Using Space-Based Observations and Lagrangian Modeling to Evaluate Urban Carbon Dioxide Emissions in the Middle East

Emily G Yang, Eric A Kort, Dien Wu, John C Lin, Tomohiro Oda, Xinxin Ye, Thomas Lauvaux

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD031922

Using Space-Based Observations and Lagrangian Modeling to Evaluate Urban Carbon Dioxide Emissions in the Middle East

Emily G. Yang[1], Eric A. Kort[1], Dien Wu[2], John C. Lin[2], Tomohiro Oda[3,4], Xinxin Ye[5,6], and Thomas Lauvaux[5,7]

[1]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA, [2]Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT, USA, [3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [4]Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD, USA, [5]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA, [6]Now at Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA, [7]Now at Laboratoire des Sciences du Climat et de l'Environnement, CEA, CNRS, UVSQ/IPSL, Université Paris-Saclay, Orme des Merisiers, Gif-sur-Yvette, France

Key Points:
- We use satellite CO_2 observations and three global emissions inventories to assess emissions representations of five Middle Eastern cities
- Emissions estimates optimized with satellite CO_2 observations are insensitive to prior inventory spatial distributions
- Emissions estimates suggest afternoon Middle East urban inventory representations are too low

Supporting Information:
- Supporting Information S1

Correspondence to:
E. G. Yang,
egyang@umich.edu

Citation:
Yang, E. G., Kort, E. A., Wu, D., Lin, J. C., Oda, T., Ye, X., & Lauvaux, T. (2020). Using space-based observations and Lagrangian modeling to evaluate urban carbon dioxide emissions in the Middle East. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031922. https://doi.org/10.1029/2019JD031922

Received 29 OCT 2019
Accepted 18 MAR 2020
Accepted article online 24 MAR 2020

Abstract Improved observational understanding of urban CO_2 emissions, a large and dynamic global source of fossil CO_2, can provide essential insights for both carbon cycle science and mitigation decision making. Here we compare three distinct global CO_2 emissions inventory representations of urban CO_2 emissions for five Middle Eastern cities (Riyadh, Mecca, Tabuk, Jeddah, and Baghdad) and use independent satellite observations from the Orbiting Carbon Observatory-2 (OCO-2) satellite to evaluate the inventory representations of afternoon emissions. We use the column version of the Stochastic Time-Inverted Lagrangian Transport (X-STILT) model to account for atmospheric transport and link emissions to observations. We compare XCO_2 simulations with observations to determine optimum inventory scaling factors. Applying these factors, we find that the average summed emissions for all five cities are 100 MtC year^{-1} (50–151, 90% CI), which is 2.0 (1.0, 3.0) times the average prior inventory magnitudes. The total adjustment of the emissions of these cities comes out to ~7% (0%, 14%) of total Middle Eastern emissions (~700 MtC year^{-1}). We find our results to be insensitive to the prior spatial distributions in inventories of the cities' emissions, facilitating robust quantitative assessments of urban emission magnitudes without accurate high-resolution gridded inventories.

Plain Language Summary Carbon dioxide (CO_2) emitted from burning fossil fuels is the most important contributor to climate change and is changing the Earth's carbon cycle. Most of these emissions can be linked to cities. Since cities around the world are quickly changing and growing, especially in developing countries, it is important to better understand the extent of urban CO_2 emissions to understand how the climate and carbon cycle will change. In this study, we evaluate three global emissions inventories —modeled estimates of CO_2 emissions on a grid spanning the globe—and their ability to capture the afternoon emissions of five Middle Eastern cities (Riyadh, Mecca, Tabuk, Jeddah, and Baghdad). This assessment relies on comparing simulations using the inventories with observations from the Orbiting Carbon Observatory-2 (OCO-2) satellite. Based on these comparisons, we see that the inventory representations have underestimated afternoon emissions of the five studied cities and that the level of underestimation is a substantial portion of total Middle Eastern emissions. Our results are unaffected by the differing spatial patterns of emissions from different inventories. This work demonstrates the ability to use satellites to evaluate subnational emissions, a valuable advance for both science and policy issues relating to climate change and the carbon cycle.

1. Introduction

Anthropogenic carbon dioxide (CO_2) emissions are greatly changing the global carbon cycle and are a main driver of climate change. Before the industrial revolution, the CO_2 mixing ratio was 280 ppm (IPCC, 2014); since then, it has surpassed 400 ppm and continues to rise (Betts et al., 2016), mainly due to emissions from burning fossil fuels (Stocker et al., 2013). To predict future climatic trends, it is

important for us to know how each part of the carbon cycle is responding to this significant perturbation to CO_2 concentrations. However, rising uncertainties in fossil fuel CO_2 ($FFCO_2$) emissions due to overall increases in absolute emissions and increases in the proportion of emissions from developing regions with fewer constraints are limiting our understanding of the extent and implications of this large and growing perturbation.

A first step to constraining such uncertainties is to focus on cities, as the majority of $FFCO_2$ emissions derive from urban areas (International Energy Agency, 2008). These in-city emissions are mostly linked to automobile tailpipe emissions, industrial emissions, and home heating; emissions also come from electricity generation from power plants that use fossil fuels, though such emissions may take place outside of the city itself (Lin et al., 2018). Presently, more than half of the global population resides in cities (United Nations, Department of Economic and Social Affairs, Population Division, 2015). Quantifying urban emissions is also becoming more important as cities pledge to take greenhouse gas mitigation actions (Bodansky, 2016; Gurney et al., 2015; Rosenzweig et al., 2010) and require a means of assessing those actions. As the developing world rapidly urbanizes, it is becoming critical to quantify and constrain the $FFCO_2$ emissions in cities, to inform such urban mitigation policy issues, and to understand the carbon cycle science implications of the $FFCO_2$ perturbation.

Global $FFCO_2$ gridded emissions inventories are valuable tools for carbon cycle science and mitigation policies but have room for improvement with respect to local- and urban-scale representations of emissions. These emissions inventories use proxies to disaggregate national-level emissions statistics onto a fine-scale grid (Oda et al., 2019; Oda & Maksyutov, 2011)—but at finer spatial scales, the proxy approach becomes insufficient to characterize the spatial distribution of emissions sources (Oda et al., 2019). For example, Gurney et al. (2019) found city-level differences between inventories for four U.S. cities of up to about 20%. Meanwhile, in their study on different inventories for the northeastern United States, Gately and Hutyra (2017) found that more than one fourth of the grid cells in urban areas at $0.1 \times 0.1°$ resolution had relative differences of 100–300%. In that study, the researchers determined that existing global-scale $FFCO_2$ inventories are unsuitable for urban emissions monitoring, as they are unable to describe the underlying spatiotemporal patterns of the activities from which urban emissions are sourced (Gately & Hutyra, 2017). Lacking in the spatiotemporal resolution needed to capture the variability in local-level emissions and exhibiting large uncertainties, global $FFCO_2$ emissions inventories need improvement not just for local-scale issues but also global-scale ones. For example, such large uncertainties are a problem for the accuracy of global carbon budgets that are determined by considering fossil fuel CO_2 emissions to be a known quantity (Hutyra et al., 2014). Additionally, errors in the spatiotemporal distribution or magnitude of $FFCO_2$ fluxes have been shown to propagate to remaining flux components of carbon inversion studies (Gurney et al., 2005; see also Hutyra et al., 2014). Uncertainties in $FFCO_2$ emissions in global inventory representations have also been shown to be a dominating factor in global CO_2 atmospheric inversion model spread and to limit the assessment of regional scale terrestrial fluxes and ocean-land partitioning (Gaubert et al., 2019). For the sake of fine-resolution emissions inventories to be useful at many scales, it is thus a critical task to evaluate and improve the inventory models, and to optimize model representations of smaller scale emissions.

Top-down, space-based observations of CO_2 concentrations are a strong contender for evaluating emissions inventories at a local scale, especially in regions with few or no on-ground observations. In recent years, space-based technological advancements have been evolving the study of urban- and local-scale emissions. Satellites have opened the doors to global, high-resolution measurements of XCO_2, the column-averaged dry-air mole fraction of CO_2. A number of satellites observing XCO_2 have already been launched, such as the Japanese Greenhouse Gases Observing SATellite (GOSAT) and its successor GOSAT-2, the American Orbiting Carbon Observatory-2 (OCO-2) and its successor (OCO-3), and the Chinese TanSat. A number of studies have already been conducted that have utilized the observations of some of these satellites to better understand local sources of CO_2 such as large cities and power plants (Hakkarainen et al., 2016; Hedelius et al., 2018; Kort et al., 2012; Nassar et al., 2017; Wu et al., 2018). That body of work—and in particular that of Wu et al. (2018), which presents a method of extracting urban XCO_2 signals from OCO-2 and evaluates this method with a case study of Riyadh, Saudi Arabia—serves as the foundation for the work of this present study.

In this study, we evaluate global $FFCO_2$ emissions inventory representations of urban CO_2 emissions in the Middle East and assess the ability of satellite observations to inform this evaluation. Middle Eastern cities serve as optimal target study domains. Limited cloud cover increases data density, and reduced biospheric signals simplifies the analysis. Additionally, these cities provide a good example of locations with few ground-based measurements making space-based approaches the only observation-based method that could be applied. To complete our assessment, we focus on three global $FFCO_2$ emissions inventories: the Fossil Fuel Data Assimilation System (FFDAS), the Open-source Data Inventory for Anthropogenic CO_2 (ODIAC), and the Emission Database for Global Atmospheric Research (EDGAR). We examine the differences in the spatial distribution and magnitudes of these inventories at the urban scale for cities in the Middle East, focusing on Riyadh, Saudi Arabia, but also analyzing Saudi Arabian cities Mecca, Tabuk, and Jeddah, as well as the city of Baghdad, Iraq. We next quantify the relationship between observations of XCO_2 from the OCO-2 satellite and simulations of XCO_2 using the column version of the Stochastic Time-Inverted Lagrangian Transport (X-STILT) model coupled with Global Data Assimilation System (GDAS) reanalysis products and the three inventories. Using this comparison between top-down observations and bottom-up simulations of XCO_2, we calculate scaling factors to adjust the flux magnitudes of the inventory estimates of city emissions. We then discuss the implications of our estimates on a regional scale, and on the ability of space-based observations to quantify urban emissions.

2. Data and Methods

2.1. Emissions Data and Methods

2.1.1. Global $FFCO_2$ Emissions Inventories

In this study, we evaluate three of the most widely used global $FFCO_2$ emissions inventories: FFDAS, ODIAC, and EDGAR. Figure 1 shows the three inventories' representations of the Middle East. Each of these inventories is formed using an approach that combines top-down elements—distribution of emissions in space by using spatial proxies like population or nighttime lights—with bottom-up elements—summing up individual fuel consumption or emissions sources to estimate total emissions at a larger scale (Hutchins et al., 2017). However, the inventories are all gridded using different mechanisms, with variations in energy statistics used, sectors included, and methods of downscaling emissions. These differences contribute to notable discrepancies between the inventories' aggregated emissions totals and spatial distributions. Table 1 summarizes the key information about the inventories.

