

A Bayesian framework for deriving sector-based methane emissions from top-down fluxes

Daniel H. Cusworth^{1,2✉}, A. Anthony Bloom², Shuang Ma², Charles E. Miller², Kevin Bowman², Yi Yin³, Joannes D. Maasakkers⁴, Yuzhong Zhang^{5,6}, Tia R. Scarpelli⁷, Zhen Qu⁸, Daniel J. Jacob^{7,8} & John R. Worden²

Atmospheric methane observations are used to test methane emission inventories as the sum of emissions should correspond to observed methane concentrations. Typically, concentrations are inversely projected to a net flux through an atmospheric chemistry-transport model. Current methods to partition net fluxes to underlying sector-based emissions often scale fluxes based on the relative weight of sectors in a prior inventory. However, this approach imposes correlation between emission sectors which may not exist. Here we present a Bayesian optimal estimation method that projects inverse methane fluxes directly to emission sectors while accounting uncertainty structure and spatial resolution of prior fluxes and emissions. We apply this method to satellite-derived fluxes over the U.S. and at higher resolution over the Permian Basin to demonstrate that we can characterize a sector-based emission budget. This approach provides more robust comparisons between different top-down estimates, critical for assessing the efficacy of policies intended to reduce emissions.

¹ Arizona Institutes for Resilience, University of Arizona, Tucson, AZ, USA. ² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ³ Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. ⁴ SRON Netherlands Institute for Space Research, Utrecht, Netherlands. ⁵ Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, Zhejiang Province, China. ⁶ Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang Province, China. ⁷ Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. ⁸ School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. ✉email: dcusworth@arizona.edu

As the second highest greenhouse gas (GHG) contributor to global radiative forcing, understanding the global budget of methane (CH_4) is a top climate priority^{1,2}. Methane is emitted from a variety of anthropogenic and natural emission sectors, including oil and gas operations, waste management, coal mining, agriculture, wetlands, and fires among others³. Article 14 of the Paris Agreement⁴ requires participating countries to report progress towards achieving their climate mitigation goals, or nationally determined contributions. Reporting progress, including any changes in the CH_4 budget, necessitates inventorying all possible emission sources. CH_4 emission inventories can be constructed from “bottom-up” or derived from “top-down” observations. Bottom-up accounting relies on a knowledge of activity data and emission factors for anthropogenic sectors and/or detailed processed-based models that predict CH_4 emissions based on a set of environmental factors for natural emission sectors. By aggregating an ensemble of bottom-up inventories and process-models, Saunio et al.⁵ calculated a global methane budget for 2008–2017 and estimated total emissions of 594–880 $\text{TgCH}_4 \text{ a}^{-1}$, with 113–154 $\text{TgCH}_4 \text{ a}^{-1}$ from fossil fuels, 191–223 $\text{TgCH}_4 \text{ a}^{-1}$ from agriculture and waste, 26–40 $\text{TgCH}_4 \text{ a}^{-1}$ from biomass and biofuel burning, 102–182 $\text{TgCH}_4 \text{ a}^{-1}$ from wetlands, and 143–306 $\text{TgCH}_4 \text{ a}^{-1}$ from other natural sources. Uncertainty and bias in bottom-up CH_4 emissions in some geographic regions may be caused by imprecise emission factors and activity data that are not readily available at necessary spatial and temporal scales, or by process-based models that perform poorly due to a host of environmental factors (e.g., wetland models rely on wetland inundation maps, biogeochemical process parameterizations, and knowledge of carbon availability^{6,7}).

Atmospheric observations of CH_4 , combined with an atmospheric transport model and a regularizing statistical approach, provide a top-down constraint on the global CH_4 budget. Typically referred to as “inversions” or top-down inventories, these methods estimate CH_4 fluxes by assimilating tower, aircraft, or satellite-based CH_4 measurements^{8–14}. Generally, these top-down methods only estimate total fluxes (i.e., sum of all emission sector contributions) explicitly, and may rely on using prior ratios or relative weights (RWs) between source categories to partition fluxes to specific source sectors. Top-down inverse models may be driven by different regularization or prior conditions, complicating direct comparison between an ensemble of inventories. Therefore, these partitioning approaches are prone to error when comparing with bottom-up inventories if the prior distribution of emissions is biased, or if different sectors have different uncertainties. For example, Saunio et al. compared an ensemble of 22 top-down CH_4 global inversions to an ensemble of bottom-up inventories, and found the bottom-up estimate to be 30% higher than the top-down ensemble mean. The study attributes much of this total discrepancy to large differences in non-wetland natural sources (e.g., lakes and rivers, oceans seeps, termites, geologic sources, wild animals, etc.). However, when integrating emissions over the whole globe, they find that other source categories from bottom-up and top-down approaches are consistent within their reported uncertainties (fossil fuel top-down: 81–131 Tg a^{-1} , bottom-up: 113–154 Tg a^{-1} ; agriculture+waste top-down: 207–240 Tg a^{-1} , bottom-up: 191–223 Tg a^{-1} ; and biomass+biofuel burning top-down: 22–36 Tg a^{-1} , bottom-up: 26–40 Tg a^{-1}). Reported uncertainties reflect the range of estimates among distinct bottom-up and top-down inventories across various emission sectors. Saunio et al. also relied on using RWs to partition top-down fluxes to individual bottom-up sectors. An explicit approach for comparison between top-down and bottom-up inventories, and between independent top-down based inventories, is needed to reduce this uncertainty in the global CH_4 budget by sector and by region.

Bayesian estimation of emissions from fluxes. We propose a comprehensive Bayesian framework that derives top-down gridded CH_4 emissions ($\hat{\mathbf{z}}$) and their error covariance ($\hat{\mathbf{Z}}$) from inverse fluxes ($\hat{\mathbf{x}}$) and their error covariance ($\hat{\mathbf{S}}$) without reliance on RWs from an inventory. This framework takes the following form:

$$\hat{\mathbf{z}} = \mathbf{z}_A + \hat{\mathbf{Z}}\mathbf{M}^T\hat{\mathbf{S}}^{-1}[(\mathbf{I} - \hat{\mathbf{S}}\mathbf{S}_A^{-1})(\mathbf{x}_A - \mathbf{M}\mathbf{z}_A) + (\hat{\mathbf{x}} - \mathbf{x}_A)] \quad (1)$$

Where the vector $\hat{\mathbf{x}}$ represents inverse CH_4 fluxes on a spatial grid with full error characterization given by the covariance matrix $\hat{\mathbf{S}}$, \mathbf{x}_A is the vector of prior fluxes, \mathbf{I} is the identity matrix, \mathbf{S}_A is the prior error covariance matrix, and \mathbf{M} is an aggregation matrix that sums emissions to fluxes (Methods section: Eqs. 8–9). The posterior emission error covariance matrix $\hat{\mathbf{Z}}$ is calculated explicitly given \mathbf{M} , \mathbf{S}_A , $\hat{\mathbf{S}}$, and prior emissions error covariance matrix \mathbf{Z}_A :

$$\hat{\mathbf{Z}} = (\mathbf{M}^T(\hat{\mathbf{S}}^{-1} - \mathbf{S}_A^{-1})\mathbf{M} + \mathbf{Z}_A^{-1})^{-1} \quad (2)$$

The full derivation is documented in the Methods section and Supporting Information (SI) and has been mechanically verified and tested using simulated emissions, concentrations, and fluxes. This approach has been previously described for atmospheric trace gas retrievals¹⁵ but is here modified and applied for flux comparisons.

