Atmospheric
Chemistry
and Physics

State-wide California 2020 carbon dioxide budget estimated with OCO-2 and OCO-3 satellite data

Matthew S. Johnson[1], Sofia D. Hamilton[2], Seongeun Jeong[2], Yu Yan Cui[3], Dien Wu[4], Alex Turner[5], and Marc Fischer[2]

[1]Earth Science Division, NASA Ames Research Center, Moffett Field, CA, USA
[2]Energy Analysis and Environmental Impacts Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]independent researcher: Sacramento, CA, USA
[4]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[5]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA

Correspondence: Matthew S. Johnson (matthew.s.johnson@nasa.gov)

Abstract. Satellite observations are instrumental in observing spatiotemporal variability in carbon dioxide (CO_2) concentrations, which can be used to derive fluxes of this greenhouse gas. This study leverages NASA's Orbiting Carbon Observatory-2 and -3 (OCO-2 and OCO-3, respectively) CO_2 observations with a Gaussian process (GP) machine learning inverse model, a Bayesian nonparametric approach well suited for integrating the unique spatiotemporal characteristics of these satellite observations, to estimate subregional CO_2 fluxes. Utilizing the GEOS-Chem chemical transport model (CTM) to simulate column-average CO_2 concentrations (XCO_2) for 2020 in California – a period marked by the coronavirus disease (COVID-19) pandemic, drought conditions, and significant wildfire activity – we estimated the state-wide CO_2 emission rates constrained by OCO-2/3. This study developed prior fossil fuel emissions to reflect reduced activities during the COVID-19 pandemic, while net ecosystem exchange (NEE) and fire emissions were derived based on satellite data. GEOS-Chem source-specific XCO_2 concentrations for fossil fuels, NEE, fire, and oceanic sources were simulated coincident to OCO-2/3 XCO_2 retrievals to estimate state-wide sector-specific and total CO_2 emissions. GP inverse model results suggest that annual posterior median fossil fuel emissions were consistent with prior estimates (317.8 and 338.4 Tg CO_2 yr^{-1}, respectively; 95 % confidence level) and that posterior NEE fluxes had less carbon uptake compared to prior fluxes (-36.8 vs. -99.2 Tg CO_2 yr^{-1}, respectively; 95 % confidence level). Posterior fire CO_2 emissions were estimated to be 68.0 Tg CO_2 yr^{-1}, which was much lower than a priori estimates (103.3 Tg CO_2 yr^{-1}). The total median annual CO_2 emissions for the state of California in 2020 were estimated to be 349.6 Tg CO_2 yr^{-1} (range of 272.8–428.6 Tg CO_2 yr^{-1}; 95 % confidence level), aligning closely with the prior total estimate of 342.5 Tg CO_2 yr^{-1}. This study, for the first time, demonstrates that OCO-2/3 XCO_2 observations can be assimilated into inverse models to estimate state-wide source-specific CO_2 fluxes on a seasonal and annual scale.

1 Introduction

Carbon dioxide (CO_2) is the most abundant greenhouse gas in Earth's atmosphere and contributes predominantly to the present-day increase in global radiative forcing (Dunn et al., 2022). Due primarily to anthropogenic emissions from fossil fuel production and usage, global concentrations of CO_2 have nearly doubled since the beginning of the preindustrial era (Gulev et al., 2021; Lan et al., 2023). A recent comprehensive budget analysis of global CO_2 fluxes suggests that, as of 2022, anthropogenic emissions are $\sim 10\,\mathrm{Gt\,C\,yr^{-1}}$ (primarily from combustion of coal, oil, and natural gas) with an oceanic and terrestrial uptake offset of $\sim 3\,\mathrm{Gt\,C\,yr^{-1}}$ and $\sim 4\,\mathrm{Gt\,C\,yr^{-1}}$, respectively (Friedlingstein et al., 2023). According to this report, the United States (US) contributes 14 % ($\sim 1.4\,\mathrm{Gt\,C\,yr^{-1}}$) of global CO_2 anthropogenic emissions. The sectors contributing the most to US anthropogenic emissions are transportation, electricity generation, and industry (United States Environmental Protection Agency, 2023). One of the larger emitters of greenhouse gases in the US is the state of California which, as of 2021, contributes $\sim 0.1\,\mathrm{Gt\,C\,yr^{-1}}$ of CO_2 (CARB, 2023). In 2006, the state of California passed Assembly Bill 32 (AB 32) which required that the state's greenhouse gas emissions must be reduced to 1990 levels by the year 2020. California was able to achieve this goal, but in order to demonstrate this, as well as the success of other future emission reduction goals, it is vital to have accurate estimates of past and present-day greenhouse gas emissions.

Bottom-up inventories of CO_2 are commonly used to derive country-level to state-wide fossil fuel anthropogenic emissions in the US (e.g., Andres et al., 2012; CARB, 2023). Calculations of natural sources and sinks (e.g., terrestrial and marine biosphere and wildfires) contributing to total CO_2 emissions are frequently estimated using model predictions (Friedlingstein et al., 2023). The California Air Resources Board (CARB) has quantified state-wide greenhouse gas emissions for California between 2000–2021 (CARB, 2023). Anthropogenic and natural bottom-up CO_2 flux estimates are typically implemented in atmospheric transport models and compared to atmospheric observations in order to assess their accuracy. In situ observations of CO_2 from ground-based, tower, and aircraft platforms, due to their high accuracy and precision, are most frequently used to evaluate the quality of these emission estimates (Graven et al., 2018; Cui et al., 2022). While highly accurate, these types of in situ observations are limited in their spatiotemporal coverage and ability to constrain large regions and annual cycles of emissions. The assimilation of the satellite-retrieved column-averaged dry-air mole fraction of CO_2 (XCO_2) (e.g., Orbiting Carbon Observatory-2, OCO-2; Orbiting Carbon Observatory-3, OCO-3; Greenhouse Gases Observing Satellite, GOSAT; GOSAT-2; Carbon Dioxide Monitoring constellation, CO_2M; TanSat) into atmospheric transport models has been demonstrated to be able to constrain emissions on a global- to country-level scale more effectively in regions which lack dense in situ measurement networks (Pandey et al., 2016; Yang et al., 2021; Peiro et al., 2022; Philip et al., 2022; Imasu et al., 2023; Byrne et al., 2023; Noël et al., 2024). This study focuses on CO_2 fluxes in the state of California, where OCO-2 and OCO-3 have been used to constrain urban-scale emissions in the megacity of Los Angeles (Hedelius et al., 2018; Ye et al., 2020; Kiel et al., 2021; Wu et al., 2022; Roten et al., 2023; Hamilton et al., 2024). However, to date, no studies have demonstrated the capability to evaluate and constrain CO_2 fluxes using satellite-retrieved information on a state-wide spatial domain such as California. For California and other states, this is important, as some state agencies only release state-wide inventories (not specifically for urban areas), and many climate programs are generated based on the state-wide inventories.

While satellite-based atmospheric inverse modeling provides a significantly enhanced method for quantifying CO_2 emissions, using satellite observations in atmospheric inversions introduces two principal challenges: (1) incorporating the spatiotemporal covariance inherent in satellite data and (2) accurately estimating the hyperparameters (such as the length scale) of this covariance. Satellite observations contain both spatial and temporal properties, meaning that the data have inherent spatial and temporal characteristics that inform us about surface emissions. However, numerous inverse modeling studies have not consistently incorporated both covariance structures (Johnson et al., 2016; Fischer et al., 2017; Cui et al., 2019; Graven et al., 2018; Nathan et al., 2018; Ye et al., 2020; Wu et al., 2022; Roten et al., 2023). While some studies have accounted for both spatial and temporal covariances, they have not determined the optimal hyperparameters that align with the satellite observations (e.g., Turner et al., 2020). For example, the length scale parameter is crucial for influencing the covariance, which in turn affects the estimation of the unknown functions; in many cases, however, this parameter is not estimated explicitly for its optimal value but instead prescribed. In this study, we applied an atmospheric inversion system which fully utilizes the spatiotemporal properties embedded in satellite data (i.e., OCO-2 and OCO-3). This system is built based on the Gaussian process (GP) machine learning (ML) approach enabled by modern probabilistic programming languages (PPLs). GP is an ML technique that treats predictions as distributions, rather than single points, providing a measure of prediction uncertainty; this is ideal for atmospheric inverse modeling (see Sect. 2.4), as posterior uncertainties are vital for providing quantitative information on the confidence level of the emission constraint. Inversion CO_2 models, other than analytical systems, cannot always provide posterior emission uncertainties, and these estimates can be unreliable and computationally expensive to calculate (e.g., Liu et al., 2014; Bousserez et al., 2015). The kernels (i.e., covariance functions) of GP models are employed to capture the intricate spatiotemporal correlation structures of OCO-2/3 data. PPLs

have been used in previous studies (e.g., Jeong et al., 2017, 2018), but modern PPLs provide significantly improved capabilities to implement GP models. Specifically, the built-in functions for GP kernels in modern PPLs enhance our ability to model the covariance structure of OCO-2/3 data.

This study applies inverse modeling techniques following the GP/ML methods described in further detail in Sect. 2.4 to estimate CO_2 fluxes in California for the full year of 2020 using XCO_2 observations from OCO-2 and OCO-3. The year 2020 had numerous anomalous features likely impacting total CO_2 fluxes in California, such as reduced anthropogenic emissions caused by coronavirus disease pandemic (COVID-19) lockdown procedures (Yañez et al., 2022), extreme wildfire activity (Jerret et al., 2022; Safford et al., 2022), and drought conditions (Steel et al., 2022). The impact that these types of events have on CO_2 fluxes are challenging to predict and difficult to replicate in bottom-up emission inventories. This study is structured as follows: Sect. 2 presents the forward and inverse models, satellite observations, and bottom-up emission inventories; Sect. 3 discusses the results of the study; and Sect. 4 contains the discussion and conclusions.

2 Methods

2.1 GEOS-Chem forward model

The forward model used to calculate atmospheric concentrations of CO_2 corresponding to OCO-2 and OCO-3 observations was the GEOS-Chem (version 14.0.1) chemical transport model (CTM) (Bey et al., 2001; Nassar et al., 2010). GEOS-Chem was used to simulate XCO_2 concentrations corresponding to each OCO-2 and OCO-3 retrieval for a nested North American domain (10–70° N, 40–140° W) driven by Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) meteorology at a $0.5° \times 0.625°$ spatial resolution using 47 vertical levels from the surface to 0.01 mbar. Chemical boundary conditions (BCs) of CO_2 used in the nested simulations were provided by a global GEOS-Chem-based 4D-Var data assimilation system that was run at a $4.0° \times 5.0°$ horizontal spatial resolution using 47 vertical levels. These global simulations of CO_2 for the year 2020 were constrained using inverse model methods through the assimilation of OCO-2 XCO_2 land nadir + land glint (LN + LG) retrievals and global in situ observations (Philip et al., 2019, 2022). The bottom-up emission inventories for CO_2 fluxes from fossil fuel (FF), net ecosystem exchange (NEE), wildfires, and oceans are described in Sect. 2.2. GEOS-Chem was initialized with chemical BCs and run for the entire year of 2020 with 2 months of spin-up time.

