**High-resolution North American methane emissions inferred from an inversion of 2019 TROPOMI satellite data**

Hannah Nesser1, Daniel J. Jacob1, Xiao Lu1, Joannes D. Maasakkers2, Melissa P. Sulprizio1, Tia R. Scarpelli1, Lu Shen1, Zhen Qu1, Margaux Winter1, A. Anthony Bloom3, Shuang Ma3, …

1Harvard University, Cambridge, Massachusetts, United States

2SRON Netherlands Institute for Space Research, Utrecht, The Netherlands

3Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, United States

**Abstract.** …

**1 Introduction**

All modelled pathways that limit global warming to 1.5°C require deep reductions in methane emissions (IPCC). However, global methane concentrations have tripled from pre-industrial concentrations and are increasing at an increasing rate (Dlugokencky). While the sources of methane emissions, including livestock, the fossil fuel industry, waste management, and wetlands, are well understood, significant uncertainties exist in their spatial and temporal distribution. Countries report their total anthropogenic methane emissions to the United Nations Framework Convention on Climate Change (UNFCCC) following procedures defined by the International Panel on Climate Change (IPCC).

Papers to cite

NA inversions

* Miller et al. 2013 – Lagrangian inversion of observations from towers and aircraft, analyzes footprints for each of 12,694 observations, geostatistical inversion (no prior), 1deg x 1deg
* Wecht et al. 2014 – SCIAMACHY, adjoint
* Turner et al. 2015 – GOSAT, GMM, EDGAR
* Janardanan et al. 2017?
* Bruhwiler et al. 2017?
* Sheng et al. 2018?
* Lan et al. 2019?
* Maasakkers et al. 2021 – GOSAT, GMM, EPA GHGI

High resolution regional inversions

* Wecht et al. 2014 – analytical inversion over western North America and Pacific? (157 grid cells at 0.5 degrees) (Spatially resolving methane emissions in California, ACP)
* Sheng et al. 2018 – Regional SEAC4RS using GMM
* Zhang et al. 2020 – Permian inversion using analytical inversion

Inversion at 25 km resolution enabled by reduced-rank approach

Paragraph on the reduced-rank approach

**2** **Data and methods**

We conduct an inversion of 2019 TROPOMI methane observations over North America. The TROPOMI observations are fit to simulated concentrations from the GEOS-Chem chemical transport model (CTM, [www.geos-chem.org](http://www.geos-chem.org)) to optimize mean methane emissions at 0.25° x 0.3125° spatial resolution. We calculate the optimal emissions and the associated error covariance and information content by finding the analytical minimum of a Bayesian cost function regularized by a prior emissions estimate (section 2.1). Sections 2.2 through 2.4 describe the components of the inversion: section 2.2 describes the state vector, prior emissions, and prior errors; section 2.3 describes the TROPOMI observations and the observing system errors; and section 2.4 describes GEOS-Chem and the novel, reduced-rank method used to calculate the Jacobian matrix. We also conduct a suite of sensitivity tests to provide additional constraints on the error of the optimized emissions, which are summarized in section 2.5.

**2.1 Analytical inversion procedure**

We optimize the state vector of gridded emissions assuming normal errors by minimizing a Bayesian cost function

where and are the prior emissions estimate and error covariance, respectively (section 2.4); and are the observational vector and error covariance, respectively (section 2.5); is the chemical transport model (CTM) that simulates observations as a function of emissions (section 2.2); and is a regularization factor that accounts for errors in the estimated observational error covariance matrix (Brasseur and Jacob, 2017). The nested methane CTM is linear so that where is the Jacobian matrix and is constant, allowing analytical minimization of the cost function yielding the optimal (posterior) state vector estimate and associated error covariance matrix :

The relative reduction in the prior error covariance defines the information content of the observing system, quantified by the averaging kernel matrix

The averaging kernel matrix describes the sensitivity of the posterior estimate to the true state vector, . The diagonal elements are therefore commonly referred to as averaging kernel sensitivities and their sum gives the degrees of freedom for signal (DOFS), the number of pieces of information independently constrained by the observing system (Rodgers, 2000).

**2.2 State vector, prior estimate, and prior error**

We optimize emissions in 23,691 grid cells at 0.25° x 0.3125° resolution over North America, including all grid cells containing land or anthropogenic methane emissions larger than 0.1 Mg km-2 a-1, representing over 99% of methane emissions in North America. Methane chemical and soil sinks are not optimized because these loss processes are slow compared to the ventilation timescale.

Table 1 summarizes the prior emissions estimates and Figure xx shows the spatial distribution of major source sectors. Anthropogenic emissions for the United States, Mexico, and Canada are given by the spatially disaggregated (gridded) versions of the EPA GHGI for 2012 (Maasakkers et al., 2016), the INECC inventory for 2015 (Scarpelli et al., 2020), and the ECCC estimates for 2018 (Scarpelli et al., 2021), respectively. To account for changes in the distribution and magnitude of oil and natural gas emissions in the United States since 2012, we update the distribution of production fields using 2018 DrillingInfo data and scale the total natural gas transmission, processing, and distribution emissions to match 2018 emissions as reported in the 2020 GHGI. All other anthropogenic emissions in the North American domain are provided by the EDGAR v4.3.2 (?) global emission inventory for 2012 (?). Anthropogenic emissions are assumed aseasonal except for manure management and rice cultivation, for which we apply seasonal scaling factors as described by Maasakkers et al. (2016) and Zhang et al. (2016), respectively.