Applying Eqs. 1 and 2 allows for the ability to “swap priors.” This means that no matter what prior was used in an initial flux inversion, we can swap it with a different prior emissions vector \mathbf{z}_A , which can include sector-based information. This a critical component for comparison between two different top-down inventories, as it removes error that may arise from choice of prior. Another advantage of this Bayesian approach is that fluxes are partitioned according to not just the prior emission state \mathbf{z}_A , but also according prior uncertainties on those emissions (i.e., \mathbf{Z}_A). For example, if we have evidence that a particular emission sector is well-characterized in bottom-up inventories (i.e., a tight prior uncertainty), Eqs. 1–2 take this knowledge into account when optimizing emissions. In this sense, our approach is similar to updated methods that use prior ratios with prior error variance when computing RWs used for partitioning¹⁶. However, any RW-based approach still assumes that correlation exists between emission sectors at the grid-level, creating a relationship that may bias results depending on prior construction.

The main caveat of this Bayesian approach is that an explicit representation of $\hat{\mathbf{S}}$ is needed. For analytical flux inversions, this matrix has a closed-form representation that is computed as part of the inversion. For adjoint-based inversions^{10,11}, a closed-form representation of $\hat{\mathbf{S}}$ is not directly computed. Though the error covariance can still be estimated^{17,18}, this is computationally expensive and is generally not done. Analytical frameworks are best suited for direct comparison between inverse products due to explicit error characterization.

Results

In what follows, we show examples of this approach using a previously performed 2010–2015 GOSAT flux inversion of the Continental United States⁹ (CONUS). We also apply Eqs. 1 and 2 to a previously performed 2018–2019 TROPOMI flux inversion over the Permian Basin in western Texas, southern New Mexico¹⁹. We compare with the partitioned results from the 2010–2015 GOSAT inversion to show how projection to a common prior can be used to assess regional emission trends since uncertainties from different priors and different spatial resolutions of the flux inversions are removed.

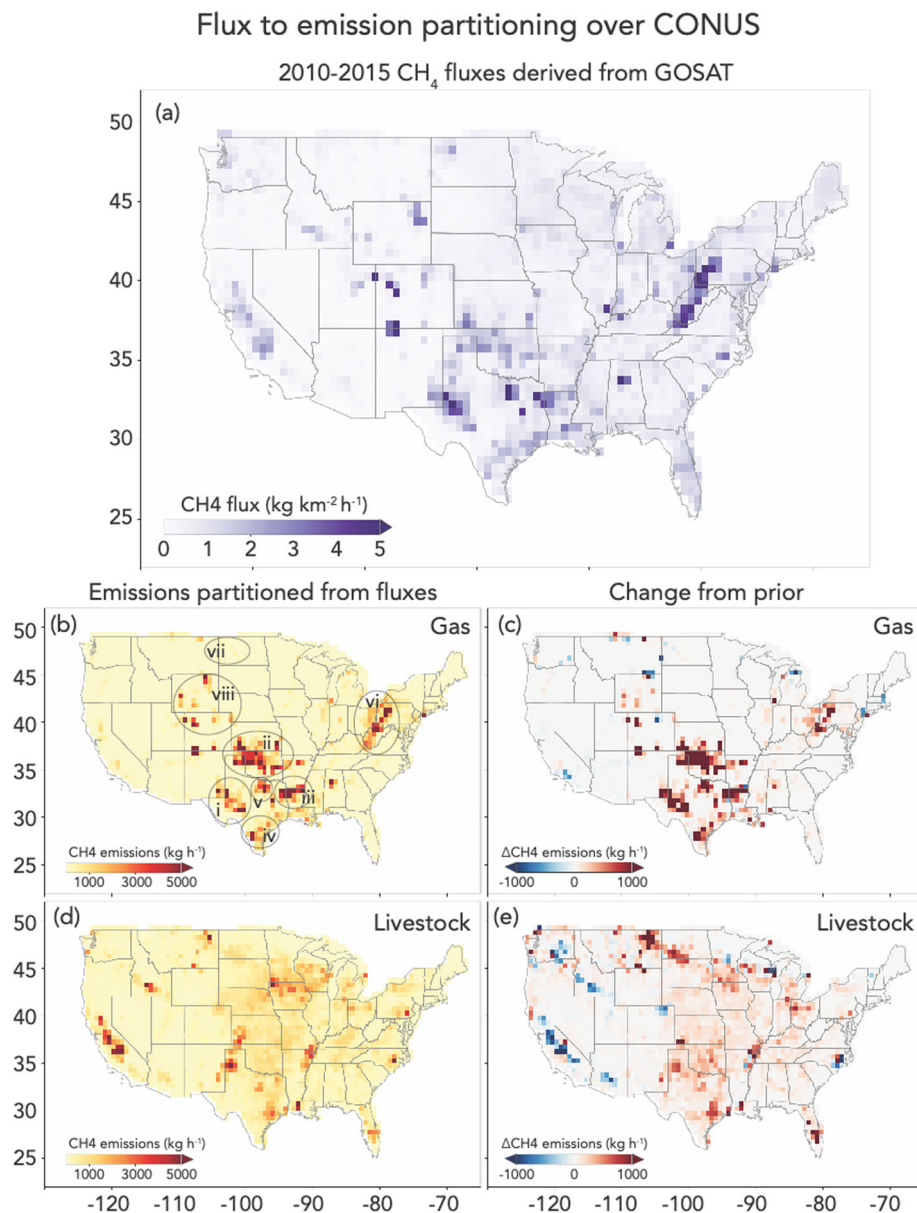


Fig. 1 Optimized total CH₄ emissions per grid cell over CONUS using 2010–2015 GOSAT inverse fluxes⁶. Panel **a** shows CH₄ fluxes at $0.5 \times 0.625^\circ$ resolution. Panels **b** and **c** show optimized emissions and the change from the prior for the gas sector, respectively at $0.1 \times 0.1^\circ$ resolution. Circled boxes in Panel **b** show rough outlines of major U.S. gas producing basins: (i) Permian, (ii) Anadarko, (iii) Haynesville, (iv) Eagle Ford, (v) Barnett Shale, (vi) Appalachia, (vii) Bakken, and (viii) Niobrara. Panels **d** and **e** show optimized emissions and the change from the prior for the livestock sector, respectively.