FFDAS is a global product with a spatial resolution of $0.1 \times 0.1°$. This inventory downscales national emissions statistics reported by the International Energy Agency (IEA), distributing emissions by constraining the Kaya identity, which uses population, GDP, energy intensity, and carbon intensity as multiplicative factors to determine emissions (Asefi-Najafabady et al., 2014; Gately & Hutyra, 2017; Hutchins et al., 2017; Rayner et al., 2010). Constraints are placed on the Kaya identity using satellite nightlights, population density information, and power plant data (publicly disclosed or from the World Electric Power Plants [WEPP] database). Sectors that are included in FFDAS are based on sectors from which IEA produces emissions statistics, including energy generation, industrial, and transportation (land transport native to FFDAS, aviation, and shipping added in from EDGAR), as well as other sectors such as residential, commercial, agriculture, and fishing (Asefi-Najafabady et al., 2014; Hutchins et al., 2017; Rayner et al., 2010). Emissions due to cement production and gas flaring are excluded from this product (Asefi-Najafabady et al., 2014). In this study, we use the 2014b (beta) data set, an early version of the FFDAS v2.0 data set. This data set is at both hourly and yearly temporal resolutions and provides emissions data for the year 2014. For 2014, total global emissions in this set come out to 8.9 PgC.

ODIAC is a monthly global product that distributes $FFCO_2$ emissions with a spatial resolution of approximately 1×1 km ($0.008333 \times 0.008333°$) (Oda et al., 2018). In the ODIAC2017 version, the inventory begins with national emissions estimates separated by fuel type (liquid, gas, solid, cement production, gas flaring, and international aviation and marine bunkers) as opposed to emissions sectors, from the Carbon Dioxide Information Analysis Center (CDIAC) at the Oak Ridge National Laboratory (ORNL). These emissions data are then re-categorized into the following categories: point source, nonpoint source, cement production, gas flare, and international aviation and marine bunkers. For the more recent years that are not included in the CDIAC estimates (2015–2016), the 2014 CDIAC emissions estimates are extrapolated using more recent BP

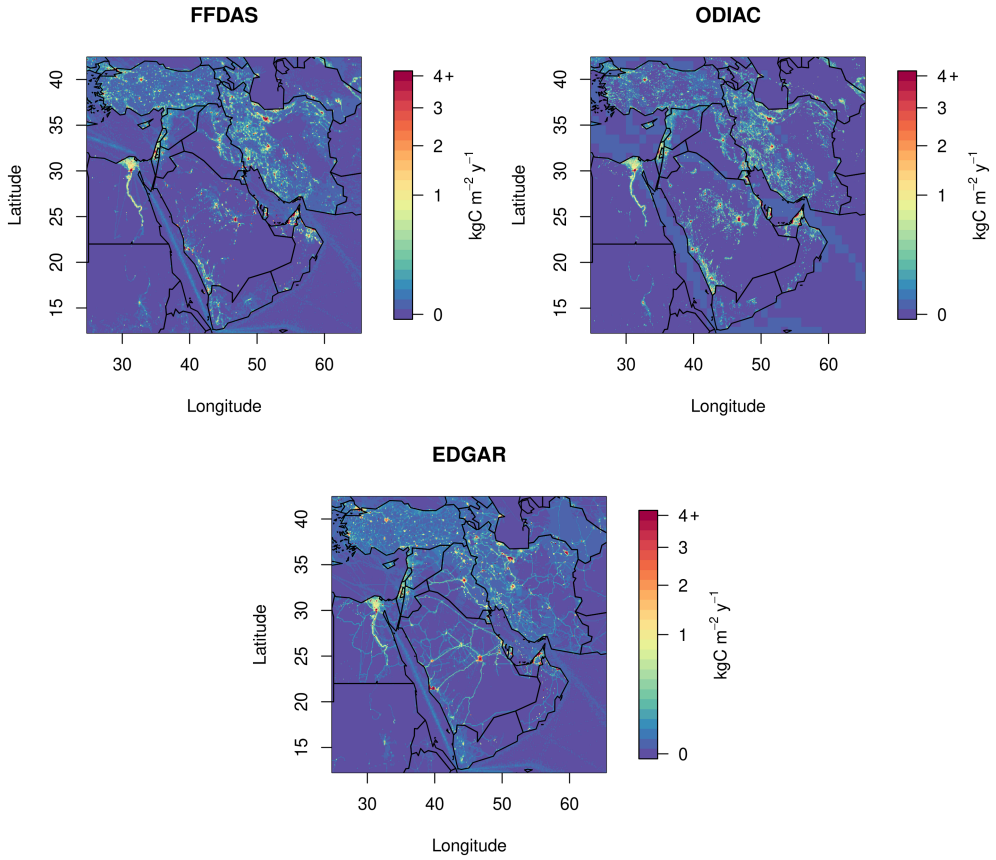

Figure 1. Global fossil fuel CO_2 emissions inventory representations of the Middle East: FFDAS, ODIAC, and EDGAR, shown with a square-root scale. The three representations differ in both spatial distribution and magnitude of emissions. Note that all inventories are shown at their native resolutions, with ODIAC having land emissions at a resolution of 1 × 1 km and international aviation and marine bunker emissions at 1 × 1°.

global fuel statistical data (BP, 2017; Oda et al., 2018). These emissions statistics are then spatially distributed with multiple spatial proxies such as the Carbon Monitoring for Action (CARMA) database for point sources, nighttime light data collected by Defense Meteorological Satellite Program (DMSP) satellites for nonpoint sources, and ship and aircraft fleet tracks. The emissions are also distributed temporally using

Table 1
Key Information About the Global Gridded $FFCO_2$ Emissions Inventories Used in This Study

Version	FFDAS 2014b (beta)	ODIAC ODIAC2017	EDGAR 4.3.2
Year(s) used	2014	2014–2016	2012
Resolution	0.1× 0.1° hourly/annually	1 × 1 km monthly	0.1 × 0.1° annually
Global total	8.9 PgC year^{-1}	9.9 PgC year^{-1}	9.5 PgC year^{-1}
Middle Eastern total	697 MtC year^{-1}	789 MtC year^{-1}	722 MtC year^{-1}
Sectors or categories included	IEA sectors: energy generation, manufacturing, industrial, transportation, and others including residential, commercial, agriculture, and fishing	CDIAC fuel types (liquid, gas, solid, cement production, gas flaring, and international aviation and marine bunkers); re-categorized as point source, nonpoint source, cement production, gas flare, and international aviation and marine bunkers	IPCC sectors: energy, fugitive, industrial processes, solvents and products use, agriculture, waste, and other (emissions due to fossil fuel fires)

seasonality derived from the CDIAC monthly gridded emissions, resulting in a monthly gridded product (Oda et al., 2018). We note that ODIAC makes available at a resolution of 1 × 1° international aviation and marine bunkers emissions information, which is recorded by CDIAC/ORNL but not included in CDIAC's gridded emissions data products. In this study, we use the ODIAC2017 version for the years 2014–2016 (Oda & Maksyutov, 2015), summing the 1 × 1 km land emissions data set with the 1 × 1° international aviation and marine bunkers data. For the year 2014, total global emissions in this data set are 9.9 PgC.

EDGAR is an annual global emissions inventory with a resolution of 0.1 × 0.1°, and it relies on international activity data provided by sources like the IEA and emissions factors to determine country-specific CO_2 emissions (Crippa et al., 2018; European Commission Joint Research Centre, 2017; Janssens-Maenhout et al., 2019; Olivier & Janssens-Maenhout, 2015). National sector totals for emissions are then distributed using a number of different proxies, including location of energy and manufacturing facilities, road networks, shipping routes, population density, and agricultural land use (European Commission Joint Research Centre, 2017; Janssens-Maenhout et al., 2013). EDGAR includes emissions sectors as defined by the Intergovernmental Panel on Climate Change (IPCC): energy (including international aviation and shipping emissions), fugitive, industrial processes (including cement production), solvents and products use, agriculture, waste, and other (emissions due to fossil fuel fires) (Janssens-Maenhout et al., 2019). Gridded $FFCO_2$ emissions for version 4.3.2 are available for years 1970–2012 (Janssens-Maenhout et al., 2019); we use only the most recent year (2012) in this study. For 2012, the total global emissions magnitude for this data set is 9.5 PgC.

2.1.2. Inventory Resolutions, Years, and Domains

Direct comparisons between the three inventories necessitate the inventories to share common spatiotemporal dimensions. For such direct comparisons, we sum the 2014 ODIAC monthly emissions to form a single annual gridded product and aggregate it from its native spatial resolution to 0.1 × 0.1°. We also use the annual 2014 FFDAS data set and the 2012 EDGAR data set. For comparisons with observations, we use ODIAC both in its native resolution and in the aggregated resolution. While the years do not match perfectly, we choose 2014 for ODIAC and FFDAS due to the availability of both inventory and OCO-2 data. We choose to match FFDAS and ODIAC and use the most recent year (2012) for EDGAR, under the assumption that EDGAR's emissions representations would not change in distribution much between the 2 years, with mostly changes in magnitude. Regarding the change in magnitude from 2012 to 2014, there is some uncertainty; while Friedlingstein et al. (2014) find that global CO_2 emissions grow 2.5% per year on average, the developing world may experience a different rate of emissions growth. However, the growth rate in population for the cities of interest are approximately 2–4% per year (for 2004 to 2010 for Saudi Arabian cities and 2009 to 2018 for Baghdad) (Brinkhoff, 2018), suggesting small emissions growth as well. These changes are expected to have minimal effects on the results of this study.

We choose separate domains of interest for each studied city: square domains that range from 0.4 × 0.4° in size to 0.7 × 0.7° to encompass the cities of interest (Figure 2). For each city, a first-pass domain is chosen by examining Google Earth to approximate the spatial boundaries of the cities. After an approximate square domain is chosen, a sensitivity analysis is done using the emissions inventories, either by reducing or expanding the domain by increments of 0.1° on each side of the domain to determine at which size the total emissions within the chosen domain changes the least.

Note that the chosen domains do not correspond exactly to the administrative boundaries of the cities themselves. Emissions sources such as nearby suburbs that would not necessarily be part of the cities themselves could be included. However, given that it would be quite difficult to distinguish OCO-2 XCO_2 enhancements as having come from either a city or its neighboring suburb, we choose to set the boundaries of the domains as squares that hold the greatest emissions relative to the rest of the neighboring area.