Total atmospheric CO_2 and source-apportioned (i.e., FF, NEE, fire, ocean, and boundary conditions) concentrations were calculated over California for all OCO-2 and OCO-3 observations. These source-attributed concentrations were calculated with sensitivity simulations by turning off indi-

vidual source fluxes or boundary conditions and comparing these results to the total atmospheric CO_2 concentration predictions from simulations with all sources included. Model-simulated XCO_2 corresponding to each OCO-2 and OCO-3 retrieval (H) were derived through the convolution of model CO_2 profiles with the column averaging kernel vector ($\boldsymbol{a}$) from OCO-2 and OCO-3 following Eq. (1):

$$H = XCO_{2a} + \boldsymbol{a}^T (\boldsymbol{f}(\sigma(\boldsymbol{x})) - \boldsymbol{c}_a), \qquad (1)$$

where prior profiles of CO_2 ($\boldsymbol{c}_a$) and prior column $CO_2 (XCO_{2a})$ represent prior information used in the OCO-2 and OCO-3 XCO_2 retrieval (O'Dell et al., 2012) and $\boldsymbol{f}(\sigma(\boldsymbol{x}))$ represents the GEOS-Chem-predicted vertical profiles of CO_2 interpolated to the retrieval levels of OCO-2 and OCO-3.

2.2 Bottom-up emission inventories

Bottom-up emission inventories used to drive GEOS-Chem simulations are described in Table 1, and seasonally averaged emission maps are displayed in Fig. S1 in the Supplement. The Vulcan version 3.0 FF emission inventory covers all anthropogenic source sectors of CO_2 in California (i.e., residential, commercial, industrial, electricity production, on-road, non-road, commercial marine vessel, airport, rail, and cement) between 2010–2015 (Gurney et al., 2020a). To create a spatially and temporally resolved Vulcan inventory in California for the year 2020 ($V2020_M$), the hourly 2015 Vulcan emissions ($V2015_M$) are scaled by an annual and a monthly scaling factor using Eq. (2). The sector-specific annual scaling factor ($R_{2020/2015}^{CARB}$) was calculated as the ratio of annual emissions from that sector in the CARB inventory for 2020 (which accounts for COVID-19 lockdown emission reductions; CARB, 2022) to the 2015 emissions. The sector-specific monthly scaling factor (R_M) was calculated from activity data from each sector, as the ratio of monthly activity to annual average activity, and used to appropriately distribute reductions due to the COVID-19 lockdown throughout the year.

$$V2020_M = V2015_M \times R_{2020/2015}^{CARB} \times R_M \qquad (2)$$

The Vulcan inventory for 2015 was then multiplied by these scaling factors to produce $V2020_M$. Both Vulcan and CARB provide the same sector-level emission estimates, so the scaling was done for each emission sector separately. The scaled 2020 Vulcan inventory was then aggregated to $0.1° \times 0.1°$ (latitude × longitude).

Natural CO_2 emission source (NEE, wildfire, and ocean) estimates were available for the year 2020, and no scaling was necessary. Biospheric fluxes of CO_2 were derived using monthly 5 km × 5 km NEE calculations from the Solar-Induced Fluorescence for Modeling Urban biogenic Fluxes version 1 (SMUrF v1; Wu et al., 2021) model. SMUrF calculates gross primary production (GPP), respiration (R_{eco}), and

Table 1. Bottom-up prior CO_2 emission inventories and the 2020 terrestrial carbon budget ($Tg\,CO_2\,yr^{-1}$) for California.

Source	Inventory name	Spatial resolution	Annual flux: California	Reference
FF	Vulcan	$1\,km \times 1\,km$	338.4	Gurney et al. (2020a)
NEE	SMUrF	$5\,km \times 5\,km$	−99.2	Wu et al. (2021)
Fire	GFED (modified)	$500\,m \times 500\,m$	103.3	van Wees et al. (2022)
Ocean	CarbonTracker (CT2022)	$100\,km \times 100\,km$	–	Jacobson et al. (2023)
Net			342.5	

NEE ($= R_{eco}$ – GPP) fluxes using (1) land cover type 500 m MODerate resolution Imaging Spectroradiometer (MODIS) data, (2) solar-induced fluorescence (SIF) from the OCO-2 sensor, (3) aboveground biomass at a 100 m resolution from GlobBiomass, (4) observed flux measurements from eddy-covariance towers, and (5) gridded soil and air temperature data products. Wildfire CO_2 emissions were implemented using a modified Global Fire Emissions Database version 4 (GFED4) dataset (van Wees et al., 2022). This modified version of GFED4 was produced using MODIS burned area and fire detection data with a spatial resolution of 500 m. Finally, oceanic CO_2 fluxes were derived from CarbonTracker (CT2022; Jacobson et al., 2023) $1° \times 1°$ output. These CT2022 coarse-spatial-scale fluxes were interpolated to match the GEOS-Chem model spatial resolution.

2.3 OCO-2 and OCO-3 observations

NASA has two operational satellites with the spatial resolution and precision necessary to constrain point-source to regional- and global-scale CO_2 emissions (i.e., OCO-2 and OCO-3). OCO-2 was launched in 2014 and is a Sun-synchronous polar-orbiting satellite which is in the Afternoon Constellation (A-train) of Earth observing satellites with a local overpass time of $\sim 13{:}30$ LT (local time) retrieving XCO_2 at a $1.3\,km \times 2.3\,km$ spatial resolution (Crisp et al., 2017). OCO-3 has been aboard the International Space Station (ISS) since 2019 and has an orbital inclination of $51.6°$, providing observations at varying times of the day (Eldering et al., 2019). OCO-3 makes orbital observations; however, it differs from OCO-2, as it has the capability to make snapshot area maps (SAMs) which cover $80\,km \times 80\,km$ at the native spatial resolution of $1.6\,km \times 2.2\,km$. The XCO_2 retrievals from OCO-2 and OCO-3 both use the Atmospheric Carbon Observations from Space (ACOS) algorithm (O'Dell et al., 2018), and this study applied version 11r and version 10.4r of OCO-2 and OCO-3, respectively. Retrievals of XCO_2 from the LN + LG retrieval modes were used for comparison to GEOS-Chem and to estimate posterior state-wide CO_2 emissions. As individual high-spatiotemporal-resolution OCO-2 and OCO-3 retrievals do not provide independent pieces of information, they are averaged to the $0.5° \times 0.625°$ spatial resolution of GEOS-Chem in this study.

In total, 1614 co-located model–satellite data points were available during 2020 to evaluate prior XCO_2 predictions and constrain posterior CO_2 emissions. The seasonal distribution (meteorological seasons: winter – December, January, and February – DJF; spring – March, April, and May – MAM; summer – June, July, and August – JJA; fall – September, October, and November – SON) of these co-locations was as follows: 386, 299, 551, and 378 for the winter, spring, summer, and fall months, respectively. The spatial distribution of the observational coverage provided by OCO-2 + OCO-3 during 2020 is displayed in Fig. S2.

2.4 Inverse model technique

The inverse model developed for this study used a GP/ML framework. GP is a flexible, nonparametric approach, distinguished by its use of hyperparameters, that defines a prior probability distribution over functions (Williams and Rasmussen, 2006; Biship, 2007; Murphy, 2022). A GP is fully characterized by its mean function $m(x)$ and kernel $\mathbf{k}(\mathbf{x}, \mathbf{x}')$:

$$f(x) \sim GP(m(x), k(x, x')), \tag{3}$$
$$y = f(x) + \epsilon, \tag{4}$$

where y is the OCO-2 and OCO-3 satellite observation vector, including additive noise ϵ (i.e., noisy version of $f(x)$). The noise term (ϵ) is modeled as $\epsilon \sim N(0, \sigma_{noise}^2 \mathbf{I})$, where $\mathbf{I}$ is the identity matrix and σ_{noise}^2 is the noise variance hyperparameter. As described in Sect. S3, σ_{noise} is assigned a half-Cauchy prior distribution, and its posterior is inferred using the No-U-Turn Sampler (NUTS; Hoffman and Gelman, 2014). Although sampling every possible value of the function $f(x)$ across a continuous domain is supported, we sample a finite set of points (i.e., OCO-2/3 observation time and locations), leading to a vector of function values, $f = [f(x_1), f(x_2), \ldots, f(x_N)]$, which follows a joint Gaussian distribution with mean vector $\mu = m(x_1), m(x_2), \ldots, m(x_N)$ and covariance matrix $[\mathbf{Cov}]_{i,j} = \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j)$. In this work, the terms "kernel" and "covariance function" are used synonymously.

For our flux inference application, we define the mean function $m(x)$ as follows:

$$m(x) = \mathbf{K}\lambda + \boldsymbol{D}, \tag{5}$$

where $\mathbf{K}$ is the input data, a $n \times k$ matrix, derived from GEOS-Chem model predictions; λ is a vector ($k \times 1$) of scaling factors, which quantify the adjustment required for our prior emission estimates to be consistent with observations; and $\boldsymbol{D}$ is the systematic bias. In this work, we estimate a single value of $\boldsymbol{D}$ for each month in 2020. Thus, each element of vector $\boldsymbol{D}$ is populated with the same value for each month of 2020. We assume that this bias term captures systematic bias due to instrument error, model transport error, and GEOS-Chem BC errors (Jeong et al., 2017). We show the probability density function of the estimated bias hyperparameter by month in Fig. S3. The median values range from -0.99 to 0.71 ppm, depending on the month. As noted, this value reflects the combined bias arising from atmospheric transport, boundary conditions, or other potential sources of error. This approach to addressing model bias has been applied in previous studies (e.g., Jeong et al., 2017). In this work, we included the bias term in the mean function, Eq. (5). As in Jeong et al. (2017), we model the bias term ($\boldsymbol{D}$) as a single component in the GP mean function due to the lack of prior information needed to separate it into identifiable sources (e.g., transport or boundary condition errors). Introducing multiple terms without such constraints would risk overfitting and model instability. Here, $\boldsymbol{D}$ and λ are considered a GP hyperparameter because they directly scale $m(x)$. This mean function has been widely adopted in atmospheric inverse analysis for estimating greenhouse gas emissions (Jeong et al., 2017; Ye et al., 2020; Ohyama et al., 2023). In GP modeling, it is important to note that the function $\mathbf{K}\lambda + \boldsymbol{D}$ is used as the mean of the latent (i.e., unknown) GP function, $f(x)$. In traditional Bayesian inversion methods (e.g., Jeong et al., 2017), the mean function is directly related to $\boldsymbol{y}$ in the form $\boldsymbol{y} = \mathbf{K}\lambda + \boldsymbol{D} + \epsilon$. The prior distributions for λ and other hyperparameters are described in Sect. S3.

The second component of a GP is the covariance function (i.e., GP kernel), which dictates how function values at different points relate. For the spatial part of the kernel, we employ the Matérn 5/2 kernel, a widely used covariance function for modeling spatial data (Bevilacqua et al., 2022). The Matérn 5/2 kernel between two spatial points can be expressed as follows:

$$\mathbf{k}\left(\mathbf{x}, \mathbf{x}'\right) = \left(1 + \frac{\sqrt{5}r l_s}{+} \frac{5r^2}{3l_s^2}\right) \exp\left(-\frac{\sqrt{5}r}{l_s}\right), \tag{6}$$

$$r = \sqrt{\left(x_1 - x_1'\right)^2 + \left(x_2 - x_2'\right)^2}, \tag{7}$$

where r is the Euclidean distance between the points x and x', x_1 and x_2 represent longitude and latitude, and l_s is the spatial length scale. The length scale is typically prescribed, estimated, or computed based on independent data (Baker et al., 2022). In this work, we estimate it simultaneously with other hyperparameters (e.g., the scaling factors). We used the squared exponential kernel for the temporal covariance to express the relationship between two temporal points:

$$\mathbf{k}_t\left(\mathbf{x}, \mathbf{x}'\right) = \exp\left(-\frac{\left(x_3 - x_3'\right)^2}{2l_t^2}\right), \tag{8}$$

where x_3 denotes the time and l_t is the temporal length scale. The spatiotemporal kernel matrix is then constructed by multiplying the spatial and temporal kernels:

$$\mathbf{k}_{st}\left(\mathbf{x}, \mathbf{x}'\right) = \sigma^2 \mathbf{k}_s\left(\mathbf{x}, \mathbf{x}'\right) \cdot \mathbf{k}_t\left(\mathbf{x}, \mathbf{x}'\right), \tag{9}$$

where σ^2 denotes the variance of the kernel, which scales the amplitude of the function values predicted by the GP. The spatiotemporal kernel, $\mathbf{k}_{st}$, is realized by element-wise multiplication of the spatial, $\mathbf{k}_s$, and temporal, $\mathbf{k}_t$, kernels. The resulting spatiotemporal kernel maintains the dimensionality of its constituent kernels.