Partitioning CH₄ fluxes over CONUS. We apply the partitioning algorithm described by Eqs. 1 and 2 to a 2010–2015 $0.50 \times 0.625^\circ$ resolution North American CH₄ flux inversion⁶ performed using Greenhouse Gas Observing Satellite²⁰ (GOSAT) dry air column mixing ratios of CH₄. We partition emission to seven distinct sectors at $0.1 \times 0.1^\circ$ resolution: oil, gas, coal, livestock, waste management, wetlands, and other emission sources (soil, fire, etc.). For oil, gas, and coal prior emissions and uncertainties, we use a global inventory for 2016 based on national reports to the United Nations Framework Convention on Climate Change (Scarpelli et al.²¹). For wetland prior emissions and uncertainties, we take the ensemble average and standard deviation of wetland models that were found to be the highest performing when compared with a global CH₄ flux inversion²². For all other emission sectors, we use the 2012 EPA gridded CH₄ inventory as the prior estimate²³, and assume a one standard deviation uncertainty equal to 50% of the mean value. For illustration and

computational tractability, the prior error covariances are represented as diagonal matrices.

Figure 1a shows the inverse fluxes \hat{x} over CONUS, optimized emissions \hat{z} for the gas and livestock sectors when applying Eqs. 1–2, and the change in emissions compared to the emission prior ($\hat{z} - z_A$). Other optimized emission sectors are shown in Fig. S1. Optimized oil and gas emissions show large changes from the prior at the basin scale in several major producing basins and shale plays: Permian (New Mexico/Texas; ΔCH_4 : gas/oil = 0.40/0.76 TgCH₄ a⁻¹), Eagle Ford (southern Texas; ΔCH_4 = 0.14/0.11 TgCH₄ a⁻¹), Haynesville (Texas/Louisiana; ΔCH_4 = 0.18/0.0 TgCH₄ a⁻¹), Barnett Shale (Texas; ΔCH_4 = 0.21/0.0 TgCH₄ a⁻¹), Anadarko (Oklahoma; ΔCH_4 = 0.52/0.05 TgCH₄ a⁻¹), and the Appalachia Basin (Ohio/Pennsylvania/West Virginia; ΔCH_4 = 0.20/0.0 TgCH₄ a⁻¹). The Niobrara (Wyoming/Colorado; ΔCH_4 = 0.05/0.0 TgCH₄ a⁻¹) region shows much less or no change from the prior. For the Bakken (Montana/North Dakota;

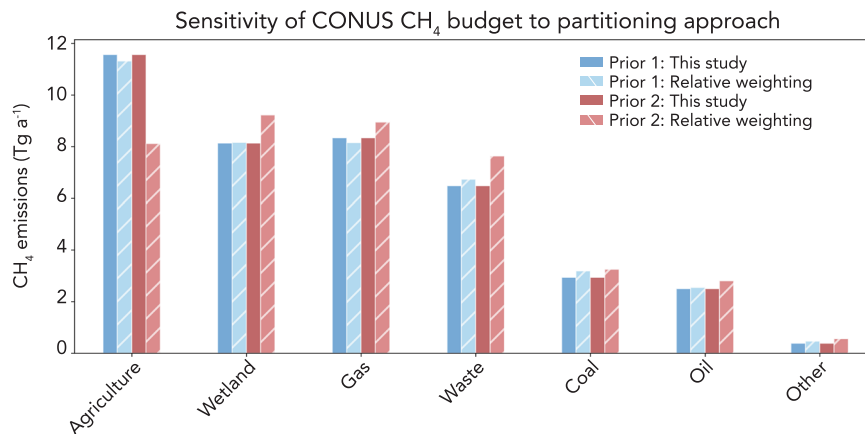


Fig. 2 Emissions partitioned or optimized from 2010–2015 GOSAT flux inversions. Two different flux products are partitioned using two different partitioning approaches. Blue bars represent fluxes that were derived using the same flux prior as Fig. 1 (Table S1). Red bars represent fluxes that were derived using EDGAR v5.0 for oil, gas, livestock, and waste sectors. Solid bars represent the optimal emission (OE) approach from Eqs. 1–2. Hatched bars represent partitioning done by computing the relative weights (RW) of emission sectors that made up the flux prior.

$\Delta\text{CH}_4 = 0.00/-0.06 \text{ TgCH}_4 \text{ a}^{-1}$), the small CH_4 flux enhancement observed in Fig. 1a is partitioned entirely to the oil sector, but produces emissions lower than the prior, which contrasts with the increasing production reported in the basin²⁴. This emission reduction may be due to an overestimate in the prior inventory or a difficulty in GOSAT sampling over that region, factors which could be verified with additional study. Posterior livestock emissions show increases from the prior that are distributed across the central and eastern United States, and a $0.07 \text{ TgCH}_4 \text{ a}^{-1}$ decrease in emissions over central California.

A major advantage of using Eqs. 1–2 to estimate emissions from independent inverse fluxes is that any discrepancies in flux priors can be accounted for when partitioning to a common emission prior. We show this through a sensitivity study whose results are summarized in Fig. 2. Here, we use two inverse flux products: (1) the 2010–2015 GOSAT CONUS flux shown in Fig. 1a, and (2) inverse fluxes from 2010–2015 GOSAT recomputed by using the EDGAR v5.0 emission inventory for oil, gas, waste, and livestock sectors for 2015²⁵. Differences in these prior inventories are summarized in Table S1 and Figs. S2–S3. Global inventories like EDGAR use consistent bottom-up aggregation methods across countries, which sometimes leads to discrepancies when compared with national inventories for particular sectors²¹. We employ two methods for optimizing/partitioning inverse fluxes: (1) using optimal estimation (OE) from this study's Eqs. 1–2 to optimize the same emissions prior from Fig. 1, and (2) by computing RWs of each emission sector in the flux prior and partitioning the inverse fluxes to emissions using these weights. In Fig. 2 we show that by employing OE on each separate inverse flux, we get the exact same answer for optimized emissions. This is a result of the $(\mathbf{I} - \hat{\mathbf{S}}\mathbf{S}_A^{-1})$ term from Eq. 1. Sometimes called the averaging kernel matrix²⁶, this term represents the spatial resolution of the flux estimate, or alternatively, the degree of smoothing of the flux prior to the estimate. Since the averaging kernel is applied to the prior swap term $(\mathbf{x}_A - \mathbf{M}\mathbf{z}_A)$, we account for discrepancies between flux and emission priors as a condition of emission optimization, so the choice of flux prior is immaterial. However, Fig. 2 shows the result when a relative weighting scheme is used for partitioning. The livestock to gas ratio in EDGAR is higher than in the EPA and Scarpelli et al. inventories. The result is that RW-partitioned livestock emissions are higher and gas emissions are lower when using EDGAR RWs ($13.2 \text{ TgCH}_4 \text{ a}^{-1}$ and $5.8 \text{ TgCH}_4 \text{ a}^{-1}$, respectively) than when using EPA–Scarpelli

