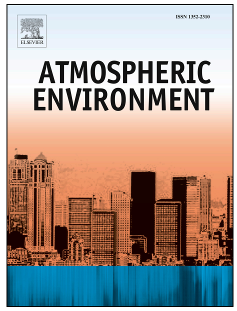


Constraining the uncertainty in emissions over India with a regional air quality model evaluation

Alexandra Karambelas, Tracey Holloway, Gregor Kiesewetter, Chris Heyes

PII: S1352-2310(17)30813-0
DOI: 10.1016/j.atmosenv.2017.11.052
Reference: AEA 15709

To appear in: *Atmospheric Environment*

Received Date: 10 July 2017
Revised Date: 17 November 2017
Accepted Date: 27 November 2017

Please cite this article as: Karambelas, A., Holloway, T., Kiesewetter, G., Heyes, C., Constraining the uncertainty in emissions over India with a regional air quality model evaluation, *Atmospheric Environment* (2017), doi: 10.1016/j.atmosenv.2017.11.052.

Constraining the uncertainty in emissions over India with a regional air quality model evaluation

Alexandra Karambelas[1,*], Tracey Holloway[1,2], Gregor Kiesewetter[3], Chris Heyes[3]

[1] Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin—Madison, 1710 University Avenue, Madison, WI, USA
[2] Department of Atmospheric and Oceanic Sciences, University of Wisconsin—Madison, 1225 West Dayton Street, Madison, WI, USA
[3] Air Pollution and Greenhouse Gases, International Institute for Applied Systems Analysis, Schloßplatz 1, 2361 Laxenburg, Austria

* *Corresponding author present address*: The Earth Institute, Columbia University, 2910 Broadway, New York, NY, USA

E-mail addresses: ak4040@columbia.edu (A. Karambelas); taholloway@wisc.edu (T. Holloway); kiesewet@iiasa.ac.at (G. Kiesewetter) heyes@iiasa.ac.at (C. Heyes)

Abstract

To evaluate uncertainty in the spatial distribution of air emissions over India, we compare satellite and surface observations with simulations from the U.S. Environmental Protection Agency (EPA) Community Multi-Scale Air Quality (CMAQ) model. Seasonally representative simulations were completed for January, April, July, and October 2010 at 36km x 36km using anthropogenic emissions from the Greenhouse Gas-Air Pollution Interaction and Synergies (GAINS) model following version 5a of the Evaluating the Climate and Air Quality Impacts of Short-Lived Pollutants project (ECLIPSE v5a). We use both tropospheric columns from the Ozone Monitoring Instrument (OMI) and surface observations from the Central Pollution Control Board (CPCB) to closely examine modeled nitrogen dioxide (NO_2) biases in urban and rural regions across India. Spatial average evaluation with satellite retrievals indicate a low bias in the modeled tropospheric column (-63.3%), which reflects broad low-biases in majority non-urban regions (-70.1% in rural areas) across the sub-continent to slightly lesser low biases reflected in semi-urban areas (-44.7%), with the threshold between semi-urban and rural defined as 400 people per km^2. In contrast, modeled surface NO_2 concentrations exhibit a slight high bias of +15.6% when compared to surface CPCB observations predominantly located in urban areas. Conversely, in examining extremely population dense urban regions with more than 5000 people per km^2 (dense-urban), we find model overestimates in both the column (+57.8) and at the surface (+131.2%) compared to observations. Based on these results, we find that existing emission fields for India may overestimate urban emissions in densely populated regions and underestimate rural emissions. However, if we rely on model evaluation with predominantly urban surface observations from the CPCB, comparisons reflect model high biases, contradictory to the knowledge gained using satellite observations. Satellites thus serve as an important emissions and model evaluation

metric where surface observations are lacking, such as rural India, and support improved emissions inventory development.

Keywords: India; model; satellite; OMI; NO2; emissions

1. Introduction

Ambient air pollution results in 3.7 million annual deaths worldwide—contributing to 6.7% of the world's total annual deaths. Of these air-pollution-related mortalities, 88% occur in developing and low-income countries [*World Health Organization*, 2014]. Ambient air pollution causes premature death most often resulting from respiratory illnesses, heart disease, cancer, and stroke [*Lim et al.*, 2012]. India suffers from some of the worst air pollution in the world, owing to its rapid economic development, increasing population, growth in energy demand, and limited air-pollution regulation.

Much of the ambient pollution in India is a result of anthropogenic emissions from biomass burning [*Reddy and Venkataraman*, 2002; *Sharma et al.*, 2015], agricultural waste burning [*Liu et al.*, in review], and fossil fuel combustion for transportation [*Apte et al.*, 2011] and industrial processing and electricity combustion [*Reddy and Venkataraman*, 2002; *Guttikunda and Jawahar*, 2014]. Industrial sources are also often coincident with urbanized regions, as evident in *Garg et al*. [2001], and are noticeable "hot spots" detectable by satellite [*Ramachandran et al.*, 2013]. India's Central Pollution Control Board (CPCB) identified 88 such industrial hot spot clusters, which are found predominantly near cities and in the industrial regions of eastern India [*Central Pollution Control Board*, 2009]. Contributions to pollution from vehicles are predominantly urban (on-road), but rural areas can also be affected by off-road sources for farming [*Guttikunda and Mohan*, 2014]. Primary particulate emissions from residential combustion sectors are more common in rural and low-income urban regions, where people rely more on traditional biomass to meet their cooking and heating needs.

Urbanization, industrialization, and population growth are leading causes of India's growing ambient pollution problem. Major industrial manufacturing and processing sources in India include smelting, cement production, sulfuric acid production, and brick kilns, sources which in total are estimated to contribute 36% of total SO_2 and 19% of total NO_2 emissions in the country [*Garg et al.*, 2001]. Brick kilns alone have been estimated to contribute more than 70% of ambient PM_{10} and up to 60% of the $PM_{2.5}$ in certain parts of India [*Muntaseer Billah Ibn Azkar et al.*, 2012]. Coal-fired power plants for electricity generation contribute 50% of total sulfur dioxide (SO_2) and 30% of total nitrogen oxides (NO_X) emissions in India [*Garg et al.*, 2006], such that coal-fired generation contributes 96% of emissions from the power sector [*Lu and Streets*, 2012]. Transportation emissions of NO_X and SO_2 contribute up to one third of $PM_{2.5}$ [*Amann et al.*, 2017], compounding the already severe problem of particulate pollution in the region. However, contributions from individual sectors vary regionally, including between major urban areas [*Guttikunda et al.*, 2014]. Better constraining the budget of NO_X emissions from all sources can address the significant uncertainties across emission inventories, sectors, and pollutant species [*Saikawa et al*., 2017].

Due to limited ground-based measurement sites in India with varying levels of data reliability, past studies have often used vertical column densities (VCDs) from satellites to inform emissions, distributions, and recent trends of tropospheric NO_2 [*Lu and Streets*, 2012; *Ghude et al.*, 2013], SO_2 [*Fioletov et al.*, 2011], and PM via the interpretation of aerosol optical depth (AOD) [*Ramachandran*, 2007]. Satellite-based approaches have informed trends over recent decades, and provided data to supplement and compare with ground-based instruments. NO_X and SO_2 pollution from power plants have increased by more than 70% from 1996 to 2014 and 2005-2012 respectively as observed by temporal trend observations from satellites [*Lu and Streets*, 2012; *Lu et al.*, 2013]. Trends in tropospheric NO_2 at selected industrial areas have been found to increase at a rate of 1 to 9% per year [*Ramachandran et al.*, 2013], with a regional average decadal increase from 2004-2015 on the order of 14% [*Zia ul-Haq et al.*, 2015]. The largest growth in VCDs is over areas of high population density in the north, attributable to enhanced electricity production, industrial activity, transportation, and crop burning, trends not as prominent in southern India [*Duncan et al.*, 2015; *Zia ul-Haq et al.*, 2015]. However, recent developments, including slight stagnation due to economic slow-down [*Hilboll et al.*, 2017], indicate the complex nature of pollution trends in the region which may be unaccounted for in current emissions inventories for the region.

In this study, we use the U.S. Environmental Protection Agency's Community Multi-Scale Air Quality Model (CMAQ) to simulate recent air quality conditions for the Indian subcontinent using anthropogenic emissions from the Greenhouse Gas-Air Pollution Interactions and Synergies (GAINS) model following version 5a of the Evaluating the Climate and Air Quality Impacts of Short-Lived Pollutants project (ECLIPSE v5a). Previous assessments of the region have relied on statistical modeling of pollution levels in urban areas [*Chaudhuri and Dutta*, 2014; *Mishra and Goyal*, 2015], urban and industrial dispersion modeling [*Kumar and Goyal*, 2014; *Saini et al.*, 2014; *Aggarwal and Jain*, 2015; *Gulia et al.*, 2015], and evaluating sector contributions [*Guttikunda and Jawahar*, 2012; *Gupta and Mohan*, 2013; *Chambliss et al.*, 2014; *Sharma et al.*, 2016]. Although a few studies have sought to use advanced chemistry and transport models to evaluate Indian air quality [*Ghude et al.*, 2013; *Guttikunda and Jawahar*, 2014], and CMAQ has previously been used in larger East Asian domains [*Chatani et al.*, 2014; *Park*, 2015], over Bangladesh [*Muntaseer Billah Ibn Azkar et al.*, 2012], and to assess ground-level O_3 in India [*Sharma et al.*, 2016], all applications of CMAQ and related models depend on the accuracy of the input emissions. Here we use CMAQ to evaluate the skill of this advanced emissions inventory, by comparing calculated ambient concentrations and VCDs with a suite of observations on a national scale and four-season basis to identify and assess regional differences in model performance.

2. Methods

2.1. Model Description

Model simulations were conducted using CMAQ v5.0.1 at 36 km x 36 km over the Indian subcontinent and surrounding countries, including parts of Afghanistan, Bangladesh, Bhutan, China, Nepal, and Pakistan (5°N to 40°N, 60°E to 100°E), for four seasonally representative months—January, April, July, and October—representing winter, pre-monsoon, monsoon, and post-monsoon fall respectively. The CMAQ model includes processes related to surface- and upper- level emissions, photolysis, gaseous and particulate chemistry, deposition, and dispersion for 36 vertical layers in the troposphere up to about 150hPa [*Byun and Schere*, 2006]. Model specifications include the Carbon Bond 05 (CB05) chemical mechanism [*Yarwood et al.*, 2005], the AERO 6 aerosol mechanism, in-line lightning NO_X production [*Allen et al.*, 2012], and the inclusion of windblown dust [*Dong et al.*, 2015]. Boundary and initial conditions are taken as the CMAQ default profiles, which assumes location and seasonal invariance in vertical chemical profiles.