We perform inversions using three distinct mean functions, as depicted in Eq. (5). Model 1 incorporates a systematic bias term $\boldsymbol{D}$, assuming a normal distribution with a mean of 0 and a standard deviation (SD) of 0.5 ppm. Model 2 resembles Model 1 but adopts a standard deviation of 1.0 ppm for the bias term. In contrast, Model 3 excludes the systematic bias term $\boldsymbol{D}$ and instead corrects the OCO-2 or OCO-3 a priori and background concentrations by applying scaling factors, thus addressing any biases in the OCO-2 or OCO-3 a priori and background concentrations multiplicatively. We evaluate the three GP models using their expected log pointwise predictive density (ELPD), a metric for model predictive performance. Further details on the model comparison through ELPD are provided in Sect. S1 and Fig. S4.

We employed the Markov chain Monte Carlo (MCMC) method to estimate the hyperparameters of the GP model framework. MCMC has been utilized in several atmospheric inverse modeling studies (Ganesan et al., 2014, 2015; Jeong et al., 2016, 2017, 2018). However, we adopt NUTS, a modern and advanced MCMC algorithm (Hoffman and Gelman, 2014). We utilized the PyMC PPL (Abril-Pla et al., 2023) to implement the NUTS algorithm for MCMC sampling, generating 4000 samples each month following a tuning phase of 3000 steps. More details on the GP model structure and the prior distribution for the hyperparameters are given in Sect. S2. Using the prior distributions specified in Sect. S3, we inferred the posterior distributions of the GP hyperparameters using the NUTS algorithm, implemented via the PyMC framework. In our approach, all hyperparameters are estimated jointly; that is, the scaling factors are estimated simultaneously with other parameters, such as the kernel length scales (Jeong et al., 2025). This method enables full posterior inference of the hyperparameters, offering a more robust characterization of uncertainty than point estimation methods. More details on hyperparameter optimization for the GP model are provided in Jeong et al. (2025).

2.5 Evaluation techniques

Prior and posterior emissions can be indirectly evaluated using atmospheric observations for accuracy and uncertainty applying normalized mean bias (NMB) and root-mean-square error (RMSE), respectively (Vermote and Kotchenova, 2008). GEOS-Chem forward and inverse model simulations were evaluated using daily OCO-2 and OCO-3 LN + LG XCO$_2$ retrievals during 2020. These model predictions were evaluated for each season to determine the accuracy of prior and posterior emissions and BCs which have large variability throughout the year (see seasonal a priori emissions in Fig. S1). General statistical parameters were used to evaluate model simulations: NMB, RMSE, correlation coefficient (R), and simple ordinary least-squares linear regression (slope, y intercept, etc.). Calculations of the NMB are normalized by OCO-2 and OCO-3 observation values, as shown in Eq. (10):

$$\text{NMB} = \frac{\sum_{i=1}^{N}(M_i - y_i)}{\sum_{i=1}^{N} y_i}, \tag{10}$$

where N is the total number of model (M_i) and OCO-2 and OCO-3 (y_i) co-locations. Equation (11) is used to calculate RMSE values:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(M_i - y_i)^2}{N}}. \tag{11}$$

3 Results

3.1 California prior emissions

According to prior emission inventories used in this study, the majority of CO$_2$ emitted in California is from anthropogenic FF sources (see Table 1). The Vulcan FF emission inventory, scaled to 2020 emissions using the CARB state-wide inventory, suggests that anthropogenic sources contributed 338.4 Tg CO$_2$ yr^{-1}, and these sources are primarily located in the Los Angeles Basin and San Francisco Bay areas, where there are highly populated cities (see Fig. S1). It is estimated that CO$_2$ emissions in 2020 were reduced by $\sim 10\,\%$ compared to 2019 due to COVID-19 restrictions (CARB, 2022). According to GFED4, a total of 103.3 Tg CO$_2$ yr^{-1} was emitted from biomass burning during 2020, which was one of the most active Californian wildfire years on record. Figure S1 shows that the majority of these emissions came from the large wildfires which occurred in northern and central California. These fire emissions were nearly offset by the biospheric uptake of CO$_2$ in California of -99.2 Tg CO$_2$ yr^{-1} estimated by the SMUrF model (i.e., our prior model). The largest NEE uptake was estimated to have occurred in the forested regions of northern California and the Sierra Nevada, while the largest respiration fluxes were in the Sacramento Valley and San Joaquin Valley areas and the Tulare Basin.

For emission sources other than FF, such as wildfire and NEE, CO$_2$ fluxes in the bottom-up data products had noticeable seasonality (see Fig. S1). Wildfires in 2020 had pronounced emissions during the summer and fall months compared to minimal emissions in the winter and spring, the latter of which comprise California's rainy season. The fire season of 2020 was exceptionally active, with multiple large complexes occurring between August and September (Keeley and Syphard, 2021). Prior emissions suggest that fires emitted 95.4 Tg CO$_2$ in California between August and September, which accounted for 92 % of the annual total. Biospheric fluxes also displayed large seasonality, with the highest uptake in the warmer growing season during the spring and summer and the highest respiration rates during the colder months of the winter and fall. NEE uptake peaked between May and June, with average monthly uptake rates of around -27.0 Tg CO$_2$, while respiration peaked between September and October, with average monthly rates of ~ 11.0 Tg CO$_2$. Less seasonality is apparent in Vulcan 2020 FF emissions for California, with monthly emission rates ranging between 23.0 and 32.0 Tg CO$_2$; however, our CARB-adjusted prior FF model does capture the decrease in anthropogenic CO$_2$ emissions upon the initiation of the COVID-19 lockdown during spring 2020.

3.2 Evaluation of model-simulated XCO$_2$ using prior emissions

To indirectly evaluate a priori bottom-up emissions, GEOS-Chem forward model simulations were evaluated with OCO-2 and OCO-3 XCO$_2$ retrievals. Figure 1 shows the comparison of modeled and satellite XCO$_2$ values using prior emissions and observations by season. A time series of daily co-located prior and posterior model-predicted XCO$_2$ compared to OCO-2/3 observations during the year 2020 is also displayed in Fig. S6 (histogram of annual prior and posterior residuals displayed in Fig. S7). For spring months, GEOS-Chem using prior emissions displayed a slight high bias (NMB = 1.1 ppm) and low correlation ($R = 0.38$), as the model did not capture the variability in XCO$_2$ retrieved by satellites. While the model captures the mean XCO$_2$ values observed, high and low values observed in the spring months were not replicated by the model (linear regression slope = 0.24). A similar evaluation was derived for the winter months, as the model had a similar high bias (NMB = 1.0 ppm), low correlation ($R = 0.39$), and relatively low linear regression slope (0.24). A somewhat different comparison was calculated between the model with prior emissions and observations for the summer and fall months. The GEOS-Chem simulations during the summer were able to capture the variability in satellite-retrieved XCO$_2$ values, with a high correlation ($R = 0.73$) and a linear regression slope of 0.75. The model and prior emissions resulted in a small negative bias during the summer months (NMB = -0.4 ppm). The prior model runs had the least bias

in the fall months (NMB $= -0.3$ ppm) and also displayed a moderate correlation ($R = 0.52$) and linear regression slope (0.34). The evaluation of the prior model displayed similar RMSE values throughout 2020, ranging between 1.4 and 1.8 ppm, with the largest random error in the fall months and the lowest values in the summer. GEOS-Chem using prior emission displayed biases and errors which varied by season, suggesting that observational constraint could improve the estimates of CO_2 emission in California. The following sections present the inversion of CO_2 emissions when assimilating satellite-derived XCO_2 values and the evaluation of posterior emissions.

3.3 Inverse GP model evaluation

This section describes the evaluation of the GP inversion model using posterior predictive checks (PPCs). PPCs ensure that the inversion results accurately represent the observed data (Gelman et al., 1996). The method involves using the posterior distribution of the model parameters to generate new datasets, which are then compared to the actual observed data. PPCs assess whether the model is capable of producing data similar to the observed data, thereby providing insight into the model's ability to capture the data-generating process accurately. Section S1 describes the comparison of the different GP inversion model setups and how Model 1 performs the most accurately. Due to the best performance by Model 1, the rest of the results in this study are based on these outputs. Figure 2 shows PPCs using probability density functions (PDFs) for the middle of each season (except January) employing Model 1. Due to an insufficient number of OCO-2 and OCO-3 XCO_2 observations ($N < 10$) in January, the PPC for February is included instead to represent the winter season. We construct the PDFs by utilizing local enhancements in XCO_2 concentrations after subtracting the OCO-2 or OCO-3 a priori XCO_2 and modeled BCs from the total satellite XCO_2 concentrations. The results in Fig. 2 demonstrate that the data generated from the Model 1 posterior parameters generally agree with observations.

The comparison between the posterior predictions from the GP inversion and the observed satellite XCO_2 data indicates an improvement in the RMSE for all seasons (posterior RMSE values on average $\sim 17\%$ lower compared to prior model simulations), suggesting a more accurate model fit than the initial prior predictions (see Fig. 1). The GP inversion was also able to remove the majority of systematic bias imposed by the prior emissions and BCs used in the nested GEOS-Chem simulations and to improve the correlation with satellite XCO_2 observations. For spring months, posterior model simulations displayed a small bias of ~ 0.3 ppm and a slightly improved correlation ($R = 0.39$) compared to prior model results. Posterior model results for the summer season displayed nearly zero bias and high correlation values of 0.79. The statistical evaluation of posterior model performance in the fall months improved compared to prior

simulations with a bias of ~ -0.1 ppm and a correlation of 0.57. Finally, for winter months, posterior results had a bias of ~ 0.1 ppm, a significant improvement on the value of 1.0 ppm from the prior result, and a moderate correlation of 0.41. Overall, posterior results from the GP inversion performed in this study proved to be more accurate compared to prior simulations, suggesting that the emission estimates from these inverse model runs are robust, as expected from the PPCs.

3.4 Posterior emissions by season and sector

We estimate state-wide posterior emissions by season and sector based on Model 1, which was evaluated to perform the best according to the ELPD metric results (see Sect. S1), and the seasonally averaged posterior emissions are displayed in Fig. 3. Figure 4 shows the seasonal state-wide total posterior CO_2 fluxes from all three GP inversion models and the prior estimates for each source sector in California during 2020 (monthly averaged prior and posterior state-wide CO_2 fluxes displayed in Fig. S8). In general, all three GP inversion models are relatively consistent with respect to median posterior emission estimates for all source sectors and seasons. This consistency suggests that the GP models are robust with respect to inferring posterior emissions, despite slight performance variations by season and sector for each model. The rest of the results discussed in this section are focused on posterior estimates from Model 1. Figure 4 shows that posterior FF emissions align closely with the prior estimates on a seasonal scale, indicating consistency between the initial assumptions and the inversion-derived results. Posterior FF emissions are most consistent with prior estimates during the spring and summer months, when COVID-19 lockdown restrictions were most strict, suggesting that corrections applied to the 2020 Vulcan data using CARB data were reasonable compared to observations. For the fall and winter seasons, posterior FF emission estimates were reduced by 10–15 Tg CO_2 compared to a priori assumptions, although the reduction is within the margin of error. Seasonal posterior 2σ uncertainty (95 % confidence level) had a range of 20–30 Tg CO_2, which is on average $\sim 30\%$ of the seasonal posterior median FF emission values. Interestingly, from the monthly averaged state-wide emissions shown in Fig. S8, it can be seen that some months in the spring, summer, and fall of 2020 had posterior FF fluxes that were further reduced compared to the prior emissions, emphasizing the strong reduction in greenhouse gas (GHG) emissions due to COVID-19 lockdown restrictions.

