Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

Geoscientific Model Development Discussions

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 v1.1")

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

[1]Department of Atmospheric Sciences, University of Utah, Salt Lake City, USA
[2]Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, Maryland/Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[3]Department of Meteorology and Atmospheric Science, Pennsylvania State University, USA
[4]Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, USA

Correspondence to: Dien Wu (Dien.Wu@utah.edu)

Abstract. Urban regions are responsible for emitting significant amounts of fossil fuel carbon dioxide ($FFCO_2$), whose emissions at finer, city scales are more uncertain than those aggregated at the global scale. Carbon-observing satellites may provide independent *top-down* emission evaluations and compensate for the sparseness of surface CO_2 observing networks, especially in urban areas. Although some previous studies have attempted to derive urban CO_2 signals from satellite column-averaged CO_2 data (XCO_2) using simple statistical measures, less work has been carried out to link upwind emission sources to downwind atmospheric columns using atmospheric models. In addition to Eulerian atmospheric models that have been customized for emission estimates over specific cities, the Lagrangian modeling approach—in particular, the Lagrangian Particle Dispersion Model (LPDM) approach—has the potential to efficiently determine the sensitivity of downwind concentration changes to upwind sources. However, when applying LPDMs to interpret satellite XCO_2, several issues—namely, uncertainties in anthropogenic XCO_2 signals due to receptor configurations and errors in atmospheric transport and background XCO_2—have yet to be addressed.

In this study, we present a modified version of the Stochastic Time-Inverted Lagrangian Transport (STILT) model, "X-STILT", for extracting urban XCO_2 signals from NASA's Orbiting Carbon Observatory 2 (OCO-2) XCO_2 data. X-STILT incorporates satellite profiles and provides comprehensive uncertainty estimates towards urban XCO_2 enhancements for selected satellite soundings. Several methods to initialize receptors/particle setups and determine background XCO_2 are presented and discussed via sensitivity analyses and comparisons. To illustrate X-STILT's utilities and applications, we examined five OCO-2 overpasses over Riyadh, Saudi Arabia, during a two-year time period and performed a simple scaling factor-based inverse analysis. As a result, the model is able to reproduce most observed XCO_2 enhancements. Conservative error estimates show that the 68 % confidence limit of XCO_2 uncertainties due to transport and emission uncertainties contribute to ~29 % and ~25 % of the averaged latitudinally-integrated urban signals, respectively, over five overpasses, given meteorological fields from the Global Data Assimilation System (GDAS). Additionally, a sizable negative "bias" of 0.56 ppm in background derived from a previous study employing simple statistics (daily median over a regional domain) leads to positive biases of 43 % in mean observed urban signal and 68 % in posterior scaling factor, from a simple inversion analysis. Based on these results, we foresee X-STILT serving as a tool for interpreting column measurements, estimating urban enhancement signals, and carrying out inverse modeling to improve quantification of urban emissions.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

1 Introduction

Carbon dioxide (CO_2) is a major greenhouse gas in the atmosphere in terms of radiative forcing, with its concentration increasing significantly over the past century (Dlugokencky and Tans, 2015). The largest contemporary net source of CO_2 to the atmosphere over the decadal time scales is anthropogenic emissions, namely from fossil fuel burning and net land-use change (Ciais et al., 2013). Urban areas play significant roles in the global carbon cycle and are responsible for over 70 % of the global energy-related CO_2 emissions (Rosenzweig et al., 2010). Global fossil fuel CO_2 ($FFCO_2$) emission uncertainty (8.4%, 2σ, Andres et al., 2014) may be smaller than other less-constrained emissions such as of wildfire (Brasseur and Jacob, 2017). Still, uncertainties associated with national $FFCO_2$ emissions derived from *bottom-up* inventories typically range from 5–20 % per year (Andres et al., 2014). These estimated emission uncertainties result primarily from differences in emission inventories, such as the emission factors and energy consumptions data used. Moreover, heightened interests in regional- and urban-scale emissions require modelers to investigate $FFCO_2$ emissions at finer spatiotemporal resolutions (Lauvaux et al., 2016; Mitchell et al., 2018) as well as uncertainties in gridded emissions (Andres et al., 2016; Gately and Hutyra, 2017; Hogue et al., 2016; Oda et al., 2018). Dramatic increases in emission uncertainties are associated with finer scales, with these uncertainties being mostly biases due to different methods disaggregating national-level emissions (Marland, 2008; Oda and Maksyutov, 2011). For instance, emission uncertainties of 20 % at regional scales increased to 50–250 % at city scales even for the northeastern United States (Gately and Hutyra, 2017), an area that is considered relatively "data-rich".

Given the large differences/discrepancies in emission inventories at urban scales, the use of atmospheric *top-down* constrain could be helpful for quantifying urban emissions and possibly providing a monitoring support (Pacala et al., 2010). Observed concentrations used in the top-down approach can often be obtained from ground-based instruments (Kim et al., 2013; Mallia et al., 2015; Wunch et al., 2011) and aircraft observations (Gerbig et al., 2003; Lin et al., 2006). Each type of measurement offers valuable information and has both advantages and disadvantages. Most ground-based measurements provide reliable, continuous CO_2 concentrations from fixed locations/heights. Unfortunately, current ground-based observing sites are too sparse to constrain urban emissions around the globe. Most National Oceanic and Atmospheric Administration (NOAA) sites are designed to measure background concentrations and few others aim at measuring concentration changes from few vertical levels within the planetary boundary layer (PBL). Other than a few notable examples (Feng et al., 2016; Lauvaux et al., 2016; Mitchell et al., 2018; Verhulst et al., 2017; Wong et al., 2015; Wunch et al., 2009), near-surface CO_2 measurements may not be available over many other cities around the world. Alternatively, airborne measurements from field campaigns provide better vertical and regional coverages (Cambaliza et al., 2014). Yet, continuous airborne operations over months to years are often impractical due to limited resources, which limits researchers' capability to track temporal variability of anthropogenic carbon emissions (Sweeney et al., 2015).

The carbon cycle community has entered a new era with advanced carbon-observing satellites—i.e., Greenhouse gases Observing SATellite (GOSAT; Yokota et al., 2009), TanSat (Liu et al., 2013) and Orbiting Carbon Observatory (OCO-2) satellite (Crisp et al., 2012)—routinely in orbit to measure variations of atmospheric column-averaged CO_2 mole fraction (XCO_2). Although most carbon-observing satellites have revisit times of multiple days (e.g., 3 days for GOSAT and 16 days for OCO-2), their global coverage, large number of retrievals and multi-year observations can fill the gaps of sparse surface observing networks. Eventually, space-borne instruments will help reduce emission uncertainties and benefit urban emissions analysis, especially over cities with no surface observations (Duren and Miller, 2012; Houweling et al., 2004; Rayner and O'Brien, 2001).

Previous studies have demonstrated the potential for detecting and deriving urban CO_2 emission signals from satellite CO_2 observations, in the form of XCO_2 enhancements above the background, without making use of much atmospheric transport information (Hakkarainen et al., 2016; Kort et al., 2012; Schneising et al., 2013; Silva and Arellano, 2017; Silva et al., 2013).

However, the linkage between their derived urban CO_2 emission signals and upstream sources is tenuous, as downwind XCO_2 can be enhanced by not only near-field upwind urban activities (e.g., traffic, houses, and power plants/industries), but regional-scale advection of upwind sources/sinks as well. Simulations using transport models are able to isolate the portion of satellite observations influenced by urban regions from the portion affected by natural fluxes or long-range transport (e.g., Ye et al., 2017). Therefore, accurate knowledge of atmospheric transport is essential in top-down assessment. As importantly, transport modeling is a necessary step within inverse modeling, which can help improve fossil fuel emission estimates and shed lights on CO_2 emissions monitoring network (Kort et al., 2013; Lauvaux et al., 2009). Uncertainties in transport modeling have been identified as a significant error source that affects inferred surface fluxes (Peylin et al., 2011; Stephens et al., 2007; Ye et al., 2017). Yet, by increasing the examined satellite overpass numbers, uncertainties from atmospheric inversions due to non-systematic transport errors in emission estimates can be reduced and confined to 5 and 15 % over Los Angeles and Riyadh (Ye et al., 2017).

Two main approaches can be considered for atmospheric transport modeling. Eulerian models, in which fixed grid cells are adopted and CO_2 concentrations within the grid cells are calculated by forward numerical integrations, have been widely utilized and customized to understand urban emissions and quantify model uncertainties over specific metropolitan regions worldwide (Deng et al., 2017; Lauvaux et al., 2013; Palmer, 2008; Ye et al., 2017). The Lagrangian approach, especially the time-reversed approach in which atmospheric transport is represented by air parcels moving backward in time from the measurement location ("receptor"), is efficient in locating upwind sources and facilitating the construction and calculation of the "footprint" (e.g., Lin et al., 2003) or "source-receptor matrix" (Seibert and Frank, 2004)—i.e., the sensitivity of downwind CO_2 variations to upwind fluxes.

In particular, the receptor-oriented Stochastic Time-Inverted Lagrangian Transport (STILT) model, one of the Lagrangian Particle Dispersion Models (LPDM), has the ability to more realistically resolve the sub-grid scale transport and near-field influences (Lin et al., 2003). STILT has been used to interpret CO_2 observations within the PBL (Gerbig et al., 2006; Kim et al., 2013; Lin et al., 2017) and, in recent years, to analyze column observations, i.e., XCO_2 (Fischer et al., 2017; Heymann et al., 2017; Macatangay et al., 2008; Reuter et al., 2014). Among STILT-based column studies, most aim at either natural CO_2 sources and sinks like wildfire emissions and biospheric fluxes, or anthropogenic emissions at regional or state scales. Very few studies focus on city-scale $FFCO_2$ using column data and LPDMs. Moreover, when applying LPDMs to interpret column CO_2 data, three key issues have hitherto yet to be carefully examined and will be addressed in this paper:

I. Uncertainty of modeled XCO_2 enhancements due to model configurations. Very few studies have examined model uncertainties resulted from model configurations—i.e., receptors and particles in LPDMs. Reuter et al. (2014) suggested negligible uncertainty on modeled biospheric XCO_2 due to STILT setups. Sensitivity tests have been performed regarding the uncertainty caused by STILT particle number releasing from one fixed level on modeled wildfire CO concentration (Mallia et al., 2015). However, when it comes to representing an atmospheric column using particle ensemble and modeling anthropogenic enhancements, minimal guidance on setup of LPDMs for modeling XCO_2 has been provided in prior studies.

II. Lack of detailed analysis on impact of transport errors on XCO_2 simulations. Flux inversions, e.g., Bayesian Inversion (Rodgers, 2000) involving LPDMs have been widely adopted to constrain emissions. Transport errors are ignored or assumed to be diagonal in some inversion studies, which leads to conclusions that are biased or overly optimistic (Lin and Gerbig, 2005; Stephens et al., 2007). Although approaches on quantifying transport errors have been proposed (Gerbig et al., 2003; Lin and Gerbig, 2005), very few studies (Lauvaux and Davis, 2014; Macatangay et al., 2008) focus on quantifying transport errors when interpreting XCO_2 data. Transport uncertainty of LPDMs and its impact on inverse estimates of $FFCO_2$ emissions should undergo more sophisticated calculations and evaluations.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

III. Determining background XCO_2 and characterizing its uncertainties. Here we define background value as the CO_2 "uncontaminated" by fossil fuel emissions from the city of interest. As urban emission signals are defined as the enhancements of XCO_2 over the background, errors in the background value introduce first-order errors into the derived urban XCO_2 signal from total XCO_2, with such errors propagating directly into fluxes calculated from atmospheric inversions (e.g., Göckede et al., 2010). Consequently, background determination is another critical task. One method in determining model boundary conditions of various species using LPDMs is the "trajectory-endpoint" method that establishes the background based on CO_2 extracted at endpoints of back trajectories based on modeled regional/global concentration fields (Lin et al., 2017; Macatangay et al., 2008; Mallia et al., 2015). Most of these studies aim at extracting relatively large CO_2 changes at a fixed level within the PBL or due to large emissions such as of wildfire, with relatively large signal-to-noise ratio. However, for studying XCO_2 that is less variable than near-surface CO_2 (Olsen and Randerson, 2004), potential errors in modeled concentration fields and atmospheric transport may pose more significant adverse impact on derived urban signals. Other ways of defining background include geographic definitions (Kort et al., 2012; Schneising et al., 2013) and simple statistical estimates (Hakkarainen et al., 2016; Silva and Arellano, 2017). In this study, we compare simple statistical background estimates against a new background determination method that combines OCO-2 observations and the STILT-simulated atmospheric transport.

In this paper, we attempt to address the aforementioned issues by extending STILT with column features and comprehensive error analyses, referred to as the column-STILT, "X-STILT". We illustrate the model's applications in extracting urban XCO_2 signals from OCO-2 retrievals (Fig. 1) and evaluate model performances via a case study focusing on Riyadh, Saudi Arabia. Riyadh with population of over 6 million by 2014 (WUP 2014) is chosen as the city of interest due to its low cloud interference, limited vegetation coverage, and isolated location in a barren area, which leads to higher data recovery rates and facilitates the background determination. Saudi Arabia has the largest CO_2 emission among Middle Eastern countries and ranks eighth globally in 2016 (Boden et al., 2017; BP, 2017; UNFCCC, 2017). We examine several satellite overpasses and focus on a small spatial domain adjacent to Riyadh for each overpass.

2 Data and methodology

2.1 STILT-based approach for XCO_2 simulation ("X-STILT")

Before demonstrating model details, Fig. 1 highlights several X-STILT characteristics, e.g., the account of column transport errors, background XCO_2 approximations, and the identification of upwind emitters using backward-time runs from column-receptors.

Sensitivities of the satellite sensor towards different parts of the atmospheric column are characterized by the averaging kernel profiles, which evaluates the relative portion between the "true" profile ($CO_{2,true}$) observed by satellite and prior profile ($CO_{2,prior}$). The averaging kernel in OCO-2 product is the product of OCO-2's normalized averaging kernel (AK_{norm}) and pressure weighting (PW) function. Column AK_{norm} peaks near the surface and exhibits values near unity throughout most of the troposphere (Boesch et al., 2011). Less-than-unity AK_{norm} values are mainly found aloft, which requires additional information from prior CO_2 profiles (Fig. 2a). To yield "apple-to-apple" comparisons against OCO-2 retrieved XCO_2, modeled column-averaged CO_2 concentrations should be properly weighted using satellite's weighting functions (Basu et al., 2013; Lin et al., 2004). Thus, the final simulated XCO_2 ($XCO_{2.sim.ak}$) are weighted between model-derived CO_2 profiles and OCO-2 a priori profiles (O'Dell et al., 2012):

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

where I is the identity vector and n stands for X-STILT release levels (Sect. 2.1.1). The first summation on the right-hand side in Eq. (1) is further comprised of the modeled XCO_2 enhancements due to $FFCO_2$ emissions (Sect. 2.1.2) and background XCO_2

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

(Sect. 2.4). Since X-STILT levels may not necessarily match the prescribed 20 levels in the retrieval, we perform interpolations on AK_{norm}, PW and $CO_{2,prior}$ from retrieval levels to model levels. Specifically, satellite profiles are treated as continuous functions and then linearly interpolated to model levels (Fig. 2). We then weight the linearly interpolated PW functions through some scaling factors—ratios of the pressure difference between adjacent model levels over that between adjacent retrieval levels (Fig. 2b).

2.1.1 X-STILT setup ("column receptors")

The linkage between the observed XCO_2 concentration by a given OCO-2 sounding and upwind carbon sources and sinks is determined by atmospheric transport. We adopt the STILT model to describe this connection. Fictitious particles, representing air parcels, are released from a "receptor" (location of interest) and are dispersed backward in time. The Lagrangian air parcels within STILT are transported along with the mean wind ($\bar{\boldsymbol{u}}$), turbulent wind component ($\boldsymbol{u}'$), and other meteorological variables, which are derived from Eulerian meteorological fields. In this study, we used meteorological fields simulated by the Weather Research and Forecasting (WRF; Skamarock and Klemp, 2008) and the 0.5° × 0.5° Global Data Assimilation System (GDAS; Rolph et al., 2017; Stein et al., 2015). When WRF fields were available, we performed two experiments of model simulations using two sets of meteorology fields—i.e., GDAS and WRF (Fig. 4). Proper WRF configurations including nesting setup, schemes for PBL, and surface layer for Riyadh have been carefully evaluated in Ye et al. (2017). Our primary focus is to assess the resulting errors given the choice of a particular wind field (i.e., GDAS 0.5°), rather than to carry out analyses of differences between WRF and GDAS.

To represent the air arriving at the atmospheric column of each OCO-2 sounding, we release air parcels from multiple vertical levels, "column receptors" (Fig. 3e), at the same lat/lon as satellite sounding at the same time and allow those parcels to disperse backward for 72 hours (see Appendix D2 for model impact from backward durations). To reduce computational cost, the air column is only simulated up to a certain height (hereinafter referred to as the maximum release height of air parcels above ground level in meters—MAXAGL). Sensitivity tests are performed regarding different setups of column receptors (Sect. 2.5). Our goal is to compare an overall modeled versus observed anthropogenic signals within a small latitudinal range for each overpass (Sect. 3.5.1). We select dozens of soundings with quality flags equal zero (QF=0) that implies selected observations have passed the cloud and aerosol screens (with removal of albedo > 0.4), and their retrievals have converged (Mandrake et al., 2013; Patra et al., 2017). Specifically, about 10–20 soundings are selected for simulations over every 0.5° latitude.

2.1.2 Modeling changes in XCO_2 ("column footprints" × fluxes)

Air parcels undergo vertical mixing within the PBL with concentrations modified by surface emissions. The sensitivity of a measurement site's concentration to potential upwind fluxes is referred to as the "source-receptor matrix" (Seibert and Frank, 2004), or, equivalently, the "footprint" (Lin et al., 2003). Longer the time an air parcel p spends ($\Delta t_{p,i,j,k}$) in a grid volume (i, j, k) from the ground to a surface dilution height h over timestamp t_m, higher its footprint value f will be (Lin et al., 2003):

$$f(x_r, t_r | x_i, y_j, t_m) \propto \frac{1}{N_{tot}} \sum_{p=1}^{Ntot} \Delta t_{p,i,j,k}, \quad (2)$$

where N_{tot} denotes the total number of air parcels. (x_r, t_r) and (x_i, y_j) describe model receptors and potential upstream sources, respectively. The vertical dilution height h is set to be half of the PBL heights (z_i); and footprint value is insensitive to h (Gerbig et al., 2003). Then, multiplying the footprint by certain gridded emissions yields change in CO_2 at a downwind receptor. Since we deal with column-averaged CO_2 concentrations, modified changes in XCO_2 take the form of,

$$XCO_{2,source/sink} = E(x_i, y_j, t_m) \times \sum_{n=1}^{nlevel} f_w(x_{n.r}, t_{n.r} | x_i, y_j, t_m), \quad (3)$$

where E broadly represents CO_2 emissions or fluxes (Sect. 2.3) and $(x_{n.r}, t_{n.r})$ denotes the column receptors. We define the entire term of $\sum_{n=1}^{nlevel} f_w$ as the "weighted column footprint" that describes the sensitivity of changes in column concentration due to potential upstream sources/sinks and incorporates satellite profiles:

$$f_w(x_{n.r}, t_{n.r}|x_i, y_j, t_m) \propto \frac{1}{N_{tot}} \sum_{p=1}^{Ntot} \Delta t_{p,i,j,k} \times AK_{norm}(n,r) \times PW(n,r). \quad (4)$$

For instance, modeled XCO_2 enhancements due to $FFCO_2$ derive from the convolution of spatially-varying f_w and ODIAC emissions (Sect. 2.3.1). Also, we account for modeled uncertainties that includes errors in prior $FFCO_2$ emissions (Sect. 2.3.1), receptor configurations (Sect. 2.5), and atmospheric transport (Sect. 2.6).