Anthropogenic emissions from the Greenhouse Gas–Air Pollution Interactions and Synergies (GAINS) integrated assessment model [*Amann et al.*, 2011], developed and operated at the International Institute for Applied Systems Analysis (IIASA), are for year 2010. Sub-national total emissions for 10 species emitted from anthropogenic sectors were calculated using detailed activity factors and combustion information as described for $PM_{2.5}$ in *Klimont et al.*, [2016]. The GAINS inventory includes energy combustion, domestic combustion, transportation, agriculture, area sources, the extraction and removal of energy sources, and other anthropogenic combustion sectors. Gridding sub-national emissions to 0.5 degree x 0.5 degree global fields was conducted according to the ECLIPSE (Evaluating the Climate and Air Quality Impacts of Short-Lived Pollutants) project which uses sector-specific spatial surrogates according to EDGAR (Emissions Database for Global Atmospheric Research) as described in *Lamarque et al.* [2010]. Annual total emissions were allocated temporally and vertically as follows: day and night emissions ratios (each 12 hours long) for each anthropogenic emission sector followed global model parameterizations described in *Simpson et al.* [2012]. Domestic combustion, industrial manufacturing, solvent emissions, and mobile sources were assumed to occur primarily in daytime. Vertical distributions are based on power plant stack height, such that power generation and industrial processing and manufacturing were distributed in the first eight model layers and dispersed through nearby layers up to ~1000m. Surface emissions sources were assigned to the lowest model layer. Values for these distributions can be found in Supplemental Tables 1 and 2. Emissions from GAINS were chemically speciated for inclusion in CMAQ from 10 to 32 species, with speciation factors adapted from the Sparse Matrix Operator Kernel Emissions (SMOKE) model, where average speciation factors were applied across all anthropogenic sectors in the same way for all combustion sources. Speciation information for VOC compounds is adapted from speciation developed by Drs. Qiang Zhang and David G. Streets for the INTEX-B project over Asia [*Li et al.*, 2014], and particulate speciation is adapted from *Chowdhury et al.* [*Chowdhury et al.*, 2007]. Detailed speciation factors can be found in Supplemental Table 3.

Global biogenic emissions are from the Model of Emissions and Gases from Nature (MEGAN) on a monthly average basis from the MEGAN website[1], calculated from the Community Land Model, which includes emissions for 25 gaseous species at 0.1° by 0.1°. These emissions were allocated to the 36 km by 36 km Lambert-conformal grid, with all emissions occurring in the lowest model layer, and during daytime hours (6 am to 6 pm local time) for simulations in each season.

Biomass burning emissions were taken from the Global Fire Emissions Database version 4.1 with small fires (GFED v4.1s) [*Randerson et al.*, 2012]. Emissions were allocated from 0.5° by 0.5° latitudinal-longitudinal grid to 36 km by 36 km. Biomass burning VOC speciation was performed following *Akagi et al.*,[2011]. Biomass burning emissions from GFED were distributed temporally according to the GFED v4.1s dataset and vertically using burned area and emissions buoyancy flux as determined by the fire size per grid cell as described in *Fu* et al. [2012] and *Pouliot* et al., [2005].

Annual total anthropogenic emissions of NOx, SO_2, and $PM_{2.5}$ (not including windblown dust) are shown in Figure 1. Emissions of NO_X (1a) are greatest in highly populated mega cities and nearby such as Delhi and Kolkata, and Mumbai. NO_X emissions “hotspots” occur scattered across India indicative of urban pollution from transportation and other combustion but for the most part highest emissions remain coincident with the largest Indian cities. Emissions of SO_X exhibit a similar pattern to that of NO_X emissions. Comparing with population densities in Figure 1d, highest emissions are coincident with highly populated cities and near combustion sources, namely industry in eastern India (1b). In contrast, primary $PM_{2.5}$ emissions are significantly lower across India, with regions of greatest $PM_{2.5}$ emissions restricted to Delhi, Kolkata, and Mumbai (1c). Primary particulates are often in the form of organic and elemental carbon from combustion

[1] http://lar.wsu.edu/megan/docs/05degree_MEGAN/

sources, which according to the gridded ECLIPSE sectoral spatial surrogates concentrates the emissions in urban regions.

Meteorology for 2010 is simulated using the Weather Research and Forecasting (WRF) model v3.2 and Preprocessing System (WPS) and ERA-Interim reanalysis from the European Center for Medium-Range Weather Forecasting (ECMWF) [*Dee et al.*, 2011]. Reanalysis weather data is globally gridded at about an 80 km resolution over 60 vertical layers and available in 6-hour increments. WRF is used to interpolate 6-hour data to hourly data. Data from WRF is simulated using Grell cumulus parameterization [*Grell and Devenyi*, 2002] with 36 vertical sigma layers from the surface to approximately 150hPa. Meteorological data was finally preprocessed for use in CMAQ with the Meteorology-Chemistry Interface Processor (MCIP). Figure 2 describes the seasonal variation in monthly average planetary boundary layer height, temperature, and total rainfall from MCIP and compares rainfall from TRMM. Figure 2 shows monthly average planetary boundary layer (PBL) heights (purple), temperatures (orange), and total rainfall (green) for January (top) and July (bottom). Generally, these seasons differ with lower (higher) PBL heights, cooler (warmer) temperatures, and less (more) rainfall in January as opposed to July. There are noticeable variations across the sub-continent: PBL heights are at their lowest along the Himalayan mountain range in January (2a), a region that also exhibits extreme temperature shifts during the year from coldest in January (2b) to warmest in July (2f). Finally, precipitation is limited in January but can exceed 100 centimeters per month in certain regions, particularly along the Himalayas and the Western Ghats mountain ranges, during July (Figures 2c and 2g). Monthly precipitation totals were validated against measurements from the Tropical Rainfall Measuring Missions (TRMM) microwave imager instrument shown in Figures 2d and 2h. MCIP reproduces January rainfall conditions fairly well however July precipitation totals are underestimated in central India. A similar image for April and October meteorology is presented in Supplemental Figure 1.

2.2. Satellite and Ground-Based Measurements

We compared CMAQ output with observational data from satellite and ground-based instruments. The OMI instrument aboard the Aura satellite [*National Aeronautics and Space Administration*, 2012] supports the calculation of tropospheric NO_2 and formaldehyde (HCHO) VCDs. Observations from OMI have been previously used in regional model evaluation over regions of the U.S. [*Canty et al.*, 2015; *Kemball-Cook et al.*, 2015]. Daily total column values for NO_2 and HCHO were downloaded from the TEMIS database[2] in a Level 2 data format, and gridded to the 36 km x 36 km model grid with the Wisconsin Horizontal

---

[2] http://www.temis.nl/airpollution/no2col/no2regioomi_v2.php

Interpolation Program for Satellites (WHIPS) [*Harkey et al.*, 2015]. An averaging kernel was applied to model simulations at the Aura overpass time of about 2PM to calculate equivalent VCDs for comparison with the satellite-derived values. Annual average evaluation with OMI NO_2 VCDs is presented in Section 3.1; seasonal OMI NO_2 VCD and HCHO evaluation is included in the Supplemental Information.

Two sets of ground-based surface observations were employed: one, from the peer-reviewed literature for NO_2 [*Carmichael et al.*, 2009; *Guttikunda et al.*, 2013; *Chaudhuri and Dutta*, 2014; *Mallik and Lal*, 2014; *Mallik et al.*, 2014], SO_2 [*Carmichael et al.*, 2009; *Guttikunda and Calori*, 2013; *Guttikunda et al.*, 2013; *Chaudhuri and Dutta*, 2014; *Mallik and Lal*, 2014; *Mallik et al.*, 2014; *Surendran et al.*, 2015], O_3 [*Guttikunda et al.*, 2013; *Mallik et al.*, 2014; *Surendran et al.*, 2015], and $PM_{2.5}$ [*Carmichael et al.*, 2009; *Guttikunda and Calori*, 2013; *Guttikunda et al.*, 2013; *Chaudhuri and Dutta*, 2014; *Mallik and Lal*, 2014; *Mallik et al.*, 2014; *Surendran et al.*, 2015]; the other accessed via the CPCB online data portal. Comparisons for NO_2 are presented here; detailed comparisons for the other pollutants are included in the Supplemental Information. Measurements reported in the literature are most often annual average concentrations, collected between 2005-2010. CPCB measurements use traditional monitoring techniques [*Central Pollution Control Board*, 2003], and when available, can be retrieved for individual monitor locations at hourly intervals at the download portal (http://www.cpcb.gov.in/CAAQM/mapPage/frmindiamap.aspx). Downloaded data from the CPCB is available from a maximum of 26 sites per January, April, July and October 2010. The values presented for comparison in this work are for all data available. Monitor locations from the literature (triangles) and CPCB (circles) sites are shown in Figure 1d. CPCB hourly monitoring data is geographically limited, with most monitors located in Delhi and immediate surroundings, hence why we also include observations from the literature for enhanced spatial comparison.