Posterior NEE fluxes from the GP inversion indicate that prior estimates assumed carbon uptake that was too strong during the drought year of 2020, suggesting an overestimation of the ecosystem's carbon sequestration capacity. Besides the fall season, posterior NEE was much less negative compared to the a priori fluxes, and it even transitioned from a small sink to a small source during the winter season
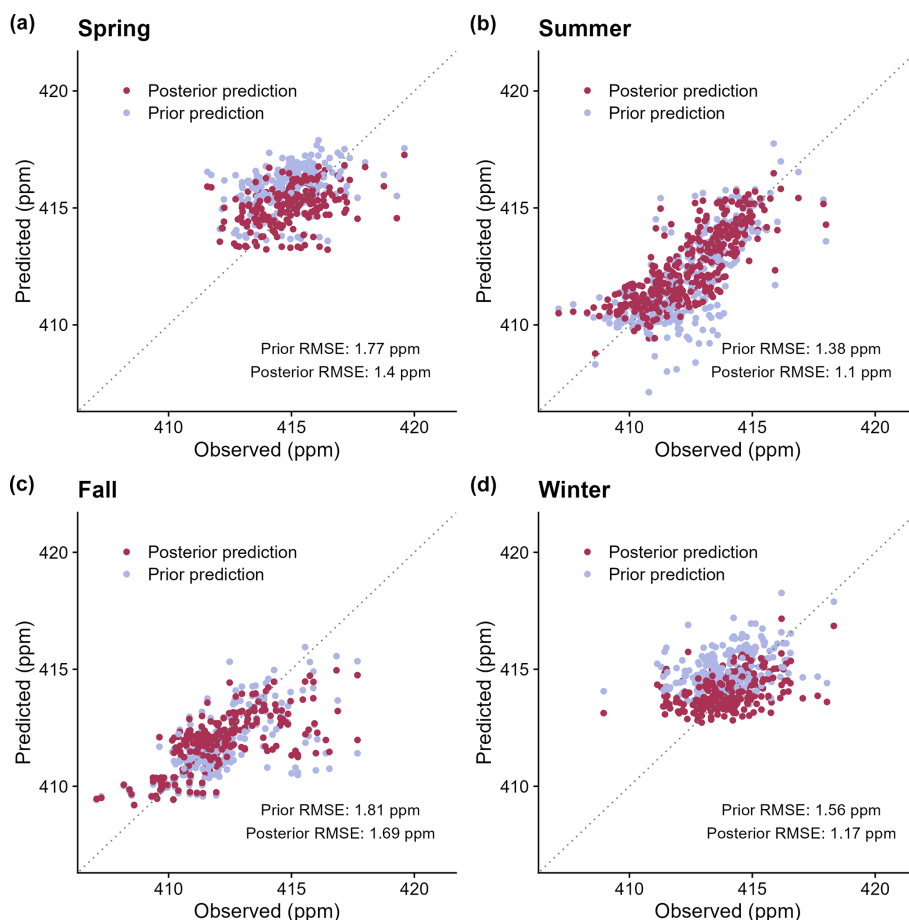
Figure 1. Comparison of seasonal GEOS-Chem XCO_2 predictions (ppm) using prior (blue) and posterior (purple) emissions and observed OCO-2 and OCO-3 concentrations (ppm). This result is based on the GP inversion model with a systematic bias of 0.5 ppm (Model 1). RMSE values for the prior and posterior model simulations are presented in the panel legends.

(see Figs. 3 and 4). Posterior NEE fluxes are 25–35 Tg CO_2 less (lower NEE) compared to prior estimates in the growing seasons of the spring and summer. From Fig. S8, it can be seen that posterior NEE fluxes were near neutral during some of the summer months, compared to the large uptake suggested by prior fluxes. This is likely due to the strong drought and hot temperatures experienced in California during 2020 greatly reducing the CO_2 uptake during the growing season. The posterior adjustments during the fall were smaller and tended to be consistent with prior estimates from SMUrF. Posterior NEE emissions were consistent with the prior estimates within a 2σ uncertainty range for the spring and fall seasons; however, they were not statistically consistent for winter and summer months. Seasonal posterior NEE displayed the largest uncertainty values of all source sectors in California, and these uncertainties were on average ∼ 95 % of the seasonal posterior median emission value.

The inversion results for fire emissions imply that the prior estimates are consistent with the posterior results within the 2σ uncertainty range, although the posterior median values for summer were lower than the prior. As expected, prior and posterior CO_2 emissions from fires were small during the winter and spring months. Posterior median seasonal total CO_2 emissions ranged between 20 and 50 Tg CO_2 for the summer and fall seasons, respectively. Constraints from OCO-2 and OCO-3 observations reduced emission estimates compared to the prior during both of these seasons with the largest reduction occurring for summer months (−21 %). Seasonal posterior fire emissions displayed moderate to high uncertainty values, and these uncertainties were on average ∼ 80 % of the seasonal posterior median emission values.

3.5 State-wide posterior total CO_2 emissions

This section describes the annual state-wide CO_2 flux estimates constrained using OCO-2 and OCO-3 observations for each source sector in 2020. Table 2 shows the results of the prior and posterior state-wide flux estimates for each source sector and the overall net terrestrial flux. The PDFs of these annual state-wide CO_2 fluxes are displayed in Fig. 5 (seasonal sector CO_2 flux PDFs shown in
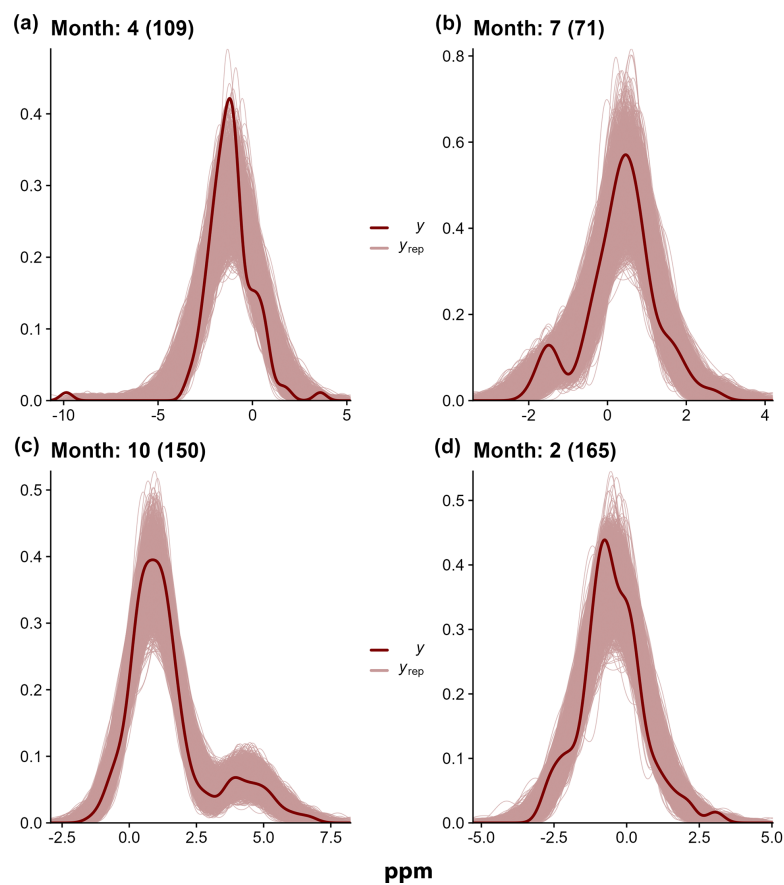
Figure 2. Evaluation of the GP inverse model (Model 1) performance using PPCs for months representative of each season in 2020 (April – spring; July – summer; October – fall; February – winter). The observed satellite XCO_2 data (y; in units of ppm) are represented by the bold lines, while the fine lines (y_{rep}) depict 4000 samples (in units of ppm) simulated with parameters drawn from the posterior distributions. Each sample (i.e., each fine line) for each month is of equivalent size to the number of the model–observation co-locations (noted in parentheses).

Fig. S9). Both the table and figure show that the net state-wide CO_2 fluxes from both prior and posterior estimates are nearly identical at between 340 and 350 $Tg\,CO_2\,yr^{-1}$. However, larger differences are evident when the state-wide annual emissions are broken down by source sector. Large constraints were imposed by OCO-2 and OCO-3 observations when focusing on NEE fluxes, where the posterior median estimate ($-36.8\,Tg\,CO_2\,yr^{-1}$; range of -71.7 to $-6.0\,Tg\,CO_2\,yr^{-1}$; 95 % confidence level) was 63 % lower (reduced carbon sink) compared to prior estimates ($-99.2\,Tg\,CO_2\,yr^{-1}$). Prior emissions from wildland fires were also reduced when constrained by satellite observations, as state-wide posterior estimates of $68.0\,Tg\,CO_2\,yr^{-1}$ (range of 24.9–126.2 $Tg\,CO_2\,yr^{-1}$; 95 % confidence level) were ~ 35 % lower compared to a priori estimates. Finally, posterior FF emissions were $317.8\,Tg\,CO_2\,yr^{-1}$ (range of 271.3–364.0 $Tg\,CO_2\,yr^{-1}$; 95 % confidence level), which is ~ 5 % lower compared to the prior estimates.

For total CO_2 fluxes, including all source sectors, California state-wide emissions are constrained with relatively high confidence using OCO-2 and OCO-3 XCO_2 observations, as

Table 2. The median prior and posterior (2σ range; 95 % confidence level) California CO_2 budget for 2020.

Source	Prior CO_2 flux ($Tg\,CO_2\,yr^{-1}$)	Posterior CO_2 flux ($Tg\,CO_2\,yr^{-1}$)
FF	338.4	317.8 (271.3 to 364.0)
NEE	−99.2	−36.8 (71.7 to −6.0)
Fire	103.3	68.0 (24.9 to 126.2)
Total	342.5	349.6 (272.8 to 428.6)

