Analysis of Oil and Gas Ethane and Methane Emissions in the Southcentral and Eastern United States Using Four Seasons of Continuous Aircraft Ethane Measurements

Z. R. Barkley[1] [iD], K. J. Davis[1,2] [iD], S. Feng[1,3] [iD], Y. Y. Cui[1] [iD], A. Fried[4] [iD], P. Weibring[4] [iD], D. Richter[4] [iD], J. G. Walega[4] [iD], S. M. Miller[5] [iD], M. Eckl[6] [iD], A. Roiger[6] [iD], A. Fiehn[6] [iD], and J. Kostinek[6]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA, [2]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA, [3]Pacific Northwest National Laboratory, Atmospheric Sciences and Global Change Division, Richland, WA, USA, [4]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA, [5]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, UK, [6]Deutsches Zentrum für Luft- und Raumfahrt e.V., Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Abstract In the last decade, much work has been done to better understand methane (CH_4) emissions from the oil and gas (O&G) industry in the United States. Ethane (C_2H_6), a gas that is co-emitted with thermogenic sources of CH_4, is emitted in the US predominantly by the O&G sector. In this study, we perform an inverse analysis on 200 h of atmospheric boundary layer C_2H_6 measurements to estimate C_2H_6 emissions from the US O&G sector. Measurements were collected from 2017 to 2019 as part of the Atmospheric Carbon and Transport (ACT) America aircraft campaign and encompass much of the central and eastern United States. We find that for the fall, winter, and spring campaigns, C_2H_6 data consistently exceeds values that would be expected based on EPA O&G leak rate estimates by more than 50%. C_2H_6 observations from the summer 2019 data set show significantly lower C_2H_6 enhancements in the southcentral region that cannot be reconciled with data from the other three seasons, either due to complex meteorological conditions or a temporal shift in the emissions. Combining the fall, winter, and spring C_2H_6 posterior emissions estimate to an inventory of O&G CH_4 emissions, we estimate that O&G CH_4 emissions are larger than EPA inventory values by 48%–76%. Uncertainties in the gas composition data limit the accuracy of using C_2H_6 as a proxy for O&G CH_4 emissions. These limits could be resolved retroactively by increasing the availability of industry-collected gas composition data.

Plain Language Summary Methane is a potent greenhouse gas responsible for a quarter of the warming the climate has experienced thus far. The oil and gas (O&G) sector is a significant source of methane through leaks in its infrastructure. Recent studies of individual basins have found emissions from O&G in the US to be greater than inventory estimates, but difficulties arise with source attribution in broader scale studies due to the numerous potential sources of methane. This study quantifies methane emissions from O&G by looking at atmospheric ethane, a gas whose emissions originate mostly from O&G in the US. Hundreds of hours of boundary layer ethane observations were collected via aircraft over the course of four seasons between 2017 and 2019. These observations are compared with model-projected ethane values based on our current knowledge of ethane emissions, and those emissions are adjusted to best match the observed data. We find ethane emissions are grossly underestimated in the US. Because ethane is co-emitted with O&G methane sources, this underestimation of ethane reflects similar underestimations in O&G methane emissions. We conclude that US inventories are underestimating methane emissions from O&G by 48%–76%.

1. Introduction

Methane (CH_4) is an important greenhouse gas with 28–35 times the warming potential of carbon dioxide over a 100 year period (Myhre et al., 2013). Global CH_4 concentrations in the atmosphere have nearly tripled since pre-industrial times, mainly driven by anthropogenic activity, and are responsible for a fourth of the increased radiative forcing on the planet (Myhre et al., 2013). Although CH_4 concentrations stabilized for a

brief period in the early 2000s, global concentrations began increasing again by 2007 and have continued to increase through the present date (Dlugokencky, 2021; Nisbet et al., 2019).

Efforts to understand causes for increasing global CH_4 trends are hampered by difficulties related to source attribution which arise from complex emission processes, limited data availability, and disagreements between bottom-up inventories and top-down assessments (Saunois et al., 2020). CH_4 has numerous anthropogenic sources, including animal agriculture, fossil fuel extraction, and waste management. In addition to anthropogenic emitters, there are natural sources of CH_4 emissions that play a large role in the global CH_4 budget. Of particular importance to the global CH_4 budget are emissions from anaerobic respiration in wetlands, which create unique challenges to CH_4 source attribution. Unlike most anthropogenic sources, whose emissions can be spatially mapped out using inventory techniques and are relatively consistent in magnitude on an annual timeframe, wetland emissions are more difficult to describe, with an uncertain spatial pattern and large seasonal variability in emissions based on soil temperature and moisture (Yvon-Durocher et al., 2014). Many prior studies have examined seasonal emissions and magnitudes of local wetland sources using flux tower measurements (Grant & Roulet, 2002; Matthes et al., 2014), but extrapolating these results to a continental or global scale is challenging.

Difficulties with source attribution extend into the US, where both anthropogenic and natural sources play a large role in the country's CH_4 budget. In the US, greenhouse gas emissions from anthropogenic sources are quantified and aggregated on a national scale through a bottom-up inventory created by the United States Environmental Protection Agency (EPA). The largest sources of anthropogenic CH_4 emissions projected by this inventory are leaks in oil and gas (O&G) infrastructure, enteric fermentation and manure management related to livestock, and anaerobic respiration occurring in landfills (US Environmental Protection Agency, 2020b). These sources were responsible for 83% of US anthropogenic CH_4 emissions in the 2018 inventory estimate. Natural sources in the US are dominated by wetlands and are not tracked by the EPA's inventory. WetCHARTs, a global wetland emissions ensemble, is generally used as the prior for CH_4 emissions from wetlands in the US (Bloom et al., 2017; Maasakkers et al., 2016; Sheng et al., 2017). WetCHARTs ensemble members estimate the country's wetland emissions to vary from values that are insignificant to totals rivaling those from anthropogenic emissions, with seasonal dependence and spatial variability between ensemble members, resulting in large uncertainties in the overall US CH_4 budget.

Contrary to recent global increases in the atmospheric growth rate of CH_4, the EPA's inventory estimates of anthropogenic CH_4 emission in the US show a 10% decrease in the last decade, from 28 Tg in 2008 to 25 Tg in 2018. This decrease is driven primarily by projected decreases in emissions from the energy sector, despite a 50% increase in gas production and a greater than 200% increase in oil production during the 10 year period (US Energy Information Administration, 2020b). Over the last several years, various atmospheric studies monitoring emissions from O&G from individual wellpads (Caulton et al., 2019; Rella et al., 2015; Robertson et al., 2017), basins (Barkley et al., 2017; Karion et al., 2013, 2015; Peischl et al., 2015, 2016, 2018; Pétron et al., 2012, 2014), and entire regions (Barkley, Davis, et al., 2019) have consistently found emission rates larger than the EPA inventory, raising concerns of a broad underestimation of leaks from the O&G sector (Alvarez et al., 2018). However, large-scale CH_4 inversion studies involving the US have not been as conclusive, with differing findings as to the accuracy of inventory emissions from O&G, animal agriculture, and wetlands (Maasakkers et al., 2019; Miller et al., 2013; Sheng et al., 2018; Yu et al., 2020). The large spread of uncertainty regarding the magnitude, spatial distribution, and seasonality of CH_4 emissions from wetlands, as well as the numerous other potential sources of CH_4, poses a challenge to large scale studies attempting source attribution of detected CH_4 signals.

One common method to disaggregate certain sources is to measure both CH_4 and ethane (C_2H_6) mixing ratios (Barkley, Lauvaux et al., 2019; McKain et al., 2015; Pétron et al., 2020). Sources of C_2H_6 coincide with thermogenic CH_4 emitters (O&G and C_2H_6 supply chain, biomass burning), but not biogenic sources (wetlands, animal agriculture, and landfills) (Yacovitch et al., 2014). Thus, a correlation (or lack of) between C_2H_6 and CH_4 can provide information on the sector responsible for observed CH_4 enhancements. While global C_2H_6 emissions are split between the energy sector and biomass burning (Xiao et al., 2008), the National Emissions Inventory (NEI) 2017 C_2H_6 inventory attributes more than two-thirds of C_2H_6 emissions in the US to the O&G sector, making it an ideal tracer for identifying O&G emissions in the region (US Environmental Protection Agency, 2020a). This technique has been used in various facility and basin-scale

studies across individual O&G basins to verify that the measured enhancements are associated with O&G activity (Peischl et al., 2018; Roscioli et al., 2015). Similarly, C_2H_6 concentrations from a network of flask measurements have been used to infer trends in emissions from O&G emissions on broader scales (Dalsøren et al., 2018; Helmig et al., 2016; Lan et al., 2019; Tzompa-Sosa et al., 2017; Xiao et al., 2008).

In addition to its role in attributing CH_4 emissions to different source types, understanding C_2H_6 emissions is important in atmospheric air quality studies. Oxidation of C_2H_6 is involved in the formation of ozone and other photochemical constituents (peroxyacetyl nitrate, acetaldehyde, formaldehyde, and carbon monoxide) giving rise to poor air quality (Aikin et al., 1982; Hodnebrog et al., 2018). Furthermore, since the decay of C_2H_6 in the troposphere occurs primarily via reactions with the hydroxyl (OH) radical, significant local enhancements of C_2H_6 from O&G operations can affect the local concentrations and lifetimes of other organic gases. Additionally, other volatile organic compounds are co-emitted with CH_4 and C_2H_6 in O&G basins, contributing to the formation of surface ozone (Edwards et al., 2014; Gilman et al., 2013). Therefore, C_2H_6 in its production of ozone acts as a direct greenhouse gas, and at the same time, its effects on CH_4 and other organics act as an indirect greenhouse gas (Kort et al., 2016).

Advancements in technology have resulted in the increased availability of instrumentation capable of measuring C_2H_6 mixing ratios precisely at high-temporal resolution (Kostinek et al., 2019; Weibring et al., 2020; Yacovitch et al., 2014), expanding our capabilities of studying C_2H_6 emissions. The Atmospheric Carbon and Transport-America (ACT-America) mission was a 5-season aircraft campaign across the central and eastern US designed to examine various trace gases and their transport in the atmosphere. During four of the seasonal campaigns, continuous C_2H_6 measurements were collected, producing more than 200 h of boundary layer C_2H_6 data capable of capturing various plume structures related to O&G activity on a regional scale. In this study, we examine the characteristics of this unprecedented C_2H_6 data set. From this analysis, we (1) provide up-to-date maps of C_2H_6 distributions in the boundary layer from 2017 to 2019 spanning four seasons over the eastern and southcentral US; (2) quantify C_2H_6 emissions from O&G sources in these regions using atmospheric inversion techniques and compare these values to existing bottom-up inventories; and (3) utilize these updated C_2H_6 inventories in determining CH_4 emissions from the O&G sector.

