Supplementary Information (SI) for Research Article

Hydrocarbon Tracers Suggest Methane Emissions from Fossil Sources Occur Predominately Before Gas Processing and That Petroleum Plays Are a Significant Source

Ariana L. Tribby[1]; Justin S. Bois[2]; Stephen A. Montzka[3]; Elliot L. Atlas[4]; Isaac Vimont[3,5]; Xin Lan[3,5]; Pieter P. Tans[3]; James W. Elkins[3,5]; Donald R. Blake[6]; Paul O. Wennberg[7,8]*

[1] Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[3] National Oceanic and Atmospheric Administration, Global Monitoring Laboratory, Boulder, CO, 80305 USA

[4] Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA

[5] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, 80309 USA

[6] Department of Chemistry, University of California – Irvine, Irvine, CA 92697, USA

[7] Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA

[8] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, 91125, USA

[*] Corresponding author: wennberg@caltech.edu

Contains six text sections in 71 pages, 4 tables, and 63 figures.

Table of Contents

1. Economics & Production of Oil & Natural Gas

The development of the hydraulic fracturing (fracking) technique has led to a widespread increase in the production of associated and non-associated natural gas,[1] as well as natural gas liquids, or NGLs (Figure S1, left and Figure S2). NGLs are extracted with raw gas and are, just after extraction, often separated from the stream to yield consumer-grade natural gas and the more valuable alkane liquids.[2] Basins with petroleum typically have a larger composition of ethane and propane and other NGLs, whereas "dry" basins, such as Fayetteville and Appalachian in the U.S., tend to provide mostly dry natural gas with low fractions of NGLs.[3] By volume, ethane, C_2H_6, is the second most abundant component of natural gas after CH_4, while propane (C_3H_8) is the third most abundant.[4] The fraction of C_2H_6 (but not C_3H_8) removed during gas processing changes significantly over time. Over the last 10 years, the dramatic increase of C_2H_6 production (Figure S1, left) has exceeded domestic demand or ability to export it abroad,[5] resulting in C_2H_6 prices generally at or below natural gas since 2012 (Figure S1, right).[6,7] As a result, it is often more economical to sell C_2H_6 as natural gas rather than separate it from the raw stream. Producers can increase the amount of C_2H_6 they sell as natural gas by "rejecting" it (not recovering it). Rejection of C_2H_6 has continued to grow almost continuously over the past decade (Figure S1, right) resulting in increasing abundance of C_2H_6 in the natural gas distribution system.

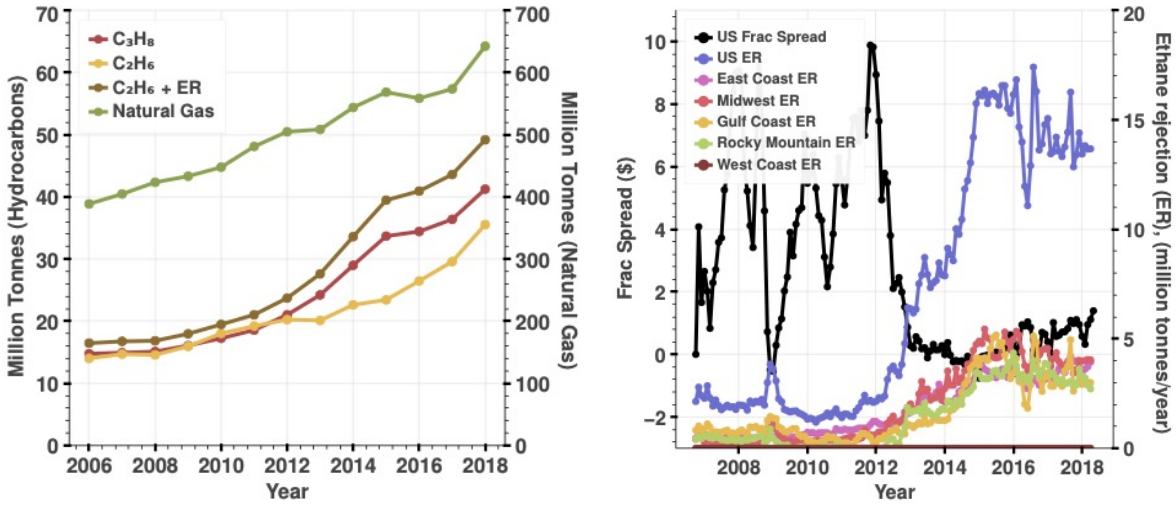


Figure S1. Economic trends of natural gas and natural gas liquids. Left: Trends in natural gas and hydrocarbon production (EIA) and total ethane summed with rejected ethane modeled by OPIS, Point Logic, provided by IHS Markit. **Right**: The value of ethane compared to natural gas represented by fractionation spread (frac spread) on the left axis. Ethane rejection in the U.S. and major U.S. refining areas is plotted on the right axis. (Data by OPIS, Point Logic, provided by IHS Markit.)

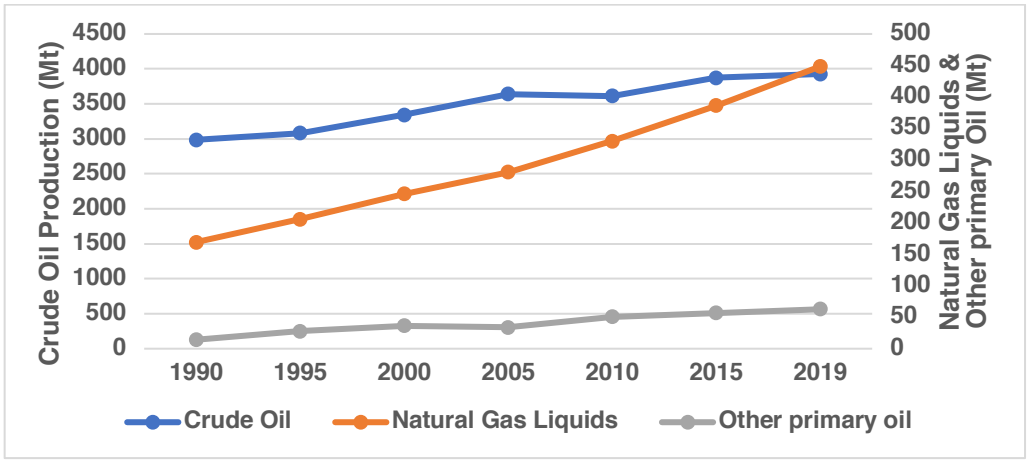


Figure S2. Global oil production. Data provided by IEA (https://www.iea.org/fuels-and-technologies/oil).

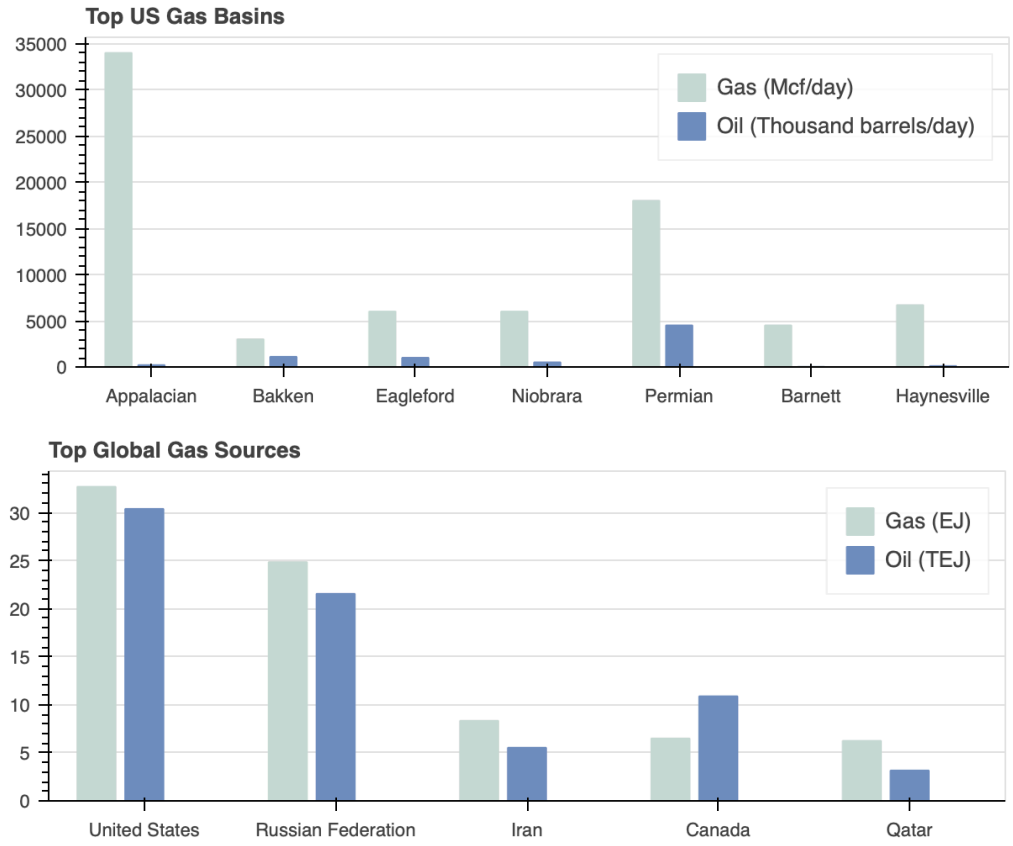


Figure S3. Top Natural Gas-Producing US basins/Countries and their corresponding oil production. Top: Oil and gas production for the top 7 natural gas producing basins that account for 86% of total U.S. natural gas production.[8,9] **Bottom**: Global oil and gas production for the top 5 natural gas producing countries that account for 50% of global natural gas production.[10,11]

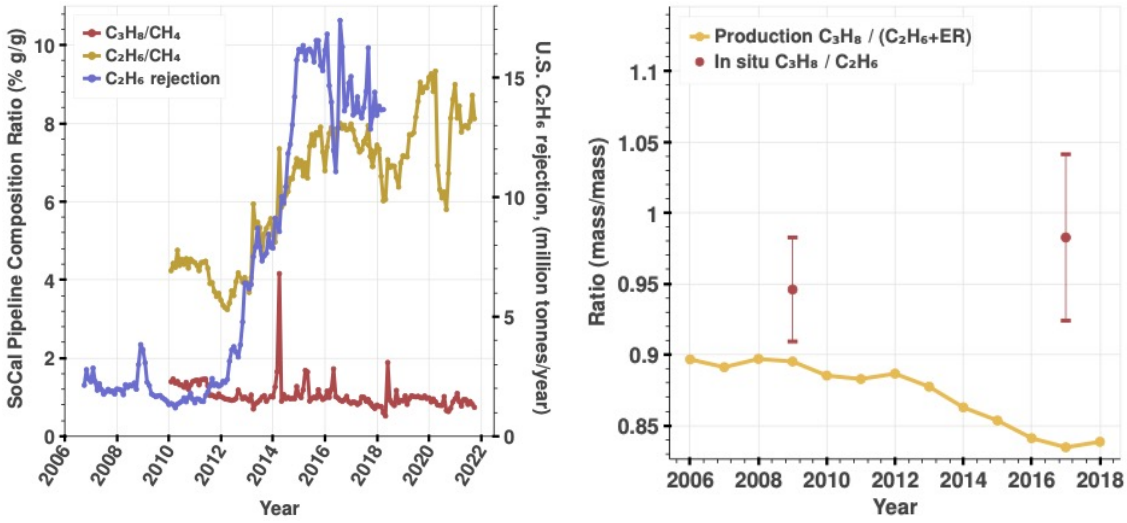


Figure S4. Observed pipeline composition in Playa del Rey and ethane rejection trends. Left: The ratio of propane/methane and ethane/methane measured in natural gas withdrawn from Playa del Rey in Southern California,[12] compared to U.S. ethane rejection (see Figure S1 for more information on rejection; data provided by IHS Markit). **Right**: The ratio of U.S. propane production and total ethane production (including rejection, ER). The production data is provided by EIA[13,14]s and the rejection data is provided by IHS Markit. The in-situ observed ratio is calculated from NOAA-ongoing observations, see Figure S9.

2. NOAA & FRAPPE Observations

2.1. Processing and statistical methods

Ethane and propane mole fraction data from aircraft measurements, tall tower, and surface flasks are publicly available at https://gml.noaa.gov/ccgg/arc/?id=155. Methane mole fraction data is publicly available at https://gml.noaa.gov/ccgg/obspack/data.php?Id=obspack_multi-species_1_CCGGAircraftFlask_v2.0_2021-02-09 , https://gml.noaa.gov/ccgg/obspack/data.php?Id=obspack_multi-species_1_CCGGSurfaceFlask_v2.0_2021-02-09, and https://gml.noaa.gov/ccgg/obspack/data.php?Id=obspack_multi-species_1_CCGGTowerInsitu_v1.0_2018-02-08.

We first discuss NOAA ongoing observations. We use measurement quality flags labeled as either preliminary or good sampling and analysis only. Some sites had an unequal number of quality measurements between alkane species, so we match UTC time stamps that are shared between each species to avoid sampling bias. We drop any corresponding paired measurements of CH_4, C_2H_6 and C_3H_8 that are labeled as NaN (all three pairs are dropped). Figure S5 shows the spatial location of the NOAA ongoing observation sites used in this analysis, and Table S1 lists the temporal and spatial coverage offered at each site. Most sites offered a few measurements each week for the years indicated.

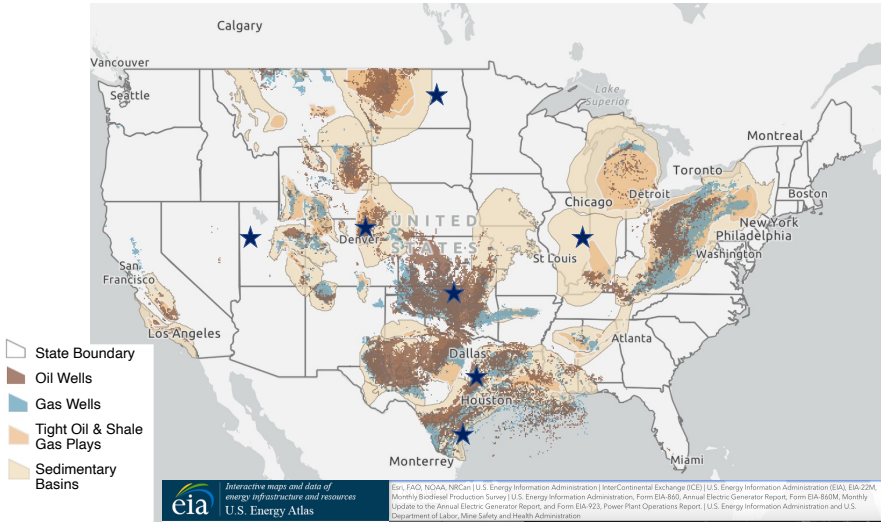


Figure S5. Major Oil/Natural Gas Shale Plays in the U.S. & NOAA ongoing measurement locations. Approximate geographical locations of NOAA ongoing measurement locations are shown in the blue stars on the map. Not pictured is East Trout Lake (ETL) site, located in Saskatchewan, Canada (54.3541N, 104.9868W). Well and basin layers provided by https://atlas.eia.gov/apps/all-energy-infrastructure-and-resources/explore .

Table S1. Sites for NOAA Ongoing Observations in the U.S.

Site Location	Site Abbreviation	Years	Measurement Method	Processing lab
Homer, IL	HIL	2015 - 2018	Aircraft/Tower	CCGG, HATS
Lamont, OK	SGP	2006 - 2017	Aircraft/Tower	CCGG
Dahlen, ND	DND	2014 - 2016	Aircraft/Tower	CCGG, HATS
East Trout Lake, Canada	ETL	2014 - 2018	Aircraft/Tower	CCGG, HATS
Wendover, UT	UTA	2006 - 2016	Tower	CCGG, ARL
Boulder, CO	BAO	2014 - 2016	Tower	CCGG, HATS
Moody, TX	WKT	2015 - 2018	Tower	CCGG, HATS
Sinton, TX	TGC	2015 - 2018	Aircraft/Tower	CCGG, HATS
Niwot Ridge, CO	NWR	2005- 2014	Tower	CCGG, ARL

For FRAPPE, we also use 1000 meters as a marker for the boundary layer, and analyze measurements taken above it. We also drop any corresponding paired measurements of CH_4, C_2H_6 and C_3H_8 that are labeled as NaN (all three pairs are dropped). Figure S6 shows the spatial location of the FRAPPE observations (after filtering) in Colorado.

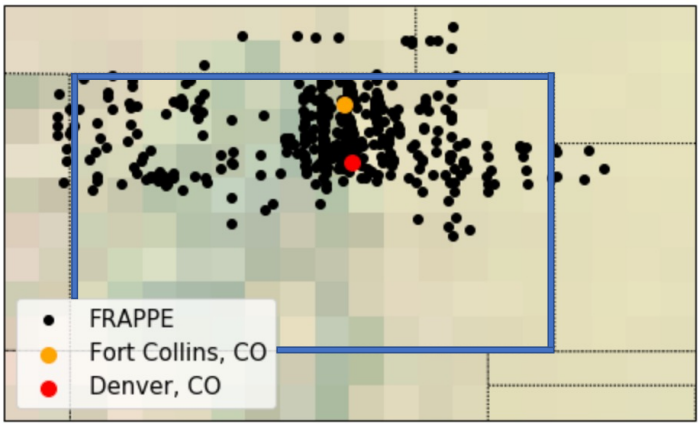


Figure S6. FRAPPE observations. The outline of Colorado state is shown in blue. We show data already pre-processed and filtered for fresh emissions, as discussed in this section.

To better quantify geophysical variability and generate a confidence interval in the correlation in measured mole fractions between C_2H_6 and C_3H_8, we implement a pairs bootstrap to generate replicates of C_2H_6 and C_3H_8 observations for NOAA and FRAPPE observations. We draw random samples of pairs of C_2H_6 and C_3H_8, where instead of drawing two random samples of each array, we draw the same indices from both arrays so we end up with paired samples, since C_2H_6 and C_3H_8 were measured together and we want to compute the correlation. We draw samples the size of the dataset, then compute the slope of the correlation. We repeat this 10,000 times. As shown for different scenarios below, we perform this bootstrap for individual sites separately, as well as all sites combined. The confidence interval reported for correlations between C_2H_6, C_3H_8, and CH_4 anomalies is the 95% CI of the 10,000 samples. The CIs calculated from the bootstrapped samples are much broader than those calculated assuming the noise in the measurements is dominated by analytical errors. This suggests that geophysical noise induced by differences in transport and chemistry dominates the statistics.

2.2. Chemical aging approach to determining methane background

We take a chemical aging approach to defining the threshold between samples associated with fresh emissions (unaged) and photochemically aged emissions. As in Parrish et al 2018,[15] we observe both fresh and aged regimes (Figure S7 below). We chose the 50th percentile of C_3H_8 as the demarcation between these regimes (about 10^3 ppt) and show in Figure S10 that our analysis of the ethane and propane ratio to methane is not terribly sensitive to the choice. This threshold will clearly depend on the fraction of samples obtained in the two regimes.

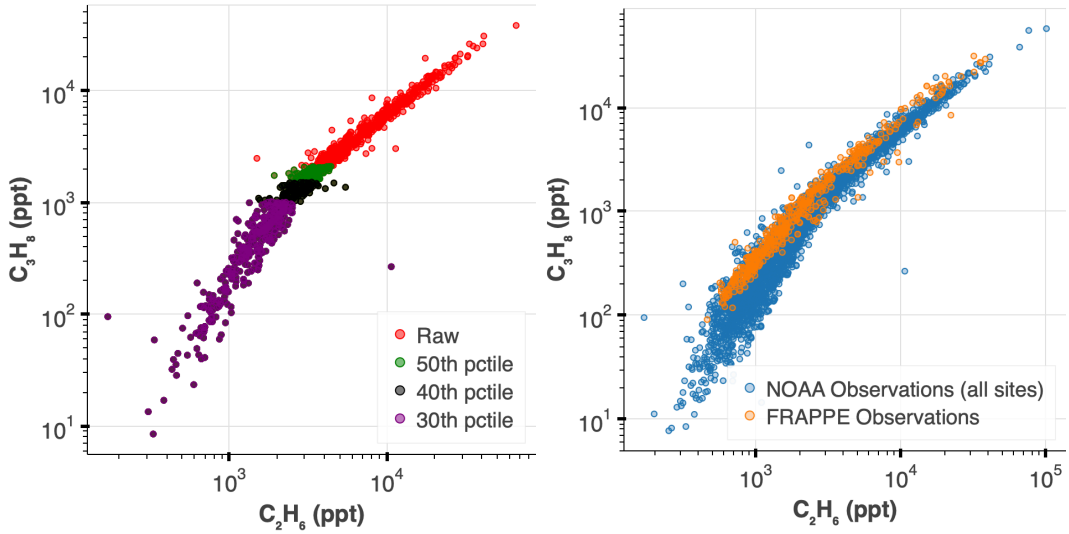


Figure S7. Identifying fresh emission chemical regime in NOAA and FRAPPE campaigns. Left: Varying C_3H_8 percentiles at the NOAA SGP site. The inflection point between aged and fresh emission regime is visually contained within varying the C_3H_8 percentile cutoff by ± 10%. We chose the 50th percentile of C_3H_8 as the demarcation between these regimes (about 10^3 ppt). **Right**: C_3H_8 vs C_2H_6 using FRAPPE data (already pre-processed, as described in the methods in the main text). FRAPPE observations are quite consistent with NOAA, hence we use the same C_3H_8 demarcation between the aged and fresh chemical regime.

After filtering for fresh emissions using C_3H_8 percentile method, two NOAA observation sites (NWR and UTA) only showed aged emissions (Figure S8). Consequently, these sites were not used in the subsequent analysis.

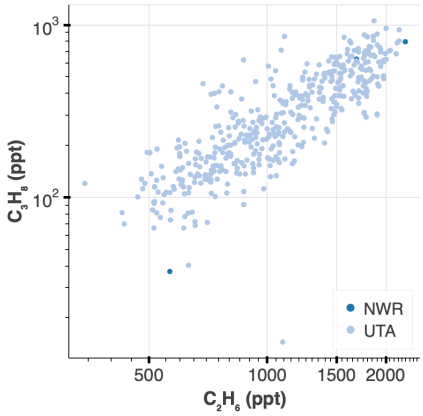


Figure S8. NOAA C_3H_8 vs C_2H_6 after filtering for fresh emissions. We used C_3H_8 50th percentile as a marker for fresh emissions (please see details in text above). Sites NWR and UTA only had C_3H_8 mole fractions below this demarcation and were assumed to be affected only by aged emissions, and as such, were excluded from further analysis.

After filtering for fresh emissions using C_3H_8 percentile method, the cross plot of C_3H_8 vs C_2H_6 is similar for all NOAA sites, NOAA SGP only, and FRAPPE observations (Figure S9).

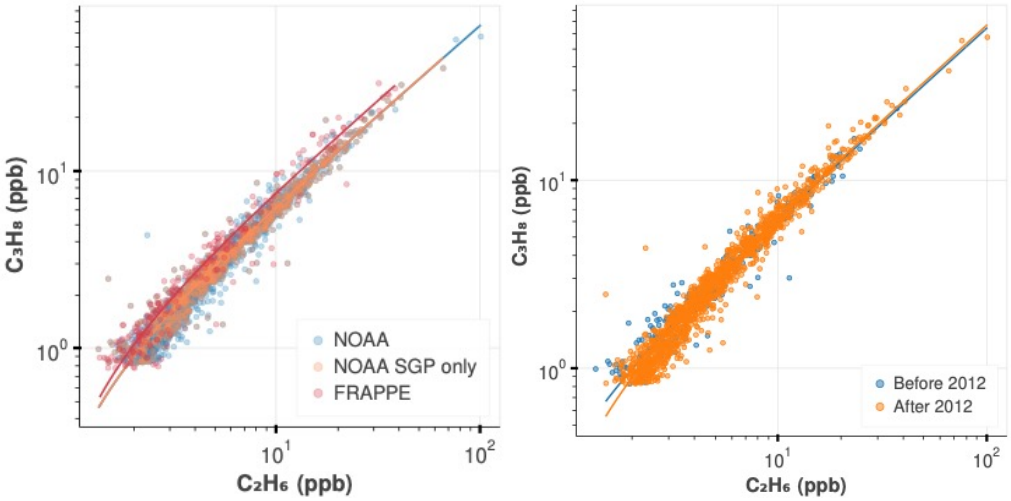


Figure S9. C_3H_8 vs C_2H_6 after filtering for fresh oil and gas emissions. Left: The filtering method is described in Section 3.1. We show observations for all NOAA sites (2005-2018, see Table S1), NOAA for SGP site only (2006-2018 Oklahoma tower and aircraft observations), and FRAPPE campaign (2014 aircraft observations around Colorado). **Top**: FRAPPE linear least squares slope 95% CI is [0.76, 0.87] (ppb/ppb, $R^2 = 0.97$) compared to [0.63, 0.70] (ppb/ppb, $R^2 = 0.98$) for all NOAA sites. Variability in the slope for both FRAPPE and NOAA is given by a pairs bootstrap analysis, described in Section 2.1. **Right**: Slope before 2012 (2005-2011): [0.62, 0.67] (95% CI), $R^2 = 0.98$. Slope after 2012 (2012-2018): [0.63, 0.71] (95% CI), $R^2 = 0.98$. We use data from SGP, TGC, ETL, HIL, DND, BAO, and WKT sites (see Table S1) before filtering for air influenced by fresh oil/gas emissions, which is shown here. We use both tower and aircraft data. We use pairs bootstrapping to arrive at confidence intervals, described in detail in Section 2.1.

After identifying the fresh emissions, we defined a background for CH_4 observations and constructed CH_4 anomalies by doing the following:

- Find corresponding co-CH_4 measurements in the aged air regime as identified by C_3H_8 mole fraction (below 10^3 ppt C_3H_8).
- Interpolate this CH_4 array to the full timeseries using time to obtain a CH_4 background.
- Subtract this interpolated background from the full CH_4 array to obtain a CH_4 anomaly. Note that because the CH_4 lifetime is much longer than either C_2H_6 or C_3H_8, the differences are much smaller.

Since we only focus on the linear part of the curve, our analysis is not terribly sensitive to how the CH_4 anomaly is determined (it simply produces varying intercepts, see our quantitative analysis on the impact on the slope in Figure S10). Again, to get C_3H_8/ C_2H_6, we only consider the fresh emission regime (beyond 10^3 ppt C_3H_8). Figure S10 shows our calculated CH_4 background and CH_4 anomalies for the NOAA SGP site (near Lamont, OK).