the 2σ standard deviation on this total flux is ~ 23 % of the annual median posterior estimate. Annual posterior emission estimates were most confident for FF sources, as the 2σ standard deviation from these sources was 47 $Tg\,CO_2$, which is ~ 15 % of the posterior median value. Natural fluxes of CO_2 (i.e., NEE and wildland fire) in California displayed higher uncertainties with respect to their posterior estimates, as indicated by the wider PDFs in Fig. 5. The 2σ standard deviation of annual posterior NEE fluxes was on average $\sim 35\,Tg\,CO_2$,
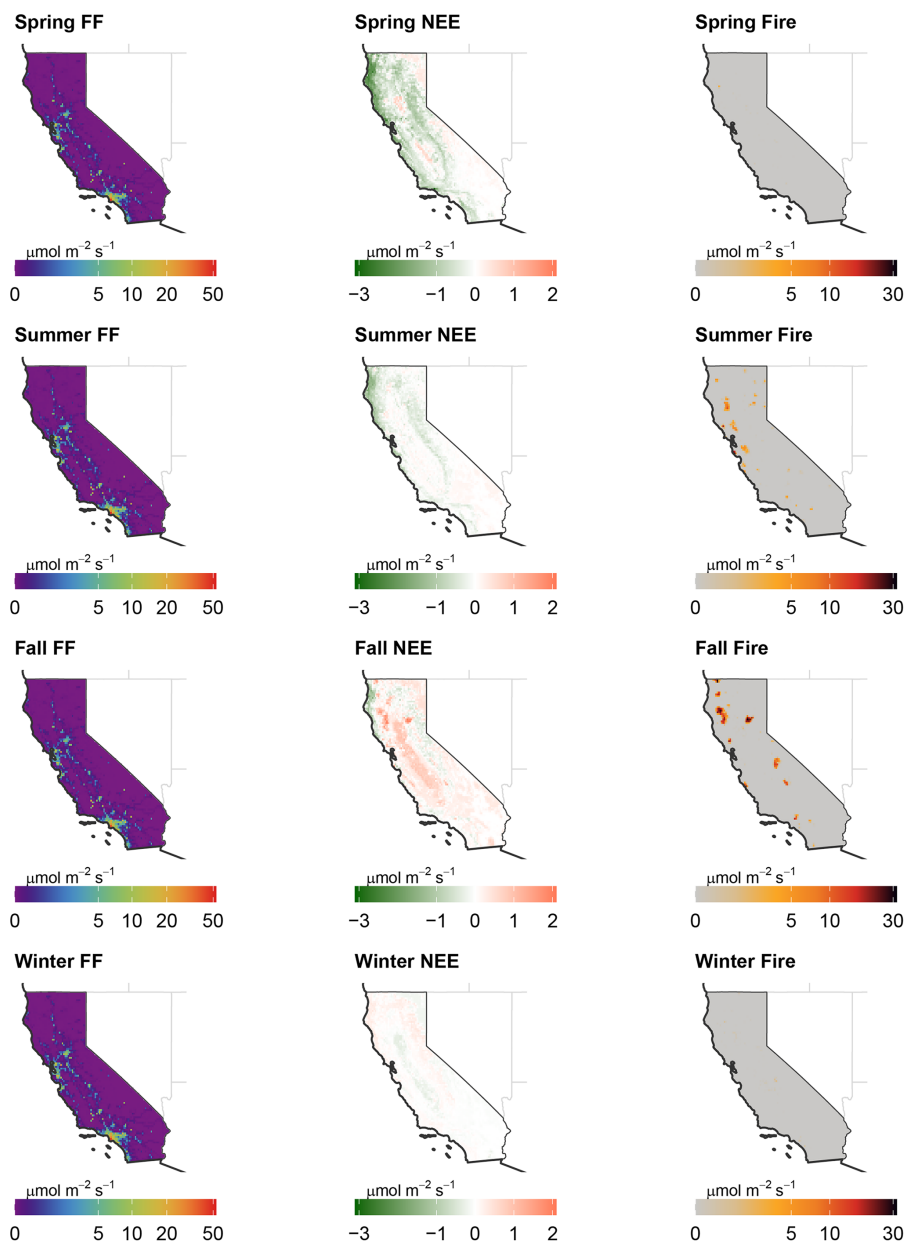
Figure 3. Seasonally averaged 2020 posterior CO_2 emissions ($\mu mol\,m^{-2}\,s^{-1}$) for the state of California. Emissions from the terrestrial portion of California are shown for FF (left column), NEE (middle column), and Fire (right column) for the spring (first row), summer (second row), fall (third row), and winter (fourth row) months.

which is $\sim 95\%$ of the posterior median value, indicating that this is the most uncertain carbon flux when using satellites to constrain emissions. Posterior annual fire emissions were also associated with larger uncertainty, as the 2σ uncertainty range was 43 Tg CO_2 (64 % of the median posterior flux).

4 Discussion and conclusions

This study presents the first attempt to constrain state-wide CO_2 fluxes from California using spaceborne XCO_2 observations employing both OCO-2 and OCO-3. We chose to focus on the year 2020, as this time period was characterized by anomalous features likely impacting total CO_2 fluxes in California, including reduced anthropogenic emissions caused by the COVID-19 lockdown (Yañez et al., 2022), elevated wildfire activity (Jerret et al., 2022; Safford et al., 2022),
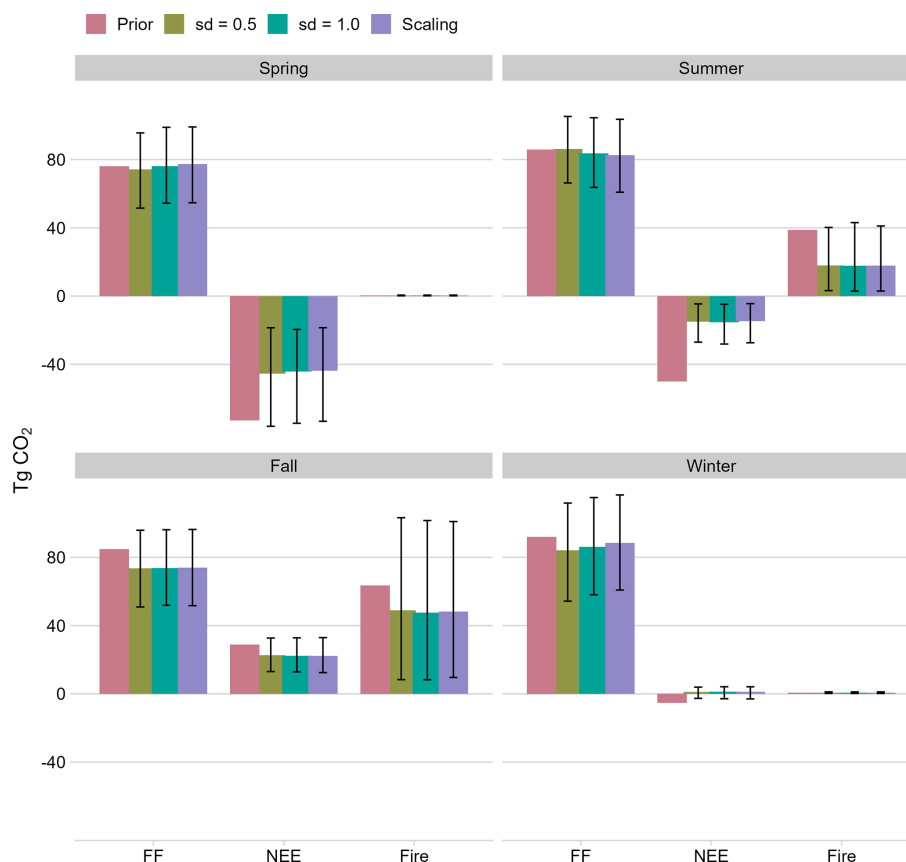
Figure 4. Sectoral emission estimates (Tg CO_2) by season from the scaled Vulcan a priori data and those using three distinct models: Model 1 ("SD = 0.5"), where a standard deviation (SD) of 0.5 ppm is applied to the prior probability distribution for the systematic bias; Model 2 ("SD = 1.0"), with a standard deviation of 1.0 ppm for the prior for the systematic bias; and Model 3 ("Scaling"), which optimizes the OCO-2 and OCO-3 a priori and model-predicted BC concentrations using scaling factors analogous to sector emission adjustments. The error bars in this figure reflect the 2σ uncertainty (i.e., 95 % confidence) values for each source sector. All 2σ confidence intervals were calculated using 4000 MCMC samples.

and drought occurrence (Steel et al., 2022). In this study, assimilating OCO-2 and OCO-3 LN + LG XCO$_2$ observations into a GP inversion framework was demonstrated to be effective for constraining state-wide CO_2 fluxes with a high degree of accuracy. The median posterior top-down annual total CO_2 flux of 349.6 Tg CO_2 yr^{-1} (range of 272.8–428.6 Tg CO_2 yr^{-1}; 95 % confidence level) was consistent with the a priori estimate and constrained with low 2σ uncertainty levels of ~ 23 %. The posterior uncertainty estimates of this work are similar to other recent studies that have used OCO-2 and OCO-3 XCO$_2$ data to constrain city-wide CO_2 emissions in California (e.g., Roten et al., 2023), other city flux estimates (e.g., Wu et al., 2020), and country-wide CO_2 budgets (e.g., Byrne et al., 2023). Our study adds to the growing evidence on how satellite XCO$_2$ data can be used to confidently estimate city- to country-scale CO_2 fluxes.

CARB inventories for the years 2019 and 2020 suggest that anthropogenic FF CO_2 emissions were reduced by ~ 10 % in 2020 compared to the year prior in California. In this study, the state-wide annual FF CO_2 source was

estimated, using the GP inversion assimilating OCO-2 and OCO-3 XCO$_2$ data for 2020, to be 317.8 Tg CO_2 yr^{-1}, which is ~ 5 % lower than the prior flux assumed. This top-down estimate is ~ 15 % higher compared to the CARB 2020 inventory, which calculated state-wide anthropogenic CO_2 emissions for 2020 to be 277.7 Tg CO_2 yr^{-1}. The state-wide FF CO_2 emissions estimated using OCO-2 and OCO-3 data in this study had posterior uncertainties of ~ 15 % on an annual scale and are, therefore, statistically consistent with the CARB 2020 inventory. The difference between the bottom-up and top-down median FF CO_2 emission estimate may be due to errors and uncertainties in the GP inversion and errors in the bottom-up CARB inventory, such as missing sources. It appears that the results in our study for FF emission estimates are robust, as they compare well to emission totals in California for the year 2020 from CARB and posterior top-down estimates are associated with low posterior uncertainty. Our PPC results, which compare the simulated data from posterior parameters with observations, provide additional confidence in our GP inversion models.
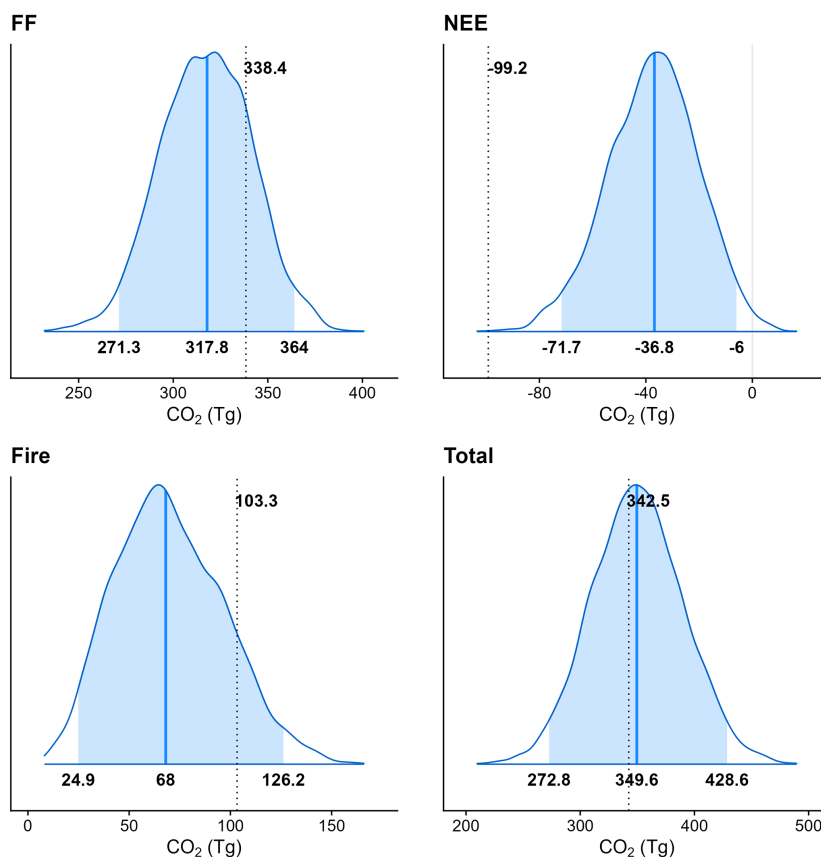
Figure 5. Annual CO_2 emission totals ($Tg\,CO_2$) for California by source sector in 2020. The numerical labels at the base of each PDF denote 2.5-, 50- (indicated by the bold vertical line), and 97.5-percentile estimates of the posterior emissions, respectively. The vertical dotted line indicates the prior emission estimate, with the corresponding value displayed. Note that the annual PDF presented here was derived by aggregating seasonal MCMC samples.