2. Methods

2.1. ACT-America Aircraft Campaign and Observational Data Set

The ACT-America campaign was a NASA-funded Earth Venture suborbital aircraft mission designed to study the transport of trace gases throughout the US, with flight activities spanning summer 2016 to summer 2019 (Davis et al., 2021). During this time, five individual campaigns covering all four seasons (summer twice) were conducted using two aircraft, a NASA Langley Research Center Beechcraft B200 King Air and a NASA Wallops Flight Facility Lockheed C130 Hercules, collecting meteorological and trace gas data within fair weather and frontal weather patterns across the central and eastern US. Flights took place during afternoon hours and were typically composed of long horizontal transects (>500 km) at multiple heights, with designated boundary layers and free troposphere legs. Altogether, a total of 1,140.7 h of aircraft observations were collected across 121 flight days. For this study, boundary layer observations are primarily used in the analysis, identified using flags from the ACT-America data set that determine boundary layer height based on potential temperature profiles from the flights, and constitute one-third of the entire ACT-America data set (Pal et al., 2020). Data from the ACT-America campaign used in this study can be found at https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=37.

For all five seasonal campaigns, high frequency CH_4 dry air mole fraction data were collected on both aircraft using a commercial PICARRO G2401-m instrument adapted with a custom inlet system for drying and conditioning the sample air (DiGangi et al., 2018). During winter 2017, fall 2017, spring 2018, and summer 2019 campaigns, high frequency C_2H_6 mole fraction data were collected on the B200 aircraft using the CAMS-2 (Compact Airborne Multi-Species Spectrometer) instrument (Weibring et al., 2020). The CAMS-2 C_2H_6 measurements when averaged over time and linearly regressed versus NOAA portable flask packages collected during the flights yielded slopes in the 0.990 to 1.031 range across seasonal campaigns, with r^2 values between the two measurements of 0.996 (Baier et al., 2020; Weibring et al., 2020), providing high confidence in the accuracy of its measurements. Additionally, during the fall 2017 and summer 2019 campaigns,
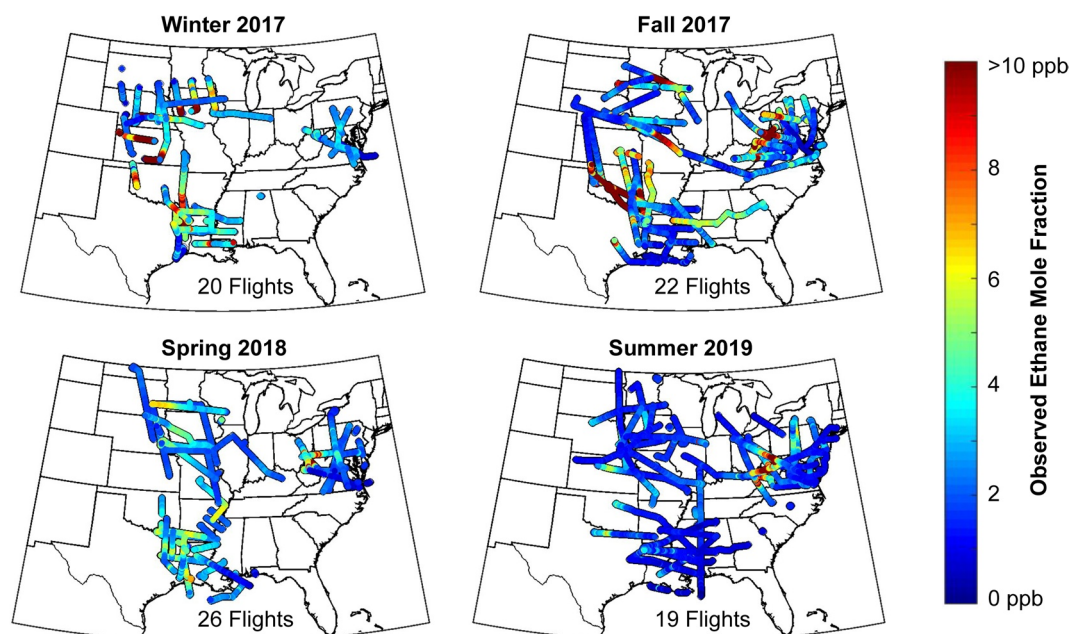
Figure 1. Continuous boundary layer C_2H_6 measurements were collected by the B200 and C130 aircraft during each seasonal campaign. The total number of flight days are shown at the bottom of each seasonal figure.

the C130 aircraft was equipped with a quantum and interband cascade laser spectrometer (QCLS) capable of continuous in situ C_2H_6 measurements (Kostinek et al., 2019). The QCLS instrument performed in-flight two-point calibrations every three to 10 minutes to ensure accurate measurements throughout the flights and, similar to the CAMS-2 instrumentation, measurements were compared to NOAA portable flask samples taken in flight (Kostinek et al., 2019). Altogether, the entire C_2H_6 data set contains more than 500 h of continuous airborne C_2H_6 observations, of which more than 200 h were within the atmospheric boundary layer (ABL), making it an ideal data set to study atmospheric C_2H_6 plumes and structures throughout the central and eastern US.

From the observational data set, large C_2H_6 plumes were consistently observed in the boundary layer downwind of Texas/Oklahoma/Louisiana in the southcentral US and along the western Appalachians in the northeast (Figure 1). C_2H_6 plumes could be observed in the midwestern flights as far north as Minnesota when consistent southerly winds were present. C_2H_6 mixing ratios within the plumes were largest during the fall and winter campaign. During the summer 2019 campaign, C_2H_6 plumes greater than 7 ppb were detected in the western Appalachian similar to other seasons, but signals in the southcentral US were substantially smaller (<3 ppb). Additionally, boundary layer C_2H_6 mixing ratios in the summer were often similar in value to their free tropospheric counterparts, a trait not observed during other seasons (Figure S1).

Influence functions for observations within the boundary layer were created using the Lagrangian particle dispersion model FLEXPART-WRF (Brioude et al., 2013) to provide information on the area captured by the ACT-America C_2H_6 data set. The model was run at a 27 km resolution over the North American domain and was meteorologically driven using WRF-Chem simulations developed as part of the ACT-America campaign (Feng et al., 2019). Every 30 s of flight time, 5,000 particles were released from the aircraft location and traced back 10 days in time. Further details on the development of the influence functions can be found in Cui et al. (2015).

For each season, we sum up the number of flight days where a surface grid cell is in the top 80% of the boundary layer influence function for a given day to examine the amount of coverage of different O&G plays captured during the seasonal campaigns (Figure 2). Much of the midwestern and southcentral US is captured across all four seasons, encompassing many of the major O&G basins including Haynesville, Fayetteville, Anadarko, Woodford, and the Barnett. The Eagle Ford and the Permian plays, located on the extreme southern and western portions of the southcentral US, have less uniform seasonal coverage, with
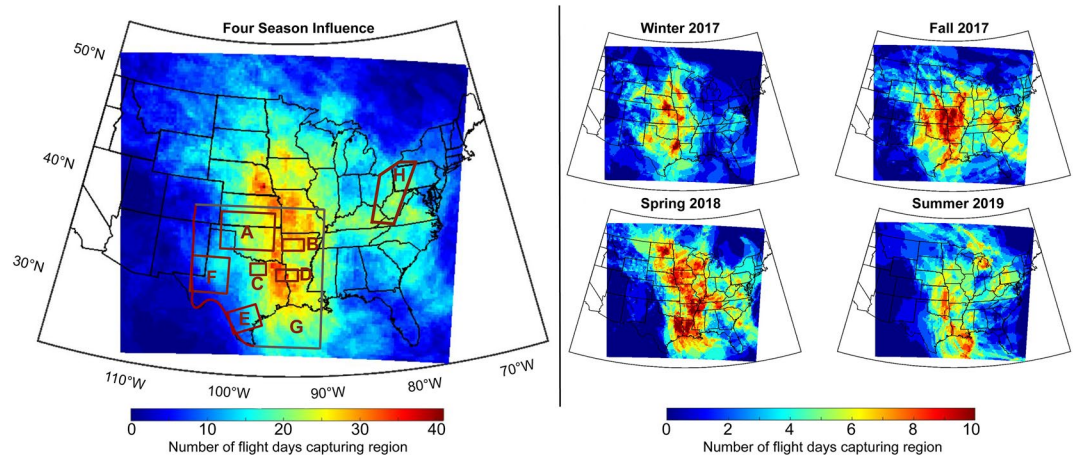
Figure 2. (Left) Map showing the frequency that regions are captured by flights from the four season ACT-America data set used in this study. A flight day is considered to have influence from a grid if the grid is within the top 80% of the boundary layer influence function for that day. O&G basins and plays pertinent to this study are highlighted in boxes and are labeled as follows: (A) Anadarko. (B) Fayetteville. (C) Barnett. (D) Haynesville. (E) Eagle Ford. (F) Permian. (G) Southcentral U.S. (H) Appalachia. Basins are defined using an approximation of the location of their O&G emissions from the EPA 2012 Gridded Methane Inventory. (Right) Same as left, but broken down into seasonal campaigns.

few flights capturing the regions in the spring and summer campaign, but better coverage during the fall and winter campaign when winds are climatologically more westerly. In the northeast, many flight tracks were designed to fly downwind of the Marcellus and Utica O&G basins in the western Appalachian regions of West Virginia and western Pennsylvania, and thus this area has significant coverage in the influence maps for all four seasons.

