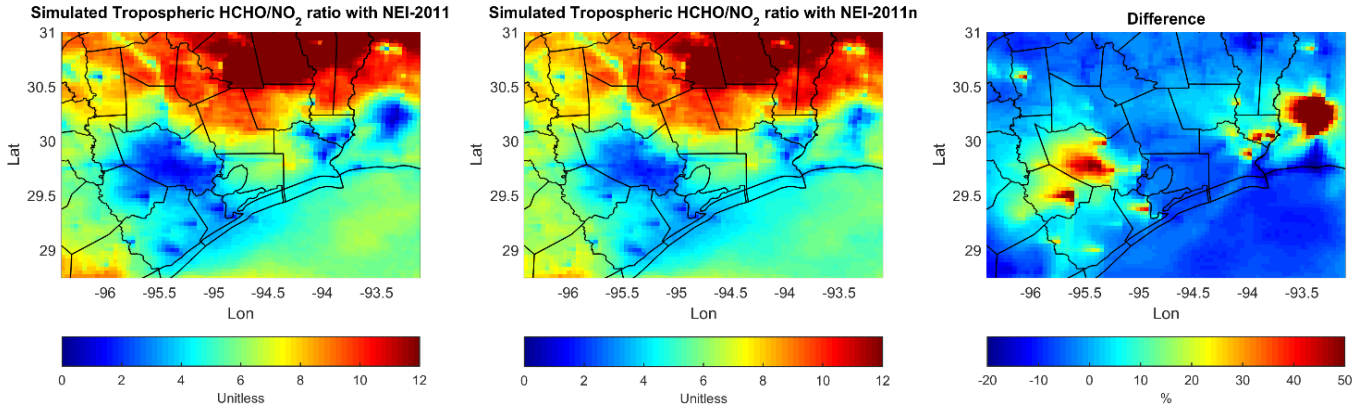


**Figure 17**. Simulated tropospheric $HCHO/NO_2$ with NEI-2011 (left panel), NEI-2011n (middle panel) and the difference (right panel). Positive percentage means becoming more NOx-sensitive.

# 5. Inverse Modeling: Impact on Ozone

Since the two principal pollutants triggering pollution events in Texas are $NO_x$ and ozone, the ultimate goal of this study is to investigate the impact of updated $NO_x$ emissions on ozone. In Texas, the HGB region is assigned a 'non-attainment' status by EPA due to ozone exceeding the National Ambient Air Quality Standards. Here we present ozone results before and after inverse modeling.

## 5.1 Surface Ozone

### *Hourly Ozone Statistics*

Hourly surface ozone statistics are displayed in Table 10. It is based on all available hourly ozone observations. Missing data points are not included. The updated emissions slightly improved surface ozone statistics, with correlation increased by 0.02 and IOA by 0.01. Model mean ozone and bias showed minimal change.

**Table 10.** Statistics of hourly surface ozone

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| NEI-2011 | 33308 | 0.74 | 0.79 | 14.6 | 12.0 | 9.3 | 24.4 | 33.7 | 16.5 | 14.2 |
| NEI-2011n | 33308 | 0.76 | 0.80 | 14.4 | 11.7 | 9.2 | 24.4 | 33.7 | 16.5 | 15.2 |

N – data points; Corr – Correlation; IOA – Index of Agreement; RMSE – Root Mean Square Error; MAE – Mean Absolute Error; MB – Mean Bias; O – Observation; M - Model; O_M – Observed Mean; M_M – Model Mean; SD – Standard Deviation; Units for RMSE/MAE/MB/O_M/M_M/O_SD/M_SD: ppb

### *Regional Ozone Average Time Series*

Figure 18 shows the time series of daily regional average ozone, which is calculated by averaging all the available hourly CAMS observations and the corresponding model values. On most days, the observed average ozone fell below 30 ppb. Since the winds after dawn consistently push precursors from the industrial area to the southwest of the city, the wind pattern does not favor local ozone production. Daytime winds also contained a persistent easterly component which moved the pollutants away from the metro Houston area. In the first 10-day period, lower background ozone coming from the Gulf of Mexico contributed to the low-ozone days. With overcast skies on the 19$^{th}$ and 20$^{th}$, ozone values dipped below 20 ppb. The two highest ozone days, characterized by post-frontal ozone events, were the 25$^{th}$ and 26$^{th}$ of September.

o

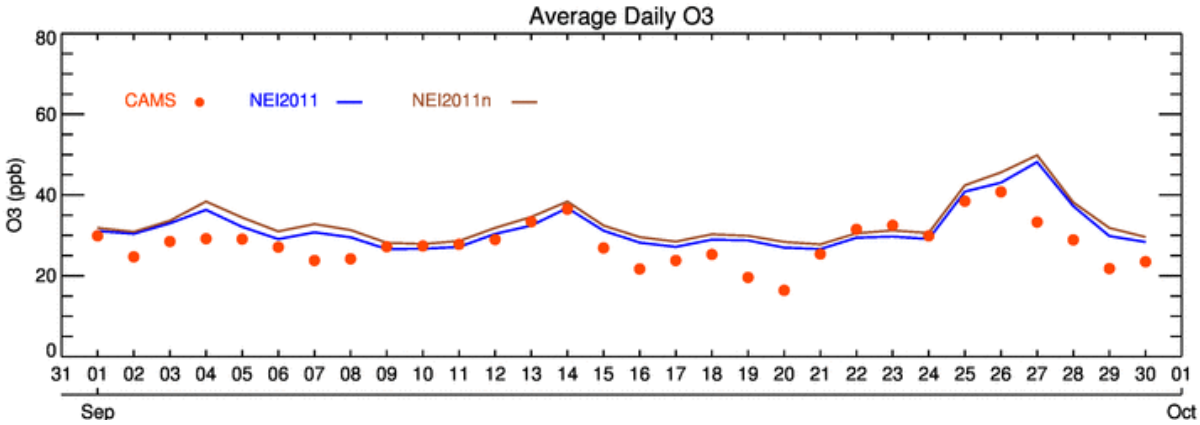


**Figure 18**. The daily regional averaged ozone over all the sites in the 4 km domain for September 1-30.

Figure 18 shows that model ozone generally followed observations reasonably well, although some overestimation by the model is visible. After examining the ozone time series of individual sites (Fig 19 to 25), we found that the coastal sites facing the Gulf usually have significant higher biases than the inland sites. Therefore, the overall positive bias, we believe, is largely the result of the high static CMAQ lateral boundary conditions (e.g., Li et al., 2014). Updated $NO_x$ emission minimally enhanced ozone which may be attributed to a slight decrease in NO titration.

## *Hourly Ozone Time Series at Selected Sites*

Figures 19-25 show hourly time series at a couple of selected sites. The ozone concentrations in the NEI-2011 typically have slightly higher peaks and lower lows. The model did well in capturing daily peaks for inland sites such as C78 (Conroe) and C26 (Northwest Harris County). On the other hand, the model tends to have trouble simulating observed nightly lows.