RWs ($11.3 \text{ TgCH}_4 \text{ a}^{-1}$ and $8.1 \text{ TgCH}_4 \text{ a}^{-1}$, respectively). Similarly, wetland and waste emissions differ depending on which prior are used for RWs. Using prior ratios assumes total correlation between emission sectors in each grid cell as each sector's RW depends on emissions from other sectors. Therefore, we see that different assumptions on the flux prior can complicate comparison even between inverse fluxes that use the same atmospheric observations and transport model.

Comparing inverse fluxes in the Permian Basin. As seen from Figs. 1 and S1, the Permian Basin shows large increases in CH_4 emissions for both oil and gas sectors compared to the prior inventory. Production data from the Energy Information Administration²⁴ (EIA) indicates that between 2010 and 2019, the Permian increased oil production by 190% and gas production by 150%. We expect that fugitive emissions from these sectors would increase proportionately to the production increases, but the actual posterior estimates do not differentiate between intentional (planned) and unintentional (unplanned) emissions. The sensitivity study in Fig. 2 shows that the application of Eqs. 1–2 can be used to compare distinct inverse fluxes when a common emissions prior is used. We compare the 2010–2015 GOSAT flux product with a Permian $0.25^\circ \times 0.325^\circ$ flux product¹⁹ based on May 2018–March 2019 TROPOMI flux inversion. We partition fluxes to the same sectors as Fig. 1, but at a finer $0.1 \times 0.1^\circ$ grid resolution.

Figure 3a, b shows the 2010–2015 GOSAT flux inversion and the 2018–2019 TROPOMI flux inversion over the Permian Basin, respectively. Within the Permian domain (black line in Fig. 3), the GOSAT inverse flux estimate is $2.01 \pm 0.01 \text{ TgCH}_4 \text{ a}^{-1}$ and the TROPOMI inverse flux estimate is $2.68 \pm 0.5 \text{ TgCH}_4 \text{ a}^{-1}$, though these inversions were performed over different time periods and used different flux priors and prior error covariance matrices to constrain their solutions. The GOSAT and TROPOMI inverse products show distinct spatial patterns. The TROPOMI inversion shows two main regions of elevated CH_4 flux, which correspond to the Delaware Basin on the western side of the Permian (west Texas, southeast New Mexico) and the Midland Basin on the eastern side of the Permian. The GOSAT inversion shows a more distributed region of flux enhancement across the Permian. The difference in spatial distribution of these CH_4 flux maps could be due to the more limited GOSAT spatial observing coverage over the Permian and the coarser spatial resolution over which the 2010–2015 GOSAT flux inversion was

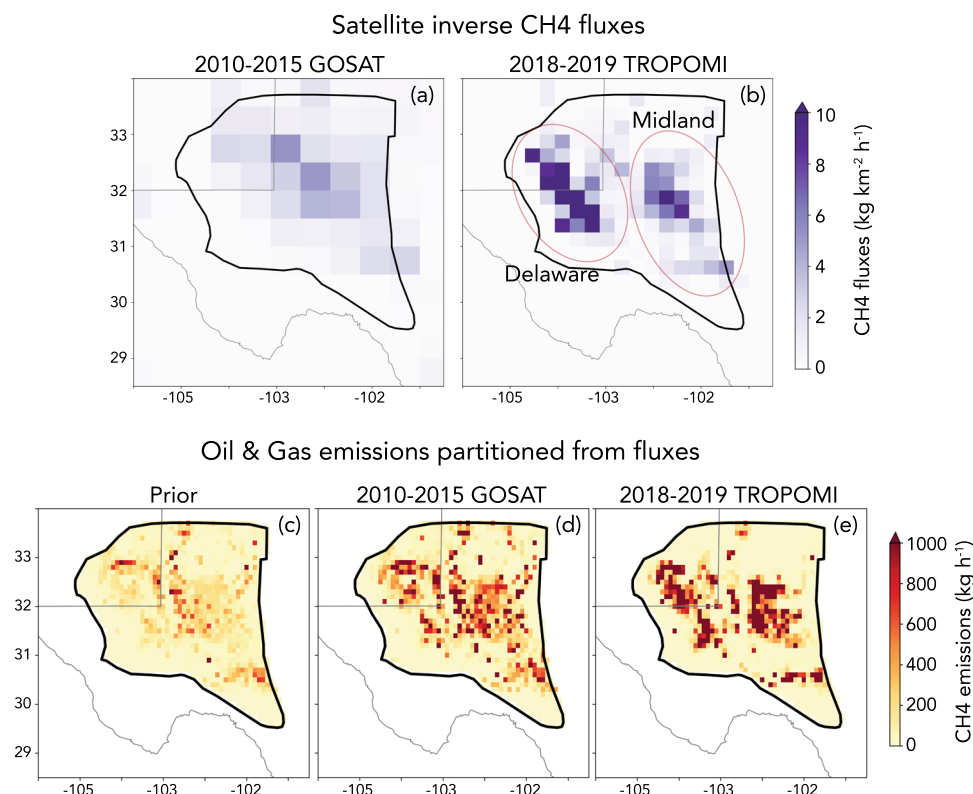
Partitioning applied to satellite CH₄ fluxes in the Permian Basin

Fig. 3 Comparison of two satellite inverse flux whose emissions were optimized using the same prior in the Permian Basin (black outline). Panel **a** shows the 2010–2015 GOSAT inverse flux result over the Permian Basin. Panel **b** shows the 2018–2019 TROPOMI inverse flux¹³, with the Delaware and Midland basins circled in red. Panel **c** is the prior oil and gas inventory used for partitioning¹⁵. Panels **d** and **e** represent posterior oil and gas emissions, optimized from fluxes in panels **a** and **b**, respectively.

performed. However, given that the two flux inversions are asynchronous and that oil and gas production increased dramatically between 2015 and 2018, spatial differences in CH₄ fluxes could also be the result of changing infrastructure throughout the basin.