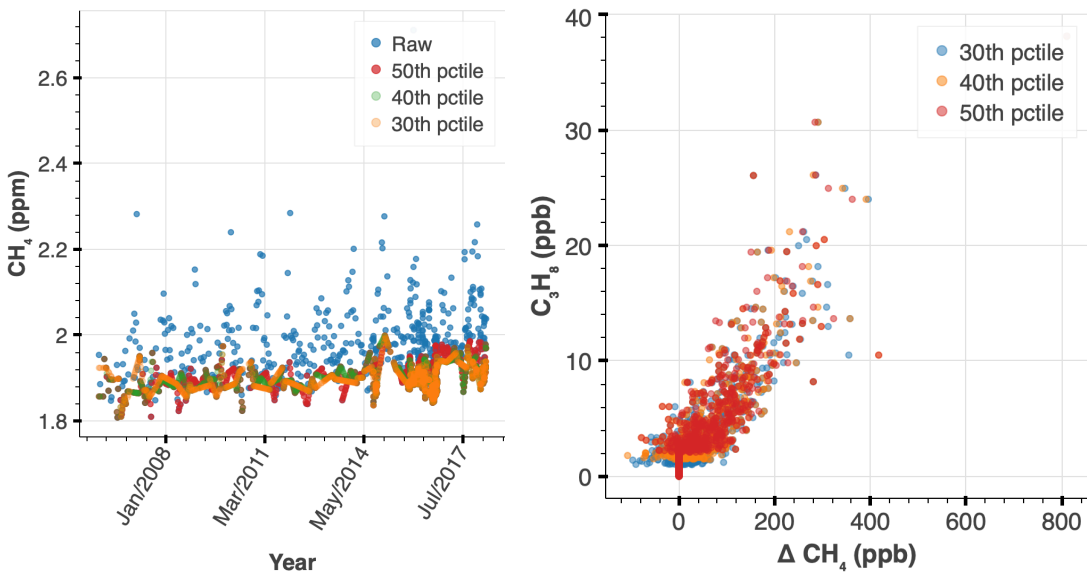
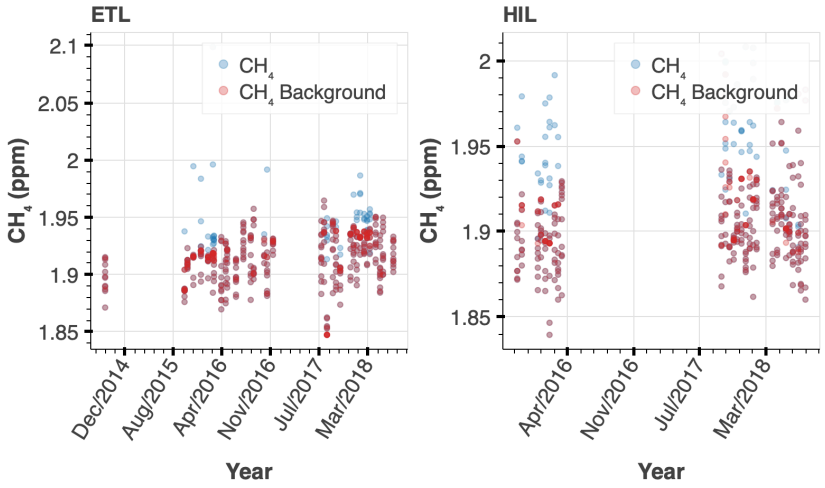


Figure S10. Left: Background estimate for CH_4 at NOAA SGP site. Left: Using 50% ±10% percentile cutoff of C_3H_8 has a minimal effect on the background CH_4 estimation. **Right**: C_3H_8 vs CH_4 Anomaly at NOAA SGP site. A CH_4 anomaly is calculated by linearly interpolating the estimated CH_4 background to the raw CH_4 measurement timescale. The interpolated background is then subtracted from the raw CH_4 measurements. Using 50% ±10% percentile cutoff of C_3H_8 has a minimal effect on CH_4 anomaly cross plots. Using a pairs bootstrap approach (see Section S2.1), we generate thousands of slope replicates and calculate the following 95% CIs for the slope using the following C_3H_8 percentile cutoffs: [0.0458, 0.0526] (30th percentile); [0.0460, 0.0534] (40th percentile); [0.0481, 0.0563] (50th percentile).

Below in Figure S11, we show the result of our CH_4 background calculations for each NOAA site.



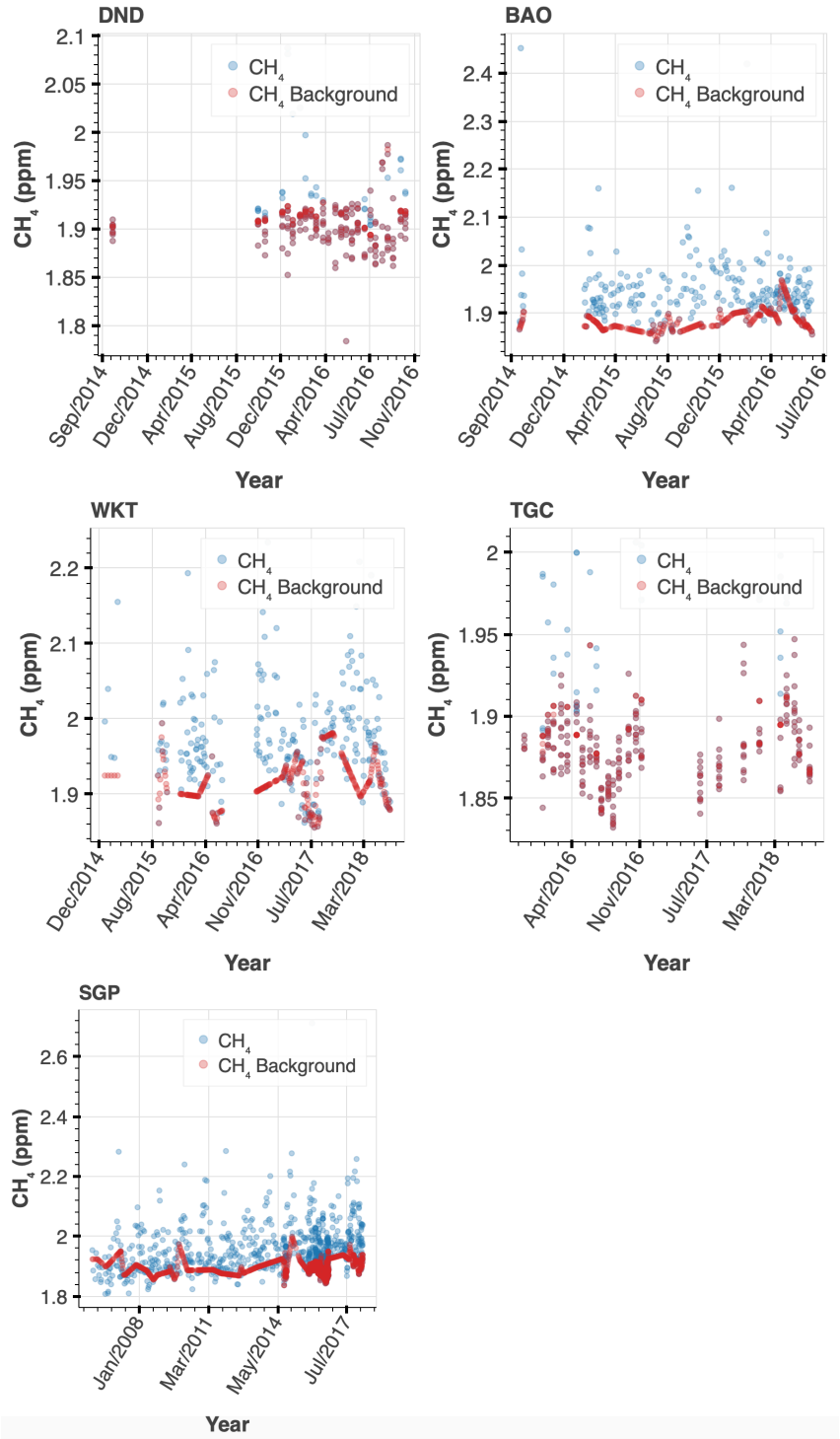


Figure S11. Methane vs time for NOAA sites. The estimated background is shown in red. Raw CH_4 data is shown in blue. The background was calculated using 50% C_3H_8 percentile cutoff method.

2.3. Methane anomaly plots for NOAA and FRAPPE campaigns

Below, we show the results of our CH_4 anomaly calculations for NOAA and FRAPPE observations using the methods described in Section 2.2 (above). First, we show correlations for

individual NOAA sites (Figures S12, S13), followed by a comparison between NOAA SGP site (Oklahoma site) and FRAPPE observations (Figure S14).

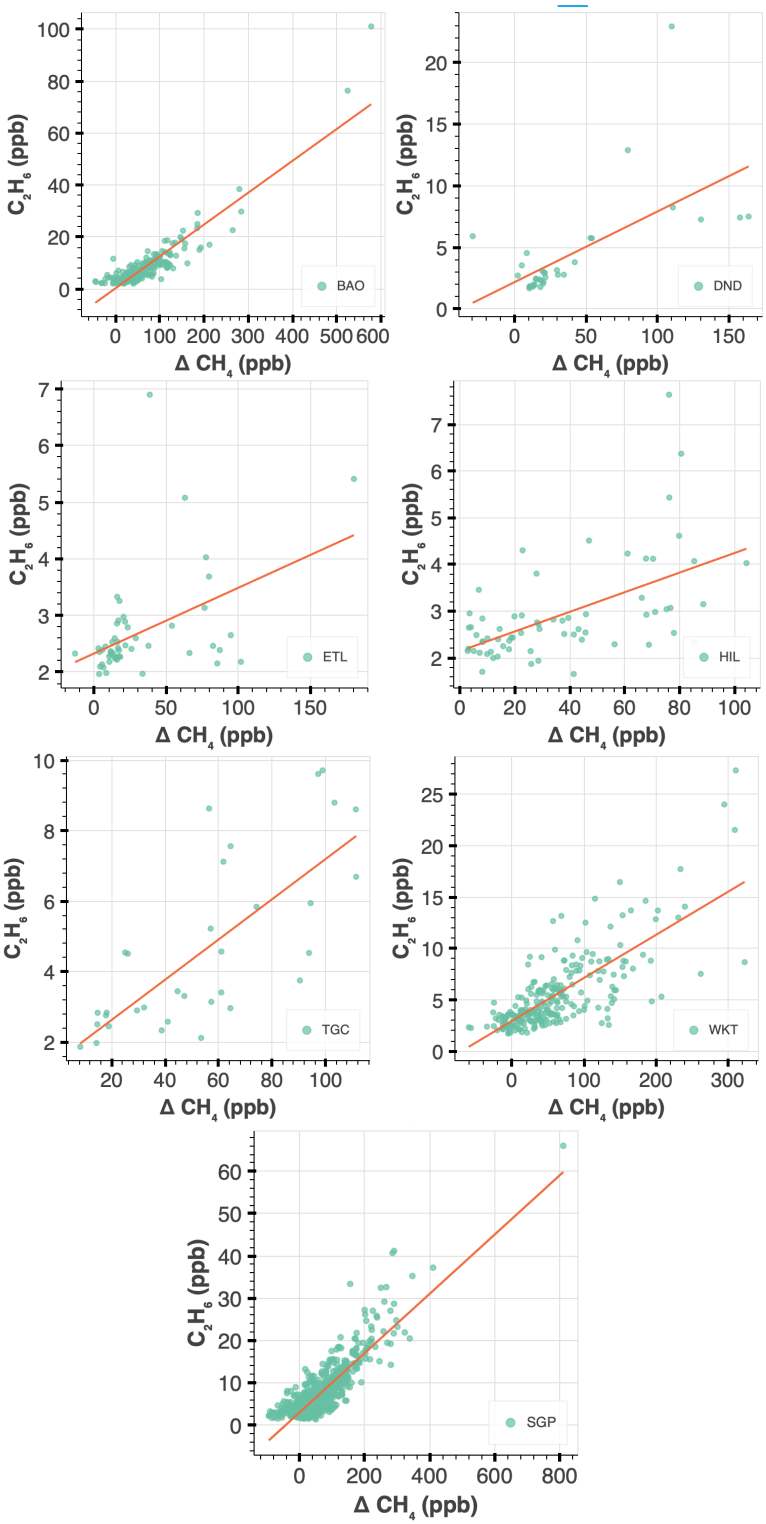


Figure S12. C_2H_6 vs CH_4 anomaly for NOAA sites. The data for each site were filtered using the chemical aging regime to filter for fresh emissions and to construct CH_4 anomalies (Section 2.2). See Table S1 for a description of site location/observation type. We also ran a bootstrap for each individual site (bootstrapping methods, main text). The 95% CI slopes (ppb/ppb) are as

follows: BAO: [0.0833, 0.1449], $R^2 = 0.91$; DND: [0.0289, 0.1205], $R^2 = 0.63$; ETL: [0.0030, 0.0176], $R^2 = 0.46$; HIL: [0.0116, 0.0313], $R^2 = 0.56$; TGC: [0.0400, 0.0730], $R^2 = 0.74$; WKT: [0.0324, 0.0510], $R^2 = 0.75$; SGP: [0.0645, 0.0749], $R^2 = 0.86$.

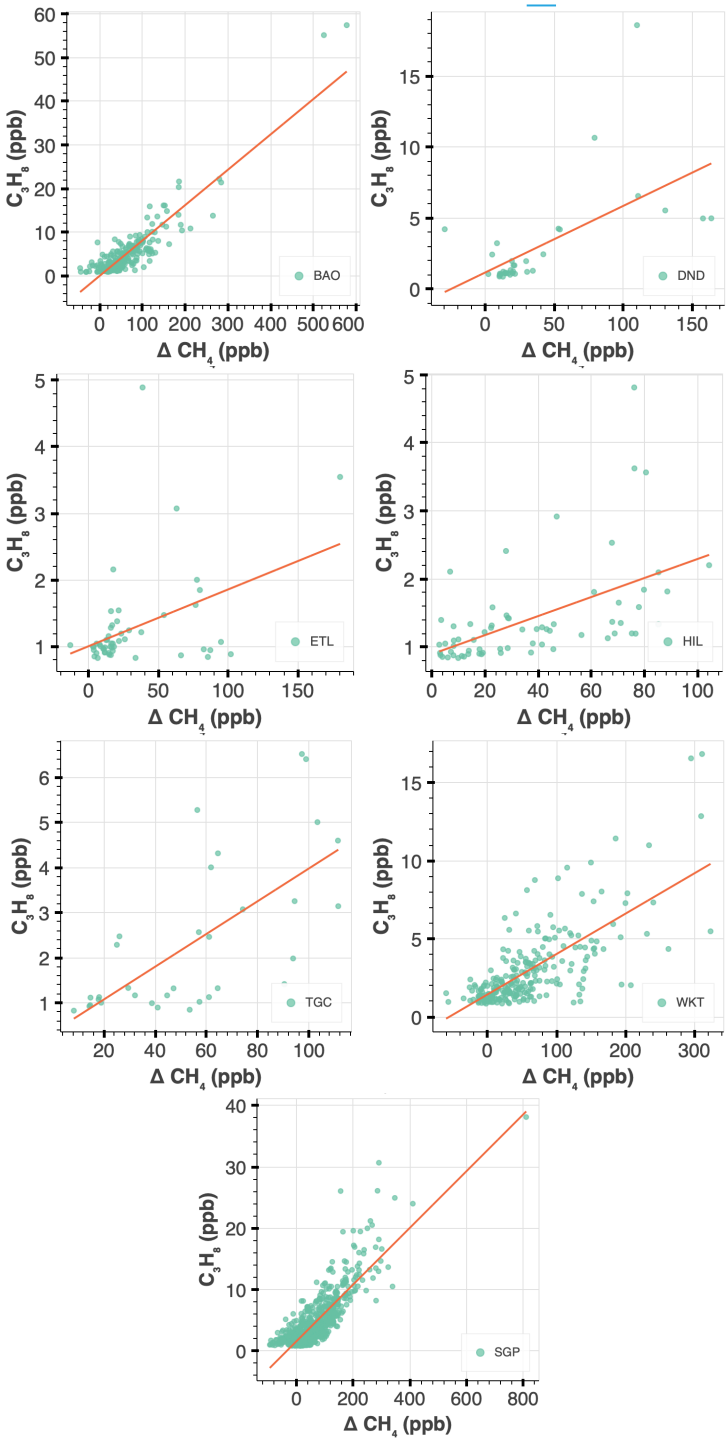


Figure S13. C_3H_8 vs CH_4 anomaly for NOAA sites. The data for each site were filtered using the chemical aging regime to filter for fresh emissions and to construct CH_4 anomalies (Section 2.2). See Table S1 for a description of site location/observation type. We also ran a bootstrap for each individual site (bootstrapping methods, main text). The 95% CI slopes (ppb/ppb) are as

follows: BAO: [0.0587, 0.0922], R^2 = 0.91; DND: [0.0221, 0.1003], R^2 = 0.61; ETL: [0.0013, 0.0136], R^2 = 0.41; HIL: [0.0078, 0.0216], R^2 = 0.54; TGC: [0.0228, 0.0506], R^2 = 0.68; WKT: [0.0195, 0.0321], R^2 = 0.71; SGP: [0.0426, 0.0499], R^2 = 0.83.

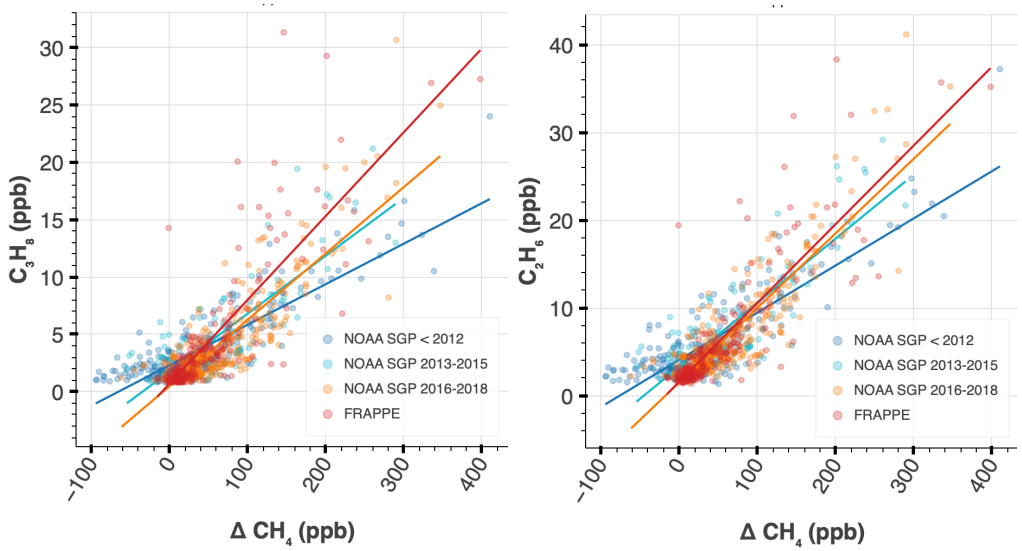


Figure S14. C_2H_6 vs CH_4 anomaly for NOAA SGP site and FRAPPE study. We use the chemical aging approach defined in Section 3.1 to identify C_3H_8 and CH_4 observations within a fresh oil and gas emissions chemical regime. We construct CH_4 background-corrected anomalies as described in SI Section S2. Results with C_2H_6 are similar and shown in Figure S9. NOAA observations for SGP site (Oklahoma, Table S1) are shown here. We show correlations between 2006-2011, labeled as "NOAA SGP < 2012" (HIPPO takes place between 2009-2011), 2013-2015 (FRAPPE takes place in 2014), and 2016-2018 (ATom time period). **Left**: The slope of the correlation between C_3H_8 and CH_4 anomaly for NOAA observations before 2012 is [0.031, 0.040] ppb/ppb, R^2 = 0.85; between 2013-2015 is [0.045, 0.084], R^2 = 0.82; and between 2016-2018 is [0.039, 0.059], R^2 = 0.86. FRAPPE is [0.063, 0.085] ppb/ppb, R^2 = 0.83. The slope of the correlation for all years of NOAA is [0.043, 0.050] ppb/ppb, R^2 = 0.83. C_3H_8 vs CH_4. **Right**: C_2H_6 vs CH_4 The FRAPPE slope (95% CI, ppb/ppb) is [0.0763, 0.1047], R^2 = 0.85. C_2H_6 vs CH_4 NOAA slope for all years is [0.0647, 0.0749], R^2 = 0.86. The C_2H_6 vs CH_4 slope before 2012 is [0.047, 0.060], R^2= 0.85; from 2013-2015 is [0.066, 0.143], R^2 = 0.85; and from 2016-2018 is [0.058, 0.084], R^2 = 0.88.

2.4. Maps of FRAPPE and NOAA SGP observations compared to oil and gas sites

Below, we show the location of the NOAA samples taken at the SGP site in Oklahoma, which include a combination of aircraft and tower measurements (Figure S15). We add the approximate location on top of a USGS map of oil and gas sites in Oklahoma, using coordinates for Lamont and Billings for reference (Figure S16), where we see that the SGP measurements are taken around a mix of oil and gas sites. We include a figure of Oklahoma oil and gas production by county (Figure S17), where we see widespread surrounding oil and gas production.

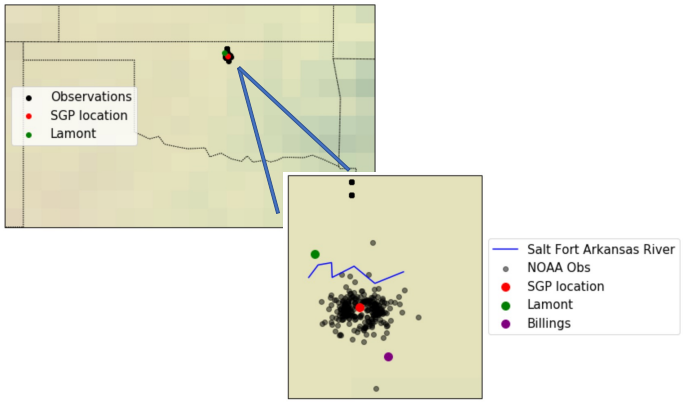


Figure S15. NOAA observations at SGP site (Oklahoma). The observations shown here are pre-processed and filtered for fresh emissions as discussed in the methods section in the main body.

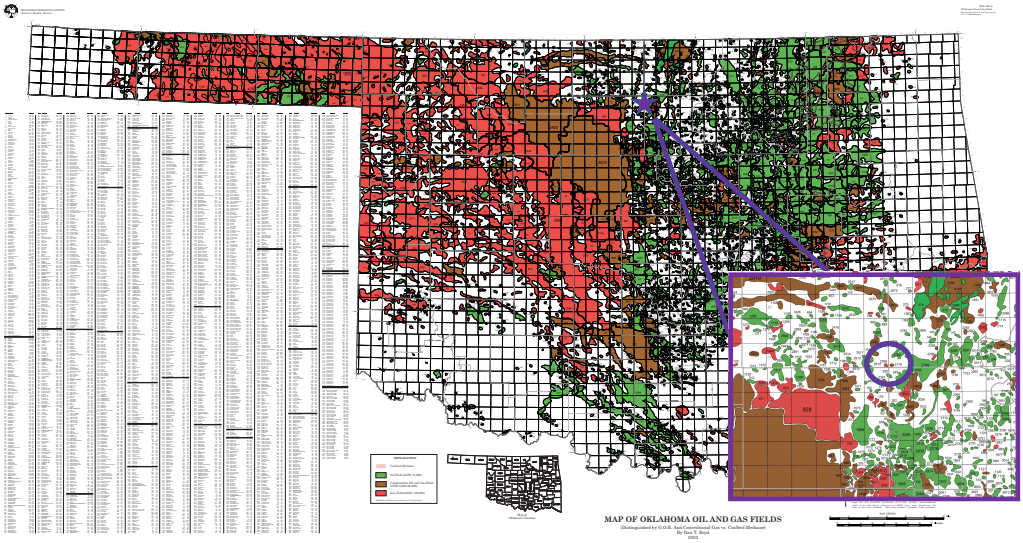


Figure S16. Oklahoma oil and gas wells. Plot adapted from Oklahoma Geological Survey[16]: http://www.ogs.ou.edu/fossilfuels/MAPS/GM-36.pdf.

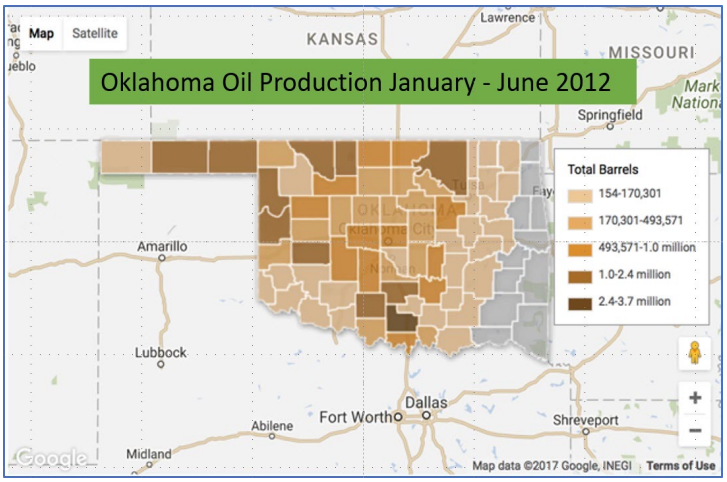


Figure S17. Oklahoma oil and gas production by county. Plot was created by Joe Wertz of StateImpact Oklahoma[17].

Below, we include a plot by Water Education Colorado showing the value of oil and gas production value by county (Figure S18). If we compare it to Figure S6, it is evident that much of the FRAPPE observations were taken around nearby oil-producing wells that produce significant revenue. This is consistent with the large C2 and C3 to C1 emission ratios observed during the FRAPPE campaign (Figure 13, main text).

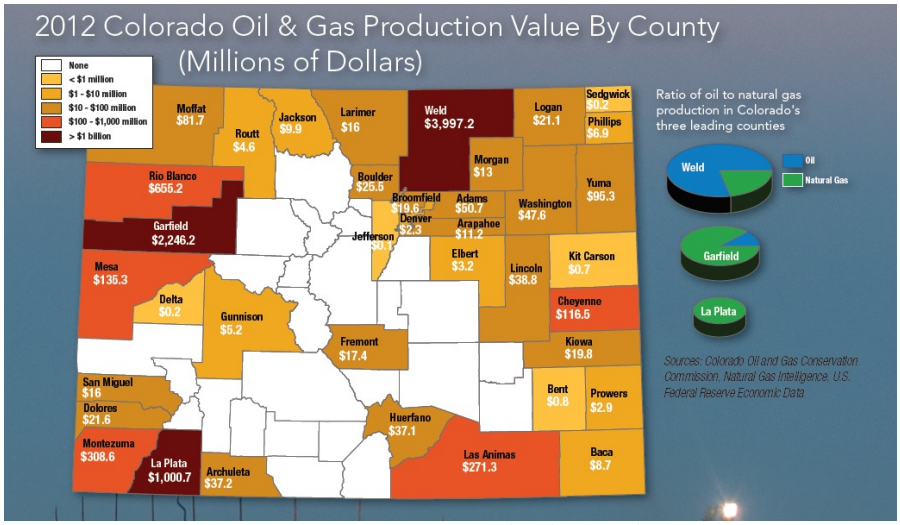


Figure S18. Oil and gas production value by county in Colorado. The plot was obtained from Water Education Colorado (founded by Colorado State Legislature).[18]

2.5. Comparison to Lan et al. 2019 study

Lan et al. 2019 investigated C_3H_8/CH_4 and C_2H_6/CH_4 ratios using NOAA-ongoing observations. Consistent with their study, we find increasing C_3H_8/CH_4 and C_2H_6/CH_4 ratios over time with relatively similar slopes. However, we find no statistically significant temporal trend in C_3H_8/C_2H_6. As shown in Figure S9 and in the additional cross plot of C_3H_8 and C_2H_6 colored by time (Figure S19, left), the correlation of these gases in the 'fresh emissions' regime is identical within error. Even when excluding aircraft data for SGP site, the ratio remains nearly the same (Figure S19, right).