2.2 OCO-2 retrieval XCO_2 and data pre-processing

The OCO-2 algorithm of retrieving XCO_2 from radiances employs an optimal estimation approach (Rodgers, 2000) involving a forward model, an inverse model, and prior information regarding the vertical CO_2 profiles (O'Dell et al., 2012). We used the bias-corrected XCO_2 values from OCO-2 Lite files (version 7R; OCO-2 Science Team/Michael Gunson, Annmarie Eldering, 2015). Measurements over Riyadh were all carried out in land Nadir mode. We then selected 5 OCO-2 overpasses during the time period of Sept 2014–Dec 2016, based on four stringent criteria (see Appendix A). For smoothing noisy observations, we binned up the observed XCO_2 with QF=0 according to the lat/lon of model receptors (served as the midpoints of each bin) and calculated the mean and standard deviation of screened measurements within each bin. Next, background values were defined (Sect. 2.4) and subtracted from the bin-averaged observed XCO_2 to estimate increase in observed XCO_2 (step 3 in Fig. 1). The impacts of different bin-widths on bin-averaged observed signals are shown in Appendix D1. Total observed errors comprise the spatial variation of observed XCO_2 in each bin, background uncertainties, and retrieval errors obtained from the Lite file.

2.3 Sources of information for CO_2 fluxes

2.3.1 Fossil fuel emission (ODIAC) and prior emission uncertainties

We used the latest (year 2017) version of the Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC2017 dataset, Oda et al., 2018; Oda and Maksyutov, 2011, 2015) with fossil fuel CO_2 emissions at 1×1 km resolution on the monthly scale (Fig. 4). ODIAC starts with annual national emission estimates by fuel types from the Carbon Dioxide Information Analysis Center (CDIAC, Andres et al., 2011), which are then re-categorized into specific ODIAC emission categories on monthly basis, i.e., point source, non-point source, cement production, international aviation and marine bunker (Oda et al., 2018). Because CDIAC only covers years up to 2013, ODIAC extrapolates emissions in 2013 for emissions in 2014 and 2015 based on BP global fuel statistical data (BP, 2017). Also, ODIAC estimates point sources emissions according to a global power plants database—the Carbon Monitoring and Action (CARMA) Database (Wheeler and Ummel, 2008), and collects and distributes non-point sources using an advanced nighttime lights dataset from the Defense Meteorological Satellite Program Operational Line Scanner (DMSP/OLS). The use of the nightlight dataset allows ODIAC to characterize the spatial patterns of the anthropogenic sources such as point sources, line sources and diffused sources. More details about ODIAC2017 are in Oda et al. (2018).

To estimate emission uncertainties around Riyadh, we followed a method similar to those reported in Oda et al. (2015) and Fischer et al. (2017). Three emission inventories derived from different methods are inter-compared: ODIAC, the Fossil Fuel Data Assimilation System (FFDASv2; Asefi-Najafabad et al., 2014; Rayner et al., 2010) and the Emission Database for Global Atmospheric Research (EDGARv4.2; http://edgar.jrc.ec.europa.eu; Janssens-Maenhout et al., 2017). Slightly different from the uncertainty estimation method proposed in Oda et al. (2015), the fractional uncertainty is solely characterized by the emission

spread (1-σ, among three inventories) and mean values (μ) of estimated emissions for each grid cell within a given region (10° N–40° N, 25° E–60° E; Fig. 5). Due to different spatial grid spacing among inventories, we aggregated ODIAC emissions from 1 km to 0.1° grid cell to match the resolution of EDGAR and FFDAS and used the emission spread as a proxy for the spatial emission uncertainty. Because the proxy data do not change often in many emission inventories, emissions for the year 2008 (when all three inventories were readily available) were used for uncertainty calculations. Ultimately, fractional emission uncertainties and ODIAC emissions are convolved with X-STILT's weighted column footprints to provide the XCO_2 uncertainties due to prior emission uncertainties. Results of XCO_2 uncertainties due to emissions are shown in Sect. 3.4.2 and Sect. 3.5.2.

2.3.2 Natural fluxes (CarbonTracker)

Although $FFCO_2$ emissions are our focus, other potential confounding factors, e.g., the oceanic and terrestrial biospheric fluxes, affecting atmospheric CO_2 are also accounted for. Both fluxes are derived from a 3-hourly 1°×1° product—the CarbonTracker-NearRealTime (CT-NRTv2016 and v2017, http://carbontracker.noaa.gov). CT-NRT, an extension of the standard CarbonTracker (Peters et al., 2007) is designed for the OCO-2 program and uses different prior flux models and "real-time" ERA-Interim reanalysis in its transport model than regular CT, which allows for more timely model results. To calculate oceanic and biospheric XCO_2 changes, we multiplied these natural fluxes with column weighted footprint according to Eq. (3). Although wildfire emissions can greatly modify atmospheric XCO_2 (Heymann et al., 2017), we expected relatively small XCO_2 contributions from wildfire and hence excluded wildfire-elevated XCO_2 estimations, considering the times (mostly wintertime overpasses) and study region (the Middle East) in this study.

2.4 Background XCO_2

Determination of background XCO_2 is crucial, as it can significantly affect the magnitude of inferred observed anthropogenic signals. If the background is underestimated, then the detected signal may be overestimated, and vice versa. In this study, we seek to develop best-estimated background values given five tracks, where 3 methods are proposed and investigated as follows,

M1. A "trajectory-endpoint" method by assigning CO_2 values extracted from global models (i.e., CT-NRT) to trajectory endpoints plus simulated biospheric, oceanic components (Sect. 2.4.1);

M2. Statistical methods estimated solely from XCO_2 observations based on two previous studies (Sect. 2.4.2);

M3. An "overpass-specific" background requires model-defined urban plume and measurements outside the plume (Sect. 2.4.3).

We devote considerable efforts to compare the aforementioned three ways (Sect. 3.3) and investigate the background impact on model-data comparisons and emission estimates (Sect. 4.2). We choose M3-based background for this study as it is designed specifically for examining a particular city and specific overpasses downwind of the city.

2.4.1 Trajectory-endpoint method (M1)

Since the background is the portion of the atmosphere that is not "contaminated" by urban emissions, modeled background XCO_2 comprises modeled boundary condition confined by four-dimensional CO_2 fields from CT-NRT and contributions from biospheric fluxes, oceanic fluxes, and OCO-2 prior profiles (M1 in Fig. 1). Specific for modeling CO_2 boundary condition, CO_2 values for upper levels above MAXAGL are estimated based on CO_2 from their adjacent CT levels. And, averaged CT CO_2 values at trajectory endpoints are used for boundary conditions at model release levels (Fig. 2c). Then, modeled boundary conditions at vertical levels

are weighted accordingly via OCO-2's averaging kernel. Model-derived background values depend on the duration of hours backward (*nhrs*, increasing in the backward direction) in the model and are expected to stabilize as *nhrs* increases.

However, potential uncertainties in transport may strongly influence the distribution of Lagrangian parcels as backward duration time *nhrs* increases and may lead to potential spatial mismatch of the background region. Furthermore, potential biases and relatively coarse resolution of 2°× 3° for the global CarbonTracker may add inaccuracies to CO_2 values at trajectory-endpoints.

2.4.2 Statistic method (M2)

Hakkarainen et al. (2016) (referred to as M2H) extracted XCO_2 anomalies over the Middle East based on the daily median of screened measured XCO_2 with QF=0 and WL<15 within a relatively broad region (0° N–60° N, 15° W–60° E). Their detected anthropogenic signals vary from 1–2 ppm over 0.5° × 0.5° gridcells near Riyadh. Silva and Arellano (2017) (referred to as M2S) used measurements within a 4°×4° combustion region centered around the "urban and dense settlements" inferred from an anthropogenic biomes dataset ("anthromes", Ellis and Ramankutty, 2008). Then, M2S derived background as the mean minus one standard deviation of available observations within their studied urban extents.

Both statistical methods are highly efficient in estimating background values but can be limited to certain applications. For instance, M2H may be less suitable for determining background values when zooming into specific cities. Measurements within a broad spatial domain are lumped together, regardless of their locations (whether over rural or urban areas) and atmospheric transport. Silva and Arellano (2017) have pointed out that their examined 4°×4° combustion region may be too coarse for studying urban emissions and is suitable for studying the "bulk" characteristics. Also, M2S may be less applicable when multiple observed peaks are tangled together caused by emitters near the target city/megacity (e.g., clusters of cities in China). Therefore, without incorporating much atmospheric transport information, accuracies in the transport from an urban center to the downwind satellite overpass cannot be guaranteed. It may be difficult for either statistical method to locate the exact XCO_2 peak elevated by target city or background region. These difficulties motivate us to introduce a third approach in the next subsection.

2.4.3 Overpass-specific background (M3)

A few other studies defined the background values as the averaged observed XCO_2 values over a "clean" upwind region that is "uncontaminated" by urban emissions. For instance, Kort et al. (2012) and Schneising et al. (2013) defined the "clean" region based on geographic information (e.g., rural area to the north of LA basin). Although OCO-2 has relatively narrow swaths, transport models can be used to differentiate the enhanced versus background portions along the overpass. For example, Janardanan et al. (2016) calculated background XCO_2 as the averaged observed XCO_2 among gridcells with modeled anthropogenic signals < 0.1 ppm. This 0.1 ppm threshold is determined from the average simulated fossil fuel abundance over desert areas worldwide using the FLEXible PARTicle dispersion model (FLEXPART; Stohl et al., 2005), a model that is similar to STILT in that both are time-reversed LPDMs.

We present an alternative method using a forward-time run from an urban box to reveal the urban influence on satellite soundings, which are more straightforwardly and efficiently than solely relaying on backward-time runs. Fictitious particles are released from a box around the city center as a feature implemented with STILT (T. Nehrkorn, personal communication) to track air parcels over a city and the transport of the urban plume (Fig. 1). Specifically, the model continuously releases air parcels over a 30-minute window from a box surrounding Riyadh (extending ± 0.2° from the city center), with multiple 30-minute releases of 1000-particle ensembles over the 10-hours ahead of the satellite overpass hour (00–10UTC). Then, an urban plume can be derived from the parcels' distribution during the ~3 minutes OCO-2 passing window (Fig. 6a). Note that air parcels are tracked forward in

time for 12 hours, allowing for equal contributions from parcels released initially from different time intervals (every 30 mins) onto defined urban plume. We are aware of potential model error and its adverse impact on defined urban plume. Therefore, we further accounted for this influence, where wind error components are added in the forward run, resulting in additional spread in the air parcels (Sect. 2.6). In addition to random errors, systematic wind errors may affect the model's capability in simulating the urban plume. Yet, we did not see large biases for Riyadh and no large latitude shifts in the modeled and observed urban enhancements for most examined events. See Sect. 2.6 for more discussions.

Next, two-dimensional kernel density estimation (Venables and Ripley, 2002) is applied to determine the boundary of city plume based on the air parcels' distributions. We normalized the two-dimensional kernel density by its maximum value and "sketch" the boundary of the city plume based on a threshold of 0.05, which is sufficient to include most air parcels. No dramatic change in the shape/size of resultant urban plume was found when testing other thresholds < 0.05. The urban-influenced latitude range is defined as the intersection of the urban plume and OCO-2 overpass (Fig. 6a). Overall, the urban-influenced latitudinal band represented by 5 % of the maximum kernel density covers from 23.5° N–26° N, given multiple overpasses for Riyadh (Fig. A2).

The background latitudinal range unaffected by Riyadh's urban plume for estimating background then extend ~100 km from the north-most and/or south-most of derived urban plume (Fig. 6b). We abandon observations with latitudes > 26° N and < 23° N, because those retrievals are too scattered (Fig. 6b) and indicate a second peak during few other events (not for event on 12/29/2014). Eventually, the mean and spatial variation (i.e., the standard deviation) of screened observations (QF=0) over background ranges serve as best-estimated background value and background uncertainty, respectively (Fig. 6b). If the near-field wind vectors point more towards the north, screened measurements over the southern background latitudinal range is utilized, vice versa. Over 80 soundings are used to estimate final background for each examined overpass.

2.5 Sensitivity analyses for X-STILT column receptors

The goal of carrying out sensitivity tests is to understand any systematic/random errors towards STILT simulations brought by receptor configurations. Under the premise of limited computational resources, proper column receptors are set up with allowable random errors. Specifically, we focused on 3 receptor/ensemble settings, i.e., the maximum release level (MAXAGL), the vertical spacing of release levels (dh), and the particle number per level (dpar). The combination of these parameters yields the total number of particles (NUMPAR) released from column receptors.

Instead of regenerating model trajectories (Jeong et al., 2013; Mallia et al., 2015), we adopted the bootstrap method to resample model ensembles. The bootstrap approach helps construct hypothesis tests and infer error statistics (Efron and Tibshirani, 1986). The initial sample size before the bootstrap should be sufficiently large to ensure the performance of this method and related statistics. Thus, we generated a "base run" of trajectories starting initially from 401 levels (from the surface to 10 km) with a vertical spacing of 25 m and 200 particles per level. For testing each parameter (MAXAGL/dh/dpar), we fixed the other two variables and randomly selected/resampled model particles for 100 times (allowing for repetitions). In other words, we obtained 100 new sets of trajectories along with 100 derived anthropogenic enhancements for each test. Basic statistics, i.e., mean values and standard deviations, describing the distribution of these anthropogenic enhancements, are used to infer systematic and random uncertainties, respectively (results in Sect. 3.1).

2.6 X-STILT column transport errors

Uncertainty in atmospheric transport modeling has been identified to significantly affect emission constraints (Cui et al., 2017; Gerbig et al., 2008; Lauvaux et al., 2016; Stephens et al., 2007). However, transport errors due to vertical wind profiles and PBL

height uncertainties are expected to be less sensitive for column-integrated measurements than in situ measurements (Lauvaux and Davis, 2014; Law et al., 1996). In this study, we only accounted for errors caused by deviations in horizontal wind fields.

Previous studies (Lin and Gerbig, 2005; Mallia et al., 2017) aimed at estimating transport error at one fixed level, whereas for XCO_2 we account for transport error in a column sense (i.e., column transport error). Macatangay et al. (2008) briefly explained the column transport error as the weighting of transport error variances with respect to pressures. Similarly, we treated each model level separately and calculated one CO_2 transport error per level, denoted as σ_{ε}^2 $(CO_{2.sim.ak.n})$, following Lin and Gerbig (2005). In short, an additional wind error component ($\boldsymbol{u_{\varepsilon}}$) is added to the mean wind ($\bar{\boldsymbol{u}}$) and turbulent wind component ($\boldsymbol{u'}$) that are embedded in normal STILT runs (Lin et al., 2003), to randomly perturb air parcels for each level. RMS errors of u- and v-component modeled wind, error correlation timescales and length scales describe the $\boldsymbol{u_{\varepsilon}}$ in space and time (more in Appendix B).

In addition to the random error component, we are aware of potential systematic wind errors in certain areas, e.g., positive wind speed bias reported over Los Angeles (Ye et al., 2017), and their impacts on both forward- and backward- time simulations. Specifically, wind biases can influence the accurate definition of urban plume (Sect. 2.4.3) or the model-data comparisons (leading to model-data discrepancies in XCO_2 shapes, Sect. 3.4.1). As an attempt to resolve these obstacles, X-STILT can incorporate a near-field wind bias correction (to both backward- and forward-time simulations). By rotating/shifting model trajectories (Fig. A4), this bias correction aims at "correcting" air parcel distributions and resultant footprints, given knowledge of the near-field wind biases can be properly interpolated. Unfortunately, only 2 radiosonde stations around Riyadh with only 3 vertical pressure levels within the PBL (and sometimes with missing data) may be insufficient to correctly interpolate the vertical wind biases. Cities with meteorological profiles sampling more levels within the PBL and higher temporal frequency in reporting observed vertical winds will be more suitable sites retrieve the near-field wind errors. Other methods include rotation and stretching of urban plumes derived from WRF-Chem (Ye et al., 2017), similar to the rotation of X-STILT air parcels, to quantify errors in wind directions and speeds. Deng et al. (2017) sought correction of wind biases in a sophisticated manner via data assimilation. Yet, the near-field correction within X-STILT can be potentially utilized in the future as a quick bias correction to the near-field wind in LPDMs, given more wind observations and relatively flat terrains. Therefore, we decided to 1) ignore the impact of wind bias on forward-trajectories defined urban plume due to minimal biases reported over Riyadh whose magnitudes fall well within the prescribed error components (Fig. A1; Ye et al., 2017), and 2) get around with the impact on model-data comparisons using a latitudinal integration (further in Sect. 3.5).

For each model level (n), we obtained two sets of parcel distributions—i.e., one without and one with the wind error component ($\boldsymbol{u_{\varepsilon}}$). Then, the difference in the spread of these two distributions, or mathematically the difference in the variances of derived CO_2 distributions among air parcels (Lin and Gerbig, 2005), serve as the XCO_2 uncertainty (in ppm) due to transport error. We tested describing this XCO_2 transport uncertainty based on either normal or log-normal statistics. Since transport error using log-normal statistics did not show very distinct improvement from that using normal statistics, we ended up adopting normal statistics for the consideration of benefiting inverse modeling. Because the parcel distribution with $\boldsymbol{u_{\varepsilon}}$ is more dispersed than the parcel distribution without $\boldsymbol{u_{\varepsilon}}$ (Fig. 7), the increase in CO_2 variance with $\boldsymbol{u_{\varepsilon}}$ from that without $\boldsymbol{u_{\varepsilon}}$ describes the transport error for each level. However, negative values of transport error can occasionally occur, due to statistical sampling from insufficient model parcel trajectories. To resolve this technical issue, we modified Lin and Gerbig (2005) by using a regression-based approach. A weighted linear regression slope is used to describe the increase in CO_2 variances and then estimate transport error. More descriptions about this regression-based method are demonstrated in Appendix B. Overall, transport errors at levels within the PBL are expected to be larger than those at higher levels that approach zero.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

Lastly, we weighted the vertical profiles of transport errors against OCO-2's weighting functions. Following the definition of modeled AK-weighted XCO_2 in Eq. (1), the weighted column transport error follows Eq. (5),

$$\sigma_{\varepsilon}^2(XCO_{2.sim.ak}) = \sum_{n=1}^{nlevel} \{w_n^2 \times \sigma_{\varepsilon}^2\,(CO_{2.sim.ak.n})\} + 2 \sum_{1 \le n < m \le nlevel} \{w_n \times w_m \times cov_{\varepsilon}(CO_{2.sim.ak.n}, CO_{2.sim.ak.m})\}, \quad (5)$$

where w_n denotes the product of AK_{norm} and PW at level n; and cov_{ε} represents the correlation of transport errors between every two levels n and m ($1 \le n < m \le nlevel$). To calculate a typical vertical error correlation length scale, we fit exponential variogram according to transport errors and their separation distances between levels. Results of transport error at each sounding and its latitudinal integration for each track are shown in Sect. 3.4.2 and Sect. 3.5.2.

3 Results

3.1 X-STILT sensitivity tests with column receptors

The sensitivity tests examined uncertainties in modeled anthropogenic enhancements influenced by different maximum release heights (MAXAGL), particle numbers per level (dpar), and vertical spacings between levels (dh). Figure 8 shows test results given a wintertime sounding on 12/29/2014 around Riyadh.

Anthropogenic enhancements increase as maximum model levels increase from 1–2 km and then approach to equilibrium (Fig. 8a). When MAXAGL is very small (< 2.5 km), the model fails to fully capture the CO_2 enhancements within the PBL and causes underestimations on the elevated XCO_2. Random errors reflect the stochastic nature of air parcels within X-STILT that results in small fluctuations in parcel distributions and resultant signals. Random uncertainties of XCO_2 signals given different MAXAGLs stay relatively the same. In this experiment, dpar and dh are fixed to 100 particles and 100 meters.

For testing particle number per level (dpar), MAXAGL is set to 6 km (well above the top of the PBL, see Appendix C for the choice of 6 km) and 8 different sizes of dpar comprise the simulated atmospheric column: 25, 50, 75, 100, 125, 150, 175 and 200 particles. No obvious bias is associated with mean XCO_2 enhancements. The decay rate of random error approximates to the square root of the growth rate of particle numbers (e.g., $\sqrt[2]{50/25}$, Fig. 8b). We ended up placing 100 particles for each level, as random errors do not change dramatically from 100 to 200 particles.