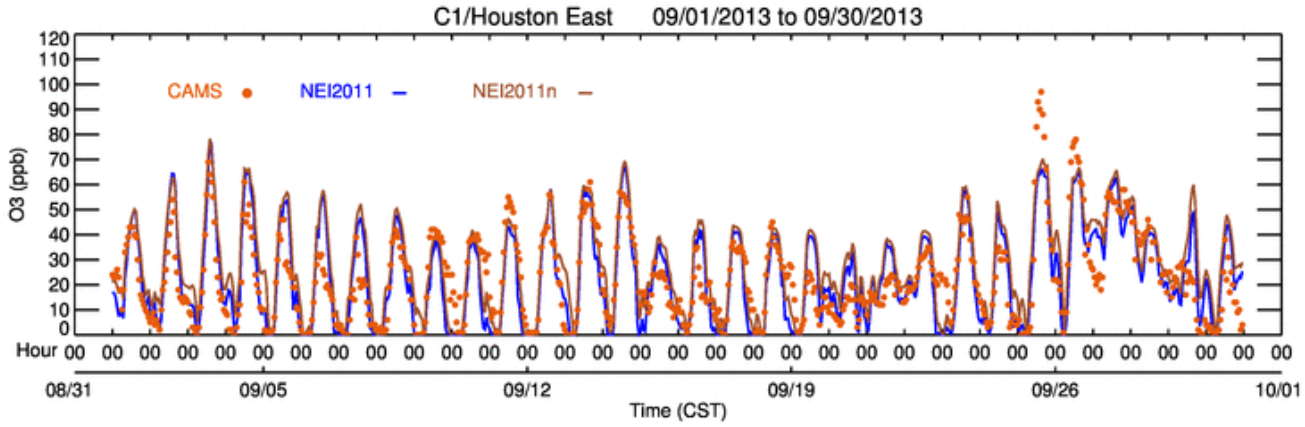


**Figure 19**. C1 – Houston East

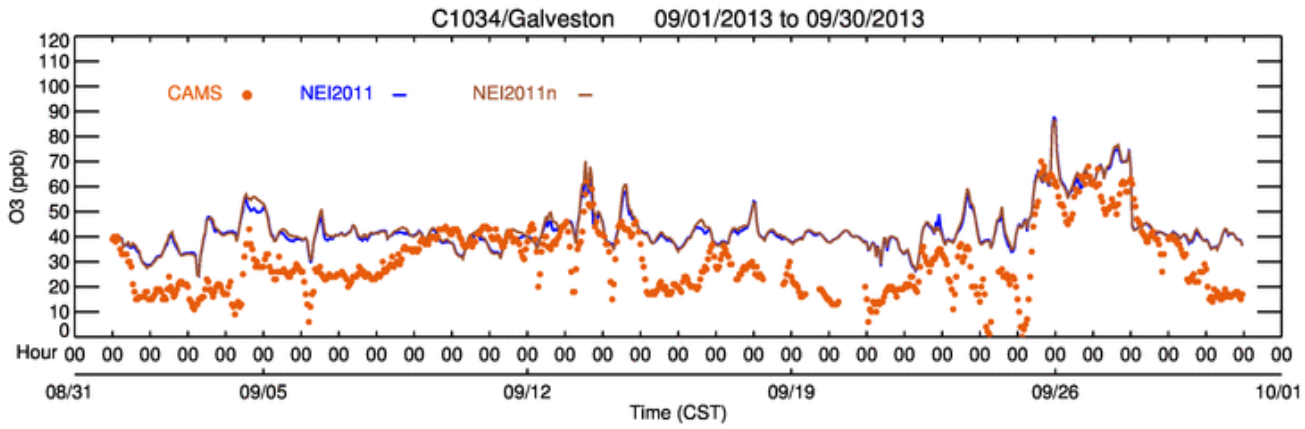


**Figure 20**. C1034 – Galveston

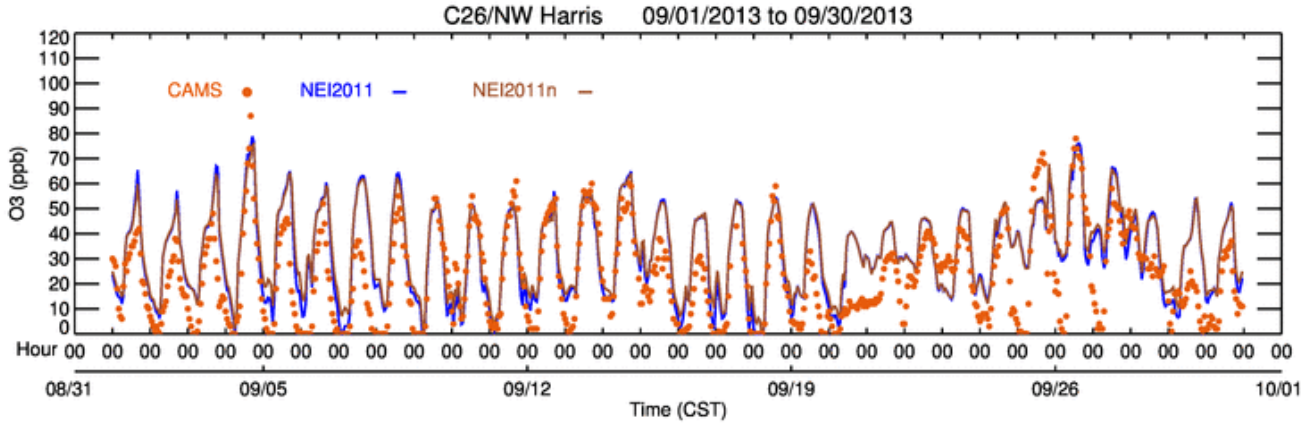


**Figure 21**. C26 – Northwest Harris County

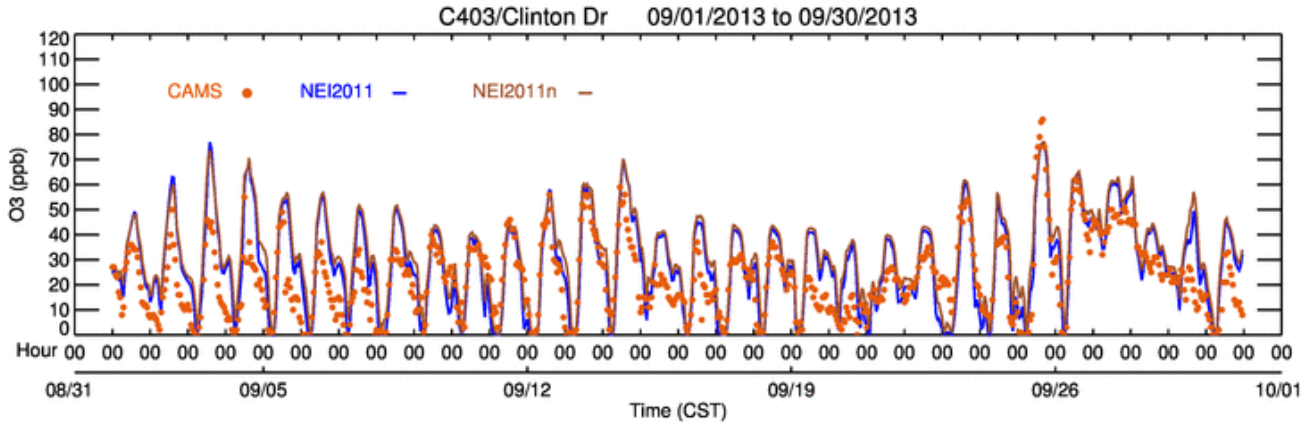


**Figure 22**. C403 – Clinton Drive

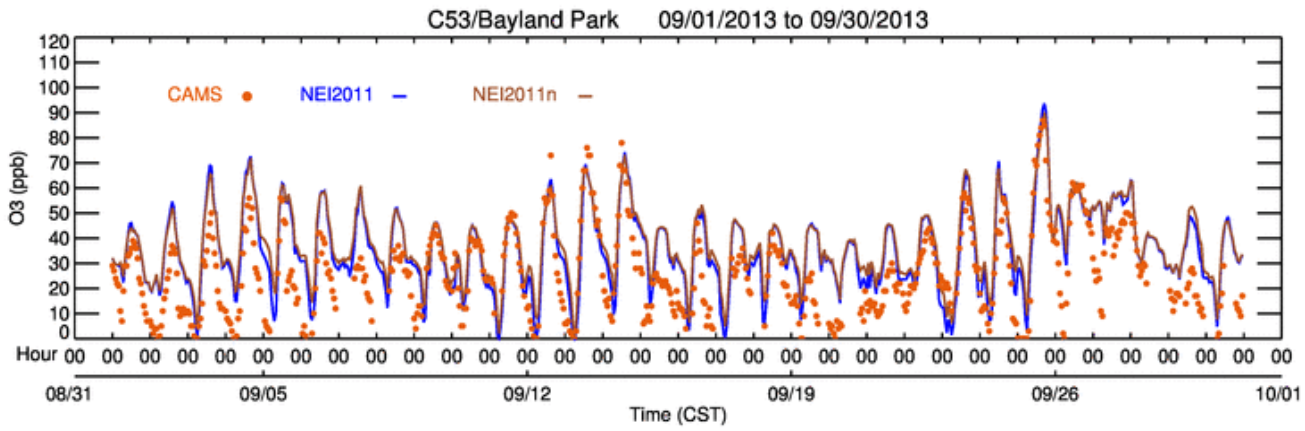


**Figure 23**. C53 – Bayland Park

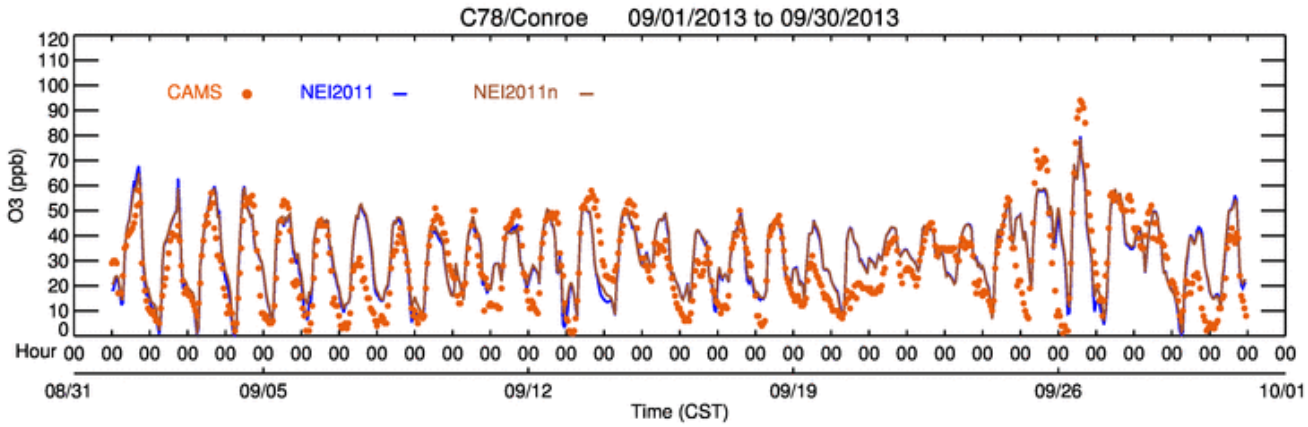


**Figure 24**. C78 – Conroe

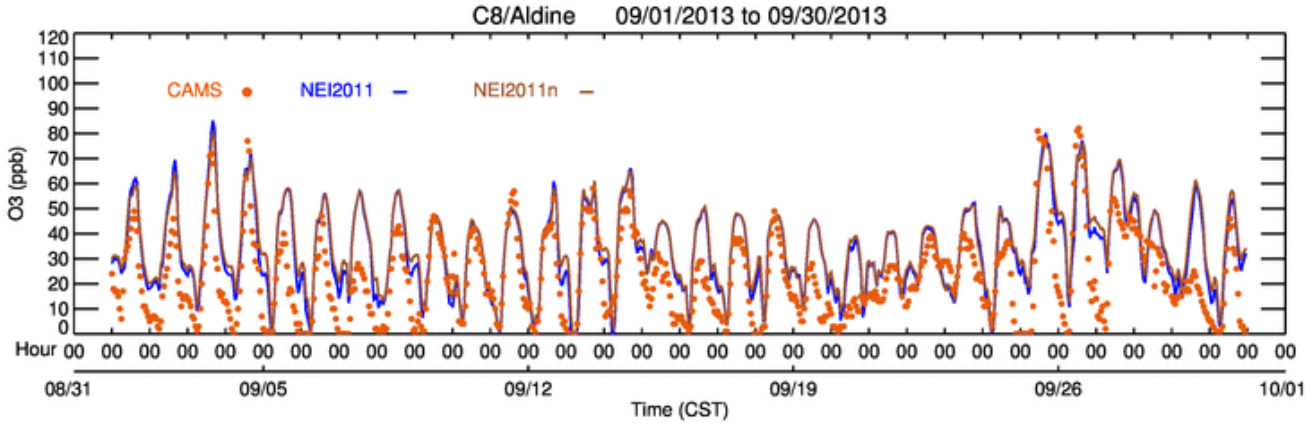


**Figure 25**. C8 – Aldine

## *Ozone on 09/25 – the Only High Ozone Episode*

In the HGB region, high ozone events during the fall season usually occur after frontal passage (e.g., Rappenglück et al., 2008; Ngan and Byun, 2011; Ngan et al., 2012). Two major factors contribute to post-front ozone events: 1) following a cold spell, winds reverse direction and subsequent light winds and sunny skies prevail, creating ideal conditions for ozone production and accumulation 2) reversal in wind direction transports back the pollutants that were blown into Galveston Bay previously, i.e., recirculation (e.g. Banta et al., 2005).