2.2. Inversion Framework

Boundary layer C_2H_6 mixing ratio data collected from the ACT-America campaign is used in this study to perform an inversion and learn about C_2H_6 emissions from O&G sources in the southcentral and eastern US. The basic inversion framework is similar to previous work (Lauvaux et al., 2012; Sheng et al., 2018) and is expressed by cost function,

$$J\left(x\right) = \frac{1}{2}(y - \mathbf{H}x)^T \mathbf{R}^{-1}(y - \mathbf{H}x) + \frac{1}{2}(x - x_0)^T \mathbf{B}^{-1}\left(x - x_0\right) \tag{1}$$

In this equation, we solve for a posterior emissions grid x that minimizes the cost function J using influence functions ($\mathbf{H}$) that translate the flux field to a modelled enhancement ($\mathbf{H}x$). In the cost function, two terms control the solution. The first term is a cost related to the mismatch between the posterior modelled enhancements versus observed mixing ratio enhancement (y), with greater discrepancies resulting in a larger cost term. Here, $\mathbf{R}$ is the observation error covariance matrix and weights the first term based on the confidence in the observations and model transport. The second term in the cost function equation is a cost related to the change between the posterior flux (x) and the prior flux (x_0), with larger changes resulting in a greater cost. Here, $\mathbf{B}$ is the flux error covariance matrix and weights the second term based on the confidence in the prior flux field. Minimizing the cost function with respect to x yields

$$x = x_0 + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(y - \mathbf{H}x_0) \tag{2}$$

and solving for x yields the posterior flux field (Brasseur & Jacob, 2017), assuming Gaussian errors. For this study, observations from each season are grouped together and a posterior flux map is solved for each of the seasonal flight campaigns.

A best guess C_2H_6 emissions map was created to serve as a prior for the inversion. To create this prior, CH_4 emissions from the O&G sector of the EPA Gridded 2012 CH_4 Emissions Inventory (Maasakkers

et al., 2016) were multiplied by expected molar C_2H_6/CH_4 ratios of each basin (Table S2), resulting in an C_2H_6 emissions map. For Eagle Ford, Haynesville, Fayetteville, Barnett, Permian, Denver-Julesburg, and the Bakken O&G basins, atmospheric measurements from NOAA aircraft studies were available to derive observed basin-wide C_2H_6/CH_4 ratios (Peischl et al., 2015, 2018; Tzompa-Sosa et al., 2017). For the Anadarko and Appalachian region where representative atmospheric measurements were not available, data from the United States Geological Survey were used to create a spatial map of C_2H_6/CH_4 ratios to apply to these regions (Kitanidis, 1997; US Geological Survey, 2018) (See Section S1 and Figure S1 for additional info). For all emissions related to transmission, storage, and distribution, an C_2H_6/CH_4 ratio of 0.027 was applied (Plant et al., 2019). For C_2H_6 emissions unrelated to O&G operations, as well as sources in Canada and Mexico, values from the US National Emissions Inventory 2017 (NEI2017) for C_2H_6 are used. These other sources when multiplied by the influence functions from the flight campaign represent 31% of the signal in the prior inventory, and are generally uncorrelated spatially with O&G sources. Overall, the described C_2H_6 emissions map represents our best guess as to representing C_2H_6 emissions based on the EPA's 2012 bottom-up O&G CH_4 emissions inventory and will be referred to as the "Default" map henceforth. Although O&G emissions from the Default map are based on EPA estimates from 2012, nationwide bottom-up O&G CH_4 emission estimates from the EPA have decreased by only 4% between 2012 and 2018 (US Environmental Protection Agency, 2020b). For this reason, we do not expect the 2012 EPA Gridded 2012 CH_4 Emissions Inventory to be unrepresentative of the EPA estimates of total O&G emissions during the years of the flight campaign (2017–2019), though individual basins may be more susceptible to higher uncertainty.

Two additional C_2H_6 emission maps were created to test the sensitivity of the inversion to different priors. The first alternative map was created by taking the EPA CH_4 emissions map used in the creation of the Default inventory and applying a flat C_2H_6/CH_4 ratio of 0.085. This ratio preserves the total C_2H_6 emissions from the "Default" map but redistributes them in a way that removes the knowledge of the unique gas compositions of different basins, and is referred to henceforth as the "Flat Ratio" inventory. The second alternative map is based on O&G C_2H_6 emissions provided by the NEI2017 inventory. In addition to having a slightly different spatial distribution compared to the Default inventory, total emissions from the O&G sector in the NEI2017 inventory are roughly half of the comparable emissions from the Default inventory. C_2H_6 emissions from the Emissions Database for Global Atmospheric Research (EDGAR) v4.3.2 were originally considered as well for this analysis, but total emissions were 8 times lower than our Default inventory and were decided to be too inaccurate to serve as a useful prior (Huang et al., 2017). For each of the three inventories used, the prior is regridded to a 27 km^2 field to match the mapping of the influence functions and are solved for at that resolution. All fluxes in both the prior and posterior are treated as a constant value for the season they are being evaluated. All three priors and their respective posterior solutions can be seen in Figure 4.

Approximately 200 h of continuous boundary layer C_2H_6 measurements from the ACT-America campaign were available and used as observational input for this study. For this study, we restrict the domain of our inversion to the area within the four corners (23.7°N, 110.7°W), (23.0°N, 77.5°W), (49.9°N, 67.3°W), and (51.1°N, 119.0°W), (domain shown as the colored region in Figure 2). Because the influence functions only provide a local enhancement inside the study domain, for each flight date a background value is determined to represent the C_2H_6 mixing ratios entering the domain. This value is chosen by taking the fifth percentile of the observed boundary layer C_2H_6 mixing ratios on a given flight and subtracting it from the observations, producing an observed C_2H_6 enhancement. The fifth percentile of model-projected enhancements along the flight track is then added onto the observed enhancements in order to align the modeled and observed background values. This final step is necessary for rare scenarios where modeled O&G enhancements are influencing the entire aircraft transect, thus impacting the observed background mixing ratios (see Barkley, Davis, et al., 2019 for further details). In calculating the modeled C_2H_6 enhancements for this study, we treat C_2H_6 as an inert gas rather than a reactive one due to its long average lifetime (weeks to months) relative to the length of time the local plumes travel from the source to the aircraft (a few hours to 1–2 days). For more information on the possible influence of C_2H_6 loss rates, see Section S2.

To run an inversion, values must be assigned to the **R** and **B** matrices related to the uncertainty in the observation/transport and prior flux fields respectively. For the observational/transport uncertainty matrix **R**, we use a method similar to the residual error method discussed in Sheng et al. (2017). For each flight, modeled

enhancements are first scaled by a constant to have the same mean as the mean observed enhancement. This step is performed to remove any existing mean bias that may exist in the prior inventory for the calculation of **R**. After removing this bias, the residual errors are calculated between the model and observations and the standard deviation of this error is used to represent the **R** value along its diagonal for a given flight, with no value assigned for the off-diagonals. This process results in flights with large observational and transport uncertainty on days with large enhancements and poor correlations between the observed and modeled values, thus giving these flights less weight in the inversion solution. Similarly, flights where observed and modeled plume structures align have a smaller **R** value assigned and thus are given greater weight in the overall solution. Values for the diagonal elements of **R** varied across flights (Figure S3), but seasonal averages for the standard deviation of the error ranged from 0.7 ppb in the spring to 1.8 ppb in the fall. This method for classifying transport uncertainty is particularly effective for a C_2H_6 inversion study since the general locations of the sources (i.e., O&G infrastructure) are known with high confidence such that misaligned plumes would most likely be caused by errors in the transport rather than problems with the spatial mapping of the flux.

For the flux uncertainty matrix **B**, there lacks a clear answer on what the uncertainty of the prior fluxes should be. Since the primary source of C_2H_6 in the US is from the O&G production sector, the location of C_2H_6 emitters in the US should be accurate. However, leak rates from O&G activity are beyond the uncertainty bounds of the EPA 2012 Gridded CH_4 Inventory (Alvarez et al., 2018; Maasakkers et al., 2016). For this reason, a value for the flux uncertainty **B** is selected based on mathematical constraints of the inversion rather than an understanding of the prior flux map. Specifically, all flux grids are assigned an error as a percentage of their prior, where the percent uncertainty is selected using a chi-square metric, defined as

$$\chi^2 = \frac{1}{m}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x_0})^T (\mathbf{HBH'} + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x_0}) \tag{3}$$

where m is the number of observations and χ^2 is the chi-square metric used to assess whether the inversion errors satisfy a Gaussian distribution. Here, we select a percent error for the flux uncertainty in **B** that brings χ^2 close to 1. For the Default, Flat Ratio, and NEI2017 inventory, the assigned flux errors along the diagonal of **B** were 50, 50, and 80% of the total grid emissions, respectively. Additionally, a correlation length with an e-folding decay length of 50 km is assigned to the off-diagonal elements in **B**. This added correlation provides consistent shifts in the posterior emissions within a basin while allowing changes across basins to behave uniquely. Implications of the **R** and **B** matrices selected for this study are explored further in a sensitivity analysis. For sources unrelated to the O&G sector, no flux uncertainty was applied, such that all changes in the posterior are reflective of changes to O&G emissions.

To test the robustness of the inversion results, a sensitivity analysis is performed for each prior inventory, adjusting various parameters of the inversion set up to examine how these changes impact the posterior solution. Among the tests performed were sensitivity to errors in the model boundary layer depth and wind speed, sensitivity to a low and high biased prior, sensitivity to the background choice, sensitivity to the correlation length scale used in the inversion, sensitivity to the chosen transport uncertainty matrix **R**, and sensitivity to potential errors in the non-O&G sources. Details on how each of these tests was performed can be found in Section S3.