3. Results

3.1. Tropospheric VCD Evaluation with OMI NO_2

Tropospheric NO_2 columns averaged over January, April, July, and October shown in Figure 3 for the model (3a) and from OMI (3b). Comparisons between modeled NO_2 tropospheric columns and those from OMI both reflect high average NO_2 VCDs across the Himalayan-bordering northern India, as well as in cities such as Delhi, Mumbai, Kolkata, and Pune (Figure 3a). Other areas of high NO_2 column VCDs exist throughout the domain including the cities of Dhaka, Bangladesh, and Lahore and Karachi in Pakistan. Monthly variations indicate the lowest NO_2 VCDs in July associated with monsoon season, and highest VCDs in January corresponding with longer NO_2 lifetimes, lower rainfall, shallow boundary-layer height, and reduced wind speeds (see Supplemental Figure 2). Increased wintertime accumulation of local and regional air pollution occur along the Himalayas because of shallow boundary layer heights and reduced mixing. Because of heavy monsoon rains in July, WHIPS algorithms that filter and remove pixels with cloud cover greater than 30% are more prevalent than in any other month, reflecting a limitation in using satellite observations during the monsoon season as opposed to winter (Figure 4). High-density NO_2 VCDs in eastern India are co-located with emissions from Hindalco's aluminum manufacturing plants, and the largest plants are located in Renukoot in the southeastern parts of Uttar Pradesh. Tropospheric column densities over Renukoot, with its population of about 350,000 people, are about as large as the densities found over Delhi, a megacity with a population surpassing 16 million people. Similar hot spots are visible in the annual average NO_2 column VCDs in the nearby states of Chhattisgarh and Odisha. These isolated regions of enhanced electricity generation and industrial processing are also observed in satellite analyses by *Duncan et al.*, [*Duncan et al.*, 2015], *Lu and Streets* [*Lu and Streets*, 2012], *Lu et al*, [2013], and most recently *Hilboll et al.*, [*Hilboll et al.*, 2017] which note this region as an area of large increases in NO_2 and SO_2 VCDs between the early 2000s and early- to mid-2010s due to electricity generation and industrial processing. Industries in India are subject to few emissions regulations, hence high NO_2 column densities in this region are unsurprising.

We first define semi-urban and rural grid cells using a threshold following the Indian census definition of urban population density of approximately 400 people per km^2 (among other classifications, [*Census-India*, 2012], to classify as urban (above 400 people) or rural (below 400 people). Because of this classification, there are 5871 grid cells in India that fall in the "Rural" category representing a population of 354 million, as opposed to 707 that fall within the "Semi-Urban" category representing a population of 858 million. The proportion of rural-designated population is approximately 69% according to the 2011 Census [*Census-India*, 2012], however our under-representation of the rural population at 29.3% occurs because of the size of each grid cell, which is too coarse to account for sub-grid cell population differences. By our definition, a grid cell is semi-urban if the population density in that grid cell is quite small, for instance a fraction of the population of New Delhi's National Capital Region 29 million individuals, meaning this distribution likely over-accounts for the sprawl of major urban centers.

Strong gradients of NO_2 VCD are visible between highly populated or industrialized areas, as compared to the rural background, in line with the rather local nature of NO_2 pollution. However, compared to OMI NO_2, CMAQ consistently underestimates column densities in both semi-urban and rural regions according to our population density distribution. The rural bias is quite large at -70.1% while the semi-urban bias is somewhat lower at -44.7%; overall, the model bias in tropospheric NO_2 columns is -63.3%. Stronger rural biases likely incorporate underestimates in industrial areas to the east (Figure 3b), which maintain low population densities and abundant electricity generating and industrial capacity, but significant underestimates in the modeled concentrations. Modeled underestimates of NO_2 tropospheric columns are evident across central and southern India as well where population density is relatively lower. This spatial and statistical comparison suggests inconsistencies in the emissions inventories, including eastern Indian electricity-generating and industrial regions, but also perhaps across semi-urban areas as well particularly with respect to the emissions column distribution. Despite these differences, monthly and spatial variations in CMAQ's NO_2 tropospheric column density mimic what is observed by OMI (Supplemental Figure 2). Seasonally, CMAQ exhibits large overestimates of NO_2 column VCDs in Nepal (for January, July, October) and Bhutan (especially in simulations for January and October). Such overestimation is likely due to difficulties in CMAQ accurately modeling the Himalayan topography. Highest VCDs occur in January and coincide with both a shallow boundary layer and low rainfall—characteristics of wintertime meteorology—and generally reach a minimum in July due to the highest levels of mixing and the great rainfalls of the monsoon season.

Statistical metrics including correlation, normalized mean bias, and normalized mean error following Eder et al., [2006] were calculated for average NO_2 VCDs for CMAQ and from OMI. The correlations between annual average OMI and CMAQ tropospheric NO_2 columns are positive, with an average spatial r^2=0.63 (Table 1). The strongest correlations are in April (r^2=0.68), after the dry, polluted winter, and weakest in July (r^2=0.39), coincident with the wet monsoon season that limits OMI retrieval availability and a low precipitation bias in MCIP, resulting in greater modeled pollution compared to OMI observations. The annual average normalized mean bias is large and negative (NMB_{OMI}=-63.3%), with a large low bias in July (-71.1%) and the smallest low bias in polluted January (-46.9%), suggesting CMAQ is better at estimating higher NO_2 VCDs as opposed to lower values. The annual average normalized mean error is large at 68.9%.

3.2. NO_2 Evaluation with Ground-Based Monitors

Four-month averaged modeled concentrations of NO_2 are overlaid with observations from the literature (hollow triangles) and from the CPCB (hollow circles) are presented in Figure 3c. Following Section 3.1.1, statistical metrics were calculated between daily average NO_2 from the model and from the CPCB. Comparisons between modeled surface concentrations and CPCB observations for SO_2, O_3, and $PM_{2.5}$ are included in the Supplemental Information.

Modeled concentrations of NO_2 are greatest along the Himalayas in northern India, stretching from Pakistan through India and into Bangladesh (3c), following relatively greater NO_2 VCDs in Figures 3a and

b. Generally, modeled concentrations follow those of population density which is greatest in the north and in urban centers outside of the Indo-Gangetic Plain (see Figure 1d). Easily identifiable urban areas in this region include Karachi in Pakistan; Mumbai, Surat, Ahmedabad, and Kolkata in India; and Dhaka in Bangladesh. The domain 4-month average modeled concentration of NO_2 is 2.0 ppb, with an average maximum of 41.4 ppb in Delhi. Outside of mega cities across the central and southern sub-continent, modeled surface concentrations exhibit significantly lower NO_2 values, on the order of four times lower than in urban areas. Concentrations of NO_2 from CMAQ are overlaid with observations from the literature (triangles) and from the CPCB (circles). Most observation locations from the CPCB (circles) are found in or downwind of Delhi (13 out of 26 monitors for 2010). Modeled concentrations are high coincident with observations from the literature (triangles) along the coast in Kolkata and inland in Delhi, while surface concentrations from CMAQ at Jodhpur (central western India; 3.7 ppb) and Nagpur (central India; 8.3 ppb) are lower than surface observations of 11.8 and 16.1 ppb respectively.

Statistical comparisons between daily modeled and observed NO_2 concentrations at CPCB monitor locations (circles) indicate CMAQ has a low spatial correlation (r^2=0.27) and an average slight model high bias (NMB_{CPCB}=+15.6%) at these monitor locations. According to our definition of Semi-Urban and Rural based on a population density threshold of 400 people per km^2, all NO_2 monitors from the CPCB are in semi-urban population density locations. Thus model high biases are reflective of modeled concentrations in high population density regions, and this comparison may not be representative of concentrations outside of urban areas, or even outside of Delhi where a majority of monitors are located, and therefore model biases in the rest of India remain uncertain when compared with CPCB observations. Finally, model errors are large, with an average daily NME at surface sites of 72.4%.

3.3. Assessing Model Performance in Urban and Rural Environments for NO_2

There are apparent inconsistencies in the statistical biases between ground level modeled NO_2 concentrations and satellite observations, namely that urban model performance exhibits a slight high bias compared only to surface observations whereas total column model comparison against OMI NO_2 VCDs indicates significant low biases in both semi-urban and rural defined grid cells (Figure 5a). Starting with the low bias in model performance compared to OMI NO_2 tropospheric VCDs (NMB_{OMI}=-63.3%), we note low model biases at both semi-urban locations (NMB_{OMI_urb}=-44.7%) and at rural locations (NMB_{OMI_rur}=-70.1%), colored in green, as compared to OMI observations, colored in light blue slash marks (Figure 5a, left two column pairs). The average is heavily weighted towards the low estimate, considering there are many more grid cells in India marked as "Rural" than there are "Semi-Urban." A similar performance disparity occurs when comparing model performance at surface monitor locations (Figure 5a, right two column pairs). Rural and urban modeled grid cells (green) corresponding to observation locations from the peer-reviewed literature (solid dark blue) and from the CPCB monitor locations (solid light blue) are shown in the right side of Figure 5 for NO_2. Only urban CPCB monitors and only rural literature observations exist, and we show comparable surface concentrations from CMAQ that best correspond with CPCB observations. The surface concentration comparison reflects the slight model high bias of +15.6% at CPCB monitor locations, contradicting the comparison across a larger compilation of modeled and OMI-observed semi-urban grid cells. From the CPCB comparison, we are unable to conclude rural model performance at the surface.

Looking only at extremely population dense urban regions with population densities greater than 5000 people per km^2, Dense-Urban and Rural modeled column VCDs and surface concentrations of NO_2 reflect different biases compared with observations (Figure 5b). With this urban-rural specification, only 4 grid cells within India are determined to be urban, representing Delhi (2 grids), Kolkata, and Mumbai, while 6574 are rural and include cities with relatively lower population densities. Model estimated columns exhibit large high biases at these locations (NMB_{high_OMI}=+57.8%), while rural modeled grid cells exhibit low biases (NMB_{low_OMI}=-63.2%) (Figure 5b, left two column pairs). Similar divergent biases are exhibited for modeled and observed surface concentrations, where some CPCB monitors are now reflected in the lower population dense rural grid cell category and all observations from the literature are in the rural grid cell category (Figure 5b, right two column pairs). Modeled surface biases averaged across these four CPCB sites reflecting extreme population density are NMB_{high_CPCB}=131.2%, while low biases across other monitor locations are NMB_{low_CPCB}=-20.3%. Through this, we find that although modeled dense-urban regions exhibit low biases on average, extremely populated modeled grid cells exhibit both column and surface overestimates. Separately, discrepancies between the two different observational datasets are unreflective of the whole modeled NO_2 concentration performance for India. Yet, combined, this analysis points to the uncertainties in the spatial allocation of existing emissions inventories used for modeling air quality in India.