Additionally, we conducted two experiments using constant and uneven vertical spacings. Vertical spacing in the constant dh experiment ranges from 50 m to 1 km. Mean anthropogenic XCO_2 enhancements decrease as vertical spacing increases (Fig. 8c), likely because fewer release levels are insufficient to represent air parcels in a column and their interactions with surface emissions, especially under strong wind shear. Two additional cases with uneven vertical spacing below and above a "cutoff level" of 3 km (see Appendix C for the choice of 3 km) are performed: 1) different lower spacings with a fixed upper spacing of 500 m; 2) different upper spacings with a fixed lower spacing of 150 m. For example, for case 2) where upper spacings vary, their resultant means and spreads of anthropogenic enhancements are similar to those using a constant spacing of 150m (Fig. 8c). This result suggests that the lower spacing below the cutoff level matters mostly to model results, because most anthropogenic XCO_2 enhancements are confined within the PBL. Thus, column receptors are placed from 0–3 km with a spacing of 100m and from 3–6 km with a spacing of 500 m (Fig. 3e).

Above 3 model parameters are closely related and can be combined as the total particle number. Fractional uncertainty in modeled enhancement reduces as total particle number increases and is about 5 % given large amounts of particles > 12500 (Fig.

8d). Our choice of column receptors and particle numbers has no noticeable bias and a fractional uncertainty of about 8 %. Overall changes in X-STILT column receptors have a relatively small impact on modeled anthropogenic signals, which is consistent with the finding (for biospheric signals) in Reuter et al. (2014).

3.2 X-STILT column footprints and upwind emission contributions

Upstream source regions and their contributions to downwind air column can be identified as the "footprint" using backward-time simulations. Here we take one sounding/example at 24.4961° N on 12/29/2014, when southwestern winds dominated, to explain the differences in parcel distributions and footprint patterns derived from 500 m, 3 km, and multiple levels. Air parcels released at 500 m are associated with large footprints in the adjacent area of Riyadh (Fig. 3a-b). While parcels released from a higher level of 3 km travel much faster, where most parcels barely get entrained into the PBL (Fig. 3c), which yields very few contacts with the surface implied from small footprint values (Fig. 3d). When air parcels are released from multiple levels (Fig. 3e), the footprint derived from each level within PBL is weighted by pressure weighting function, which results in overall smaller footprint (Fig. 3f) than that derived from 500 m. Yet, column footprint covers a broader spatial domain than the footprint derived from any fixed level. The intention here is to illustrate the difference in upwind influences from a PBL-based tower-like measurement versus a column-integrated measurement (e.g., satellite). As expected, surface influence arriving at an air column can be one or few orders of magnitude smaller than that arriving at a given location. Consequently, CO_2 changes within the PBL are expected to be larger than column changes. If zooming into the near-field land surface, westerly winds dominated during the 12/29/2014 event. XCO_2 contribution maps indicate large contributions due to urban emissions of Riyadh (Fig. 9b-c, 9e-f) and small contributions from regions to the south of Riyadh (Fig. 9a, 9d), regardless adopted meteorological fields.

3.3 Comparisons between methods to calculate background XCO_2

We now compare the magnitudes and temporal variations of background values under 4 definitions (introduced in Sect. 2.4.1 to Sect.2.4.3), which are a trajectory-endpoint definition (M1), statistical definitions proposed as in Silva and Arellano (2017) and Hakkarainen et al. (2016) (M2S and M2H); and an overpass-specific definition using X-STILT and OCO-2 measurements over background region (M3). As Silva and Arellano (2017) have pointed out their 4°×4° urban extent may be too coarse for studying urban emissions and is suitable for studying the "bulk" characteristics, we only borrowed their statistical assumption (μ-σ). We used the same examined latitudinal range (see Sect. 3.5) as M3, for computing M2S-based background.

M2S and M3 calculate background values from observations at the local scales, whereas M1 and M2H derive background values based on modeled or observed XCO_2 over broad spatial domains. Thus, the temporal variation of background XCO_2 using M2S agree better with that using M3 (Fig. 8e). M2S- and M3-based background often converge with a difference of < 0.5 ppm for examined events. Background using M1 differs significantly from background values using other approaches and exhibits positive biases spanning from 0.5–1.5 ppm (Fig. 8e). Possible reasons of this discrepancy may be the accumulated transport uncertainties as backward duration increases and potential uncertainties in input concentration fields with relatively coarse resolutions.

Some limitations of M1, M2H and M2S have been discussed in previous sections (Sect. 2.4.1 and 2.4.2). Here we emphasis the advantages and limitations of our final choice—M3 method. Comparing against methods based solely on models or simple statistics, M3 incorporates both and accounts for near-field transport from the target city to downwind satellite soundings. The background values are calculated based on screened measurements away from the urban plume, which represent more of the localized background and are specific to each overpass. Meanwhile, as wind errors may introduce errors/biases in defined urban plume, we added a wind error component to broaden the urban plume (Sect. 2.4.3 and Sect. 2.6) that helps reduce the inclusion of

enhanced values in the background region. Therefore, M3 is designed for identifying the urban plume found within each track and defining an overpass-specific background.

On average, background derived from simple statistics (e.g., regional daily median from M2H) may be "biased" low by 0.56 ppm, comparing against the localized "overpass-specific" background using M3. While a bias of 0.56 ppm in the background XCO_2 may seem small, this relatively small discrepancy in background values, on average, can lead to larger differences in estimated observed urban signals and emission evaluations (Sect. 4.2).

3.4 Anthropogenic enhancements and associated uncertainties

The application of X-STILT to OCO-2 yields the observed anthropogenic enhancements above background values as well as the modeled enhancements based on weighted column footprint and ODIAC emissions (Fig. 1). Observed uncertainties include natural variations of observations in each bin, background uncertainties, and retrieval errors; simulated uncertainties comprise XCO_2 uncertainties due to column model configurations, atmospheric transport, and prior emissions.

3.4.1 Comparisons against OCO-2 XCO_2 at selected soundings

We compare both the magnitude and shape of modeled and observed anthropogenic enhancements along the track. Models using GDAS and WRF report relatively higher and lower XCO_2 compared against bin-averaged observations, respectively, for the 12/29/2014 overpass (Fig. 10). The model-to-model discrepancy between GDAS versus WRF can be attributed to their estimated surface influences. Specifically, parcels driven by GDAS are less dispersed along the meridional direction (Fig. 9b) than parcels driven by WRF (Fig. 9e), which yields relatively stronger surface influence from GDAS. As the satellite overpassed Riyadh, influences from the city attenuate and both observed and modeled XCO_2 approach to the background. In terms of the spatial XCO_2 shape, large enhancements > 1.5 ppm inferred from bin-averaged observations cover a wide range from 24.5° N–24.8° N; whereas model-derived enhancements are narrower (Fig. 10). Also, modeled versus observed enhancements exhibit a 0.1° latitudinal shift.

These small discrepancies in the enhancement widths and locations of XCO_2 peaks between observation and models can also be found in other events. Similarly, for events on 12/27/2014, 12/16/2015 and 01/15/2016, observed enhancements are more continuous and moderate; whereas GDAS-modeled enhancements are narrower and sharper (similar to the 12/29/2014 case, Fig. 9). Maximum modeled enhancements can reach > 5 ppm for few soundings on 12/16/2015 and 01/15/2016. On the contrary, weaker modeled enhancements with wider coverage along latitudes (as opposed to the 12/29/2014 event) are reported during the last event. Latitudinal shifts of XCO_2 peaks vary from 0.1°–0.4° among total 5 events.

Based on these findings, we suspect that mismatch in the model-data enhancement widths is primarily due to errors in wind speeds; while latitudinal mismatch in model-data XCO_2 peaks results from errors in the wind directions near the site. Simulations with strong near-field influences can be sensible to potential biases in wind speeds and directions. This challenge leads us to introduce a method that integrates the urban XCO_2 enhancements over a latitudinal band (Sect. 3.5) to reduce near-field sensitivity on model-data comparisons and emission evaluations.

3.4.2 Transport errors, prior emission uncertainties and observed uncertainties

Time-dependent transport uncertainties of XCO_2 depend partially on interpolated regional wind error statistics as well as the inhomogeneous distribution of urban emissions. RMS errors associated with the GDAS u- and v-component winds are mainly less than 2 m s^{-1} (Fig. A1) over relatively flat terrain around Riyadh, which is much smaller than values from previous studies over relatively more complex terrains (Henderson et al., 2015; Lin et al., 2017). Even though biases in GDAS u- and v-component

winds for each track can be positive or negative, overall biases in u- and v-component winds are relatively small, with absolute values close to zero, given a dozen tracks. That is, no obvious systematic error over times/tracks is found in GDAS wind field around Riyadh. Similarly, Ye et al. (2017) reported no bias in the transport for Riyadh using WRF-Chem.

Because of the spatial inhomogeneity in urban emissions, wider parcel distributions after randomization may have higher possibility in making contact with more emission sources than parcel distributions before perturbation. Take the overpass on 12/29/2014 as an example. Small transport errors can often be found over less polluted latitudinal range (< 24.3° N and > 24.9° N in Fig. 10). Then, transport errors start to increase as few randomized parcels tend to "hit" some emission sources, although mean enhancements are small, e.g., 24° N–24.5° N and 24.7° N–24.8° N in Fig. 10. Although air parcels at higher altitudes are also under perturbations, the change in parcel distribution may have very small impact on column transport errors due to little contact of parcels aloft with surface emissions. As a result, the transport error per sounding for the 12/29/2014 overpass ranges from 0.07–2.87 ppm (Fig. 10). Within the enhanced latitudinal band (Fig. 6b), the median of transport errors is 0.56 ppm. For the other tracks with more intense urban enhancements, maximum transport error per sounding can reach > 5 ppm and median of transport errors over polluted latitudinal band vary from 0.60–1.50 ppm among tracks.

Around Riyadh, fractional uncertainties in gridded emissions mostly range from ~60–130 % (Fig. 5). While gridcells with large fractional uncertainties >150 % can be found especially over North Africa, most of these gridcells possess very small estimated emissions. Our spatial fractional uncertainties in gridded emissions over the Middle East can be comparable to few studies focusing on the gridded emission uncertainties even over different regions. For instance, in the northeastern U.S., several commonly-used inventories differ by > 100 % over half of examined 0.1° gridcells (Gately and Hutyra, 2017). As a result, gridded emissions uncertainties around Riyadh (Fig. 5) contribute to XCO_2 uncertainties of 0.04–2.40 ppm per sounding for the 12/29/2014 overpass. The median of emission errors for soundings within the enhanced latitudinal band is about 0.66–1.41 ppm for all 5 events.

Retrieval errors are reported for each retrieved sounding according to OCO-2 Lite file and exhibit a Gaussian-like distribution with most frequent values of 0.45–0.5 ppm among 5 overpasses. These retrieval error variances are then averaged within each observed bin to obtain bin-averaged retrieval errors. Background uncertainty varies from 0.65 to 0.84 ppm, depending on how scattering the observations are over background latitudinal range in each case. A third observed error source depicts the natural variation of noisy observations in each averaged bin. Overall observed uncertainty associated with each sounding varies from 0.8–1.27 ppm. Worden et al. (2017) accounted for the natural variability in observed XCO_2, the measurement noise errors with error covariance within spatial domain of 100 km×10.5 km. They found the overall precision of a measurement (WL< 10) over the land is ~0.75 ppm. Our larger observed uncertainties per sounding may be attributed to no filter of observations using WL, different examined regions and time periods and inclusion of background uncertainty (for the purpose of inverse analysis).

On a per sounding basis, XCO_2 uncertainties due to atmospheric transport are comparable to or slightly higher than those caused by emission uncertainties. Both uncertainties are higher than observed uncertainties within the urban-enhanced region. Still, reductions in uncertainties are expected as sounding-level uncertainties are aggregated along the track (Sect. 3.5.2).

3.5 Latitudinally-integrated anthropogenic signals and uncertainties

Because shapes and locations for XCO_2 peaks between models and observations did not line up perfectly (Sect. 3.4.1), direct comparison of observed versus simulated anthropogenic enhancements at each sounding may lead to significant deviations within small latitudinal bins. Instead, we compared anthropogenic signals integrated over their corresponding latitudes for each overpass.

Firstly, we integrated bin-averaged observed or modeled anthropogenic enhancements (i.e., differences between total XCO_2 and overpass-specific background) along their latitudes. While multiple degrees of freedom are sacrificed by this integration, this calculation gains a larger benefit of potentially reducing the impact of near-field wind bias on emission evaluations, as long as the latitude band for aggregation is representative. Secondly, a representative latitudinal range for integration (e.g., ~24° N–25.2° N in Fig. 10) can neither be too wide to include a second XCO_2 enhancements due to emissions from nearby emitters other than emissions from the target city; nor be too narrow to exclude observed or modeled enhancements due to emissions of the target city. In addition, negative observed urban enhancements may occur when the bin-averaged total observed XCO_2 is slightly lower than background value. The occurrence of these negative values is partially caused by the natural variations of observed XCO_2 and background uncertainties, which have been included in the uncertainties related to observed signals (Sect. 3.4.2). To minimize the inclusion of those negative values, we start with the enhanced latitudinal range (e.g., 24.24° N–24.92° N in Fig. 6b) and further account for latitudinal mismatch in model-data XCO_2 peaks (e.g., 0.1° in Fig. 10). To further include urban enhancements over the "tails" outside the distinct XCO_2 peaks, we further extend previous latitudinal range by an assumed 20 % of its width on both sides. We tested percentages other than 20 % and found no large increase in estimated signals due to small enhancements outside the plume (Appendix D2). Changes in the angle between near-field wind direction and satellite overpass may fluctuate the width of enhanced latitudinal band as well as the final integration latitudinal ranges (i.e., 1.08°–2.27°). The mean latitudinal range for integration is about 1.56° (~173 km) over 5 tracks.

For integrating uncertainties over one overpass, error variance-covariance matrices can be built. We took transport error as an example (Fig. A5a). Diagonal elements comprise transport error variance per sounding with off-diagonal elements filled with error covariance between each two soundings/receptors. A typical correlation length scale of transport errors along the overpass is about 25 km by fitting exponential variograms (Fig. A5b), given the transport errors (further driven by plume structures) and our choice of grid spacing between examined receptors.

3.5.1 Comparisons of latitudinally-integrated anthropogenic XCO_2 signals

Modeled anthropogenic XCO_2 signals range from 0.83–2.32 ppm and observed signals vary from 1.01–2.16 ppm. Averaged over 5 overpasses, X-STILT using GDAS yields a mean anthropogenic signal of about 1.59 ppm that is comparable to the signal of 1.57 ppm detected by OCO-2. The magnitudes of latitudinally-integrated observed signals can be affected slightly by how observations are binned up or selected (Appendix D). The slope of the linear regression line for best-estimated model-data urban signals is over one. Specifically, aggregated modeled signals are slightly lower/higher than aggregated observed signals, especially for those small/large observed signals (Fig. 11). Although best-estimated model-data urban signals line up well with a close-to-unity correlation coefficient, the correlation coefficient is dampened by uncertainties in both observations and models. Thus, we performed Monte Carlo experiments for 5000 times to fit regression lines (Fig. 11) based on randomly sampled model-data XCO_2 signals given their best-estimations and uncertainties (assuming normal distributions). The median of correlation coefficients using overpass-specific background is 0.61 (Fig. 11a).

3.5.2 Uncertainties associated with simulated and observed signals

The random uncertainties due to the choices of column receptors/parcels are small. For the one sounding showed in Fig. 8d, our model setup is associated with ~9 % on the modeled anthropogenic signal. If assuming this fraction ensemble uncertainty does not change for other examined soundings and no correlation between soundings, latitudinally-integration of this error source is < 0.02 ppm and contributes to an averaged 1 % of latitudinally-integrated modeled urban signal.

The latitudinally-integration of transport errors (σ) varies from 0.41–1.41 ppm for each track. The transport error variations can be attributed to a combined effect from several underlying factors: regional wind errors and correlation scales within the PBL, sounding-level transport error variance-covariance, the inhomogeneous distribution of anthropogenic emissions around a city and how air parcels interact with surface emissions. For the same soundings over Riyadh, our transport errors in ppm are comparable to those reported in Ye et al. (2017). Moreover, we aggregated these 5 sounding-level transport errors, assuming errors are independent, given the large separation time between the overpasses (one or several months for most cases). Therefore, the 68 % confidence limit (1σ) of transport error is 0.46 ppm, which is ~29 % of the averaged modeled signal (~1.59 ppm) over 5 events.

Similar steps for integration are taken for understanding XCO_2 uncertainties due to prior emissions (Sect. 2.3.1 and Sect. 3.4.2). As a result, the 68 % confidence limit of XCO_2 uncertainties due to emissions is about 0.40 ppm, leading to ~25 % of the mean urban signal (~1.59 ppm) over 5 tracks. Modeled uncertainties due to aggregate effects of all three error sources contribute to about 40.1 % of the mean urban signal, averaged over total 5 overpasses.

Retrieval errors between OCO-2 soundings are found to be correlated in both space and time, with correlation coefficients (for land nadir) of 0.45 and 0.31 as a function of satellite footprint and time, respectively (Worden et al., 2017). Uncertainties of bin-averaged observed XCO_2 share similar source as the background uncertainties, both of which rely on spatial variation in noisy observations (in each bin or over background region). Different types of observed uncertainties are assumed to be uncorrelated. Because observations along with their uncertainties have been binned up and the satellite footprints for bin-averaged observed uncertainties are hard to track, we only account for the temporal correlation of retrieval errors between every two soundings. As a result, total observed uncertainty per track vary from 0.30–0.44 ppm and the 68 % confidence limit of observed error is about 0.14 ppm, i.e., ~ 9 % of the mean observed signals (1.57 ppm) over total 5 overpasses.

4 Discussions

4.1 Model capabilities and performances

In this study, we demonstrate the coupling of forward- and backward-time Lagrangian particle dispersion model simulations within a new modeling framework ("X-STILT") and their applications in locating the urban plume, determining background XCO_2, identifying upwind sources, and estimating enhanced XCO_2 caused by sources/sinks (Fig. 1). Specifically, backward-time simulations over an atmospheric column connect upwind emission sources with downwind atmospheric columns and generate spatial maps of this connection with additional information from satellite retrieval profiles. Although forward-time simulations from an urban box are an alternative and optional portion of X-STILT, these simulations help gain information regarding the location and size of the time-varying urban plume (Fig. 6a) and locate downwind polluted range on a satellite overpass.

Model sensitivity tests suggest two main implications on simulating anthropogenic XCO_2 enhancements using LPDMs: 1) Receptor levels need to reach levels exceeding certain averaged PBL height to fully capture influences from surface emissions. 2) The model may capture a larger anthropogenic signal as number of levels increases. But, to minimize computational costs, one may try sparser and denser levels above and within a representative mean PBL height (the cutoff level) over upwind regions. Users can adopt their own setup of receptors in X-STILT according to combined results from sensitivity tests (Fig. 8d).

Additionally, X-STILT offers alternative solutions in dealing with errors in the meteorological fields, including regional random wind error perturbations and potential near-site wind bias corrections on model trajectories. For several satellite overpasses over Riyadh, models using WRF and GDAS are capable of capturing XCO_2 enhancements due to urban emissions, even though there remains small mismatch in the locations of model-data XCO_2 peaks. Model-to-model discrepancy between GDAS and WRF

in latitudinally-integrated urban signals is not large, benefiting from relatively flat terrain and similar interpolated terrain heights around Riyadh. No noticeable difference in overall wind error statistics (e.g., RMSE) derived from radiosonde comparisons with WRF versus GDAS is reported in this case. Thus, global meteorological fields such as 0.5° GDAS can be used for studying "flat cities" like Riyadh.

When dealing with enhancements in column concentration with small signal-to-noise ratio, careful examination to modeled background XCO_2 should be taken care of. X-STILT isolates the urban plume, and additional information from transport errors (with potential near-field wind bias correction) implemented in X-STILT accounts for uncertainty in model-defined urban plume and background latitudinal range not impacted by emissions of target city. Although one can possibly "eyeball" the city plume from observed XCO_2 (especially when a signal XCO_2 peak is visually distinctive), forward-time simulations with additional accounts for transport errors implemented in X-STILT may provide a more objective and efficient way (in that valuable human time is unnecessary) in figuring out the potential downwind sections along track that are affected by the city plume and extrapolating background region and its value. These advantages of overpass-specific background will become more important as more satellite tracks are incorporated within the analyses and future flux inversions.