Figure 3d, e show the optimized emissions for the oil, gas, and livestock sectors in the Permian, respectively, and compared to the prior inventory (Fig. 3c). Optimized emissions are derived from GOSAT and TROPOMI fluxes (Fig. 3a, b) that were swapped with a consistent prior (Fig. 3c) using Eqs. 1 and 2. The partitioned GOSAT emissions continue to show more distributed oil and gas CH₄ emissions across the basin when compared to the partitioned TROPOMI emissions, which are mostly concentrated to the Delaware and Midland Basins. Figure 4 shows the aggregated top-down oil and gas emissions in the Permian compared with reported trends from EIA production reports. The mean 2010–2015 oil and gas production in the Permian was 1.3 million bbl per day and 5.0 million Mcf per day, respectively. This increased to 3.8 million bbl per day and 12.7 million Mcf per day between May 2018 and March 2019 mean production, respectively²⁴. Though not linear with increased production, comparing partitioned 2010–2015 GOSAT oil and gas emissions and 2018–2019 TROPOMI emissions, we see a $0.52 \pm 0.29 \text{ TgCH}_4 \text{ a}^{-1}$ change in CH₄ emissions, representing a 29% increase. Optimized emissions summed across all sectors increased by $0.42 \pm 0.33 \text{ TgCH}_4 \text{ a}^{-1}$, with the increase driven mostly by gas, some contribution by oil, and offset by a small decrease in livestock emissions. Originally reported posterior flux estimates showed a $0.67 \pm 0.5 \text{ TgCH}_4 \text{ a}^{-1}$ difference between TROPOMI and GOSAT inversions, larger than the $0.42 \pm 0.33 \text{ TgCH}_4 \text{ a}^{-1}$ difference we quantify here after reprojected to a common prior. This difference

discrepancy between top-down estimates corresponds explicitly to differences in flux priors used in the original inversions that is now accounted for with this approach. Therefore, we can quantify how much the choice of flux prior directly impacts any quantified changes between top-down inventories.

Though consistent with the increase in gas production, the quantified $0.42 \text{ TgCH}_4 \text{ a}^{-1}$ increase in oil and gas emissions from 2010–2015 to 2018–2019 could also be due to observing constraints and/or biases in satellite retrievals. For example, Qu et al.²⁸ performed global $2 \times 2.5^\circ$ CH₄ inversions for 2019 using GOSAT and TROPOMI using the same prior and atmospheric chemistry-transport model, and derived Permian net fluxes of $2.36 \text{ TgCH}_4 \text{ a}^{-1}$ and $2.43 \text{ TgCH}_4 \text{ a}^{-1}$, respectively, showing consistency between signals observed by these independent remote sensing platforms for this region. Therefore, we conclude that the trends observed in Fig. 3 are likely due to changes in gas operations, and not bias in observing systems. However, in other global regions where the surface is less bright and homogeneous than the Permian, flux results derived from TROPOMI and GOSAT inversion may not agree²⁸.

Discussion

Having robust intercomparison methods in place are needed for interpreting the ever increasing number of atmospheric observations of CH₄, particularly with regard to the expected launches of several CH₄-observing satellites in the 2020s²⁹. As described in this study, inverse fluxes derived from these satellite observations will depend on the observations themselves, the chemical transport model, and the prior constraint. Having the ability to directly quantify how these terms influence emissions is needed for

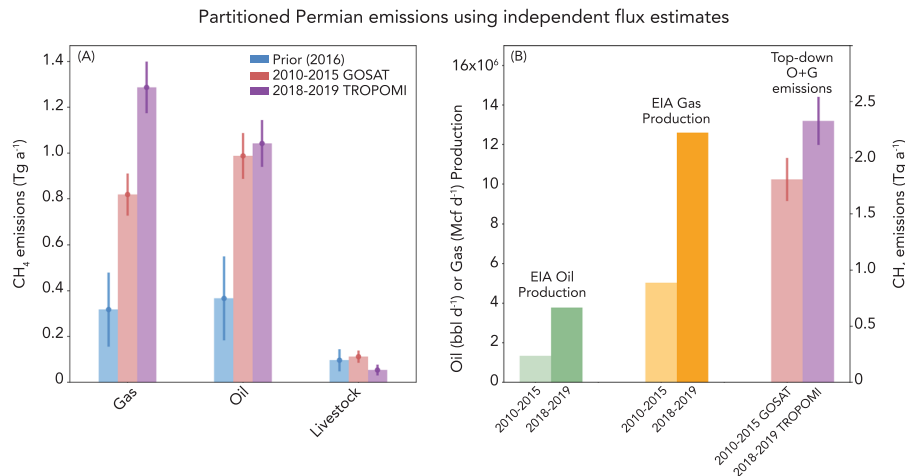


Fig. 4 Aggregated CH₄ emissions and reported production from the Energy Information Administration (EIA) in the Permian Basin. Panel **a** shows top-down emissions optimized from the independent flux products from Fig. 3 aggregated to the basin. Error bars are an integration of prior and posterior error covariance matrices for the respective emission sectors. Panel **b** compares mean 2010–2015 and May 2018–March 2019 EIA production data²¹ for oil and gas sectors to the oil+gas emissions from panel **a**.

diagnosing uncertainty in the estimated methane budget. Ultimately, this information can be combined to provide better global understanding of methane emissions and finer and more policy-relevant spatial and temporal scales. Likewise, the Bayesian framework we describe in this study can be applied to other atmospheric gas fluxes like carbon dioxide (CO₂). Partitioning between biospheric and anthropogenic CO₂ emissions remains highly uncertain³⁰, so incorporating this framework that directly optimizes emission sectors could be useful for reconciling the budget. This approach does require a move from traditional ensemble or adjoint-based inversions, created to reduce cost of this computationally expensive problem, to an analytic or optimal estimation inversion as an explicit representation of the posterior covariance is required and this covariance is not easily calculated from ensemble or 4D-var methods.