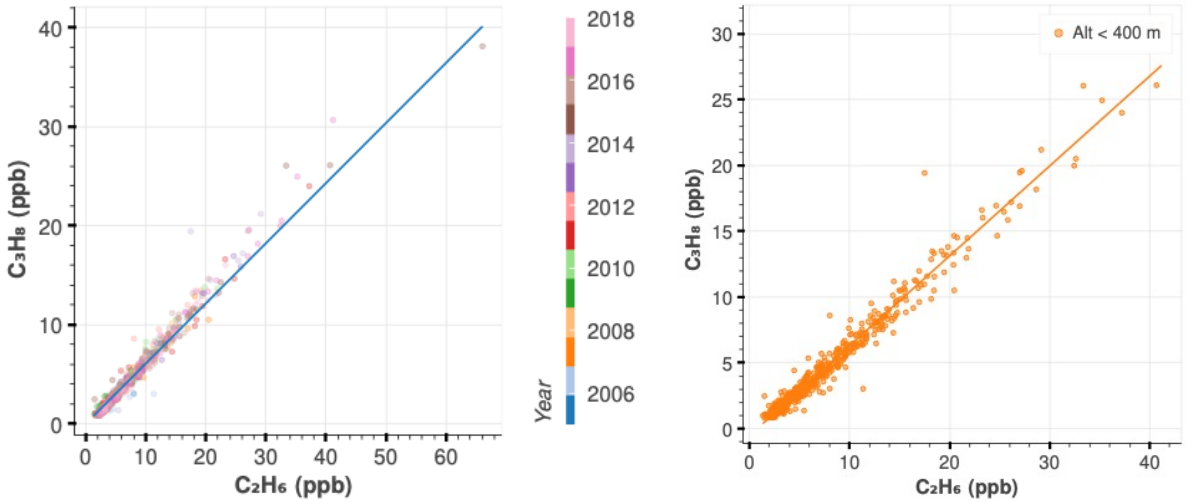


Figure S19. C_3H_8 and C_2H_6 correlation at NOAA SGP site– yearly and tower observations. Left: C_3H_8 vs C_2H_6 colored by all years for the NOAA SGP site. Slope: [0.63, 0.70] (95% CI), R^2=0.98. **Right**: Slope of C_3H_8 vs C_2H_6 for ground- and tower-based measurements NOAA SGP site. (The highest tower sampling is 374m sampling at SGP.) The slope is [0.66, 0.70], and R^2=0.99, comparable to [0.63, 0.70] 95% CI slope of the correlation that includes both aircraft and tower observations (this Figure, left side). We bootstrapped the samples to obtain a 95% CI (see methods, main text).

Our C_3H_8 anomalies are calculated in a different fashion than in Lan et al. Given the curvature of the correlation (Figure 1, main text) and its seasonal dependence, we determine the slope of the C_3H_8 and CH_4 anomalies only within the fresh emission regime ($C_3H_8 > 10^3$ ppt, along with co-measurements of CH_4, see Section 2.2). Since we only focus on the linear part of the curve, our analysis is not terribly sensitive to how the CH_4 anomaly is determined (it simply produces varying intercepts). To estimate C_3H_8/C_2H_6, we also only consider the fresh emission regime (beyond 10^3 ppt C_3H_8).

We replicate Figure 3a,b from Lan et al. 2019 using our methods in Figure S20. Even with very different methodology, our results for the central value of the emissions ratio between C_3H_8 and C_2H_6 and CH_4 anomalies are similar, albeit Lan et al. claims a much smaller uncertainty in these ratios (51.2 $\pm$ 0.6ppt/ppb and 80.5 $\pm$ 2.5ppt/ppb, respectively) such that the interannual variation and the trend over the record far exceed the stated uncertainty. Our 95% CI of the slope for C_3H_8/CH_4 is [42.57, 49.87] and for C_2H_6/CH_4 is [64.66, 74.89] (both in ppt/ppb) where most of the CI spread results from the temporal trend (see Figure S20).

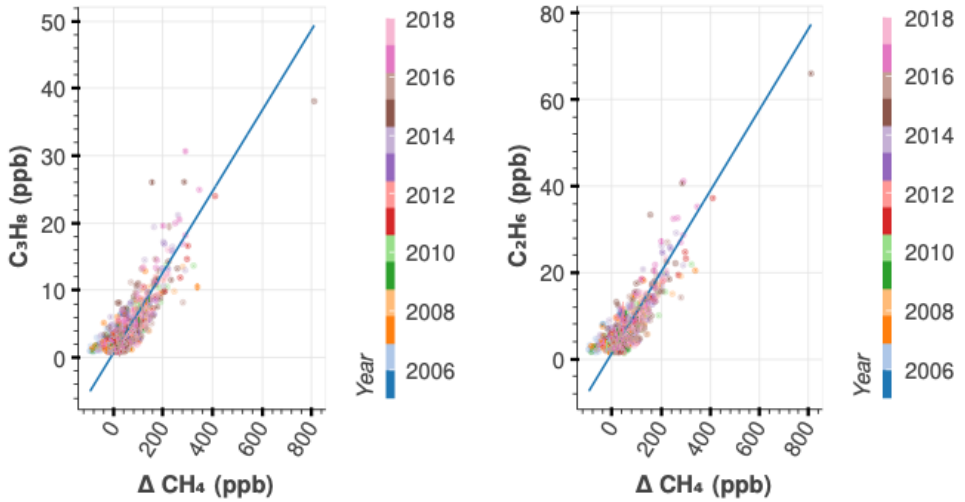


Figure S20. NOAA C_3H_8 and C_2H_6 vs CH_4 anomaly colored by year. Data is for SGP site only. $C_3H_8/\Delta CH_4$ slope: [0.43, 0.50] ppb/ppb, R^2= 0.83. $C_2H_6/\Delta CH_4$ slope: [0.65, 0.75] ppb/ppb, R^2= 0.86. We use data within the fresh emission regime (see Section 2.2). Our methods for determining CH_4 anomalies are described in detail in section 2.3, and our methods for determining the 95% CI via bootstrapping is described in the methods section of the main text.

We reproduce Figure 3e from Lan et al. 2019 in Figure S21, below. In Figure S21, the variability in ratios each year is constructed from the 95% confidence interval of the slopes from samples of a pairs bootstrap, described in more detail in section 2.1 As in Lan et al., we find significant trends in C_3H_8/CH_4 anomalies (3.12 ± 0.63 ppt/ppb/year), and C_2H_6/CH_4 anomalies (3.89 ± 0.84 ppt/ppb/year), which are comparable to their result. On the right side of Figure S21, we plot the fractional change relative to the mean. Instead, we find that both ratios are fractionally increasing at the same rate. The reported error in the slope is simply the standard error calculated from a linear regression of the yearly slopes vs year (that includes the upper and lower CI points). We use the linear trend of anomalies/year (Figure S21, left) to calculate mean anomaly ratios for SGP during 2016-2018 to be [0.060, 0.061] ppb/ppb (and C_2H_6/CH_4 to be 0.086, 0.088]), where the interval is determined using the standard error in the slope.

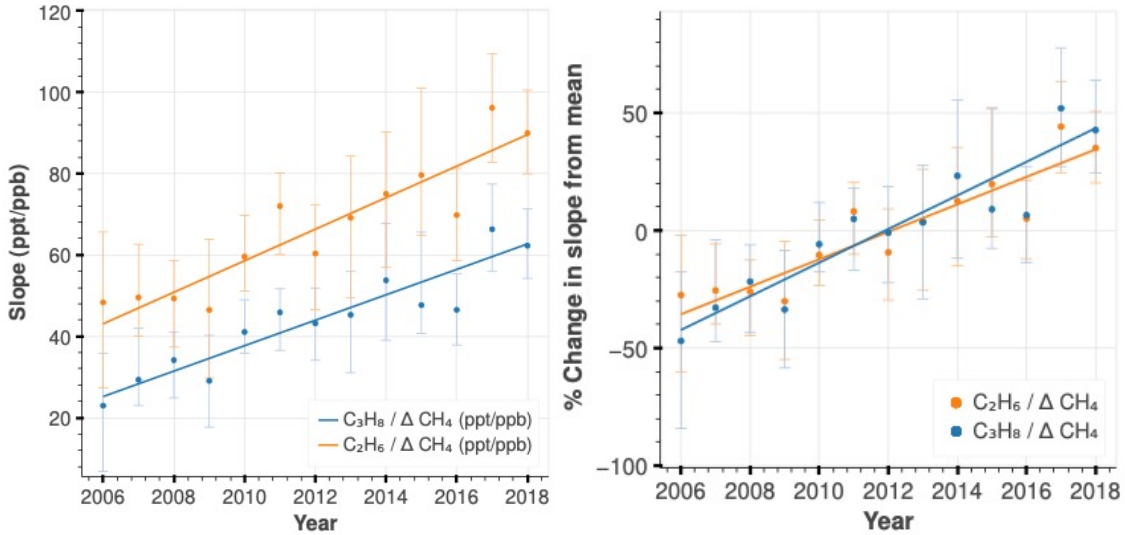


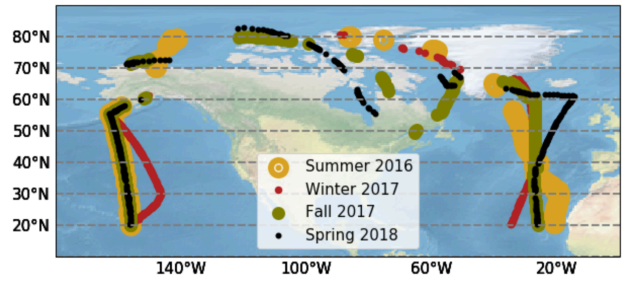
Figure S21. Yearly correlation between NOAA hydrocarbon vs CH_4 anomaly. Left: Average hydrocarbon vs CH_4 anomaly for each year for NOAA SGP site. C_3H_8/CH_4 anomaly slope: 3.12 ± 0.63 ppt/ppb/year (R^2=0.71), and C_2H_6/CH_4 anomaly trend is 3.89 ± 0.84 ppt/ppb/year (R^2=0.69). The variability in the trend (ppt/ppb/year) comes from the standard error of a linear regression. The variability in the yearly slope (ppt/ppb) comes from the 95% confidence interval of a pairs bootstrap (we ran a

pairs bootstrap for co-measurements of C_3H_8 and ΔCH_4 and compute the slope of the correlation for each bootstrap sample and repeated this for every year in the data; please see the methods section of the main text for more information about pairs bootstrapping). **Right**: Same as left, but in units of percent change with respect to the mean hydrocarbon and methane anomalies. The resulting trend for C_3H_8/CH_4 is 7.13 ± 1.44 % with an R^2 of 0.71. The trend for C_2H_6/CH_4 is 5.87 ± 1.26 % with an R^2=0.69. Both trends are calculated in the same way as the left figure.

Our analysis of the NOAA data suggests that the C_3H_8/C_2H_6 ratio is quite static in the U.S. over this 12-year record. That the ratios C_2H_6/CH_4 (and C_3H_8/CH_4) are increasing over time is completely consistent with Lan et al. 2019, and as they point out, studies that assume these ratios are invariant will overestimate the rate of oil/gas CH_4 emissions. Here, we use C_3H_8 vs CH_4 and C_2H_6 vs CH_4 between 2012 and 2018 to estimate the CH_4 emissions from the US as this is the period when most of the top-down and bottom-up estimates of CH_4 have been performed. That the ratios are getting "wetter" (higher hydrocarbon content in pre-processed gas) over time is consistent with an increasing contribution from oil exploration. That the atmospheric C_2H_6 and C_3H_8 increase fractionally the same, suggests that the ratio of the alkanes in the reservoirs producing these emissions do not change significantly over the time of this record.

3. ATom & HIPPO aircraft observations

The HIPPO campaign was a sequence of five global measurement campaigns which sampled from near the North Pole to the coastal waters of Antarctica, covering different seasons and years: HIPPO 1: 8-30 January 2009, HIPPO 2: 31 October – 22 November 2009, HIPPO 3: 24 March – 16 April, HIPPO 4: 14 June – 11 July 2011. ATom was a sequence of four global campaigns that took place from 29 July – 12 August 2016, 26 January – 10 February 2017, 28 September 2017- 11 October 2017, and 24 April 2018 – 6 May 2018. Flight paths of HIPPO and ATom campaigns are illustrated in Figure S22.



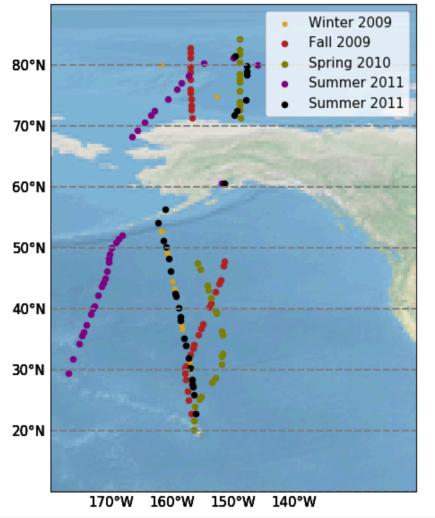


Figure S22. Truncated ATom and HIPPO flight paths. Flight paths used in this analysis are shown above (Top: ATom, Bottom: HIPPO). We split the data into Pacific (left column) and Atlantic (right column) "curtains" shown above for ATom, but HIPPO only offered Pacific curtains over remote ocean. The flight paths shown above do not encompass the entire dataset due to filtering out measurements south of 20 latitude north, those obtained over land, and those associated with very recent emissions. A summary of the filtering parameters we use in the main text are shown in Table S2.

Table S2. Filters for Aircraft Measurements.

Parameter	ATom Aircraft	HIPPO Aircraft
Altitude	> 1000 meters	> 1000 meters
N_2O	> 0.327 ppb	> 0.320 ppb
Tropopause Pressure	> 100 hPa	> 100 hPa
Summer season	exclude	exclude

HIPPO data were accessed from https://www.eol.ucar.edu/field_projects/hippo on 12/18/18, using the "discrete continuous merge" file. ATom data were accessed from the WAS-Discrete merged file from espo.nasa.gov/atom/archive/browse/atom/DC8/MER-WAS in August, 2021. Data taken over landmass for both aircraft campaigns were filtered away using global-land-mask version 1.0.0,[19] available from Python Package Index.

Here, we outline additional information on processing ATom and HIPPO aircraft observations used to compare with the GEOS-Chem model. HIPPO measurements were filtered for AWAS/UM instrument measurements to avoid measurement bias in C_3H_8 over C_2H_6, as the NOAA instrument only measured one of those species. All measurement species of HIPPO and ATom were filtered for consistent measurements of C_3H_8 and C_2H_6 within their respective campaigns; i.e., when those constituents were either both null or non-zero. This requires filtering data to remove plumes from highly local sources (including both energy infrastructure and wildfires), and to exclude regions and times where the lifetime of the alkanes is very short and thus regional / local sources dominate the variance. To reduce the influence of local sources, we only analyze observations in the free troposphere over the ocean at altitudes above 1000 meters (this filter excludes less than 20% of the dataset). To diagnose tropical air, we use tropopause

pressure as a filter. We use the NASA Global Modeling and Assimilation Office GEOS FP-IT (version 5.12.4) tropopause pressure product at 0.5 x 0.67 resolution,[20] and linearly interpolate it to the HIPPO aircraft path (for ATom, we use the product already included in the dataset). We only analyze measurements with tropopause pressure above 100 hPa (about 5% of the data was excluded under this constraint) for both ATom and HIPPO, which was sufficient to reduce the influence of tropical intrusions.

Because C_2H_6 and C_3H_8 are relatively short-lived gases, their abundance in the stratosphere is low and poorly connected to the underlying fluxes. To exclude stratospheric observations, we use N_2O which is inert and generally well-mixed in the troposphere, but is destroyed in the stratosphere by photolysis and reaction with O^1D.[21] Thus, we exclude data associated with low N_2O mole fraction (Figure S23). We use nearest neighbor interpolation to estimate missing N_2O observations. In Figure S23, we compare N_2O observations and GEOS-Chem simulations of N_2O and determine a common filter for both datasets. (Note that we generate GEOS-Chem N_2O simulations shown in Figure S23 by interpolating GEOS-Chem to aircraft latitude, longitude, time and potential temperature, but for all subsequent analysis, we filter GEOS-Chem by N_2O before interpolating to the aircraft potential temperature). To account for biomass burning, we use HCN as a tracer and did not use data with high HCN (Figure S24-26) for ATom observations only, as HIPPO did not provide HCN observations.

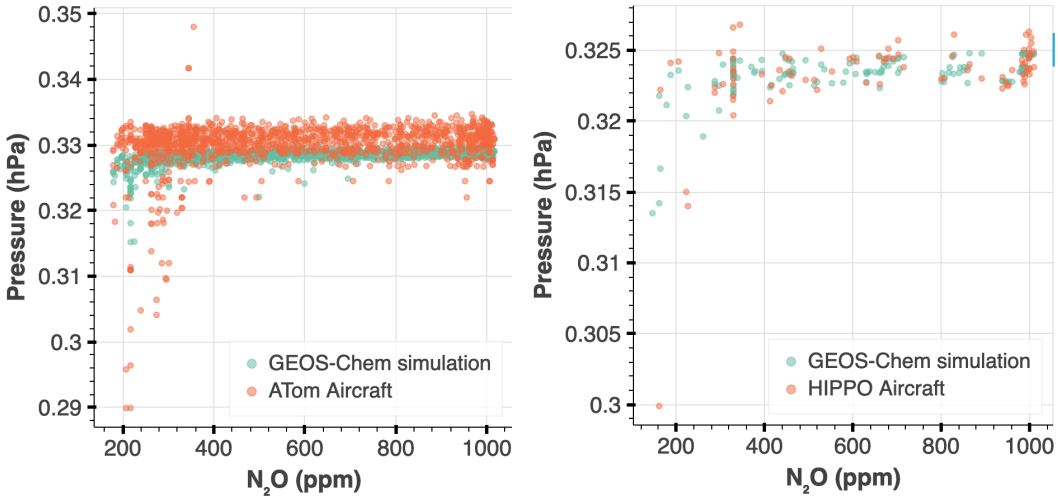


Figure S23. Stratospheric filter using N_2O. Left: ATom. **Right**: HIPPO. In both figures, GEOS-Chem simulations were interpolated to aircraft latitude, longitude, time, and potential temperature in order to compare N_2O. However, for all subsequent analysis, GEOS-Chem was filtered by N_2O before interpolating simulations to aircraft potential temperature. We use these figures to arbitrarily choose 0.327 and 0.320 N_2O mole fractions as a filter cutoff for ATom and HIPPO, respectively, as described in the main text.

To account for biomass burning, we use HCN as a tracer. We see elevated HCN over the Atlantic ocean on several campaigns (Figure S21). In the cross plot of HCN and C_2H_6 (Figure S23), we observe distinct plumes of elevated HCN and C_2H_6 that suggest biomass burning, and we de-weighted and excluded samples with high HCN (see description in Figure S23).

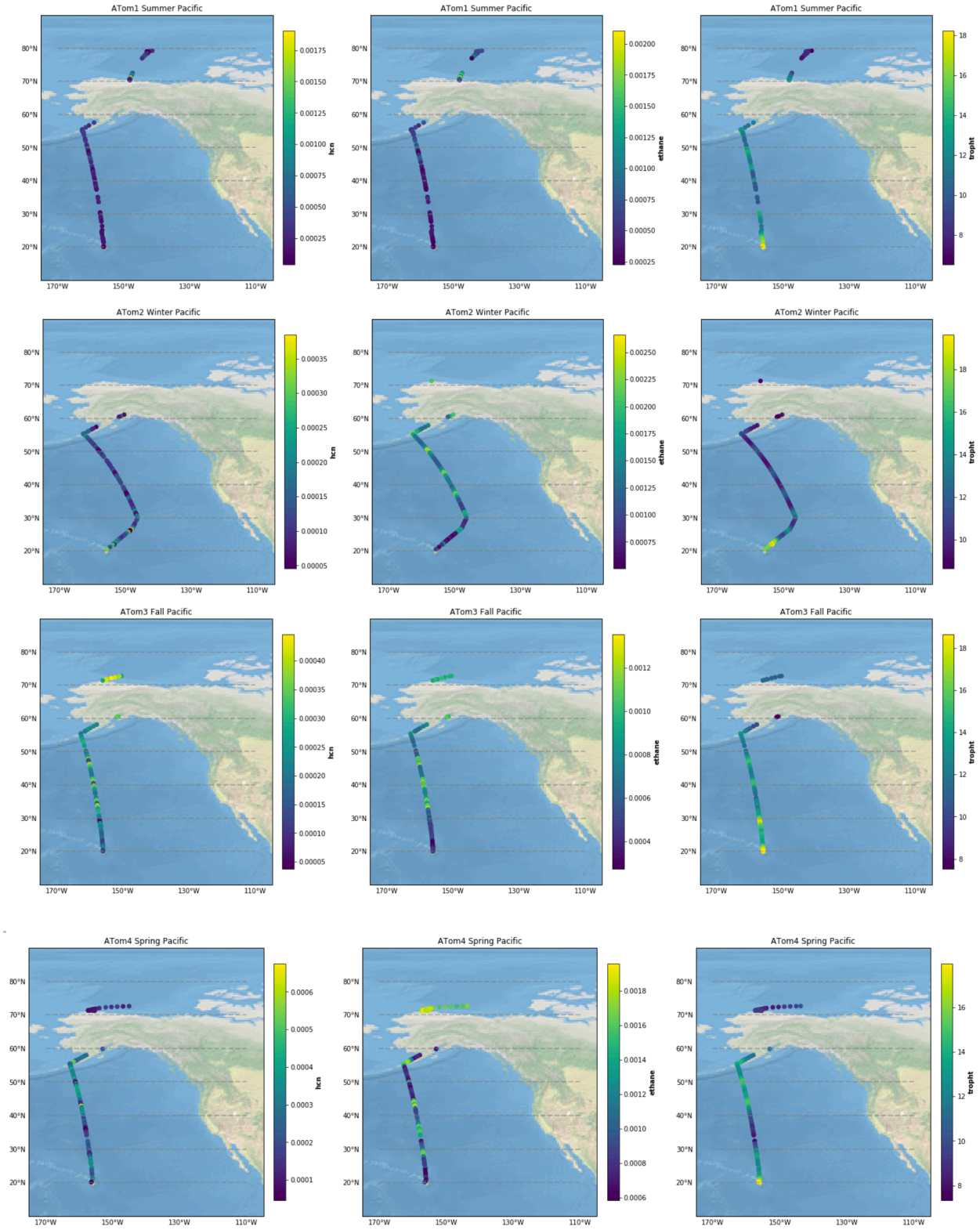


Figure S24. ATom HCN Pacific transects. HCN (left column), Ethane (middle column), and tropopause height (right column).

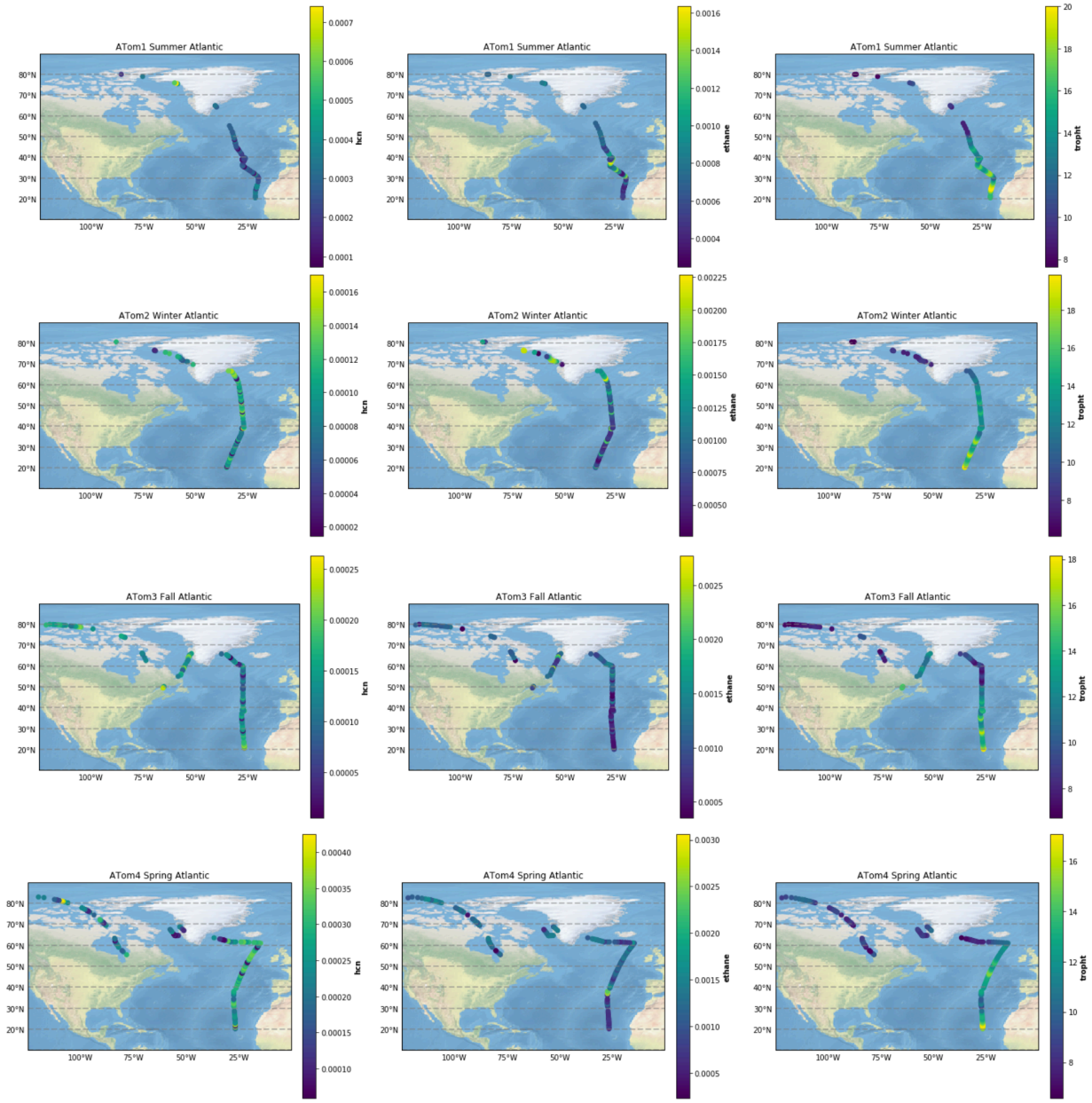


Figure S25. ATom HCN Atlantic transects. HCN (left column), Ethane (middle column), and tropopause height (right column).