Wetlands are the dominant natural source of methane emissions. We use the high-performance WETCharts ensemble version 1.3.1, which selects 10 process-based models on the basis of their correspondence with GOSAT satellite observations (Ma et al. 2021). Other natural methane emission sources include open fires, termites, and geological seeps. Open fire emissions are from the Global Fire Emissions Database version 4 (GFED4, van der Werf et al., 2017), termite emissions from Fung et al. (1991), and geological seepage from Etiope et al. (2019) scaled to the 2 Tg a-1 global emission magnitude given by Hmiel et al. (2020).

We assume uniform relative errors of 50% in the absence of better information. We test this assumption in a series of sensitivity inversions in which we increase, decrease, and vary the errors over the domain.

**2.3 Forward model**

We use the nested version of the GEOS-Chem chemical transport model (CTM) v12.7 at 0.25° x 0.3125° resolution over North America as the forward model for the inversion. Earlier versions of the methane simulation were described by Wecht et al. (2014) and Turner et al. (2015). The model is driven by \_\_\_ meteorological fields from the \_\_\_. Methane loss from OH, Cl, soil uptake, and stratospheric oxidation is described in Maasakkers et al. (2019). Initial conditions for January 2019 and 3-hourly boundary conditions for the year are given by the methane concentration fields from the global 2° ⨉ 2.5° TROPOMI inversion conducted by Qu et al. (2021). These concentration fields are unbiased with respect to the global TROPOMI data and are informed predominantly by observations outside of North America.

We validate GEOS-Chem by comparison to surface and aircraft methane observations for May 2018. We use observations from the Atmospheric Tomography Mission (ATom), the Atmospheric Carbon and Transport – America (ACT-America) campaign, and the NOAA Observation Package (ObsPack). We find a mean model-observation bias of 6.36 and a correlation of R = 0.45. We also find no significant latitudinal bias in the model-observation difference, although the observations used provide significant coverage only between 30°N and 50°N. We expect no systematic bias with respect to albedo because GEOS-Chem does not simulate radiative transfer.

**2.4 TROPOMI observations**

The Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5 Precursor satellite has provided daily, global observations of dry column methane mixing ratios at 7 ⨉ 7 km2 nadir pixel resolution since May 2018 and at 5.5 ⨉ 7 km2 nadir pixel resolution since August 2019 (citation). TROPOMI is in sun-synchronous orbit with a local overpass time of 13:30 (Veefkind et al. 2012). TROPOMI retrieves methane concentrations at the 2.3 μm absorption feature using a full-physics retrieval, which is limited by cloud cover, variable topography, albedo, and high aerosol loading (citation). As a result, TROPOMI has a xx% retrieval rate over North America for 2019. We use the retrieval described by Lorente et al. (2021), which has a -3.4 ± 5.6 ppb bias relative to the Total Carbon Column Observing Network (TCCON). We use only high-quality retrievals as indicated by the quality assessment flag (qa = 1).

We find that systematic albedo biases in the GEOS-Chem – TROPOMI difference are reduced by filtering out scenes snow- or ice-cover or with low albedo values, which are more likely have large errors (Lorente et al. 2021). We identify snow- and ice-covered scenes using the blended albedo, an empirical parameter that is a function of the shortwave and near-infrared albedo and that correlates with snow- and ice-cover when greater than about 1 (Wunch et al. 2011). We remove scenes with blended albedo > 0.75 in months that show the characteristic cluster of points …. Finally, we remove scenes with shortwave albedo less than 0.05, which disproportionately account for the remaining unphysical TROPOMI observations (XCH4 < 1700) and which exhibit larger GEOS-Chem – TROPOMI biases (de Gouw et al., 2020). The resulting data preserves 70% of the original high-quality data.

Figure xx (solid lines) shows the GEOS-Chem – TROPOMI difference with respect to albedo, season, and latitude for the filtered data. We find no bias with respect to albedo or season and an aseasonal latitudinal bias. This bias has been noted and corrected previously by Turner et al. (xxxx), Maasakkers et al. (xxxx), and Zhang et al. (xxxx). We define a latitudinal correction term (ppb) using the first-order polynomial

where is the degrees latitude.

Figure xx shows the correlation between the prior GEOS-Chem simulation and corrected TROPOMI observations. We find good agreement between the model and observations (R = 0.77). The top row of figure xx shows the remaining 2946216 observations, regridded onto the GEOS-Chem grid and averaged seasonally, that constitute our observation vector **y**. The observations exhibit seasonally heterogenous coverage that is the basis for the reduced-rank approach we take to construct the Jacobian matrix for the inversion (Section 2.3.4).