This relationship in modeled NO_2 biases at urban and rural grid cells compared with OMI NO_2 tropospheric columns and largely-urban surface observations from the CPCB emphasize two things: (1) the need for better detailed spatial information for gridding anthropogenic emissions, and (2) the utility provided by using remote sensing observations for model analysis and evaluation. There are limitations to this kind of comparison. Population density is highly variable within a 36 km x 36 km grid cell, and our estimates describe urban as very highly populated grid cells when in reality there is significant variation in population density and NO_2 across an area. Another limitation is in the observational datasets. Observations from the peer-reviewed literature are meant to be spatially representative of high and low regions concentrations, as they were taken across different seasons between 2005 and 2010 and do not reflect a true temporal comparison against our CMAQ simulations. For a direct temporal comparison, the CPCB observations are more suitable, yet there are systemic issues among the collection of data including monitor reliability, human error with no regular bias correction factors known or applied, and monitor placement mostly in urban areas. Given this, an urban-rural observational analysis is able to inform modelers and emissions inventory developers of geographic variations in pollution trends that can be integrated into spatial gridding fields for emissions inventories.

3.4 Urban-Rural Influences for Other Pollutants

To determine if there are urban-rural bias differences across pollutants in addition to NO_2, we compare observations for SO_2 and O_3 at Semi- and Dense-Urban locations (Figure 6). In general, pollutant concentrations of SO_2 are lower than those of both NO_2 and O_3. Similar to NO_2, at semi-urban CPCB monitor locations (Figure 6a), CMAQ tends to overestimate both SO_2 and O_3, with positive model biases of 16.2% and 4.39% respectively. However, the opposite occurs at Dense-Urban monitor locations (Figure 6b); both modeled SO_2 and O_3 are under-estimated with respect to surface observations from the CPCB, with biases of -7.82% and -84.7% respectively, where one monitor is used for the O_3 comparison in the Dense-Urban scenario.

The comparison of surface concentrations across all urban monitors in the Semi-Urban scenario and the Dense-Urban monitors indicates the differences in concentrations in these two regions. For instance, modeled NO_2 increases substantially between the Semi-Urban (19.1 ppb) and the Dense-Urban (35.6 ppb) yet only increases slightly in the surface observations from the CPCB from 16.8 ppb to 18.5 ppb in the Semi- to Dense-Urban respectively. Modeled SO_2 concentrations increase slightly from 5.27 in the Semi-Urban to 6.94 ppb in the Dense-Urban, and are slightly high compared to the Semi-Urban observed concentrations (4.87 ppb) at the Dense-Urban monitors yet are low compared to the Dense-Urban observed concentrations (7.69 ppb). Modeled concentrations of O_3 decrease from 33.6 ppb to 19.3 ppb between Semi- and Dense-Urban areas due to modeled NO_X titration, yet the observations in these regions note a considerable increase in O_3 at Dense-Urban areas (from 32.1 ppb to 126 ppb). Discrepancies in modeled surface biases of these gas phase pollutants may indicate transport deficiencies in CMAQ at this resolution.

4. Discussion and Conclusion

To the best of our knowledge we present the first analysis of CMAQ model performance for NO_2 over India using three observational datasets measuring tropospheric VCDs from OMI and surface observations from two datasets collected from the peer-reviewed literature and the CPCB. Annual anthropogenic and monthly biogenic and biomass burning emissions combined with modeled meteorology for 2010 were used for four monthly simulations for January, April, July, and October to evaluate CMAQ's daily performance metrics under seasonally representative conditions. Model evaluation was conducted using tropospheric VCDs of NO_2 and HCHO at overpass time and with an averaging kernel applied to model data and limited ground measurements available across the domain. Guided by contradictory modeled NO_2 biases compared to our surface and tropospheric column observational datasets, we identify differences in model performance at urban and rural areas, most noticeably the underestimate of NO_2 across relatively lower population-dense rural regions (NMB_{OMI}=-63.3%) compared to very large model high biases in dense urban regions (NMB_{high_CPCB}=+131.2%), and suggest these biases result from large underestimates in rural regions of the emissions inventory.

Given inherent limitations in both emission inventories estimates and gridding proxies used for the region, model performance informs locations of regional biases. Anthropogenic emissions tend to coincide with regions of high population density or large point source emissions, regions which emit the greatest quantity of pollutants such as NO_X. Negative model bias of -63.3% against OMI NO_2 tropospheric column densities are larger than biases found comparing the DOMINO product against output from other regional and global models and ensembles (-9% to -23%) [*Huijnen et al.*, 2010]. Limitations to this analysis exist for both the model in the form of a limited number of time steps available for comparison and limitations in the spatial distribution and quantity of NO_X emissions, as well as for satellite retrievals in the form of a priori profiles used to calculate tropospheric NO_2 columns and uncertainties due to cloud fraction [*Huijnen et al.*, 2010; *Boersma et al.*, 2011], factors which contribute to the air mass factor calculations. Greatest model low biases occur in non-urban regions and parts of the industrial east.

In contrast, model biases in comparison to surface observations suggest a modeled NO_2 high bias of NMB_{CPCB}=+15.6%. Differences in average model bias when evaluated with OMI NO_2 tropospheric VCDs or surface observations arise due to spatial variations in biases. In particular, NO_2 high biases appear predominantly in and downwind of densely populated urban areas, often where there are surface monitors, and low biases occur everywhere else across the much broader rural areas. Urban-rural differences in biases have been reported before [*Huijnen et al.*, 2010; *Allen et al.*, 2012; *Kemball-Cook et al.*, 2015], where low model biases against OMI NO_2 VCDs across rural regions may result from a misrepresentation of NO_X transport [*Gilliland et al.*, 2008] or lifetimes of organic nitrates in the CB05 chemical mechanism [*Canty et al.*, 2015]. Geographic differences in model biases occur for SO_2 and O_3 as well, though in less of a coherent urban and rural sense as NO_2, suggesting emissions inventory improvements for these gas phase species and relevant precursors are needed.

Many of these discrepancies between modeled and observed concentrations exist as a result of uncertainties in emissions inventories. Emissions inventories incorporate regional combustion and activity information often at a coarse resolution, such as at the state- or district-level. Issues can arise in the gridding process when coarse data must be allocated to a higher-resolution domain. At present, spatial proxies following EDGAR v4 described in *Lamarque et al.* [2010] for individual emission sectors are used to grid emissions from GAINS, including population distribution, stack locations, and detailed emissions factors for particular combustion process. In this case, emissions tend to be allocated in highly populated regions, such as across the Indo-Gangetic Plain and in Delhi where there are more people, leading to greater ambient concentrations in this region as compared to other locally populous and polluted areas across the subcontinent including cities such as Jodhpur and Nagpur and electricity generating facilities in the east. As our results indicate, this in turn leads to lower modeled concentrations across rural regions which often remain unmonitored at the surface, making it difficult to measure pollution in the region. Model biases to satellite-derived NO_2 columns shown in this study suggest that the concentration of NO_X emissions in extremely urban environments as opposed to rural and many lower population dense urban areas may be too high in the ECLIPSE gridded emissions, pointing to possible lack of information on urban-rural distribution of modes of transportation or domestic combustion, a significant source of uncertainty among emissions inventories [*Saikawa et al.*, 2017]. Furthermore, the results demonstrate that missing or outdated information on the location of large point sources such as power or industrial plants can lead to strong local underestimation of NO_2 levels, as seen across industrial regions in eastern India.

Informed by contradictory modeled NO_2 biases between evaluation with satellite VCDs (NMB_{OMI}=-63.3%) and surface observations at urban monitoring locations (NMB_{CPCB}=+15.6%) for a population density threshold of approximately 400 people per km^2, we find that there are unique differences in model performance between our Dense Urban classification and all other grid cells, defined as exceeding a population density threshold of 5000 people per km^2. In particular, there are large negative NO_2 biases at rural locations compared in the tropospheric column (NMB_{OMI_rur}=-63.2%) and large positive NO_2 biases at surface urban areas (NMB_{CPBC_urb}=+131.2%), with similar urban and rural bias discrepancies in modeled SO_2 and O_3 compared to observations. Considering much of the domain is classified as "rural" (6574 rural grid cells to 4 Dense Urban grid cells), this estimate thereby excludes sub-grid variations in population density across urban sprawl. This analysis is limited by the coarse resolution of CMAQ at 36 km by 36 km, which can encompass a broad variety of population densities with highly varying localized effects that are diminished at most regional resolutions.

Further work to improve model performance include the recommendation of using higher resolution model simulations to differentiate across high-resolution urban and rural regions. In addition, emissions inventories allocated to a grid using region specific activity and population information, particularly for highly uncertain sectors, will lead to improved detailed for spatially distributing state or country level inventory totals. Higher levels of emissions detail will in turn support high-resolution CMAQ modeling over India where there remains limited observational coverage, research which is useful for assessing region-specific questions pertaining to air quality and related implications.

Acknowledgements

The authors would like to thank Markus Amann for project guidance and Wolfgang Schoepp for providing ancillary data. A.K. and T.H. were supported by the NASA Air Quality Applied Sciences Team (AQAST, NASA Grant #NNX11AI50G). A.K. received additional support from the Wisconsin Space Grant Consortium Graduate Research Fellowship and participated in the IIASA Young Scientists Summer Program through a grant from the National Academy of Sciences Board on International Scientific Organizations, funded by the National Science Foundation (Grant #OISE-1148655).

References

Aggarwal, P., and S. Jain (2015), Impact of air pollutants from surface transport sources on human health: A modeling and epidemiological approach, *Environ. Int.*, *83*, 146–157, doi:10.1016/j.envint.2015.06.010.

Akagi, S. K., R. J. Yokelson, C. Wiedinmyer, M. J. Alvarado, J. S. Reid, T. Karl, J. D. Crounse, and P. O. Wennberg (2011), Emission factors for open and domestic biomass burning for use in atmospheric models, *Atmos. Chem. Phys.*, *11*(9), 4039–4072, doi:10.5194/acp-11-4039-2011.

Allen, D. J., K. E. Pickering, R. W. Pinder, B. H. Henderson, K. W. Appel, and a. Prados (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*(4), 1737–1758, doi:10.5194/acp-12-1737-2012.

Amann, M. et al. (2011), Cost-effective control of air quality and greenhouse gases in Europe: Modeling and policy applications, *Environ. Model. Softw.*, *26*(12), 1489–1501, doi:10.1016/j.envsoft.2011.07.012.