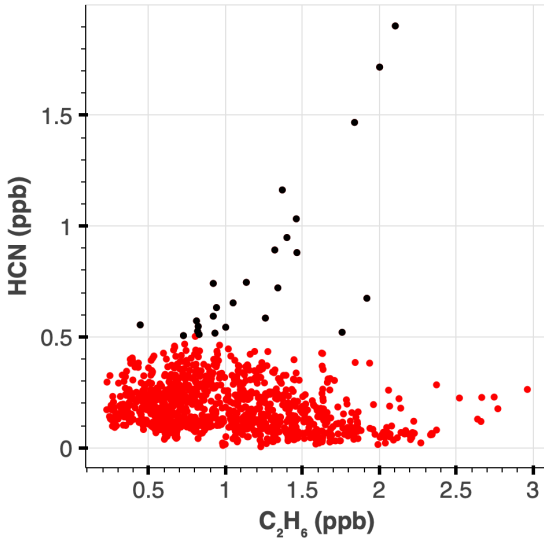


Figure S26. ATom HCN vs C_2H_6. This data includes all four ATom campaigns and ocean transects and has been filtered using the specifications outlined in the methods section in the main text. The few points with very high HCN and C_2H_6 are associated with biomass burning. We assigned a weak percentile score for each datapoint, and the values greater than or equal to equal to the 87th percentile are highlighted in black. Those points were replaced with NaN and then interpolated using the "backfill" method for C_3H_8 and C_2H_6.

4. GEOS-Chem Simulations

We simulated HIPPO and ATom measurements using the GEOS-Chem "classic" global 3-D chemical transport model in v12.1.1 (doi:10.5281/zenodo.2249246). The simulations were driven by MERRA-2 reanalysis meteorology product by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center.[22] MERRA-2 has a native resolution of 0.5 lat x 0.625 lon x 72 hybrid sigma/pressure levels, of which we degrade to a 4 x 5 x 72 resolution. We simulate time periods that encompass the HIPPO and ATom aircraft measurements with a 1-year spin-up period. In all cases, we use a standard chemistry simulation with no changes to the chemistry regimes. Simulations were collected over every hour over every day of each campaign period (~2 months of results for each campaign time period). Hydrocarbons were converted from units of carbon to $mol_x/mol_{dry\ air}$.

All emissions are computed using the Harmonized Emissions Component (HEMCO) Standalone[23] version 3.0.0 (DOI: 10.5281/zenodo.4429214)[24] with GEOS-Chem development version 13.0.0, cloned on 9/2020 at https://github.com/geoschem/geos-chem). This Standalone utilizes the most up-to-date versions of emissions as of September, 2020. As such, even though we utilize an older version of GEOS-Chem classic for the simulations, we implement up-to-date emissions. Relevant inventories that cover the oil and natural gas sector that are used in the default emissions configuration for GEOS-Chem v13.0.0 are Tzompa-Sosa et al. 2017 for C_2H_6, and Xiao et al. 2008 for C_3H_8.[4,25]

In Figure S27, we show that GEOS-Chem "synoptic replicates" (GEOS-Chem sampled several days before and after the aircraft in situ sampling time) show less consistency in latitude compared to the coordinate, potential temperature.

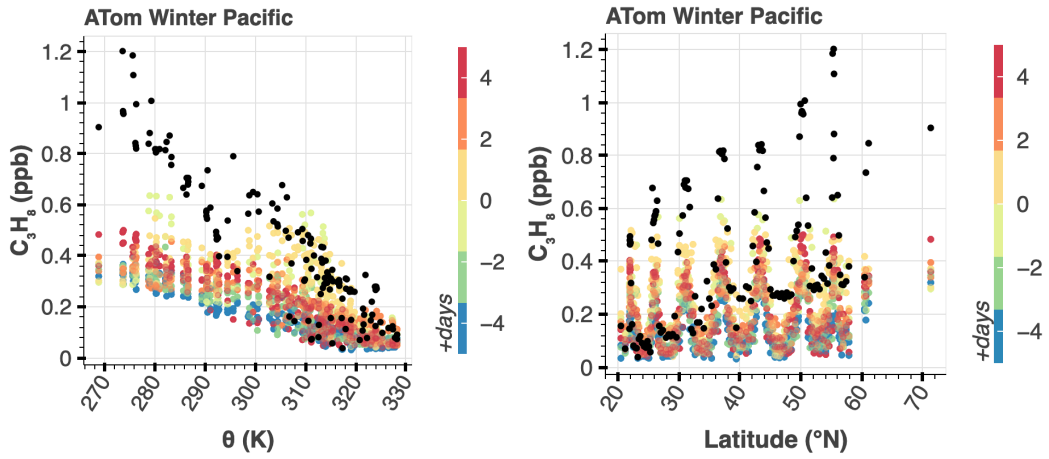
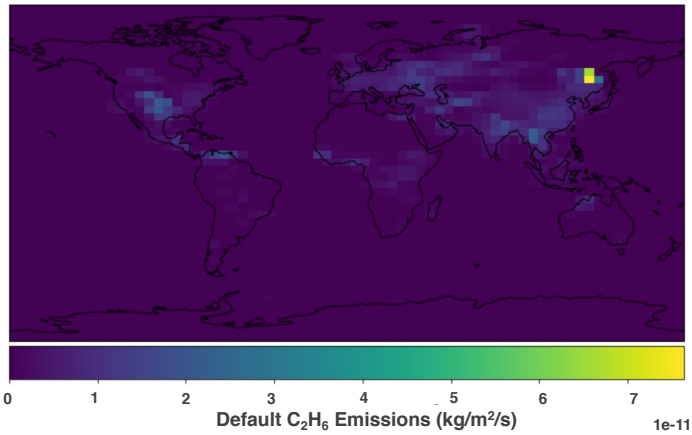


Figure S27. GEOS-Chem simulated C_2H_6 vs potential temperature and latitude during 2017. These data are analyzed for Pacific and Atlantic transects during January-Feburary 2017. Aircraft observations are shown in black. GEOS-Chem simulations are colored by the sampling time, in days beyond the flight path. (We found the median time of in-situ sampling of the aircraft, and then sampled the GEOS-Chem model for several days before and after the median to generate what we call "synoptic replicates" here. Each of the synoptic replicates were sampled along the flight path latitude, longitude, time and potential temperature using nearest neighbor interpolation.) In this figure we include $\pm$ 5 days to demonstrate the variance, but we use up to $\pm$ 2 days of the GEOS-Chem replicates in the Bayesian model. All remaining simulations of ATom and HIPPO C_2H_6 and C_3H_8 are included in the SI.

As discussed in Section 3.2 in the main text, we revise C_3H_8 emissions using GEOS-Chem v13.0.0 default C_2H_6 emissions scaled by the observed C_3H_8/C_2H_6 ratio estimated from the NOAA data. In Figure S28, we show the default C_2H_6 and C_3H_8 emissions, as well as the revised C_3H_8 emissions.



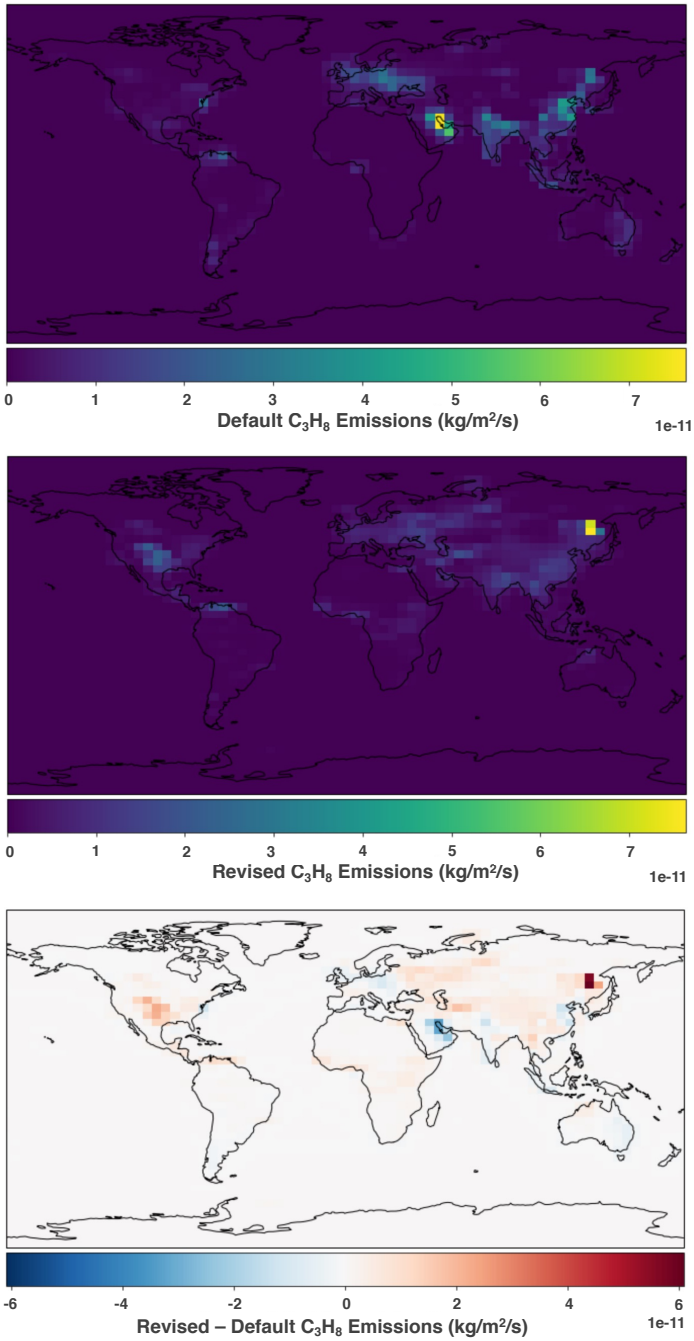


Figure S28. GEOS-Chem v13.0.0 default and revised emissions. Top: Default C_3H_8 emissions. **Upper middle**: Default C_2H_6 emissions. **Bottom middle**: Revised C_3H_8 emissions scaled by the observed NOAA C_3H_8/C_2H_6 ratio (0.67 mol/mol, Figure S9). **Bottom**: Difference between revised and default C3H8 emissions used by GEOS-Chem v13.0.0.

In the GEOS-Chem simulations, both ethane and propane have the most variability and lowest mole fraction, as expected, since their oxidative chemistry is much faster. During the summer, tropical intrusions with very low mixing ratios are prominent (see Atom summer Pacific transect, Figure S29.).

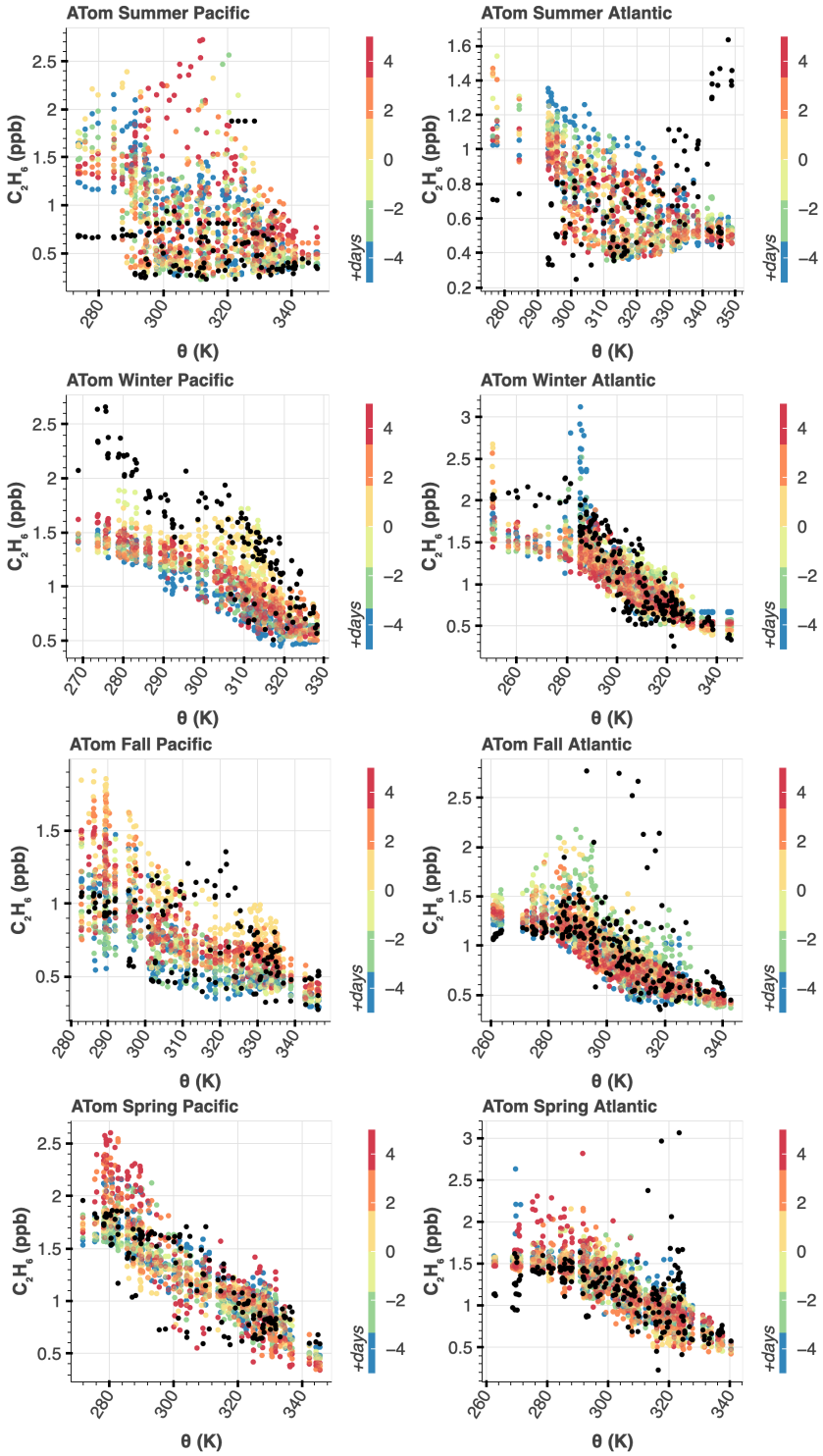


Figure S29. GEOS-Chem simulated C_2H_6 vs potential temperature during ATom campaign. Each plot is specific to the ocean transect. GEOS-Chem simulations are colored by the sampling time, in the number of days from the day of the flight. (We found the median time of in-situ sampling of the aircraft, and then sampled the GEOS-Chem model for several days before and after the median to generate "synoptic replicates". Each of the synoptic replicates were sampled along the flight path latitude, longitude, time of day and potential temperature using nearest neighbor interpolation.)

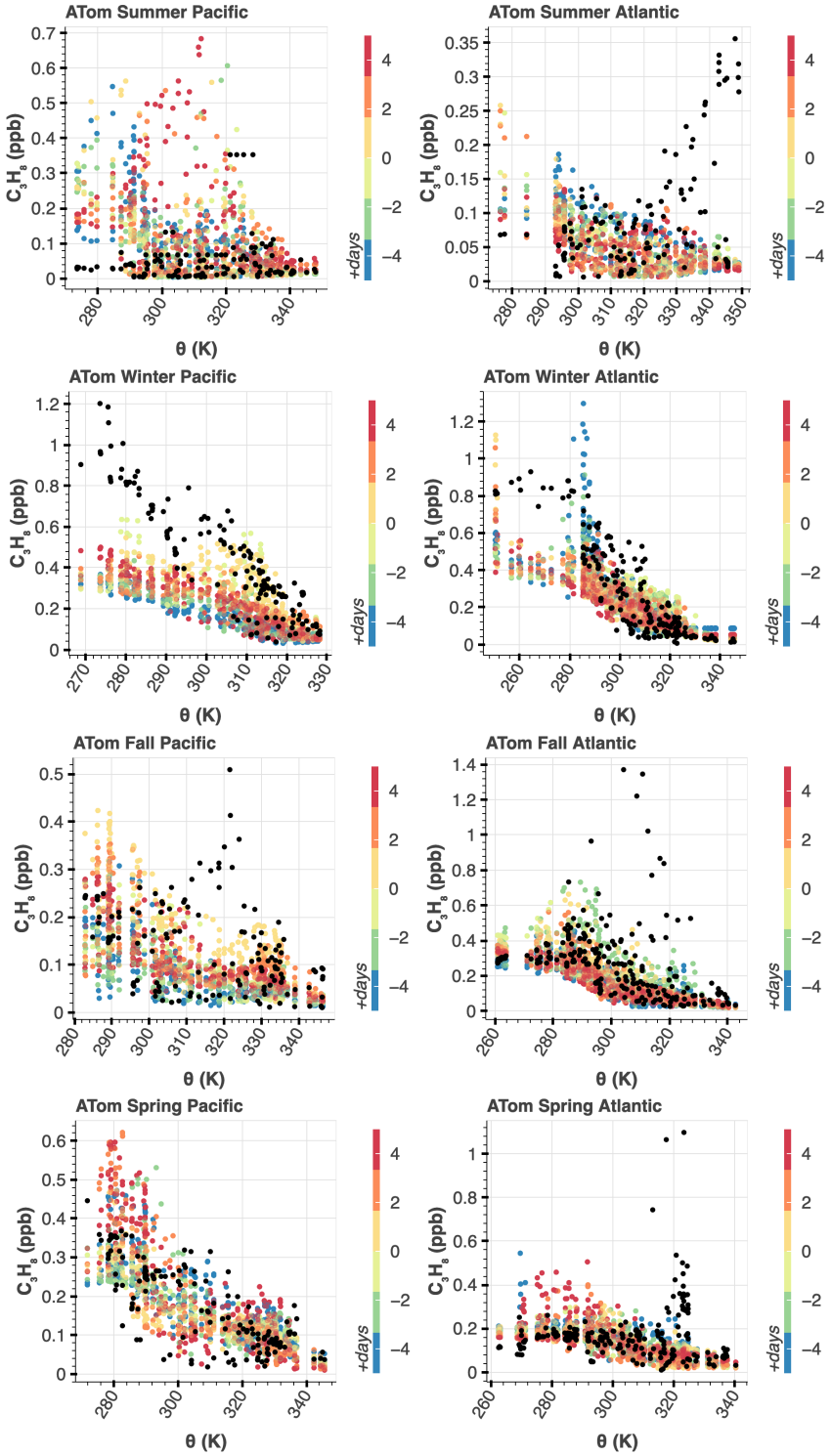


Figure S30. GEOS-Chem simulated C_3H_8 vs potential temperature during ATom campaign. Each plot is specific to the ocean transect. GEOS-Chem simulations are colored by the sampling time, in the number of days from the day of the flight. (We found the median time of in-situ sampling of the aircraft, and then sampled the GEOS-Chem model for several days before and after the median to generate "synoptic replicates". Each of the synoptic replicates were sampled along the flight path latitude, longitude, time of day and potential temperature using nearest neighbor interpolation.)

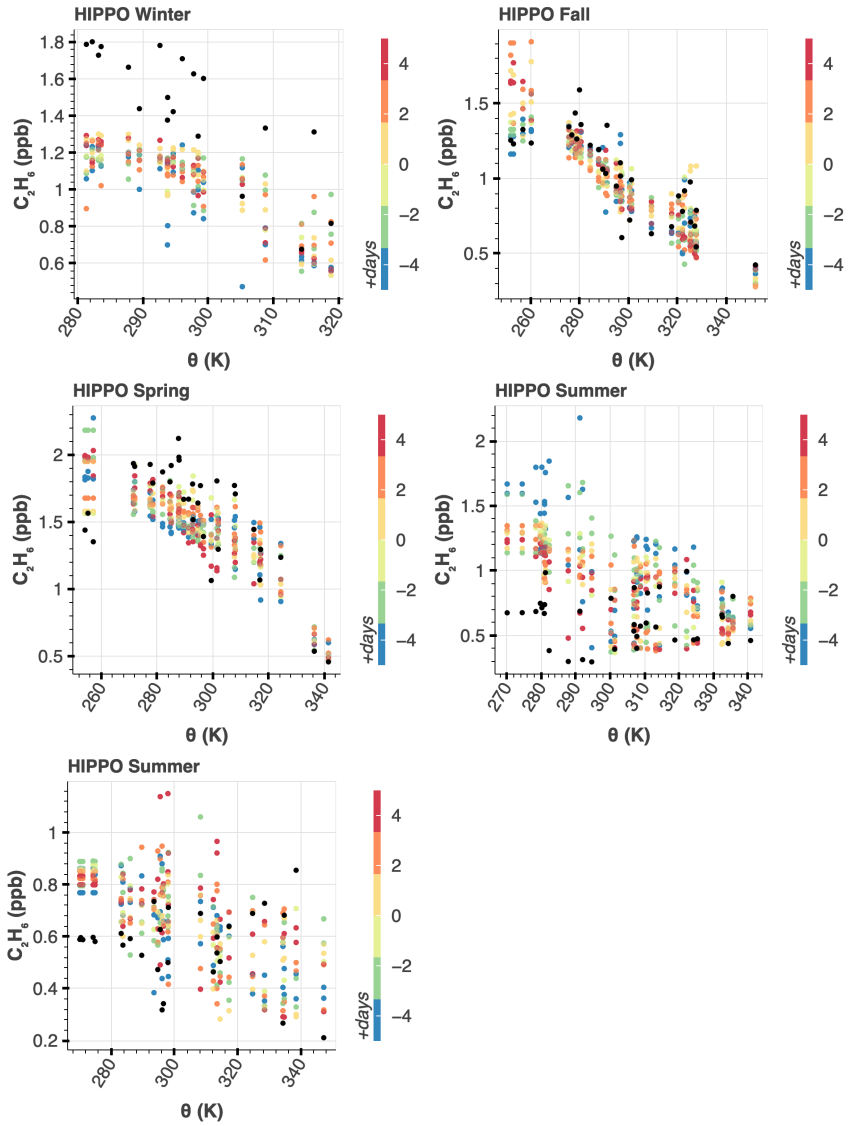


Figure S31. GEOS-Chem simulated C_2H_6 vs potential temperature during HIPPO aircraft campaign. Each plot is specific to the ocean transect. GEOS-Chem simulations are colored by the sampling time, in the number of days from the day of the flight. (We found the median time of in-situ sampling of the aircraft, and then sampled the GEOS-Chem model for several days before and after the median to generate "synoptic replicates". Each of the synoptic replicates were sampled along the flight path latitude, longitude, time of day and potential temperature using nearest neighbor interpolation.)

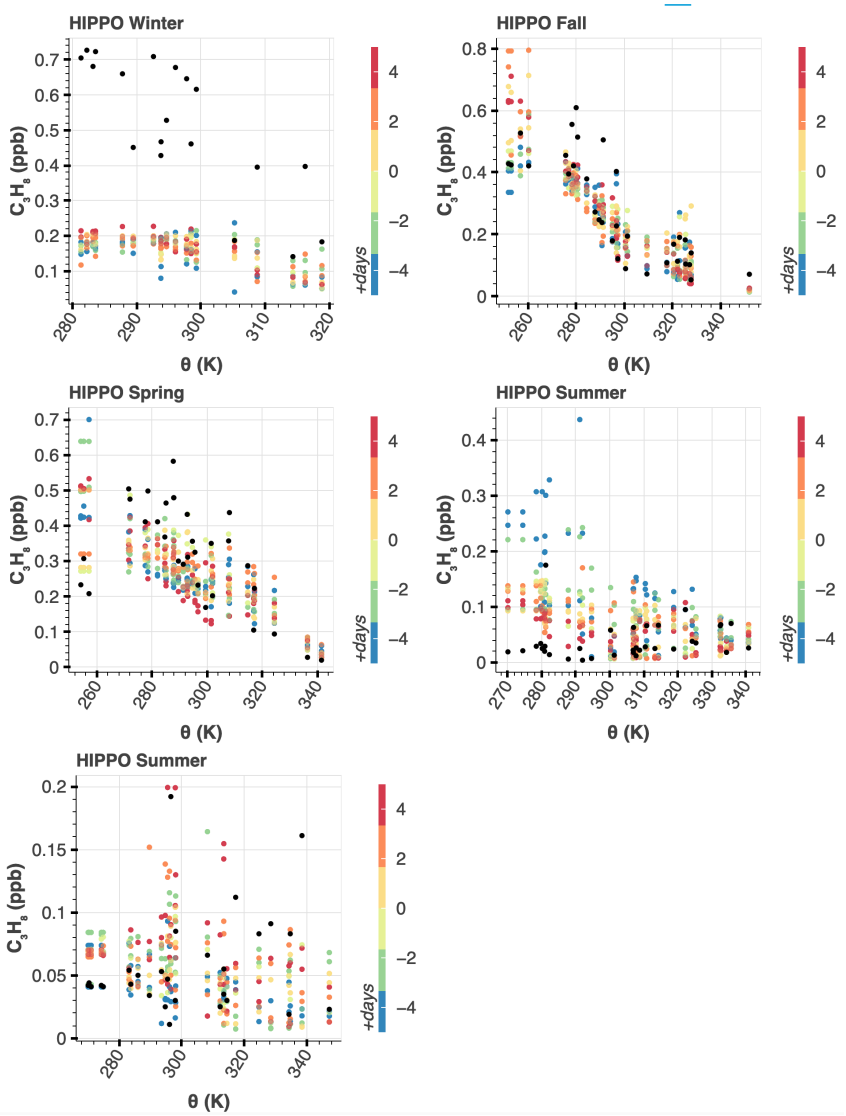


Figure S32. GEOS-Chem simulated C_3H_8 vs potential temperature during HIPPO aircraft campaign. Each plot is specific to the ocean transect. GEOS-Chem simulations are colored by the sampling time, in the number of days from the day of the flight. (We found the median time of in-situ sampling of the aircraft, and then sampled the GEOS-Chem model for several days before and after the median to generate "synoptic replicates". Each of the synoptic replicates were sampled along the flight path latitude, longitude, time of day and potential temperature using nearest neighbor interpolation.)

We compare the cross plot of C_3H_8 to C_2H_6 from the aircraft measurements and GEOS-Chem simulations to the NOAA measurements. As expected, both the aircraft and GEOS-Chem simulations fall under the photochemically aged emissions part of the NOAA curve. While the aircraft data overlays the NOAA measurements almost perfectly (especially in the winter when the lifetimes of both gases are longest), GEOS-Chem underestimates C_3H_8, particularly over the Atlantic curtain (Figure S33). The same conclusion is drawn for HIPPO time periods (Figure S 34).