During September 2013, 09/25 and 09/26 had the highest ozone with hourly ozone reaching 151 ppb on 09/25. However, the two days exhibit different patterns. The maximum ozone on 09/25 was caused primarily by favorable weather conditions: sunny, overall light winds and shifting winds in the industrial area. The light morning land breeze carried pollutants from the Ship Channel area to Galveston Bay. As the day warmed up, a bay breeze starts to develop, carrying pollutants back over land. This localized circulation is described by Banta et al. (2005). Ngan et al. (2012) reported the same phenomenon in their Texas Air Quality Study-II 2006 study.

Observed ozone (shown as circles) started at a low value low in the Houston metropolitan area at 0600 CST (Figure 26). However, the model shows a large area of elevated ozone located in the Gulf of Mexico. Model ozone predictions from the updated $NO_x$ emission inventory differed very little from that in the original one.

By 0900 CST (Figure 27), the observed ozone level increased to 30-70 ppb at most CAMS sites higher than normal. Predicted ozone in the NEI-2011n case started to show a difference (generally higher) from the NEI2011 case.

A bay breeze started to develop at 1000 CST, as shown in Figure 28, and a wind convergence zone around La Porte was formed. Rapid ozone production was observed around the La Porte area from 1000 CST to 1300 CST. In fact, the largest 1-hour increase, 46 ppb to 110 ppb was observed from 1000 to 1100 CST at C556 (La Porte Sylvan Beach). However, the bay breeze and wind convergence are missing in the model leading to much lower model ozone production in the area. By 1200 CST (Figure 29), ozone increased to 70-150 ppb at most CAMS sites with C556 having the highest value.

Overall, the predicted ozone concentrations in the two cases (NEI2011 and NEI2011n) are quite similar on 09/25, especially in metro Houston. We can see small increases in ozone concentrations in the rural area, which is consistent with changes in $NO_x$ emissions there. This indicates that $NO_x$ is likely to be saturated even after a decrease in NEI-2011n in the cities. Therefore ozone is not very sensitive to changes in $NO_x$ in this case.

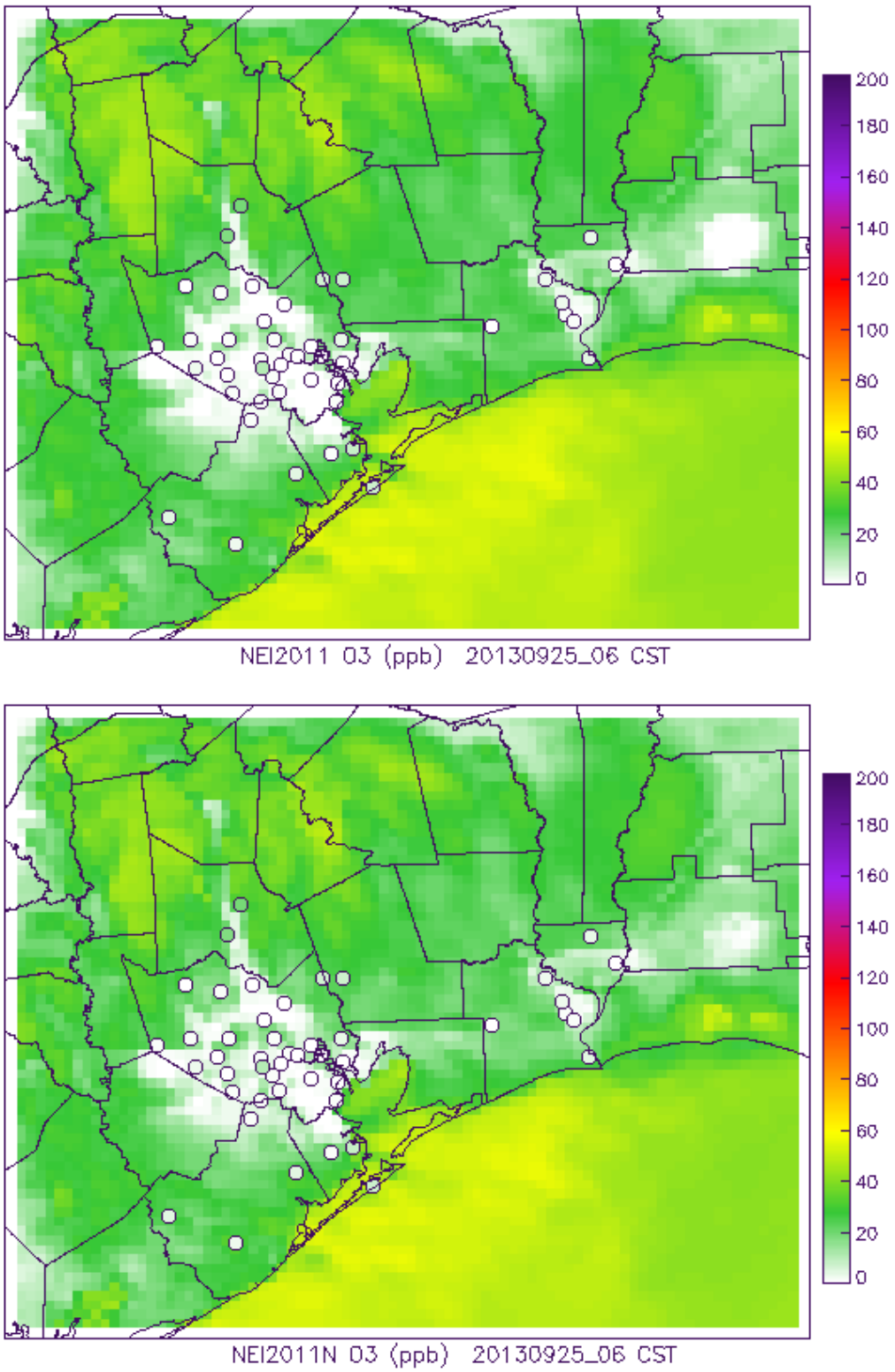


**Figure 26.** Spatial predicted ozone concentrations at 20130925_06 CST – NEI-2011 (top) and NEI-2011n (bottom). Observed ozone at CAMS sites are shown in circles.