4.2 Implications on error analysis and future inversion using LPDMs

Column transport uncertainties have not been rigorously examined in most XCO_2 studies employing Lagrangian particle dispersion models like STILT. In this study, we conducted comprehensive analysis towards, 1) observed errors including spatial XCO_2 variability, background and retrieval uncertainties; 2) simulated errors comprise uncertainties caused by model configurations, atmospheric transport and prior emissions. With the help of OCO-2, X-STILT is able to provide both the time- or latitude-dependent uncertainties (Sect. 3.4.2) and latitudinally-integrated uncertainties for each satellite overpass (Sect. 3.5.2).

On average, column transport uncertainties (with 68 % confidence limit) contribute to 29.1 % of the mean modeled urban signal over 5 overpasses. Still, transport error on a per track basis can be substantial with fractional transport uncertainties < 65 % for the first four overpasses but > 80 % for the last overpass. We accounted for transport error correlations between X-STILT release levels and sounding locations when calculating overall transport uncertainty on the per sounding and per track basis. For instance, transport error covariance between selected soundings contributes to about 70 % of the integrated transport errors, emphasizing the importance of error covariance on model evaluations (e.g., Lin and Gerbig, 2005).

To further demonstrate X-STILT's role in inverse modeling, we conducted a simple scaling factor inversion (Rodgers, 2000), based on 5 pairs of model-data urban signals. The prior emissions from ODIAC are assumed to be "unbiased", which yields a prior scaling factor of unity (λ_a = 1). The prior uncertainty (S_a) is simply one number representing the overall uncertainties of the sum and spatial spread of ODIAC emissions around Riyadh (further calculated from the inter-comparisons against FFDAS and EDGAR). Observational error covariance matrix (S_λ) includes the observed errors and latitudinally-integrated transport errors with a dimension of 5 × 5. Errors between every two overpasses are assumed to be independent. Our conservative results based on GDAS suggest that the posterior scaling factor ($\hat{\lambda}$) and its uncertainty ($\hat{S}$) is about 1.04 and 0.26, respectively, which are comparable to the WRF-Chem-based emission estimate in Ye et al. (2017), given 5 satellite tracks over Riyadh.

Estimated background uncertainty is represented by the spread of screened observations and may be reduced given large sampling size. However, potential errors in background XCO_2 defined by other methods may affect resultant observed signals. The regional daily median background may be "biased" low and the trajectory-endpoint derived background may be "biased" high, compared against the X-STILT overpass-specific background (Sect. 3.3). Background values using daily median or trajectory-endpoint methods can result in changes in mean observed emission signal by about + 0.68 ppm or – 0.77 ppm (Fig. 11b and 11d);

contributing to ~43 % increase and 49 % decrease in the mean observed signal using overpass-specific background (i.e., 1.57 ppm). Furthermore, even large impact on posterior scaling factor can be caused by using background derived from simplistic statistic. For instance, the posterior scaling factor ($\hat{\lambda}$) calculated using daily median background is larger than that using overpass-specific background by 68 %. These results again emphasize the significant role of background definitions played in estimated observed signals and emission estimates. In particular, simple statistical approaches without considering the atmospheric transport may lead to erroneous conclusions.

4.3 Limitations and future plans

We note that the modeled and observed anthropogenic signals reported in this work may represent more of the signals during the wintertime. No summertime anthropogenic signal has been examined or derived, because the lack of screened observations reported in OCO-2 Lite file over most summertime tracks (Fig. A1) are insufficient to either determine robust background values or group/bin noisy observations. We intended to select tracks with sufficient amount of measurements passed the quality check to ensure the robustness in derived background and observed signals.

Robust constraints on anthropogenic emission can be hampered by due to their alternating-sign nature and signals potentially comparable to anthropogenic emissions (Shiga et al., 2014; Ye et al., 2017), which are also inferred from tracks we modeled over Cairo with non-negligible biomass (results not shown in this paper). If examining summertime tracks, the large gradients from urban to rural due to local gradients in biospheric fluxes should be considered. Although biospheric fluxes or their resultant changes in XCO_2 concentrations are beyond the scope of this work, many studies have been working to address this challenge. Ye et al. (2017) incorporated biospheric fluxes from the North American Carbon Program (NACP) Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP; Fisher et al., 2016; Huntzinger et al., 2013) and performed downscaling on biospheric fluxes using MODIS-derived Green Vegetation Fraction (GVF), to provide high-resolution biospheric flux fields and estimated background XCO_2 by modeling. Besides, radiocarbon and terrestrial solar-induced chlorophyll fluorescence (SIF) data are helpful to isolate fossil fuel CO_2 and biospheric CO_2 (Fischer et al., 2017; Levin et al., 2003; Sun et al., 2017). In particular, recent studies have identified SIF as a better indicator/proxy of gross or net primary production than some other greenness indices over several different vegetation types (Shiga et al., 2018; Sun et al., 2017), which improves biospheric fluxes estimation in ecosystem models and benefits the interpretation of OCO-2/OCO-3 retrievals (Dayalu et al., 2017; Luus et al., 2017).

X-STILT extends its way to account for transport errors and particle statistics in a column sense within LPDMs. Admittedly, the transport error analysis and near-field correction may work the best with the assistance of denser meteorological observing networks to characterize the error structures of transport errors. Increasing the density of surface networks may modify the wind error statistics including the wind error variances and horizontal correlated length-scale, and further impact the model transport uncertainties and inversed fluxes. Yet, this shortcoming is not inherent to X-STILT and applies to other means of quantifying the transport errors based on real data as well. The trade-off of choosing a city in the Middle East like Riyadh to minimize cloud and vegetation influences is the relatively sparse observations of surface meteorological network or aircraft. The most recent OCO-2 version 8 Lite files include retrieved surface winds for each sounding. Unfortunately, most of those surface wind retrievals are not available over Riyadh, but the retrieved surface winds for other urban areas, if available, may be used for assimilation and assisting X-STILT error analysis.

Emission evaluations for different regions can be different and affected by different observational constraints. Therefore, more comprehensive Bayesian inversions on the spatially distributed emissions are needed, given more sampled satellite overpasses over Riyadh or more sampled cities over the Middle East. We expect the inclusion of more column observations in stationary (target)

modes, e.g., by scanning over megacities by OCO-3 (Eldering et al., 2016), which may offer more concrete spatial XCO_2 variabilities and more benefits in flux inversions. Many nations are devoting considerable resources in launching carbon-observing satellites that can potentially be coordinated in a larger monitoring system (Tollefson, 2016). Given that X-STILT can potentially work with most satellites (given satellite-specific information like AK, PW, CO_2 prior profiles, etc.), we expect enhanced capability in emission constraints of urban emissions by combining satellite data with X-STILT.

Code availability. Modifications on initial STILT code for X-STILT are archived on https://github.com/wde0924/X-STILT.

Appendices

Appendix A: Four conservative criteria to select overpasses over Riyadh

In short, we accounted for four factors, including 1) prevailing wind directions and downwind regions, 2) the portion of soundings with QF=0, 3) the distance between satellite track and the city center, and 4) regional wind errors in modeled meteorological fields.

I. First of all, we defined a broad spatial domain (23.5° N–26° N, 46.5° E–48° E) around the Riyadh city center and count the total sounding numbers that fall into this domain for each overpass. This spatial domain can be determined by examining prevailing wind directions and locating downwind regions based on wind rose plot from radiosonde stations at the city center and the airport of Riyadh (with 4-character international ID of OERK and OERY) during each overpass date. Alternatively, forward-time model runs starting from a box around the center of Riyadh allows us to determine polluted latitudinal ranges on satellite overpass (Fig. 1). Detailed demonstrations about forward-time runs can be found in Sect. 2.4.3. As a result, total 51 overpasses with at least one measurement fall into this designed spatial domain for Riyadh (Fig. A1).

II. Next, we ensured the amount of screened observed data using warn levels/quality flags (QF). Because high warn level is generally associated with high total aerosol optical depth inferred from soundings near Riyadh, we only used quality flag to control data quality in this study. After selecting overpasses with > 100 soundings with QF=0, about a third of total overpasses remain. Most spring- and summer- time tracks (during Mar–Aug) fail to satisfy this criterion (Fig. A1), primarily due to high aerosol loading and cloud contaminations.

III. Also, overpasses with distinct enhancements in retrieved XCO_2 due to urban emissions are preferred. Near-field domain affected by PBL processes may extend over 100–1000 km based on the globally averaged ventilation time for PBL (Lin et al., 2003). Due to relatively smaller SNR of extracting column signals and the preference for a more distinct XCO_2 peak, we made a conservative assumption on the impacted near-field domain, which is a circle with a radius of 50 km around the city center. Thus, we calculated the smallest distance between soundings and city center (e.g., 24.71° N, 46.74° E for Riyadh) and rule out additional 3 overpasses on 01/30/2015, 01/17/2016 and 02/18/2016.

IV. As a final step, because model results can potentially be affected by meteorological fields, we ruled out another 3 overpasses on 04/02/2015, 11/30/2016 and 12/02/2016, where the averaged regional u- and v- wind RMS errors below 3 km are above 2 m s^{-1} (Fig. A1). More details on calculating wind errors are in Sect. 2.6.

Appendix B: Wind error calculation and regression-based transport error method in X-STILT

In terms of the wind error component ($\boldsymbol{u_\varepsilon}$) mentioned in Sect. 2.6, two sets of parameters are used to describe, 1) σ_{uverr}, the standard deviation of horizontal wind errors (root-mean-squared errors, RMSE) describing to what extent should we randomly

perturb air parcels; and 2) horizontal and vertical length-scales and time-scales (Lx, Lz, and Lt) determining how wind errors are correlated and decayed in space and time. We calculated different sets of wind error statistics over 3 vertical bins, i.e., 0–3 km, 3–6 km and 6–10 km, for randomizing air parcels. To obtain σ_{uverr}, observed winds at mandatory levels (i.e., 925, 850, 700, 500, 400, 300 mb) from surrounding radiosonde sites (Fig. 4) are compared against WRF- or GDAS-interpolated winds. Then, we averaged wind errors at different mandatory levels over the aforementioned three vertical bins. In addition, wind errors are considered to be spatiotemporally correlated. To determine error correlation scales, differences in the wind errors are calculated and wind errors at different radiosonde stations or different reported hours (00UTC or 12UTC) are paired up based on their separation length- or time-scales. An exponential variogram is then applied to estimate the horizontal, vertical and temporal correlation scales, which are the separation scales when errors become statistically uncorrelated.

Technical issue: The CO_2 variance before the randomizations ($\sigma_{u'}^2$) can be larger than that after the randomization ($\sigma_{\varepsilon+u'}^2$) for few levels occasionally, due to insufficient parcel numbers (green dots in Fig. A3). Instead of flagging these data, we adopted a regression-based method for further accounting for those enhancements in CO_2 variance. Specifically, we applied linear regression lines to the two CO_2 variances before and after the randomizations, with weights of $1/\sigma_{\varepsilon+u'}^2$. That means the relatively large variance will be weighted less. We also tried several ways to apply linear regression (e.g., without the weights), resultant regression slopes can be extremely large, and the y-intercept can be negative, potentially leads to unreasonable large transport errors (in ppm) at lower levels within the PBL and negative transport errors aloft. Then, we scaled and recalculate $\sigma_{\varepsilon+u'}^2$ based on weighted regression slope S_{WLR} and $\sigma_{u'}^2$. The regression line indicates the overall increase in CO_2 variance that serve as transport error in ppm:

$$\sigma_{\varepsilon}^2\,(CO_{2.sim.ak.n}) = (S_{WLR} - 1) \times \sigma_{u'}^2\,(CO_{2.sim.ak.n}), \quad \text{(A1)}$$

where the weighted linear regression is fitted for variances with versus without wind error component (Fig. A3). And, extremely large anthropogenic enhancement (e.g., >1000 ppm) for a given parcel may exist for a few cases. Thus, outliers (i.e., the upper 1st percentile of both parcel distributions before and after the randomizations) are removed for each level, before calculating variances in both CO_2 distributions.

Appendix C: The determinations of MAXAGL and cutoff level

MAXAGL and a cutoff level (below which more model levels are placed) are the most important factors in determining modeled urban signals and can be determined based on few model trajectories starting from few satellite soundings for each overpass. Modeled PBL height reported for an individual air parcel at a timestamp, $z(p, t_m)$, can be very high over the upwind desert region near Riyadh. We determine MAXAGL to be the maximum PBL height for each individual air parcel. To determine a cutoff level, we calculate the averaged PBL heights over all parcels as a function of backward time, as follows

$$\overline{z_i}\,(t_m) = \frac{1}{N_{tot}} \sum_{p=1}^{N_{tot}} z_i(p, t_m), \quad \text{(A2)}$$

where t_m represents the backward timestamp, ranging from 0 to 72 hours back. The averaged modeled PBL heights among air parcels at each timestamp $\overline{z_i}\,(t_m)$ exhibit a diurnal cycle, where expected high values present during the daytime. Also, $\overline{z_i}\,(t_m)$ typically display relatively high values where parcels are more concentrated within a day backward, and low values as parcels disperse outwards few days back. We ended up using the highest value of mean PBL heights over parcels and over time as a representative cutoff level.

Following these algorithms, the maximum $z_i(p, t_m)$ and maximum $\overline{z_i}\,(t_m)$ are 5816 m and 2420 m, for the one sounding we showed in Fig. 8. Considering potential small uncertainties in modeled PBL heights (Zhao et al., 2009), we rounded these maximum

numbers to 6 km and 3 km as a representative MAXAGL and cutoff level. In addition, we generalize the rules for placing *column receptors* to other seasons, based on aforementioned calculations. Maximum $z_i(p, t_m)$ and maximum $\overline{z_i}\ (t_m)$ over the upwind region vary slightly for different soundings during different seasons. Typically, maximum $z_i(p, t_m)$ are mostly under 6 km for wintertime soundings in Dec, Jan, and Feb, but can reach ~7 km and 10 km for soundings in spring/fall and summer. The highest $\overline{z_i}\ (t_m)$ are less than 3 km for wintertime overpasses and ~4 km and 6 km for overpasses in spring/fall and summer. Therefore, *column receptors* are placed from the surface to 3 km with 100 m spacing and 3–6 km with 500 m spacing for wintertime overpasses with 100 parcels per level (Fig. 3e). For other seasons such as the summertime, additional receptors are placed from 6–10 km with 1 km spacing, to ensure the model captures entire contributions from surface emissions. Although we expect similar values for MAXAGL and cutoff level for most soundings over the Middle East due to overlaps in upwind regions, these values should be recalculated when other cities are examined (Eq. A2).

Appendix D: Factors that may influence observed or modeled enhancements/signals

In Section 3.5, we integrated XCO_2 enhancements along latitudes to estimate modeled and observed signals within a certain latitudinal band for each overpass. This latitudinal band starts with enhanced latitudinal range, then gets corrected based on model-data latitudinal shift in XCO_2 peaks, and finally extends by 20% of its length. Also, we tested the impact of different percentages other than 20 % on latitudinally-integrated signals. Because of relatively small XCO_2 enhancements over background range, the impact due to different percentages (i.e., 10 %, 15 %, 20 %, 25 %) are relatively small—i.e., with changes of 0.03 ppm and 0.06 ppm in averaged modeled and observed signals, respectively. These small changes self-explain that our integration latitudinal band is representative as it does not include a second peak or misses any large XCO_2 enhancements.

D1 Influences on observed signals (bin-widths and warn levels)

These modeled and observed signals reported in Sect. 3.5.1 are calculated based on the uneven sampling choice for model receptor lat/lon described in Sect. 2.1.1; i.e., with smaller bin withds of 0.025° and larger bin widths of 0.05° over which urban influences are stronger and weaker. In addition, we tested the impact on observed signals resulted from different bin widths with constant values starting from 0.01° to 0.5°. Both the latitudinal variation and the overall observed signal for an overpass generally decrease as bin widths increase, because bin-averaged observed XCO_2 enhancements get smoothed out, especially over latitudes with strong urban influences. Some information is lost in latitude-integrated observed signals based on our sampling choices when comparing against the signals calculated using constant bin widths such as 0.02°. Yet, binning observations based on the lat/lon of model receptors benefit a fair comparison with the model and our uneven sampling choices may better resolve XCO_2 enhancements within much finer grid spacing (particularly under urban influences) in the premise of limited computational resources.

In addition, warn levels (WLs) may impact the filtering of observed data, bin-averaged observed XCO_2, defined background and conclusion regarding the model-data comparisons. Based on three simply tests by selecting measurements with QF=0 and additional WL filters (i.e., WL < 10, 12, and 15), observed signals increase a little, as more conservative WL filtering is applied. Changes in linear regression slopes and correlation between best-estimated modeled and observed signals due to sample choices and WL filtering are small.

D2 Influences on modeled signals

An additional set of hourly scaling factors (Nassar et al., 2013) can be applied to ODIAC to downscale the monthly mean emissions down to hourly values. In this study, we use monthly mean $FFCO_2$ emissions from ODIAC and apply TIMES to only 1 of the total 5 overpasses. Simulations including TIMES are slightly larger than those without the hourly scaling factors. Also, numbers of

hours may impact the modeled enhancements at each sounding/receptor. We also conducted another simulation for 12/27/2014 event using model trajectories with only 24 hours back (different from 72 hours used in main text). The decrease in anthropogenic enhancements is < 0.05 ppm per sounding, which is small due to very small surface influence from far-away emission sources. Lastly, we report overall discrepancy in the modeled anthropogenic enhancements with or without weighting by OCO-2 prior profiles to be small. The difference is about 1–2 % of the weighted modeled anthropogenic enhancements, which is much smaller than impact caused by uncertainties in transport, emissions, or different setups. Note that XCO_2 portion from OCO-2's prior profile is zero and averaging kernel is simply unity everywhere for non-AK weighted simulations.

Acknowledgements. This work is based upon work supported by the National Aeronautics and Space Administration funding under Grant No. NNX15AI41G and the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260. We gratefully acknowledge Thomas Nehrkorn for providing modifications code to facilitate the forward-time box runs and thank Derek Mallia, Feng Deng, Arlyn Andrews, Andy Jacobson for their valuable advice on STILT-based modeling. The OCO-2 data were produced by the OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. The authors acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and/or READY website (http://www.ready.noaa.gov) used in this publication. CarbonTracker CT-NRT.v2017 results provided by NOAA ESRL, Boulder, Colorado, USA from the website at http://carbontracker.noaa.gov. The support and resources from the Center for High Performance Computing (CHPC) at the University of Utah are gratefully acknowledged.

References

Andres, R. J., Gregg, J. S., Losey, L., Marland, G. and Boden, T. A.: Monthly, global emissions of carbon dioxide from fossil fuel consumption, Tellus, Ser. B Chem. Phys. Meteorol., 63(3), 309–327, doi:10.1111/j.1600-0889.2011.00530.x, 2011.

Andres, R. J., Boden, T. A. and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission, Tellus B Chem. Phys. Meteorol., 66(1), 23616, doi:10.3402/tellusb.v66.23616, 2014.

Andres, R. J., Boden, T. A. and Higdon, D. M.: Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example, Atmos. Chem. Phys., 16(23), 14979–14995, doi:10.5194/acp-16-14979-2016, 2016.

Asefi-Najafabad, S., Rayner, P. J., Gurney, K. R., Mcrobert, A., Song, Y., Coltin, K., Huang, J., Elvidge, C. and Baugh, K.: A multiyear, global gridded fossil fuel emission data product: Evaluation and analysis of results, J. Geophys. Res. Atmos., 119(17), 10213–10231, doi:10.1002/2013JD021296, 2014.

Brasseur, G. P., and Jacob, D. J. Modeling of Atmospheric Chemistry. Cambridge University Press, 2017.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A. and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13(17), 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.

Boden, T. A., Marland, G., and Andres, R. J.: Global, Regional, and National Fossil-Fuel CO2 Emissions, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., USA, doi:10.3334/CDIAC/00001_V2017, 2017.