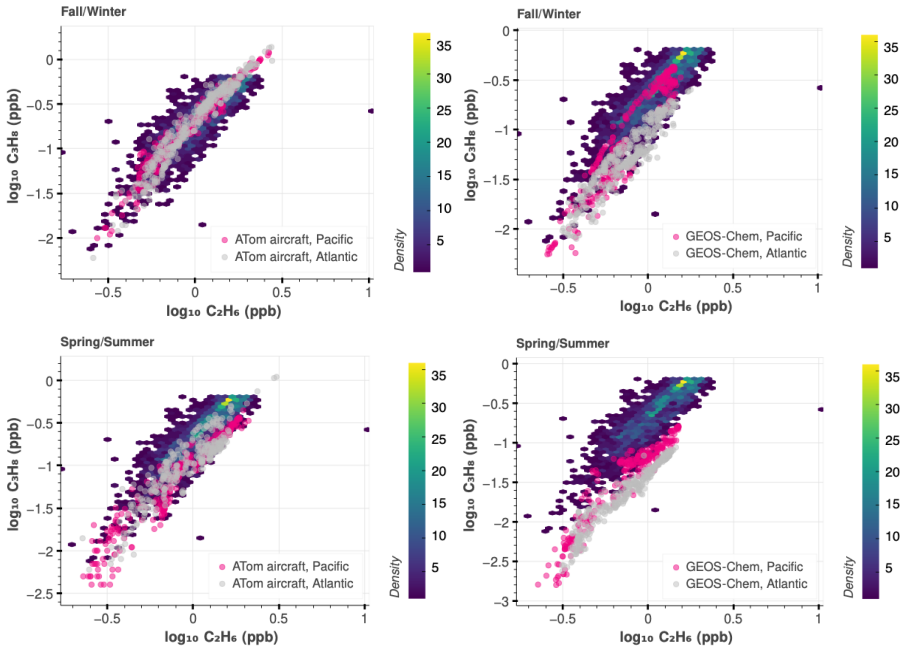


Figure S33. Comparison of C_3H_8 vs C_2H_6 for NOAA, ATom aircraft, and GEOS-Chem (GC) simulations. NOAA photochemically-aged measurements (all sites, 2005-2018), as explained in the text, are shown on the heat map (colored by density of data). Note the distinction between winter/fall and spring/summer seasons. HIPPO is included in the SI, Section 3.

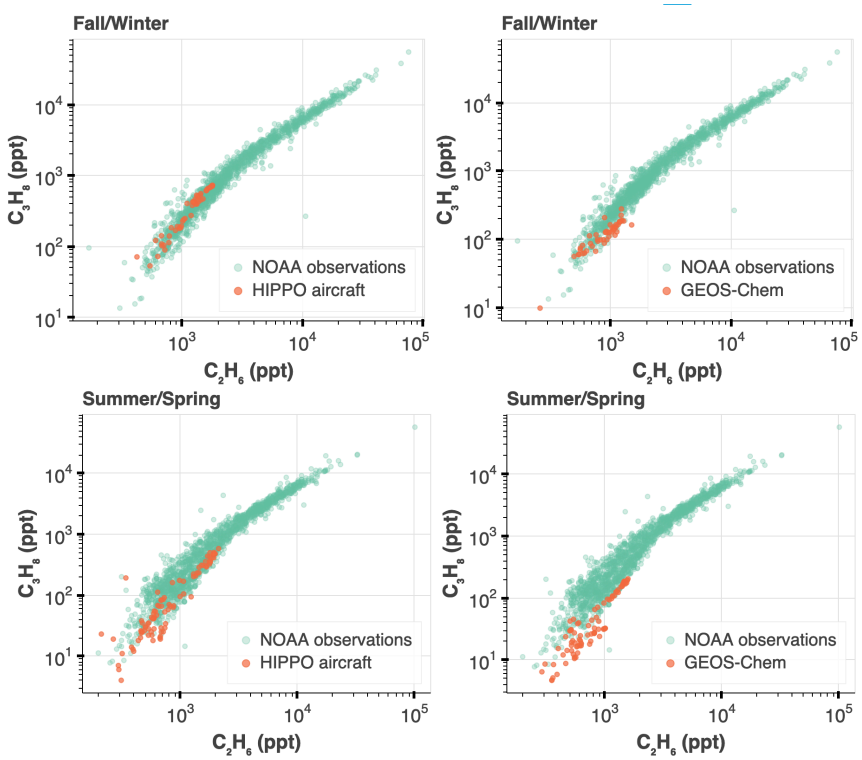


Figure S34. C_3H_8 vs C_2H_6 for HIPPO aircraft and GEOS-Chem simulations. Please see section 3.2 in the main text for a discussion.

In the main text, we show, in Figure 4, the impact of the revised C_3H_8 emissions on GEOS-Chem simulations during the ATom 4 campaign time period. Below, in Figure S35, we show the impact of the revised C_3H_8 emissions for all four ATom campaigns. In Figure S36, we show the impact of the revised C_3H_8 emissions during all 5 HIPPO campaigns.

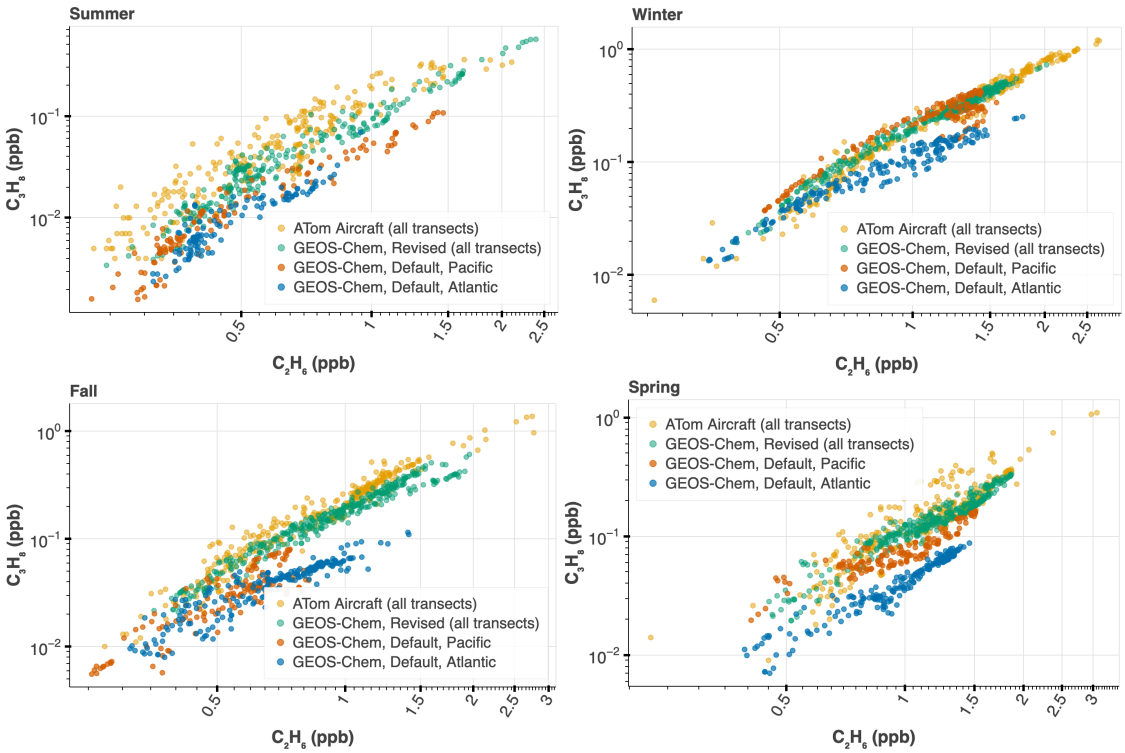


Figure S35. GEOS-Chem simulations using the default and revised C_3H_8 emissions during all four ATom campaigns. Please see section 3.2 in the main text for more discussion.

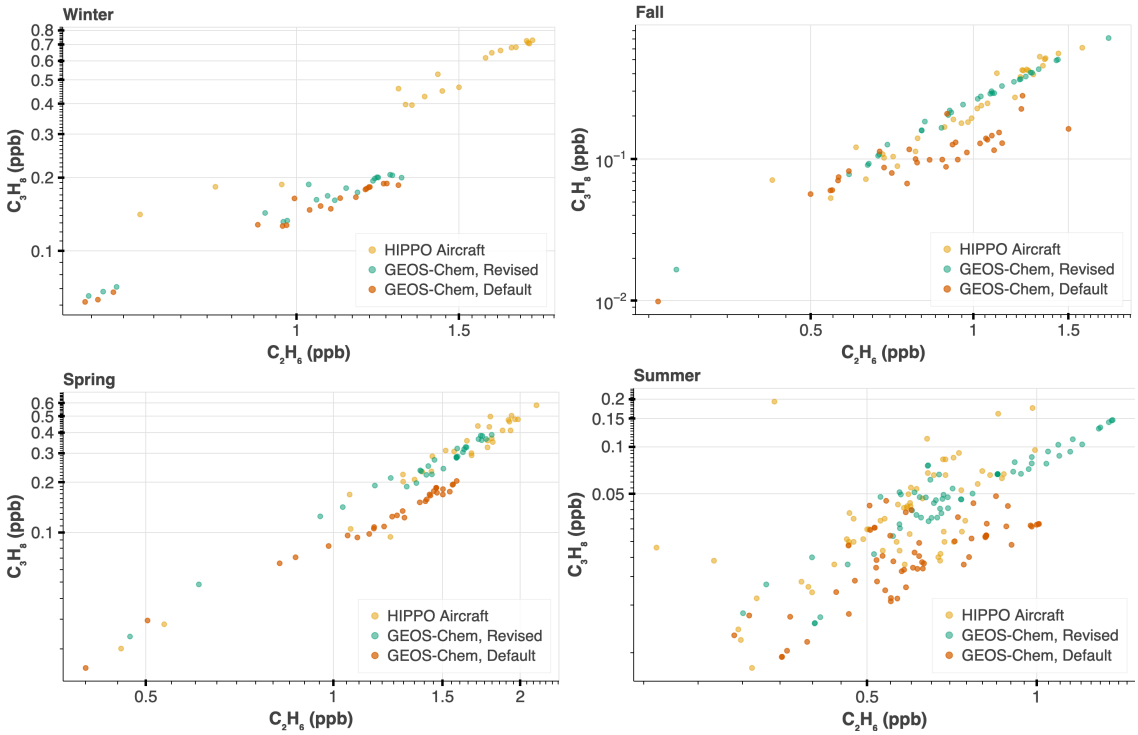


Figure S36. GEOS-Chem simulations using the default and revised C_3H_8 emissions during all 5 HIPPO campaigns. Please see section 3.2 in the main text for more discussion.

5. Bayesian Inference

5.1. Background and Priors

We wish to use the ATom and HIPPO aircraft observations to quantify C_3H_8 and C_2H_6 emissions. We use hierarchical Bayesian modeling to estimate what global scalar would minimize the difference between the simulated C_3H_8 and C_2H_6 from the updated GEOS-Chem v13.0.0 emissions and the observations made during the ATom and HIPPO aircraft campaigns. Using Bayesian probability, we can quantify a degree of certainty about a hypothesis or parameter value. Using probability rules, one can derive Bayes's Theorem:

$$P(\theta \mid y) = \frac{P(y \mid \theta)P(\theta)}{P(y)}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

The likelihood tells us how likely it is to acquire the observed data, y, given the parameter, θ. The prior is a measure of plausibility of the hypothesis θ before the experiment was conducted. The evidence is a marginal likelihood that is computed from the likelihood and the prior. The posterior contains the information we want about the parameters we are after. The ambient mole

fraction of C_3H_8 and C_2H_6 is more linearly related to its underlying emissions pattern during the winter/fall/spring when there is decreased sunlight/oxidation. As such, we assume differences between the GEOS-Chem simulations and the aircraft observations can largely be attributed to the underlying emissions grid, such that,

$$\mathrm{a} = \mathrm{gcs} \cdot \alpha \tag{1}$$

where a is the aircraft C_2H_6 or C_3H_8, gcs stands for GEOS-Chem simulation of C_2H_6 or C_3H_8, and α is a scalar that represents the most likely mismatch between the underlying C_3H_8 and C_2H_6 emissions as transported through GEOS-Chem v13.0.0 relative to ATom and HIPPO observations. This assumption forms the heart of our hierarchical Bayesian model.

We can reasonably approximate C_2H_6 and C_3H_8 measured by the aircraft to be Lognormally distributed with an approximate error. Lognormal distributions have longer tails, which is appropriate given the outliers we see in the measurements. We can model the GEOS-Chem simulated results as follows:

$$\begin{aligned} \mathrm{gcs}_j &\sim \mathrm{LogNorm}(\beta \cdot \mathrm{a}_j, \sigma) \\ \sigma &\sim \mathrm{Prior} \\ \beta &\sim \mathrm{Prior} \end{aligned} \tag{2}$$

Where gcs_j represents GEOS-Chem simulated C_3H_8 and C_2H_6, a_j represents the *j*th datum of aircraft-observed C_3H_8 or C_2H_6. Here, β parameter (equivalent to $1/\alpha$) estimates error in the default GEOS-Chem emissions, and σ is the approximate uncertainty in the GEOS-Chem simulations. In this case, we expect β to be less than one since the aircraft observations are usually greater than the GEOS-Chem simulations. We organized the Bayesian model this way because we consider the aircraft observations to be unchanging, while treating the GEOS-Chem simulations as the experimental dataset. To obtain the GEOS-Chem missing emissions, we can invert the β parameter.

We sampled the GEOS-Chem model several days before and after the aircraft path to estimate uncertainty in the simulations due to meteorology, as explained in Section 2.3 in the main text. If we were to pool all the data together, each experiment would be governed by identical parameters. However, each replicate is subject to differences mainly due to meteorology and we conclude that the parameters in each replicate experiment should vary from one another, such that we have i separate models to fit, each looking like equations 2, above. Under this scenario, we organize our model into a hierarchical structure, pictured in Figure S37.

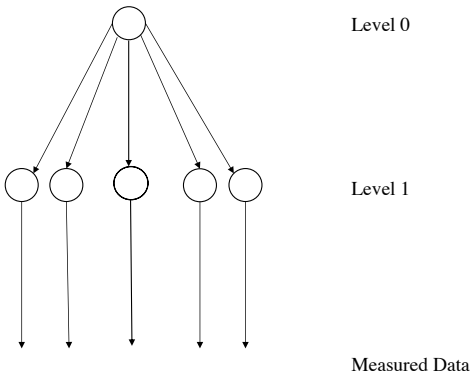


Figure S37. Schematic of hierarchical Bayesian model. Level 0 contains the hyperparameter β, the parameter we ultimately wish to get estimates for. Level 1 corresponds to the day the GEOS-Chem model was sampled (there are 5 days because we sampled 2 days before and after the mean flight path). There will be variability from day to day, and the location and scale parameters for a given day are conditioned on the hyperparameters.

We can consider a hierarchical model in which there is a hyperparameter, which we call β (corresponds to level 0, Figure S32), and the values of the scaling parameters of the replicates, which we now call β_1 (corresponds to level 1, Figure S37), may vary from this β according to some probability distribution, $g(\beta_{1,i}|\beta)$. We now have parameters $\beta_{1,1}, \beta_{1,2}, \ldots\ \beta_{1,i}$ and β. The posterior can be written using Bayes' theorem, defining $\beta_1 = (\beta_{1,1}, \beta_{1,2}, \ldots)$,

$$g(\beta, \beta_1|\mathrm{a}, \mathrm{gcs}) = \frac{f(\mathrm{a}, \mathrm{gcs}|\beta, \beta_1)g(\beta, \beta_1)}{f(\mathrm{a}, \mathrm{gcs})} \tag{3}$$

Note though that the observed values of gcs do not directly depend on β, only on β_1 and as such, the observations are only indirectly dependent on β. So, we can write:

$$g(\beta, \beta_1|\mathrm{a}, \mathrm{gcs}) = \frac{f(\mathrm{a}, \mathrm{gcs}|\beta_1)g(\beta, \beta_1)}{f(\mathrm{a}, \mathrm{gcs})} \tag{4}$$

Next, we can rewrite the prior using the definition of conditional probability:

$$g(\beta, \beta_1) = g(\beta_1|\beta)g(\beta) \tag{5}$$

Substituting this back into the previous expression for the posterior, we have:

$$g(\beta, \beta_1|\mathrm{a}, \mathrm{gcs}) = \frac{f(\mathrm{a}, \mathrm{gcs}|\beta_1)g(\beta_1|\beta)g(\beta)}{f(\mathrm{a}, \mathrm{gcs})} \tag{6}$$

In the numerator, we see a chain of dependencies. The gcs simulations depend on β_1. Parameters β_1 depend on hyperparameter β. Hyperparameter β then has some hyperprior distribution. As such, this hierarchical model captures both the sample day-to-day variability, as well as the hyperparameter.

We must specify a hyperprior, and a conditional prior, $g(\beta_1|\beta)$. Here, we have no reason to believe that we can distinguish any one $\beta_{1,i}$ from another prior to the experiment. As such, we can assume the conditional prior to behave in an exchangeable manner, where the label i is not dependent on the permutations of the indices. Our expression for the posterior is:

$$g(\beta_1, \beta \mid \mathrm{a}, \mathrm{gcs}) = \frac{f(\mathrm{a}, \mathrm{gcs} \mid \beta_1)\left(\prod_{i=1}^{k} g(\beta_{1,i} \mid \beta)\right) g(\beta)}{f(\mathrm{a}, \mathrm{gcs})} \tag{7}$$

The full hierarchical model is given by one additional level above Level 0 in Figure 37 that corresponds to individual aircraft campaign/season and ocean transect. This additional level contains the parameter α, the overall parameter that is a result of the sampling replicates and the season/ocean transect. It is very difficult to sample the full hierarchical model with so many levels, and for practical reasons, we were not able to. However, given that each season/ocean transect is assumed to be independent, we can treat each season/ocean transect with a separate hierarchical model (that is shown in Figure 37), and then sample the posterior samples of those separate models to define a credible interval for overall α. Our statistical model is defined for campaign/transect, i, and the observed mole fraction, j, as follows:

$$\begin{gathered}
\tau_{ij} = 0.05|\Delta t|_{ij} + 0.01 \\
\beta_i \sim \mathrm{Norm}(0.7, 0.2) \\
\beta_{1,ij} \sim \mathrm{Norm}(\beta_i, \tau_{ij}) \\
\alpha_i = 1/\beta_i \\
\sigma_{ij} = 0.14 \cdot \mathrm{tropht}_{ij} + 0.8 \\
\mathrm{gcs}_{ij} \sim \mathrm{LogNorm}(\beta_{1,ij} \cdot \mathrm{a}_{ij}, \sigma_{ij})
\end{gathered} \tag{8}$$

The likelihood is given by gcs_{ij}, which represents the *ij*th mole fraction of GEOS-Chem simulated C_3H_8 and C_2H_6, and a_{ij} represents the *ij*th C_3H_8 or C_2H_6 mole fraction observed by the aircraft. The uncertainty in gcs_{ij} is given by σ_{ij}, which increases linearly with tropopause height, tropht_{ij}, since we expect more variability in C_3H_8 or C_2H_6 mole fraction with high tropopause height that is often related with tropical intrusions. The conditional parameter, $\beta_{1,ij}$, depends on a hyperprior distribution for the emissions scalar, β_i (which is equivalent to $1/\alpha_i$), and τ_{ij}, which describes variability in the emissions scalar due to transport errors in GEOS-Chem. As such, τ_{ij} depends on the difference between the original aircraft sampling time and the GEOS-Chem sampling replicate time (Δt_{ij}) and increases with deviation between them. Development of this hierarchical model and our process for selecting priors are included in the SI.

We assume each α_i parameter from individual aircraft campaign season and ocean transect to be independent of one another. As such, to estimate a credible interval for an overall α, we draw a random sample of the posterior of hyperparameter α_i for each campaign season and ocean transect. We take the mean of these samples and repeat this 10,000 times. (Note that we do not use the summer estimates for this calculation, for reasons described in Section 2.2, main text.) Details on the software used are described below.

The τ_{ij} parameter has the effect of weighting the GEOS-Chem simulation replicate that falls on the plane path higher than the other replicates that do not. As such, GEOS-Chem simulation replicate that falls on the exact aircraft data collection time has more influence on the final result of β_i parameter (the emission scalar). We would not expect meteorology to cause more than $\pm$ 0.5 variation in parameter $\beta_{1,ij}$, as this would imply a very large variation in Tg after scaling emissions. As such, we vary τ_{ij} from 0.05 (the plane path day), to 0.33 for the furthest day from the plane path ($\pm$ 2 days), (Figure S38).

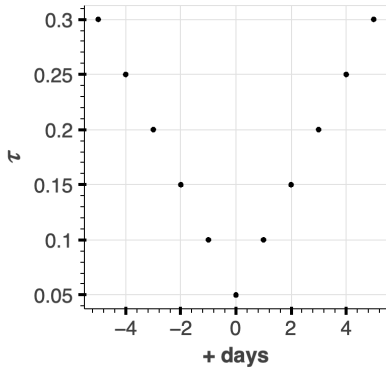


Figure S38. Prior for τ_{ij} parameter for the GEOS-Chem simulations in the Bayesian hierarchical model. The x-axis corresponds to the GEOS-Chem sample replicate, in units of days above or below the aircraft path (day 0).

We define σ_{ij} parameter as a function of tropopause height. As discussed earlier, tropical air masses are characterized by very low mole fractions of C_3H_8 and C_2H_6 because in the tropics they have short lifetimes relative to transport. We assign higher σ_{ij} (lower weight) to GEOS-Chem simulations that have higher tropopause height, which extends the width of the lognormal distribution likelihood for those measurements. This de-weights the samples that have more tropical influence. (In practice, the model is quite robust against changes in σ_{ij}, so this implementation has a small impact – see results of the sampling the posterior in Section 4.7). We somewhat arbitrarily define $\sigma_{ij} = 0.14 \cdot \mathrm{tropht}_{ij} + 0.8$, where tropht_{ij} is the tropopause height in km associated for the *j*th aircraft observation. This equation results in σ_{ij} typically ranging from 1- 1.5 and implementing this range as the variance in a lognormal distribution centered at $1x10^{-3}$ or $1x10^{-4}$ (typical of C_2H_6 or C_3H_8 measured mole fractions in remote atmospheres) results in a broad distribution that is consistent with mole fractions that we would expect for short-lived gases in remote atmospheres. The 1st and 99th percentiles of the resulting C_3H_8 distribution are $3x10^{-6}$ and $3x10^{-3}$ respectively, which is two orders of magnitude below and one order of magnitude above an average C_3H_8 mole fraction we would expect in the remote atmosphere. We find similar results for C_2H_6, except that it is centered at a value that is one

magnitude larger. We use the same σ_{ij} when modeling both C_3H_8 and C_2H_6. An example of the value of σ_{ij} for ATom 2 observations over the Atlantic is shown in Figure S39.

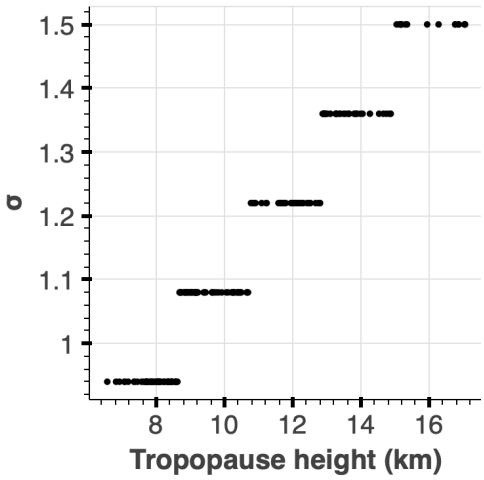


Figure S39. Prior for the σ_{ij} parameter for ATom 2 observations in the Bayesian hierarchical model.

We expect the β_i to be less than 1 since GEOS-Chem typically underestimates the aircraft observations. The prior for β_i is weakly informative, centered at 0.7 with the 1st percentile at 0.23 and 99th percentile at 1.17 (Figure S40).

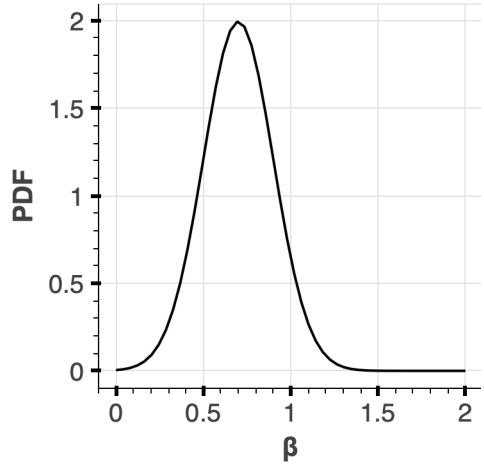


Figure S40. Prior distribution for β_i parameter in the Bayesian hierarchical model.

We performed prior predictive checks[26] to visualize the data our Bayesian model would generate given our priors. This check includes drawing parameter values from the prior distributions, plugging those parameters into the likelihood to generate pseudo data, and saving those data. These simulations gave us insight as to whether this was an appropriate model given our prior knowledge. The results of are satisfactory, as the empirical cumulative distribution functions are within what we expect given our prior knowledge. The results are shown in Figure S41.

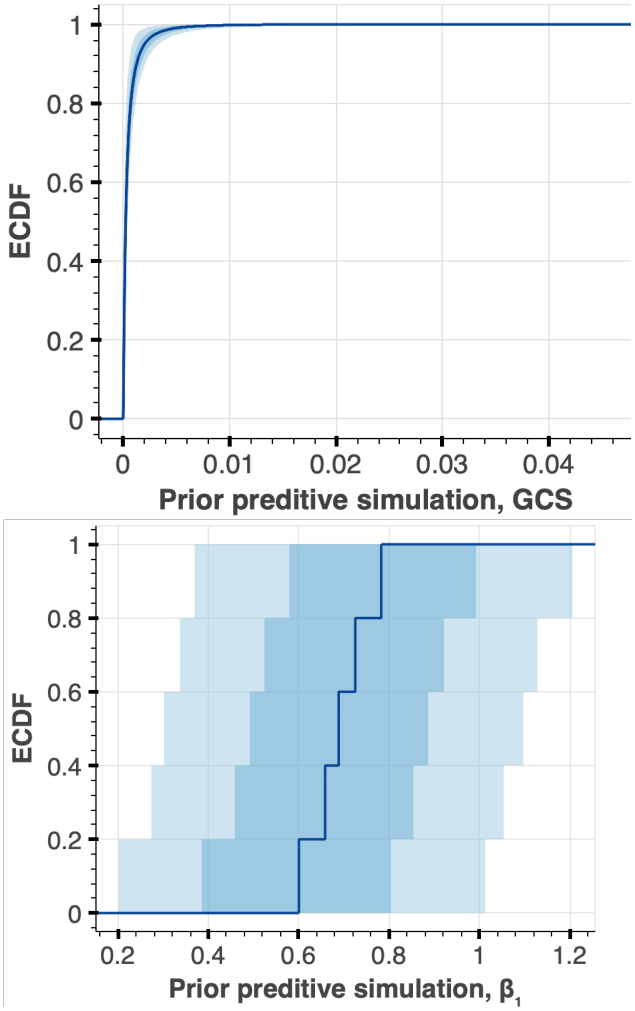


Figure S41. Prior predictive checks during ATom 2, Atlantic curtain. Top: Empirical cumulative distribution function of pseudo data of GEOS-Chem simulations given our priors. Bottom: Empirical cumulative distribution function of pseudo data of the $\beta_{1,ij}$ parameter given our priors.