While our estimates account for the spatial resolution and error associated with the inversion of observations to fluxes, we do not explicitly account for error in model transport and chemistry. These errors could be important when comparing emissions between seasons or in regions where transport is poorly modeled such as in the tropics. For example, a study of global carbon monoxide emissions, a trace gas that (like CH₄) is affected by reaction with the hydroxyl radical (OH) and transport, found that convective mass flux in the tropics is likely responsible for errors in emissions³¹. While our approach can account for this error if a corresponding posterior covariance is provided, we emphasize that studies that both characterize and mitigate this part of the flux error budget are needed to better use satellite observations of methane and of other trace gases. Ultimately, improved emission and error characterization from top-down information will allow for better updates and comparisons with bottom-up inventories, which can guide progress towards CH₄ mitigation.

Methods

In this section we derive a method to estimate CH₄ emissions from atmospheric observations. The SI contains an alternate derivation (Section S1) and conceptual examples to further clarify the mechanics of prior-swapping. Emissions can be represented as a vector *i.e.*, ($\mathbf{z} \in \mathbb{R}^m$) that contains both sectoral and spatial information. Atmospheric observations *i.e.*, ($\mathbf{y} \in \mathbb{R}^p$) often represent concentrations of CH₄ observed by surface, satellite, airborne, or some other observing system. Generally, atmospheric inversions do not directly optimize CH₄ sectoral emissions from observations, and instead optimize CH₄ fluxes *i.e.*, ($\mathbf{x} \in \mathbb{R}^n$), which represent the summation of CH₄ emissions within a grid cell. In the flux inversion setup, atmospheric CH₄ observations *y* are used to estimate CH₄ fluxes *x*. We estimate this optimal state by finding the mode of the posterior flux distribution

$p(\mathbf{x}|\mathbf{y})$, or $\hat{\mathbf{x}}$. A transformation or Jacobian matrix *K* can be derived from atmospheric transport simulations (e.g., GEOS-Chem), such that we can represent the relationship between fluxes and observations:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{n} \quad (3)$$

Where *n* represents noise. We apply Bayes Theorem to estimate the posterior distribution $p(\mathbf{x}|\mathbf{y})$:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \quad (4)$$

Where $p(\mathbf{y}|\mathbf{x})$ is the maximum likelihood given by Eq. 3, and $p(\mathbf{x})$ is the prior distribution. If we assume that $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$ are Gaussian distributions, and *y* and *x_A* represent the modes of those respective distributions, then the mode of the posterior distribution $\hat{\mathbf{x}}$ has a closed form solution, known as the Maximum A Posteriori (MAP; Rodgers, 2000) solution:

$$\hat{\mathbf{x}} = \mathbf{x}_A + \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_y^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_A) \quad (5)$$

$$\hat{\mathbf{S}} = (\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_A^{-1})^{-1} \quad (6)$$

For policy-relevance and CH₄ budget quantification, we really wish to optimize emissions using atmospheric observations, *i.e.*, we want to compute the explicit posterior representation $p(\mathbf{z}|\mathbf{y})$ without re-simulation of an atmospheric transport model. The relationship between *z* and *x* is simple aggregation, and can be represented by matrix *M*:

$$\mathbf{x} = \mathbf{M}\mathbf{z}. \quad (7)$$

If *z* and *x* share the same grid resolution (*i.e.*, if $\mathbf{z} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$ and *s* is the number of emission sectors, then $m = ns$), the matrix *M* $\in \mathbb{R}^{n \times m}$ is represented with the following terms:

$$m_{ij} = \begin{cases} 1 & (i,j) \text{ pertain to same grid} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

If *z* and *x* pertain to different grids, the relationship *M* is defined by the geographic area (*Ω*) and intersections (*∩*) of grid cells:

$$m_{ij} = \frac{\Omega_{x_i \cap z_j}}{\Omega_{z_j}}. \quad (9)$$

Since *M* is simply a summation matrix, we assume there is no noise associated with its application. Using *M*, we can update Eqs. 5 and 6 to find the optimal emission state vector $\hat{\mathbf{z}}$ and its posterior error covariance $\hat{\mathbf{Z}}$:

$$\hat{\mathbf{z}} = \mathbf{z}_A + \hat{\mathbf{Z}}(\mathbf{KM})^T\mathbf{S}_y^{-1}(\mathbf{y} - (\mathbf{KM})\mathbf{z}_A) \quad (10)$$

$$\hat{\mathbf{Z}} = ((\mathbf{KM})^T\mathbf{S}_y^{-1}(\mathbf{KM}) + \mathbf{Z}_A^{-1})^{-1} \quad (11)$$

Equations 10 and 11 provide an explicit closed form solution for $\hat{\mathbf{z}}$, which is sufficient for emission optimization without reliance on RWs. Therefore, application of Eqs. 10 and 11 into existing inverse frameworks would provide posterior emission estimates constrained by atmospheric observations. However, these equations require the computation of the matrix (*KM*), which can be large in cases where many atmospheric observations exist. An exactly equivalent solution is possible with just the products of the flux inversion, specifically $\hat{\mathbf{S}}$, $\hat{\mathbf{x}}$, *S_A*, and *x_A*. We show one derivation below and provide an alternative derivation in the SI:

Equation 5 can be shown to have an equivalent form that is often used in atmospheric trace gas retrievals²⁶:

$$\hat{\mathbf{x}} = \mathbf{x}_A + \mathbf{A}(\mathbf{x} - \mathbf{x}_A) + \mathbf{G}\mathbf{n} \quad (12)$$

Where \mathbf{A} is the averaging kernel matrix $\mathbf{A} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}$.

$$\mathbf{A} = \mathbf{I} - \hat{\mathbf{S}}\mathbf{S}_A^{-1} \quad (13)$$

And \mathbf{G} is the Gain Matrix $\mathbf{G} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}}$.

$$\mathbf{G} = \mathbf{S}_A \mathbf{K}^T (\mathbf{K} \mathbf{S}_A \mathbf{K}^T + \mathbf{S}_y)^{-1} \quad (14)$$

And \mathbf{x} are the true atmospheric fluxes. Therefore, Eq. 12 shows that the optimal solution $\hat{\mathbf{x}}$ is a combination of the truth, smoothed by some prior and includes noise. From Eq. 12, we can create a flux to posterior flux operation \mathbf{H} , given that our relationship \mathbf{M} is known:

$$\mathbf{H}(\mathbf{x}) = \mathbf{x}_A + \mathbf{A}(\mathbf{x} - \mathbf{x}_A) \quad (15)$$

$$\mathbf{H}(\mathbf{Mz}) = \mathbf{x}_A + \mathbf{A}(\mathbf{Mz} - \mathbf{x}_A) \quad (16)$$