Boesch, H., Baker, D., Connor, B., Crisp, D. and Miller, C.: Global characterization of CO_2 column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 mission, Remote Sens., 3(2), 270–304, doi:10.3390/rs3020270, 2011.

BP: Statistical Review of World Energy, available at http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html, 2017.

Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L., and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, Atmos. Chem. Phys., 14(17), 9029-9050, doi:10.5194/acp-14-9029-2014, 2014.

Ciais, P., Sabine, C., Govindasamy, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R., Piao, S., and Thornton, P.: Chapter 6: Carbon and Other Biogeochemical Cycles, in: Climate Change 2013 The Physical Science Basis, Cambridge University Press, Cambridge, 2013.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Mandrake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D. and Yung, Y. L.: The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization, Atmos. Meas. Tech., 5(4), 687–707, doi:10.5194/amt-5-687-2012, 2012.

Cui, Y. Y., Brioude, J., Angevine, W. M., Peischl, J., McKeen, S. A., Kim, S. W., Neuman, J. A., Henze, D. K., Bousserez, N., Fischer, M. L., Jeong, S., Michelsen, H. A., Bambha, R. P., Liu, Z., Santoni, G. W., Daube, B. C., Kort, E. A., Frost, G. J., Ryerson, T. B., Wofsy, S. C. and Trainer, M.: Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The San Joaquin Valley, J. Geophys. Res. Atmos., 122(6), 3686–3699, doi:10.1002/2016JD026398, 2017.

Dayalu, A., Munger, W., Wofsy, S. C., Wang, Y., Nehrkorn, T., Zhao, Y., McElroy, M. B., Nielsen, C. and Luus, K.: VPRM-CHINA: Using the Vegetation, Photosynthesis, and Respiration Model to partition contributions to CO_2 measurements in Northern China during the 2005–2009 growing seasons, Biogeosciences Discuss., in review, 2017.

Deng, A., Lauvaux, T., Brian, Gaudet, Kenneth, James, Davis, Kevin, Robert, Gurney, Risa, Patarasuk, Robert, Michael, Hardesty, And, Alan and Brewer: Toward Reduced Transport Errors in a High Resolution CO_2 Inversion System, Environ. Sci. Technol., 5, 5–20, doi:10.1021/es3011282, 2017.

Dlugokencky, E. and Tans, P.: Trends in atmospheric carbon dioxide, National Oceanic & Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL), available at: http://www.esrl.noaa.gov/gmd/ccgg/trends, 2015.

Duren, R. M. and Miller, C. E.: Measuring the carbon emissions of megacities, Nat. Clim. Chang., 2(8), 560–562, doi:10.1038/nclimate1629, 2012.

Efron, B. and Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75, 1986.

Eldering, A., Bennett, M. and Basilio, R.: The OCO-3 Mission: overview of science objectives and status. In EGU General Assembly Conference Abstracts, 18, 5189, 2016.

Ellis, E.C. and Ramankutty, N.: Putting people in the map: anthropogenic biomes of the world. Frontiers in Ecology and the Environment, 6(8), 439-447, 2008.

Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., Díaz-Isaac, L. I., Duren, R. M., Fischer, M. L., Gerbig, C., Gurney, K. R., Huang, J., Jeong, S., Li, Z., Miller, C. E., O'Keeffe, D., Patarasuk, R., Sander, S. P., Song, Y., Wong, K. W. and Yung, Y. L.: Los Angeles megacity: A high-resolution land-atmosphere modelling system for urban CO_2 emissions, Atmos. Chem. Phys., 16(14), 9019–9045, doi:10.5194/acp-16-9019-2016, 2016.

Fischer, M. L., Parazoo, N., Brophy, K., Cui, X., Jeong, S., Liu, J., Keeling, R., Taylor, T. E., Gurney, K., Oda, T. and Graven, H.: Simulating estimation of California fossil fuel and biosphere carbon dioxide exchanges combining in situ tower and satellite column observations, J. Geophys. Res. Atmos., 122(6), 3653–3671, doi:10.1002/2016JD025617, 2017.

Fisher, J. B., Sikka, M., Huntzinger, D. N., Schwalm, C. and Liu, J.: Technical note: 3-hourly temporal downscaling of monthly global terrestrial biosphere model net ecosystem exchange, Biogeosciences, 13(14), 4271–4277, doi:10.5194/bg-13-4271-2016, 2016.

Gately, C. K. and Hutyra, L. R.: Large Uncertainties in Urban-Scale Carbon Emissions, J. Geophys. Res. Atmos., 242–260, doi:10.1002/2017JD027359, 2017.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S. and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res. Atmos., 108(D24), 4757, doi:10.1029/2003JD003770, 2003.

Gerbig, C., Lin, J. C., Munger, J. W. and Wofsy, S. C.: What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, Atmos. Chem. Phys., 6(2), 539–554, doi:10.5194/acp-6-539-2006, 2006.

Gerbig, C., Lin, J. C. and Körner, S.: Vertical mixing in atmospheric tracer transport models: error characterization and propagation, Atmos. Chem. Phys., 8, 591–602, doi:10.5194/acp-8-591-2008, 2008.

Göckede, M., Turner, D. P., Michalak, A. M., Vickers, D. and Law, B. E.: Sensitivity of a subregional scale atmospheric inverse CO_2 modeling framework to boundary conditions, J. Geophys. Res. Atmos., 115(D24), 2010.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Clais, P., Fan, S., Fung, I. Y., Gloor, M., Helmann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Park, B. C., Randerson, J., Sarmiento, J., Tuguchi, S., Takahashi, T. and Yuen, C.-W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415(6872), 626–630, 2002.

Hakkarainen, J., Ialongo, I. and Tamminen, J.: Direct space-based observations of anthropogenic CO_2 emission areas from OCO-2, Geophys. Res. Lett., 43(21), 11,400-11,406, doi:10.1002/2016GL070885, 2016.

Henderson, J. M., Eluszkiewicz, J., Mountain, M. E., Nehrkorn, T., Chang, R. Y. W., Karion, A., Miller, J. B., Sweeney, C., Steiner, N., Wofsy, S. C. and others: Atmospheric transport simulations in support of the Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE), Atmos Chem Phys, 15(8), 4093–4116, 2015.

Heymann, J., Reuter, M., Buchwitz, M., Schneising, O., Bovensmann, H., Burrows, J. P., Massart, S., Kaiser, J. W. and Crisp, D.: CO_2 emission of Indonesian fires in 2015 estimated from satellite-derived atmospheric CO_2 concentrations, Geophys. Res. Lett., 44(3), 1537–1544, doi:10.1002/2016GL072042, 2017.

Hogue, S., Marland, E., Andres, R. J., Marland, G. and Woodard, D.: Uncertainty in gridded CO_2 emissions estimates, Earth's Futur., 4(5), 225–239, doi:10.1002/2015EF000343, 2016.

Houweling, S., Breon, F. M., Aben, I., Rodenbeck, C., Gloor, M., Heimann, M. and Ciais, P.: Inverse modeling of CO_2 sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their performance as a function of space and time, Atmos. Chem. Phys., 4, 523–538, doi: 10.5194/acp-4-523-2004, 2004.

Huntzinger, D. N., Schwalm, C. R., Michalak, A. M., Schaefer, K., King, A. W., Wei, Y., Jacobson, A. R., Liu, S., Cook, R. B., Post, W. M., Berthier, G., Hayes, D., Huang, M., Viovy, N., Lu, C., Tian, H., Ricciuto, D. M., Mao, J. and Shi, X.: The north american carbon program multi-scale synthesis and terrestrial model intercomparison project – Part 2: Environmental driver data, Geosci. Model Dev., 6(6), 2121–2133, doi:10.5194/gmd-7-2875-2014, 2013.

Janardanan, R., Maksyutov, S., Oda, T., Saito, M., Kaiser, J. W., Ganshin, A., Stohl, A., Matsunaga, T., Yoshida, Y. and Yokota, T.: Comparing GOSAT observations of localized CO_2 enhancements by large emitters with inventory-based estimates, Geophys. Res. Lett., 43(7), 3486–3493, doi:10.1002/2016GL067843, 2016.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U. and Petrescu, A. M. R.: EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970–2012, Earth Syst. Sci. Data Discuss., 10.5194/essd-2017-79, in review, 2017.

Jeong, S., Hsu, Y. K., Andrews, A. E., Bianco, L., Vaca, P., Wilczak, J. M. and Fischer, M. L.: A multitower measurement network estimate of California's methane emissions, J. Geophys. Res. Atmos., 118(19), 11339–11351, doi:10.1002/jgrd.50854, 2013.

Kim, S. Y., Millet, D. B., Hu, L., Mohr, M. J., Griffis, T. J., Wen, D., Lin, J. C., Miller, S. M. and Longo, M.: Constraints on carbon monoxide emissions based on tall tower measurements in the US Upper Midwest, Environ. Sci. Technol., 47(15), 8316–8324, 2013.

Kort, E. A., Frankenberg, C., Miller, C. E. and Oda, T.: Space-based observations of megacity carbon dioxide, Geophys. Res. Lett., 39(17), 1–5, doi:10.1029/2012GL052738, 2012.

Kort, E. A., Angevine, W. M., Duren, R. and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity, J. Geophys. Res. Atmos., 118(3), 1–8, doi:10.1002/jgrd.50135, 2013.

Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J. and Rayner, P. J.: Structure of the transport uncertainty in mesoscale inversions of CO_2 sources and sinks using ensemble model simulations, Biogeosciences, 6(6), 1089–1102, doi:10.5194/bg-6-1089-2009, 2009.

Lauvaux, T. and Davis, K. J.: Planetary boundary layer errors in mesoscale inversions of column-integrated CO_2 measurements. Journal of Geophysical Research: Atmospheres, 119(2), 490-508, 2014.

Lauvaux, T., Miles, N. L., Richardson, S. J., Deng, A., Stauffer, D. R., Davis, K. J., Jacobson, G., Rella, C., Calonder, G. P. and DeCola, P. L.: Urban emissions of CO_2 from Davos, Switzerland: The first real-time monitoring system using an atmospheric inversion technique. Journal of Applied Meteorology and Climatology, 52(12), 2654-2668, 2013.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O'Keefe, D., Song, Y., Karion, A., Oda, T., Patarasuk, R., Razlivanov, I., Sarmiento, D., Shepson, P., Sweeney, C., Turnbull, J.

and Wu, K.: High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX), J. Geophys. Res. Atmos., 121(10), 5213–5236, doi:10.1002/2015JD024473, 2016.

Law, R. M., Rayner, P. J., Denning, A. S., Erickson, D., Fung, I. Y., Heimann, M., Piper, S. C., Ramonet, M., Taguchi, S., Taylor, J. A., Trudinger, C. M. and Watterson, I. G.: Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, Global Biogeochem. Cycles, 10(4), 783–796, doi:10.1029/96GB01892, 1996.

Levin, I., Kromer, B., Schmidt, M. and Sartorius, H.: A novel approach for independent budgeting of fossil fuel CO_2 over Europe by $^{14}CO_2$ observations, Geophys. Res. Lett., 30(23), 2194, doi:10.1029/2003GL018477, 2003.

Lin, J. C. and Gerbig, C.: Accounting for the effect of transport errors on tracer inversions, Geophys. Res. Lett., 32(1), 2005.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J. and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res. Atmos., 108(D16), 2003.

Lin, J. C., Gerbig, C., Daube, B. C., Wofsy, S. C., Andrews, A. E., Vay, S. A. and Anderson, B. E.: An empirical analysis of the spatial variability of atmospheric CO_2: Implications for inverse analyses and space-borne sensors, Geophys. Res. Lett., 31(23), 1–5, doi:10.1029/2004GL020957, 2004.

Lin, J. C., Gerbig, C., Wofsy, S. C., Daube, B. C., Matross, D. M., Chow, V. Y., Gottlieb, E., Andrews, A. E., Pathmathevan, M. and Munger, J. W.: What have we learned from intensive atmospheric sampling field programmes of CO_2?, Tellus, Ser. B Chem. Phys. Meteorol., 58(5), 331–343, doi:10.1111/j.1600-0889.2006.00202.x, 2006.

Lin, J. C., Mallia, D. V., Wu, D. and Stephens, B. B.: How can mountaintop CO_2 observations be used to constrain regional carbon fluxes?, Atmos. Chem. Phys., 17(9), 5561–5581, doi:10.5194/acp-17-5561-2017, 2017.

Liu, Y., Yang, D. and Cai, Z.: A retrieval algorithm for TanSat XCO_2 observation: Retrieval experiments using GOSAT data, Chin. Sci. Bull. 58(13), 1520–1523, doi:10.1007/s11434-013-5680-y, 2013.

Luus, K. A., Commane, R., Parazoo, N. C., Benmergui, J., Euskirchen, E. S., Frankenberg, C., Joiner, J., Lindaas, J., Miller, C. E., Oechel, W. C., Zona, D., Wofsy, S. and Lin, J. C.: Tundra photosynthesis captured by satellite-observed solar-induced chlorophyll fluorescence, Geophys. Res. Lett., 44(3), 1564–1573, doi:10.1002/2016GL070842, 2017.

Macatangay, R., Warneke, T., Gerbig, C., Ahmadov, R., Heimann, M. and Notholt, J.: A framework for comparing remotely sensed and in-situ CO_2 concentrations, Atmos. Chem. Phys., 8, 2555–2568, 2008.

Mallia, D. V, Lin, J. C., Urbanski, S., Ehleringer, J. and Nehrkorn, T.: Impacts of upwind wildfire emissions on CO, CO_2, and $PM_{2.5}$ concentrations in Salt Lake City, Utah, J. Geophys. Res. Atmos., 120(1), 147–166, 2015.

Mallia, D. V., Kochanski, A., Wu, D., Pennell, C., Oswald, W. and Lin, J. C.: Wind-blown dust modeling using a backward-Lagrangian particle dispersion model, J. Appl. Meteorol. Climatol., 56(10), 2845–2867, doi:10.1175/JAMC-D-16-0351.1, 2017.

Mandrake, L., Frankenberg, C., O'Dell, C. W., Osterman, G., Wennberg, P. and Wunch, D.: Semi-autonomous sounding selection for OCO-2, Atmos. Meas. Tech., 6(10), 2851–2864, 2013.

Marland, G.: Uncertainties in accounting for CO_2 from fossil fuels, J. Ind. Ecol., 12(2), 136–139, doi:10.1111/j.1530-9290.2008.00014.x, 2008.

Mitchell, L., Lin, J. C., Bowling, D. R., Pataki, D. E., Strong, C., Schauer, A. J., Bares, R., Bush, S., Stephens, B. B., Mendoza, D., Mallia, D. V, Holland, L., Gurney, K. R. and Ehleringer, J. R.: Long-term urban carbon dioxide observations reveal spatial and temporal dynamics related to urban characteristics and growth, Proc. Natl. Acad. Sci., 115(12), 2912–2917, doi:10.1073/pnas.1702393115, 2018.

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R. and Deng, F.: Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets, J. Geophys. Res. Atmos., 118(2), 917–933, doi:10.1029/2012JD018196, 2013.

OCO-2 Science Team/Michael Gunson, Annmarie Eldering, OCO-2 Level 2 bias-corrected solar-induced fluorescence and other select fields from the IMAP-DOAS algorithm aggregated as daily files, Retrospective processing V7r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://disc.gsfc.nasa.gov/datacollection/OCO2_L2_Lite_SIF_7r.html, 2015.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O. and Wunch, D.: The ACOS CO_2 retrieval algorithm-Part 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5(1), 99–121, doi:10.5194/amt-5-99-2012, 2012.

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

Oda, T., Ott, L., Topylko, P., Halushchak, M., Bun, R., Lesiv, M., Danylo, O. and Horabik-Pyzel, J.: Uncertainty associated with fossil fuel carbon dioxide (CO_2) gridded emission datasets, 2015.

Oda, T. and Maksyutov, S.: ODIAC Fossil Fuel CO_2 Emissions Dataset (Version name: ODIAC2017), Center for Global Environmental Research, National Institute for Environmental Studies, doi:10.17595/20170411.001, 2015.

Oda, T., Maksyutov, S. and Andres, R. J.: 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, , 2, 87–107, 2018.

Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO_2 and implications for carbon cycle research, J. Geophys. Res., 109(D2), D02301, doi:10.1029/2003JD003968, 2004.

Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I., Gunson, M. R., Heddle, G., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, P., and Wofsy, S. C.: Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, Tech. rep., Committee on Methods for Estimating Greenhouse Gas Emissions, Washington, DC, 2010.

Palmer, P. I.: Quantifying sources and sinks of trace gases using space-borne measurements: current and future science, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 366(1885), 4509–4528, doi:10.1098/rsta.2008.0176, 2008.

Patra, P. K., Crisp, D., Kaiser, J. W., Wunch, D., Saeki, T., Ichii, K., Sekiya, T., Wennberg, P. O., Feist, D. G., Pollard, D. F., Griffith, D. W. T., Velazco, V. A., De Maziere, M., Sha, M. K., Roehl, C., Chatterjee, A. and Ishijima, K.: The Orbiting Carbon Observatory (OCO-2) tracks 2–3 peta-gram increase in carbon release to the atmosphere during the 2014–2016 El Niño, Sci. Rep., 7(1), 13567, doi:10.1038/s41598-017-13459-0, 2017.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C. and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proc. Natl. Acad. Sci., 104(48), 18925–18930, doi:10.1073/pnas.0708986104, 2007.

Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P. and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: Model intercomparison, Atmos. Chem. Phys., 11(13), 6607–6622, doi:10.5194/acp-11-6607-2011, 2011.

Rayner, P. J. and O'Brien, D. M.: The utility of remotely sensed CO_2 concentration data in surface source inversions, Geophys. Res. Lett., 28(1), 175–178, doi:10.1029/2000GL011912, 2001.

Rayner, P. J., Raupach, M. R., Paget, M., Peylin, P. and Koffi, E.: A new global gridded data set of CO_2 emissions from fossil fuel combustion: Methodology and evaluation, J. Geophys. Res., 115(D19), D19306, doi:10.1029/2009JD013439, 2010.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Schneising, O., Pillai, D., Bovensmann, H., Burrows, J. P., Bösch, H., Parker, R., Butz, A., Hasekamp, O., O'Dell, C. W., Yoshida, Y., Gerbig, C., Nehrkorn, T., Deutscher, N. M., Warneke, T., Notholt, J., Hase, F., Kivi, R., Sussmann, R., Machida, T., Matsueda, H. and Sawa, Y.: Satellite-inferred European carbon sink larger than expected, Atmos. Chem. Phys., 14(24), 13739–13753, doi:10.5194/acp-14-13739-2014, 2014.

Rodgers, C.D.: Inverse methods for atmospheric sounding: theory and practice. World scientific, 2000.

Rolph, G., Stein, A. and Stunder, B.: Real-time Environmental Applications and Display sYstem: READY, Environ. Model. Softw., 95, 210–228, doi:10.1016/j.envsoft.2017.06.025, 2017.

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

Schneising, O., Heymann, J., Buchwitz, M., Reuter, M., Bovensmann, H. and Burrows, J. P.: Anthropogenic carbon dioxide source areas observed from space: Assessment of regional enhancements and trends, Atmos. Chem. Phys., 13(5), 2445–2454, doi:10.5194/acp-13-2445-2013, 2013.

Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmos. Chem. Phys., 4(1), 51–63, doi:10.5194/acp-4-51-2004, 2004.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L. and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41(12), 4381–4388, doi:10.1002/2014GL059684, 2014.

Shiga, Y. P., Tadić, J. M., Qiu, X., Yadav, V., Andrews, A. E., Berry, J. A. and Michalak, A. M.: Atmospheric CO_2 Observations Reveal Strong Correlation Between Regional Net Biospheric Carbon Uptake and Solar-Induced Chlorophyll Fluorescence, Geophys. Res. Lett., 1122–1132, doi:10.1002/2017GL076630, 2018.

Silva, S. and Arellano, A.: Characterizing Regional-Scale Combustion Using Satellite Retrievals of CO, NO_2 and CO_2, Remote Sens., 9(7), 744, doi:10.3390/rs9070744, 2017.