We run our statistical model using Stan software[27] (version 2.26) with CmdStanPy Python interface (version 0.9.67)[28]. We parse Markov chain sampling using ArviZ (version 0.11.1).[29] We validate our hierarchical model using simulation-based calibration,[30] and posterior predictive checks[26] (described more below). We use bebi103 package (version 0.1.0)[31] to execute simulation-based calibration, prepare data for Stan sampling, parse MCMC samples, plot posteriors and plot posterior predictive checks. We also use iqplot (version 0.1.6)[32] to visualize empirical cumulative distribution functions of our priors. Finally, other software we use in our analysis includes Holoviews version 1.14.5,[33] Bokeh version 2.3.3,[34] Pandas version 1.3.1,[35] SciPy version 1.6.2,[36] and NumPy version 1.20.3.[37]

5.2. Simulation based calibration

Often, the posterior distribution is impossible to calculate analytically. Markov chain Monte Carlo (MCMC) allows us to sample out of an arbitrary probability distribution, where the probability of choosing a given value of a parameter is proportional to the posterior probability or probability density. Here, we use Stan to sample the posterior. Stan is a free, open source,

state-of-the-art probabilistic programming language that has interfaces for many other programming languages. Stan translates the model into C++, which is then compiled into machine code. It uses Hamiltonian Monte Carlo (HMC),[26] which allows for more efficient sampling of the posterior by taking large step sizes while taking into account the shape of the target distribution and tracing trajectories along it. We use CmdStanPy to install Stan, version 0.9.67. We use the bebi103 package[31] to execute simulation-based calibration.

Simulation-based calibration[30] consists of the following general steps: 1) Draw a parameter set $\tilde{\theta}$ out of the prior; 2) Use $\tilde{\theta}$ to draw a data set $\tilde{y}$ out of the likelihood; 3) Perform HMC sampling of the posterior using $\tilde{y}$ as if it were the actual measured data set, and draw L HMC samples of the parameters; 4) Do steps 1-3 N times, on order of N = 1000. In step 3, we are using a data set for which we know the underlying parameters that generated it. Because the data were generated using $\tilde{\theta}$ as the parameter set, $\tilde{\theta}$ is now the ground truth parameter set. As such, we can check to see if we uncover the ground truth in the posterior sampling by calculating the z-score. We can also check whether the posterior is narrower than the prior (shrinkage), indicating that the data are informing the model. We compute a z-score for each parameter, θ_i, which measures how close the mean sampled parameter value is to the ground truth, relative to the posterior uncertainty in the parameter value:

$$z_i = \frac{\langle \theta_i \rangle_{post} - \tilde{\theta}_i}{\sigma_{i,post}} \tag{9}$$

Here, $\langle \theta_i \rangle_{post}$ is the average value of θ_i over all posterior samples, and $\sigma_{i,post}$ is the standard deviation of θ_i over all posterior samples. The z-score should be symmetric about zero to indicate that there is no bias in estimating the ground truth, and should have a magnitude less than 5.[30] Our z-score calculations are satisfactory, shown in Figure S42.

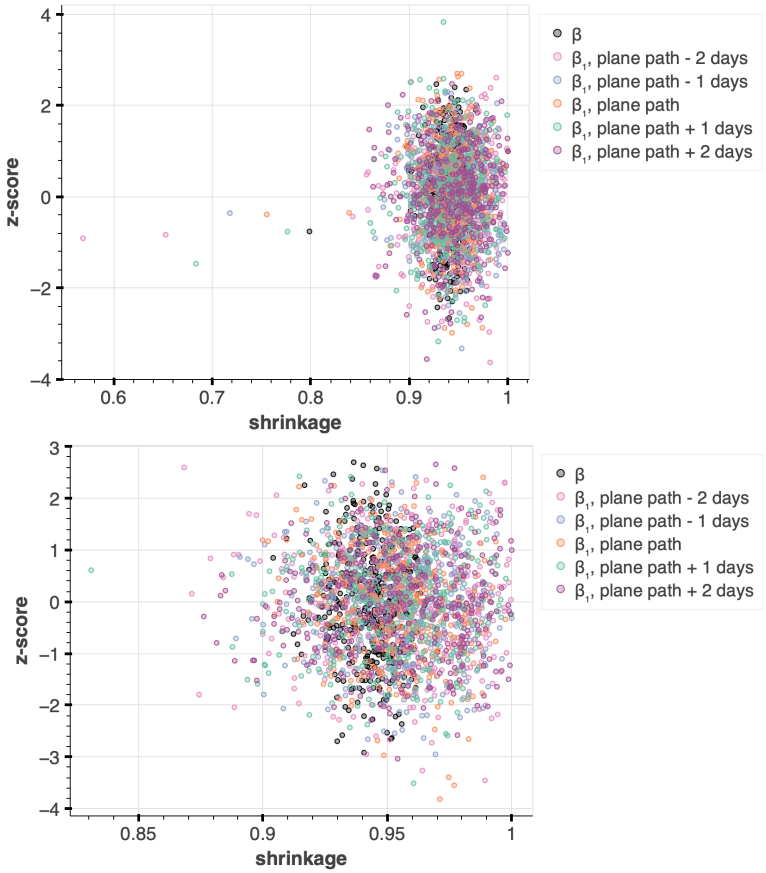


Figure S42. Z-score and shrinkage. Top: C_2H_6 during ATom 2 Atlantic curtain. Bottom: C_3H_8 during ATom 4 Atlantic curtain. Satisfactory z-score is symmetric about zero with a magnitude less than 5, while shrinkage should be around 1.

5.3. Posterior samples – ATom observations

Below are our results for our HMC sampling of the posterior (Figure S43 and S44). The posterior of hierarchical models inherently has regions of high curvature, which can cause difficulties for HMC sampling. If HMC trajectories veer sharply due to this curvature, the Monte Carlo step ends in a divergence. We decreased the step size of the sampler to sample the areas of high curvature (increased the adapt_delta parameter to 0.99 in Stan). To further reduce problems with high curvature, we implemented a non-centered parametrization of $\beta_{1,ij}$. We also set the warmup iterations to 2000 and conducted 1000 samples. Using bebi103's stan.check_all_diagnostics() function, our sampling had effective sample size for all parameters (based on the suggestion of 50 effective samples per split chain):[30] 0 out of 4000 iterations ended with a divergence or saturated the maximum tree depth and the energy-Bayes fraction of missing information indicated no pathological behavior. We achieved these diagnostics for all runs of ATom and HIPPO.

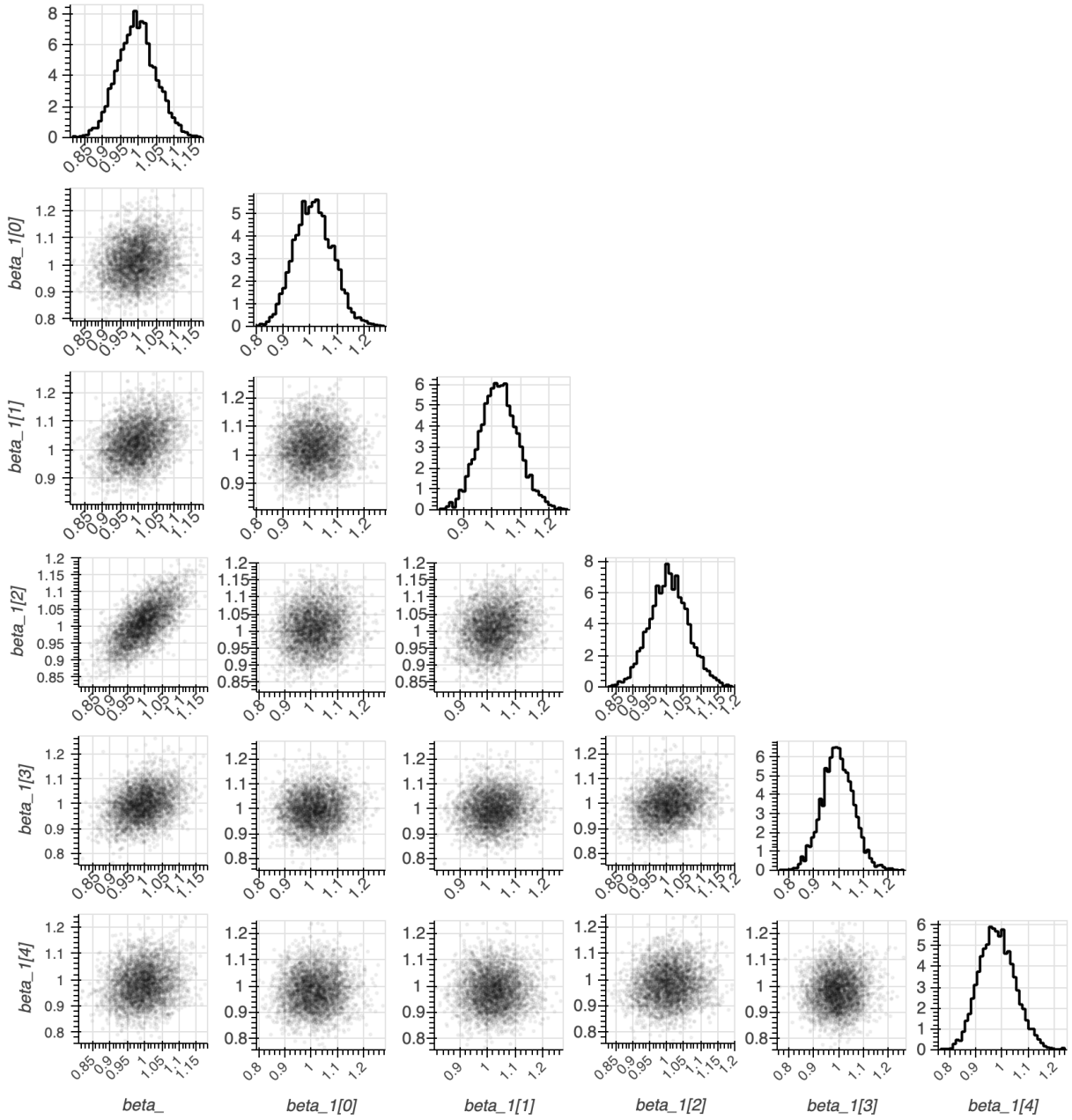


Figure S43. HMC Posterior samples for $\boldsymbol{\beta_{1,ij}}$ and $\boldsymbol{\beta_i}$ parameters using C_2H_6 ATom 4 aircraft and GEOS-Chem simulations. Beta_1 parameter is a vector of length 5, corresponding to the synoptic replicates of GEOS-Chem. Beta_1[0] and beta_1[1] correspond to 2 days before the aircraft, beta_1[2] is the plane path, and beta_1[3] and beta_1[4] correspond to 2 days after the aircraft. The hyperparameter β_i is represented by beta_.

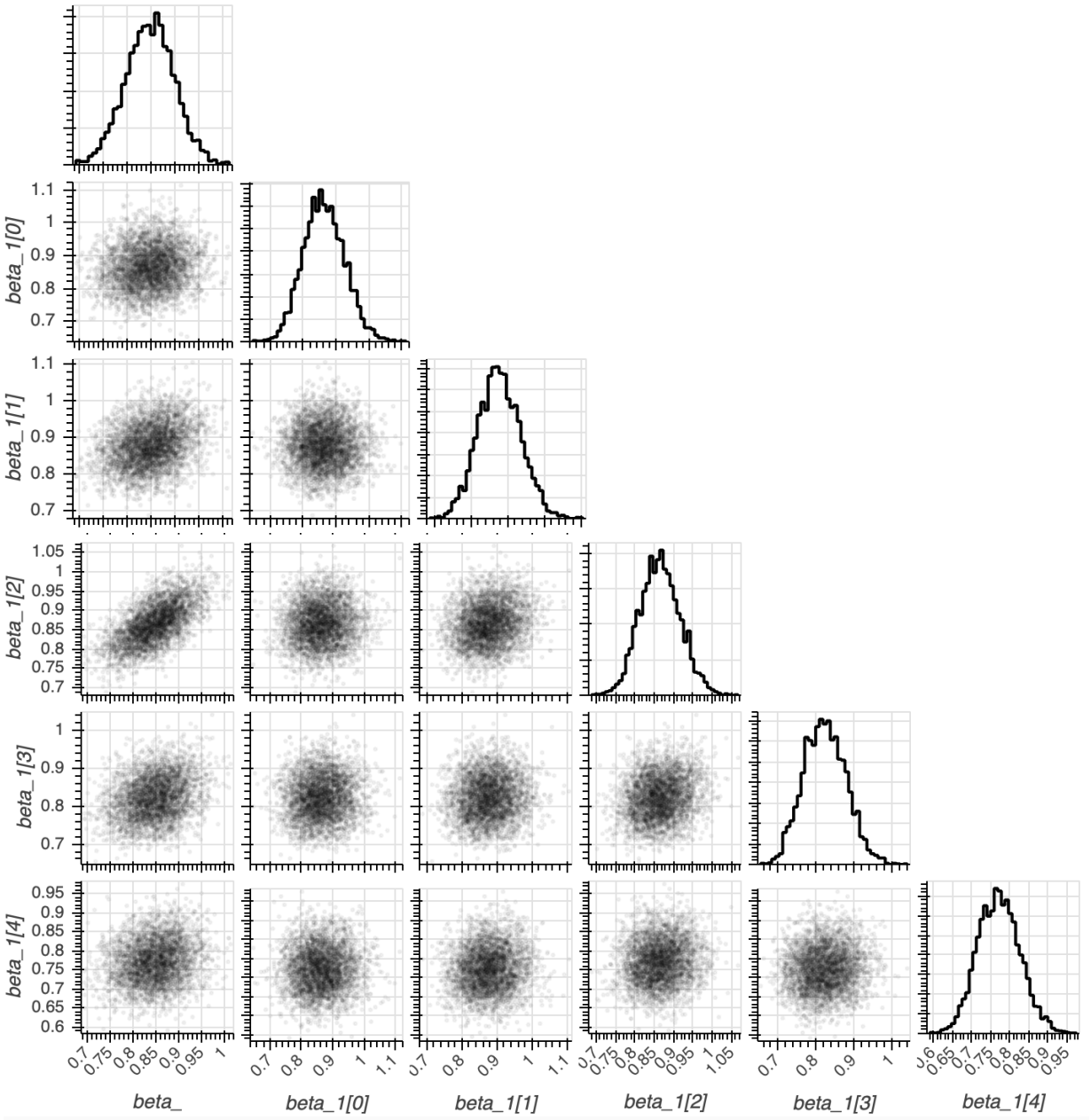


Figure S44. HMC Posterior samples for $\boldsymbol{\beta_{1,ij}}$ and $\boldsymbol{\beta_i}$ parameters using C_3H_8 ATom 4 aircraft and GEOS-Chem simulations. Beta_1 parameter is a vector of length 5, corresponding to the synoptic replicates of GEOS-Chem. Beta_1[0] and beta_1[1] correspond to 2 days before the aircraft, beta_1[2] is the plane path, and beta_1[3] and beta_1[4] correspond to 2 days after the aircraft. The hyperparameter, β_i, is represented by beta_.

In Figure S45, we show a cross plot of our hyperparameter, β_i and $\beta_{1,ij}$, in inverse form, for a single ATom campaign, which directly corresponds to the scaling of our adjusted default emissions under the GEOS-Chem v13.0.0 simulations. This shows an example of the variability due to GEOS-Chem meteorology compared to the hyperparameter that we use to scale the emissions. In Figure S46, we show the results of the Bayesian model for the α_i emissions scalar estimate for all 4 ATom campaigns and ocean transects.

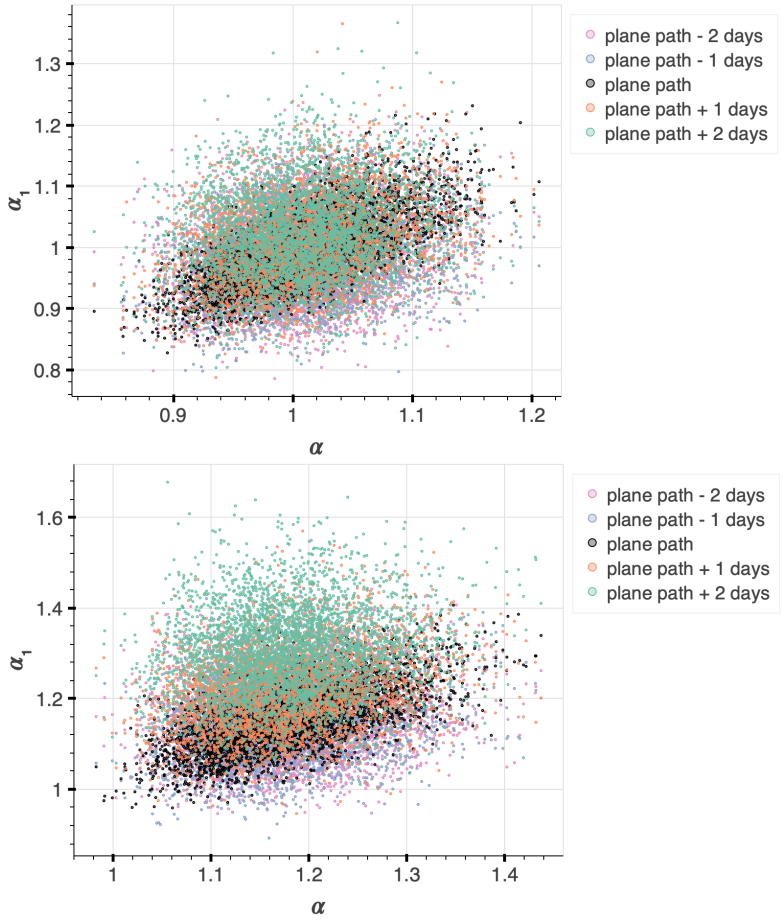


Figure S45. Posterior samples of $\alpha_{1,ij}$ vs α_i during ATom 4. Top: C_2H_6 observations. Bottom: C_3H_8 observations.

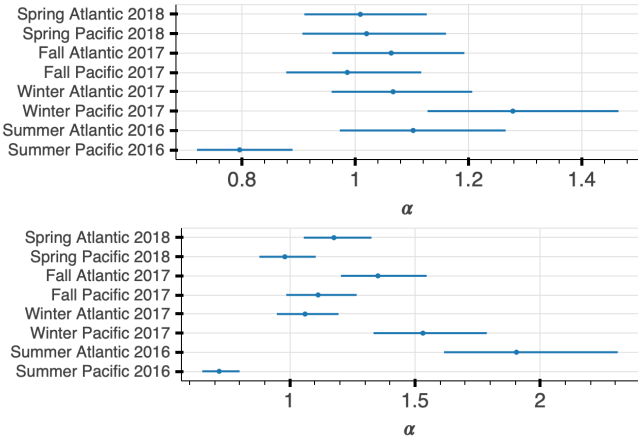
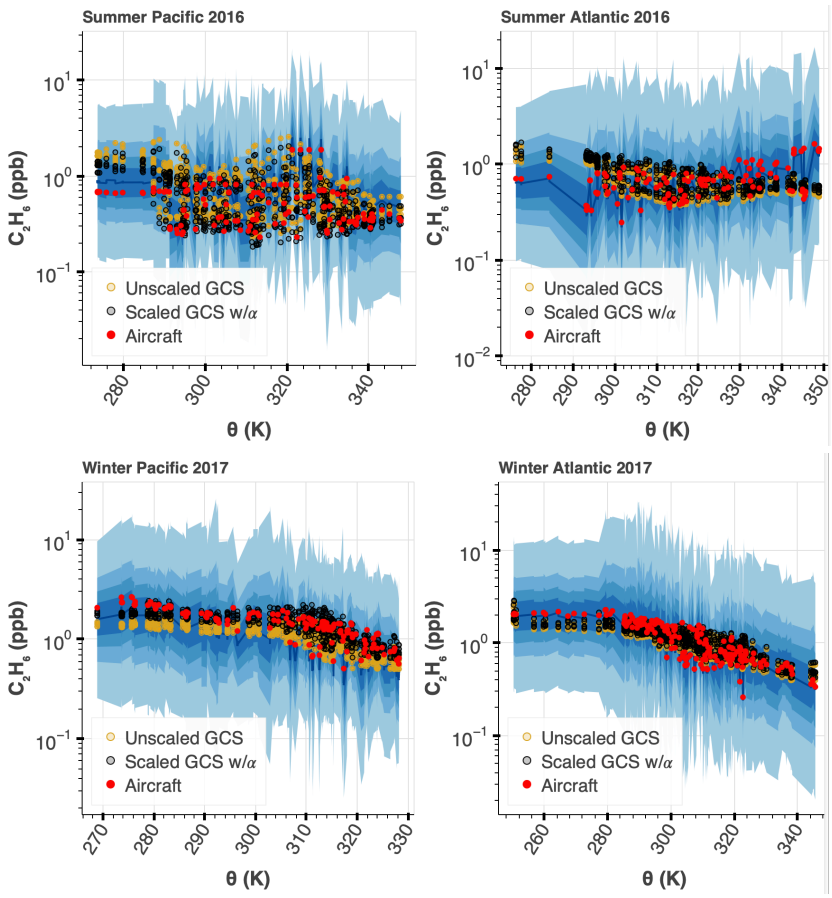


Figure S46. α_i hyperparameter estimate for each season during the ATom campaign. We do not include the summer values to calculate an overall α estimate as discussed in the methods in the main text. **Top**: C_2H_6, **Bottom**: C_3H_8.

5.4. Posterior predictive check – ATom observations

Posterior predictive checks involve drawing parameter values out of the posterior, using those parameters in the likelihood to generate a pseudo dataset, and repeat. We can see whether our Bayesian model can produce the observed data. Below, we show all posterior predictive checks for all ATom aircraft campaigns (Figures S47, S48). The majority of the measured data fell into the 30th and 50th percentile of the simulated Bayesian model data. The exception to this was the summer season, where the Bayesian model does not capture the measured aircraft data. This is expected, since during the summer we do not observe a robust relationship between potential temperature and C_3H_8 or C_2H_6.



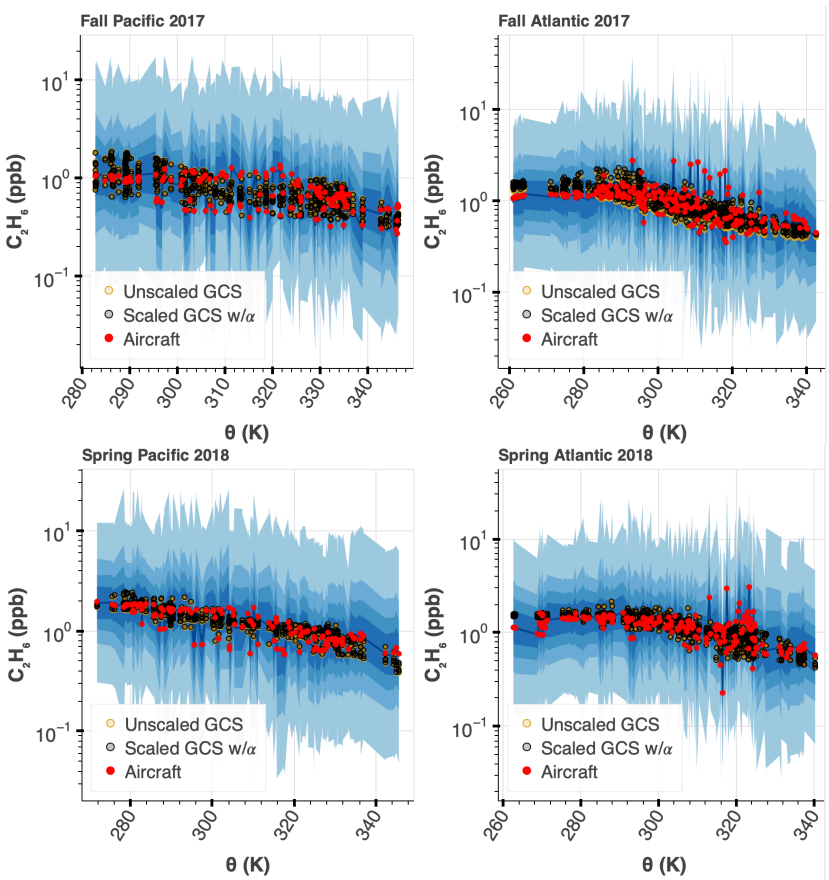
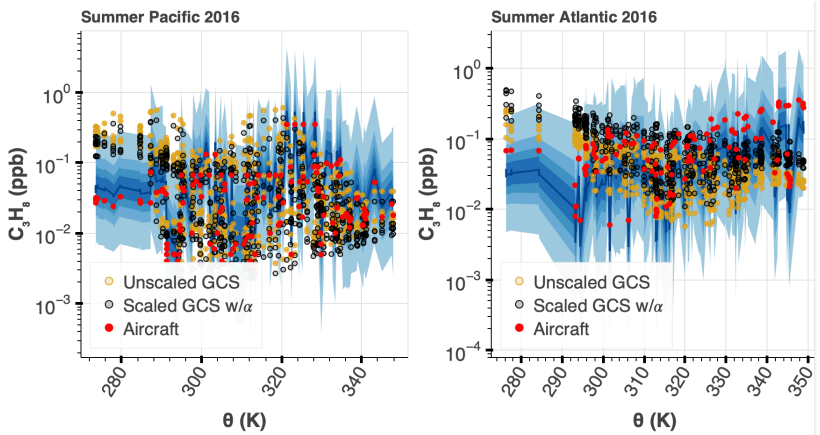


Figure S47. Posterior predictive check of C_2H_6 using ATom data. Posterior predictive checks are explained in the text above. The pseudo data are shown in blue with 30, 50, 70, 99th percentiles. Please see Figure S46 for the estimated values of α_i that were used to scale the GCS data in each season/transect.