Amann, M. et al. (2017), Managing future air quality in megacities: A case study for Delhi, *Atmos. Environ.*, *161*, doi:10.1016/j.atmosenv.2017.04.041.

Apte, J. S., T. W. Kirchstetter, A. H. Reich, S. J. Deshpande, G. Kaushik, A. Chel, J. D. Marshall, and W. W. Nazaroff (2011), Concentrations of fine, ultrafine, and black carbon particles in auto-rickshaws in New Delhi, India, *Atmos. Environ.*, *45*(26), 4470–4480, doi:10.1016/j.atmosenv.2011.05.028.

Boersma, K. F. et al. (2011), An improved tropospheric NO2 column retrieval algorithm for the Ozone Monitoring Instrument, *Atmos. Meas. Tech.*, *4*(9), 1905–1928, doi:10.5194/amt-4-1905-2011.

Byun, D., and K. L. Schere (2006), Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, *Appl. Mech. Rev.*, *59*(2), 51, doi:10.1115/1.2128636.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson (2015), Ozone and NOx chemistry in the eastern US: Evaluation of CMAQ/CB05 with satellite (OMI) data, *Atmos. Chem. Phys.*, *15*(19), 10965–10982, doi:10.5194/acp-15-10965-2015.

Carmichael, G. R. et al. (2009), Asian aerosols: Current and year 2030 distributions and implications to human health and regional climate change, *Environ. Sci. Technol.*, *43*, 5811–5817, doi:10.1021/es8036803.

Census-India (2012), *Census of India, 2011*, New Delhi, India.

Central Pollution Control Board (2003), Guidelines for Ambient Air Quality Monitoring, *Natl. Ambient Air Qual. Monit., NAAQMS*, 2003–4.

Central Pollution Control Board (2009), *Environmental Assessment of Industrial Clusters*.

Chambliss, S. E., R. Silva, J. J. West, M. Zeinali, and R. Minjares (2014), Estimating source-attributable health impacts of ambient fine particulate matter exposure: global premature mortality from surface transportation emissions in 2005, *Environ. Res. Lett.*, *9*(10), 104009, doi:10.1088/1748-9326/9/10/104009.

Chatani, S. et al. (2014), Photochemical roles of rapid economic growth and potential abatement strategies on tropospheric ozone over South and East Asia in 2030, *Atmos. Chem. Phys.*, *14*(17), 9259–9277, doi:10.5194/acp-14-9259-2014.

Chaudhuri, S., and D. Dutta (2014), Mann-Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models, *Environ. Monit. Assess.*, *186*(8), 4719–4742, doi:10.1007/s10661-014-3733-6.

Chowdhury, Z., M. Zheng, J. J. Schauer, R. J. Sheesley, L. G. Salmon, G. R. Cass, and A. G. Russell (2007), Speciation of ambient fine organic carbon particles and source apportionment of PM2.5 in Indian cities, *J. Geophys. Res. Atmos.*, *112*(15), D15303, doi:10.1029/2007JD008386.

Dong, X., J. S. Fu, K. Huang, and D. Tong (2015), Model development of dust emission and heterogeneous chemistry within the Community Multiscale Air Quality modeling system and its application over East

Asia, *Atmos. Chem. Phys. Discuss.*, *15*(24), 35591–35643, doi:10.5194/acpd-15-35591-2015.

Duncan, B. N. B. N., L. N. Lamsal, A. M. Thompson, Y. Yoshida, Z. Lu, D. G. Streets, M. M. Hurwitz, and K. E. Pickering (2015), A space-based, high-resolution view of notable changes in urban NOx pollution around the world (2005-2014), *J. Geophys. Res. Atmos.*, *121*(X), 976–996, doi:10.1002/2015JD024121.Received.

Eder, B., and S. Yu (2006), A performance evaluation of the 2004 release of Models-3 CMAQ, *Atmos. Environ.*, *40*(26), 4811–4824, doi:10.1016/j.atmosenv.2005.08.045.

Fioletov, V. E., C. a. McLinden, N. Krotkov, M. D. Moran, and K. Yang (2011), Estimation of SO2 emissions using OMI retrievals, *Geophys. Res. Lett.*, *38*(21), 1–5, doi:10.1029/2011GL049402.

Fu, J. S., N. C. Hsu, Y. Gao, K. Huang, C. Li, N. H. Lin, and S. C. Tsay (2012), Evaluating the influences of biomass burning during 2006 BASE-ASIA: A regional chemical transport modeling, *Atmos. Chem. Phys.*, *12*(9), 3837–3855, doi:10.5194/acp-12-3837-2012.

Garg, A., P. R. Shukla, S. Bhattacharya, and V. K. Dadhwal (2001), Sub-region (district) and sector level SO2 and NO(x) emissions for India: Assessment of inventories and mitigation flexibility, *Atmos. Environ.*, *35*(4), 703–713, doi:10.1016/S1352-2310(00)00316-2.

Garg, A., P. R. Shukla, and M. Kapshe (2006), The sectoral trends of multigas emissions inventory of India, *Atmos. Environ.*, *40*(24), 4608–4620, doi:10.1016/j.atmosenv.2006.03.045.

Ghude, S. D., G. G. Pfister, C. Jena, R. J. Van Der A, L. K. Emmons, and R. Kumar (2013), Satellite constraints of nitrogen oxide (NOx) emissions from India based on OMI observations and WRF-Chem simulations, *Geophys. Res. Lett.*, *40*(x), 423–428, doi:10.1029/2012GL053926.

Gilliland, A. B., C. Hogrefe, R. W. Pinder, J. M. Godowitch, K. L. Foley, and S. T. Rao (2008), Dynamic evaluation of regional air quality models: Assessing changes in O3 stemming from changes in emissions and meteorology, *Atmos. Environ.*, *42*(20), 5110–5123, doi:10.1016/j.atmosenv.2008.02.018.

Gulia, S., A. Kumar, and M. Khare (2015), Performance evaluation of CALPUFF and AERMOD dispersion models for air quality assessment of an industrial complex, *J. Sci. Ind. Res.*, *74*(May), 302–307.

Gupta, M., and M. Mohan (2013), Assessment of contribution to PM10 concentrations from long range transport of pollutants using WRF/Chem over a subtropical urban airshed, *Atmos. Pollut. Res.*, *4*, 405–410, doi:10.5094/apr.2013.046.

Guttikunda, S. K., and G. Calori (2013), A GIS based emissions inventory at 1 km × 1 km spatial resolution for air pollution analysis in Delhi, India, *Atmos. Environ.*, *67*, 101–111, doi:10.1016/j.atmosenv.2012.10.040.

Guttikunda, S. K., and P. Jawahar (2012), Application of SIM-air modeling tools to assess air quality in

Indian cities, *Atmos. Environ.*, *62*, 551–561, doi:10.1016/j.atmosenv.2012.08.074.

Guttikunda, S. K., and P. Jawahar (2014), Atmospheric emissions and pollution from the coal-fired thermal power plants in India, *Atmos. Environ.*, *92*, 449–460, doi:10.1016/j.atmosenv.2014.04.057.

Guttikunda, S. K., and D. Mohan (2014), Re-fueling road transport for better air quality in India, *Energy Policy*, *68*, 556–561, doi:10.1016/j.enpol.2013.12.067.

Guttikunda, S. K., R. V. Kopakka, P. Dasari, and A. W. Gertler (2013), Receptor model-based source apportionment of particulate pollution in Hyderabad, India, *Environ. Monit. Assess.*, *185*(7), 5585–5593, doi:10.1007/s10661-012-2969-2.

Guttikunda, S. K., R. Goel, and P. Pant (2014), Nature of air pollution, emission sources, and management in the Indian cities, *Atmos. Environ.*, *95*, 501–510, doi:10.1016/j.atmosenv.2014.07.006.

Harkey, M., T. Holloway, J. Oberman, and E. Scotty (2015), An evaluation of CMAQ NO2 using observed chemistry-meteorology correlations, *J. Geophys. Res. Atmos.*, (2), 1–19, doi:10.1002/2014JD022994.Received.

Hilboll, A., A. Richter, and J. P. Burrows (2017), NO2 pollution over India observed from space and the impact of rapid economic growth, and a recent decline, *Atmos. Chem. Phys. Discuss.*, *20*(2), 1–18, doi:10.5194/acp-2017-101.

Huijnen, V. et al. (2010), Comparison of OMI NO 2 tropospheric columns with an ensemble of global and European regional air quality models, *Atmos. Chem. Phys.*, *10*(7), 3273–3296, doi:10.5194/acp-10-3273-2010.

Kemball-Cook, S., G. Yarwood, J. Johnson, B. Dornblaser, and M. Estes (2015), Evaluating NOx emission inventories for regulatory air quality modeling using satellite and air quality model data, *Atmos. Environ.*, *117*, 1–8, doi:10.1016/j.atmosenv.2015.07.002.

Klimont, Z., K. Kupiainen, C. Heyes, P. Purohit, J. Cofala, P. Rafaj, J. Borken-Kleefeld, and W. Schöpp (2016), Global anthropogenic emissions of particulate matter including black carbon, *Atmos. Chem. Phys. Discuss.*, (October), 1–72, doi:10.5194/acp-2016-880.

Kumar, A., and P. Goyal (2014), Air Quality Prediction of PM10 through an Analytical Dispersion Model for Delhi, *Aerosol Air Qual. Res.*, (2009), 1487–1499, doi:10.4209/aaqr.2013.07.0236.

Lamarque, J. F. et al. (2010), Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application, *Atmos. Chem. Phys.*, *10*(15), 7017–7039, doi:10.5194/acp-10-7017-2010.

Li, M. et al. (2014), Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms, *Atmos. Chem. Phys.*, *14*(11), 5617–5638, doi:10.5194/acp-14-5617-2014.

Lim, S. S. et al. (2012), A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010, *Lancet*, *380*(9859), 2224–2260, doi:10.1016/S0140-6736(12)61766-8.

Lu, Z., and D. G. Streets (2012), Increase in NOx emissions from Indian thermal power plants during 1996 − 2010 : unit-based inventories and multi-satellite observations, *Environ. Sci. Technol.*, *46*(x), 7463–7470, doi:dx.doi.org/10.1021/es300831w.