2.1.3. Analytical Methods for Direct Comparisons Between Inventories

To evaluate the differing urban representations of the emissions as modeled by the inventories, we conduct direct comparisons between them using methods to discriminate between the magnitudes and the spatial distributions of the inventories. Some of these methods were also conducted by Hutchins et al. (2017) in their study focusing on the continental United States and Gately and Hutyra (2017) in their study of the northeast United States. We sum up the emissions over the domains of interest and compare their magnitudes. The emissions of the three inventories are plotted over each domain of interest on a linear scale and compared

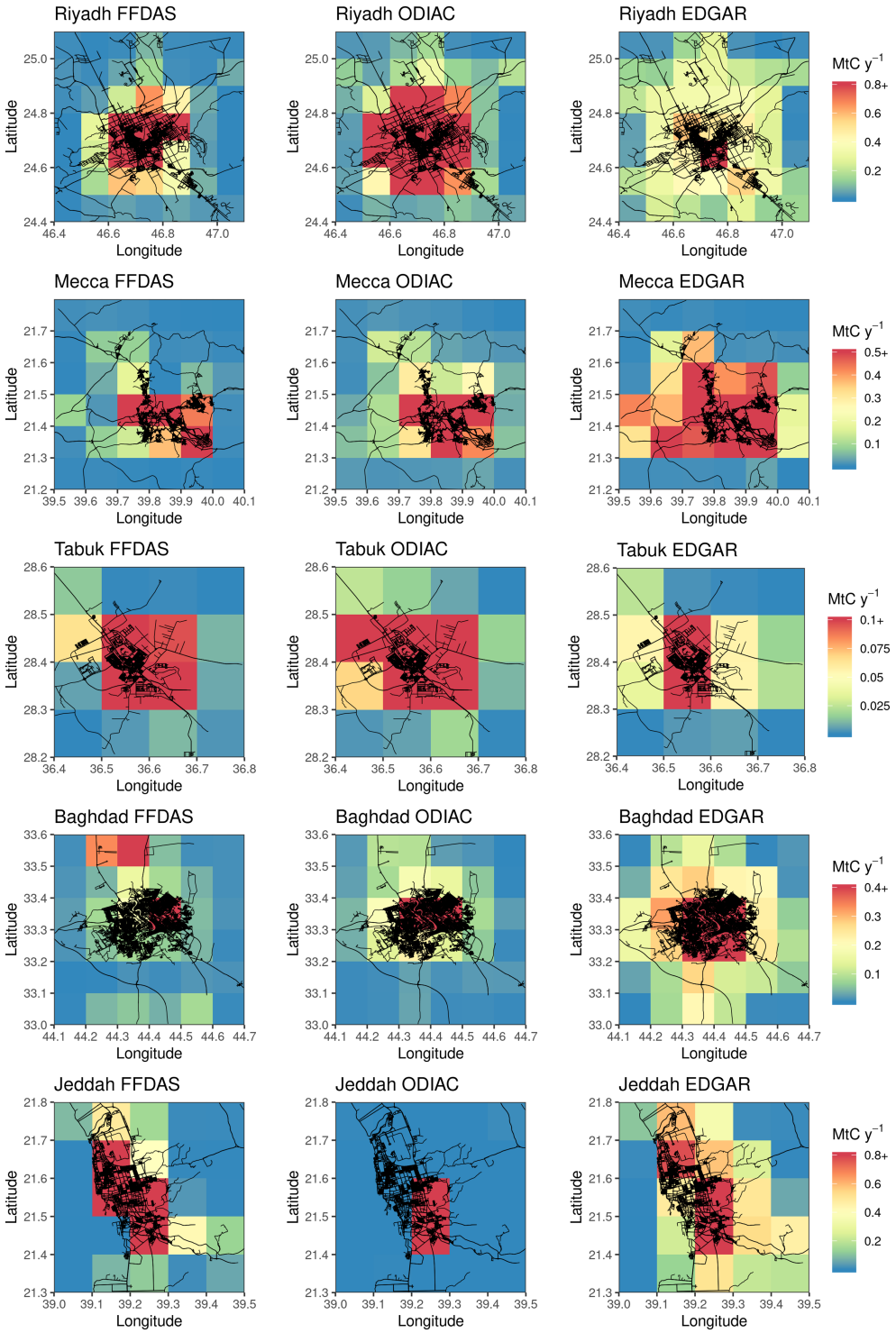

Figure 2. Emissions representations of each city of interest for each inventory at a spatial resolution of 0.1 × 0.1° (colors) and roads in that domain (black). At the urban scale, these representations show more clear differences in the spatial distribution and magnitudes of the emissions than at the regional scale. Note that the ODIAC representation for Jeddah has an error due to a mismatch between two gas flare nightlight data sources at that particular location (see Text S2); we proceed in this paper by treating it as though there is no error to understand how our methods handle the mismatch.

visually. The correlation between maps is also determined by finding the Kendall rank correlation coefficients between pairs of inventory representations of each city. We also generate cumulative emissions curves by arranging emissions for each inventory in ascending order, taking the cumulative sums of the emissions, and plotting these cumulative sums against each other.

2.2. Atmospheric Data and Methods

2.2.1. OCO-2 XCO_2 Data and Preprocessing

We use bias-corrected OCO-2 Lite, version 8, data from September 2014 to April 2017 (O'Dell et al., 2018). Overpasses are initially chosen by proximity to the cities of interest; the original net cast for overpasses is quite wide for exploratory purposes, such that any observations within an $8 \times 8°$ box surrounding the location coordinates of the city of interest are selected. These are further filtered by location of observations, to ensure that there are observations present near the city center (at least within 50–75 km away) as well as further out for the purposes of background calculation (within 200–350 km away). Then, overpasses are screened visually for sufficient number, distribution (no or few gaps), and density of observations. The data are then preprocessed by finding the medians of the up to eight cross-track footprints at each measurement time step, so as to remove outliers. Then a block average of the medians is taken in bins of 0.1° by latitude. Most of the overpasses are observed in Nadir Mode, though seven are in Glint Mode. For Riyadh, Mecca, Tabuk, Baghdad, and Jeddah, we use 11, 6, 3, 3, and 3 overpasses, respectively. The OCO-2 data are not filtered by warn level (see supporting information Text S1 for details). Table S1 details the 26 overpasses used in this study, and Figure S1 maps the overpasses in relation to the cities.

2.2.2. Simulations of XCO_2

We simulate XCO_2 with the inventories such that the simulated XCO_2 is directly comparable to the values observed by OCO-2; in this way we can quantify the relationship between the observations and the models. Most of the steps used to simulate XCO_2 follow the methodology of the "X-STILT" approach described by Wu et al. (2018), which utilizes the STILT model to simulate urban XCO_2 values. Here we summarize our process to simulate XCO_2, noting any deviations in the methodology from the X-STILT formulation.

2.2.2.1. Summary of Approach

We simulate XCO_2 values along the OCO-2 tracks of interest at a resolution of 0.1° by latitude. The basic process of these simulations at each along-track location involves using GDAS reanalysis products at $0.5 \times 0.5°$ resolution to drive the STILT model, which releases air parcels from each of n prescribed column receptor locations backward in time and calculates a "footprint." Footprints are a measure of the sensitivity of the mixing ratio to surface fluxes and are an indicator of the upstream influences of the chosen receptor locations; they are in units of ppm/(μmol m^{-2} s^{-1}). The footprints are then convolved with fluxes given by the three emissions inventories to determine an enhancement at that column receptor location. Combining these enhancements with a background value and a biospheric enhancement gives a full, simulated concentration $CO_{2.sim,n}$ at the nth level in the column (section 2.2.2.2). These concentrations are then weighted with the satellite's averaging kernel profiles to determine a single column value of XCO_2 ($XCO_{2.sim.ak}$) that is directly comparable to an OCO-2 XCO_2 value (section 2.2.2.3).

2.2.2.2. Simulating CO_2 Concentrations at Each nth Level

To reduce computational time, we differentiate between simulating CO_2 concentrations at each nth level either above a height designated as *MAXAGL*, or below or at this height. We define *MAXAGL* to be the maximum release height of air parcels with STILT. For this study, we use a *MAXAGL* of 6 km for wintertime overpasses, following Wu et al. (2018), and 10 km for non-wintertime overpasses.

At each level at or below *MAXAGL*, we simulate a value of $CO_{2.sim,n}$ by following the basic approach described in section 2.2.2.1. Using STILT, we release 100 air parcels at each column receptor 48 hr backward in time, with our column receptors located at every 100 m up from the surface to 3 km for wintertime overpasses (5 km for non-winter), then every 500 m up to the 6 km. Non-wintertime overpasses additionally have receptors located every 1 km up to the 10 km *MAXAGL*. This results in a footprint for each column receptor to be convolved with the emissions inventories to produce enhancements.

To capture the temporal variations of emissions, we use the Temporal Improvements for Modeling Emissions by Scaling (TIMES) scaling factors in combination with the inventory fluxes (Nassar et al., 2013). The TIMES scaling factors can be combined with monthly $FFCO_2$ emissions inventories to account for both

diurnal and within-week variability, with 24 hourly grids capturing the diurnal variations and seven daily grids capturing the weekly variations.

The STILT model simulations span several years to match observations from late 2014 to mid-2017, whereas the emissions inventories do not have the same availability. Since FFDAS has an hourly temporal resolution for just the year 2014, we aggregate the data up to monthly emissions by day of week and hour, such that we have 24 grids per day of week in each month to maintain both the diurnal and weekly cycles internal to the 2014 data set. (Note that these internal diurnal and weekly cycles are derived through application of the TIMES scaling factors.) For ODIAC, we use years 2014–2016 to correspond with the appropriate years' simulations and use the 2016 grids for the 2017 simulations. As these are monthly grids, we use both the diurnal and weekly TIMES scaling factors along with the ODIAC data sets, which we keep in their native spatial resolutions of 1 × 1 km and also aggregate up to 0.1 × 0.1°. For EDGAR, the most recent year available is 2012; we use this annual file along with both the diurnal and weekly TIMES scaling factors. Each of these adjusted emissions inventories are then convolved with footprints to find the fossil fuel contribution to $CO_{2.sim,n}$, or the enhancements.

We also simulate the natural contribution to $CO_{2.sim,n}$ by using biospheric flux data from CarbonTracker Near-Real Time (CT-NRT) v2016 and v2017. These biospheric fluxes are convolved with footprints in a similar fashion to the inventory $FFCO_2$ fluxes. Fluxes from oceanic and fire-related sources are deemed negligible in comparison to the anthropogenic and biospheric fluxes for this domain and not included.

The boundary conditions to add to the natural and fossil fuel enhancements are determined using the trajectory endpoint technique. CO_2 concentrations sourced from CT-NRT 3-D mole fractions of CO_2 that correspond with the endpoints of the STILT trajectories serve as the boundary conditions. These are then combined with the enhancements to find the full $CO_{2.sim,n}$ value at each level. This technique differs from that suggested by Wu et al. (2018). Above *MAXAGL*, each simulated level *n* is defined by the OCO-2 retrieval levels. $CO_{2.sim,n}$ is found by finding the CT-NRT 3-D mole fractions of CO_2 corresponding to the OCO-2 levels.