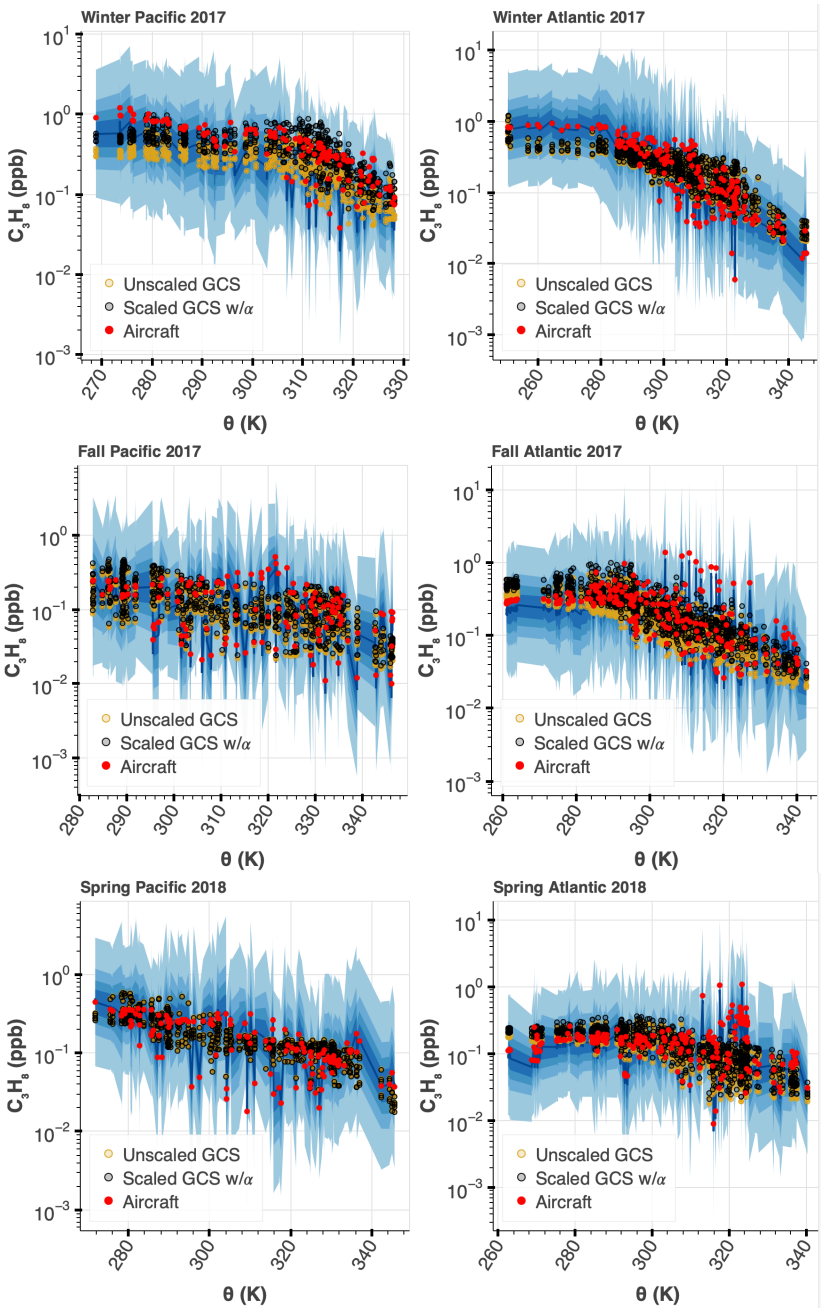


Figure S48. Posterior predictive check of C_3H_8 using ATom data. (Posterior predictive checks are explained in the text above.) The pseudo data are shown in blue with 30, 50, 70, 99th percentiles. Please see Figure S46 for the estimated values of α_i that were used to scale the GCS data in each season/transect.

Both C_3H_8 and C_2H_6 aircraft observations feature high mole fractions during ATom 2 winter measurements at low potential temperature. The largest differences between the aircraft and GEOS-Chem simulations occur at high latitude and low altitude (Figure S49), subject to low altitude and cold environments. GEOS-Chem does not able to capture this variability, as discussed in Section 3.3 in the main text.

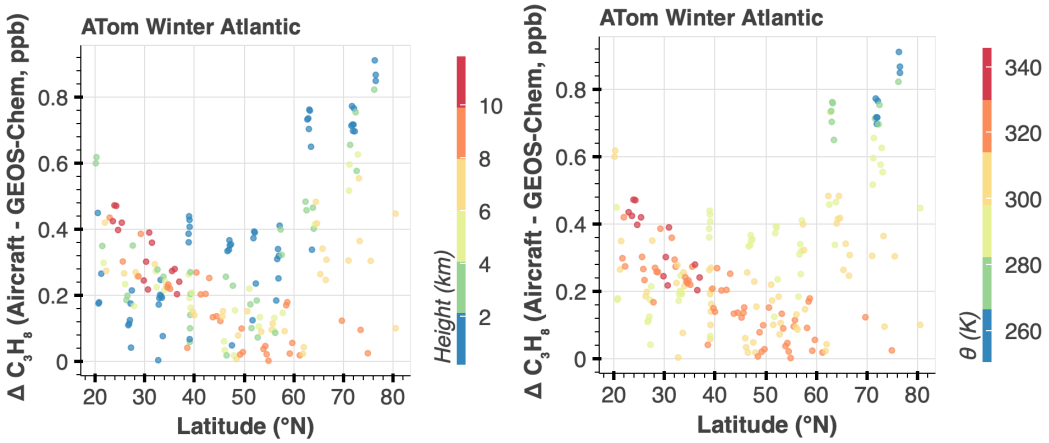


Figure S49. Difference between aircraft and GEOS-Chem C_3H_8 simulations. Simulations and aircraft observations during ATom 2, Atlantic transect, are shown. Points are colored by altitude (left) and potential temperature (right).

5.5. Posterior samples – HIPPO observations

We show an example of our posterior sampling for our Bayesian model using HIPPO C_2H_6 and C_3H_8 observations in Figures S50 and S51 below.

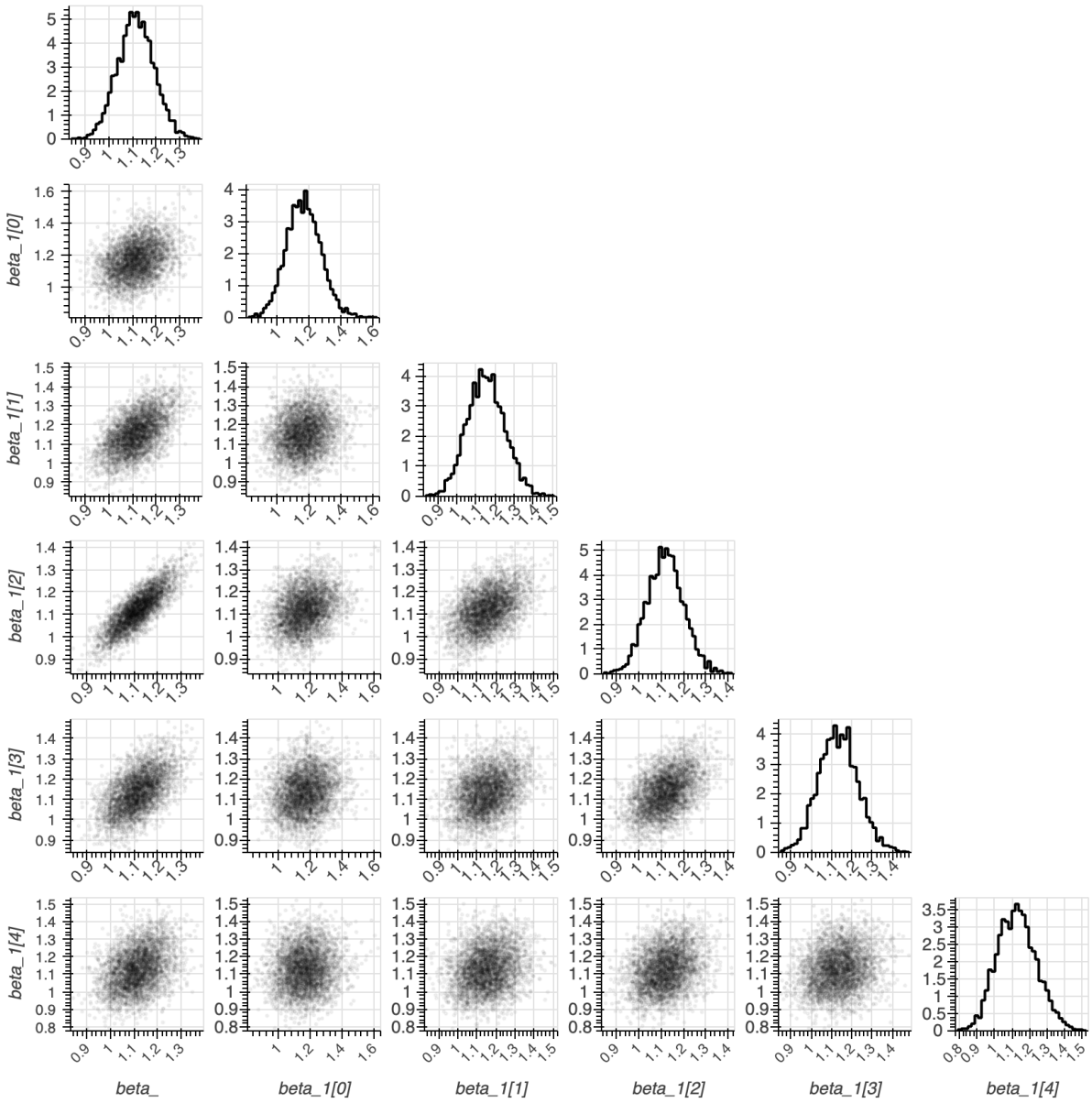


Figure S50. HMC Posterior samples for $\boldsymbol{\beta_{1,ij}}$ and $\boldsymbol{\beta_i}$ parameters using C_2H_6 HIPPO 5 aircraft and GEOS-Chem simulations. Beta_1 parameter is a vector of length 5, corresponding to the synoptic replicates of GEOS-Chem. Beta_1[0] and beta_1[1] correspond to 2 days before the aircraft, beta_1[2] is the plane path, and beta_1[3] and beta_1[4] correspond to 2 days after the aircraft. The hyperparameter, β_i, is represented by beta_.

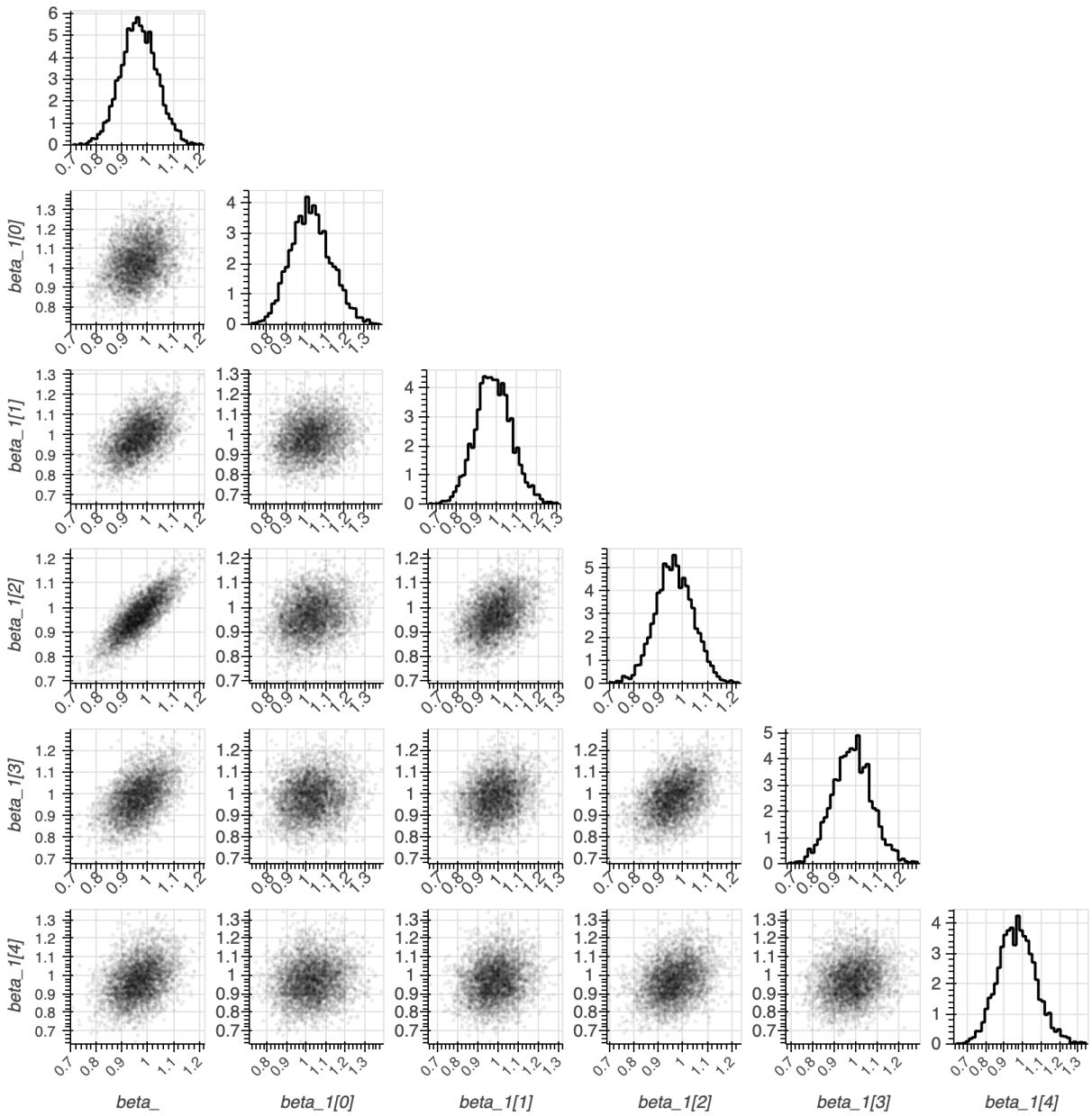


Figure S51. HMC Posterior samples for $\beta_{1,ij}$ and β_i parameters using C_3H_8 HIPPO 5 aircraft and GEOS-Chem simulations. Beta_1 parameter is a vector of length 5, corresponding to the synoptic replicates of GEOS-Chem. Beta_1[0] and beta_1[1] correspond to 2 days before the aircraft, beta_1[2] is the plane path, and beta_1[3] and beta_1[4] correspond to 2 days after the aircraft. The hyperparameter, β_i, is represented by beta_.

In Figure S52, we show a cross plot of our hyperparameter, $\beta_{1,ij}$ and β_i, in inverse form, for a single HIPPO campaign, which directly corresponds to the scaling of the GEOS-Chem v13.0.0 emissions. This shows an example of the variability due to GEOS-Chem meteorology compared to the hyperparameter that we use to scale the emissions. In Figure S53, we show the results of the Bayesian model for the α_i emissions scalar estimate for all 5 HIPPO campaigns.

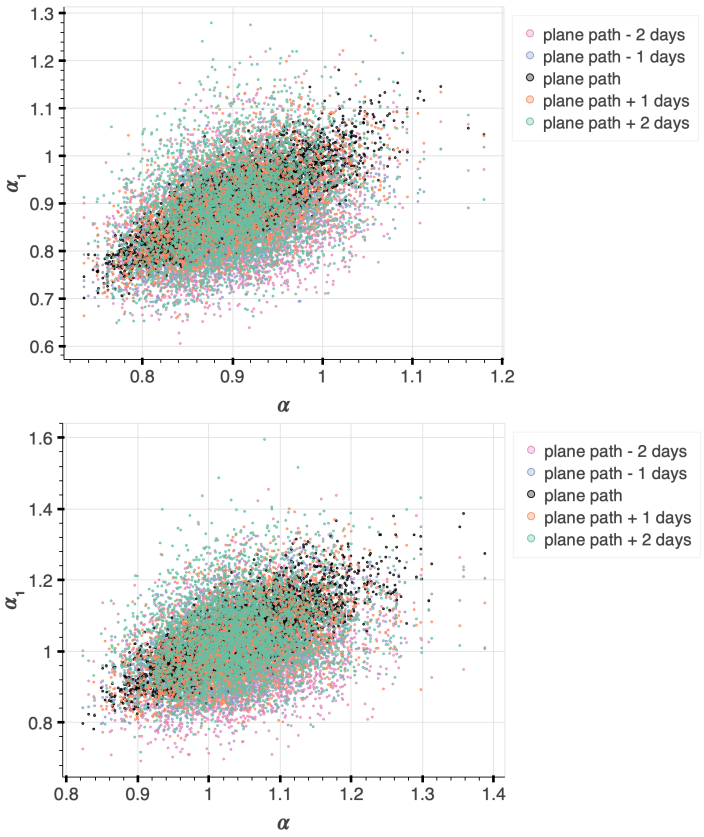


Figure S52. Posterior samples of $\alpha_{1,ij}$ vs α_i during HIPPO 5. Top: C_2H_6 observations. Bottom: C_3H_8 observations.

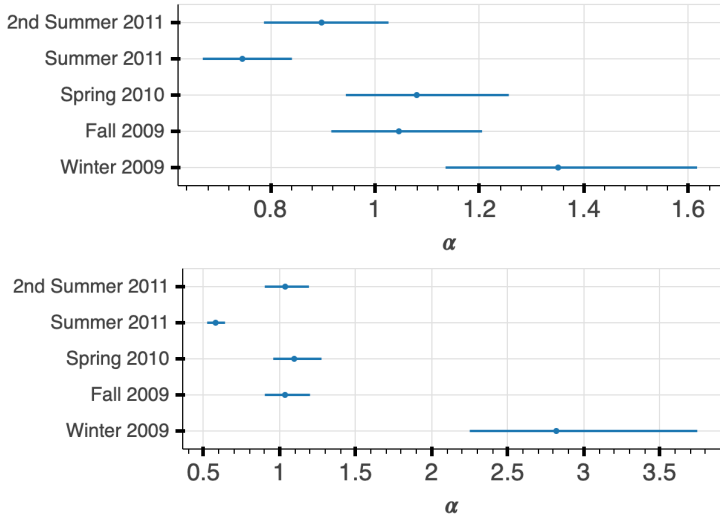


Figure S53. α_i hyperparameter estimate for each season during the HIPPO campaign. We do not include the summer values to calculate an overall α estimate as discussed in the methods in the main text. Top: C_2H_6, Bottom: C_3H_8. There are many fewer observations during HIPPO than ATom resulting in a much larger spread and bigger uncertainty in defining α_i.

5.6. Posterior predictive check – HIPPO observations

Posterior predictive checks involve drawing parameter values out of the posterior, using those parameters in the likelihood to generate a pseudo dataset, and repeat. We can see whether

our Bayesian model can produce the observed data. Below, we show all posterior predictive checks for all HIPPO aircraft campaigns (Figures S54, S55).

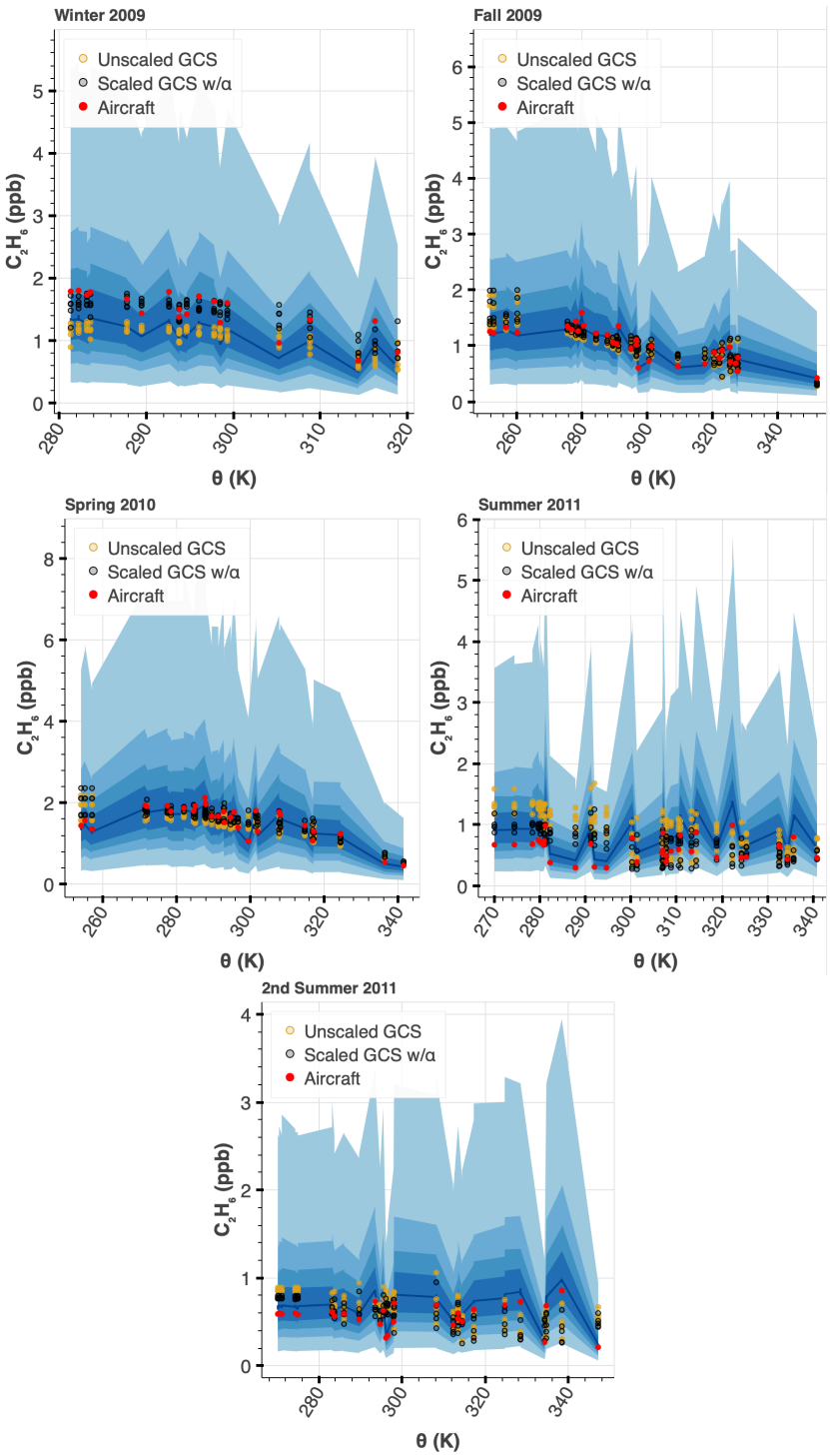


Figure S54. Posterior predictive check of C_2H_6 using HIPPO data. (Posterior predictive check method is described in the text above.) The pseudo data are shown in blue with 30, 50, 70, 99th percentiles. Please see Figure S53 for the estimated values of α_i that were used to scale the GCS data in each season/transect.

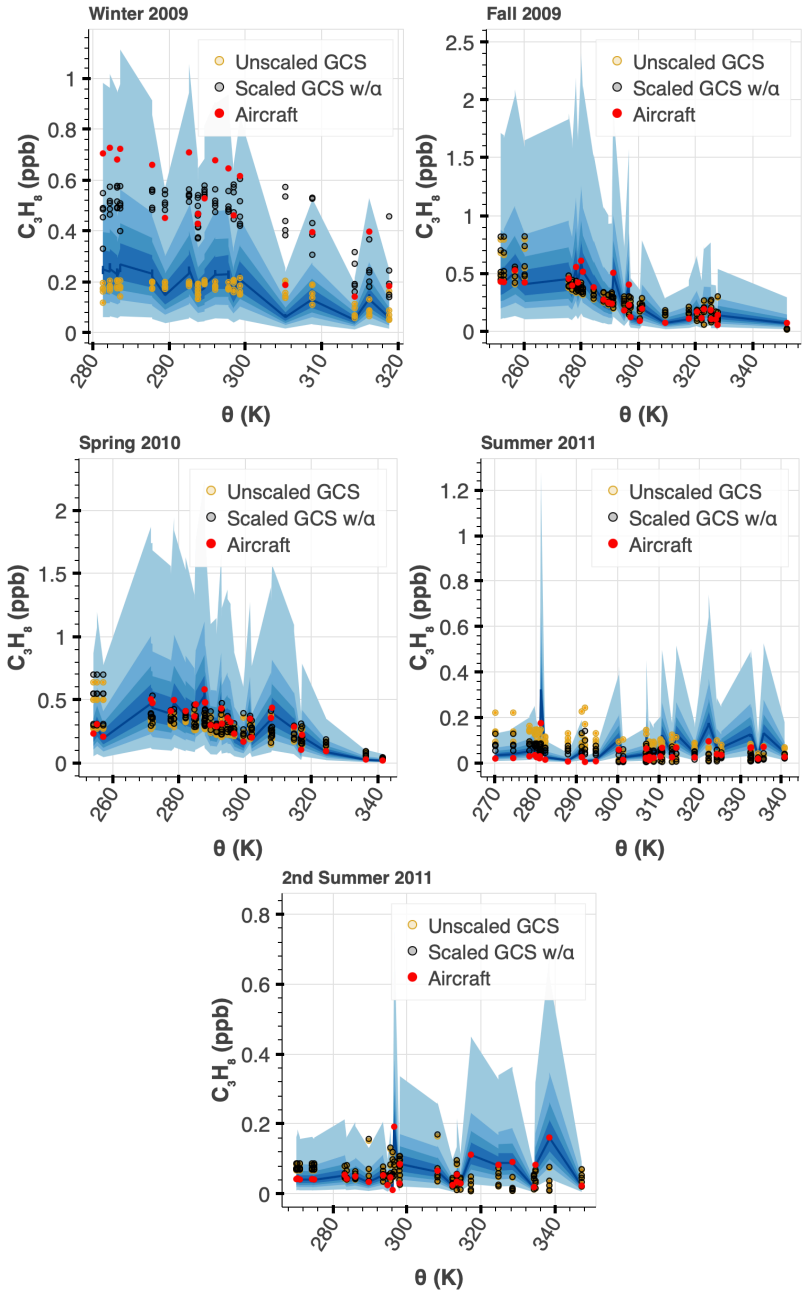


Figure S55. Posterior predictive check of C_3H_8 using HIPPO data. (Posterior predictive check method is described in the text above.) The pseudo data are shown in blue with 30, 50, 70, 99th percentiles. Please see Figure S53 for the estimated values of α_i that were used to scale the GCS data in each season/transect.

During HIPPO 1 (winter 2009), observations are biased towards high latitudes and are subject to arctic conditions. Furthermore, the revised emissions may be missing a high latitude source. This, combined with a relatively lower number of HIPPO aircraft observations at lower latitudes, results in a substantial bias on the overall Bayesian emissions scalar estimate for C_3H_8 during winter 2009 (Figure S53). To illustrate sampling biases at high latitudes during the winter

2009 campaign, we obtain two Bayesian estimates of α_i during each season: one estimate using aircraft observations restricted above 300 K (potential temperature) and those below 300 K. We see that observations restricted to values less than 300 K result in very high α_i estimates that bias the overall α scalar estimate (Figure S56-S57).