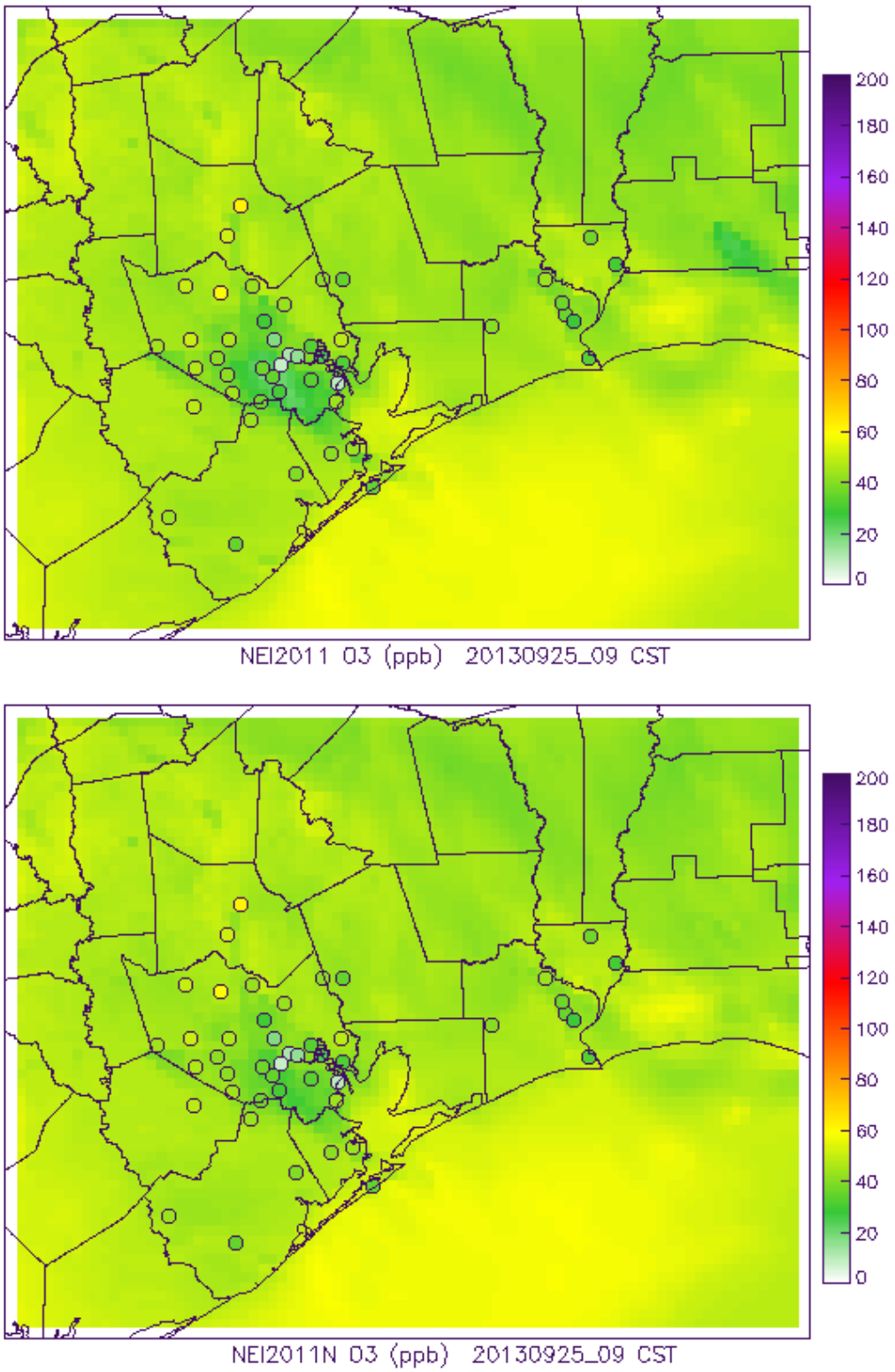


**Figure 27.** Spatial predicted ozone concentrations at 20130925_09 CST – NEI-2011 (top) and NEI-2011n (bottom)

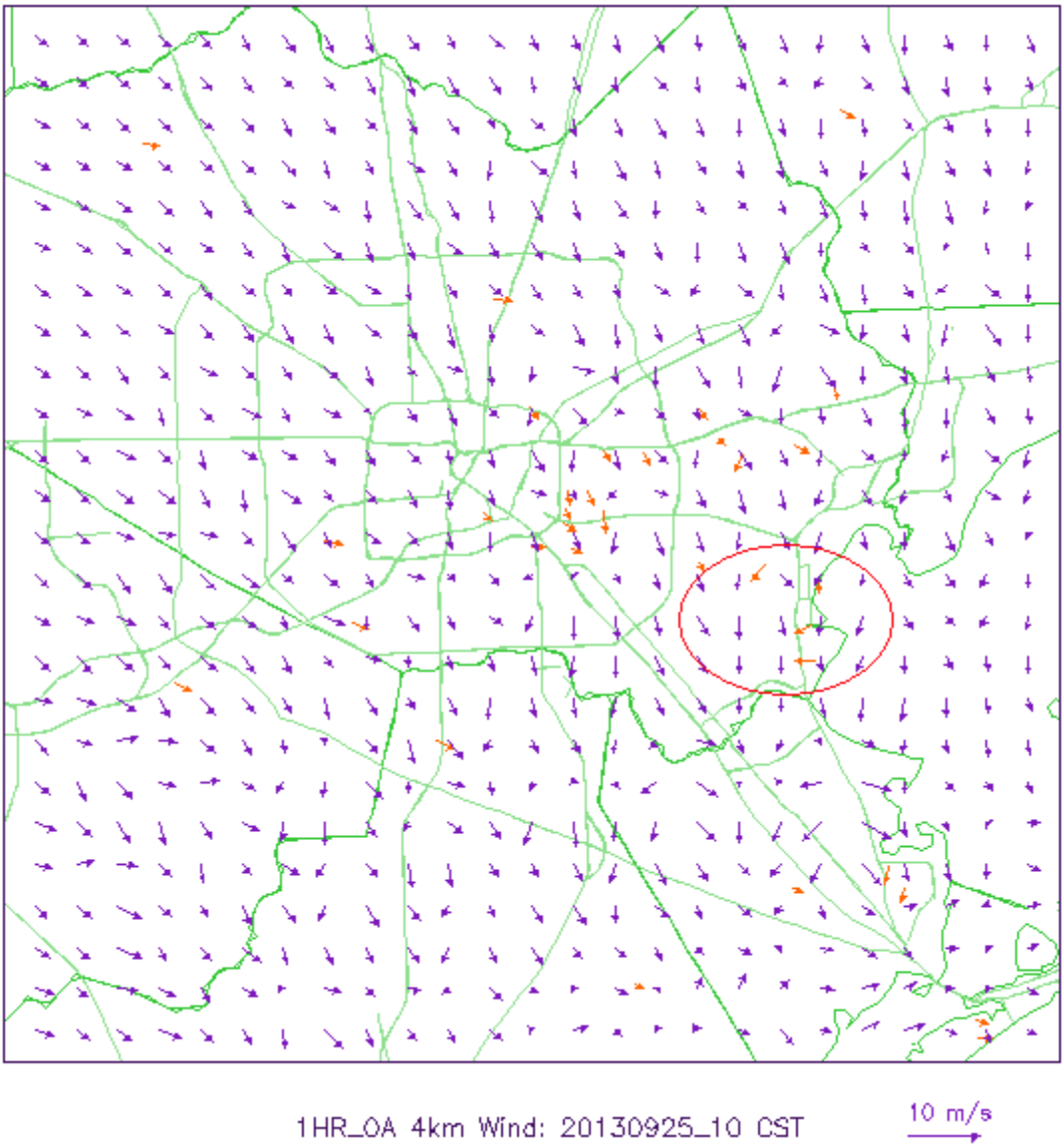


**Figure 28.** Spatial winds at 20130925_10 CST – bay breeze shown in encircled oval

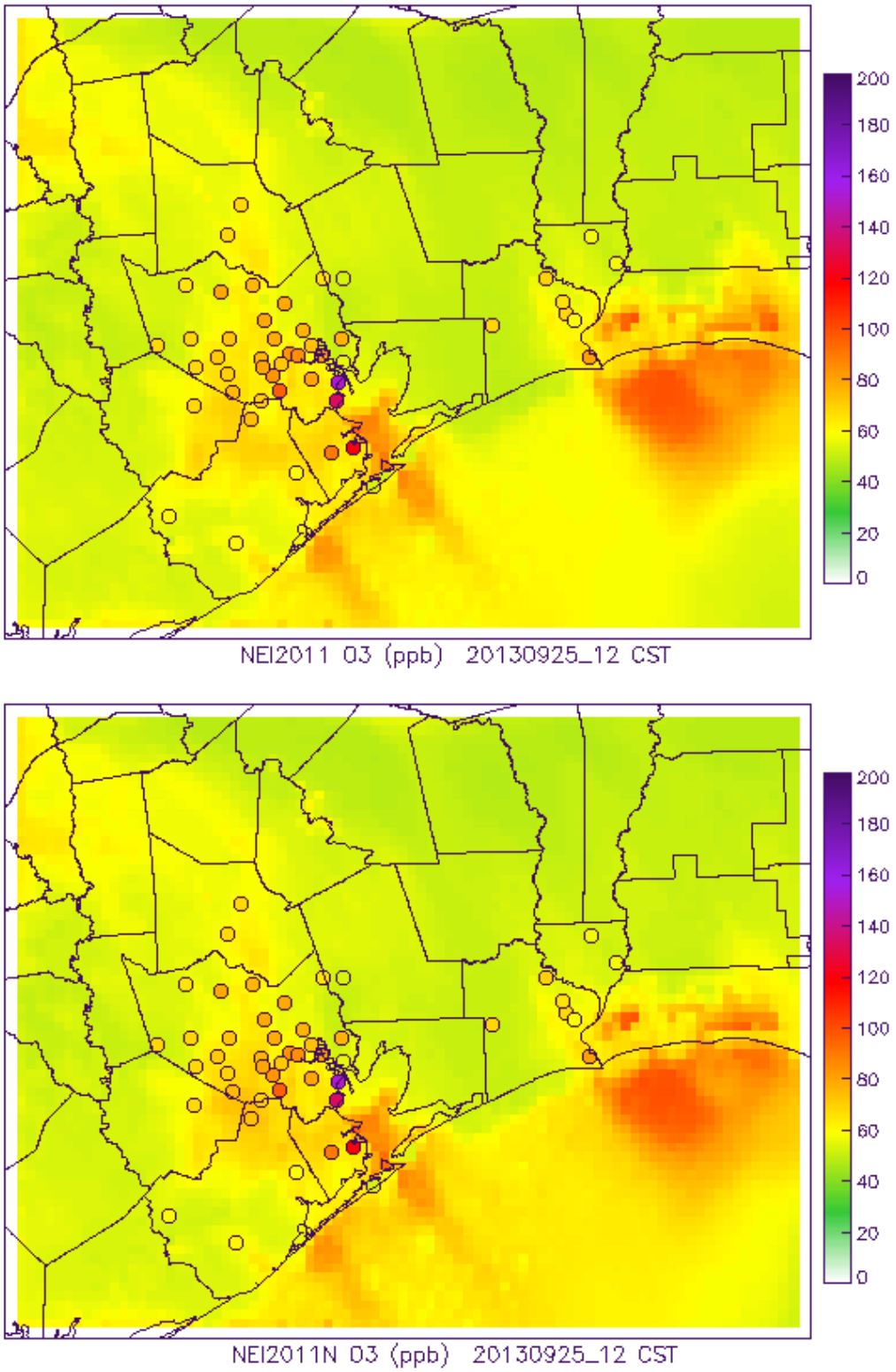


**Figure 29.** Spatial ozone at 20130925_12 CST – NEI-2011 (top) and NEI-2011n (bottom)

## 5.2 Ozone Aloft

Figure 30 shows ozone vertical biases (vis-a-vis aircraft measurements) for the old (NEI2011) and updated (NEI2011n) emissions. Overall the differences are quite small even after the emission update. The only day with more visible changes is 09/26, when NEI-2011n case showed a smaller negative bias at 200-300 meters level. The ozone profiles display significant swings from day to day, reflecting the varying weather and changing ozone lateral boundary conditions. Actual ozone biases depend on the individual day.

Above 1.5 km, the two cases are virtually the same. This is reasonable as the $NO_2$ emissions are only adjusted near the surface. At higher altitude, the $NO_2$ concentration is quite low and does not affect the ozone level.