2.2.2.3. Simulating XCO_2 to be Directly Comparable to OCO-2 Observations

The OCO-2 sensor has sensitivities to CO_2 that are different at varying heights of the atmosphere, which are characterized by the satellite's averaging kernel profiles. OCO-2 thus retrieves XCO_2 values by using these averaging kernel profiles, which are the product of the satellite's normalized averaging kernel profiles (AK_{norm}) and pressure weighting function (*PW*), to find the relative weight between the observed "true" profile and the a priori profile ($CO_{2,ap}$) (Wu et al., 2018). To make a direct, 1:1 comparison between a simulated profile and the OCO-2 retrieved profile, we thus use the same OCO-2 weighting functions, with the simulated, unweighted column serving as the "true" profile. Our simulated profile weighted with the averaging kernel, $XCO_{2.sim.ak}$, is thus determined as follows, adapted from O'Dell et al. (2012) and reproduced from Wu et al. (2018):

$$XCO_{2.sim.ak} = \sum_{n=1}^{nlevel} \left\{ AK_{norm,n} \times PW_n \times CO_{2.sim,n} + \left\{ I - AK_{norm,n} \right\} \times PW_n \times CO_{2.ap,n} \right\}, \quad (1)$$

where *I* is the identity vector and *n* is the X-STILT release level. Since the X-STILT levels do not match the 20 levels prescribed by OCO-2, we linearly interpolate AK_{norm}, *PW*, and $CO_{2.ap}$ values from the OCO-2 levels to the X-STILT levels. Note that these values undergo the preprocessing described in section 2.2.1.

2.2.3. Further Data Filtering

After the overpasses are simulated with our X-STILT model, they are filtered by leveraging the STILT output to calculate footprint (influence) values from the city of interest; they pass through the filter if any modeled point at 100 m AGL in the simulated overpass pass a threshold for footprint values, an average of 0.01 ppm/(μmol m^{-2} s^{-1}). In this way, overpasses only pass the filter and are analyzed if they indicate influence from within the city of interest.

Footprint-filtered overpasses are then further filtered by returning to the observations. These remaining overpasses are filtered by their viability for differentiating enhancements from what we refer to as the tail values, which are XCO_2 values that are not enhanced and help define the background specific to each overpass when averaged. In part, this differentiation is done through visual inspection, as well

as by selecting overpasses whose enhanced values on average exceeded the tail values by at least 0.2 ppm.

2.2.4. Determining Directly Comparable OCO-2 and Simulated Enhancements

Our interest in both the simulations and the observations of OCO-2 are the enhancements due to fossil fuel combustion within the cities of interest. Thus, our main points of comparison are between the simulated fossil enhancements and observed enhancements, which are determined by subtracting the simulated boundary conditions and natural contributions of XCO_2 from both the OCO-2 observations ($XCO_{2.obs}$) and the fully simulated $XCO_{2.sim}$ from section 2.2.2.3. (Note that subtracting the boundary conditions and natural contributions from $XCO_{2.sim}$ is not exactly the same as solely convolving the footprints with the emissions inventory fluxes; such a convolution ignores the contribution to $XCO_{2.sim}$ from the averaging kernel and a priori profiles as described in section 2.2.2.3.)

An additional constant value that corrects the bias between the CarbonTracker-derived background and the observed background is also either added to or subtracted from the full XCO_2 values to determine final enhancements. This bias correction term is determined as follows. For Riyadh, a swath of 3° of latitudinal points for each observed overpass of interest is examined, differentiating between the enhanced points and the other points, the tails of the overpass, by means of expert judgment and comparison with the simulated enhancements from the city of interest. The other cities undergo a similar process, except not all overpasses use 3° of data. The other cities have enhancements from other nearby cities, artifacts to be avoided, or their own large-scale variability, so the length of the tails may be shorter for non-Riyadh overpasses to only capture the local background. Next, for all cities, the tails are averaged to a single constant. This constant serves as a bias correction term between the CarbonTracker-derived background and the OCO-2 observed background.

This methodology for determining the background differs from that described in Wu et al. (2018). However, we choose to use this methodology for two reasons: computation time savings and the ability to capture upwind variability using CarbonTracker. Regarding the former reason, Wu et al. (2018) opts for a method that requires a plume defined through forward-modeling with STILT that is computationally expensive, especially given the number of overpasses analyzed in this study. Regarding the latter reason, we use CarbonTracker to define our boundary conditions in this study to fully capture variability upwind of the urban plume. Any biases between the OCO-2 observations and the CarbonTracker-derived background are then corrected for using observations with little local influence. In this way, we account for large-scale variation in XCO_2, while also correcting biases using the observations.

2.2.5. Quantifying the Relationship Between the Modeled and Observed Enhancements

We evaluate the relationship between the simulated and observed enhancements for each overpass of interest by generating cumulative enhancement curves, which is a modified area-under-the-curve technique. In this technique, the data points that are enhanced for each day or in aggregate are placed in ascending order, and their cumulative sums are plotted against each other, in a similar fashion as to the cumulative emissions curves described in section 2.1.3. This method allows for the examination of the spatial distribution of enhancements while simultaneously latitudinally integrating the areas under the curves generated by the enhancements. Each modeled curve is also scaled such that the total summed enhancements for that particular overpass was equivalent to that of the corresponding OCO-2 sum, to make scaled cumulative enhancement plots.

The scaling factors used to scale the cumulative enhancement plots are combined using a bootstrap method to determine a single, mean scaling factor that relates the modeled enhancements to the observed ones—one scaling factor for each inventory per city. The bootstrap method is as follows. For each overpass analyzed, there is a set of scaling factors (with one scaling factor per inventory representation) and a single associated uncertainty. Each scaling factor and its corresponding uncertainty are assumed to have a Gaussian distribution, with the scaling factor as the mean and the uncertainty as the standard deviation. It is resampled based on this distribution, outputting a single resampled scaling factor. All of the resampled scaling factors for an inventory per city further undergo a bootstrap resampling, using the mean statistic. This dual set of resamplings is repeated 1 million times for each inventory per city. The mean of the vector of 1 million resamplings is the mean scaling factor for each inventory per city. Meanwhile, the 5th and 95th percentiles serve as the bounds of the 90% confidence interval.

The mean scaling factors are then multiplied by the corresponding prior emissions magnitude, in order to scale the prior inventory estimates by the OCO-2 observations.

2.2.6. Evaluation of Uncertainties

We determine uncertainties for both the modeled and OCO-2 integrated enhancements. Combining these two sources of uncertainty also results in an uncertainty for the scaling factors and for our estimates of $FFCO_2$ emissions for each city.

For the model integrated enhancements, we leverage the 33% fractional uncertainty over five overpasses due to horizontal and vertical transport evaluated by Wu et al. (2018), which was estimated by quantifying the effect of the inclusion in the STILT model of a wind error component derived from radiosonde observations (horizontal) and calculating the root-mean-squared errors between enhancements found with different rescaled planetary boundary layer heights (vertical). To find an average fractional uncertainty per overpass, we multiply this value by the square root of 5, resulting in a ~74% fractional uncertainty to describe the transport uncertainty for a single overpass's latitudinally integrated modeled enhancements. While their study focused on Riyadh, we apply this uncertainty to all of the Middle Eastern cities in the study. This transport uncertainty is the only one we use for the modeled integrated enhancements. Since the emissions inventories do not have their own uncertainties and this study attempts to constrain those very emissions, we do not include a measure of the inventory uncertainties for the modeled integrated enhancements.

For the OCO-2 integrated enhancements, we first evaluate the uncertainties for each binned XCO_2 value at each 0.1° of latitude. We approximate this by considering the standard deviation of the medians in each bin.

The uncertainties related to the background also are included, since the OCO-2 enhancements are determined by subtracting the background from the total XCO_2. Contributions to the background uncertainty come from both the OCO-2 tails and from CarbonTracker. The tails' uncertainty contribution is determined by first only using the bin-level standard deviations that correspond only to the tails. Those values are added in quadrature and divided by the number of values used to find the average tail values. Meanwhile, the CarbonTracker contribution to the uncertainty is calculated by finding the standard deviation of the CarbonTracker-derived values (both boundary condition and biospheric influence) within the enhanced region of each overpass.

The contributions to the OCO-2 uncertainties (OCO-2 spread, tail spread, and CarbonTracker spread) are then added in quadrature, resulting in uncertainty values corresponding to each binned XCO_2 enhancement at each 0.1° of latitude. The combined binned uncertainties are then also added in quadrature, resulting in an uncertainty corresponding to the OCO-2 enhancements for a single overpass. A fractional OCO-2 uncertainty is found by dividing this single uncertainty by the integrated enhancements for each overpass.

To determine the uncertainties for each scaling factor corresponding to a single overpass, the 74% transport uncertainty and the fractional OCO-2 uncertainties (generally substantially smaller than the transport uncertainties) are combined by taking the square root of the sum of the two squared fractional uncertainties.

The overpass-level uncertainties are then used as inputs into the bootstrap method described in section 2.2.5 to determine 90% confidence intervals for the mean scaling factors.

3. Results

3.1. Inventory Representations of the Middle East in Emissions-Space

On a regional scale, we determine that the three studied emissions inventories represent the Middle East quite differently, from the perspective of both spatial distributions of emissions and magnitudes of emissions. In Figure 1, we show inventory representations of a domain that approximates the Middle East. A visual appraisal of these representations suggests that the way cities, roads, and transport are distributed throughout the Middle East vary from inventory to inventory. The total emissions magnitudes for this domain varies between inventories as well. For FFDAS, ODIAC, and EDGAR, respectively, the emissions represented in this domain come out to 697, 789, and 722 MtC $year^{-1}$.

Table 2
Kendall Rank Correlation Coefficients Between Pairs of Inventories for Each Studied City

	FFDAS-ODIAC	ODIAC-EDGAR	EDGAR-FFDAS
Riyadh	0.81	0.81	0.78
Mecca	0.70	0.82	0.69
Tabuk	0.77	0.85	0.78
Baghdad	0.64	0.67	0.51
Jeddah	−0.010*	−0.15*	0.71

Note. For all paired inventory domains other than those including the ODIAC representation of Jeddah (labeled with *), the inventory representations are correlated with each other.

3.2. Inventory Representations of Urban Centers in the Middle East

Emissions representations of each city of interest for each inventory at $0.1 \times 0.1°$ resolution are displayed in Figure 2, with emissions scaling linearly. Note that the displayed scales are the same for each representation for a city, though the maximum values vary for each city.

Figure 2 illustrates that the emissions within the chosen domains are distributed quite differently, to the point that it is not immediately apparent that different inventories are even representing the same city, apart from overlapping of the highest-valued grid cells in each representation. Same-city patterns across the inventories are not apparent in this form. A sense of these distributions can be gained by determining the Kendall rank correlations between pairs of inventories for each city, as shown in Table 2. Apart from those correlation coefficients associated with Jeddah's representation by ODIAC, which has an error with its representation (see Text S2), the remaining correlation coefficients suggest that the same-city inventory representations are correlated with each other to varying degrees without being identical.

The magnitudes of the emissions within each representation, that is, the sums of the emissions in each displayed domain, are listed in Table 3. Other than in the case of Jeddah, the FFDAS and ODIAC summed magnitudes for all other cities are more similar to each other than with EDGAR.