The main natural sources/sinks of CO_2 in California (i.e., NEE and wildland fire) were also estimated in this study using OCO-2 and OCO-3 XCO_2 data. The year 2020 was a year of drought in California and also resulted in extremely high levels of wildfire activity. The GP inversion resulted in posterior NEE fluxes that were greatly reduced compared to the initial best-guess a priori data. On an annual scale, the posterior estimate for NEE was $-36.8\,Tg\,CO_2\,yr^{-1}$, which was 63 % lower (reduced carbon sink) than prior estimates driven by satellite SIF retrievals. It is important to note that 2020 was towards the end of a multiyear drought that plagued California, and it would be expected that the terrestrial biosphere would be less effective in its uptake of carbon (Fu et al., 2022). It should also be noted that the median annual posterior NEE estimates derived in this study with satellite retrievals were associated with uncertainty levels of $\sim 95\,\%$. The larger uncertainty value associated with our posterior NEE estimates, compared to FF sources, is expected, as satellite retrievals are less sensitive to small, diffuse signals of CO_2 enhancements associated with the terrestrial biosphere compared to larger FF point sources. Wild-

fire activity was elevated in California during the time of this study, and we estimated that these sources contributed $68.0\,Tg\,CO_2\,yr^{-1}$ to the total state-wide annual carbon budget. The posterior estimate derived in our GP inversion was $\sim 35\,\%$ lower compared to the prior estimate; however, this estimate still represents highly elevated CO_2 emissions from this natural source. CARB estimated that $\sim 100\,Tg\,CO_2\,yr^{-1}$ was emitted from wildfires in 2020 (https://ww2.arb.ca.gov/wildfire-emissions, last access: 22 December 2023), which is in line with the prior estimate from GFED4 used in our study. CARB uses an emissions model that is similar to GFED4, so this is to be expected. The lower posterior wildfire estimate using our GP inversion system was associated with uncertainty levels ($\sim 64\,\%$ of the median posterior flux) that were lower compared to NEE and statistically consistent with the CARB 2020 state-wide estimate. Prior emission estimates from wildfires are generally uncertain, and satellite observations of the CO_2 resulting from these episodic events are challenging; thus, it is not surprising that posterior fire emissions are one of the more uncertain components of the 2020 California CO_2 budget.

Given that individual state- and country-wide CO_2 flux datasets generally have over a year of latency, satellite data become vital, as these spaceborne data are well equipped to provide more real-time estimates of these emissions. This is an important aspect of satellite data, especially during times of anomalous CO_2 fluxes due to economic activity, wildfire, or flood/drought. Both this study and the work by Roten et al. (2023) clearly demonstrated the ability of OCO-2 and, in particular, OCO-3 to help constrain FF emission estimates in California during the COVID-19 lockdown. OCO-3 is particularly effective for estimating city-wide (or other point-to area-source) fluxes using the data extracted from SAMs. These area-wide observations ($\sim 80\,\text{km} \times 80\,\text{km}$) greatly improve the observational coverage compared to OCO-2 (narrow swath of only $\sim 10\,\text{km}$). These SAMs allow for observations that reduce errors in assumptions about mixing between the sources and observations and illustrate intra-city variability in $X\text{CO}_2$, which has been shown to allow for sector-based emission constraints in California (Roten et al., 2023). The recent launch of satellites and future plans for spaceborne instruments that retrieve greenhouse gas concentrations (e.g., GHGSat, CO_2M, Carbon Mapper, etc.) at high spatial resolution and precision, some of which will apply SAM observational approaches, should greatly improve the ability to accurately estimate CO_2 emissions at city to global scales. As demonstrated, our GP inverse model has the potential to utilize these new satellite datasets to estimate surface emissions in a near-real-time fashion, effectively incorporating the unique spatiotemporal coverage of space-based information.

In evaluating the GP inversion method used in this study compared to linear inverse classical Bayesian inversion (CBI) models (e.g., $\mathbf{y} = \mathbf{K}\lambda + \epsilon$, commonly used in atmospheric inversions; Rodgers, 2000), advantages and disadvantages become apparent. The GP-based inversion method employed in this work offers several advantages over classical methods, as highlighted by Jeong et al. (2025). Jeong et al. (2025) presents the advantages of the GP method through inversion results from different approaches, and we briefly describe them here, focusing on the key points.

First, via simulations using multiple inverse modeling methods for constraining CO_2 fluxes in California when assimilating OCO-2 + OCO-3 $X\text{CO}_2$ observations, Jeong et al. (2025) demonstrated that the GP inversion yields superior results compared to the CBI method. Specifically, the CBI method failed to capture the FF scaling factor accurately at the 68 % confidence level. They repeated the inversion multiple times, and this result was consistent. Their work also showed that, without a proper prior distribution, a simple linear regression produced a physically implausible scaling factor for fire emissions. In our full Bayesian approach, we specify prior distributions for all parameters, including the scaling factors and kernel hyperparameters. Overall, the GP inversion offers substantial flexibility in modeling intri-

cate, nonlinear dependencies without needing a prespecified model framework (Ebden, 2015).

Second, the CBI method typically relies on analytical solutions and lacks robust techniques for estimating crucial parameters, such as hyperparameters (e.g., variance of the diagonal elements) for the covariance matrix (i.e., the kernel). Consequently, many previous inversion studies have used prescribed values for hyperparameters. For instance, these studies often utilized known values from other work (e.g., Roten et al., 2023) or estimations derived from sensitivity analyses to construct the model–data mismatch covariance (Gerbig et al., 2003; Jeong et al., 2013; Johnson et al., 2016). Such approaches do not guarantee that the estimates are consistent with the observed data. In contrast, Jeong et al. (2025) showed that the GP method can infer the noise variance, with its median value closely aligning with the true value. This represents a significant advancement over previous approaches, as it enables the direct estimation of true noise variance from the input data. Moreover, the GP inversion intrinsically provides quantification of uncertainty, which proves advantageous in scenarios with limited data, such as in atmospheric inversions.

Third, the GP method includes the spatiotemporal kernel as its essential component, as shown in this work. While some previous work (e.g., Turner et al., 2020) used a covariance with both spatial and temporal components, many previous inversion studies have not used a fully spatiotemporal covariance. This is because it is not straightforward to estimate the hyperparameters for the spatiotemporal covariance in the CBI method based on analytical solutions (Jeong et al., 2025). For example, the work by Turner et al. (2020) did not estimate the covariance parameters in a way consistent with the data. Incorporating the spatiotemporal covariance as a core component of the inversion is a significant advantage of the GP method over the CBI method.

While the GP-based inversion has many advantages over CBI methods, the computational demands of the GP method increase significantly with larger datasets, potentially restricting its application in certain contexts (Williams and Rasmussen, 2006; Murphy, 2022), although recent development for high-performance computing (e.g., GPU-enabled tools) can alleviate this issue. Conversely, the linear inverse model, while less computationally demanding, assumes linearity and typically requires explicit assumptions about the underlying distribution, which may not always be valid and can lead to underestimation of model uncertainty (Wang, 2023). Overall, this study demonstrates the clear advantages of using GP-based inversion techniques, and this modeling framework should be considered for application in future studies for constraining GHG fluxes when assimilating satellite retrievals.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., and Zinkov, R.: PyMC: a modern, and comprehensive probabilistic programming framework in Python, PeerJ Computer Science, 9, e1516, https://doi.org/10.7717/peerj-cs.1516, 2023.

Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., and Treanton, K.: A synthesis of carbon dioxide emissions from fossil-fuel combustion, Biogeosciences, 9, 1845–1871, https://doi.org/10.5194/bg-9-1845-2012, 2012.

Baker, D. F., Bell, E., Davis, K. J., Campbell, J. F., Lin, B., and Dobler, J.: A new exponentially decaying error correlation model for assimilating OCO-2 column-average CO_2 data using a length scale computed from airborne lidar measurements, Geosci. Model Dev., 15, 649–668, https://doi.org/10.5194/gmd-15-649-2022, 2022.

Bevilacqua, M., Caamaño-Carrillo, C., and Porcu, E.: Unifying compactly supported and Matérn covariance functions in spatial statistics, J. Multivariate Anal., 189, 104949, https://doi.org/10.1016/j.jmva.2022.104949, 2022.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q. B., Liu, H. G. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res.-Atmos., 106, 23073–23095, https://doi.org/10.1029/2001JD000807, 2001.

Biship, C. M.: Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, Berlin, Heidelberg, Germany, ISBN 978-0-387-31073-2, 2007.

Bousserez, N., Henze, D. K., Perkins, A., Bowman, K. W., Lee, M., Liu, J., Deng, F., and Jones, D. B. A.: Improved analysis error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model, Q. J. Roy. Meteor. Soc., 141, 1906–1921, https://doi.org/10.1002/qj.2495, 2015.

Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., Crisp, D., Crowell, S., Deng, F., Deng, Z., Deutscher, N. M., Dubey, M. K., Feng, S., García, O. E., Griffith, D. W. T., Herkommer, B., Hu, L., Jacobson, A. R., Janardanan, R., Jeong, S., Johnson, M. S., Jones, D. B. A., Kivi, R., Liu, J., Liu, Z., Maksyutov, S., Miller, J. B., Miller, S. M., Morino, I., Notholt, J., Oda, T., O'Dell, C. W., Oh, Y.-S., Ohyama, H., Patra, P. K., Peiro, H., Petri, C., Philip, S., Pollard, D. F., Poul-

ter, B., Remaud, M., Schuh, A., Sha, M. K., Shiomi, K., Strong, K., Sweeney, C., Té, Y., Tian, H., Velazco, V. A., Vrekoussis, M., Warneke, T., Worden, J. R., Wunch, D., Yao, Y., Yun, J., Zammit-Mangion, A., and Zeng, N.: National CO_2 budgets (2015–2020) inferred from atmospheric CO_2 observations in support of the global stocktake, Earth Syst. Sci. Data, 15, 963–1004, https://doi.org/10.5194/essd-15-963-2023, 2023.

California Air Resources Board: GHG Emissions Inventory (GHG EI) 2000–2020, California Air Resources Board [data set], https://ww2.arb.ca.gov/ghg-inventory-data (last access: 15 March 2024), 2022.

California Air Resources Board: California Greenhouse Gas Emissions from 2000 to 2021: Trends of Emissions and Other Indicators, California Air Resources Board [data set], https://ww2.arb.ca.gov/ghg-inventory-data (last access: 21 December 2023), 2023.

Crisp, D., Pollock, H. R., Rosenberg, R., Chapsky, L., Lee, R. A. M., Oyafuso, F. A., Frankenberg, C., O'Dell, C. W., Bruegge, C. J., Doran, G. B., Eldering, A., Fisher, B. M., Fu, D., Gunson, M. R., Mandrake, L., Osterman, G. B., Schwandner, F. M., Sun, K., Taylor, T. E., Wennberg, P. O., and Wunch, D.: The on-orbit performance of the Orbiting Carbon Observatory-2 (OCO-2) instrument and its radiometrically calibrated products, Atmos. Meas. Tech., 10, 59–81, https://doi.org/10.5194/amt-10-59-2017, 2017.

Cui, X., Newman, S., Xu, X., Andrews, A. E., Miller, J., and Lehman, S.: Atmospheric observation-based estimation of fossil fuel CO_2 emissions from regions of central and southern California, Sci. Total Environ., 664, 381–391, https://doi.org/10.1016/j.scitotenv.2019.01.081, 2019.

Cui, Y. Y., Zhang, L., Jacobson, A. R., Johnson, M. S., Philip, S., Baker, D., Chevallier, F., Schuh, A. E., Liu, J., Crowell, S., Peiro, H. E., Deng, F., Basu, S., and Davis, K. J.: Evaluating global atmospheric inversions of terrestrial net ecosystem exchange CO_2 over North America on seasonal and subcontinental scales, Geophys. Res. Lett., 49, e2022GL100147, https://doi.org/10.1029/2022GL100147, 2022.