3. Results and Discussion

3.1. C_2H_6 Inversion Results

For the seasonal inversion results, we focus specifically on changes to grid cells that had at least two flights with overlapping influence functions in a season (area shown in Figure 3). For each season, posterior C_2H_6 flux maps were successful in reducing both the mean absolute error and bias and increasing the correlation between the observed and modeled signal for that season (Table 1). In the fall, winter, and spring campaigns there was an overall positive bias between the modeled prior and observed C_2H_6 enhancements (observed enhancements larger than modeled enhancement), resulting in seasonal posterior maps that generally increased C_2H_6 emissions in order to compensate (Figure 3). Regionally, observed C_2H_6 plumes were consistently underestimated in the western Appalachian region during the fall, winter, and summer
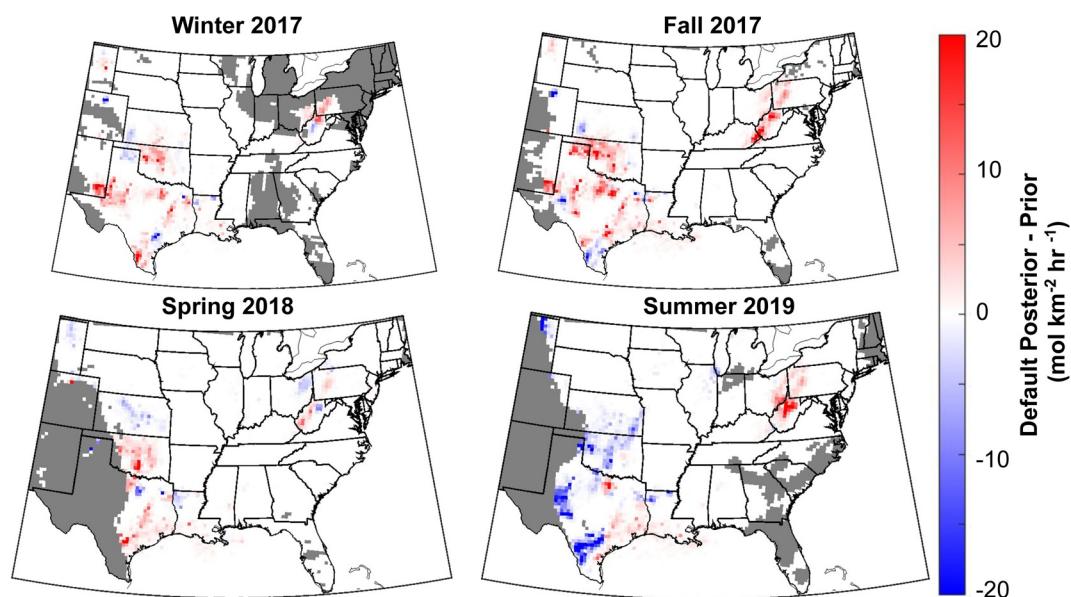
Figure 3. Change in emissions between the Default posterior and prior C_2H_6 flux map for each individual season. Areas with less than two flights that captured their influence in a season (see Figure 2) are greyed out.

season, and posterior C_2H_6 emissions were increased by 50%–110% more than the prior to correct for the underestimation. In the southcentral US, a similar negative bias was observed in the winter, fall, and spring campaigns. This led to posterior solutions for these three seasons that show a consistent increase to the total posterior flux between 30% and 50%. The increase in the southcentral region is largest and most con-

Table 1
Table Describing the Performance of Each of the Seasonal Posteriors Relative to the Observations From That Season

	Default Inventory Winter 2017 Performance		Default Inventory Fall 2017 Performance		Default Inventory Spring 2018 Performance		Default Inventory Summer 2019 Performance	
	Prior	**Posterior**	**Prior**	**Posterior**	**Prior**	**Posterior**	**Prior**	**Posterior**
Southcentral O&G C_2H_6 (mol s^{-1})	917 (100%)	1246	905 (98%)	1375	413 (45%)	574	681 (74%)	239
Western Appalachia O&G C_2H_6 (mol s^{-1})	88 (70%)	131	124 (99%)	262	125 (100%)	128	125 (100%)	265
Mean Absolute Error (ppb)	1.25	0.98	1.49	1.14	0.64	0.55	1.05	0.69
Mean Bias (ppb)	−0.75	−0.33	−1.12	−0.53	−0.19	−0.15	0.22	0.05
y, Hx Correlation	0.82	0.89	0.74	0.82	0.53	0.65	0.37	0.71

Note. Definitions for the regions encompassed in "Southcentral" and "Western Appalachia" can be seen in boxes G and H of Figure 2. Seasonal total emissions calculations for the Southcentral and Appalachia totals only include emissions from grid cells that had at least 2 flights that captured their influence in a season (shown in Figure 3). The percentage of regional emissions captured in each season is given in parenthesis under the prior column for each region. The total prior C_2H_6 emissions contained in the entire Southcentral and Western Appalachia regions are 917 and 125 mol s^{-1}, respectively.

Table 2
Table Describing the Performance of Each of the Individual Seasonal Posterior Fluxes and the Three Season Combined Posterior When Applied to all Four Seasons of Observations

	Default Inventory Prior	Default Inventory Winter 2017 Posterior	Default Inventory Fall 2017 Posterior	Default Inventory Spring 2018 Posterior	Default Inventory Summer 2019 Posterior	Default Inventory 3 Season Posterior
Four Season Mean Absolute Error (ppb)	1.13	1.07	1.05	1.09	**1.2**	1.01
Four Season Mean Bias (ppb)	−0.44	−0.17	0.07	−0.36	**−0.75**	−0.14
Four Season y,Hx Correlation	0.67	0.71	0.76	0.67	**0.46**	0.76
Percent of Flights with Reduced Mean Absolute Error	N/A	65	59	64	**45**	77

Note. Green areas highlight statistical improvement compared to the prior, whereas red boxes show degradation.

sistent in the western Anadarko region and northwestern Permian basin for seasons with coverage in those regions. A more detailed breakdown of the seasonal posterior performance and emissions can be found in Figures S4–S11.

Observations from the southcentral and midwestern US during the Summer 2019 campaign behave like outliers compared to the rest of the data set. Of the 87 flights used in the inversion analysis, only 12 had an overall negative bias when comparing the observed C_2H_6 to the model prior (observed enhancement less than modelled enhancement), eight of which occurred in the southcentral and midwestern portion of the Summer 2019 campaign. These low observations have a profound effect on the posterior emissions in the southcentral US for the Summer 2019 campaign, with total emissions decreasing by 65% from the prior compared to the 30% or greater increase in all other seasonal campaigns. To demonstrate how unrepresentative the Summer 2019 results are compared to the rest of the data set, we take the posterior inventory derived for each season and apply it to all entire four seasons of observations (Table 2). In doing this, we find that the winter, fall, and spring posteriors all produce similar statistical improvements to the overall data set compared to the prior. However, when applying the summer C_2H_6 posterior to the four season data set, not only does it perform substantially worse than each of the other seasonal posteriors, it also performs worse than the prior, with increases to the absolute error, bias, and a substantial decrease in the model-obs correlation coefficient from 0.67 to 0.46.

Differences between observed and modeled C_2H_6 enhancements can occur for various reasons, including both errors in the prior emissions as well as errors in the imperfect model transport used to create the influence functions. One possible explanation for the discrepancy in summer could be related to errors in the model transport simulation. Daily errors in the modeled boundary layer wind direction compared to flight observations were largest during the Summer 2019 campaign, likely related to the slower wind speeds observed during summer (Tables S3–S6). These directional transport errors can create misalignments in observed versus modeled plumes, leading to an overall reduction in the posterior emissions relative to the truth. Compounding on these wind issues is the possibility of increased convective activity in summer. While boundary layer statistics are tracked in the model and compared to observations, it is more difficult to assess how much of the signal is lost out the top of the boundary layer due to summertime convective processes. This possibility is supported by the fact that the gradient between the free troposphere and boundary layer C_2H_6 mixing ratios was smallest during the Summer 2019 campaign (Figure S2), and dates with the largest positive model bias occurred on days with southerlies bringing hot and humid air from the Gulf of Mexico into the flight path. Another, nontransport-related possibility is that there could be a non-trivial loss of C_2H_6 due to an increase in the OH chemical sink during the summer months. In this study, the impacts of C_2H_6 loss were not considered, as C_2H_6 has an average lifetime of two months and most plumes were captured within 48 h of release from the source (Burkholder et al., 2015; Thompson et al., 2003). However,
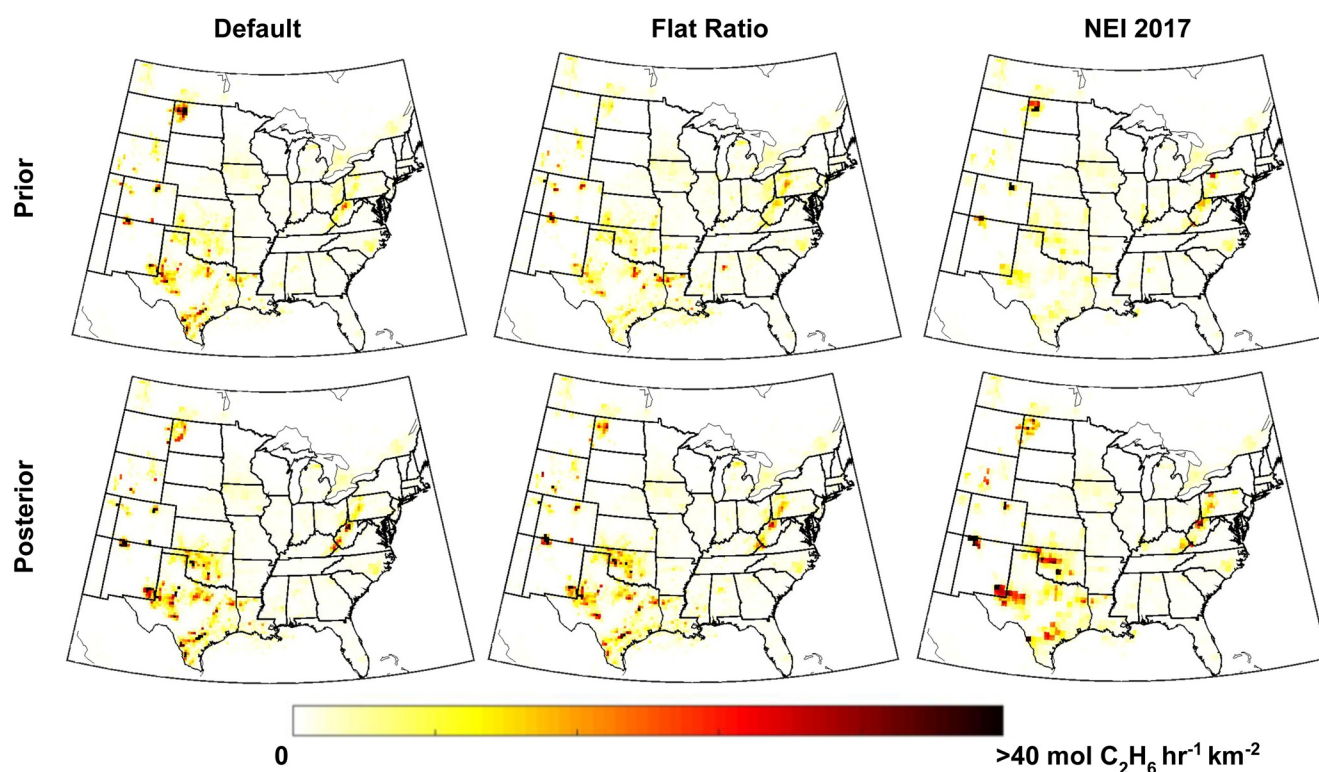
Figure 4. Prior C_2H_6 inventories used in this study and their respective three season mean posteriors. "Default" represents the best guess prior from this study based on multiplying the O&G sector of the Gridded EPA 2012 CH_4 Emissions Inventory by best-available C_2H_6/CH_4 ratios of emissions in individual basins. "Flat Ratio" multiplies the same CH_4 inventory by a uniform rate C_2H_6/CH_4 ratio of 0.085, producing a similar total as "Default" with a different spatial representation. "NEI 2017" comes from the NEI 2017 C_2H_6 inventory.