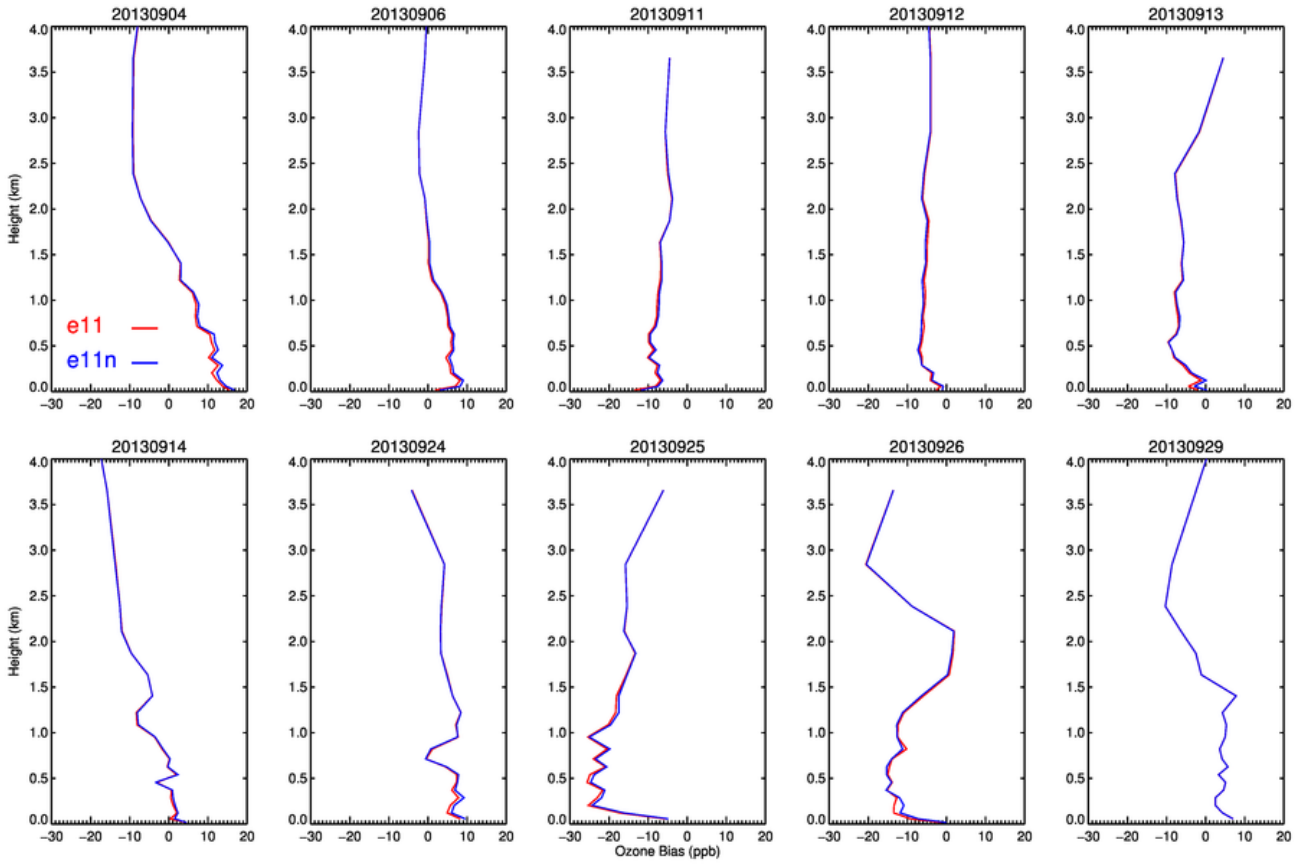


**Figure 30**. Ozone vertical biases for NEI-2011 (‘e11’) and NEI-2011n (‘e11n’). Measurements are aggregated over model grid cells.

# 6. Summary and Conclusions

In this project, we successfully implemented inverse modeling to reduce known $NO_x$ biases in the NEI-2011 emission inventory using satellite remote sensing data as a constraint. We showed that a comprehensive pre-processing part of OMI tropospheric $NO_2$ including geometric and radiometric corrections is essential to update the emission inventory reasonably. The total anthropogenic $NO_x$ emissions in the urban area were brought down while biogenic soil emissions in rural areas were increased in the updated emission inventory. The increase in biogenic emissions in rural areas may be a response to underestimation of $NO_x$ in the BEIS model. Therefore, the change in biogenic emissions is a compensation of existing biases in a model component rather than a reflection of physical reality.

Accurate meteorology is needed for effective air quality modeling. We improved the WRF simulation using the Objective Analysis technique, especially the wind fields which are critical to pollutant transport. Comparing the 1Hr-OA case to the No-OA case, the IOA improved by 6% for U-wind and 11% for V-wind. For temperature, the correlation coefficient increased by about 10% while the Index of Agreement rose by 8%. These statistics indicate that the ozone modelling performance improved moderately. It should be noted that there is a limit within current WRF nudging process to rectify problematic winds. The inability of OA to reproduce the bay breeze and subsequent wind convergence on 09/25 is an example – despite the fact that the day's winds were better after OA was performed.

Our CMAQ simulation using posterior emission showed that the model $NO_2$ column matched better with satellite $NO_2$ data. At the surface level, $NO_2$ in the three urban areas (Houston, Beaumont and Lake Charles) decreased. As a result, overall $NO_2$ biases were reduced although they still remain positive. In rural areas, there are few observation sites and $NO_2$ level was seen to increase slightly. Such increase was also evident by examining the priori and posterior spatial $NO_2$ distribution in our 4-km domain. By comparing model results to the 10-day aircraft data, we found that $NO_2$ aloft barely changed. This may be explained by two facts: all the modifications to the emission inventory occur close to surface level and the $NO_2$ concentration aloft diminishes quickly as altitude rises.

One important contribution of the project was to explore the impact of updated $NO_x$ emissions on ozone concentrations in southeast Texas; since ozone is the primary air quality concern in the region. For ozone both near surface and aloft, we detected only small changes with the updated $NO_x$ emissions. Overall, a small enhancement of ozone is found for most urban sites. Results indicated that the model simulated urban regions as being more $NO_x$-saturated than their actual chemical conditions due to overestimation and underestimation of $NO_x$ and VOC emissions respectively, which led to suppressing ozone production. Therefore, the reduction in $NO_x$ emissions did shift the inaccurate $NO_x$-saturation condition to more $NO_x$-sensitive regime conditions. This demonstrates that besides updating $NO_x$ emissions, VOC anthropogenic emissions also should be updated to reflect more realistic chemical conditions of the region.

# 7. Future Works

Here are a couple of future directions which can originate from this project:

1) Improvement of Jacobian matrix (K) to consider the impact of different emitters on a receptor precisely. This can be done by CMAQ-Adjoint which is likely to be available for public-use in near future.
2) The TROPOMI satellite instrument is about to be launched in 2015 which has significant potential for improving our capability to monitor $NO_x$ emissions from space. This will provide global daily coverage of tropospheric $NO_2$ columns with $7\times7$ $km^2$ nadir resolution. This would be a rich data source to constrain $NO_x$ emissions in a regional scale.
3) The method presented here can be extended for our 12 km Continental USA domain. This might shed more lights on the accuracy of NEI-2011 and how OMI can improve it.

# Acknowledgments

We thank all the mission scientists and principal investigators who prepared and provided the OMI satellite data and DISCOVER-AQ aircraft campaign data used in this study. The authors appreciate Vince Torres, David Westenbarger, Cyril Durrenberger, Jim Smith, and Chris Owen for their comments. This study is funded by TCEQ Air Quality Research Program project (AQRP 14-014). The contents of the report are of the authors' sole responsibility. They do not represent the views of the TCEQ or its staff. Any inquiries related to the report's content and all permission to cite should be directed to the PI.

# Appendix A

## Observation Nudging

Observation nudging is regarded as a low-cost and effective method in improving meteorological model performance, but it requires additional observational data. In this study, we acquired the input observation data and generating files in "little_r" format using similar procedures found in Ngan et al. (2012) and Czader et al. (2013). Observational data came from the MADIS and TCEQ CAMS. MADIS, a National Oceanic and Atmospheric Administration (NOAA) program, collects, integrates, quality-controls, and distributes observations from NOAA and other organizations. The four MADIS datasets used for observation nudging were NOAA Profiler Network (NPN), Cooperative Agency Profilers (CAP), Meteorological Terminal Aviation Routine (METAR) weather report and NOAA Radiosonde (RAOB). The METAR dataset was collected by mostly first-order METAR reporting surface monitoring stations. NPN, RAOB and CAP were the most commonly used upper air datasets.

The "little_r" files from previous step were fed into the WRF OBSGRID module to update the input domain analysis ("met_em") files and generate additional surface analyses ("sffdda") and text nudging files ("OBS_DOMAIN"). Actual observation nudging was performed by the main WRF program by properly setting observation nudging namelist variable. The namelist for OBSGRID and relevant WRF section settings came largely from recommended values of WRF User's Guide and a previous study by Ngan et al. (2012).

Theoretically, observation nudging update at a higher frequency should enhance model performance. A typical frequency of input analysis data is 3-hourly while the frequency for observational data is hourly. The 3-hourly frequency of input analyses may be the reason for the default 3-hour time-interval in WRF's OBSGRID settings for generating the observation nudging files. Since there were few studies on the temporal frequency in existing OA studies and we have not seen a reference of 1-hour interval, we assumed that they used the default 3-hour interval. However, the interval may be set to 1-hour or smaller when corresponding observational data were available. We studied the difference between 1-hour and "default" 3-hour OA; the results indicated 1-hour OA slightly outperformed the 3-hour one. As a result, we adopted 1-hour temporal frequency for observation nudging.

It should be noted that the default time interval for modified gridded analyses, i.e., the "metoa_em" and "sgfdda" files has to match input analysis data in OBSGRID. The namelist variable was called "interval", with a default value of "10800" seconds. The time-interval for output nudging files was set by namelist variable "int4d", with the same default value of "10800" seconds. To output the observation nudging files hourly, "int4d" should be set to "3600" seconds. This means that the OBSGRID output files, "metoa_em" and "OBS_DOMAIN", did not have the same interval in our study.

In WRF, there were a few namelist variables controlling the frequency of grid nudging and observation nudging. The namelist file "namelist.input" is attached at the end of this appendix. The first one was "interval_seconds", which should match the interval of input grid nudging files ("met_em"). The second one was "sgfdda_interval_m", matching the interval of surface grid nudging files ("sgfdda"). In our simulation, both intervals were equal to 3-hours. The third one

was “auxinput11_interval”, controlling the updating interval for observation nudging files (“OBS_DOMAIN”). The last one, “obs_ionf”, determined the nudging frequency relative to internal integration time-step. For example, if the integration time-step for the coarse domain is 30 seconds, setting “obs_ionf” to 1 means performing OA every 30 seconds, while setting "obs_ionf" to 3 means performing OA every 90 seconds. In our simulation, “obs_ionf” is set to 1. There are other namelist variables in “fdda” section controlling the grid nudging behavior. We usually adhered to the default settings. For example, the mass fields (T and Q) were nudged only above the PBL, while wind fields (U and V) were adjusted at all levels including the surface layer.