Dunn, R. J. H., F. Aldred, N. Gobron, J. B Miller, and K. M. Willett (Eds.): Global Climate [in "State of the Climate in 2021"], B. Am. Meteorol. Soc., 103, S11–S142, https://doi.org/10.1175/BAMS-D-22-0092.1, 2022.

Ebden, M.: Gaussian Processes: A Quick Introduction, arXiv [preprint], https://doi.org/10.48550/arXiv.1505.02965, 2015.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data, Atmos. Meas. Tech., 12, 2341–2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

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, 3653–3671, https://doi.org/10.1002/2016JD025617, 2017.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Cham-

berlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

Fu, Z., Ciais, P., Prentice, I. C., Gentine, P., Makowski, D., Bastos, A., Luo, X., Green, J. K., Stoy, P. C., Yang, H., and Hajima, T.: Atmospheric dryness reduces photosynthesis along a large range of soil water deficits, Nat. Commun., 13, 989, https://doi.org/10.1038/s41467-022-28652-7, 2022.

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K.-R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P., and Weiss, R. F.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmos. Chem. Phys., 14, 3855–3864, https://doi.org/10.5194/acp-14-3855-2014, 2014.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. E., Sturges, W. T., Moncrieff, J. B., and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmos. Chem. Phys., 15, 6393–6406, https://doi.org/10.5194/acp-15-6393-2015, 2015.

Gelman, A., Meng, X. L., and Stern, H.: Posterior predictive assessment of model fitness via realized discrepancies, Stat. Sinica, 733–760, 1996.

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., 108, 4757, https://doi.org/10.1029/2003JD003770, 2003.

Graven, H., Fischer, M. L., Lueker, T., Jeong, S., Guilderson, T. P., Keeling, R. F., Bambha, R., Brophy, K., Callahan, W., Cui, X., Frankenberg, C., Gurney, K. R., LaFranchi, B. W., Lehman, S. J., Michelsen, H., Miller, J. B., Newman, S., Paplawsky, W., Parazoo, N. C., Sloop, C., and Walker, S. J.: Assessing fossil fuel CO_2 emissions in California using atmospheric observations and models, Environ. Res. Lett., 13, 065007, https://doi.org/10.1088/1748-9326/aabd43, 2018.

Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., Gong, D. S., Kaufman, S., Nnamchi, H.

C., Quaas, J., Rivera, J. A., Sathyendranath, S., Smith, S. L., Trewin, B., von Shuckmann, K., and Vose, R. S.: Changing State of the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 287–422, https://doi.org/10.1017/9781009157896.004, 2021.

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., and Roest, G.: The Vulcan Version 3.0 High-Resolution Fossil Fuel CO_2 Emissions for the United States, J. Geophys. Res.-Atmos., 125, e2020JD03297, https://doi.org/10.1029/2020jd032974, 2020.

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., and Roest, G.: Vulcan: High-Resolution Hourly Fossil Fuel CO_2 Emissions in USA, 2010–2015, Version 3, ORNL DAAC [data set], https://doi.org/10.3334/ORNLDAAC/1810, 2020.

Hamilton, S. D., Wu, D., Johnson, M. S., Turner, A. J., Fischer, M. L., Dadheech, N., and Jeong, S.: Estimating carbon dioxide emissions in two California cities using Bayesian inversion and satellite measurements, Geophys. Res. Lett., 51, e2024GL111150, https://doi.org/10.1029/2024GL111150, 2024.

Hedelius, J. K., Liu, J., Oda, T., Maksyutov, S., Roehl, C. M., Iraci, L. T., Podolske, J. R., Hillyard, P. W., Liang, J., Gurney, K. R., Wunch, D., and Wennberg, P. O.: Southern California megacity CO_2, CH_4, and CO flux estimates using ground- and space-based remote sensing and a Lagrangian model, Atmos. Chem. Phys., 18, 16271–16291, https://doi.org/10.5194/acp-18-16271-2018, 2018.

Hoffman, M. D. and Gelman, A.: The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo, J. Mach. Learn. Res., 15, 1593–1623, 2014.

Imasu, R., Matsunaga, T., Nakajima, M., Yoshida, Y., Shiomi, K., Morino, I., Saitoh, N., Niwa, Y., Someya, Y., Oishi, Y., Hashimoto, M., Noda, H., Hikosaka, K., Uchino, O., Maksyutov, S., Takagi, H., Ishida, H., Nakajima, T. Y., Nakajima, T., and Shi, C.:Greenhouse gases Observing SATellite 2 (GOSAT-2): mission overview, Progress in Earth and Planetary Science, 10, 33, https://doi.org/10.1186/s40645-023-00562-2, 2023.

Jacobson, A. R., Schuldt, K. N., Tans, P., Arlyn Andrews, Miller, J. B., Oda, T., Mund, J., Weir, B., Ott, L., Aalto, T., Abshire, J. B., Aikin, K., Aoki, S., Apadula, F., Arnold, S., Baier, B., Bartyzel, J., Beyersdorf, A., Biermann, T., Biraud, S. C., Boenisch, H., Brailsford, G., Brand, W. A., Chen, G., Huilin Chen, Lukasz Chmura, Clark, S., Colomb, A., Commane, R., Conil, S., Couret, C., Cox, A., Cristofanelli, P., Cuevas, E., Curcoll, R., Daube, B., Davis, K. J., De Wekker, S., Coletta, J. D., Delmotte, M., DiGangi, E., DiGangi, J. P., Di Sarra, A. G., Dlugokencky, E., Elkins, J. W., Emmenegger, L., Shuangxi Fang, Fischer, M. L., Forster, G., Frumau, A., Galkowski, M., Gatti, L. V., Gehrlein, T., Gerbig, C., Francois Gheusi, Gloor, E., Gomez-Trueba, V., Goto, D., Griffis, T., Hammer, S., Hanson, C., Haszpra, L., Hatakka, J., Heimann, M., Heliasz, M., Hensen, A., Hermansen, O., Hintsa, E., Holst, J., Ivakhov, V., Jaffe, D. A., Jordan, A., Joubert, W., Karion, A., Kawa, S. R., Kazan, V., Keeling, R. F., Keronen, P.,

Kneuer, T., Kolari, P., Kateřina Komínková, Kort, E., Kozlova, E., Krummel, P., Kubistin, D., Labuschagne, C., Lam, D. H. Y., Lan, X., Langenfelds, R. L., Laurent, O., Laurila, T., Lauvaux, T., Lavric, J., Law, B. E., Lee, J., Lee, O. S. M., Lehner, I., Lehtinen, K., Leppert, R., Leskinen, A., Leuenberger, M., Levin, I., Levula, J., Lin, J., Lindauer, M., Loh, Z., Lopez, M., Luijkx, I. T., Lunder, C. R., Machida, T., Mammarella, I., Manca, G., Manning, A., Manning, A., Marek, M. V., Martin, M. Y., Matsueda, H., McKain, K., Meijer, H., Meinhardt, F., Merchant, L., Mihalopoulos, N., Miles, N. L., Miller, C. E., Mitchell, L., Mölder, M., Montzka, S., Moore, F., Moossen, H., Morgan, E., Morgui, J.-A., Morimoto, S., Müller-Williams, J., Munger, J. W., Munro, D., Myhre, C. L., Nakaoka, S.-I., Necki, J., Newman, S., Nichol, S., Niwa, Y., Obersteiner, F., O'Doherty, S., Paplawsky, B., Peischl, J., Peltola, O., Piacentino, S., Pichon, J.-M., Pickers, P., Piper, S., Pitt, J., Plass-Dülmer, C., Platt, S. M., Prinzivalli, S., Ramonet, M., Ramos, R., Reyes-Sanchez, E., Richardson, S. J., Riris, H., Rivas, P. P., Ryerson, T., Saito, K., Sargent, M., Sasakawa, M., Scheeren, B., Schuck, T., Schumacher, M., Seifert, T., Sha, M. K., Shepson, P., Shook, M., Sloop, C. D., Smith, P., Stanley, K., Steinbacher, M., Stephens, B., Sweeney, C., Thoning, K., Timas, H., Torn, M., Tørseth, K., Trisolino, P., Turnbull, J., van den Bulk, P., van Dinther, D., Vermeulen, A., Viner, B., Vitkova, G., Walker, S., Watson, A., Wofsy, S. C., Worsey, J., Worthy, D., Young, D., Zaehle, S., Zahn, A., and Zimnoch, M.: CarbonTracker CT2022, NOAA Global Monitoring Laboratory, https://doi.org/10.25925/Z1GJ-3254, 2023.

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, 11339–11351, 2013.

Jeong, S., Newman, S., Zhang, J., Andrews, A. E., Bianco, L., Bagley, J., Cui, X., Graven, H., Kim, J., Salameh, P., LaFranchi, B. W., Priest, C., Campos-Pineda, M., Novakovskaia, E., Sloop, C. D., Michelsen, H. A., Bambha, R. P., Weiss, R. F., Keeling, R., and Fischer, M. L.: Estimating methane emissions in California's urban and rural regions using multitower observations, J. Geophys. Res.-Atmos., 121, 13031–13049, https://doi.org/10.1002/2016JD025404, 2016.

Jeong, S., Cui, X., Blake, D. R., Miller, B., Montzka, S. A., Andrews, A., Guha, A., Martien, P., Bambha, R. P., LaFranchi, B., Michelsen, H. A., Clements, C. B., Glaize, P., and Fischer, M. L.: Estimating methane emissions from biological and fossil-fuel sources in the San Francisco Bay Area, Geophys. Res. Lett., 44, 486–495, https://doi.org/10.1002/2016GL071794, 2017.

Jeong, S., Newman, S., Zhang, J., Andrews, A. E., Bianco, L., Dlugokencky, E., Bagley, J., Cui, X., Priest, C., Campos-Pineda, M., and Fischer, M. L.: Inverse Estimation of an Annual Cycle of California's Nitrous Oxide Emissions. J. Geophys. Res.-Atmos., 123, 4758–4771, https://doi.org/10.1029/2017JD028166, 2018.

Jeong, S., Hamilton, S. D., Johnson, M. S., Wu, D., Turner, A. J., and Fischer, M. L.: Applying Gaussian Process Machine Learning and Modern Probabilistic Programming to Satellite Data to Infer CO_2 Emissions, Environ. Sci. Technol., 59, 9, 4376–4387, https://doi.org/10.1021/acs.est.4c09395, 2025.

Jerret, M., Jina, A. S., and Marlier, M. E.: Up in smoke: California's greenhouse gas reductions could be wiped out by 2020 wildfires, Environ. Pollut., 310, 119888, https://doi.org/10.1016/j.envpol.2022.119888, 2022.

Johnson, M. S., Xi, X., Jeong, S., Yates, E. L., Iraci, L. T., Tanaka, T., Loewenstein, M., Tadic, J. M., and Fischer, M. L.: Investigating seasonal methane emissions in Northern California using airborne measurements and inverse modeling, J. Geophys. Res.-Atmos., 121, 13753–13767, https://doi.org/10.1002/2016JD025157, 2016.

Keeley, J. and Syphard, A.: Large California wildfires: 2020 fires in historical context, Fire Ecol., 17, 22, https://doi.org/10.1186/s42408-021-00110-7, 2021.

Kiel, M., Eldering, A., Roten, D. D., Lin, J. C., Feng, S., Lei, R., Lauvaux, T., Oda, T., Roehl, C. M., Blavier, J.-F., and Iraci, L. T.: Urban-focused satellite CO_2 observations from the Orbiting Carbon Observatory-3: A first look at the Los Angeles megacity, Remote Sens. Environ., 258, 112314, https://doi.org/10.1016/j.rse.2021.112314, 2021.