in conditions with excessive heat and large OH mixing ratios that can occur in the summer, the lifetime of C_2H_6 can be reduced to as little as 4 days in the most extreme conditions (see Section S2 for more details). Even so, a lifetime of 4 days would only have minor impacts on local plumes, and the extreme conditions leading to high C_2H_6 loss would only exist for a short period in the afternoon hours. Furthermore, summer C_2H_6 observations from the northeast were elevated to levels similar to those observed during the other seasons, seemingly unaffected by summertime conditions. For these reasons, it is unlikely that a chemical loss could be solely responsible for the low values observed in the southcentral US. One final possibility is that the low C_2H_6 enhancements observed in the Summer 2019 campaign are due to a real and significant temporal change in C_2H_6 emissions in the southcentral US during this period. However, the near-zero emission rate solution provided by the inversion posterior in the various southcentral plays lacks any sort of real-world explanation for such a large shift compared to previous seasons. Production data shows the Texas and Oklahoma were at peak O&G production rates during the period and well counts were stable (US Energy Information Administration, 2020a). Regardless of the reason, the low C_2H_6 observations measured in the southcentral US during the Summer 2019 months are not representative of data from the winter, fall, and spring campaigns, which show better consistency in the location and expected magnitudes of regional C_2H_6 enhancements. For this work, we choose to discard summer data in evaluating our best estimate of US C_2H_6 emissions but consider it a source of uncertainty and intrigue for future research.

To create our best guess regarding C_2H_6 emissions across the southcentral and eastern US, we combine the observations from the winter, fall, and spring campaigns and rerun the inversion using the Default prior to coming up with a posterior solution that best describes data from all three seasons. In combining the seasonal data sets, influence functions for available flights provide coverage over the entire southcentral US and the western Appalachian region, allowing the inversion to solve for posterior fluxes for all grid cells in each region. This combined posterior solution provides consistent improvement compared to the prior, reducing the absolute error between the model and observations in 86% of flights from those three seasons.
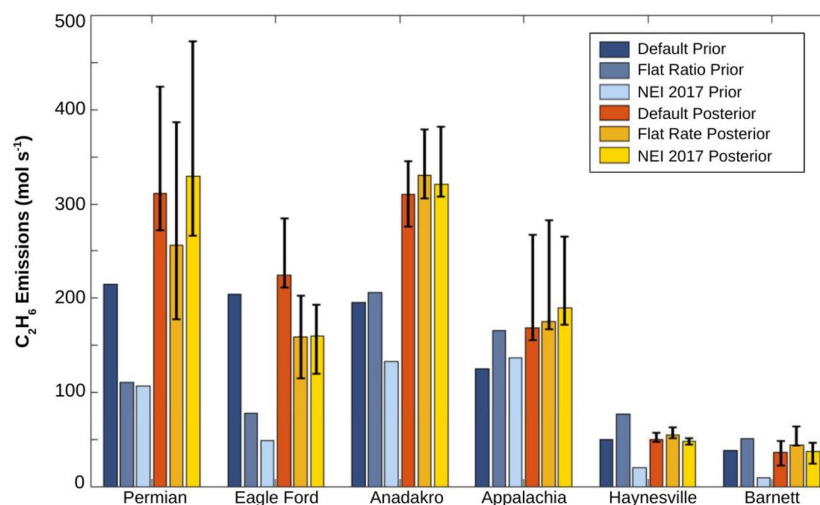
Figure 5. Priors and their respective 3 season posterior C_2H_6 emissions for major basins and plays in this study. Error bars on the posteriors show the maximum and minimum values derived for each region from the sensitivity analysis (Tables S7–S12).

C_2H_6 emissions from the 3 season posterior are almost universally increased compared to the prior, with a 42% increase overall in the southcentral US and a 34% increase in the Appalachia. With these large increases in the emissions, the mean model-obs bias is improved from −0.80 to −0.41 ppb. The inability to eliminate this bias is a natural result of errors in the transport preventing the inversion from solving for misaligned observed plumes, and thus the emission increases in the 3 season posterior are likely still an underestimation of the true emissions.

To better understand the sensitivity of our three season posterior to the prior, the inversion was rerun using the two alternative prior maps discussed in Section 2.2 and a three season posterior was created from each set of priors (Figure 4) For every basin and play, posteriors from the three inventories converged towards a similar solution compared to their priors, despite starting from very different initial states. An example of this is the Haynesville basin, where the two alternative prior inventories have a factor of 4 spread between their emissions, but their posteriors converge to within 15% of the Default posterior solution. Ultimately, the two alternative posteriors produced solutions for the entire southcentral US that were within 45 mol/s of each other, despite starting 375 mol/s apart. In the northeast, discrepancies between the three inventory priors were much smaller, and all three posteriors converged to similar values (168–190 mol/s). Overall the NEI2017 prior, whose total C_2H_6 emissions are 88% less than the Default prior, leads to a total posterior C_2H_6 emissions estimate that is within 10% of the Default posterior.

Though the posteriors generated from the Default and NEI2017 prior inventories converge to a similar emissions map, the interpretation of the values they converge to are vastly different. The NEI2017 prior is an official product provided by the EPA and specifically provides C_2H_6 emissions for the year 2017, whereas the Default inventory from this study was created using CH_4 emissions data from O&G sources for the year 2012 combined with knowledge of C_2H_6/CH_4 gas composition data of different basins. Despite both priors attempting to represent the same C_2H_6 emissions, their representations of the southcentral US in particular are starkly different. The posterior solution of 1,313 mol/s from the Default posterior is a 43% increase compared to its prior of 917 mol/s, but the posterior solution of 1,176 mol/s from the NEI2017 posterior is a 182% increase compared to its prior of 417 mol/s. The convergence of solutions provides confidence in the ability of the ACT data set data set to constrain C_2H_6 emissions but reveals a large discrepancy between the NEI2017 C_2H_6 product and atmospheric data which points towards significantly larger emissions, particularly in the southcentral US. Given that the Default inventory prior from this study is based on EPA (2012) O&G CH_4 emissions data but contains more than double the O&G C_2H_6 emissions compared the NEI2017 prior, the low bias in the NEI2017 C_2H_6 inventory may be partially explained by some methodological flaw in its creation that underestimates C_2H_6 emissions from documented O&G sources rather than some unknown or missing sources.
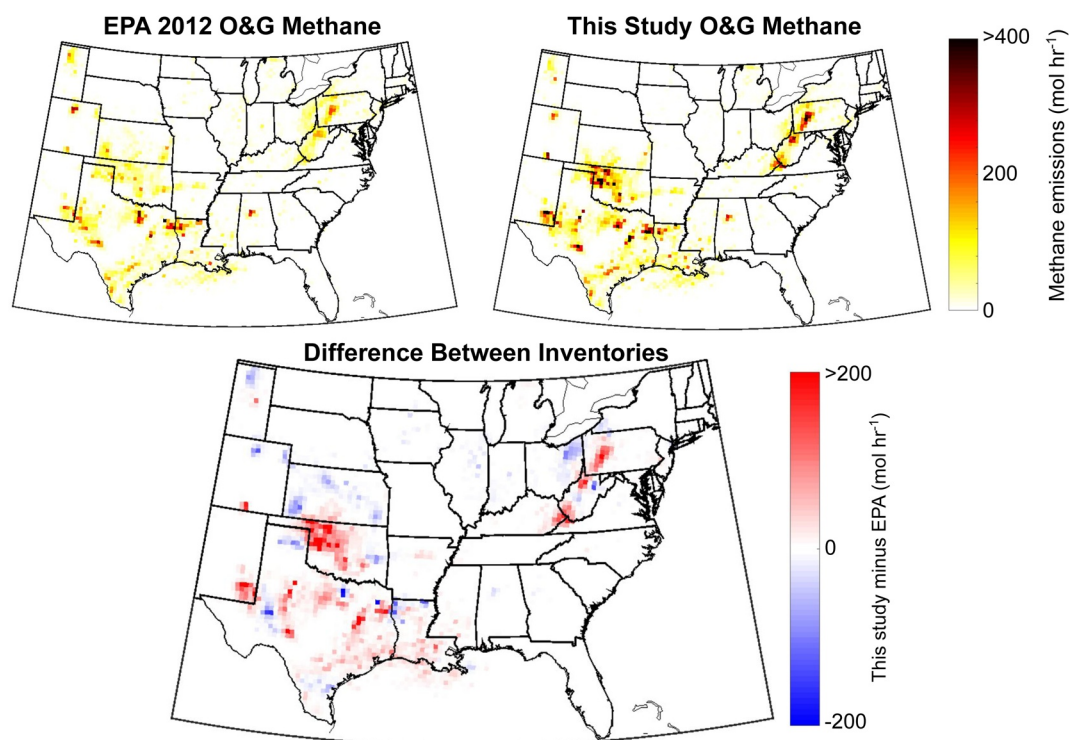
Figure 6. Top-left: Oil and gas CH_4 emissions from the EPA (2012) Gridded CH_4 Inventory. Top-right: Oil and gas CH_4 emissions estimated from the C_2H_6 posterior in this study. Bottom: Difference between the two inventories.