One departure from the default OA setting in WRF was that the moisture OA was turned off with “obs_nudge_mois” set to 0. This was based on our past experiences since performing moisture OA sometimes trigger excessive artificial thunderstorms which disrupted model flow fields.

For observation nudging, all the data used are in-situ measurements at various sites. No averaging is performed on our side although the hourly data from MADIS and CAMS may be pre-averaged. For comparison to surface data, we used “exact match” – by extracting the model grid cell value for which an observation site is located in. For aircraft data, there is a mismatch on spatial and temporal scales. Therefore, we averaged multiple observation data points within one grid cell and 1-hour timeframe.

The obs-nudging impact on meteorology has been described in the final report. Next we show its impact on ozone.

## Impact on Ozone

The sensitivity of observation nudging on ozone is performed for NEI2008 emission inventory. For NEI-2011 and NEI-2011n, we only run CMAQ for the 1Hr-OA case due to time constraint. Hence we show the comparison of surface ozone and aloft ozone for NEI2008 runs.

Table A.1 shows the statistics of surface ozone for the 1Hr-OA and No-OA cases. Interestingly, both cases had very close correlation of 0.72 and 0.73. However, the mean biases in the OA case were lower by 3.2 ppb, which helped raise the IOA from 0.78 to 0.83. It was interesting that the model standard deviation increased in the OA case and matched better with observation. The improvement in IOA was slightly less in temperature and winds.

**Table A.1.** Statistics of ozone for CMAQ simulations – NEI2008, header information is in Table 5 of Pre-Final Report.

| Case | N | Corr | IOA | RMSE | MAE | MB | O_M | M_M | O_SD | M_SD |
|---|---|---|---|---|---|---|---|---|---|---|
| No-OA | 33308 | 0.72 | 0.78 | 14.9 | 12.3 | 9.3 | 24.4 | 33.7 | 16.5 | 14.1 |
| 1Hr-OA | 33308 | 0.73 | 0.83 | 13.8 | 11.0 | 6.1 | 24.4 | 30.6 | 16.5 | 17.4 |

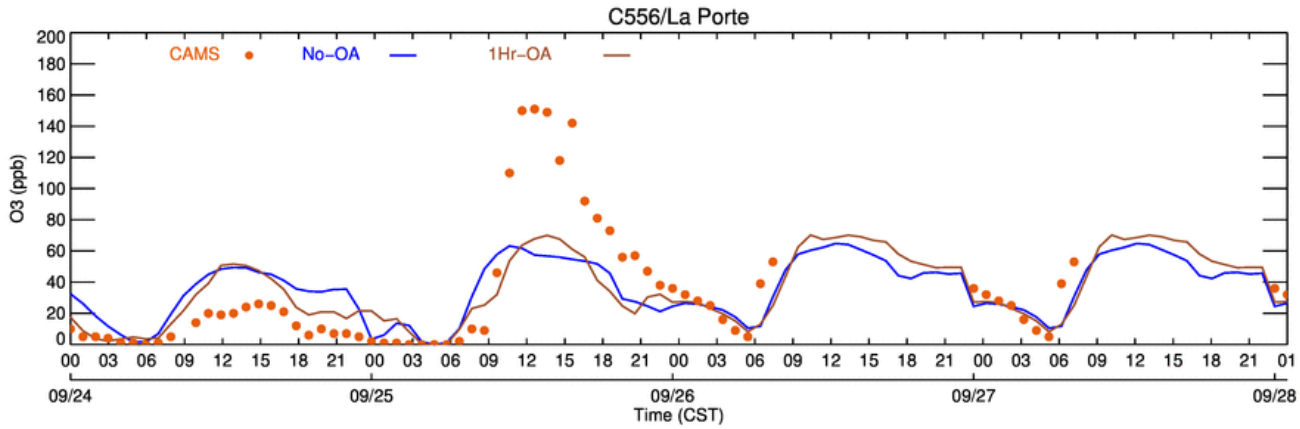


**Figure A.1**. Ozone time series of La Porte (C556) between 09/24_00 to 09/28_00 CST of 2013

CAMS 556 (La Porte) has the highest hourly ozone reading among all the sites in September, which occurred at 13 CST on 09/25. Figure A.1 plots the ozone time series during 09/24-09/27. The model failed to recreate the peak ozone on the 25$^{th}$ – which is likely due to model's inability to reproduce the small-scale wind shifts during the morning hours around the ship channel and the Galveston Bay. Still the 1hr-OA case matched better than the No-OA case in late 24$^{th}$ and around noon of 25$^{th}$. The peak hours of 1hr-OA case matched well to observations: 12 to 15 CST, though much less pronounced. On the other hand, the ozone in No-OA case peaked at 10 CST and trended lower for the rest of the day.

Figure A.2 showed hourly ozone vertical profiles from 08 CST to 16 CST on September 25$^{th}$, with ozone being displayed on x-axis and height on y-axis. One observation dot was averaged over all the grid cells in the same model layer. For example, during 08-09 CST, aircraft flew passing 30 cells located in the model's 5$^{th}$ layer. This layer had a mid-layer height of 287.5 m. The averaged ozone of the 30 cells was 56 ppb. It should be noted that the observed ozone were averaged over multiple measurements in the same model cell, such that they could be properly compared to model values. The 08 and 09 CST profiles showed there was a high ozone layer with average ozone of ~65 ppb, stretching from 450 m to 1200 m height. In comparison, all model runs had lower ozone in this layer. The model biases were about -10 ppb at 08 CST and grew to -20 ppb at 09 CST. The discrepancies between low surface ozone and ozone aloft may be explained by the earlier reversal of aloft winds: winds at surface layer still showed light northwesterly in the early morning while winds aloft already changed to southerly. The observed ozone rose continuously in following hours yet model simulated ozone stagnated around 60 ppb from surface up to 2000 m until 15 CST. At 16 CST, the ozone of OA case in 0-1 km layer rose 20 ppb over previous hours yet the base case ozone increased only a few ppb. Although different in magnitude, ozone aloft had a few similar features to the surface ozone. First, the model missed the observed high ozone in the afternoon by a large margin. For example, the base case underpredicted 0-1 km level ozone by up to 50 ppb. The primary cause for the lower ozone production was likely model's wind fields as both model and observation had clear sky in industrial area and Galveston Bay. Second, nudging clearly helped reducing the ozone biases aloft. In most plots of Figure A.2, the OA case had lower biases than the base case. The largest difference was at 16 CST, when nudging reduced biases from ~45 ppb to ~30 ppb in the 300 – 1000 m layer.

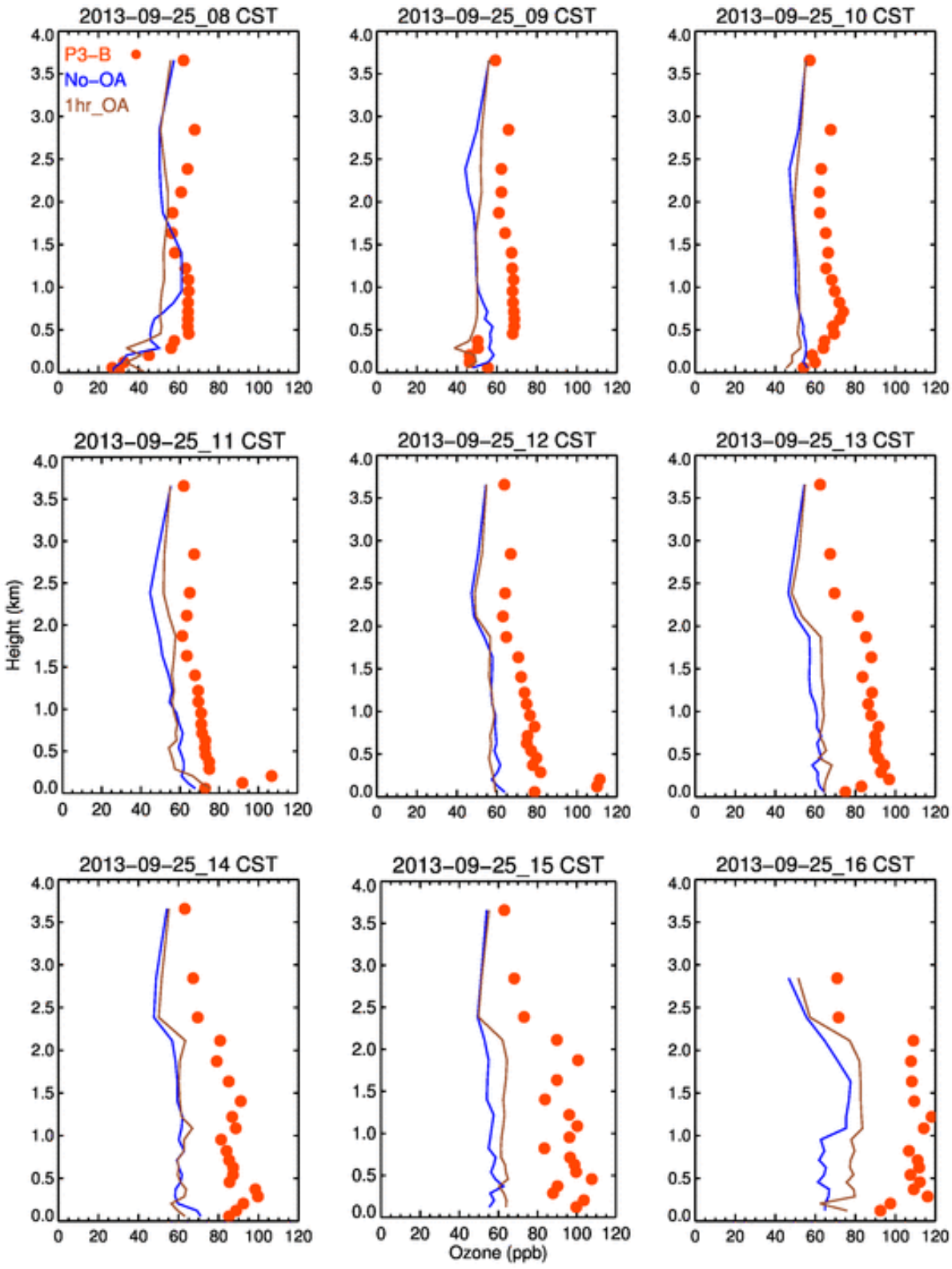


**Figure A.2**. Vertical ozone profiles from 09/25_08 CST to 09/25_16 CST of 2013 for two cases of No-OA and 1Hr-OA compared with corresponding observations.