The operation \mathbf{H} allows for emissions to be smoothed with an averaging kernel, allowing for direct comparison with $\hat{\mathbf{x}}$. With this operator relationship, we treat $\hat{\mathbf{x}}$ as an observable. The error covariance $\hat{\mathbf{S}}$ includes both smoothing (\mathbf{S}_s) and measurement error (\mathbf{S}_m):

$$\hat{\mathbf{S}} = \mathbf{S}_s + \mathbf{S}_m \quad (17)$$

For flux partitioning, we want to isolate the \mathbf{S}_m error component $\hat{\mathbf{x}}$, as the \mathbf{H} operator already accounts for smoothing via the averaging kernel. The matrix \mathbf{S}_m can be represented³² using \mathbf{G} and \mathbf{S}_y :

$$\mathbf{S}_m = \mathbf{G} \mathbf{S}_y \mathbf{G}^T. \quad (18)$$

While \mathbf{S}_s has the following representation:

$$\mathbf{S}_s = (\mathbf{I} - \mathbf{A}) \mathbf{S}_A (\mathbf{I} - \mathbf{A})^T \quad (19)$$

We can combine Eqs. 17 and 19 to get an alternate form of \mathbf{S}_m that does not require \mathbf{S}_y and \mathbf{G} explicitly:

$$\mathbf{S}_m = \hat{\mathbf{S}} - (\mathbf{I} - \mathbf{A}) \mathbf{S}_A (\mathbf{I} - \mathbf{A})^T \quad (20)$$

Using \mathbf{S}_m for observational error covariance, we can apply the MAP solution to derive $\hat{\mathbf{z}}$ and $\hat{\mathbf{Z}}$:

$$\hat{\mathbf{z}} = \mathbf{z}_A + \hat{\mathbf{Z}} \mathbf{M}^T \mathbf{A}^T \mathbf{S}_m^{-1} (\hat{\mathbf{x}} - \mathbf{H}(\mathbf{Mz}_A)) \quad (21)$$

$$\hat{\mathbf{z}} = \mathbf{z}_A + \hat{\mathbf{Z}} \mathbf{M}^T \mathbf{A}^T \mathbf{S}_m^{-1} (\hat{\mathbf{x}} - \mathbf{x}_A - \mathbf{A}(\mathbf{Mz}_A - \mathbf{x}_A)) \quad (22)$$

$$\hat{\mathbf{Z}} = (\mathbf{M}^T \mathbf{A}^T \mathbf{S}_m^{-1} \mathbf{A} \mathbf{M} + \mathbf{Z}_A^{-1})^{-1}. \quad (23)$$

Now we have a direct solution for $\hat{\mathbf{z}}$ and $\hat{\mathbf{Z}}$ derived only from the products of a flux inversion (specifically, \mathbf{x}_A , \mathbf{S}_A , $\hat{\mathbf{x}}$, $\hat{\mathbf{S}}$), an emissions prior (\mathbf{Z}_A), and an aggregation matrix \mathbf{M} . Equations 22 and 23 can be shown to be of the same form as Eqs. 1 and 2 by showing that $\mathbf{A}^T \mathbf{S}_m^{-1} = \hat{\mathbf{S}}^{-1}$. Decomposing \mathbf{A}^T and recognizing that $\hat{\mathbf{S}}$ and \mathbf{S}_A are symmetric matrices, we have

$$\mathbf{A}^T = (\mathbf{I} - \hat{\mathbf{S}} \mathbf{S}_A^{-1})^T = \mathbf{I} - (\mathbf{S}_A^{-1})^T \hat{\mathbf{S}}^T = \mathbf{I} - \mathbf{S}_A^{-1} \hat{\mathbf{S}}. \quad (24)$$

We first expand Eq. 19 for an alternative expression of \mathbf{S}_m :

$$\mathbf{S}_m = \hat{\mathbf{S}} - (\mathbf{I} - \mathbf{A}) \mathbf{S}_A (\mathbf{I} - \mathbf{A})^T \quad (25)$$

$$= \hat{\mathbf{S}} - (\mathbf{I} - \mathbf{I} + \hat{\mathbf{S}} \mathbf{S}_A^{-1}) \mathbf{S}_A (\mathbf{I} - \mathbf{I} + \hat{\mathbf{S}} \mathbf{S}_A^{-1})^T \quad (26)$$

$$= \hat{\mathbf{S}} \mathbf{A}^T. \quad (27)$$

Taking the inverse of \mathbf{S}_m using the Eq. 27, we have:

$$\mathbf{S}_m^{-1} = (\hat{\mathbf{S}} \mathbf{A}^T)^{-1} \quad (28)$$

$$\mathbf{S}_m^{-1} = (\mathbf{A}^T)^{-1} \hat{\mathbf{S}}^{-1}. \quad (29)$$

The algebra for 28 and 29 are only possible if $(\mathbf{A}^T)^{-1}$ exists. For overdetermined systems (i.e., dimension of \mathbf{y} » dimension of \mathbf{x}), this is generally valid. Multiplying both sides of Eq. 29 by \mathbf{A}^T finishes the proof:

$$\mathbf{A}^T \mathbf{S}_m^{-1} = \hat{\mathbf{S}}^{-1}. \quad (30)$$

In a similar fashion, we can show that Eqs. 2 and 23 are equivalent if $\mathbf{A}^T \mathbf{S}_m^{-1} \mathbf{A} = \hat{\mathbf{S}}^{-1} - \mathbf{S}_A^{-1}$. To do this, we multiply Eq. 30 by \mathbf{A} :

$$\mathbf{A}^T \mathbf{S}_m^{-1} \mathbf{A} = \hat{\mathbf{S}}^{-1} \mathbf{A} \quad (31)$$

$$= \hat{\mathbf{S}}^{-1} (\mathbf{I} - \hat{\mathbf{S}} \mathbf{S}_A^{-1}) \quad (32)$$

$$= \hat{\mathbf{S}}^{-1} - \mathbf{S}_A^{-1}. \quad (33)$$

Data availability

Fossil fuel prior emission inventories are available for download at <https://doi.org/10.7910/DVN/HH4EUM>. Wetland emission prior inventories are available at (<https://doi.org/10.3334/ORNLDAC/1502>). The EPA gridded methane inventory is available for download at <https://www.epa.gov/ghgemissions/gridded-2012-methane-emissions>.

Code availability

Code used for this analysis can be found at https://github.com/dcusworth/partition_fluxes_to_emissions.

