Anthropogenic emissions of methane in the United States

Scott M. Miller, Steven C. Wofsy, Anna M. Michalak, Eric A. Kort, Arlyn E. Andrews, Sebastien C. Biraud, Edward J. Dlugokencky, Janusz Eluszkiewicz, Marc L. Fischer, Greet Janssens-Maenhout, Ben R. Miller, John B. Miller, Stephen A. Montzka, Thomas Nehrkorn, and Colm Sweeney

Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138; Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305; Department of Atmospheric, Ocean, and Space Sciences, University of Michigan, Ann Arbor, MI 48109; Global Monitoring Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO 80305; Earth Sciences Division, and Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; Atmospheric and Environmental Research, Lexington, MA 02421; Institute for Environment and Sustainability, European Commission Joint Research Centre, 21027 Ispra, Italy; and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309

Edited by Mark H. Thiemens, University of California, San Diego, La Jolla, CA, and approved October 18, 2013 (received for review August 5, 2013)

This study quantitatively estimates the spatial distribution of anthropogenic methane sources in the United States by combining comprehensive atmospheric methane observations, extensive spatial datasets, and a high-resolution atmospheric transport model. Results show that current inventories from the US Environmental Protection Agency (EPA) and the Emissions Database for Global Atmospheric Research underestimate methane emissions nationally by a factor of ~1.5 and ~1.7, respectively. Our study indicates that emissions due to ruminants and manure are up to twice the magnitude of existing inventories. In addition, the discrepancy in methane source estimates is particularly pronounced in the south-central United States, where we find total emissions are ~2.7 times greater than in most inventories and account for 24 ± 3% of national emissions. The spatial patterns of our emission fluxes and observed methane–propane correlations indicate that fossil fuel extraction and refining are major contributors (45 ± 13%) in the south-central United States. This result suggests that regional methane emissions due to fossil fuel extraction and processing could be 4.9 ± 2.6 times larger than in EDGAR, the most comprehensive global methane inventory. These results cast doubt on the US EPA’s recent decision to downscale its estimate of national natural gas emissions by 25–30%. Overall, we conclude that methane emissions associated with both the animal husbandry and fossil fuel industries have larger greenhouse gas impacts than indicated by existing inventories.

Significance

Successful regulation of greenhouse gas emissions requires knowledge of current methane emission sources. Existing state regulations in California and Massachusetts require ~15% greenhouse gas emissions reductions from current levels by 2020. However, government estimates for total US methane emissions may be biased by 50%, and estimates of individual source sectors are even more uncertain. This study uses atmospheric methane observations to reduce this level of uncertainty. We find greenhouse gas emissions from agriculture and fossil fuel extraction and processing (i.e., oil and/or natural gas) are likely a factor of two or greater than cited in existing studies. Effective national and state greenhouse gas reduction strategies may be difficult to develop without appropriate estimates of methane emissions from these source sectors.


This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: scott.m.miller@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1314392110/-/DCSupplemental.
values from Kort et al. are 47 ± 20% higher than EPA (13). Assessments of CH₄ sources to inform policy (e.g., regulating emissions or managing energy resources) require more accurate, verified estimates for the United States.

This study estimates anthropogenic CH₄ emissions over the United States for 2007 and 2008 using comprehensive CH₄ observations at the surface, on telecommunications towers, and from aircraft, combined with an atmospheric transport model and a geostatistical inverse modeling (GIM) framework. We use auxiliary spatial data (e.g., on population density and economic activity) and leverage concurrent measurements of alkanes to help attribute emissions to specific economic sectors. The work provides spatially resolved CH₄ emissions estimates and associated uncertainties, as well as information by source sector, both previously unavailable.

**Model and Observation Framework**

We use the Stochastic Time-Inverted Lagrangian Transport model (STILT) to calculate the transport of CH₄ from emission points at the ground to measurement locations in the atmosphere (21). STILT follows an ensemble of particles backward in time, starting from each observation site, using wind fields and turbulence modeled by the Weather Research and Forecasting (WRF) model (22). STILT derives an influence function (“footprint”) units: ppb CH₄ per unit emission flux) linking upwind emissions to each measurement. Inputs of CH₄ from surface sources along the ensemble of back-trajectories are averaged to compute the CH₄ concentration for comparison with each observation.