# Appendix B

## Audits of data quality

### Audit Personnel

Dr. Xiangshang Li, Research Scientist, University of Houston

### Data to be audited

Model simulation period covered 30 days. We audited 10% data, i.e., a 3-day period.

The ozone concentration reached maximum in HGB on September 25, 2013. We audited 3 days: from 09/24 to 09/26. All the model input files and output files during the 3-day period were examined by Dr. Wonbae Jeon, University of Houston.

### Audit Procedures

Audit procedures were conducted on WRF input files, which are NARR files, and WRF output files which are "wrfout" files. Important parameters such as wind and temperature simulations were plotted to determine whether they are reasonable. The inputs for CMAQ are emission files and MCIP output. Again, important parameters such as $NO_2$ emissions have been checked and analyzed. Important output parameters such as $O_3$, HCHO, $NO_2$ and NO have been checked.

For observation data model and observation comparison plots were created and statistical evaluation for the 3-day period were performed. Further, the 30-day period simulations results were included in the final report.

# Appendix C

## WRF Namelist

```
&time_control
 run_days                            = 4,
 run_hours                           = 0,
 run_minutes                         = 0,
 run_seconds                         = 0,
 start_year                          = 2013, 2013,
 start_month                         = 08, 08,
 start_day                           = 29, 29,
 start_hour                          = 00, 00
 start_minute                        = 00, 00
 start_second                        = 00, 00
 end_year                            = 2013, 2013,
 end_month                           = 10, 10,
```

```
end_day                             = 03, 03,
end_hour                            = 00, 00
end_minute                          = 00, 00
end_second                          = 00, 00
interval_seconds                    = 10800
input_from_file                     = .true.,.true.,
history_interval                    = 60, 60,
frames_per_outfile                  = 12, 12,
auxinput11_interval                 = 60, 60,
auxinput11_end_h                    = 186, 186,
restart                             = .false.,
restart_interval                    = 720,
io_form_history                     = 2
io_form_restart                     = 2
io_form_input                       = 2
io_form_boundary                    = 2
debug_level                         = 0
/

&domains
time_step                           = 60,
time_step_fract_num                 = 0,
time_step_fract_den                 = 1,
max_dom                             = 2,
e_we                                = 161, 97,
e_sn                                = 145, 79,
e_vert                              = 28,   28,
p_top_requested                     = 10000,
num_metgrid_levels                  = 30,
num_metgrid_soil_levels             = 4,
dx                                  = 12000, 4000,
dy                                  = 12000, 4000,
grid_id                             = 1,     2,
parent_id                           = 0,     1,
i_parent_start                      = 1,     93,
j_parent_start                      = 1,     45,
parent_grid_ratio                   = 1,     3,
parent_time_step_ratio              = 1,     3,
feedback                            = 1,
smooth_option                       = 0
eta_levels                          = 1.000, 0.996, 0.990, 0.980, 0.970,
                                      0.960, 0.950, 0.940, 0.930, 0.920,
                                      0.910, 0.895, 0.880, 0.865, 0.850,
                                      0.825, 0.800, 0.775, 0.750, 0.720,
                                      0.660, 0.570, 0.475, 0.370, 0.250,
```

```
                                  0.145, 0.045, 0.000
/

&physics
mp_physics                          = 2,     2,
ra_lw_physics                       = 4,     4,
ra_sw_physics                       = 5,     5,
radt                                = 10,    10,
sf_sfclay_physics                   = 1,     1,
sf_surface_physics                  = 2,     2,
bl_pbl_physics                      = 1,     1,
bldt                                = 0,     0,
cu_physics                          = 1,     1,
cudt                                = 5,     5,
isfflx                              = 1,
ifsnow                              = 1,
icloud                              = 1,
surface_input_source                = 1,
num_soil_layers                     = 4,
sf_urban_physics                    = 0,     0,
maxiens                             = 1,
maxens                              = 3,
maxens2                             = 3,
maxens3                             = 16,
ensdim                              = 144,
cugd_avedx                          = 3,
/

&fdda
grid_fdda                           = 1,      1,
gfdda_inname                        = "wrffdda_d<domain>",
gfdda_end_h                         = 186,    186,
gfdda_interval_m                    = 180,    180,
fgdt                                = 0,      0,
if_no_pbl_nudging_uv                = 0,      0,
if_no_pbl_nudging_t                 = 1,      1,
if_no_pbl_nudging_q                 = 1,      1,
if_zfac_uv                          = 0,      0,
 k_zfac_uv                          = 10,     10,
if_zfac_t                           = 0,      0,
 k_zfac_t                           = 10,     10,
if_zfac_q                           = 0,      0,
 k_zfac_q                           = 10,     10,
guv                                 = 0.0003,   0.0003,
gt                                  = 0.0003,   0.0003,
```

```
gq                                  = 0.0003,   0.0003,
if_ramping                          = 1,
dtramp_min                          = 60.0,
io_form_gfdda                       = 2,
grid_sfdda                          = 1, 1,
sgfdda_inname                       = "wrfsfdda_d<domain>",
sgfdda_end_h                        = 186, 186,
sgfdda_interval_m                   = 60, 60,
io_form_sgfdda                      = 2,
guv_sfc                             = 0.0003,0.0003,
gt_sfc                              = 0.0003, 0.0003,
gq_sfc                              = 0.0003, 0.0003,
rinblw                              = 250.0,
obs_nudge_opt                       = 1, 1,
max_obs                             = 150000,
fdda_start                          = 0. 0.,
fdda_end                            = 99999. 99999.,
obs_nudge_wind                      = 1,1,
obs_coef_wind                       = 4.E-4, 6.E-4,
obs_nudge_temp                      = 1, 1,
obs_coef_temp                       = 4.E-4, 6.E-4,
obs_nudge_mois                      = 0, 0,
obs_coef_mois                       = 4.E-4, 6.E-4,
obs_rinxy                           = 150., 120.
obs_rinsig                          = 0.05,
obs_twindo                          = 0.6666667, 0.6666667,
obs_npfi                            = 10,
obs_ionf                            = 1, 1,
obs_idynin                          = 0,
obs_dtramp                          = 40.
obs_prt_freq                        = 50, 10,
obs_prt_max                         = 5,
obs_ipf_errob                       = .true.
obs_ipf_nudob                       = .true.
obs_ipf_in4dob                      = .true.
obs_ipf_init                        = .true.
/

&dynamics
w_damping                           = 1,
diff_opt                            = 1,
km_opt                              = 4,
diff_6th_opt                        = 0,      0,
diff_6th_factor                     = 0.12,   0.12,
base_temp                           = 290.
```

```
damp_opt                            = 0,
zdamp                               = 5000., 5000.
dampcoef                            = 0.2,    0.2,
khdif                               = 0,      0,
kvdif                               = 0,      0,
non_hydrostatic                     = .true., .true.,
moist_adv_opt                       = 1,      1,
scalar_adv_opt                      = 1,      1,
/

&bdy_control
spec_bdy_width                      = 5,
spec_zone                           = 1,
relax_zone                          = 4,
specified                           = .true., .false.,
nested                              = .false., .true.,
/

&grib2
/

&namelist_quilt
nio_tasks_per_group = 0,
nio_groups = 1,
/
```

## Publications from This Study


Pan, S., Choi, Y., Roy, A., Li, X., Jeon, W., and Souri A.H., Modeling the uncertainty of several VOC and its impact on simulated VOC and ozone in Houston, Texas, 2015, Atmospheric Environment, DOI:10.1016/j.atmosenv.20015.09.029

Li, X., Choi, Y., Czader, B., Kim, H., Lefer, B., and Pan, S., The impact of observation nudging on simulated meteorology and ozone concentrations during DISCOVER-AQ 2013 Texas campaign, 2015, Atmos. Chem. Phys. Discuss., 15, 27357-27404

Roy, A. and Choi, Y., Temperature dependence of source specific Volatility Basis Sets for motor vehicle exhaust, 2015, Atmospheric Environment, DOI:10.106/j.atmosenv.2015.08.035

Choi, Y. and Souri, A.H., Chemical condition and Surface Ozone in Large Cities of Texas During the Last Decade: Observational Evidence from OMI, CAMS, and Model analysis, 2015, Remote Sensing of Environment, 168, 90-101, DOI:10.1016/j.rse.2015.06.026

Jeon, W., Choi, Y., Lee, H-W., Lee, S-H., Yoo, J-W., Park, J., and Lee H-J., A quantitative analysis of grid nudging effect on each process of PM2.5 production in Korean Peninsula, 2015, Atmospheric Environment, doi:10.1016/j.atmosenv.2015.10.050

# References


Allen, D.J., Pickering, K.E., Pinder, R.W., Henderson, B.H., Appel, K.W., and Prados, A., 2012, Impact of lightning-NO on eastern United States photochemistry during the summer of 2006 as determined using the CMAQ model, Atmos. Chem. Phys., 12, 1737-1785

Anderson, D. C., Loughner, C. P., Diskin, G., Weinheimer, A., Canty, T. P., Salawitch, R. J. & Dickerson, R. R., 2014, Measured and modeled CO and NOy in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US, Atmospheric Environment, 96, 78-87.

Atkinson, R. and Arey, J., 1998, Atmospheric Chemistry of Biogenic Organic. Compounds, Acc. Chem. Res., 31, 574-583

Banta, R.M., Senff, C.J., Nielsen-Gammon, J., Darby, L.S., Ryerson, T.B., Alvarez, R.J., Sandberg, S.R., Williams, E.J., Trainer, M., 2005. A bad air day in Houston. Bulletin of the American Meteorological Society 86, 657-669.