Silva, S. J., Arellano, A. F. and Worden, H. M.: Toward anthropogenic combustion emission constraints from space-based analysis of urban CO_2/CO sensitivity, Geophys. Res. Lett., 40(18), 4971–4976, doi:10.1002/grl.50954, 2013.

Skamarock, W. C. and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys., 227(7), 3465–3485, 2008.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D. and Ngan, F.: Noaa's hysplit atmospheric transport and dispersion modeling system, Bull. Am. Meteorol. Soc., 96(12), 2059–2077, doi:10.1175/BAMS-D-14-00110.1, 2015.

Stephens, B.B., Gurney, K.R., Tans, P.P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T. and Aoki, S., Machida, T., Inoue, G., Vinnichenko, N., Lloyd, J., Jordan, A., Heimann, M., Shibistova, O., Langenfelds, R.L., Steele, L.P., Francey, R.J., Denning, A.S.: Weak Northern and Strong Tropical Land Carbon Uptake from Vertical Profiles of Atmospheric CO_2, Science, 316(5832), 1732–1736, doi:10.1126/science.1137004, 2007.

Stohl, A., Forster, C., Frank, A., Seibert, P. and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5(9), 2461–2474, doi:10.5194/acp-5-2461-2005, 2005.

Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J. and Gu, L., Magney, T. S., Kohler, P., Evans, B. and Yuen, K.: OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. Science, 358(6360), eaam5747, 2017.

Sweeney, C., Karion, A., Wolter, S, Newberger, T, Guenther, D, Higgs, J.A., Andrews, A.E., Lang, P.M., Neff, D., Dlugokencky, E., Miller, J.B.: Seasonal climatology of CO_2 across North America from aircraft measurements in the NOAA/ESRL Global Greenhouse Gas Reference Network. Journal of Geophysical Research: Atmospheres, 120(10), 5155-5190, 2015.

Tollefson, J.: Carbon-sensing satellite system faces high hurdles, Nature, 533, 446–447, 2016.

UNFCCC, 2017. National Inventory Submissions 2017. United Nations Framework Convention on Climate Change. Available at: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/9492.php; accessed June 2017.

Venables, W. N. and Ripley, B.D.: Random and mixed effects. In Modern applied statistics with S, Springer, New York, NY, 271-300, 2002.

Verhulst, K. R., Karion, A., Kim, J., Salameh, P. K., Keeling, R. F., Newman, S., Miller, J., Sloop, C., Pongetti, T., Rao, P., Wong, C., Hopkins, F. M., Yadav, V., Weiss, R. F., Duren, R. and Miller, C. E.: Carbon Dioxide and Methane Measurements from the Los Angeles Megacity Carbon Project: 1. Calibration, Urban Enhancements, and Uncertainty Estimates, Atmos. Chem. Phys., 17, 8313-8341, doi:10.5194/acp-17-8313-2017, 2017.

Wheeler, D. and Ummel, K.: Calculating CARMA: Global Estimation of CO_2 Emissions from the Power Sector, Work. Pap., (145), 37, 2008.

Wong, K. W., Fu, D., Pongetti, T. J., Newman, S., Kort, E. A., Duren, R., Hsu, Y. K., Miller, C. E., Yung, Y. L. and Sander, S. P.: Mapping CH_4 : CO_2 ratios in Los Angeles with CLARS-FTS from Mount Wilson, California, Atmos. Chem. Phys., 15(1), 241–252, doi:10.5194/acp-15-241-2015, 2015.

Worden, J., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., O'Dell, C. and Bowman, K.: Evaluation And Attribution Of OCO-2 XCO_2 Uncertainties, Atmos. Meas. Tech., 10, 2759-2771, doi:10.5194/amt-10-2759-2017, 2017.

Wunch, D., Wennberg, P. O., Toon, G. C., Keppel-Aleks, G. and Yavin, Y. G.: Emissions of greenhouse gases from a North American megacity, Geophys. Res. Lett., 36(15), 1–5, doi:10.1029/2009GL039825, 2009.

Wunch, D., Wennberg, P. O., Toon, G. C., Connor, B. J., Fisher, B., Osterman, G. B., Frankenberg, C., Mandrake, L., O'Dell, C., Ahonen, P., Biraud, S. C., Castano, R., Cressie, N., Crisp, D., Deutscher, N. M., Eldering, A., Fisher, M. L., Griffith, D. W. T., Gunson, M., Heikkinen, P., Keppel-Aleks, G., Kyrö, E., Lindenmaier, R., MacAtangay, R., Mendonca, J., Messerschmidt, J., Miller, C. E., Morino, I., Notholt, J., Oyafuso, F. A., Rettinger, M., Robinson, J., Roehl, C. M., Salawitch, R. J., Sherlock, V., Strong, K., Sussmann, R., Tanaka, T., Thompson, D. R., Uchino, O., Warneke, T. and Wofsy, S. C.: A method for evaluating bias in global measurements of CO_2 total columns from space, Atmos. Chem. Phys., 11(23), 12317–12337, doi:10.5194/acp-11-12317-2011, 2011.

World Urbanization Prospects: The 2014 Revision (WUP 2014), United Nations, Department of Economic and Social Affairs, Population Division (2014)., CD-ROM Edition.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., Yang, E., and Wu, D.: Constraining fossil fuel CO_2 emissions from urban area using OCO-2 observations of total column CO_2, Atmos. Chem. Phys. Discuss., in review, doi: 10.5194/acp-2017-1022, 2017.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H. and Maksyutov, S.: Global Concentrations of CO_2 and CH_4 Retrieved from GOSAT: First Preliminary Results, Sola, 5, 160–163, doi:10.2151/sola.2009-041, 2009.

Zhao, C., Andrews, A. E., Bianco, L., Eluszkiewicz, J., Hirsch, A., MacDonald, C., Nehrkorn, T. and Fischer, M. L.: Atmospheric inverse estimates of methane emissions from Central California, J. Geophys. Res. Atmos., 114(16), 1–13, doi:10.1029/2008JD011671, 2009.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

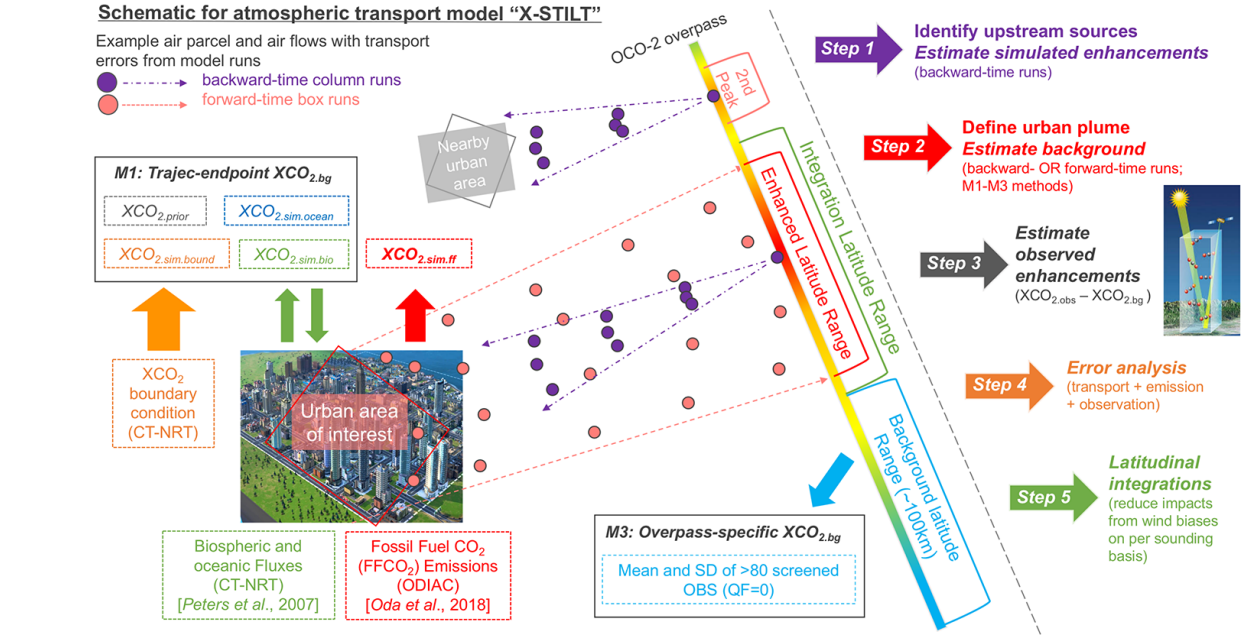

Figure 1. A schematic of X-STILT. Pink and purple dots and arrows represent the air parcels and overall air flows based on forward-time box runs and backward-time column runs with wind error component accounted for. Rainbow band is an example of one OCO-2 overpass with warmer color indicating higher observed XCO_2.

Step 1: Modeled XCO_2 enhancements due to fossil fuel CO_2 emissions along the track are derived from backward-time trajectories (and further column weighted footprint) and fossil fuel CO_2 emissions from ODIAC.

Step 2: Demonstrations of trajectory-endpoint and overpass-specific background XCO_2 are labeled as black boxes. M1 include modeled-derived biospheric, oceanic XCO_2 changes (with fluxes from CarbonTracker-NearRealTime, CT-NRT), CO_2 boundary conditions derived from CT-NRT 3D concentration field, and prior CO_2 portion from OCO-2. M3 requires enhanced latitude range based on either backward-time XCO_2 enhancements or forward-time urban plume. M2 is statistically-derived background based on two previous studies, which is not shown in this figure.

Step 3: Bin up screened observations (QF=0) according to receptor lat/lon and calculate the bin-averaged total XCO_2. Background XCO_2 is then subtracted from those bin-averaged observed XCO_2 to estimate observed enhancements.

Step 4: Error analysis of atmospheric transport uncertainty (Sect. 2.6), emission uncertainty (Sect. 2.3.1) and observed uncertainty (Sect. 2.2) yield XCO_2 uncertainties for each sounding/receptor.

Step 5: Modeled and observed urban enhancements from step 1 and 3, together with transport, prior emissions errors and observed errors (with error covariance included, from step 4) on per sounding basis are integrated along their latitudes, to finally estimate the overall observed and modeled urban signals and uncertainties (Sect. 3.5).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

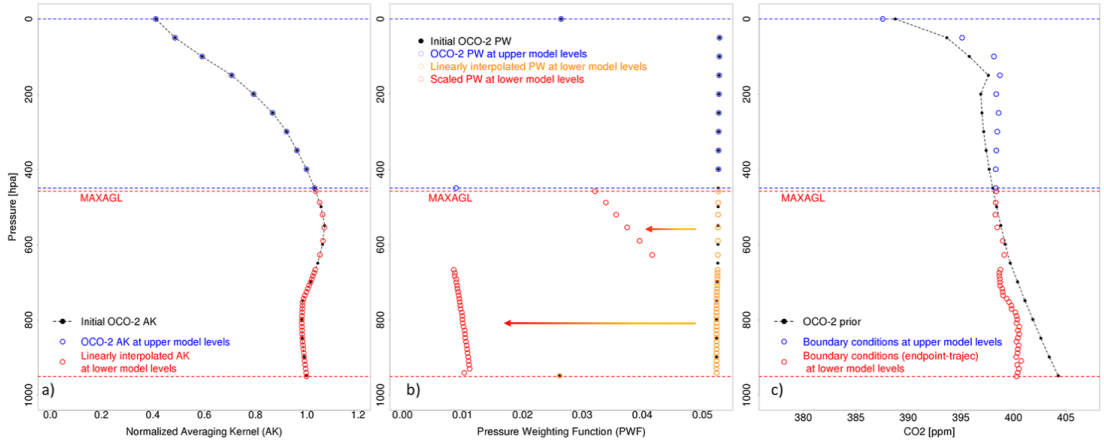

Figure 2. Demonstration of interpolations on **a)** normalized averaging kernel profiles, **b)** pressure weighting function and **c)** modeled boundary conditions (derived from trajectory endpoint using CT-NRT) and OCO-2 a priori profiles, for a sounding (lat/lon same as column receptors). Note that column receptors are placed only up to a certain level (referred to as the MAXAGL).

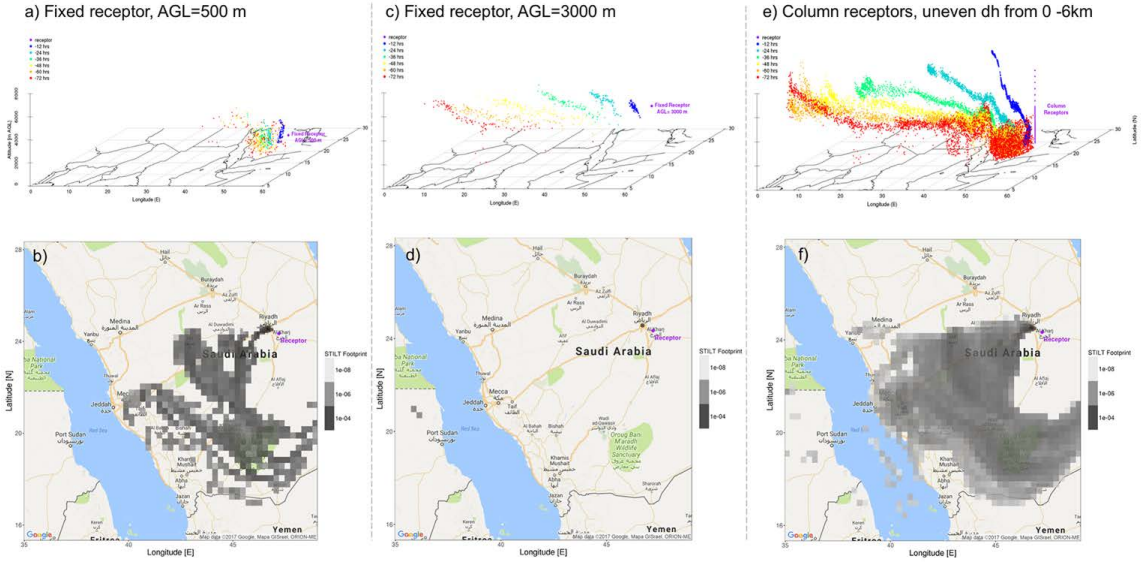

Figure 3. Upper panels (**a, c, e**): Demonstrations of X-STILT runs from fixed receptor or "column receptors" for Riyadh at 1000UTC on 12/29/2014. 3D scatter plot of locations of STILT ensembles that initially released from the same coordinate as one OCO-2 sounding (latitude=24.4961N). Colors differentiate hours backwards (i.e., -2min, -12, -24, -36, -48, -60, and -72 hours) of each trajectory. Note that column receptors are placed every 100 m within 3 km and every 500 m from 3–6 km. Lower panels (**b, d, f**): Modeled fixed footprints vs. column footprints are plotted in grey. Column footprints are only weighted by interpolated pressure weighting functions. Only footprints values >1E-8 ppm/(μmol m^{-2} s^{-1}) are indicated using the greyscale.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

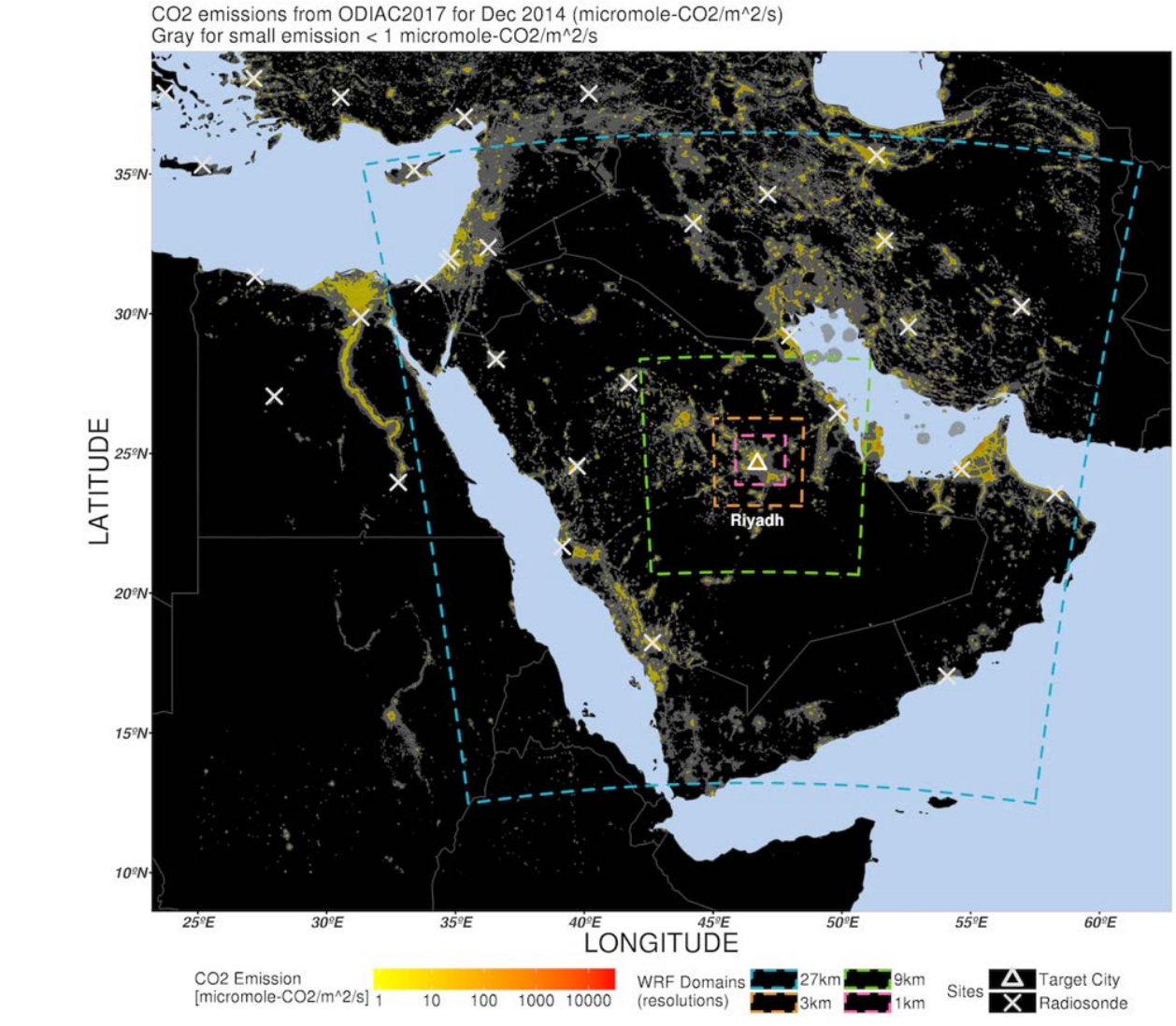

Figure 4. Map of $FFCO_2$ emissions from ODIAC (log-scale, 1km×1km spatial grid spacing) in Dec 2014, along with 4 nested WRF domains. White crosses and triangle denote the radiosonde networks used to evaluate the wind fields and provide wind error statistics, and the center of Riyadh. Regions shaded in grey indicate are areas with small emissions (< 1μmole m^{-2} s^{-1}).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