The maximum values in each domain are also listed in Table 3. For each city, the maximum valued grid cell among the three inventories is almost always represented by FFDAS. Those maximum values are often significantly higher than those of the other inventories, sometimes multiple times higher. This suggests that for FFDAS, large point sources carry more weight than they do for the other two inventories, whether that is due to the method of distribution of emissions by proxy, or due to large point source data containing much higher values for FFDAS than those used by the other inventories.

In Figure 3, we show cumulative emissions curves for each inventory and city. These curves reaffirm that the total magnitudes of emissions for FFDAS and ODIAC tend to be more similar than they are for EDGAR. These figures also give us a further understanding of the distribution of the emissions for each inventory representation of a city. For each city (other than Jeddah), the FFDAS and ODIAC curves are more rounded out than the EDGAR curves, suggesting that more of the total emissions in the domain of interest are due to fewer and higher-emission grid cells. This also means that EDGAR's urban representations distribute emissions more evenly than the other two inventories. This could be attributed to EDGAR distributing on-road emissions more homogeneously than the other inventories, which was discussed by Gately and Hutyra (2017) for an earlier version of EDGAR. Regarding the greater similarity in distribution between the FFDAS and ODIAC curves, this could be due to their shared use of nighttime lights as a proxy for distributing emissions. In any case, all curves show EDGAR as having the most evenly spread emissions, suggesting that all studied cities follow similar patterns of emissions distributions for the same inventories, in spite of relative differences in emissions magnitudes.

Table 3
City Populations, Total Sums (Magnitudes) of CO_2 Emissions, and Maximum Values of CO_2 Emissions Within Each City's Domain

	Population (million people)	Emissions sums (MtC $year^{-1}$)			Maximum values (MtC $year^{-1}$)		
		FFDAS	ODIAC	EDGAR	FFDAS	ODIAC	EDGAR
Riyadh	5.2	29.0	28.4	18.2	17.6	6.36	7.26
Mecca	1.5	5.81	6.54	7.89	3.10	1.88	0.616
Tabuk	0.4	2.48	2.37	0.548	1.65	0.837	0.137
Baghdad	6.7	4.74	3.60	7.00	1.91	1.26	1.38
Jeddah	3.4	13.4	4.08*	19.4	5.74	2.73*	8.97

Note. Population data include April 2010 census data for Saudi Arabian cities and October 2009 estimates for Baghdad (Brinkhoff, 2018). Jeddah values marked with * indicate the error in their representation by ODIAC (see Text S2).

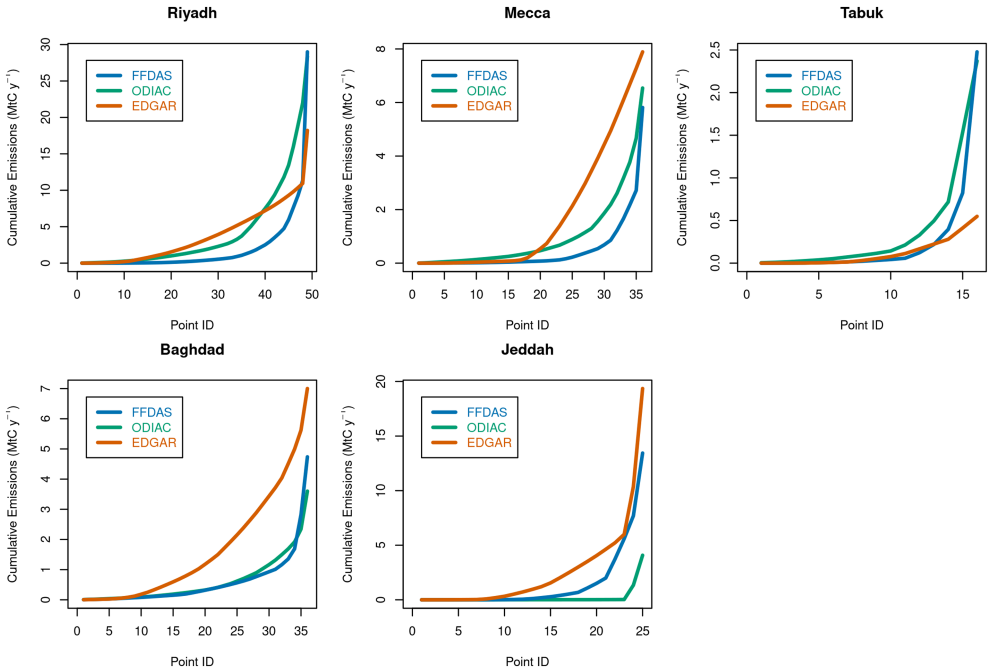

Figure 3. Cumulative emissions curves for each inventory and city. These figures simultaneously represent magnitudes and spatial distributions of emissions for each city and inventory. The total emissions magnitudes are represented by the rightmost point, whereas the distributions are represented by the shape of the curves, with straighter curves being more evenly spread than those that are more rounded out. Based on these curves, the EDGAR representations of the cities of interest differ most from the other inventories, with more variant magnitudes and more evenly spread emissions.

3.3. Simulated Atmospheric Concentrations of Urban Centers in the Middle East

The magnitude and spatial distribution differences between the different inventories across the cities of interest are found not only in their emissions but also in their respective simulations of concentrations. As an example to show the manifestations of these differences into concentration-space, Figure 4 shows the enhancements for six different observed and modeled overpasses, for all five cities (two overpasses for Riyadh) and all on different dates. Note that we have included ODIAC in its native resolution ("ODIAC" in green) and at the resolution aggregated to match the other inventories ("Agg ODIAC" in pink). In these sample overpasses, all of the modeled overpasses capture the urban plumes, though the distribution and the magnitudes of the captured enhancements vary by inventory used in the model. While these modeled overpasses' attributes are not representative of all modeled overpasses for their respective cities, they do show how the differences in emissions magnitudes and distributions have manifested themselves into concentration-space on these particular days due to atmospheric transport. A similar figure to Figure 4 but with observational uncertainties included is in the supporting information as Figure S2. Corresponding observed and simulated total XCO_2 plots are found in Figure S3.

3.4. Comparison Between Observed and Simulated Enhancements

When comparing the simulated enhancements by latitude to the observed ones, such as in Figure 4, we continue to see that the magnitude and distribution of enhancements in the models differ from what OCO-2 has observed. In this figure, we show enhancements versus latitude, where the black squares are the binned observed enhancements and the other colored points show the different modeled enhancements. In certain cases, the magnitudes of the enhancements seem fairly consistent between the observations and models (e.g., Riyadh on 16 February 2016); in others, they are quite different (e.g., Mecca on 22 March 2016), which we attribute to daily variations in emissions that are not captured by the temporal resolution of the models. However, the models capture the urban plumes that the satellite has observed, suggesting that there is a fidelity to the models. There are also cases in which the observed and the modeled plumes are latitudinally shifted from each other, as in the overpass of Riyadh on 27 December 2014. This latitudinal shift, which is

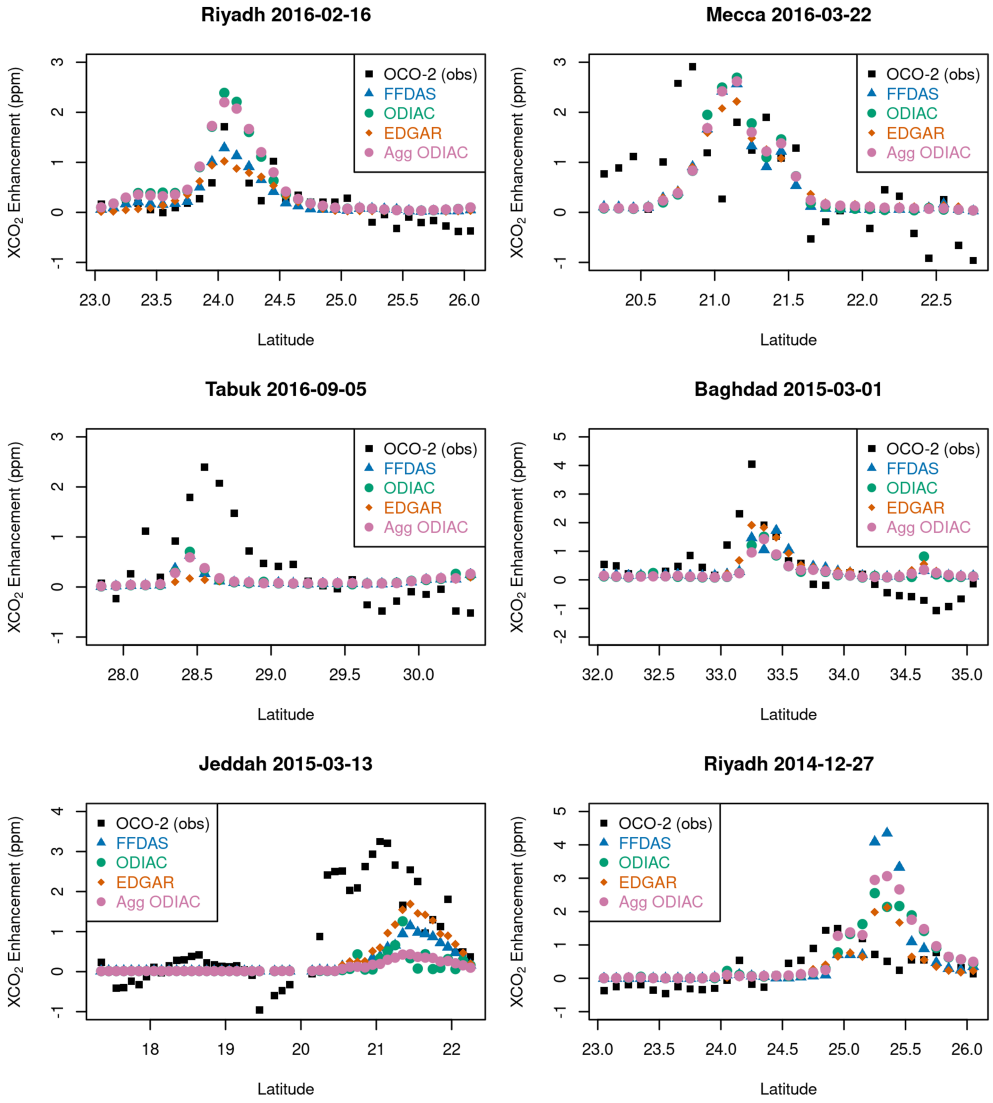

Figure 4. Enhancements of sample observed (black) and modeled (other colors) overpasses for different cities and days. Note that we have included ODIAC in its native resolution ("ODIAC" in green) and at the resolution aggregated to match the other inventories ("Agg ODIAC" in pink). All modeled overpasses capture the urban plume depicted in the observations. Differences in magnitudes and spatial distributions of the emissions manifest themselves in the differing representations of the enhancements. The sample overpasses for Jeddah on 13 March 2015 and Riyadh on 27 December 2014 depict latitudinal shifts in the urban plume as a result of transport errors that offset the location of the plumes. Our integral method of comparison between the observed and modeled enhancements is not inhibited by these latitudinal shifts.

likely due to transport errors, demonstrates the model's ability to capture the urban plume, even when transport errors offset the plume's location from the satellite-observed plume.