Bucsela, E. J., et al., 2013, A new stratospheric and tropospheric $NO_2$ retrieval algorithm for nadir-viewing satellite instruments: application to OMI, Atmos. Meas. Tech., 6, 2607-2626, doi: 10.5194/amt-6-2607-2013

Choi, Y., & Souri, A. H., 2015, Chemical condition and surface ozone in large cities of Texas during the last decade: Observational evidence from OMI, CAMS, and model analysis. Remote Sensing of Environment, 168, 90-101.

Choi, Y., Kim, H., Tong, D., and Lee, P., 2012, Summertime weekly cycles of observed and modeled NOx and O3 concentrations as a function of satellite-derived ozone production sensitivity and land use types over the Continental United States, Atmos. Chem. Phys., 6291-6307.

Choi, Y., Kim, J., Eldering, A., Osterman, G., Yung, Y.L., Gu, Y., Liou, K.N., 2009, Lightning and anthropogenic NOx sources over the United States and the western north Atlantic Ocean: impact on OLR and radiative effects. Geophys. Res. Lett., 36, L17806, doi: 10.1029/2009GL039381

Choi, Y., Osterman, G., Eldering A., Wang, Y., and Edgerton, E., 2010, Understanding the contributions of anthropogenic and biogenic sources to CO enhancements and outflow observed over North America and the western Atlantic Ocean by TES and MOPITT, Atmospheric Environment, 44, 2033-2042

Choi, Y., 2014, The impact of satellite-adjusted NOx emissions on simulated NOx and O3 discrepancies in the urban and outflow areas of the Pacific and Lower Middle US, Atmos. Chem. Phys., 14, 675-690

Choi, Y., Wang, Y., Yang, Q., Cunnold, D., Zeng, T., Shim, C., Luo, M., Eldering, A., Bucsela, E., Gleason, J., 2008b, Spring to summer northward migration of high O3 over the western North Atlantic, 2008b, Geophys. Res. Lett., 35, L04818, doi: 10.1029/2007GL032276

Choi, Y., Wang, Y., Zeng, T., Cunnold, D., Yang, E., Martin, R., Chance, K., Thouret, V., Edgerton, E., 2008a. Springtime transition of NO2, CO and O3 over North America: Model

evaluation and analysis, Journal of Geophysical Research, 113, D20311, doi: 10.1029/2007JD009632

Choi, Y., Wang, Y., Zeng, T., Martin, R.V., Kurosu, T.P., Chance, K., 2005, Evidence of lightning NOx and convective transport of pollutants in satellite observations over North America. Geophys. Res. Lett., 32, L02805, doi: 10.1029/2004GL021436

Czader, B. H., Choi, Y., Li, X., Alvarez, S., and Lefer, B., 2015, Impact of updated traffic emissions on HONO mixing ratios simulated for urban site in Houston, Texas, Atmos. Chem. Phys., 15, 1253-1263, doi:10.5194/acp-15-1253-2015.

Czader, B.H., Li, X.S., Rappenglueck, B., 2013, CMAQ modeling and analysis of radicals, radical precursors, and chemical transformations. J Geophys Res-Atmos 118, 11376-11387.

Duncan, B.N., Yoshida, Y., Olson, J.R., Silman, S., Martin, R.V., Lamsal, L., Hu, Y., Pickering, K.E., Retscher, C., Allen, D.J., Crawford, J.H., 2010, Application of OMI observations to a space-based indicator of NOx and VOC controls on surface ozone formation, Atmospheric Environment, 44(18), 2213-2223.

Guenther, A., Baugh, B., Brasseur, G., Greenberg, J., Harley, P., Klinger, L., Serca, D., and Vierling, L., 1999, Isoprene emission estimates and uncertainties for the Central African EXPRESSO study domain, J. Geophys. Res., 104, D23, 30625-30639

Guenther, A., Hewitt, C.N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W.A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P., 1995, A global model of natural volatile organic compound emissions, J. Geophys. Res., 100, D5, 8873-8892

Houyoux, M.R., Vulkovich, J.M., Coats, C.J., Wheeler, N.M., and Kasibhatla, P.S., 2000, Emission inventory development and processing for the Seasonal Model for Regional Air Qaulity (SMRAQ) project, J. Geophy. Res., 105, 9079-9090

Hudman, R.C., Murray, L.T., Jacob, D.J., Millet, D.B., Turquety, S., Wu, S., Blake, D.R., Goldstein, A.H., Holloway, J., and Sachse, G.W., 2009, Biogenic versus anthropogenic sources of CO in the United States, Geophys. Res. Lett., 35, L04801, doi: 10.1029/2007GL032393

Hudman, R.C., Russell, A.R., Valin, L.C., and Cohen, R.C., 2010, Interannual variability in soil nitric oxide emissions over the United States as viewed from space, Atmos. Chem. Phys., 10, 9943-9952

Kemball-Cook, S., Yarwood, G., Johnson, J., Dornblaser, B., & Estes, M., 2015, Evaluating NOx emission inventories for regulatory air quality modeling using satellite and air quality model data. Atmospheric Environment, 117, 1-8.

Kim, S.-W., McKeen, S.A., Frost, G.J., Lee, S.-H., Trainer, M., Richer, A., Angevine, W.M., Atlas, E., Bianco, L., Boersma, K.F., and et al., 2011, Evaluations of NOx and highly reactive VOC emission inventories in Texas and their implications for ozone plume simulations during the Texas Air Quality Study 2006. Atmos. Chem. Phys., 11, 11361-11386, doi: 10.5194/acp-11-11361-2011.

Kuhlmann, G., Hartl, A., Cheung, H. M., Lam, Y. F., and Wenig, M. O., 2014, A novel gridding algorithm to create regional trace gas maps from satellite observations, Atmos. Meas. Tech., 7, 451-467, doi:10.5194/amt-7-451-2014.

Li, X. and Rappengluck, B., A WRF-CMAQ study on spring time vertical ozone structure in Southeast Texas, 2014, Atmos. Environ., 97, 363-385

Li, X., Choi, Y., Czader, B., Kim, H., Lefer, B., and Pan, S., The impact of observation nudging on simulated meteorology and ozone concentrations during DISCOVER-AQ 2013 Texas campaign, 2015, Atmos. Chem. Phys. Discuss., 15, 27357-27404

Martin, R.V., Fiore, A.M., and Donkelaar, A.V., 2004, Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions, Geophys. Res. Lett., 31, L06120, doi: 10.1029/2004GL019416

Millet, D.B., Goldstein, A.H., Holzinger, R., Williams, B.J., Allan, J.D., Jimenez, J.L., Worsnop, D.R., Roberts, J.M., White, A.B., Hudman, R.C., Bertschi, I.T., Stohl, A., 2006, Chemical characteristics of north American surface layer outflow: insights from Chebogue point, Nova Scotia, J. Geophys. Res., 111, D23S53

Ngan, F., Byun, D., Kim, H., Lee, D., Rappengluck, B., Pour-Biazar, A., 2012. Performance assessment of retrospective meteorological inputs for use in air quality modeling during TexAQS 2006. Atmos. Env. 54, 86-96.

Noxon, J. F., 1978, Tropospheric $NO_2$, J. Geophys. Res., 83(C6), 3051–3057, doi: 10.1029/JC083iC06p03051.

Palmer, P.I., Jacob, D.J., Fiore, A.M., Martin, R.V., Chance, K., and Kurosu, T.P., 2003, Mapping isoprene emissions over North America using formaldehyde column observations from space, J. Geophys. Res., 108, D6, 4180, doi: 10.1029/2002JD002153

Pan, S., Choi, Y., Roy, A., Li, X., Jeon, W., and Souri, A.H., 2015, Modeling the uncertainty of several VOC and its impact on simulated VOC and ozone in Houston, Texas, Atmos. Environ., 120, 404-416

Pfister, G.G., Emmons, L.K., Hess, P.G., Lamarque, J.-F., Orlando, J-J., Walters, S., Guenther, A., Palmer, P., Lawrence, P.J., 2008, Contribution of isoprene to chemical budgets: a model tracer study with the NCAR CTM MOZART-4, J. Geophys. Res., 113, D05308

Rappenglück, B., B. Lefer, B. Czader, X. Li, J. Golovko, S. Alvarez, J. Flynn, C. Haman, and N. Grossberg. 2011. University of Houston Moody Tower 2010 Ozone Formation Research Monitoring; Project Grant No. 582-5-64594-FY10-14, Prepared for the Texas Commission on Environmental Quality, TX 2011

Rodgers, C.D., 2000, Inverse Methods for Atmospheric Sounding: Theory and Practice, World Sci., Hackensack, N.J.

Seinfeld, J. H. and Pandis, S. N., 2012, Atmospheric chemistry and physics: from air pollution to climate

Sharkey, T.D., Singsaas, E.L., Lerdau, T., Geon, C.D., 1999, Weather effects on isoprene emission capacity and applications in emissions in emission algorithms, Ecological Applications, 9, 1131-1137

Shim, C., Wang, Y., Choi, Y., Palmer, P.I., Abbot, D.S., and Chance, K., 2005, Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, J. Geophys. Res., 110, D24301, doi: 10.1029/2004JD005629

Sillman, S., 1995. The use of NOy, $H_2O_2$, and $HNO_3$ as empirical indicators for ozone-$NO_x$-hydrocarbon sensitivity in urban locations. J. Geophys. Res., 100, 14,175–14,188.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., 2006, and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423-3441, doi:10.5194/acp-6-3423-2006.

Yienger, J. J., and H. Levy, 1995, Empirical model of global soil-biogenic NOχ emissions, J. Geophys. Res., 100(D6), 11447–11464, doi: 10.1029/95JD00370. John Wiley & Sons.

Zhao, C., Wang, Y., Choi, Y., and Zeng, T., 2009, summertime impact of convective transport and lightning NOx production over North America: modeling dependence on meteorological simulations, Atmos. Chem. Phys., 9, 4315-4327.