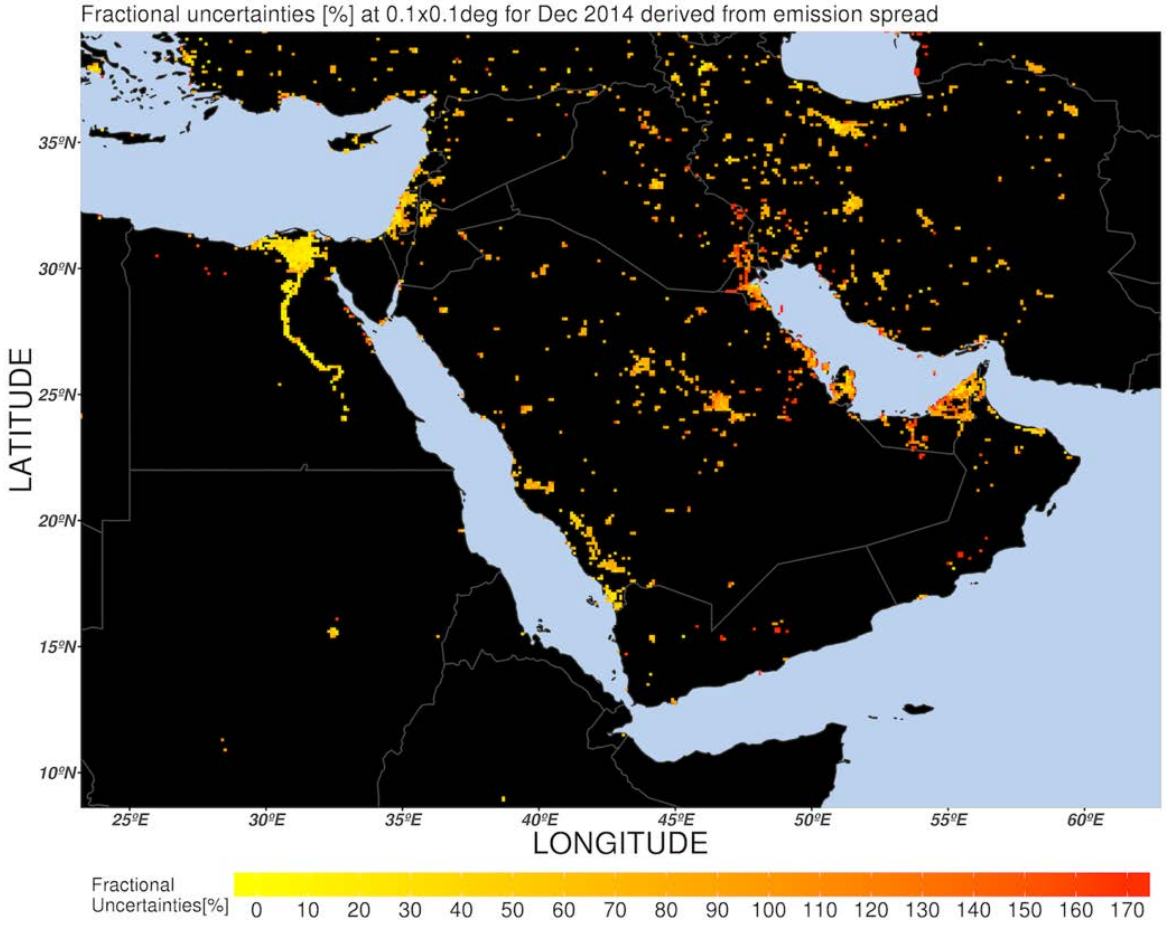

Figure 5. Map of 0.1° × 0.1° fractional uncertainties (%) of $FFCO_2$ emissions derived from the 1-σ among 3 emission inventories (ODIAC, FFDAS and EDGAR). Only fractional uncertainties with large ODIAC emissions (>1 μmole m^{-2} s^{-1}) using this spread method are displayed.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

Geoscientific Model Development Discussions

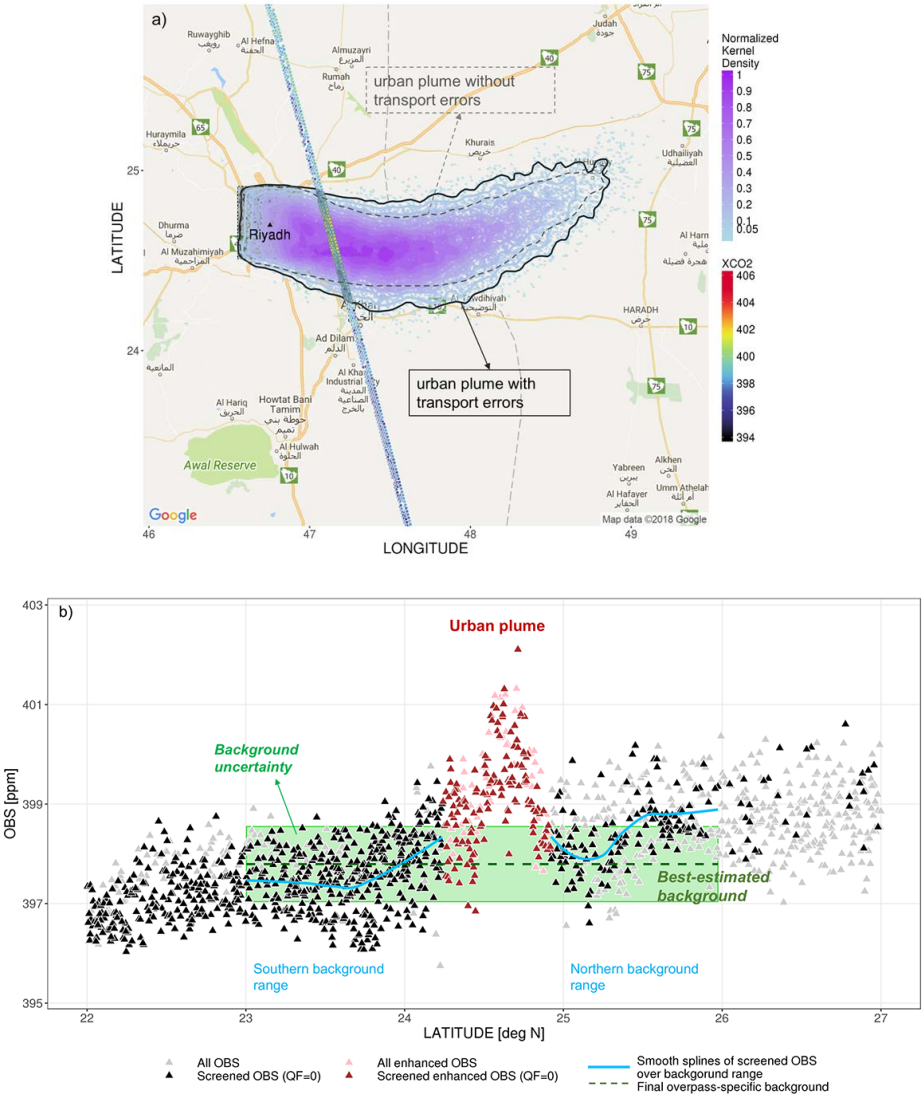

Figure 6. Demonstrations of overpass-specific background with an example of 12/29/2014 overpass for Riyadh. **a)** Forward particle distributions and their normalized kernel density with wind error component during OCO-2 overpass time with observed XCO_2. Urban plumes defined based on 5 % of the max 2D kernel density from parcels' distributions without and with transport errors are shown in grey and black solid lines. The intersection between the urban plume (black solid line) and the OCO-2 overpass (colorful dots) yields the downwind enhanced latitude range. **b)** Latitude-series of observed XCO_2. Smooth splines are applied to visually reveal the variation of observed XCO_2 over background latitudinal band. Screened observations with latitudes > 26° N and < 23° N are abandoned due to its large spread and a second large enhancement seen in other overpasses (not in this overpass). Final overpass-specific background XCO_2 and its uncertainty are calculated as the mean and one standard deviation of screened observations (> 400 soundings) on northern and southern part outside the urban plume (red triangles) in this case. For other cases, if the northern/southern portion is defined as the downwind side, the other side (southern/northern) serve as background region.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

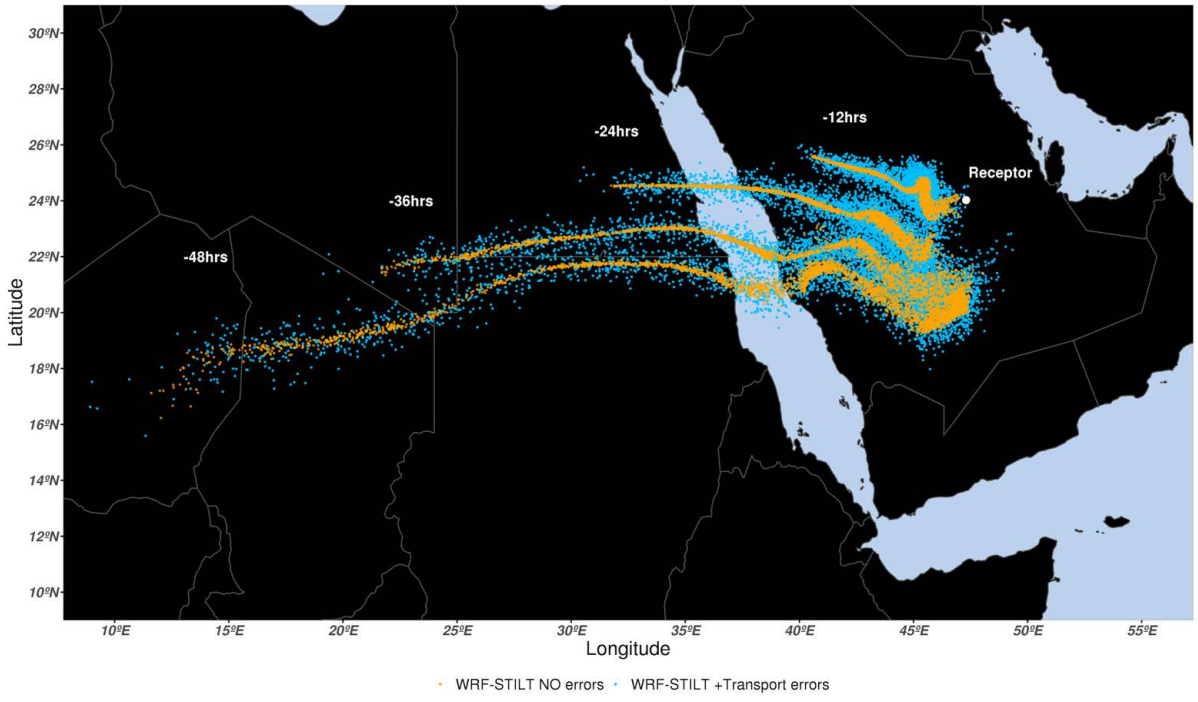

Figure 7. Spatial map of backward particle distributions without and with regional wind errors (at 12, 24, 36, 48 hours back) are shown in orange and blue dots, respectively. Note that only particles released from receptors below 3 km are plotted. For each time step, particles at higher vertical levels locate to the west of those near the surface.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

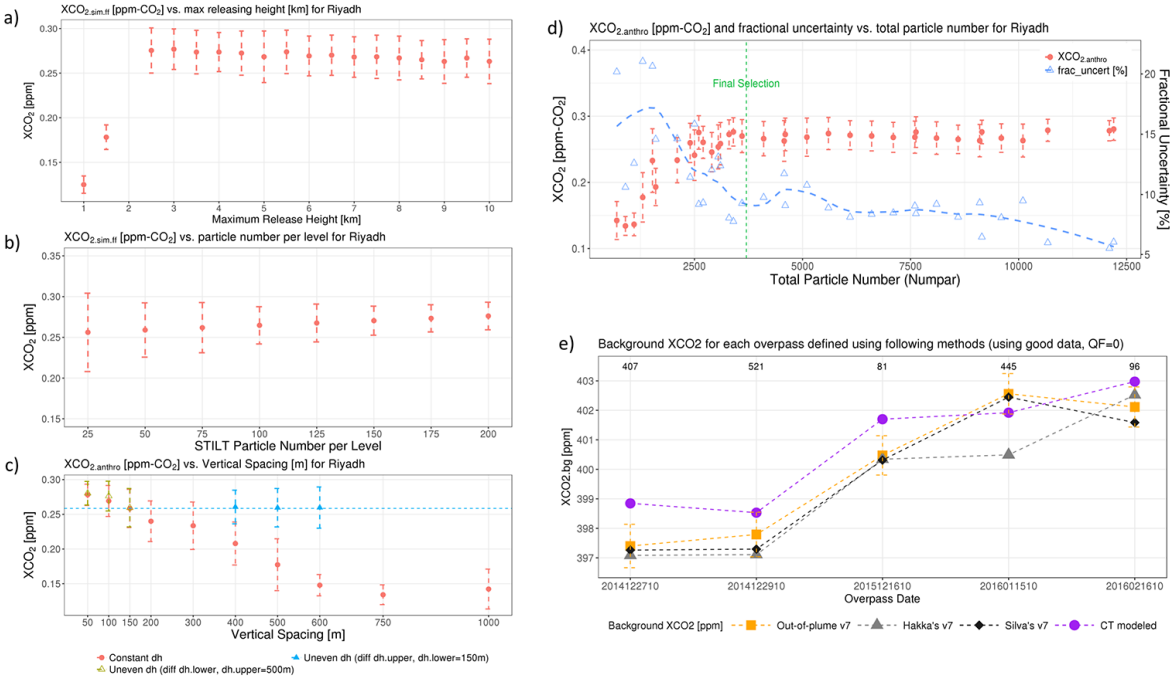

Figure 8. a-c) Three tests (MAXAGL, dpar and dh) on the sensitivities of XCO_2 enhancements to different model parameters, shown for the one sounding with the largest retrieved XCO_2 for Riyadh. For vertical spacing test **(c)**, tests with constant and uneven dh are carried out: case 1 with upper dh = 500 m and 3 different lower dh (50, 100 and 150m) and case 2 with lower dh = 150m and three upper dh (400, 500, 600m). **d)** A summary plot of XCO_2 enhancements and fractional uncertainties (%) versus total particle number. Green dashed vertical line denotes the configuration used in this study. **e)** Background comparisons using three methods for 5 tracks, with trajectory-endpoint derived background (M1), statistical background derived from Hakkarainen et al. (2016) (M2H) and Silva and Arellano (2017) (M2S) and X-STILT overpass-specific background (M3). The amount of screened observations for M3 is labeled as numbers. The uncertainty of overpass-specific background is denoted as dashed yellow error bars.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

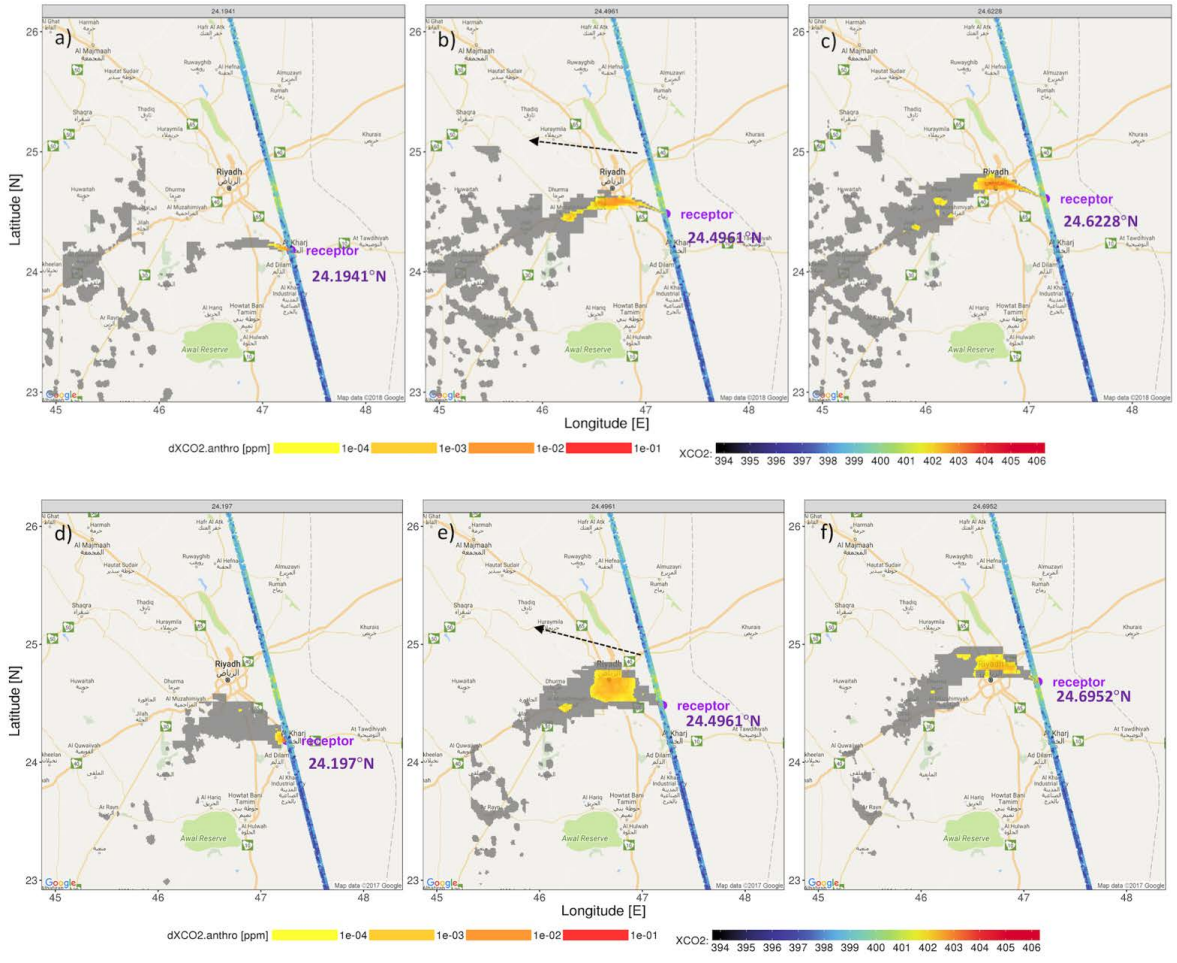

Figure 9. Spatial maps of anthropogenic contributions (> 10^{-10} ppm) with 1km×1km grid spacing from 3 selected soundings with all observations for the one overpass on 12/29/2014 1000UTC over Riyadh. Note that grey region stands for small contributions < 10^{-4} ppm. $FFCO_2$ contributions are calculated from surface emissions and weighted column footprint (derived from original trajectories without transport errors or wind bias corrections) with meteorological fields driven by GDAS (a-c) and WRF (d-f).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

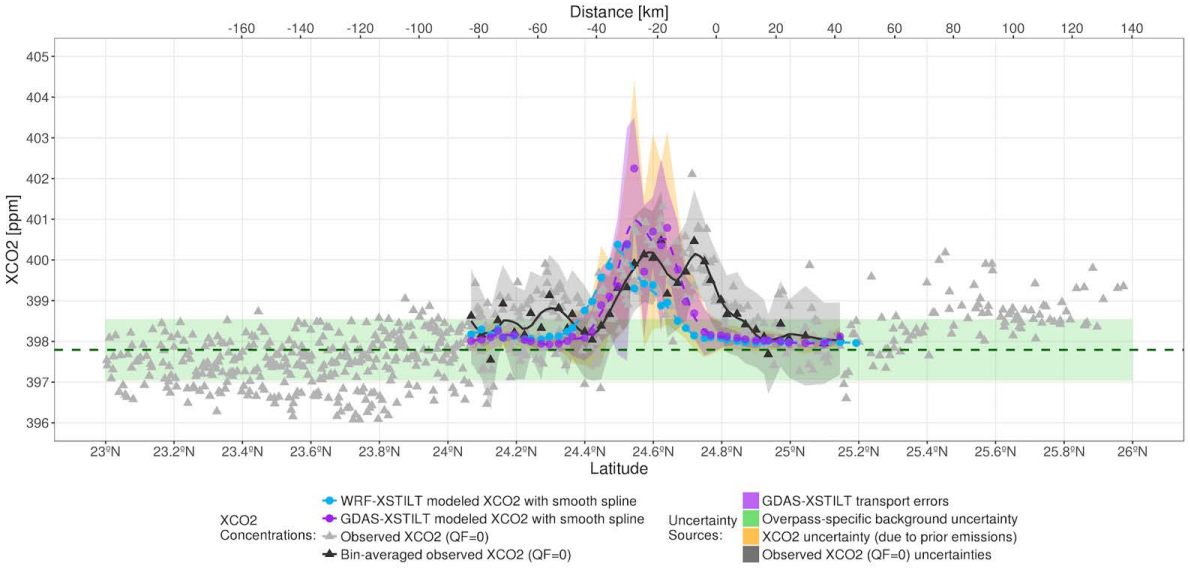

Figure 10. Latitude-series of simulated XCO_2 against OCO-2 retrieved XCO_2 (ppm) for Riyadh. Screened observations with QF=0 and bin-averaged observed XCO_2 are in grey and black triangles, respectively. GDAS- and WRF- derived XCO_2 are displayed in purple and light blue dots. The latitudinal range (from ~24° N to 25.1° N) is the integration range. Smooth splines are applied to both modeled and observed XCO_2 to visually reveal the main variations (purple and blue dashed lines and black solid line), not to calculate latitudinal-integrated signals.