The differing distributions of enhancements and latitudinal shifts in the plumes between the simulations and the observations lend themselves to an integral method of comparing enhancements. This process is captured in the cumulative enhancement curves shown in the four unscaled examples (left) in Figure 5, which correspond to four of the overpasses shown in Figure 4. The cumulative sum or magnitude of the enhancements for each overpass is captured by the rightmost points on each curve, visually representing the differences in magnitude between the different models and the observation for each respective overpass.

In the scaled cumulative enhancement curves of Figure 5 (right), the summed modeled enhancements are all scaled to match those of the observed enhancements for each overpass. These curves make the

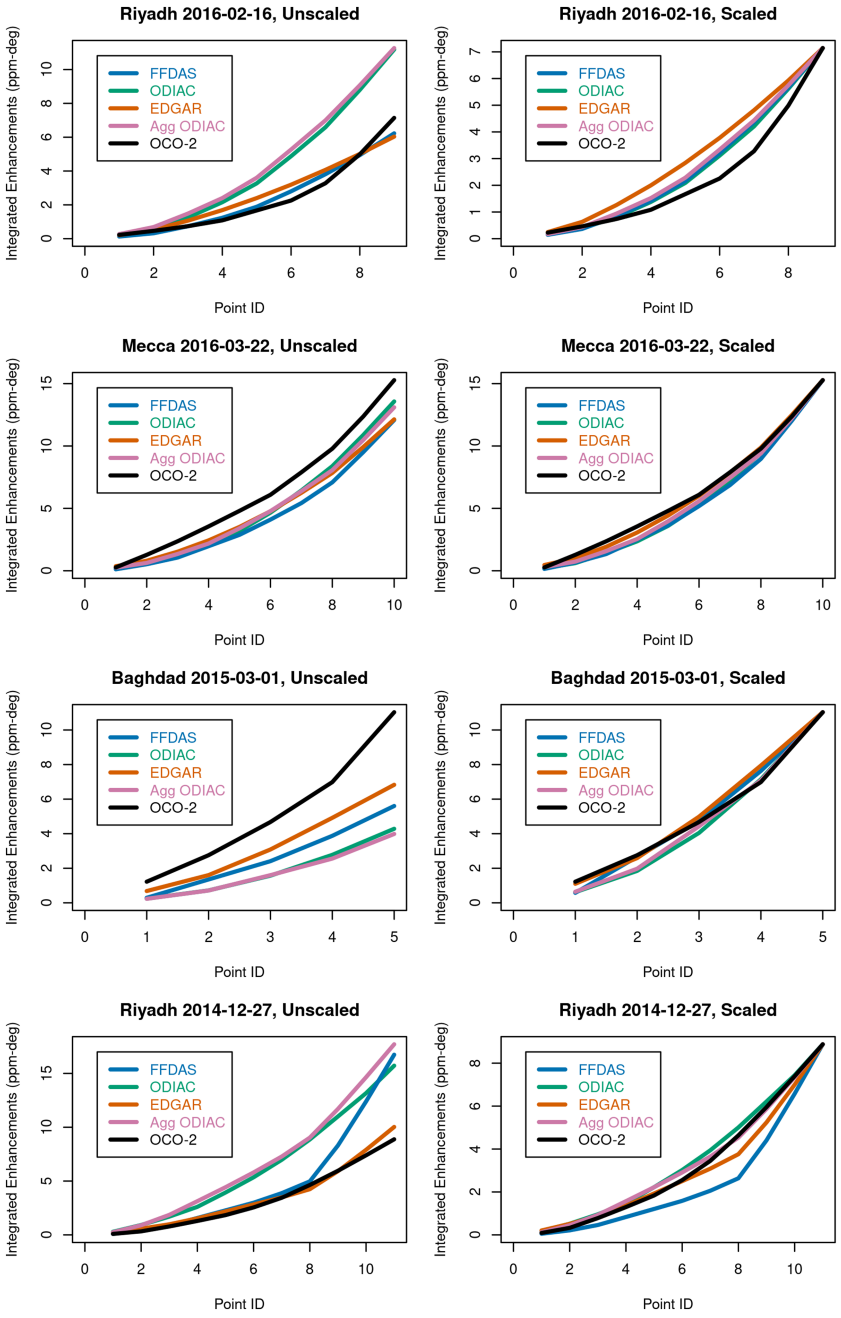

Figure 5. Cumulative enhancement curves, unscaled (left) and scaled (right), corresponding to four of the overpasses shown in Figure 4. The unscaled curves illustrate magnitude differences between the models and observations, while the scaled illustrate the spatial differences in the enhancements. The unscaled modeled curves on the left are scaled on the right to match the integral of the OCO-2 enhancements. We use those scaling factors to quantify the relationship between the modeled and observed enhancements.

differences in distribution of enhancements within the urban plume clearer. Despite the distributions of the enhancements still differing, this method normalizes the magnitudes of the modeled enhancements to those of the OCO-2 observations, resulting in overpass-specific scaling factors for each inventory, for each studied day. For the overpasses in Figure 5, we have scaling factors ranging from 0.50 (aggregated ODIAC simulation of overpass for Riyadh on 27 December 2014) to 2.77 (aggregated ODIAC simulation of overpass for Baghdad on 1 March 2015). Scaling factors vary from day to day for a specific city.

Table 4
Model-Observation Mean Scaling Factors

	FFDAS	ODIAC	EDGAR	Agg. ODIAC
Riyadh	1.1 (0.5, 1.8)	1.2 (0.5, 1.9)	1.8 (1.0, 2.6)	1.1 (0.4, 1.8)
Mecca	1.7 (0.7, 2.7)	1.7 (0.6, 2.8)	1.4 (0.5, 2.4)	1.6 (0.6, 2.6)
Tabuk	3.4 (0.9, 5.8)	3.1 (0.7, 5.4)	8.7 (4.6, 12.7)	3.1 (0.7, 5.4)
Baghdad	3.0 (1.3, 4.9)	3.0 (1.6, 4.4)	2.2 (0.8, 3.7)	4.0 (1.8, 6.4)
Jeddah	3.3 (2.1, 4.4)	6.7 (5.5, 7.9)	1.9 (0.9, 2.9)	6.6 (4.9, 8.5)

Note. These mean scaling factors represent the relationship between the inventory modeled enhancements for each city and the satellite observations. Their respective 90% confidence intervals are in parentheses.

All overpass-level scaling factors are combined into a singular scaling factor for each inventory and city with the bootstrap method discussed in section 2.2.5. These mean scaling factors are listed in Table 4 with their 90% confidence intervals. These scaling factors are determined by comparison with midday satellite observations, meaning we correct the whole inventory based on midday comparisons, an approach that relies on the diurnal model representation being accurate.

3.5. Estimates of Urban Emissions in the Middle East

The mean scaling factors multiplied with their respective inventories result in new, scaled estimates of urban emissions in the Middle East. These estimates are shown in the first five panels of Figure 6. In these panels, for each city, the colored bars represent the inventory magnitudes, while the gray bars represent the corresponding scaled emissions estimates, with 90% confidence intervals. Though some of the confidence intervals are quite large, we find that many of the scaled inventory estimates and their corresponding confidence intervals are either higher than the prior inventory emissions magnitudes or overlap only slightly. This suggests that, as compared to the OCO-2 observations, the global emissions inventories are underestimating several of the urban representations of afternoon emissions. This is the case for the EDGAR representation of Riyadh, and all prior inventory representations of Tabuk, Baghdad, and Jeddah. The mean scaled estimates all exceed the prior inventory estimates for all cities.

As discussed in section 3.2, all of the prior inventory representations of each city demonstrate varying spatial distributions and magnitudes of emissions, with no specific measure of uncertainty. However, our work has determined emissions estimates for each city that converge across inventories and are indistinguishable within the confines of the confidence intervals. Based on a limited number of observations, we have thus found emissions magnitude estimates for each city that are insensitive to the distributions of the prior

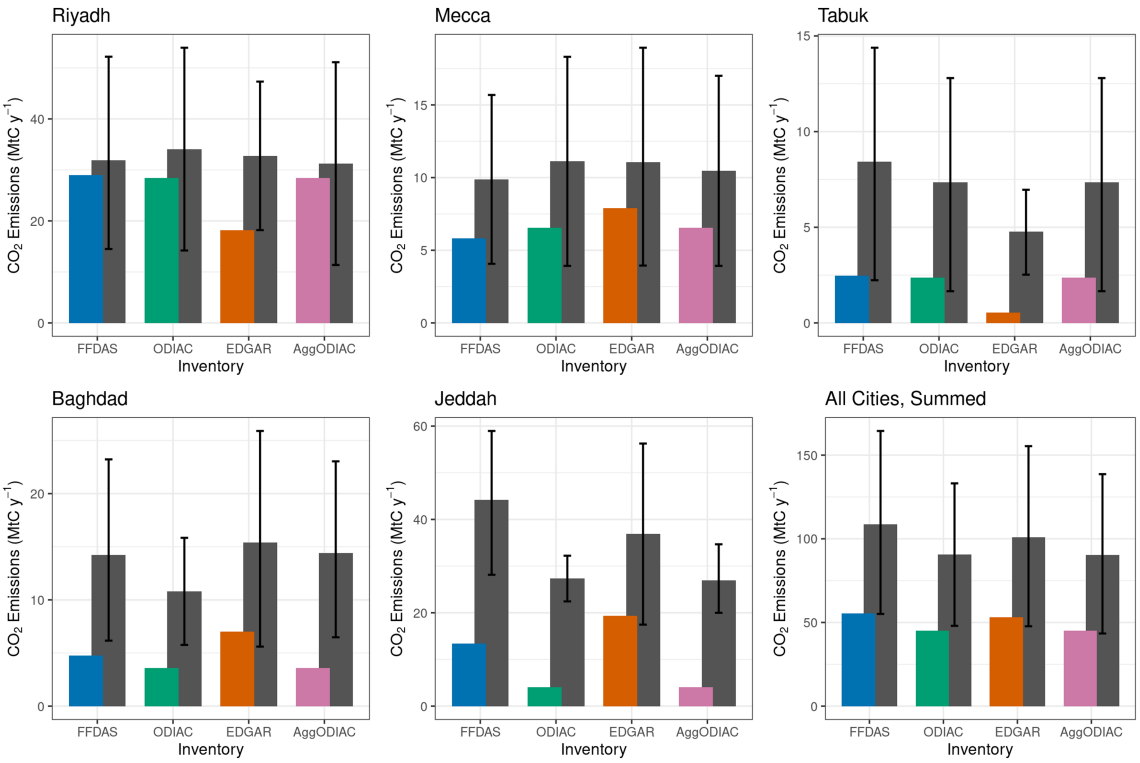

Figure 6. Prior (colored) and scaled (gray) emissions magnitude estimates for our five cities of interest: Riyadh, Mecca, Tabuk, Baghdad, and Jeddah, as well as the sums of the emissions of all five cities. The black lines on the gray bars represent the 90% confidence intervals. When taken in aggregate, the prior emissions magnitudes underestimate emissions as compared to those scaled by our emissions scaling factors.

inventory emissions and that adjust the magnitudes of the emissions, in spite of the differences between the priors. This is a surprisingly powerful result, especially given the significance placed on the accuracy of high-resolution inventories for previous ground-based inversion studies, (e.g., Lauvaux et al., 2016; Oda et al., 2017). While the space-based column measurements of CO_2 concentrations have not given us information to quantify the spatial distribution of emissions, it has allowed us to bypass the differences in spatial distributions by providing an integral constraint to quantify urban magnitudes of CO_2 emissions.