Lu, Z., D. G. Streets, B. De Foy, and N. A. Krotkov (2013), Ozone Monitoring Instrument Observations of Interannual Increases in SO 2 Emissions from Indian Coal-Fired Power Plants during 2005 − 2012, *Environ. Sci. Technol.*

Mallik, C., and S. Lal (2014), Seasonal characteristics of SO2, NO2, and CO emissions in and around the Indo-Gangetic Plain, *Environ. Monit. Assess.*, *186*(2), 1295–1310, doi:10.1007/s10661-013-3458-y.

Mallik, C., D. Ghosh, D. Ghosh, U. Sarkar, S. Lal, and S. Venkataramani (2014), Variability of SO2, CO, and light hydrocarbons over a megacity in Eastern India: Effects of emissions and transport, *Environ. Sci. Pollut. Res.*, *21*(14), 8692–8706, doi:10.1007/s11356-014-2795-x.

Mishra, D., and P. Goyal (2015), Quantitative Assessment of the Emitted Criteria Pollutant in Delhi Urban Area, *Aerosol Air Qual. Res.*, *15*(4), 1601–1612, doi:10.4209/aaqr.2014.05.0104.

Muntaseer Billah Ibn Azkar, M. A., S. Chatani, K. Sudo, and M. Ibn Azkar (2012), Simulation of urban and regional air pollution in Bangladesh, *J. Geophys. Res. Atmos.*, *117*(D7), n/a-n/a, doi:10.1029/2011JD016509.

National Aeronautics and Space Administration (2012), *Ozone Monitoring Instrument (OMI) Data User's Guide*.

Park, S.-U. (2015), Spatial distributions of aerosol loadings and depositions in East Asia during the year 2010, *Atmos. Environ.*, *107*, 244–254, doi:10.1016/j.atmosenv.2015.02.046.

Pouliot, G., T. Pierce, W. Benjey, S. M. O. Neill, and S. A. Ferguson (2005), Wildfire Emission Modeling : Integrating BlueSky and SMOKE, *14th Int. Emiss. Invent. Conf. "Transforming Emiss. Invent. Meet. Futur. Challenges Today"*, (2001), 1–8.

Ramachandran, A., J. Pallipad, S. a. Sharma, and N. K. Jain (2013), Recent trends in tropospheric NO2 over India observed by SCIAMACHY: Identification of hot spots, *Atmos. Pollut. Res.*, *4*(4), 354–361, doi:10.5094/APR.2013.040.

Ramachandran, S. (2007), Aerosol optical depth and fine mode fraction variations deduced from Moderate Resolution Imaging Spectroradiometer (MODIS) over four urban areas in India, *J. Geophys. Res. Atmos.*, *112*(16), 1–11, doi:10.1029/2007JD008500.

Randerson, J. T., Y. Chen, G. R. Van Der Werf, B. M. Rogers, and D. C. Morton (2012), Global burned area and biomass burning emissions from small fires, *J. Geophys. Res. Biogeosciences*, *117*(4), doi:10.1029/2012JG002128.

Reddy, M. S., and C. Venkataraman (2002), Inventory of aerosol and sulphur dioxide emissions from India: I—Fossil fuel combustion, *Atmos. Environ.*, *36*(4), 677–697, doi:10.1016/S1352-2310(01)00463-0.

Saikawa, E. et al. (2017), Uncertainties in emissions estimates of greenhouse gases and air pollutants in India and their impacts on regional air quality, *Environ. Res. Lett.*, *12*, doi:10.1088/1748-9326/aa6cb4.

Saini, R., P. Singh, B. B. Awasthi, K. Kumar, and A. Taneja (2014), Ozone distributions and urban air quality during summer in Agra – a world heritage site, *Atmos. Pollut. Res.*, *5*(4), 796–804, doi:10.5094/APR.2014.089.

Sharma, S., A. Goel, D. Gupta, A. Kumar, A. Mishra, S. Kundu, S. Chatani, and Z. Klimont (2015), Emission inventory of non-methane volatile organic compounds from anthropogenic sources in India, *Atmos. Environ.*, *102*(x), 209–219, doi:10.1016/j.atmosenv.2014.11.070.

Sharma, S., S. Chatani, R. Mahtta, A. Goel, and A. Kumar (2016), Sensitivity analysis of ground level ozone in India using WRF-CMAQ models, *Atmos. Environ.*, *131*, 29–40, doi:10.1016/j.atmosenv.2016.01.036.

Simpson, D. et al. (2012), The EMEP MSC-W chemical transport model – Technical description, *Atmos. Chem. Phys.*, *12*(16), 7825–7865, doi:10.5194/acp-12-7825-2012.

Surendran, D. E., S. D. Ghude, G. Beig, L. K. Emmons, C. Jena, R. Kumar, G. G. Pfister, and D. M. Chate (2015), Air quality simulation over South Asia using Hemispheric Transport of Air Pollution version-2 (HTAP-v2) emission inventory and Model for Ozone and Related chemical Tracers (MOZART-4), *Atmos. Environ.*, *122*, 357–372, doi:10.1016/j.atmosenv.2015.08.023.

World Health Organization (2014), Burden of disease from Ambient Air Pollution for 2012, , (1), 2012–2014.

Yarwood, G., S. Rao, M. Yocke, and G. Z. Whitten (2005), *Updates to the Carbon Bond chemical mechanism: CB05 Final Report to the US EPA, RT-0400675*.

Zia ul-Haq, S. Tariq, and M. Ali (2015), Tropospheric NO 2 Trends over South Asia during the Last Decade (2004 – 2014) Using OMI Data, *Adv. Meteorol.*, *2015*(2), 1–29, doi:10.1155/2015/959284.

Table and Figure Captions

Table 1 Spatial correlations, normalized mean biases, and normalized mean errors for CMAQ and OMI NO_2 tropospheric columns annually and for January, April, July, and October monthly averages. Correlations differ across seasons due to meteorology or changes in non-anthropogenic emission inventories. A land mask has been applied to both datasets, and statistics are only take for grid cells with land cover.

Figure 1 Annual total emissions for total (a) NO_X, (b) SO_X, and (c) $PM_{2.5}$ in tons per km^2. Population density (people per km^2) is shown for comparison in (d) and overlaid with surface observation locations from the Central Pollution Control Board (CPCB) of India for 2010 (circles) and at locations from the peer-reviewed literature for 2005-2010 (triangles).

Figure 2 Meteorology from MCIP: PBL, Temperature, Total Rainfall. January (top), July (bottom), and observed precipitation for January and July from TRMM on the right.

Figure 3 (a) 4-month average (January, April, July, October) tropospheric vertical column densities of NO_2 from the Ozone Monitoring Instrument (OMI) (10^{15} molecules per cm^2); (b) 4-month average NO_2 VCDs from CMAQ, taken at OMI overpass time and processed with a vertical averaging kernel; (c) modeled surface concentrations of NO_2 overlaid with observations from the Central Pollution Control Board (hollow circles) and from the literature (hollow triangles).

Figure 4 Total valid pixel counts per domain grid cell for NO_2 tropospheric vertical column densities (VCDs) from the Ozone Monitoring Instrument (OMI) aboard the Aura satellite. OMI overpasses at about 2PM each day and retrievals can be obscured by clouds or extremely high levels of pollution. Here we show the difference in quantity of valid pixels available in January (top) and July (bottom). Note that total valid pixel counts exceed 31, the number of days in January and July, because of oversampling techniques in WHIPS to apply OMI observations to the model grid.

Figure 5 (a) Average (January, April, July, October) CMAQ NO_2 VCDs (green) in urban and rural areas compared with OMI NO_2 VCDs (blue slash) on the left, and on the right are the urban and rural surface concentration splits for annual average CMAQ NO_2 (green), observations from the CPCB (light blue) and from the peer-reviewed literature (dark blue). Urban threshold defined as greater than 400 people per km^2. Discrepancies between biases in surface and satellite observation evaluation are due to variations in modeled urban and rural concentrations. Only literature observations are categorized as rural, hence there are zero instances of CPCB or comparable CMAQ rural values. (b) Same as (a) except for a population threshold of 5,000 people per km^2.

Figure 6 Bar charts comparing concentrations in (a) semi-urban and (b) dense-urban grid cells for NO_2 (left), SO_2 (middle), and O_3 (right) from CMAQ (green), observations from the Central Pollution Control Board (CPCB) of India (light blue), and observations from the peer-reviewed literature (blue). Modeled concentrations of gas-phase species exhibit high biases compared to observations from the CPCB.

NO_2	Annual	January	April	July	October
r^2	0.63	0.65	0.68	0.39	0.55
NMB	-63.3%	-46.9%	-71.3%	-71.1%	-59.8%
NME	68.9%	65.6%	73.2%	76.0%	68.0%