XCO_2 uncertainties due to emissions and transport are drawn as yellow and purple ribbons. Background latitudinal band (extending northward to 26° N and southward to 23° N) is the same band as in Fig. 6b. Uncertainty in overpass-specific background that accounts for the spatial gradient in XCO_2 over background latitudinal band is shown as the light green ribbon. Total observed uncertainties due to natural XCO_2 variability, background and satellite retrievals are shown as the light grey ribbon.

Overpass-specific background XCO_2 is drawn as dark green dashed line. The displayed latitudinal range is the range for integrating XCO_2 enhancements and their uncertainties. For this event, the background XCO_2 is 397.79 ppm and integrated urban signals are 0.83, 0.71 and 0.98 ppm, for models using GDAS, WRF and observations, respectively. The latitudinally-integrated XCO_2 uncertainties due to transport and emission uncertainties (areas of purple and yellow ribbons) are 0.41 and 0.52 ppm, respectively.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

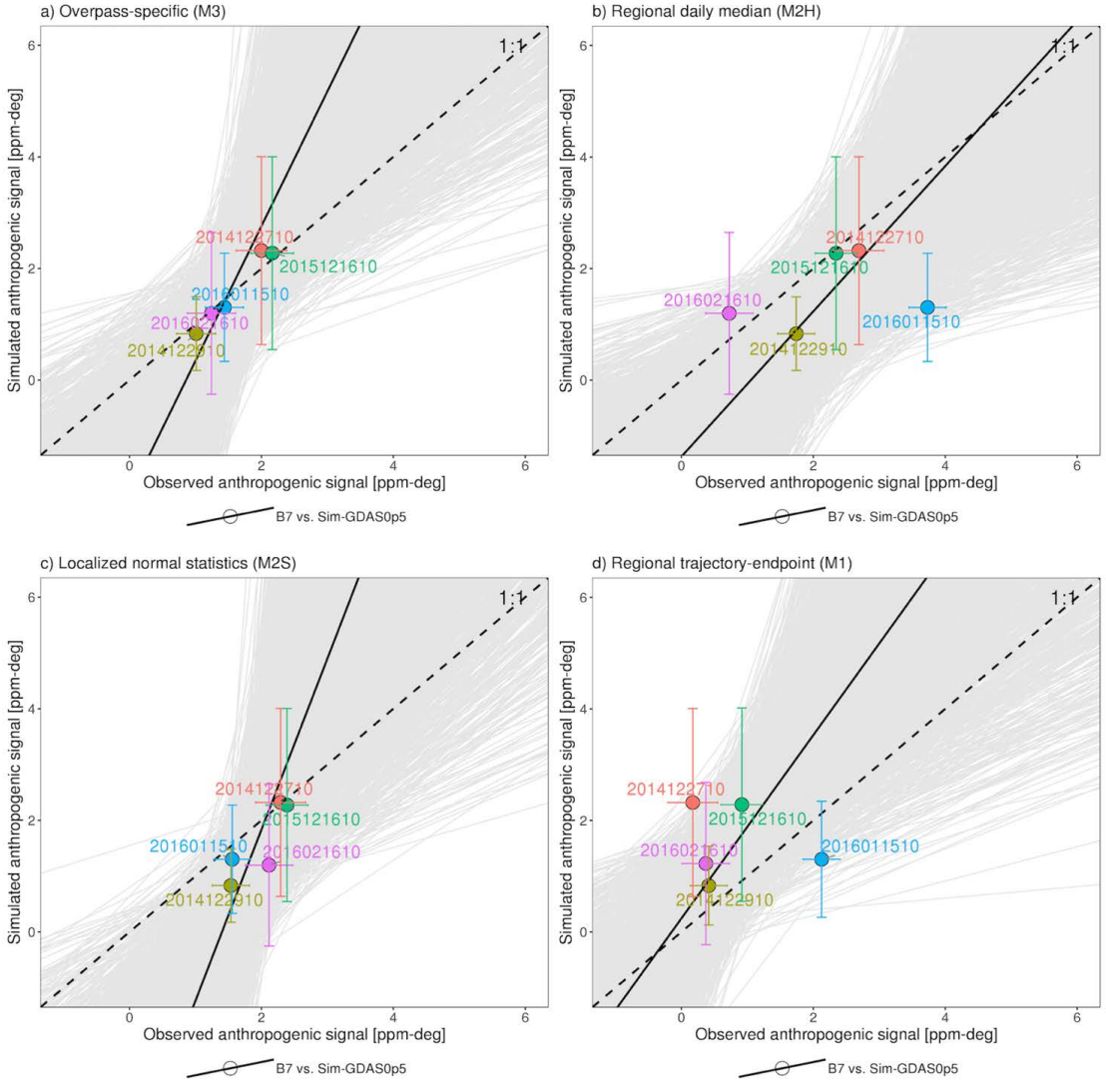

Figure 11. Correlation between observed and simulated anthropogenic XCO_2 signals for 5 overpasses. Colors differentiate different satellite overpass dates. Model-data comparisons using GDAS-derived XCO_2 signals and observed signals based on different background methods described in Section 2.4, a) overpass-specific (M3), b) daily median based on Hakkarainen et al. (2016) (M2H), c) normal statistics based on Silva and Arellano (2017) (M2S) and d) trajectory-endpoint (M1). Error bars along x-axis and y-axis represent the overall observed uncertainty (represented as 1-σ, including XCO_2 spatial variability, background uncertainty and retrieval errors) associated with observed signals and the overall modeled uncertainty (σ, including emission uncertainty and transport uncertainty) around modeled signals. Dotted dashed line represents the 1:1 line. Monte Carlo experiments are performed to fit linear regression lines based on sampled model-data signals (given best estimates and uncertainties in both x- and y-directions). Regression lines with positive slopes are shaded in light grey. Median values of slopes and y-intercepts from those multiple regression lines (with positive slopes) are used to draw a linear regression (black solid line).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

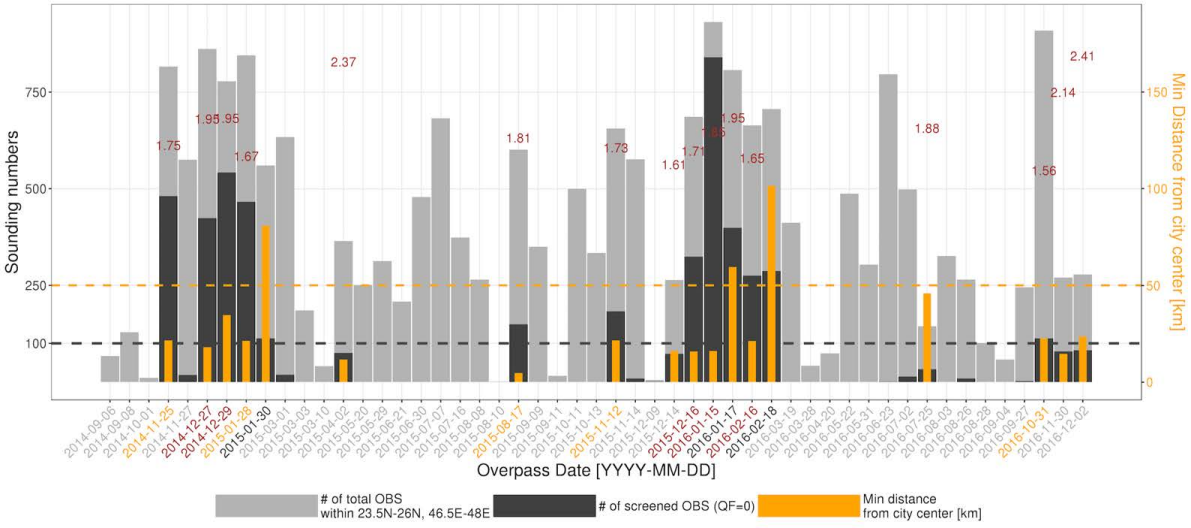

Figure A1. Four steps to choose candidates of satellite overpasses for X-STILT modeling. Grey and black bars indicate the total numbers of observations that fall into a designed spatial domain (i.e., 23.5° N–26° N, 46.5° E–48° E) and numbers of screened observations (QF=0) for each overpass, respectively, with y-axis to the left. Orange bars indicate the closest distance (in km) between each sounding and Riyadh city center (24.71° N, 46.71° E), with orange y-axis to the right. Numbers labeled in brown are the averaged u- and v-component wind errors [m/s] below 3 km at regional scale during 3-day-period. Overpasses are narrowed down using at least 100 good soundings (black dashed line), within the circle of a 50 km radius around the city center (orange dashed line), and relatively small regional wind errors (< 2 m s^{-1}, in brown text). In the end, we selected and examined 5 overpasses with their dates labeled in brown on the x-axis, via manual check.

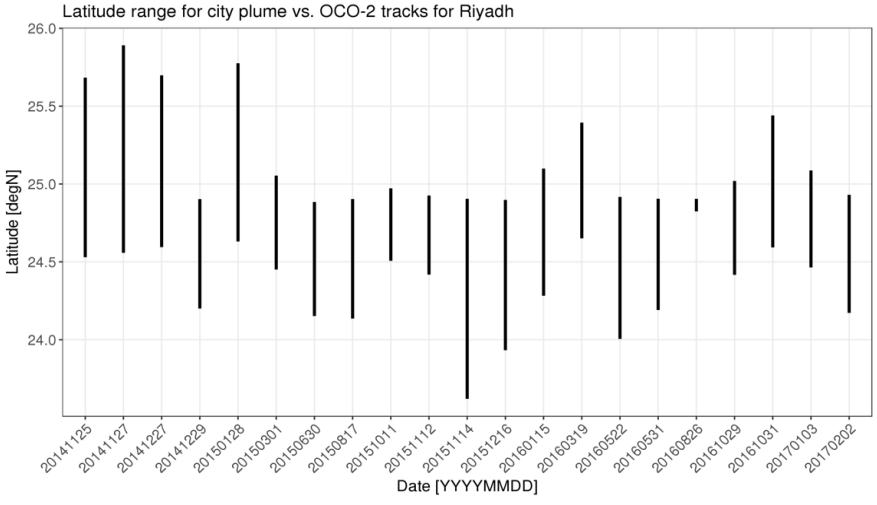

Figure A2. Time series of forward-time run interpolated latitude range for city plume, with near-field transport errors included, for several overpasses over Riyadh.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

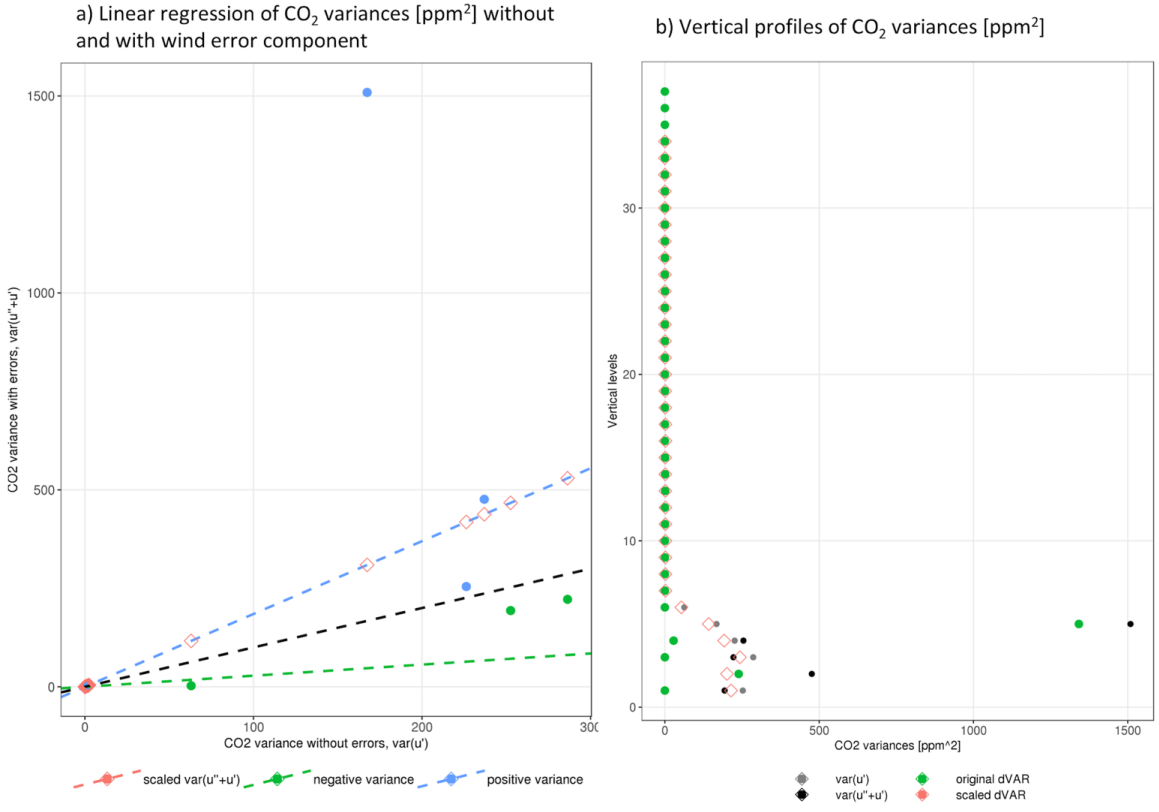

Figure A3. Demostrations of the new regression-based transport error algorithm, to resolve the techinial issue where negative difference in variance occur when parcels are statistically insufficient (green dots). Note that u'' in the figure simple means ε. **a)** Solid dots represent the errors in CO_2 among air parcels without ($\sigma^2_{u'}$) and with wind error component ($\sigma^2_{\varepsilon+u'}$) for each model release level. Linear regression line (green dashed line) is fitted for levels where $\sigma^2_{u'}$ is larger than $\sigma^2_{\varepsilon+u'}$ (green dots). Weighted linear regression line (blue dashed line) is fitted for normal cases where $\sigma^2_{u'}$ is smaller than $\sigma^2_{\varepsilon+u'}$ (blue dots), with weights of $1/\sigma^2_{\varepsilon+u'}$. The weighted regression line descibes the overall increase in CO2 variances due to the randomization over all X-STILT levels. Then, we recalculate the $\sigma^2_{\varepsilon+u'}$ based on $\sigma^2_{u'}$ and weighted regression line, as scaled $\sigma^2_{\varepsilon+u'}$ (red squares). **b)** Vertical profiles of CO_2 variances without or with the wind error component (grey or black dots), difference between $\sigma^2_{u'}$ and original $\sigma^2_{\varepsilon+u'}$ (green sqaures) and difference between $\sigma^2_{u'}$ and scaled $\sigma^2_{\varepsilon+u'}$ (red squares). If differeces between $\sigma^2_{u'}$ and original $\sigma^2_{\varepsilon+u'}$ are negative for certain lower levels, we assigned them as 0. The final transport error per level is calculated as the difference between $\sigma^2_{u'}$ and scaled $\sigma^2_{\varepsilon+u'}$ (red squares). Note that the notations of var(u' +u'') and var(u') in the plot legend represent $\sigma^2_{\varepsilon+u'}$ and $\sigma^2_{u'}$, respectively.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-123
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 14 June 2018

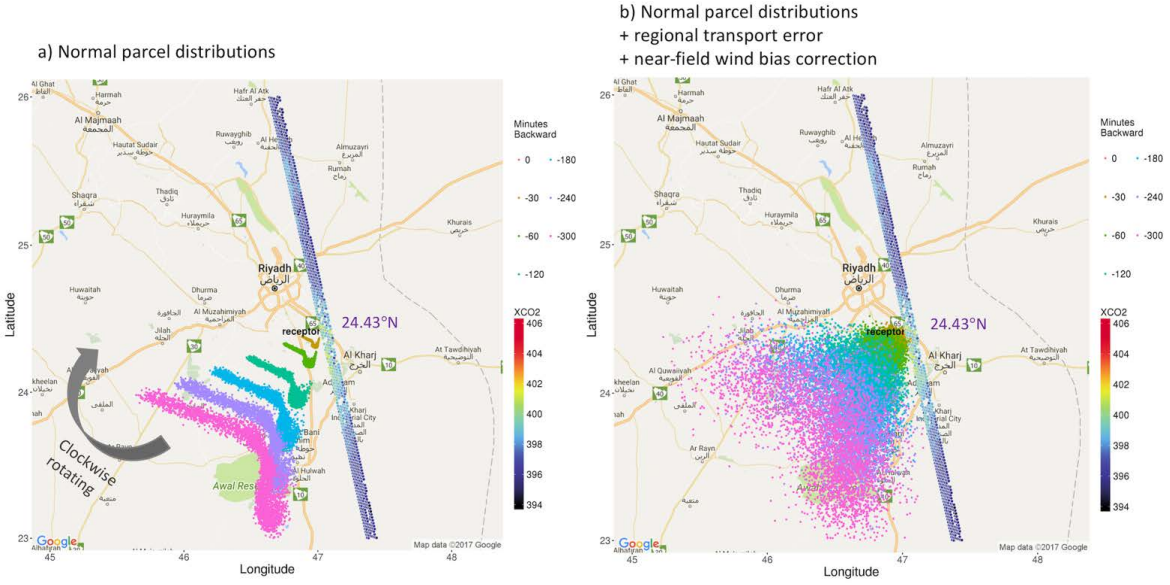

Figure A4. Demonstration of impacts of regional transport errors and near-field wind biases on parcel distributions. Original backward trajectories (**a**) and trajectories with both wind error perturbation and near-field wind corrections (**b**), at 0.5, 1, 2, 3, 4, 5 hours back in time (different colors) released from latitude at ~24.43° N, along with observed XCO_2. In this example, we rotated model trajectories based on prescribed wind biases (e.g., u = +0.3m/s; v = -1.1m/s).

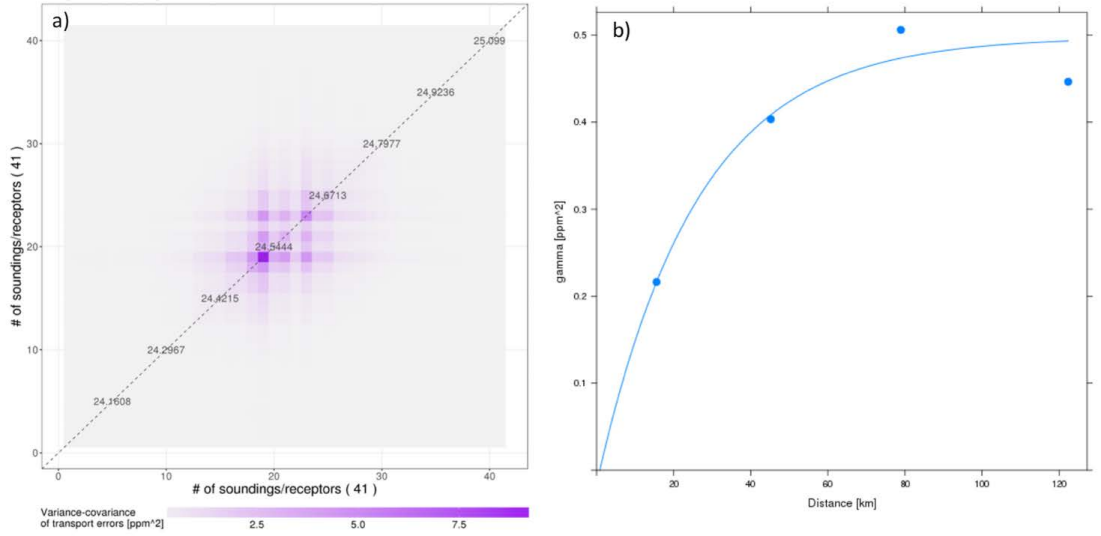

Figure A5. a) An example of transport error covariance matrix with a horizontal correlation lengthscale of 25 km for overpass on 12/29/2014. **b)** Exponential variogram for estimating the horizontal lengthscale (km) of transport errors between each two modeled receptors/sampled soundings, for the overpass on 12/29/2014.