We use observations for 2007 and 2008 from diverse locations and measurement platforms. The principal observations derive from daily flask samples on tall towers (4,984 total observations) and vertical profiles from aircraft (7,710 observations). Tower-based observations are collected as part of the National Oceanic and Atmospheric (NOAA)/Department of Energy (DOE) cooperative air sampling network, and aircraft-based data are obtained from regular NOAA flights (23), regular DOE flights (24), and from the Stratosphere-Troposphere Analyses of Regional Transport 2008 (START08) aircraft campaign (25); all data are publicly available from NOAA and DOE. These observations are displayed in Fig. 2 and discussed further in the SI Text (e.g., Fig. S2). We use a GIM framework (26, 27) to analyze the footprints for each of the 12,694 observations, and these footprints vary by site and with wind conditions. In aggregate, the footprints provide spatially resolved coverage of most of the continental United States, except the southeast coastal region (Fig. S3).

The GIM framework, using footprints and concentration measurements, optimizes CH₄ sources separately for each month of 2007 and 2008 on a 1° × 1° latitude-longitude grid for the United States. The contributions of fluxes from natural wetlands are modeled first and subtracted from the observed CH₄ (2.0 TgC yr⁻¹ for the continental United States); these fluxes are much smaller than anthropogenic sources in the United States and thus would be difficult to independently constrain from atmospheric data (SI Text). The GIM framework represents the flux distribution for each month using a deterministic spatial model plus a stochastic spatially correlated residual, both estimated from the atmospheric observations. The deterministic component is given by a weighted linear combination of spatial activity data from the EDGAR 4.2 inventory; these datasets include any economic or demographic data that may predict the distribution of CH₄ emissions (e.g., gas production, human and ruminant population densities, etc.). Both the selection of the activity datasets to be retained in the model and the associated weights (emission factors) are optimized to best match observed CH₄ concentrations. Initially, seven activity datasets are included from EDGAR 4.2, (i) population, (ii) electricity production from power plants, (iii) ruminant population count, (iv) oil and conventional gas production, (v) oil refinery production, (vi) rice production, and (vii) coal production.

We select the minimum number of datasets with the greatest predictive ability using the Bayesian Information Criterion (BIC) (SI Text) (28). BIC numerically scores all combinations of available datasets based on how well they improve goodness of fit and applies a penalty that increases with the number of datasets retained.

The stochastic component represents sources that do not fit the spatial patterns of the activity data (Fig. S4). GIM uses
a covariance function to describe the spatial and temporal correlation of the stochastic component and optimizes its spatial and temporal distribution simultaneously with the optimization of the activity datasets in the deterministic component (SI Text, Fig. S5) (26–28). Because of the stochastic component, the final emissions estimate can have a different spatial and temporal distribution from any combination of the activity data.

If the observation network is sensitive to a broad array of different source sectors and/or if the spatial activity maps are effective at explaining those sources, many activity datasets will be included in the deterministic model. If the deterministic model explains the observations well, the magnitude of CH₄ emissions in the stochastic component will be small, the assignment to specific sectors will be unambiguous, and uncertainties in the emissions estimates will be small. This result is not the case here, as discussed below (see Results).

A number of previous studies used top-down methods to constrain anthropogenic CH₄ sources from global (29–33) to regional (13–15, 34–38) scales over North America. Most regional studies adopted one of three approaches: use a simple box model to estimate an overall CH₄ budget (14), estimate a budget using the relative ratios of different gases (15, 37–39), or estimate scaling factors for inventories by region or source type (13, 34–36). The first two methods do not usually give explicit information about geographic distribution. The last approach provides information about the geographic distribution of sources, but results hinge on the spatial accuracy of the underlying regional or sectoral emissions inventories (40).

Here, we are able to provide more insight into the spatial distribution of emissions; like the scaling factor method above, we leverage spatial information about source sectors from an existing inventory, but in addition we estimate the distribution of emissions where the inventory is deficient. We further bolster attribution of regional emissions from the energy industry using the observed correlation of CH₄ and propane, a gas not produced by biogenic processes like livestock and landfills.

**Results**

**Spatial Distribution of CH₄ Emissions.** Fig. 3 displays the result of the 2-y mean of the monthly CH₄ inversions and differences from the EDGAR 4.2 inventory. We find emissions for the United States that are a factor of 1.7 larger than the EDGAR inventory. The optimized emissions estimated by this study bring the model closer in line with the observations (Fig. 4, Figs. S6 and S7). Posterior emissions fit the CH₄ observations \( R^2 = 0.64, \text{root mean square error (RMSE) } = 31 \text{ ppb} \) much better than EDGAR v4.2 \( R^2 = 0.23, \text{RMSE} = 49 \text{ ppb} \). Evidently, the spatial distribution of EDGAR sources is inconsistent with emissions patterns implied by the CH₄ measurements and associated footprints.