Sums of the different emissions estimates for each inventory (prior and scaled with confidence intervals) across the five cities are shown at the bottom right in Figure 6. As shown in this figure, our scaled estimates, with minimal or no overlap between our 90% confidence intervals and the prior emissions, demonstrate larger emissions than those estimated by the unscaled inventories. In other words, the five cities are, when summed, underestimated by their prior inventory representations as compared to their observation-scaled counterparts.

4. Discussion

4.1. Implications for Middle Eastern Emissions and Their Representations

Our scaled emissions and their respective 90% confidence intervals constrain emissions of five Middle Eastern cities. We find that for Riyadh, EDGAR representations underestimate afternoon emissions as compared to what our observations suggest. Moreover, for Tabuk, Baghdad, and Jeddah, almost all prior inventory representations underestimate afternoon emissions. After summing emissions for all five cities, prior estimates for all inventories overlap minimally or not at all with lower bounds of the scaled emissions confidence intervals. Thus, we can highlight that, in sum, inventory representations across all five cities underestimate afternoon emissions. If prior and scaled summed emissions are averaged across three inventories, the average scaled summed emissions of all cities are 100 MtC $year^{-1}$ (50, 151), which is 2.0 (1.0, 3.0) times greater than the average prior summed emissions.

Our scaled estimates have regional significance. The entire Middle East as a whole emits approximately 700 MtC $year^{-1}$ (Boden et al., 2016). The difference between the average scaled and prior summed emissions for our five cities is 49 (−1, 100) MtC $year^{-1}$, which is equivalent to ~7% (0%, 14%) of total Middle Eastern emissions. Meanwhile, the emissions of Saudi Arabia and Iraq are approximately 200 MtC $year^{-1}$ (Boden et al., 2016). Prior and scaled summed emissions come out to ~26% and 50% (25%, 76%) of the Saudi Arabian and Iraqi total, respectively. The difference between the average scaled and prior summed emissions for the five studied cities comes out to ~25% (−1%, 50%) of the sum of the two countries' total emissions. However, further work at a national or regional scale is needed to determine whether the underestimation in these urban emissions representations suggests either underestimation in the national totals or if our estimated discrepancy is more a result of misallocation of the emissions throughout the studied countries.

This work demonstrates the ability to use satellite observations to constrain urban emissions of cities of a variety of sizes. Given the much smaller magnitude of emissions of Tabuk as compared to Riyadh, for example, it is a surprising result that we can observe XCO_2 enhancements from a city like Tabuk from space (as in Figure S4). It does appear that with appropriate isolation from strong XCO_2 signals, and with sufficient concentration of emissions and favorable atmospheric dynamics, even relatively small cities (with populations of hundreds of thousands) can be identified and have emissions quantified.

This study provides an objective evaluation of the emission downscale approaches used by the gridded emissions inventories. For the five studied cities, EDGAR's prior estimates overlap substantially with the confidence intervals of our scaled emissions estimates only for Mecca. For every other city, the EDGAR representations have emissions lower than or minimally overlapping with the lower bound of the confidence intervals we estimate. As such, EDGAR does not seem to accurately capture urban emissions of these studied cities. Why is this? EDGAR tends to distribute emissions across the urban domain more than the other two inventories. Of the studied cities, Riyadh is the largest, and EDGAR's emissions estimate is significantly lower than those of other inventories. It is possible that the distributed nature of how EDGAR grids emissions is less able to capture the high emissions intensity of such a large city than alternate gridding methods that rely on proxies such as nightlights.

FFDAS and ODIAC (and its aggregated version) both also appear to underestimate emissions in comparison with the OCO-2 data, though their emission magnitudes are within the large confidence intervals of three of our optimized estimates of individual cities. These two inventories share the use of nighttime lights as a proxy for emissions distribution, and this work is somewhat suggestive this may be a useful proxy for global representation of urban domains, but further work is required to further parse out their similarities, differences, and reasons for their underestimation of the studied urban emissions.

Our work also has the potential to identify larger errors in inventory representations of cities. For example, the ODIAC representation of Jeddah has an error in it due to a mismatch between two nightlight data sources (one for distributed emissions and the other for gas flare emissions, see Oda et al., 2018) at that particular location. This error was identified with our methods. We have not corrected for the spatial distribution error in this manuscript, but we have estimated the magnitude of emissions for Jeddah that can be used to take into account when correcting ODIAC for future versions of the inventory.

4.2. The Spatial Resolution of the Prior Emissions

When making direct comparisons between the emissions inventories themselves, we scale ODIAC to match the spatial resolution of the other inventories. However, in its native spatial resolution, ODIAC has 12 times higher resolution in each spatial dimension. In order to assess the effect of using the two different spatial resolutions, all modeled results include both native resolution results ("ODIAC") and results in which the resolution is aggregated up to $0.1 \times 0.1°$ ("Agg ODIAC"). We can see in plots such as Figure 4 that the differences in spatial resolution do affect the individual modeled enhancement values, as the points for the two ODIAC resolutions do not perfectly overlap with each other. However, within the scope of our study, the differences in spatial resolution do not greatly affect the outcome of our scaled emissions estimates. For each studied city, the scaled estimates across the different ODIAC resolutions provide convergent estimates. Thus, our estimates are insensitive to both the differences in the studied resolutions as well as the spatial distributions in the priors.

4.3. Limitations and Their Implications

The OCO-2 satellite follows a sun-synchronous orbit, meaning that all the observations used in this study are temporally limited to local afternoon. Consequently, we cannot assess the diurnal cycle of our results, and our results are dependent on the accuracy of the diurnal model representation by the TIMES scaling factors. Our results thus suggest an emissions underestimation by the modeled inventory representations and their corresponding parts—namely, the convolution of the inventories themselves and the diurnal model representations. (Note that the FFDAS product has its own internal diurnal cycle embedded into the product based on the TIMES factors.) We cannot disentangle the TIMES factors from the inventory representations, which suggests that this underestimation could be a result of the inventories having too-low emissions, the TIMES diurnal cycle not fully capturing afternoon emissions, or a combination of these two factors.

Given the widespread usage of the TIMES factors for capturing diurnal cycles, it is possible that there is a systematic bias across many model representations of diurnality of emissions. The TIMES factors are potentially unable to fully correct for the lack of diurnality in the emissions inventories, particularly at the urban scale, a scale for which these factors were not designed. While this potential bias cannot be tested using OCO-2 data alone, as is done in this study, observations from multiple times of day from other satellites such as the recently launched OCO-3 could allow for assessments of the diurnal cycle represented by the TIMES factors. Moreover, there is great potential for carbon-observing satellites in geostationary orbit to assess the diurnal cycle of CO_2 emissions, as they would be able to make multiple observations of the same locations throughout the day. For example, the upcoming Geostationary Carbon Cycle Observatory (GeoCARB) mission, which will continuously monitor column concentrations of CO_2, CH_4, and CO in the Americas throughout the day from geostationary orbit, is an important step forward in this respect (Moore et al., 2018).

Assessments of the emissions magnitudes of inventory representations are also limited by the signal-to-noise ratio (SNR) of the OCO-2 observations. The methods used in this paper to analyze observed overpasses rely on high SNR and optimal conditions; it is possible that overpasses with higher SNR simultaneously also have stronger signals, suggesting that our results may be biased toward higher scaling factors and thus suggesting greater underestimation by the modeled representations of afternoon emissions. This potential bias highlights the limitations of not just this study but for other studies involving space-based observations for the

use of urban emissions assessments. The use of space-based observations of XCO_2 can be hindered by a number of factors, such as clouds, high albedo, topography, and wind conditions; for regions at urban scales, these factors in turn limit the types of studies for which the observations can be used. Given these limitations, our study does not seek to quantify absolute emissions in the studied cities but rather assess the inventory representations' ability to capture the variations in urban emissions that are suggested by the observations used in this study. The underestimation by these inventory representations that we have evaluated thus suggest that the inventories can be improved to better capture such variations. The larger observed swaths available at urban scales in both the target and snapshot area mapping (SAM) modes of OCO-3 will provide for much more data at urban scales and could lessen some of the limitations associated with OCO-2 data (Eldering et al., 2019).

5. Summary and Conclusions

Global $FFCO_2$ emissions inventories need to be evaluated and improved at the urban scale for the sake of carbon cycle science and urban mitigation policies. In this study, we evaluate global $FFCO_2$ emissions inventory representations of afternoon urban CO_2 emissions in the Middle East and assess the ability of satellite observations to inform this evaluation. We find the relationship between top-down satellite observations and bottom-up simulations of XCO_2 to calculate scaling factors to adjust the prior inventory estimates of five Middle Eastern cities' emissions. Based on our findings, we estimate that the sum of the studied cities' scaled emissions are on average 2.0 (1.0, 3.0) times the prior inventory magnitudes. The underestimation of these five cities' emissions by the inventories comes out to ~7% (0%, 14%) of total Middle Eastern emissions. Our results are insensitive to the spatial differences in the inventory representations of the cities' emissions, facilitating robust quantitative assessments of urban emissions inventory representations. This is in contrast with atmospheric inverse analyses carried out with ground-based observations, which would be heavily dependent on the spatial distribution of the prior inventories. Using space-based XCO_2 observations allows us to constrain urban emissions in a fashion not previously available to us, enabling evaluation of inventories and downscaling methods at subnational scales. These results are based only on a few overpasses for each city; with additional observations, urban emissions could be even further constrained, and more cities could be included in a future study. In particular, the recent launch of the OCO-3 is especially exciting: It opens up new potential pathways for urban CO_2 emissions studies to move forward, with observations during different times of the day (instead of only the afternoon as with OCO-2), and the new SAM mode that could capture the entire urban plume (Eldering et al., 2019). The upcoming launch of the geostationary GeoCARB mission will also help to fill in some of the observational gaps of the sun-synchronous OCO-2 and OCO-3 satellites by mapping column concentrations of CO_2 in the Americas with multiple measurements of the same locations every day. Constraining urban emissions is critically important, and satellites are paving the way for such work to be done.