**2.5 Observing system errors**

The bottom row of figure xx shows the error standard deviations of the observing system, including contributions from the forward model, the instrument, and representation error, regridded onto the GEOS-Chem grid and averaged seasonally. We calculate the variances using the residual error method (Heald et al. 2004). This method assumes that the mean difference between the TROPOMI observations and the prior GEOS-Chem simulation in each grid cell and month is caused by errors in emissions that will be corrected by the inversion. The residual standard deviation after subtracting the mean monthly, gridded error gives an estimate of the observational error standard deviation. In the 1% of scenes where the residual standard deviation is less than the reported instrument error standard deviation (Lorente et al., 2021), we use the latter instead. We find a mean observational error standard deviation of 9.5 ppb that does not vary significantly with albedo, season, or latitude. We take the corresponding variances as the diagonal elements of our observational error covariance matrix . Off-diagonal terms are assumed zero in the absence of better information.

*[I realize now that given that we are optimizing emissions annually, perhaps the mean errors should be calculated for each grid box, but not for each month.]*

We introduce a regularization factor (section 2.1) to account for the lack of covariance structure in . *[Isn’t there a source that says that scaling up the diagonal produces equivalent results to including off-diagonal elements?]* ….

Houweling et al. (2017):

“Errors that behave quasi-random and affect neighboring retrievals in a coherent way can in theory be accounted for by specifying the off-diagonal terms in the data error covariance matrix. In practice, there are many ways to do this, but quantitative information to justify a specific choice is lacking. In general, correlated uncertainty reduces the number of independent measurements, which justifies averaging retrievals within a certain distance of each other. Usually the uncertainty of the mean is calculated using a lower bound representing the contribution of purely systematic error. An alternative approach, referred to as ”error inflation”, is to increase the error of individually assimilated retrievals such that the uncertainty of a mean of surrounding retrievals does not drop below this minimum level (Chevallier, 2007). The advantage of this approach is that it avoids subjective decisions about which samples to combine into an average. Error inflation, or similar methods that compensate the neglect of off diagonals in the data error covariance matrix by increasing the (diagonal) uncertainty, lead to a χ 2 below 1. Although this may seem suboptimal from a statistical point of view, Chevallier (2007) demonstrated that this de-weighing of data nevertheless leads to uncertainty reductions that are closer to those obtained when off diagonals in R had been accounted for. Therefore, this approach avoids over constraining the problem by neglecting the contribution of data error covariance.”

**2.5 Jacobian matrix**

The relationship between simulated methane concentrations and emissions in the nested version of GEOS-Chem is strictly linear and is described by the Jacobian matrix . The Jacobian matrix is typically constructed by conducting a forward model simulation for each state vector element. While this is an embarrassingly parallel problem, constructing this matrix for the 23691 0.25° x 0.3125° resolution grid cells optimized by this inversion is computationally intractable. We take advantage of the heterogeneous information content of the TROPOMI observations to construct the Jacobian matrix at substantially decreased computational cost using the reduced-rank method introduced by Nesser et al. (2021). This method updates an initial, low-cost estimate of the Jacobian matrix by perturbing the patterns that best explain the information content of the observing system rather than grid cells, constructing a reduced-rank Jacobian matrix while optimally preserving information content.

We construct the initial estimate of the Jacobian matrix using the mass-balance approach introduced by Nesser et al. (2021). We assume that a perturbation of methane emissions in grid cell *j* produces column mixing ratio enhancements in nearby observation *i* according to

where is a dimensionless, mass-conserving coefficient providing a crude representation of turbulent diffusion that decreases the sparsity of , and are the molecular weights of dry air and methane, respectively, is a ventilation length scale equal to the square root of the grid cell area, is gravitational acceleration, is the local wind speed taken here as 5 km h-1, and is the surface pressure taken here as 1000 hPa. We assume decreases exponentially as from the inner to the outer ring of grid cells surrounding *j*, normalized and divided by the number of grid cells in each ring.

We use to calculate the patterns of information content perturbed in the forward model. The corresponding averaging kernel matrix (equation 4) captures the dominant patterns of information content because of its the dependence on the prior error covariance matrix and on the observational density as quantified by the observational error covariance matrix and by the sparsity structure of (Nesser et al. 2021). The patterns of information content are then given by the eigenvectors of the averaging kernel matrix described by the columns of

where is the matrix of the first *k* eigenvectors of .

We then construct the Jacobian matrix on the basis of these eigenvectors. We select the eigenvectors that have at least twice as much signal as noise and perturb those patterns in the forward model. We multiply the resulting Jacobian matrix by the matrix

to restore the original state dimension, yielding an updated Jacobian matrix estimate . We then recompute the eigenvectors using perturb the eigenvectors that explain xx% of the information content, and construct the updated Jacobian matrix . This update scheme follows Nesser et al. (2021), which found that the information content of the inversion was more sensitive to perturbation simulations in the second update. The information content associated with both and contains common contributions from the prior and observational error covariance matrices and from the forward model. As a result, additional iterations would not add new information content and we take as the Jacobian matrix for our inversion.

[Paragraph analyzing the eigenvectors of the resulting Jacobian matrix.]

**3 Results and discussion**

**4 Conclusions**