Table 1

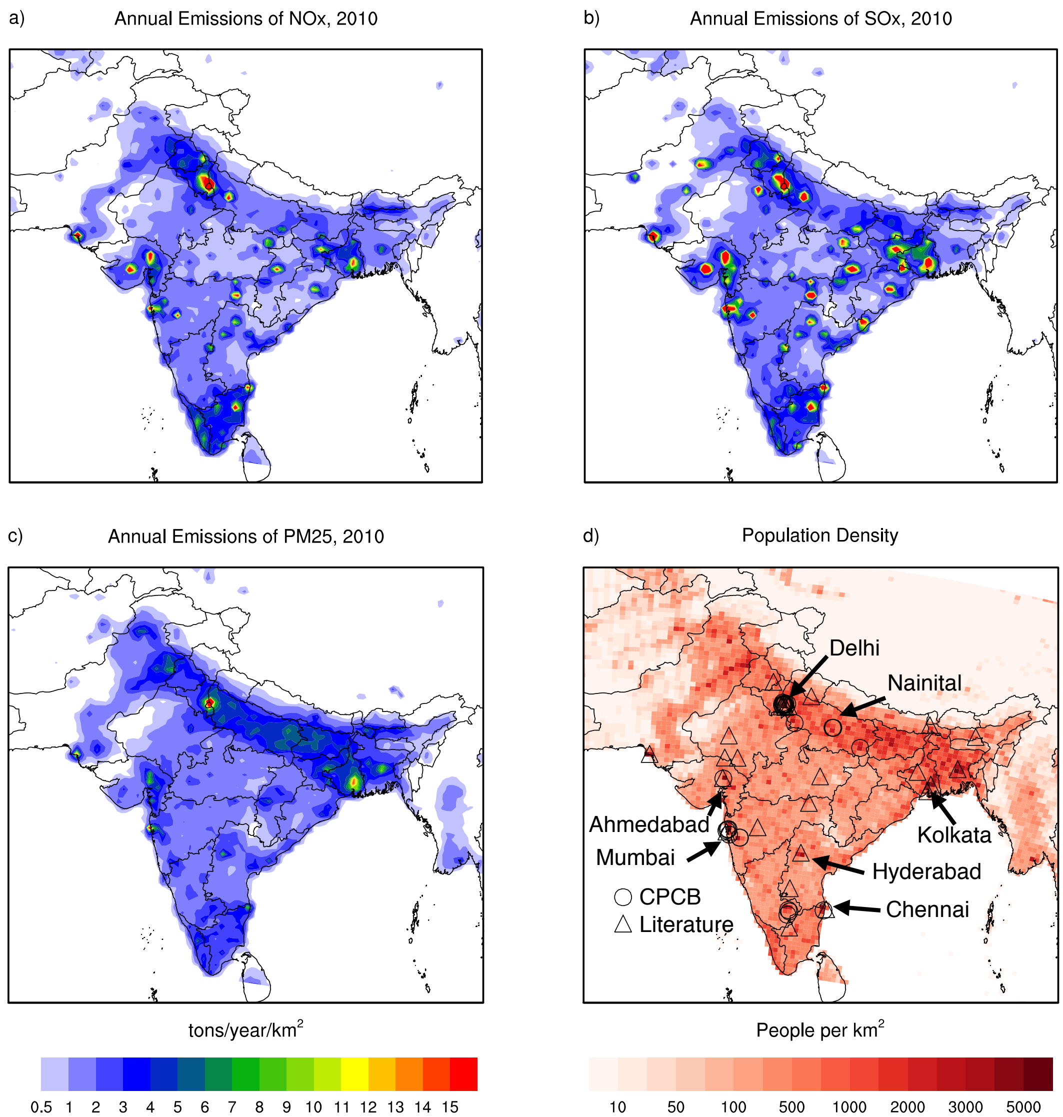

a) Annual Emissions of NOx, 2010
b) Annual Emissions of SOx, 2010
c) Annual Emissions of PM25, 2010
d) Population Density
Delhi
Nainital
Ahmedabad
Mumbai
Kolkata
Hyderabad
Chennai
CPCB
Literature
tons/year/km2
0.5 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
People per km2
10 50 100 500 1000 2000 3000 5000

Figure 1

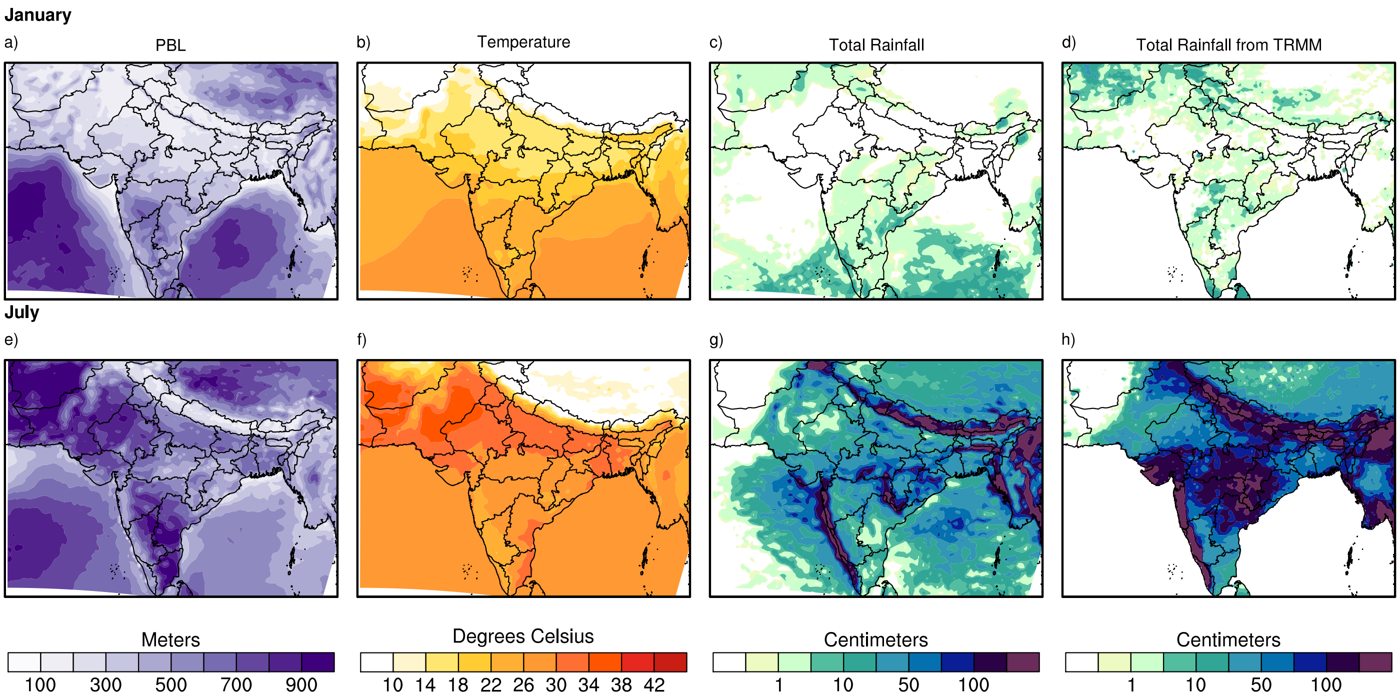

January
a)
PBL
b)
Temperature
c)
Total Rainfall
d)
Total Rainfall from TRMM
July
e)
f)
g)
h)
Meters
100 300 500 700 900
Degrees Celsius
10 14 18 22 26 30 34 38 42
Centimeters
1 10 50 100
Centimeters
1 10 50 100

Figure 2

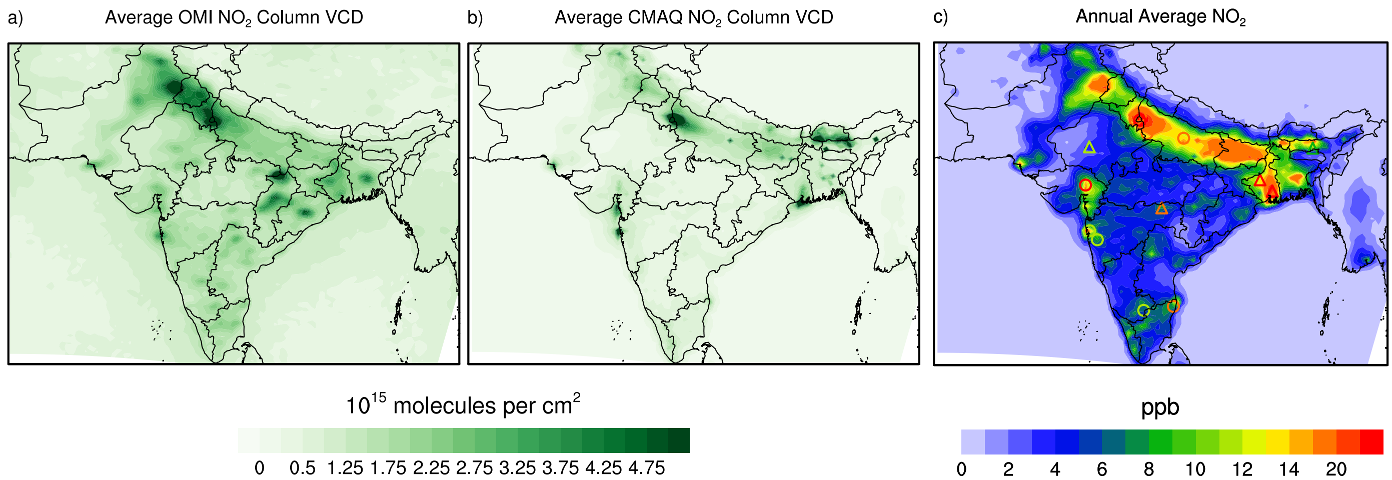

a)
Average OMI NO_2 Column VCD
b)
Average CMAQ NO_2 Column VCD
c)
Annual Average NO_2
10^{15} molecules per cm^2
0 0.5 1.25 1.75 2.25 2.75 3.25 3.75 4.25 4.75
ppb
0 2 4 6 8 10 12 14 20