In addition to testing the sensitivity of the inversion posterior to different priors, we also test its sensitivity to numerous other conditions, including adjustments to model meteorology errors, adjustments to the magnitude of the prior, changes to the selection of the background C_2H_6 term, elimination of the correlation lengths applied to the prior flux uncertainty, adjustment of the observation error covariance matrix term to a constant (giving equal uncertainty to all observations), and allowing for the adjustment of all C_2H_6 sources rather than only O&G sources. Details and tables with values from this sensitivity test can be found in Section S3 and Tables S7–S12, and the range of solutions is shown in Figure 5. Most posteriors produce similar results regardless of the methodology, providing further confidence in the solution. One particular concern was the inherent low bias in all of the prior inventories compared to the posterior solutions. However, multiplying the Default prior by 3 (making it much larger than the original posterior solution) and rerunning the inversion resulted in a total posterior solution that was only 2% higher than the original posterior, reducing concerns that using a prior with a low bias could be restricting the posterior solution. Of all the sensitivity tests, the one that had the largest impact for all three priors was adjusting the observation error covariance matrix **R**. Based on its original setup, larger uncertainty is given on days with larger C_2H_6 plumes, as any transport error on these days could potentially create larger mismatches between modeled and observed enhancements. This leads to less weight being given to days with the biggest signals, particularly in the fall and winter seasons. By assigning a constant value to **R** and giving equal weight to all flights, the overall Default posterior emissions increase by 21% compared to the original posterior, or 72% compared to the prior.

One final source of uncertainty that is difficult to address in our sensitivity tests are errors in the transport used to create the influence functions. Observed minus modelled boundary layer wind speed and direction biases were small during the winter, fall, and spring aircraft campaigns. However, modelled boundary layer heights were shallow compared to observations, particularly during the spring campaign where observed boundary layer heights were on average 46% higher than modeled heights. A negative bias in the modeled boundary layer would result in a proportionally higher value in the influence functions and model-projected enhancements. As a way of testing these impacts, we rerun the inversion applying a unique correction

factor to the influence functions on each day to account for biases in the modeled versus observed wind speed and boundary layer heights from each flight day. Doing this correction produces an overall Default posterior that is 16% higher than the original Default posterior. Consistent with our past work that has applied corrections to simulated errors in ABL depth and winds (Barkley, Davis, et al., 2019; Barkley, Lauvaux, et al., 2019), we consider the posterior created using the ABL depth correction to be an equally plausible solution in our best estimate of the C_2H_6 emissions, as the boundary layer bias is a source of error with a known and somewhat correctable bias on the solution. The boundary layer bias, along with a negative enhancement bias still present in the model versus observational comparison of the posterior C_2H_6 enhancements, is both potential reasons to suspect the Default posterior on its own may still be underestimating overall C_2H_6 emissions despite the large increase in the emissions relative to the prior.

3.2. Interpretation of CH_4 Emissions From the O&G Sector

The Default prior inventory developed in this study is created by multiplying the EPA (2012) Gridded CH_4 Emissions Inventory for O&G sources by the assumed mean C_2H_6/CH_4 ratio of each grid's emissions. If the applied ratios are correct, and emissions primarily occur in situations where the gas content is unaltered (i.e., processes unrelated to gas separation), then changes between the posterior and prior Default C_2H_6 inventory should proportionally reflect changes in the EPA (2012) CH_4 inventory for O&G. Thus, we can use our Default posterior C_2H_6 inventory to create our best interpretation of O&G CH_4 emissions in the southcentral and eastern US.

Using the converted C_2H_6 posterior as a proxy for O&G CH_4 emissions, our inventory projects that O&G CH_4 emissions are almost universally larger than the 2012 EPA Inventory estimates (Figure 6). In the southcentral US, we estimate emissions to be 48% higher than inventory estimates (77% using the meteo-corrected posterior). Of this increase, more than half of it is driven by increases in the Anadarko and Permian plays, both of which had the largest proportional changes in the C_2H_6 posterior and are some of the largest sources of CH_4 emissions in the southcentral US. Of the remaining southcentral basins captured in this study, the Haynesville and Barnett basins are the only ones that did not see an increase in their emissions relative to the prior.

In the Appalachians, we estimate CH_4 emissions to be 32% higher than EPA (2012) inventory estimates (84% using the meteo-adjusted posterior). Part of the discrepancy between inventory results and the posterior may be related to the increased presence of unconventional natural gas activity in the Marcellus shale. Between 2012 and 2018 Pennsylvania and West Virginia underwent some of the largest gas production growth in the US, with annual production tripling during the period (US Energy Information Administration, 2020a), a change that would not be captured in the EPA (2012) inventory. However, Pennsylvania state inventories, which provide annual inventory estimates of unconventional natural gas activity in the state using methodologies similar to the EPA, show CH_4 emissions from unconventional activity only increased by 20% during this period, and that these unconventional wells only represent a small fraction of O&G emissions in the region (Omara et al., 2016), with much of the emissions coming from pre-existing conventional activity. Thus, it is unlikely that changes in unconventional activity between 2012 and the time of this study would be responsible for the increase in regional emission rates found from the inversion results relative to the EPA (2012) Gridded Inventory, and that the discrepancy would still be present in an updated EPA inventory.

The three-season CH_4 inventory estimates for individual basins from this study generally align with top-down estimates of corresponding basins from the NOAA SONGNEX campaign (Peischl et al., 2018) and other studies. In the Haynesville basin, we calculate an O&G emission rate in our Default posterior ensemble of 50–59 Mg/hr, compared to 42 Mg/hr from Peischl et al. (2018) (year of study: 2015) and 76 Mg/hr from Cui et al. (2019) (year of study: 2013), which includes non-O&G values in its total as well. In the Barnett, we calculate emissions to be 29–42 Mg/hr, close to the 46 Mg/hr found in Peischl et al. (2018) (year of study: 2015) but less than the 60 Mg/hr found in Karion et al. (2015) (year of study: 2013). In Eagle Ford, both the western and eastern basin in this study had a combined emission rate of 54–72 Mg/hr versus 83 Mg/hr in Peischl et al. (2018) (year of study: 2015). In the Appalachian, our results support findings that show emissions from unconventional O&G infrastructure in the Marcellus are greatly underestimated by

EPA inventory estimates (Barkley, Lauvaux et al., 2019; Caulton et al., 2019). One area where we see strong disagreement is in the Permian, where this study's emissions estimate of 117–182 Mg/hr is substantially less than top-down estimates from Zhang et al. (2020) of 308 Mg/hr (years of study: 2018–2019) and is much closer to their updated bottom-up estimate of 114 Mg/hr. Overall, this study finds emissions from the southcentral and eastern US to be 48%–76% greater than the EPA 2012 Gridded Methane Inventory for O&G sources, agreeing with national estimates from Alvarez et al. (2018). However, in all of these comparisons, we note that the year each study took place varies, such that differences in the solution may include changes due to temporal variability in the O&G activity of a basin. The comparison of emissions in the Permian is an example of where one could expect large discrepancies. Analysis of the Permian play in this study is mostly representative of the year 2017 due to a lack of coverage from influence functions during the Spring 2018 campaign, whereas in Zhang et al. (2020) the analysis is from 2018 to 2019. Between 2017 and 2019 the Permian was undergoing rapid development, and both O&G production doubled in the 2 year period (US Energy Information Administration, 2021). This change may be partially responsible for the discrepancy between emission estimates in the Permian from this work and Zhang et al. (2020).

The interpretation of O&G CH_4 emissions using C_2H_6 observations has a unique advantage compared to more traditional methodologies that rely on CH_4 measurements due to the simplicity of C_2H_6 sources. In the US where CH_4 emissions have near equal contributions from fossil fuels, agriculture, and wetlands, each of which has their own uncertainties, C_2H_6 emissions are dominated by the O&G sector. Furthermore, there is high confidence in the spatial mapping of O&G sources in the US due to extensive documentation of the various components associated with O&G extraction, simplifying the interpretation of atmospheric C_2H_6 data. As an example, in the ACT-America campaign, the majority of aircraft flight tracks were hundreds to thousands of kilometers away from O&G basins. Despite this, the model prior was able to consistently track C_2H_6 enhancements from these sources in the winter, fall, and spring, with correlations between the model versus observed boundary layer C_2H_6 enhancements of 0.82, 0.74, and 0.53 respectively. The high skill in tracking enhancements from a single sector with well-defined locations creates a scenario where a stable posterior solution can be generated through various inverse methodologies (Table S4).

Despite high confidence in the C_2H_6 posterior, the conversion of this posterior to O&G CH_4 emissions is entirely dependent on the quality and availability of information related to the C_2H_6/CH_4 ratio for each basin. During the time observations from this study were collected, there were numerous recent flights available from a separate study that quantified the mean C_2H_6 and CH_4 emissions of various basins (Peischl et al., 2018), providing more confidence in a mean C_2H_6/CH_4 ratio to apply to those locations. However, geographically broad plays, such as the Anadarko and the Appalachia, are more difficult to characterize a mean ratio for using atmospheric data. Furthermore, the average gas composition of a basin can change over time, making ratios found in older studies less applicable (Lan et al., 2019). Additionally, sources such as C_2H_6 cracker plants that do not adhere to any of the assigned C_2H_6/CH_4 ratios could create additional uncertainties in this conversion, but we assume emissions from these facilities are small relative to emissions from O&G production activity and mostly uncorrelated with the locations of specific basins. Publicizing upstream and midstream gas composition data collected by major O&G companies would be one immediate solution to reduce uncertainty introduced in a C_2H_6 to CH_4 inventory conversion.

One other concern arises from the age of the inventory used to perform this C_2H_6 to CH_4 inventory conversion. Comparisons in this study are made relative to the EPA bottom-up estimate from 2012, the most recent year in which gridded emissions data is available, whereas flights for this study took place between 2017 and 2019. A theoretically updated EPA-gridded inventory may show differences compared to the existing 2012 gridded inventory, particularly in basins that have undergone substantial changes to their O&G activity. However, total O&G emissions from the EPA bottom-up inventory from 2012 to 2018 have been stable, with only a small decrease from 6.6 Tg CH_4 in 2012 to 6.4 Tg CH_4 in 2018 (US Environmental Protection Agency, 2020b). Thus, while it can be argued that increases to CH_4 emissions in certain basins observed in this study compared to the 2012 Gridded Methane Inventory may be related to changes in activity since 2012, the 48%–76% increase in overall CH_4 emissions across the southcentral and eastern US relative to the 2012 inventory is not compatible with current EPA bottom-up inventory estimates and indicates that current inventory estimates are undercounting CH_4 emissions from O&G activity.