Received: 19 May 2021; Accepted: 27 October 2021;

Published online: 25 November 2021

References

- Myhre, G. et al. Anthropogenic and Natural Radiative Forcing Supplementary Material. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Available from www.climatechange2013.org and www.ipcc.ch. (2013).
- United Nations Environment Programme and Climate and Clean Air Coalition (2021). *Global Methane Assessment: Benefits and Costs of Mitigating Methane Emissions*. Nairobi: United Nations Environment Programme (2021).
- Kirschke, S. et al. Three decades of global methane sources and sinks. *Nature Geoscience* **6**, 813–823 (2013).
- UNFCCC. Adoption of the Paris Agreement. Report No. FCCC/CP/2015/L.9/Rev.1, <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf>, 2015
- Saunio, M. et al. The global methane budget 2000–2017. *Earth Syst. Sci. Data* **12**, 1561–1623 (2020).
- Melton, J. R. et al. Present state of global wetland extent and wetland methane modelling: Conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences* **10**, 753–788 (2013).
- Poulter, B. et al. Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics. *Environ. Res. Lett.* <https://iopscience.iop.org/article/10.1088/1748-9326/aa8391> (2013).
- Zhang, Y. et al. Attribution of the accelerating increase in atmospheric methane during 2010–2018 by inverse analysis of GOSAT observations. *Atmos. Chem. Phys.* **21**, 3643–3666 (2021).
- Maasakkers, J. D. et al. 2010–2015 North American methane emissions, sectoral contributions, and trends: a high-resolution inversion of GOSAT satellite observations of atmospheric methane. *Atmos. Chem. Phys.* **21**, 4339–4356 (2021).
- Bergamaschi, P. et al. Inverse modelling of European CH₄ emissions during 2006–2012 using different inverse models and reassessed atmospheric observations. *Atmos. Chem. Phys.* **18**, 901–920 (2018).
- Alexe, M. et al. Inverse modelling of CH₄ emissions for 2010–2011 using different satellite retrieval products from GOSAT and SCIAMACHY. *Atmos. Chem. Phys.* **15**, 113–133 (2015).
- Yadav, V. et al. Spatio-temporally resolved methane fluxes from the Los Angeles Megacity. *J. Geophys. Res. Atmos.* **124**, 5131–5148 (2019).
- Miller, S. M. et al. Anthropogenic emissions of methane in the United States. *Proc. Natl Acad. Sci. USA* **110**, 20018–20022 (2013).
- Ganesan, A. L. et al. Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network. *Atmos. Chem. Phys.* **15**, 6393–6406 (2015).
- Rodgers, C. D. & Connor, B. J. Intercomparison of remote sounding instruments. *J. Geophys. Res. Atmos.* **108**, 4116 <https://doi.org/10.1029/2002JD002299>, D3 (2003).
- Shen, L. et al. Unravelling a large methane emission discrepancy in Mexico using satellite observations. *Remote Sens. Environ.* **260**, 112461 (2021).
- Meirink, J. F., Bergamaschi, P. & Krol, M. C. Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion. *Atmos. Chem. Phys.* **8**, 6341–6353 (2008).
- Bousserez, N. & Henze, D. K. Optimal and scalable methods to approximate the solutions of large-scale Bayesian problems: theory and application to atmospheric inversion and data assimilation. *Q. J. R. Meteorol. Soc.* **144**, 365–390 (2018).
- Zhang, Y. et al. Quantifying methane emissions from the largest oil-producing basin in the United States from space. *Sci. Adv.* **6**, eaaz5120 (2020).
- Kuze, A. et al. Update on GOSAT TANSO-FTS performance, operations, and data products after more than 6 years in space. *Atmos. Meas. Tech.* **9**, 2445–2461 (2016).
- Scarpelli, T. R. et al. A global gridded (0.1° × 0.1°) inventory of methane emissions from oil, gas, and coal exploitation based on national reports to the

- United Nations Framework Convention on Climate Change. *Earth System Science. Data* **12**, 563–575 (2020).
22. Ma, S. et al. Satellite constraints on the latitudinal distribution and temperature sensitivity of wetland methane emissions. *AGU Adv.* **2**, e2021AV000408 (2021).
 23. Maasakkers, J. D. et al. Gridded national inventory of US methane emissions. *Environ. Sci. Technol.* **50**, 13123–13133 (2016).
 24. EIA. Drilling Productivity Report, URL <https://www.eia.gov/petroleum/drilling/>, Last Accessed 2 Feb 2021 (2020).
 25. Crippa, M. et al. Fossil CO₂ and GHG emissions of all world countries - 2019 Report, EUR 29849 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-11100-9, 10.2760/687800, JRC117610. (2019).
 26. Rodgers, C. D. Inverse methods for atmospheric sounding: theory and practice (Vol. 2). *World Scientific* (World Scientific Publishing Co. Pte. Ltd, 2000).
 27. Veefkind, J. P. et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sens. Environ.* **120**, 70–83 (2012).
 28. Qu, Z. et al. Global distribution of methane emissions: a comparative inverse analysis of observations from the TROPOMI and GOSAT satellite instruments. *Atmos. Chem. Phys. Discussions* **21**, 14159–14175 (2021).
 29. Crisp, D. et al. A constellation architecture for monitoring carbon dioxide and methane from space. Prepared by the CEOS Atmospheric Constellation Greenhouse Gas Team, Version, 1(8), https://ceos.org/document_management/Virtual_Constellations/ACC/Documents/CEOS_AC-VC_GHG_White_Paper_Version_1_20181009.pdf (2018).
 30. Friedlingstein, P. et al. Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
 31. Jiang, Z. et al. Impact of model errors in convective transport on CO source estimates inferred from MOPITT CO retrievals. *J. Geophys. Res. Atmos.* **118**, 2073–2083 (2013).
 32. Bowman, K. W. et al. Tropospheric emission spectrometer: retrieval method and error analysis. *IEEE Trans. Geosci. Remote Sens.* **44**, 1297–1307 (2006).

Acknowledgements

Portions of this work research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). Some of the work was supported by NASA's Carbon Monitoring System program. Yuzhong Zhang acknowledges funding by NSFC (42007198). Government sponsorship acknowledged.

Author contributions

D.H.C. and J.R.W. designed the study. D.H.C. performed the analysis and wrote the manuscript. D.H.C., Y.Y., J.R.W., A.A.B., and K.B. developed/derived the algebraic equations. S.M., J.D.M., C.E.M., T.R.S., Z.Q., D.J.J., and Y.Z. provided and guided interpretation and integration of prior and posterior inventories. All authors provided feedback and edits on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-021-00312-6>.

Correspondence and requests for materials should be addressed to Daniel H. Cusworth.

Peer review information *Communications Earth & Environment* thanks Luke Western, Hossein Maazallahi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Joshua Dean and Clare Davis. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021