Figure 3

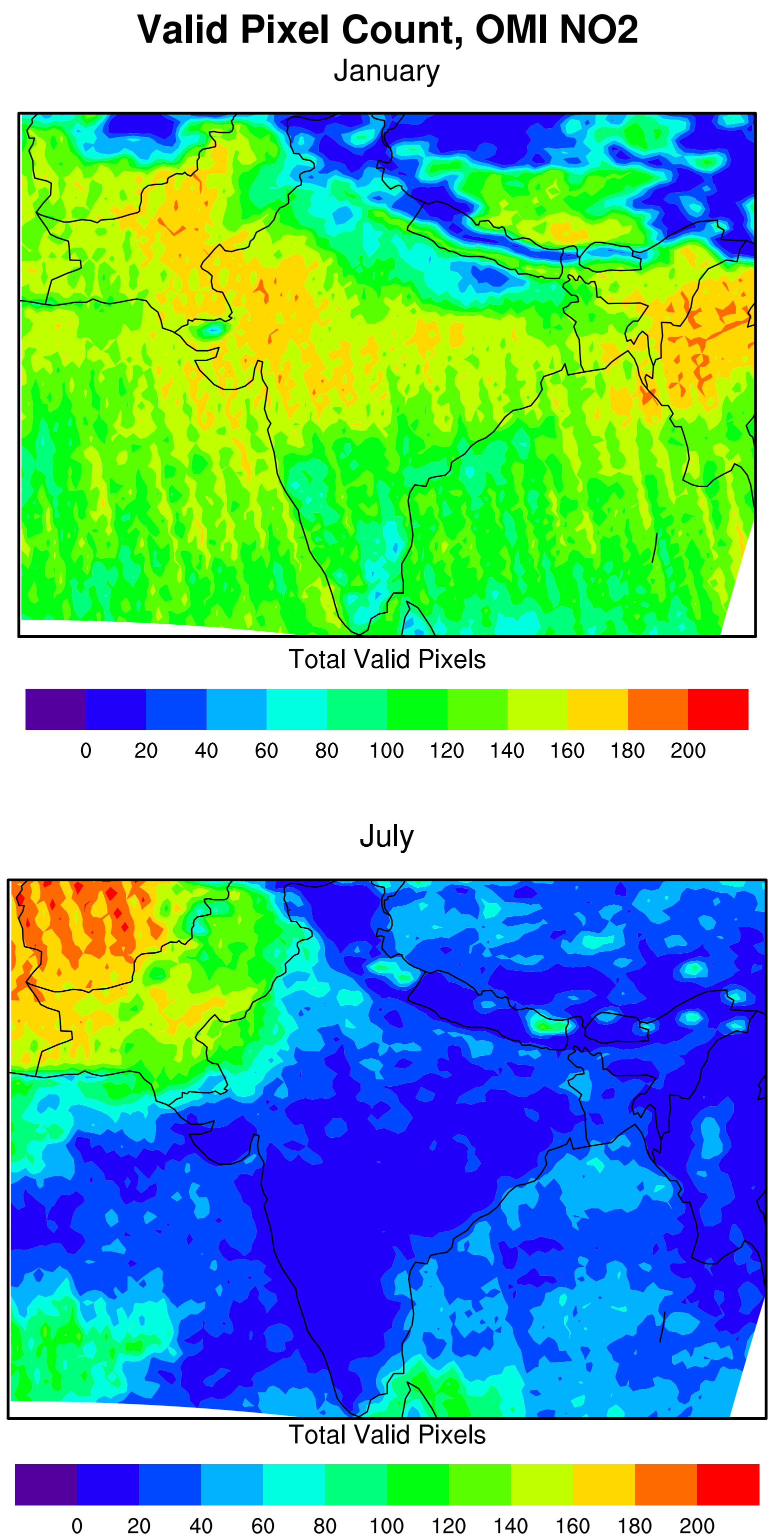
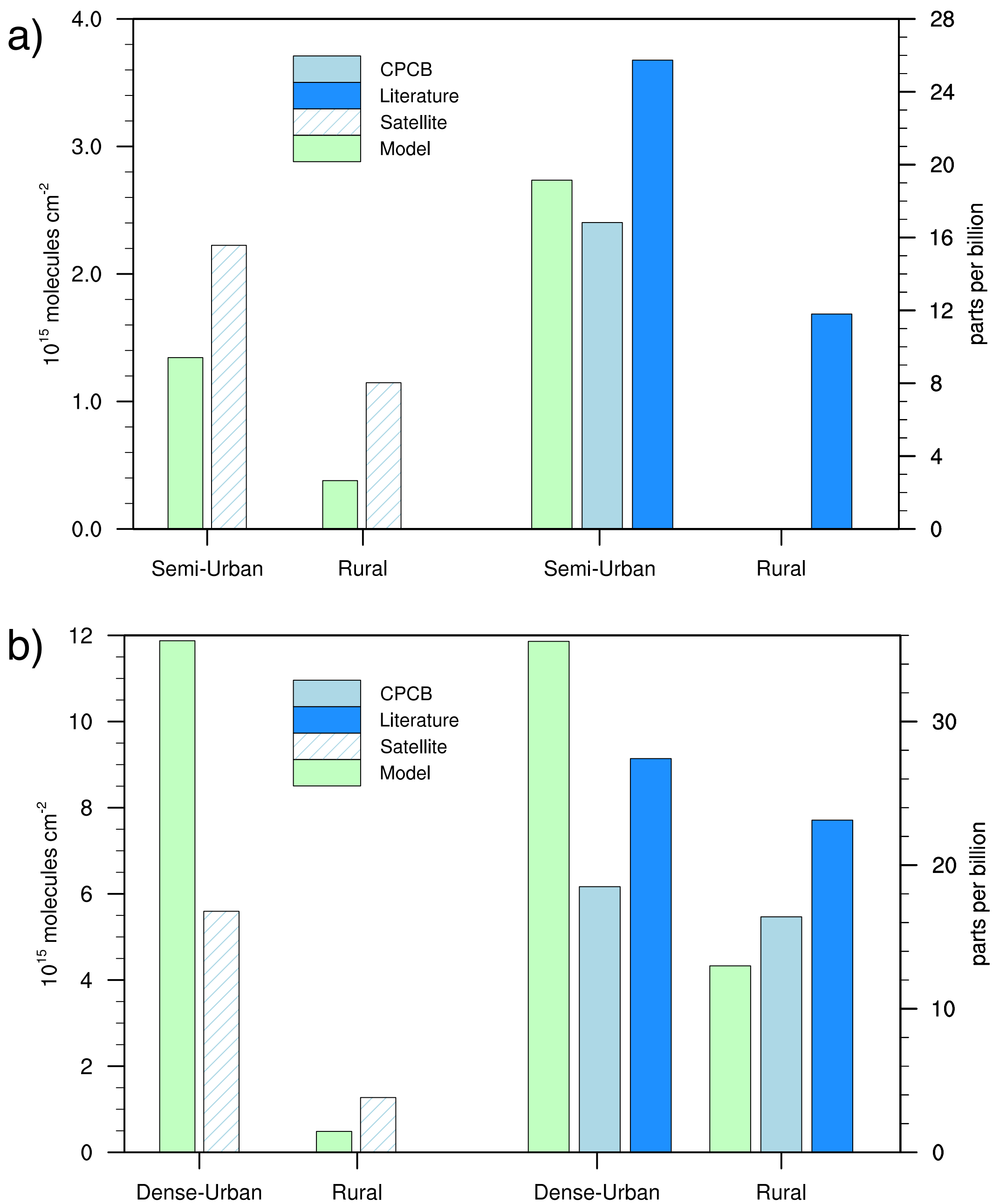

Valid Pixel Count, OMI NO2
January
Total Valid Pixels
0 20 40 60 80 100 120 140 160 180 200
July
Total Valid Pixels
0 20 40 60 80 100 120 140 160 180 200

Figure 4



a)
4.0
3.0
2.0
1.0
0.0
10^{15} molecules cm^{-2}
CPCB
Literature
Satellite
Model
28
24
20
16
12
8
4
0
parts per billion
Semi-Urban
Rural
Semi-Urban
Rural
b)
12
10
8
6
4
2
0
10^{15} molecules cm^{-2}
CPCB
Literature
Satellite
Model
30
20
10
0
parts per billion
Dense-Urban
Rural
Dense-Urban
Rural

Figure 5

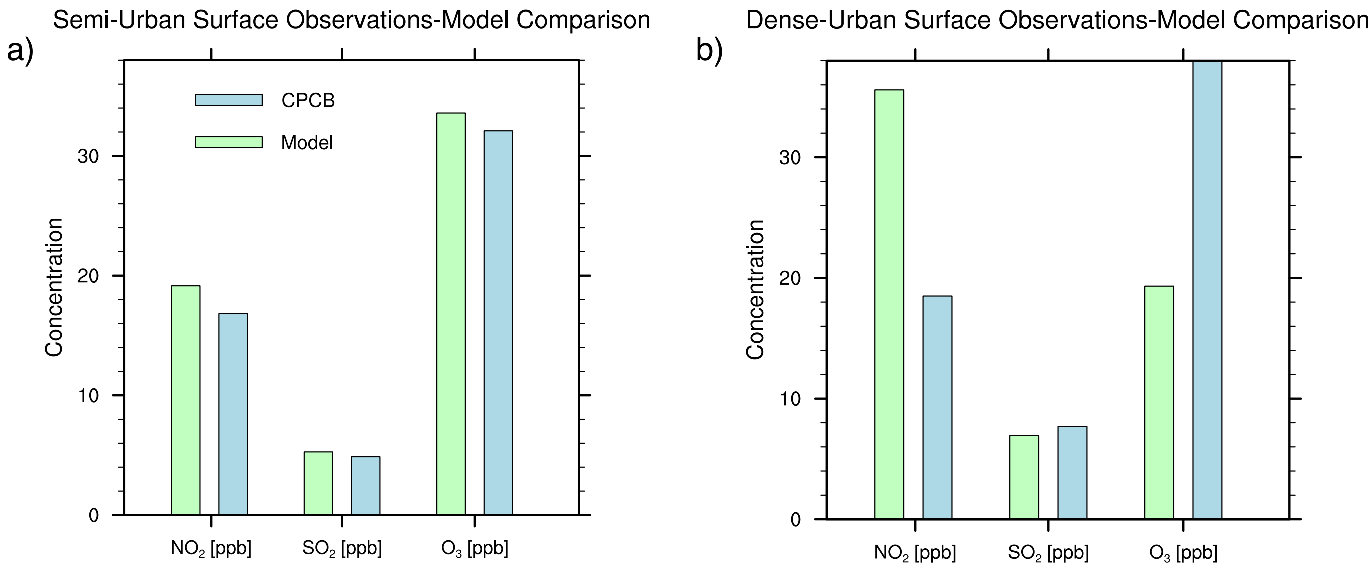

Semi-Urban Surface Observations-Model Comparison
a)
CPCB
Model
Concentration
0
10
20
30
NO2 [ppb]
SO2 [ppb]
O3 [ppb]
Dense-Urban Surface Observations-Model Comparison
b)
Concentration
0
10
20
30
NO2 [ppb]
SO2 [ppb]
O3 [ppb]

Figure 6

Highlights for the manuscript titled "Constraining the uncertainty in emissions over India with a regional air quality model evaluation," authored by Alexandra Karambelas, Tracey Holloway, Gregor Kiesewetter, and Chris Heyes.

- This is one of the earliest uses and evaluation of CMAQ for investigating India's air quality.
- Tropospheric and surface observations are used to evaluate CMAQ across urban and rural regions.
- Rural model-estimated NO_2 concentrations exhibit low biases compared to observations.
- Dense-urban regions exhibit large model high biases.
- Evaluating with OMI data exposes region-specific biases hidden by limited surface observations.