Lan, X., Tans, P., and Thoning, K. W.: Trends in globally-averaged CO_2 determined from NOAA Global Monitoring Laboratory measurements, Version 2023-09. National Oceanic and Atmospheric Administration, Global Monitoring Laboratory (NOAA/GML), https://gml.noaa.gov/ccgg/trends/global.html (last access: 12 December 2023), 2023.

Liu, J., Bowman, K., Lee, M., Henze, D., Bousserez, N., Brix, H., Collatz, G. J., Menemenlis, D., Ott, L., Pawson, S., Jones, D., and Nassar, R.: Carbon monitoring system flux estimation and attribution: impact of ACOS-GOSAT XCO_2 sampling on the inference of terrestrial biospheric sources and sinks, Tellus B, 66, 22486, https://doi.org/10.3402/tellusb.v66.22486, 2014.

Murphy, K. P.: Probabilistic Machine Learning: An Introduction, MIT Press, 864, ISBN 9780262046824, https://probml.github.io/pml-book/book1.html (last access: 7 February 2024), 2022.

Nassar, R., Jones, D. B. A., Suntharalingam, P., Chen, J. M., Andres, R. J., Wecht, K. J., Yantosca, R. M., Kulawik, S. S., Bowman, K. W., Worden, J. R., Machida, T., and Matsueda, H.: Modeling global atmospheric CO_2 with improved emission inventories and CO_2 production from the oxidation of other carbon species, Geosci. Model Dev., 3, 689–716, https://doi.org/10.5194/gmd-3-689-2010, 2010.

Nathan, B. J., Lauvaux, T., Turnbull, J. C., Richardson, S. J., Miles, N. L., and Gurney, K. R.: Source Sector Attribution of CO_2 Emissions Using an Urban CO CO_2 Bayesian Inversion System, J. Geophys. Res.-Atmos., 123, 13611–13621, https://doi.org/10.1029/2018JD029231, 2018.

NOAA: Global Monitoring Laboratory, NOAA [data set] https://gml.noaa.gov/aftp/products/carbontracker/co2/CT-NRT.v2022-1/fluxes/daily/ (last access: 14 March 2024), 2024.

Noël, S., Buchwitz, M., Hilker, M., Reuter, M., Weimer, M., Bovensmann, H., Burrows, J. P., Bösch, H., and Lang, R.: Greenhouse gas retrievals for the CO2M mission using the FOCAL method: first performance estimates, Atmos. Meas. Tech., 17, 2317–2334, https://doi.org/10.5194/amt-17-2317-2024, 2024.

OCO Science Team, Gunson, M., and Eldering, A.: OCO-3 Level 2 geolocated XCO_2 retrievals results, physical model, Retrospective Processing V10r, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://doi.org/10.5067/D9S8ZOCHCADE, 2021.

OCO Science Team, Payne, V., and Chatterjee, A.: OCO-2 Level 2 bias-corrected XCO_2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing V11.1r, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://doi.org/10.5067/8E4VLCK16O6Q, 2022.

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, 99–121, https://doi.org/10.5194/amt-5-99-2012, 2012.

O'Dell, C. W., Eldering, A., Wennberg, P. O., Crisp, D., Gunson, M. R., Fisher, B., Frankenberg, C., Kiel, M., Lindqvist, H., Mandrake, L., Merrelli, A., Natraj, V., Nelson, R. R., Osterman, G. B., Payne, V. H., Taylor, T. E., Wunch, D., Drouin, B. J., Oyafuso, F., Chang, A., McDuffie, J., Smyth, M., Baker, D. F., Basu, S., Chevallier, F., Crowell, S. M. R., Feng, L., Palmer, P. I., Dubey, M., García, O. E., Griffith, D. W. T., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Notholt, J., Ohyama, H., Petri, C., Roehl, C. M., Sha, M. K., Strong, K., Sussmann, R., Te, Y., Uchino, O., and Velazco, V. A.: Improved retrievals of carbon dioxide from Orbiting Carbon Observatory-2 with the version 8 ACOS algorithm, Atmos. Meas. Tech., 11, 6539–6576, https://doi.org/10.5194/amt-11-6539-2018, 2018.

Ohyama, H., Frey, M. M., Morino, I., Shiomi, K., Nishihashi, M., Miyauchi, T., Yamada, H., Saito, M., Wakasa, M., Blumenstock, T., and Hase, F.: Anthropogenic CO_2 emission estimates in the Tokyo metropolitan area from ground-based CO_2 column observations, Atmos. Chem. Phys., 23, 15097–15119, https://doi.org/10.5194/acp-23-15097-2023, 2023.

Pandey, S., Houweling, S., Krol, M., Aben, I., Chevallier, F., Dlugokencky, E. J., Gatti, L. V., Gloor, E., Miller, J. B., Detmers, R., Machida, T., and Röckmann, T.: Inverse modeling of GOSAT-retrieved ratios of total column CH_4 and CO_2 for 2009 and 2010, Atmos. Chem. Phys., 16, 5043–5062, https://doi.org/10.5194/acp-16-5043-2016, 2016.

Peiro, H., Crowell, S., Schuh, A., Baker, D. F., O'Dell, C., Jacobson, A. R., Chevallier, F., Liu, J., Eldering, A., Crisp, D., Deng, F., Weir, B., Basu, S., Johnson, M. S., Philip, S., and Baker, I.: Four years of global carbon cycle observed from the Orbiting Carbon Observatory 2 (OCO-2) version 9 and in situ data and comparison to OCO-2 version 7, Atmos. Chem. Phys., 22, 1097–1130, https://doi.org/10.5194/acp-22-1097-2022, 2022.

Philip, S., Johnson, M. S., Potter, C., Genovesse, V., Baker, D. F., Haynes, K. D., Henze, D. K., Liu, J., and Poulter, B.: Prior biosphere model impact on global terrestrial CO_2 fluxes estimated from OCO-2 retrievals, Atmos. Chem. Phys., 19, 13267–13287, https://doi.org/10.5194/acp-19-13267-2019, 2019.

Philip, S., Johnson, M. S., Baker, D. F., Basu, S., Tiwari, Y. K., Indira, N. K., Ramonet, M., and Poulter, B.: OCO-2 satellite-imposed constraints on terrestrial biospheric CO_2 fluxes over South Asia, J. Geophys. Res.-Atmos., 127, e2021JD035035, https://doi.org/10.1029/2021JD035035, 2022.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, Atmospheric, Oceanic and Planetary Physics, edited by: World-Scientific, Singapore, London, 240 pp., ISBN 978-9810227401, 2000.

Roten, D., Lin, J. C., Das, S., and Kort, E. A.: Constraining Sector-Specific CO_2 Fluxes Using Space-Based XCO_2 Observations Over the Los Angeles Basin, Geophys. Res. Lett., 50, e2023GL104376. https://doi.org/10.1029/2023GL104376, 2023.

Safford, H. D., Paulson, A. K., Steel, Z. L., Young, D. J. N., and Wayman, R. B.: The 2020 California fire season: A year like no other, a return to the past or a harbinger of the future?, Global Ecol. Biogeogr., 31, 2005–2025, https://doi.org/10.1111/geb.13498, 2022.

Steel, Z. L., Jones, G. M., Collins, B. M., Green, R., Koltunov, A., Purcell, K. L., Sawyer, S. C., Slaton, M. R., Stephens, S. L., Stine, P., and Thompson, C.: Mega-Disturbances Cause Rapid Decline of Mature Conifer Forest Habitat in California, Ecol. Appl., 33, e2763, https://doi.org/10.1002/eap.2763, 2022.

The International GEOS-Chem User Community: geoschem/GC-Classic: GCClassic 14.4.1 (14.4.1), Zenodo [code], https://doi.org/10.5281/zenodo.12584192, 2024.

Turner, A. J., Kim, J., Fitzmaurice, H., Newman, C., Worthington, K., Chan, K., Wooldridge, P. J., Köehler, P., Frankenberg, C., and Cohen, R. C.: Observed impacts of covid-19 on urban CO_2 emissions, Geophys. Res. Lett., 47, e2020GL090037, https://doi.org/10.1029/2020GL090037, 2020.

United States Environmental Protection Agency: Inventory of U.S. Greenhouse Gas Emissions and Sinks, https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks (last access: 21 December 2023), 2023.

van Wees, D., van der Werf, G. R., Randerson, J. T., Rogers, B. M., Chen, Y., Veraverbeke, S., Giglio, L., and Morton, D. C.: Global biomass burning fuel consumption and emissions at 500 m spatial resolution based on the Global Fire Emissions Database (GFED), Geosci. Model Dev., 15, 8411–8437, https://doi.org/10.5194/gmd-15-8411-2022, 2022.

van Wees, D., van der Werf, G. R., Randerson, J. T., Rogers, B. M., Chen, Y., Veraverbeke, S., Giglio, L., and Morton, D. C.: Model data for "Global biomass burning fuel consumption and emissions at 500-m spatial resolution based on the Global Fire Emissions Database (GFED)" (Version v2) [Data set], Zenodo [data set], https://doi.org/10.5281/zenodo.12670427, 2024.

Vermote, E. F. and Kotchenova, S.: Atmospheric correction for the monitoring of land surfaces, J. Geophys. Res., 113, D23S90, https://doi.org/10.1029/2007JD009662, 2008.

Wang, J.: An Intuitive Tutorial to Gaussian Processes Regression, arXiv [preprint], https://doi.org/10.48550/arXiv.2009.10862, 2023.

Williams, C. K. and Rasmussen, C. E.: Gaussian processes for machine learning, in: vol. 2, MIT Press, Cambridge, MA, ISBN 026218253X, 2006.

Wu, D. and Lin, J. C.: Urban Biogenic CO_2 fluxes: GPP, Reco and NEE Estimates from SMUrF, 2010–2019, ORNL DAAC [code], https://doi.org/10.3334/ORNLDAAC/1899, 2021.

Wu, D., Lin, J., Oda, T., and Kort, E.: Space-based quantification of per capita CO_2 emissions from cities, Environ. Res. Lett., 15, 035004, https://doi.org/10.1088/1748-9326/ab68eb, 2020.

Wu, D., Lin, J. C., Duarte, H. F., Yadav, V., Parazoo, N. C., Oda, T., and Kort, E. A.: A model for urban biogenic CO_2 fluxes: Solar-Induced Fluorescence for Modeling Urban biogenic Fluxes (SMUrF v1), Geosci. Model Dev., 14, 3633–3661, https://doi.org/10.5194/gmd-14-3633-2021, 2021.

Wu, D., Liu, J., Wennberg, P. O., Palmer, P. I., Nelson, R. R., Kiel, M., and Eldering, A.: Towards sector-based attribution using intra-city variations in satellite-based emission ratios between CO_2 and CO, Atmos. Chem. Phys., 22, 14547–14570, https://doi.org/10.5194/acp-22-14547-2022, 2022.

Yañez, C. C., Hopkins, F. M., Xu, X., Tavares, J. F., Welch, A., and Czimczik, C. I.: Reductions in California's urban fossil fuel CO_2 emissions during the COVID-19 pandemic, AGU Advances, 3, e2022AV000732. https://doi.org/10.1029/2022AV000732, 2022.

Yang, D. X., Liu, Y., Feng, L., Wang, J., Yao, L., Cai, Z. N., Zhu, S. H., Lu, N. M., and Lyu, D. R.: The First Global Carbon Dioxide Flux Map Derived from TanSat Measurements, Adv. Atmos. Sci., 38, 1433–1443, 2021.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., Yang, E. G., and Wu, D.: Constraining Fossil Fuel CO_2 Emissions From Urban Area Using OCO-2 Observations of Total Column CO_2, J. Geophys. Res.-Atmos., 125, e2019JD030528, https://doi.org/10.1029/2019JD030528, 2020.