Acknowledgments
Funding for this study was provided by the National Aeronautics and Space Administration (NASA) (Grants NNX15AI42G, NNX14AM76G, NNX15AI40G, 80NSSC19K0092, and 80NSSC18K1307) and the National Science Foundation Graduate Research Fellowship (Grant DGE 1256260). The authors thank Dr. Genevieve Plant for assistance in error calculations, and the OCO-2 science team for general feedback. The OCO-2 data are hosted at the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center and can be found online (https://disc.gsfc.nasa.gov/datasets?project=OCO). The most recent version of FFDAS (v2.2) data can be found online (http://ffdas.rc.nau.edu/Data.html). The ODIAC emission data product is archived at the Center for Global Environmental Research, National Institute for Environmental Studies (NIES), Japan (http://db.cger.nies.go.jp/dataset/ODIAC/DL_odiac2017.html). EDGAR is provided and archived by the European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL) (http://edgar.jrc.ec.europa.eu/overview.php?v=432_GHG&SECURE=123). Derived data and model products can be found at the University of Michigan Deep Blue archive (https://doi.org/10.7302/rbjy-7e98).

References

Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., et al. (2014). A multiyear, global gridded fossil fuel CO_2 emission data product: Evaluation and analysis of results. *Journal of Geophysical Research: Atmospheres*, *119*, 10,213–10,231. https://doi.org/10.1002/2013JD021296

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., & Kennedy, J. J. (2016). El Nino and a record CO_2 rise. *Nature Climate Change*, *6*(9), 806–810. https://doi.org/10.1038/nclimate3063

Bodansky, D. (2016). The Paris climate change agreement: A new hope? *American Journal of International Law*, *110*(2), 288–319. https://doi.org/10.5305/amerjintelaw.110.2.0288

Boden, T., Andres, R., & Marland, G. (2016). Global, regional, and national fossil-fuel CO_2 emissions (1751-2013) (V. 2016) [Data set]. Environmental System Science Data Infrastructure for a Virtual Ecosystem; Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). https://doi.org/10.3334/cdiac/00001_v2016

BP. (2017). Statistical review of world energy. Retrieved July 19, 2019, from https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html

Brinkhoff, T. (2018, May 21). City population. Retrieved January 17, 2020, from http://www.citypopulation.de/

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., et al. (2018). Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth System Science Data*, *10*(4), 1987–2013. https://doi.org/10.5194/essd-10-1987-2018

Eldering, A., Taylor, T. E., O'Dell, C. W., & Pavlick, R. (2019). The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data. *Atmospheric Measurement Techniques*, *12*(4), 2341–2370. https://doi.org/10.5194/amt-12-2341-2019

European Commission Joint Research Centre. (2017). Methodology Retrieved April 10, 2018, from http://edgar.jrc.ec.europa.eu/methodology.php

Friedlingstein, P., Andrew, R. M., Rogelj, J., Peters, G. P., Canadell, J. G., Knutti, R., et al. (2014). Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience*, *7*(10), 709–715. https://doi.org/10.1038/ngeo2248

Gately, C. K., & Hutyra, L. R. (2017). Large uncertainties in urban-scale carbon emissions. *Journal of Geophysical Research: Atmospheres*, *122*, 11,242–11,260. https://doi.org/10.1002/2017JD027359

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., et al. (2019). Global atmospheric CO_2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate. *Biogeosciences*, *16*(1), 117–134. https://doi.org/10.5194/bg-16-117-2019

Gurney, K. R., Chen, Y., Maki, T., Kawa, S. R., Andrews, A., & Zhu, Z. (2005). Sensitivity of atmospheric CO_2 inversions to seasonal and interannual variations in fossil fuel emissions. *Journal of Geophysical Research*, *110*, D10308. https://doi.org/10.1029/2004JD005373

Gurney, K. R., Liang, J., O'Keeffe, D., Patarasuk, R., Hutchins, M., Huang, J., et al. (2019). Comparison of global downscaled versus bottom-up fossil fuel CO_2 emissions at the urban scale in four U.S. urban areas. *Journal of Geophysical Research: Atmospheres*, *124*, 2823–2840. https://doi.org/10.1029/2018JD028859

Gurney, K. R., Romero-Lankao, P., Seto, K. C., Hutyra, L. R., Duren, R., Kennedy, C., et al. (2015). Climate change: Track urban emissions on a human scale. *Nature*, *525*(7568), 179–181. https://doi.org/10.1038/525179a

Hakkarainen, J., Ialongo, I., & Tamminen, J. (2016). Direct space-based observations of anthropogenic CO_2 emission areas from OCO-2. *Geophysical Research Letters*, *43*, 11,400–11,406. https://doi.org/10.1002/2016GL070885

Hedelius, J. K., Liu, J., Oda, T., Maksyutov, S., Roehl, C. M., Iraci, L. T., et al. (2018). Southern California megacity CO_2, CH_4, and CO flux estimates using ground- and space-based remote sensing and a Lagrangian model. *Atmospheric Chemistry and Physics*, *18*(22), 16271–16291. https://doi.org/10.5194/acp-18-16271-2018

Hutchins, M. G., Colby, J. D., Marland, G., & Marland, E. (2017). A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States. *Mitigation and Adaptation Strategies for Global Change*, *22*(6), 947–972. https://doi.org/10.1007/s11027-016-9709-9

Hutyra, L. R., Duren, R., Gurney, K. R., Grimm, N., Kort, E. A., Larson, E., & Shrestha, G. (2014). Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective. *Earth's Future*, *2*, 473–495. https://doi.org/10.1002/2014EF000255

International Energy Agency. (2008). World energy outlook 2008. International Energy Agency.

IPCC. (2014). Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change (P. 151). Geneva, Switzerland: IPCC.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., et al. (2019). EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012. *Earth System Science Data*, *11*(3), 959–1002. https://doi.org/10.5194/essd-11-959-2019

Janssens-Maenhout, G., Pagliari, V., Guizzardi, D., & Muntean, M. (2013). Global emission inventories in the Emission Database for Global Atmospheric Research (EDGAR) - Manual (I). Gridding: EDGAR emissions distribution on global gridmaps (JRC Technical Reports) (pp. 1–33). Joint Research Centre of the European Commission. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/75050af2-850c-4b87-ae35-eaa82de96489/language-en

Kort, E. A., Frankenberg, C., Miller, C. E., & Oda, T. (2012). Space-based observations of megacity carbon dioxide. *Geophysical Research Letters*, *39*, L17806. https://doi.org/10.1029/2012GL052738

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., et al. (2016). High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX). *Journal of Geophysical Research: Atmospheres*, *121*, 5213–5236. https://doi.org/10.1002/2015JD024473

Lin, J. C., Mitchell, L., Crosman, E., Mendoza, D. L., Buchert, M., Bares, R., et al. (2018). CO_2 and carbon emissions from cities: Linkages to air quality, socioeconomic activity, and stakeholders in the Salt Lake City urban area. *Bulletin of the American Meteorological Society*, *99*(11), 2325–2339. https://doi.org/10.1175/BAMS-D-17-0037.1

Moore, B. III, Crowell, S. M., Rayner, P. J., Kumer, J., O'Dell, C. W., O'Brien, D., et al. (2018). The potential of the geostationary carbon cycle observatory (GeoCarb) to provide multi-scale constraints on the carbon cycle in the Americas. *Frontiers in Environmental Science*, *6*. https://doi.org/10.3389/fenvs.2018.00109

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., & Crisp, D. (2017). Quantifying CO_2 emissions from individual power plants from space. *Geophysical Research Letters*, *44*, 10–45. https://doi.org/10.1002/2017GL074702

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., & Deng, F. (2013). Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets. *Journal of Geophysical Research: Atmospheres*, *118*, 917–933. https://doi.org/10.1029/2012JD018196

Oda, T., Bun, R., Kinakh, V., Topylko, P., Halushchak, M., Marland, G., et al. (2019). Errors and uncertainties in a gridded carbon dioxide emissions inventory. *Mitigation and Adaptation Strategies for Global Change.*, *24*(6), 1007–1050. https://doi.org/10.1007/s11027-019-09877-2

Oda, T., Lauvaux, T., Lu, D., Rao, P., Miles, N. L., Richardson, S. J., & Gurney, K. R. (2017). On the impact of granularity of space-based urban CO_2 emissions in urban atmospheric inversions: A case study for Indianapolis, IN. *Elementa: Science of the Anthropocene*, *5*, 28. https://doi.org/10.1525/elementa.146

Oda, T., & Maksyutov, S. (2011). A very high-resolution (1 km×1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmospheric Chemistry and Physics*, *11*(2), 543–556. https://doi.org/10.5194/acp-11-543-2011

Oda, T., & Maksyutov, S. (2015). ODIAC fossil fuel CO_2 emissions dataset (ODIAC2017). *Center for Global Environmental Research, National Institute for Environmental Studies.*. https://doi.org/10.17595/20170411.001

Oda, T., Maksyutov, S., & Andres, R. J. (2018). The open-source data inventory for anthropogenic CO_2, version 2016 (ODIAC2016): A global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions. *Earth System Science Data*, *10*(1), 87–107. https://doi.org/10.5194/essd-10-87-2018

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., et al. (2012). The ACOS CO_2 retrieval algorithm—Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, *5*(1), 99–121. https://doi.org/10.5194/amt-5-99-2012

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M., Fisher, B., et al. (2018). Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm. *Atmospheric Measurement Techniques*, *11*(12), 6539–6576. https://doi.org/10.5194/amt-11-6539-2018

Olivier, J. G. J., & Janssens-Maenhout, G. (2015). CO_2 emissions from fuel combustion Part III: Total greenhouse gas emissions; 2015 Edition (pp. 1–43). International Energy Agency. Retrieved from http://www.pbl.nl/sites/default/files/cms/publicaties/pbl-2015-iea-co2-emissions-from-fuel-combustion-1971-2013-part-3-total-greenhouse-gas-emissions.pdf

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P., & Koffi, E. (2010). A new global gridded data set of CO_2 emissions from fossil fuel combustion: Methodology and evaluation. *Journal of Geophysical Research*, *115*, D19306. https://doi.org/10.1029/2009JD013439

Rosenzweig, C., Solecki, W., Hammer, S. A., & Mehrotra, S. (2010). Cities lead the way in climate-change action. *Nature*, *467*(7318), 909–911. https://doi.org/10.1038/467909a

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., et al. (2013). Technical summary. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change*, (pp. 33–115). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.005

United Nations, Department of Economic and Social Affairs, Population Division. (2015). World urbanization prospects: The 2014 revision.

Wu, D., Lin, J. C., Fasoli, B., Oda, T., Ye, X., Lauvaux, T., et al. (2018). A Lagrangian approach towards extracting signals of urban CO_2 emissions from satellite observations of atmospheric column CO_2 (XCO_2): X-Stochastic Time-Inverted Lagrangian Transport model ("X-STILT v1"). *Geoscientific Model Development*, *11*(12), 4843–4871. https://doi.org/10.5194/gmd-11-4843-2018