4. Conclusions

Using the largest collection of airborne boundary layer C_2H_6 data to date, an inversion was performed to estimate C_2H_6 and CH_4 emissions from various O&G basins across the southcentral and eastern US. From this data set, we estimate that a large portion of C_2H_6 emissions cannot be explained using O&G emission data from the EPA (2012) Gridded CH_4 Inventory and existing C_2H_6/CH_4 ratio data. We conclude that the EPA CH_4 emissions inventory underestimates CH_4 emissions overall from O&G sources in the southcentral and Appalachian regions by 48%–76%, an offset similar in magnitude to national estimates calculated in Alvarez et al. (2018). In particular, this study finds the largest discrepancies occurring in the Anadarko, western Appalachian, and Permian O&G plays. Though some of these discrepancies may be related to regional changes in O&G activity since 2012, bottom-up O&G inventory emission estimates nationally have decreased by 4% from 2012 to 2018 and cannot explain the higher overall emissions observed in this study. We also find C_2H_6 emissions from O&G are more than two times higher than the NEI2017 C_2H_6 inventory, a fact that should be taken into account in any future efforts to investigate O&G CH_4 emissions using atmospheric C_2H_6 observations, as well as studies relying on existing C_2H_6 inventories to account for certain chemical reactions.

This study reveals the potential to use broad-scale continuous C_2H_6 data to constrain CH_4 emissions from the O&G sector on a nationwide scale. Our confidence in the spatial distribution of emissions associated with the O&G sector and the dominant role of O&G in US C_2H_6 emissions allows for modeling and interpretation of observed signals without large concerns for source attribution. Furthermore, misalignment in observed versus modeled plumes may serve as a useful diagnostic of model transport errors due to the well-documented spatial knowledge of C_2H_6 emitters in the US. One current weakness with relying on C_2H_6 observations to understand CH_4 emissions from O&G is the requirement of knowing the C_2H_6/CH_4 ratios of various O&G fields in order to convert C_2H_6 emissions into CH_4 emissions. Gas composition data is collected by individual companies but not shared publicly. Should some form of this information be made available for the public, it would greatly enhance the confidence of CH_4 emission estimates of the O&G industry using C_2H_6 and other trace gases.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Aircraft data from the ACT-America campaign used in this study can be found at https://doi.org/10.3334/ORNLDAAC/1556.

References

Aikin, A. C., Herman, J. R., Maier, E. J., & McQuillan, C. J. (1982). Atmospheric chemistry of ethane and ethylene. *Journal of Geophysical Research*, *87*(C4), 3105–3118. https://doi.org/10.1029/JC087iC04p03105

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., et al. (2018). Assessment of methane emissions from the U.S. oil and gas supply chain. *Science*, eaar7204.

Baier, B. C., Sweeney, C., Choi, Y., Davis, K. J., DiGangi, J. P., Feng, S., et al. (2020). Multispecies assessment of factors influencing regional CO_2 and CH_4 enhancements during the winter 2017 ACT-America Campaign. *Journal of Geophysical Research - D: Atmospheres*, *125*(2), e2019JD031339. https://doi.org/10.1029/2019JD031339

Barkley, Z. R., Davis, K. J., Feng, S., Balashov, N., Fried, A., DiGangi, J., et al. (2019). Forward modeling and optimization of methane emissions in the south central United States using aircraft transects across frontal boundaries. *Geophysical Research Letters*, *46*(22), 13564–13573. https://doi.org/10.1029/2019GL084495

Barkley, Z. R., Lauvaux, T., Davis, K. J., Deng, A., Fried, A., Weibring, P., et al. (2019). Estimating methane emissions from underground coal and natural gas production in southwestern Pennsylvania. *Geophysical Research Letters*, *46*(8), 4531–4540. https://doi.org/10.1029/2019GL082131

Barkley, Z. R., Lauvaux, T., Davis, K. J., Deng, A., Miles, N. L., Richardson, S. J., et al. (2017). Quantifying methane emissions from natural gas production in north-eastern Pennsylvania. *Atmospheric Chemistry and Physics*, *17*(22), 13941–13966. Retrieved from http://www.atmos-chem-phys.net/17/13941/2017/

Bloom, A. A., Bowman, K. W., Lee, M., Turner, A. J., Schroeder, R., Worden, J. R., et al. (2017). A global wetland methane emissions and uncertainty dataset for atmospheric chemical transport models (WetCHARTs version 1.0). *Geoscientific Model Development*, *10*(6), 2141–2156. https://doi.org/10.5194/gmd-10-2141-2017

Brasseur, G. P., & Jacob, D. J. (2017). *Modeling of atmospheric chemistry*. Cambridge University Press.

Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., et al. (2013). The Lagrangian particle dispersion model FLEX-PART-WRF version 3.1. *Geoscientific Model Development*, *6*, 1889–1904. https://doi.org/10.5194/gmd-6-1889-2013

Burkholder, J. B., Sander, S. P., Abbatt, J., Barker, J. R., Huie, R. E., Kolb, C. E., & Wine, P. H. (2015). *Chemical Kinetics and photochemical data for Use in atmospheric studies: Evaluation No. 18*. JPL Publication 15-10. Jet Propulsion Laboratory.

Caulton, D. R., Lu, J. M., Lane, H. M., Buchholz, B., Fitts, J. P., Golston, L. M., et al. (2019). Importance of superemitter natural gas well pads in the Marcellus shale. *Environmental Science and Technology*, *53*, 4747–4754. https://doi.org/10.1021/acs.est.8b06965

Cui, Y. Y., Brioude, J., McKeen, S. A., Angevine, W. M., Kim, S.-W., Frost, G. J., et al. (2015). Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The South Coast Air Basin. *Journal of Geophysical Research - D: Atmospheres*, *120*(13), 6698–6711.

Cui, Y. Y., Henze, D. K., Brioude, J., Angevine, W. M., Liu, Z., Bousserez, N., et al. (2019). Inversion estimates of lognormally distributed methane emission rates from the Haynesville-Bossier oil and gas production region using airborne measurements. *Journal of Geophysical Research - D: Atmospheres*, *124*(6), 3520–3531.

Dalsøren, S. B., Myhre, G., Hodnebrog, Ø., Myhre, C. L., Stohl, A., Pisso, I., et al. (2018). Discrepancy between simulated and observed ethane and propane levels explained by underestimated fossil emissions. *Nature Geoscience*, *11*(3), 178–184.

Davis, K., Browell, E., Feng, S., Lauvaux, T., Obland, M., Pal, S., et al. (2021). The atmospheric carbon and transport (ACT) -America Mission. *Earth and Space Science Open Archive*, *54*. https://doi.org/10.1002/essoar.10505721.1

DiGangi, J. P., Choi, Y., Nowak, J. B., Halliday, H., & Yang, M. M. (2018). *ACT-America: L_2 in situ atmospheric CO_2, CO, CH4, and O_3 concentrations, eastern USA*. (ORNL DAAC). https://doi.org/10.3334/ORNLDAAC/1556

Dlugokencky, E. (2021). *Trends in atmospheric methane*. Retrieved from www.esrl.noaa.gov/gmd/ccgg/trends_ch4/

Edwards, P. M., Brown, S. S., Roberts, J. M., Ahmadov, R., Banta, R. M., Degouw, J. A., et al. (2014). High winter ozone pollution from carbonyl photolysis in an oil and gas basin. *Nature*, *514*(7522), 351–354. https://doi.org/10.1038/nature13767

Feng, S., Lauvaux, T., Davis, K. J., Keller, K., Zhou, Y., Williams, C., et al. (2019). Seasonal characteristics of model uncertainties from biogenic fluxes, transport, and large-scale boundary inflow in atmospheric CO_2 simulations over North America. *Journal of Geophysical Research - D: Atmospheres*, *124*(24), 14325–14346. https://doi.org/10.1029/2019JD031165

Gilman, J. B., Lerner, B. M., Kuster, W. C., & de Gouw, J. A. (2013). Source signature of volatile organic compounds from oil and natural gas operations in northeastern Colorado. *Environmental Science and Technology*, *47*(3), 1297–1305. https://doi.org/10.1021/es304119a

Grant, R. F., & Roulet, N. T. (2002). Methane efflux from boreal wetlands: Theory and testing of the ecosystem model ecosys with chamber and tower flux measurements. *Global Biogeochemical Cycles*, *16*(4). https://doi.org/10.1029/2001GB001702

Helmig, D., Rossabi, S., Hueber, J., Tans, P., Montzka, S. A., Masarie, K., et al. (2016). Reversal of global atmospheric ethane and propane trends largely due to US oil and natural gas production. *Nature Geosciences*, *9*(7), 490–495. https://doi.org/10.1038/ngeo2721

Hodnebrog, Ø., Dalsøren, S. B., & Myhre, G. (2018). Lifetimes, direct and indirect radiative forcing, and global warming potentials of ethane (C_2H_6), propane (C_3H_8), and butane (C_4H_{10}). *Atmospheric Science Letters*, *19*(2), e804. https://doi.org/10.1002/asl.804

Huang, G., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., et al. (2017). Speciation of anthropogenic emissions of non-methane volatile organic compounds: A global gridded data set for 1970-2012. *Atmospheric Chemistry and Physics*, *17*(12), 7683–7701. https://doi.org/10.5194/acp-17-7683-2017

Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., et al. (2015). Aircraft-based estimate of total methane emissions from the Barnett shale region. *Environmental Science and Technology*, *49*(13), 8124–8131. https://doi.org/10.1021/acs.est.5b00217

Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., et al. (2013). Methane emissions estimate from airborne measurements over a Western United States natural gas field. *Geophysical Research Letters*, *40*(16), 4393–4397. https://doi.org/10.1002/grl.50811

Kitanidis, P. (1997). *Introduction to geostatistics: Applications in hydrogeology*. Cambridge University Press. https://doi.org/10.1017/cbo9780511626166

Kort, E. A., Smith, M. L., Murray, L. T., Gvakharia, A., Brandt, A. R., Peischl, J., et al. (2016). Fugitive emissions from the Bakken shale illustrate role of shale production in global ethane shift. *Geophysical Research Letters*, *43*(9), 4617–4623. https://doi.org/10.1002/2016GL068703

Kostinek, J., Roiger, A., Davis, K. J., Sweeney, C., DiGangi, J. P., Choi, Y., et al. (2019). Adaptation and performance assessment of a quantum and interband cascade laser spectrometer for simultaneous airborne in situ observation of CH_4, C_2H_6, CO_2, CO and N_2O. *Atmospheric Measurement Techniques*, *12*(3), 1767–1783.