Several diagnostic measures preclude the possibility of major systematic errors in WRF–STILT. First, excellent agreement between the model and measured vertical profiles from aircraft implies little bias in modeled vertical air mixing (e.g., boundary-layer heights) (Fig. 4). Second, the monthly posterior emissions estimated by the inversion lack statistically significant seasonality (Fig. S8). This result implies that seasonally varying weather patterns do not produce detectable biases in WRF–STILT. SI Text discusses possible model errors and biases in greater detail.

CH₄ observations are sparse over parts of the southern and central East Coast and in the Pacific Northwest. Emissions for these regions therefore rely more strongly on the deterministic component of the flux model, with weights constrained primarily by observations elsewhere. Therefore, emissions in these areas, including from coal mining, are poorly constrained (SI Text).

**Contribution of Different Source Sectors.** Only two spatial activity datasets from EDGAR 4.2 are selected through the BIC as meaningful predictors of CH₄ observations over the United States: population densities of humans and of ruminants (Table S1). Some sectors are eliminated by the BIC because emissions are situated far from observation sites (e.g., coal mining in West Virginia or Pennsylvania), making available CH₄ data insensitive to these predictors. Other sectors may strongly affect observed concentrations but are not selected, indicating that the spatial datasets from EDGAR are poor predictors for the distribution of observed concentrations (e.g., oil and natural gas extraction and oil refining). Sources from these sectors appear in the stochastic component of the GIM (SI Text).

The results imply that existing inventories underestimate emissions from two key sectors: ruminants and fossil fuel extraction and/or processing, discussed in the remainder of this section.

We use the optimized ruminant activity dataset to estimate the magnitude of emissions with spatial patterns similar to animal husbandry and manure. Our corresponding US budget of 12.7 ± 5.0 TgC·y⁻¹ is nearly twice that of EDGAR and EPA (6.7 and 7.0, respectively). The total posterior emissions estimate over the northern plains, a region with high ruminant density but little fossil fuel extraction, further supports the ruminant estimate (Nebraska, Iowa, Wisconsin, Minnesota, and South Dakota). Our total budget for this region of 3.4 ± 0.7 compares with 1.5 TgC·y⁻¹ in EDGAR. Ruminants and agriculture may also be

![Fig. 3.](image-url) The 2-y averaged CH₄ emissions estimated in this study (A) compared against the commonly used EDGAR 4.2 inventory (B and C). Emissions estimated in this study are greater than in EDGAR 4.2, especially near Texas and California.
partially responsible for high emissions over California (41). EDGAR activity datasets are poor over California (42), but several recent studies (34, 36–38, 41) have provided detailed top-down emissions estimates for the state using datasets from state agencies.

Existing inventories also greatly underestimate CH$_4$ sources from the south-central United States (Fig. 3). We find the total CH$_4$ source from Texas, Oklahoma, and Kansas to be $8.1 \pm 0.96$ TgC·y$^{-1}$, a factor of 2.7 higher than the EDGAR inventory. These three states alone constitute $\sim 24 \pm 3\%$ of the total US anthropogenic CH$_4$ budget or 3.7% of net US greenhouse gas emissions [in CO$_2$ equivalents (10)].

Texas and Oklahoma were among the top five natural gas producing states in the country in 2007 (18), and aircraft observations of alkanes indicate that the natural gas and/or oil industries play a significant role in regional CH$_4$ emissions. Concentrations of propane (C$_3$H$_8$), a tracer of fossil hydrocarbons (43), are strongly correlated with CH$_4$ at NOAA/DOE aircraft monitoring locations over Texas and Oklahoma ($R^2 = 0.72$) (Fig. 5). Correlations are much weaker at other locations in North America ($R^2 = 0.11$ to 0.64).

We can obtain an approximate CH$_4$ budget for fossil-fuel extraction in the region by subtracting the optimized contributions associated with ruminants and population from the total emissions. The residual (Fig. S4C) represents sources that have spatial patterns not correlated with either human or ruminant density in EDGAR. Our budget sums to $3.7 \pm 2.0$ TgC·y$^{-1}$, a factor of $4.9 \pm 2.6$ greater than oil and gas emissions in EDGAR v4.2 ($0.75$ TgC·y$^{-1}$) and a factor of $6.7 \pm 3.6$ greater than EDGAR sources from solid waste facilities ($0.55$ TgC·y$^{-1}$), the two major sources that may not be accounted for in the deterministic component. The population component likely captures a portion of the solid waste sources so this residual methane budget more likely represents natural gas and oil emissions than landfills. SI Text discusses in detail the uncertainties in this sector-based emissions estimate. We currently do not have the detailed, accurate, and spatially resolved activity data (fossil fuel extraction and processing, ruminants, solid waste) that would provide more accurate sectorial attribution.