Lan, X., Tans, P., Sweeney, C., Andrews, A., Dlugokencky, E., Schwietzke, S., et al. (2019). Long-term measurements show little evidence for large increases in total U.S. methane emissions over the past decade. *Geophysical Research Letters*, *46*(9), 4991–4999. https://doi.org/10.1029/2018GL081731

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., et al. (2012). Constraining the CO_2 budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system. *Atmospheric Chemistry and Physics*, *12*(1), 337–354.

Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Scarpelli, T. R., Nesser, H., Sheng, J.-X., et al. (2019). Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010-2015. *Atmospheric Chemistry and Physics*, *19*(11), 7859–7881.

Maasakkers, J. D., Jacob, D. J., Sulprizio, M. P., Turner, A. J., Weitz, M., Wirth, T., et al. (2016). Gridded national inventory of U.S. methane emissions. *Environmental Science and Technology*, *50*(23), 13123–13133. https://doi.org/10.1021/acs.est.6b02878

Matthes, J. H., Sturtevant, C., Verfaillie, J., Knox, S., & Baldocchi, D. (2014). Parsing the variability in of CH_4 flux at a spatially heterogeneous wetland: Integrating multiple eddy covariance towers with high-resolution flux footprint analysis. *Journal of Geophysical Research: Biogeosciences*, *119*(7), 1322–1339. https://doi.org/10.1002/2014JG002642

McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., et al. (2015). Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 1941–1946. https://doi.org/10.1073/pnas.1416261112

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., et al. (2013). Anthropogenic emissions of methane in the United States. *Proceedings of the National Academy of Sciences*, *110*(50), 20018–20022. https://doi.org/10.1073/pnas.1314392110

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., & Zhang, H. (2013). Anthropogenic and natural radiative forcing [Book Section], In T. Stocker, et al., (Eds.) 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 (pp. 659–740). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

Nisbet, E. G., Manning, M. R., Dlugokencky, E. J., Fisher, R. E., Lowry, D., Michel, S. E., et al. (2019). Very strong atmospheric methane growth in the 4 years 2014-2017: Implications for the Paris agreement. *Global Biogeochemical Cycles*, *33*(3), 318–342.

Omara, M., Sullivan, M. R., Li, X., Subramanian, R., Robinson, A. L., & Presto, A. A. (2016). Methane emissions from conventional and unconventional natural gas production sites in the Marcellus shale basin. *Environmental Science and Technology*, *50*(4), 2099–2107.

Pal, S., Davis, K. J., Lauvaux, T., Browell, E. V., Gaudet, B. J., Stauffer, D. R., et al. (2020). Observations of greenhouse gas changes across summer frontal boundaries in the eastern United States. *Journal of Geophysical Research - D: Atmospheres*, *125*(5), e2019JD030526. https://doi.org/10.1029/2019JD030526

Peischl, J., Eilerman, S. J., Neuman, J. A., Aikin, K. C., de Gouw, J., Gilman, J. B., et al. (2018). Quantifying methane and ethane emissions to the atmosphere from central and western U.S. oil and natural gas production regions. *Journal of Geophysical Research - D: Atmospheres*, *123*(14), 7725–7740. https://doi.org/10.1029/2018JD028622

Peischl, J., Karion, A., Sweeney, C., Kort, E. A., Smith, M. L., Brandt, A. R., et al. (2016). Quantifying atmospheric methane emissions from oil and natural gas production in the Bakken shale region of North Dakota. *Journal of Geophysical Research - D: Atmospheres*, *121*(10), 6101–6111. https://doi.org/10.1002/2015JD024631

Peischl, J., Ryerson, T. B., Aikin, K. C., Gouw, J. A., Gilman, J. B., Holloway, J. S., et al. (2015). Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions. *Journal of Geophysical Research - D: Atmospheres*, *120*(5), 2119–2139. https://doi.org/10.1002/2014JD022697

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., et al. (2012). Hydrocarbon emissions characterization in the colorado front range: A pilot study. *Journal of Geophysical Research*, *117*(D4). https://doi.org/10.1029/2011JD016360

Pétron, G., Karion, A., Sweeney, C., Miller, B. R., Montzka, S. A., Frost, G. J., et al. (2014). A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. *Journal of Geophysical Research - D: Atmospheres*, *119*(11), 6836–6852. https://doi.org/10.1002/2013JD021272

Pétron, G., Miller, B., Vaughn, B., Thorley, E., Kofler, J., Mielke-Maday, I., et al. (2020). Investigating large methane enhancements in the U.S. San Juan Basin. *Elementa: Science of the Anthropocene*, *8*(1). https://doi.org/10.1525/elementa.038

Plant, G., Kort, E. A., Floerchinger, C., Gvakharia, A., Vimont, I., & Sweeney, C. (2019). Large fugitive methane emissions from urban centers along the U.S. east coast. *Geophysical Research Letters*, *46*(14), 8500–8507. https://doi.org/10.1029/2019GL082635

Rella, C. W., Tsai, T. R., Botkin, C. G., Crosson, E. R., & Steele, D. (2015). Measuring emissions from oil and natural gas well pads using the mobile flux plane technique. *Environmental Science and Technology*, *49*(7), 4742–4748. https://doi.org/10.1021/acs.est.5b00099

Robertson, A. M., Edie, R., Snare, D., Soltis, J., Field, R. A., Burkhart, M. D., et al. (2017). Variation in methane emission rates from well pads in four oil and gas basins with contrasting production volumes and compositions. *Environmental Science and Technology*, *51*(15), 8832–8840. https://doi.org/10.1021/acs.est.7b00571

Roscioli, J. R., Yacovitch, T. I., Floerchinger, C., Mitchell, A. L., Tkacik, D. S., Subramanian, R., et al. (2015). Measurements of methane emissions from natural gas gathering facilities and processing plants: Measurement methods. *Atmospheric Measurement Techniques*, *8*(5), 2017–2035.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., & Zhuang, Q. (2020). The global methane budget 2000–2017. *Earth System Science Data*, *12*(3), 1561–1623.

Sheng, J.-X., Jacob, D. J., Maasakkers, J. D., Sulprizio, M. P., Zavala-Araiza, D., & Hamburg, S. P. (2017). A high-resolution (0.1° × 0.1°) inventory of methane emissions from Canadian and Mexican oil and gas systems. *Atmospheric Environment*, *158*, 211–215. https://doi.org/10.1016/j.atmosenv.2017.02.036

Sheng, J.-X., Jacob, D. J., Turner, A. J., Maasakkers, J. D., Sulprizio, M. P., Bloom, A. A., et al. (2018). High-resolution inversion of methane emissions in the Southeast US using SEAC4RS aircraft observations of atmospheric methane: Anthropogenic and wetland sources. *Atmospheric Chemistry and Physics*, *18*(9), 6483–6491.

Thompson, A., Rudolph, J., Rohrer, F., & Stein, O. (2003). Concentration and stable carbon isotopic composition of ethane and benzene using a global three-dimensional isotope inclusive chemical tracer model. *Journal of Geophysical Research*, *108*(D13). https://doi.org/10.1029/2002JD002883

Tzompa-Sosa, Z. A., Mahieu, E., Franco, B., Keller, C. A., Turner, A. J., Helmig, D., & Fischer, E. V. (2017). Revisiting global fossil fuel and biofuel emissions of ethane. *Journal of Geophysical Research: Atmospheres*, *122*(4), 2493–2512.

US Energy Information Administration. (2020a). *Natural gas gross withdraws and production*. Retrieved from https://www.eia.gov/dnav/ng/ng_prod_sum_a_EPG0_VGM_mmcf_a.htm

US Energy Information Administration. (2020b). *United States energy information administration homepage*. Retrieved from https://www.eia.gov/

US Energy Information Administration. (2021). *Drilling productivity report: Permian region*. Retrieved from https://www.eia.gov/petroleum/drilling/pdf/permian.pdf

US Environmental Protection Agency. (2020a). *2017 national emissions inventory data*. Retrieved from https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei

US Environmental Protection Agency. (2020b). *Inventory of U.S. Greenhouse gas emissions and sinks: 1990-2018*. Retrieved from https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks

US Geological Survey. (2018). *Geochemistry laboratory database*. Retrieved from https://certmapper.cr.usgs.gov/data/apps/geochem-db/

Weibring, P., Richter, D., Walega, J. G., Fried, A., DiGangi, J., Halliday, H., et al. (2020). Autonomous airborne mid-ir spectrometer for high precision measurements of ethane during the NASA ACT-America studies. *Atmospheric Measurement Techniques Discussions*, *2020*, 1–42.

Xiao, Y., Logan, J. A., Jacob, D. J., Hudman, R. C., Yantosca, R., & Blake, D. R. (2008). Global budget of ethane and regional constraints on U.S. sources. *Journal of Geophysical Research*, *113*(D21). https://doi.org/10.1029/2007JD009415

Yacovitch, T. I., Herndon, S. C., Roscioli, J. R., Floerchinger, C., Mcgovern, R. M., Agnese, M., et al. (2014). Demonstration of an ethane spectrometer for methane source identification. *Environmental Science and Technology*, *48*, 8028–8034. https://doi.org/10.1021/es501475q

Yu, X., Millet, D. B., Wells, K. C., Henze, D. K., Cao, H., Griffis, T. J., et al. (2020). Aircraft-based inversions quantify the importance of wetlands and livestock for upper midwest methane emissions. *Atmospheric Chemistry and Physics Discussions*, *2020*, 1–34.

Yvon-Durocher, G., Allen, A. P., Bastviken, D., Conrad, R., Gudasz, C., St-Pierre, A., et al. (2014). Methane fluxes show consistent temperature dependence across microbial to ecosystem scales. *Nature*, *507*, 488–491. https://doi.org/10.1038/w1316410.1038/nature13164

Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J. D., Sadavarte, P., et al. (2020). Quantifying methane emissions from the largest oil-producing basin in the United States from space. *Science Advances*, *6*(17), eaaz5120.