Katzenstein et al. (2003) (14) were the first to report large regional emissions of CH$_4$ from Texas, Oklahoma, and Kansas; they cover an earlier time period (1999–2002) than this study. They used a box model and 261 near-ground CH$_4$ measurements taken over 6 d to estimate a total Texas–Oklahoma–Kansas CH$_4$ budget (from all sectors) of $3.8 \pm 0.75$ TgC·y$^{-1}$. We revise their...
estimate upward by a factor of two based on the inverse model and many more measurements from different platforms over two full years of data. SI Text further compares the CH\textsubscript{4} estimate in Katzenstein et al. and in this study.

Discussion and Summary

This study combines comprehensive atmospheric data, diverse datasets from the EDGAR inventory, and an inverse modeling framework to derive spatially resolved CH\textsubscript{4} emissions and information on key source sectors. We estimate a mean annual US anthropogenic CH\textsubscript{4} budget for 2007 and 2008 of 33.4 ± 1.4 TgC year\textsuperscript{-1} or ~7-8% of the total global CH\textsubscript{4} source. This estimate is a factor of 1.5 and 1.7 larger than EPA and EDGAR v4.2, respectively. CH\textsubscript{4} emissions from Texas, Oklahoma, and Kansas alone account for 24% of US methane emissions, or 3.7% of the total US greenhouse gas budget.

The results indicate that drilling, processing, and refining activities over the south-central United States have emissions as much as 4.9 ± 2.6 times larger than EDGAR, and livestock operations across the US have emissions approximately twice that of recent inventories. The US EPA recently decreased its CH\textsubscript{4} emission factors for fossil fuel extraction and processing by 25-30% (for 1990-2011), (10), but we find that CH\textsubscript{4} data from across North America instead indicate the need for a larger adjustment of the opposite sign.

ACKNOWLEDGMENTS. For advice and support, we thank Roisin Commane, Elizabeth Kort, Hendrik van der Laan-Luijben, and Matthew Hayek (Harvard University); Robert Hardtits (Environmental Defense Fund); Hanqin Tian and Bowen Zhang (Auburn University); Jed Kaplan (École Polytechnique Fédérale de Lausanne); Kimberly Mueller and Christopher Weber (Institute for Defense Analyses Science and Technology Policy Institute); Nadia Oussaye; and Gregory Carmichael for their contributions. In addition, we thank the National Aeronautics and Space Administration (NASA) Advanced Supercomputing Division for computing help, P. Lang, K. Sours, and C. Siso for analysis of National Oceanic and Atmospheric Administration (NOAA) flask; and B. Hall for calibration standards work. This work was supported by the American Meteorological Society Graduate Student Fellowship/Department of Energy (DOE) Atmospheric Radiation Measurement Program, a DOE Computational Science Graduate Fellowship, and the National Science Foundation Graduate Research Fellowship Program. NOAA measurements were funded in part by the Atmospheric Composition and Climate Program and the Carbon Cycle Program of NOAA’s Climate Program Office. Support for this research was provided by NASA Grants NNX08ARATG and NNX11AAGT4, NOAA Grants NAOA101222 and NAOA1101558, National Science Foundation (NSF) Grant ATM-0628575, and Environmental Defense Fund Grant 0146-10100 (to Harvard University). Measurements at Walnut Grove were supported in part by a California Energy Commission Public Interest Environmental Research Program grant to Lawrence Berkeley National Laboratory through the US Department of Energy under Contract DE-AC02-05CH11231. DOE flights were supported by the Office of Biological and Environmental Research of the US Department of Energy under Contract DE-AC02-05CH11231 as part of Atmospheric Radiation Measurement (ARM) National Aeronautics and Space Administration (NASA) Aerial Facility, and Terrestrial Ecosystem Science Program. Weather Research and Forecasting-Stochastic Time-Inverted Lagrangian Transport model development at Atmospheric and Environmental Research has been funded by NSF Grant ATM-0836153, NASA, NOAA, and the US intelligence community.