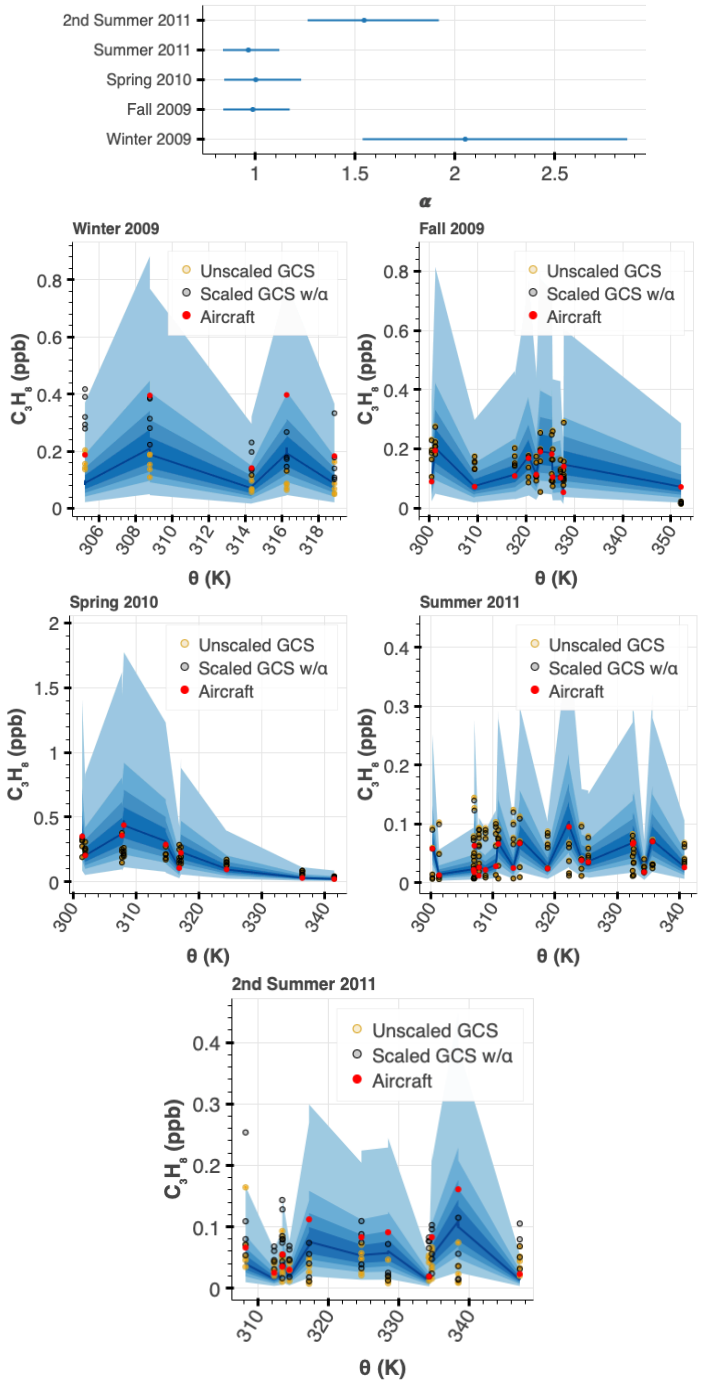


Figure S56. Bayesian α_i hyperparameter estimate and posterior predictive checks using HIPPO aircraft observations > 300 K.

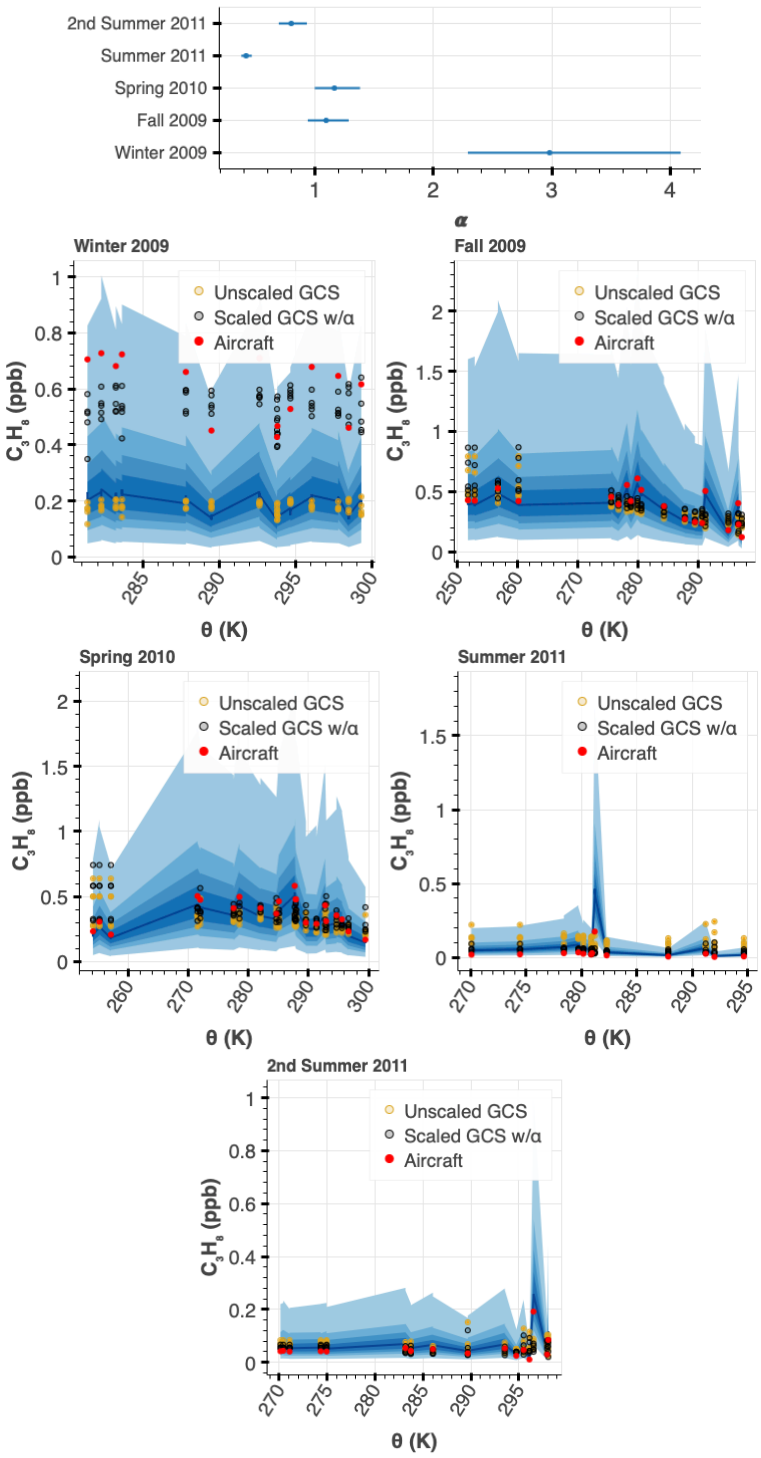


Figure S57. Bayesian α_i hyperparameter estimate and posterior predictive checks using HIPPO aircraft observations < 300 K.

5.7. Sigma parameter sensitivity analysis

We use ATom aircraft/GEOS-Chem simulations of C_3H_8 to observe the effect of implementing an unchanging, large σ_{ij} parameter. (As a reminder, the σ_{ij} parameter has the

effect of de-weighting GEOS-Chem simulations with higher tropopause height, since samples with high tropopause height tend to originate from the tropics, which is not useful for the purposes of our study. Please see Section 5.1 for more background on the parameters and the selected prior.) Here, we show that in practice, our model is quite robust against changes in σ_{ij}.

Using an unchanging, relatively large σ_{ij} parameter equal to 3.5 in the lognormal likelihood yields α_i hyperparameter estimates (Figure S58, S59) that are nearly identical to our results shown previously in which σ_{ij} varies according to tropopause height (Figure S45, S46). Furthermore, we obtain similar posterior predictive checks (Figure S60) as our previous results (Figure S48).

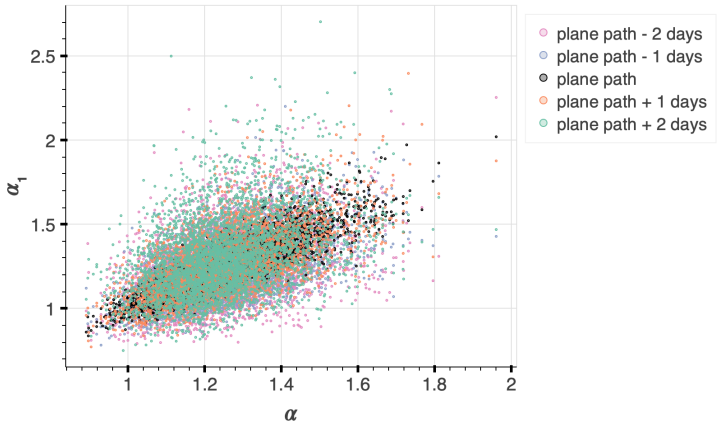


Figure S58. Posterior samples of $\alpha_{1,ij}$ vs α_i using a scalar σ_{ij} parameter. We use σ_{ij} = 3.5 instead of the usual distribution in the lognormal likelihood during ATom 4 time period.

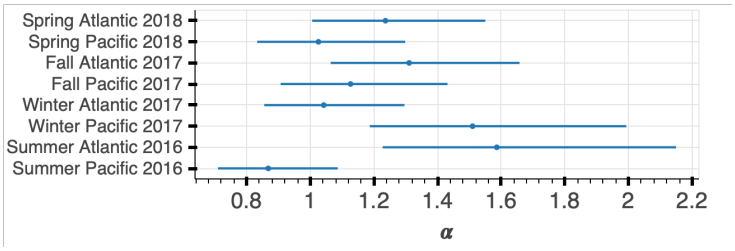
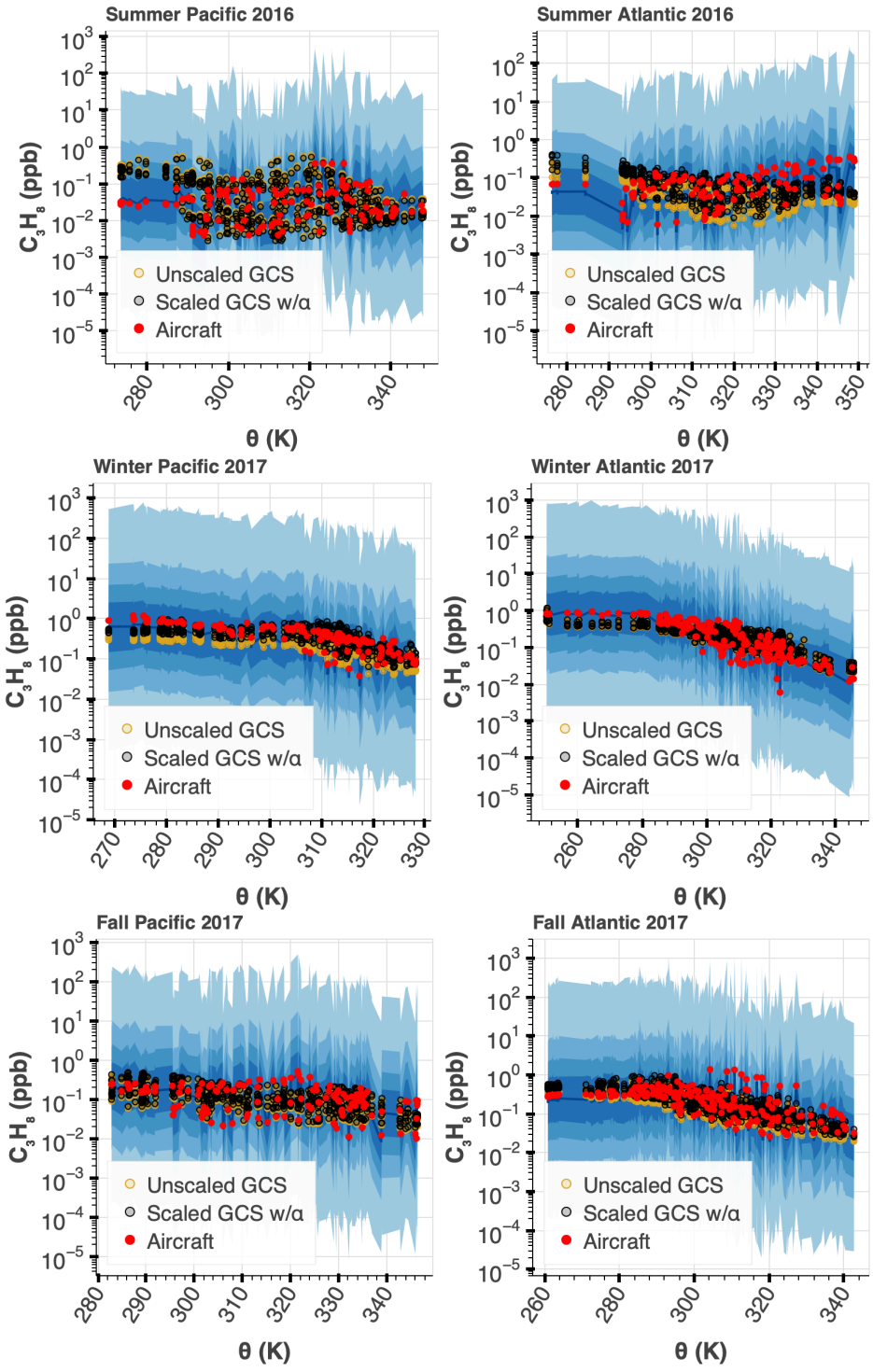


Figure S59. Estimate of α hyperparameter after using a scalar σ_{ij} parameter. We use σ_{ij} = 3.5 in the lognormal likelihood during ATom 4 time period.


Summer Pacific 2016
Summer Atlantic 2016
Winter Pacific 2017
Winter Atlantic 2017
Fall Pacific 2017
Fall Atlantic 2017
C_3H_8 (ppb)
θ (K)
Unscaled GCS
Scaled GCS w/α
Aircraft

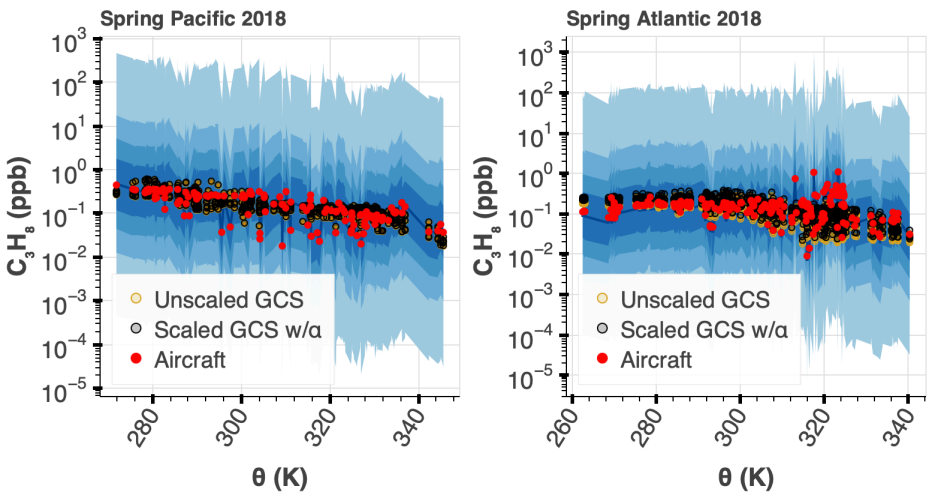


Figure S60. Posterior predictive check for C_3H_8 using a scalar σ_{ij} parameter. We use σ_{ij} = 3.5 in the lognormal likelihood.

5.8. Estimating an overall emissions scalar

To estimate a credible interval for an overall α, we draw a random sample of the posterior of hyperparameter α_i for each campaign season and ocean transect. We take the mean of these samples and repeat this 10,000 times. (Note that we do not use the summer estimates for this calculation, for reasons described in the main text.) For ATom, the 95% confidence interval for α of C_2H_6 is [1.02, 1.13] and for α of C_3H_8, [1.15, 1.27]. For HIPPO, the confidence interval for α of C_2H_6 is [1.06, 1.28] and for α of C_3H_8, [1.45, 1.98].

5.9 Estimating C_2H_6 and C_3H_8 Emissions

GEOS-Chem v13.0.0 emissions were calculated using the Harmonized Emissions Component (HEMCO) v3.0.0, as described in section 2.3 in the main text. To estimate emissions over the Northern Hemisphere and the U.S., we calculate a simple integration by defining a rectangle that describes the latitude and longitude boundaries that approximately encloses the geographical region of interest. (Table S3 shows those boundary estimates.) We approximate latitude and longitude to meters using the Haversine formula. We then integrate the region of the emission grid computed by HEMCO of the anthropogenic variable using trapezoidal integration along latitude and longitude. When estimating global emissions, we do not impose any boundaries on latitude or longitude. After integration, we simply convert the resulting units of kg/second to Tg/year.

Table S3. Boundary estimates for emissions grid.

Region	Latitude Min	Latitude Max	Longitude Min	Longitude Max
US	20	50	-130	-60
Northern Hemisphere	0	80	-165	180

The anthropogenic variable does not include biomass burning or biofuel emissions, according to GEOS-Chem documentation. Finally, we scale the emissions estimate for each boundary region with an overall hyperparameter, α, estimated during ATom from section 5.8. We report fossil emissions using 95% CI of α to define the variability. Our estimates for global fossil fuel emissions of C_2H_6 and C_3H_8 are [12.67, 13.98] (13.3 $\pm$ 0.7, 95% CI) and [13.89, 15.44] (14.7 $\pm$ 0.8, 95% CI) Tg/year, respectively (during the median year of 2017). Northern hemisphere emissions of C_2H_6 and C_3H_8 from fossil fuel production to be [11.18, 12.30] and [12.23, 13.60] Tg/year, respectively. In the U.S., we estimate C_2H_6 and C_3H_8 fossil fuel emissions at [1.29, 1.42] and [1.41, 1.56] Tg/year. Note that the global C_2H_6 emissions estimated in 2016-2018 are about 15% larger than in 2009-2011 ([10.55, 12.57] Tg/yr, Figure S61). Our C_3H_8 emissions are about 65% larger than in 2009-2011 ([7.31, 12.2] Tg/yr when not including the biased winter 2009 estimate that is impacted by arctic conditions and few observations. Our results are comparable to other studies (Figure S62).

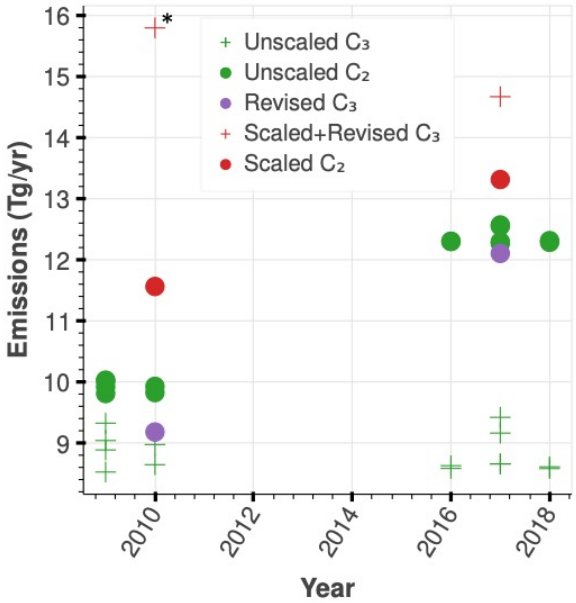


Figure S61. Global ethane and propane emissions during 2009-2018. "Unscaled" represent integrated default emissions from GEOS-Chem v13.0.0. "Revised C_3" represent the revised C_3H_8 emissions after implementing the default v13.0.0 C_2H_6 proxy. "Scaled+Revised C_3" represents the revised C_3H_8 emissions after scaling with our mean Bayesian estimate (Section 5.8). "Scaled C_2" represent the revised emissions after scaling with our mean Bayesian estimate (Section 5.8). $*$: Note that our mean scaled C_3 estimate shown here are skewed, as the 2009 winter HIPPO observations are latitudinally biased. We show that C_3 emissions increase by 65% from 2010 to 2017 when excluding the bias below in Figure S62.

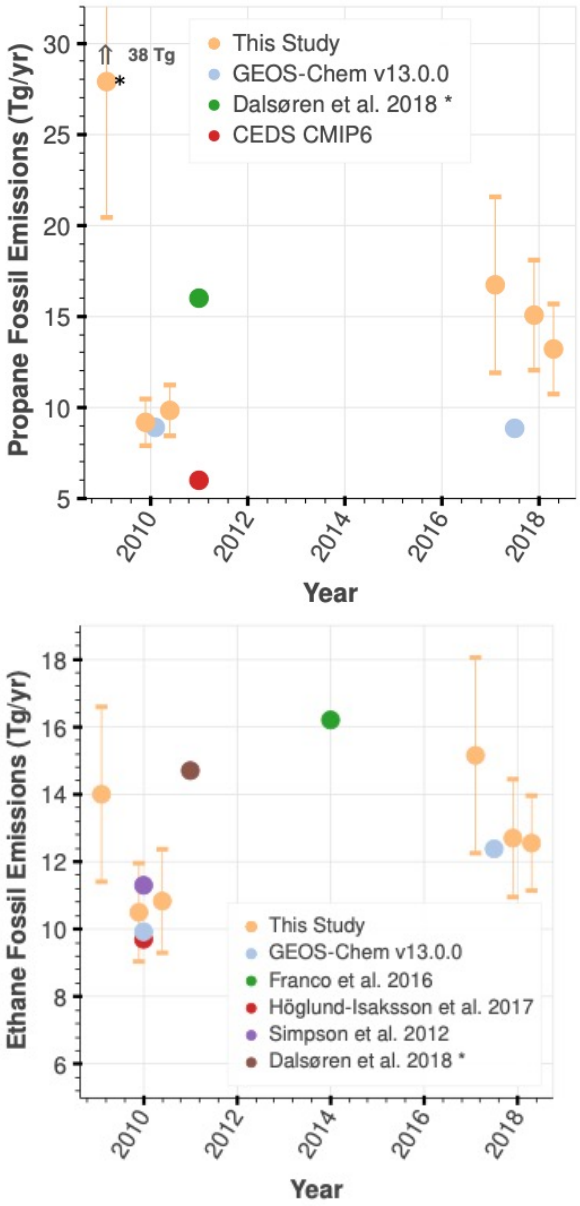


Figure S62. Global revised ethane and propane anthropogenic fossil emissions compared to other studies. Our emissions estimate in 2016-2018 (during ATom) and 2009-2011 (during HIPPO) includes GEOS-Chem v13.0.0 emissions for winter, fall and spring seasons scaled by α, that we determined with our Bayesian model during each season. As discussed in the text, fewer samples were obtained during HIPPO, resulting in a sampling bias that we test by restricting observations and simulations to $\pm$ 300K potential temperature (Figure S51-S52). This test affects the estimate about $\pm$ 1 Tg during 2010-2011 but affects our estimate by up to 12 Tg in 2009. $*$: This 2009 estimate is highly biased, as the latitudinal coverage of aircraft observations is not representative of the global spatial distribution of methane emissions from oil and gas processes and the confidence interval stretches to nearly 40 Tg (please see Section 3.3 text and Figure S51-52). We compare our revised ethane and propane emissions to the default emissions from GEOS-Chem v13.0.0 (relevant anthropogenic inventories include Tzompa-Sosa et al. 2017[25] for C_2H_6, and Xiao et al. 2008[4] for C_3H_8). ‡: The studies included here[38–41] represent anthropogenic fossil emissions, except Dalsøren et al. 2018 which also includes biofuel, agriculture, and waste. We obtained the CEDS CMIP6 estimate from Dalsøren et al. 2018. Our emissions estimates do not include biomass burning or biofuels. Please see Section 5 in the SI for more information on estimating these emissions.

6. Oil & Gas Emission Ratios

6.1 Hydrocarbon wellhead composition

We gathered hydrocarbon wellhead compositions reported in the literature for the top five natural gas-producing countries in the world (Table S4 and S5).

Table S4. Statistical summaries of hydrocarbon wellhead composition for the globe. Units are in mole % for the US,[42–49] Russia,[50–56] Qatar,[57] Iran,[58] and Canada [59–62]. **Top**: C_1; **Middle**: C_2; **Bottom**: C_3.

region	count	mean	std	min	25%	50%	75%	max
CANADA	22.0	74.090909	10.962747	54.00	65.2500	77.000	83.000	89.00
IRAN	45.0	89.531333	4.071936	74.16	89.2700	90.870	91.860	93.53
QATAR	1.0	89.900000	-	89.90	89.9000	89.900	89.900	89.90
RUSSIA	18.0	88.278778	12.873191	65.00	75.6875	95.685	98.500	100.00
US	63.0	85.718413	9.687639	52.67	78.3550	88.170	93.835	98.44

region	count	mean	std	min	25%	50%	75%	max
CANADA	22.0	14.909091	4.849242	7.0000	10.250	15.00	18.00	24.00
IRAN	45.0	5.624667	1.506998	4.0300	5.110	5.32	5.60	11.71
QATAR	1.0	6.000000	-	6.0000	6.000	6.00	6.00	6.00
RUSSIA	18.0	7.075722	7.677314	0.0000	0.750	4.50	11.95	26.00
US	63.0	7.029257	5.513786	0.0912	2.525	5.30	11.84	24.60

region	count	mean	std	min	25%	50%	75%	max
CANADA	22.0	11.000000	7.244045	4.00	5.0000	6.500	15.75	24.00
IRAN	45.0	2.250222	1.081225	1.37	1.7500	1.970	2.12	6.78
QATAR	1.0	2.200000	-	2.20	2.2000	2.200	2.20	2.20
RUSSIA	18.0	6.242111	7.98688	0.00	0.5600	1.945	11.95	26.00
US	63.0	2.792530	3.044868	0.00	0.1535	1.930	4.88	12.86

Table S5. Statistical summaries of hydrocarbon wellhead composition in the U.S. Units are in mole %.[42–49] **Top**: C_1; **Middle**: C_2; **Bottom**: C_3.

region	count	mean	std	min	25%	50%	75%	max
Appalachian	22.0	88.089091	7.121542	76.38	82.5250	88.180	94.1925	98.440
Bakken	1.0	52.670000	-	52.67	52.6700	52.670	52.6700	52.670
Barnett	5.0	84.962000	7.248256	77.81	80.3000	81.200	91.8000	93.700
Eagleford	13.0	80.726923	7.325718	66.20	77.8000	80.270	82.9900	91.790
Haynesville	15.0	94.344000	2.323018	89.40	92.5950	95.000	96.2100	97.100
Niobrara	4.0	77.447500	6.680047	68.49	74.2725	79.350	82.5250	82.600
Permian	3.0	70.140000	3.86114	66.26	68.2190	70.178	72.0800	73.982

region	count	mean	std	min	25%	50%	75%	max
Appalachian	22.0	6.558182	4.71379	1.1000	2.6325	5.235	10.6600	16.10
Bakken	1.0	24.600000	-	24.6000	24.6000	24.600	24.6000	24.60
Barnett	5.0	7.544000	3.983802	2.6000	4.4000	8.100	10.8200	11.80
Eagleford	13.0	10.045385	4.711809	0.2500	9.1900	11.880	12.6200	16.30
Haynesville	15.0	1.585613	1.58007	0.0912	0.2160	0.419	2.8800	4.10
Niobrara	4.0	10.877500	2.185869	8.5000	9.7000	10.655	11.8325	13.70
Permian	3.0	12.786333	1.172344	11.4540	12.3495	13.245	13.4525	13.66

region	count	mean	std	min	25%	50%	75%	max
Appalachian	22.0	2.069295	1.88617	0.020	0.15175	1.9000	3.3750	6.070
Bakken	1.0	12.860000	-	12.860	12.86000	12.8600	12.8600	12.860
Barnett	5.0	2.584000	2.465011	0.000	0.40000	2.3000	5.0200	5.200
Eagleford	13.0	3.813077	2.550161	0.030	2.06000	4.0500	5.1000	8.520
Haynesville	15.0	0.370527	0.417251	0.000	0.00850	0.0325	0.8355	0.927
Niobrara	4.0	5.777500	3.298488	2.700	4.05000	5.0000	6.7275	10.410
Permian	3.0	8.795667	1.469655	7.352	8.04850	8.7450	9.5175	10.290

Using Equation 2 (main text), we combine literature estimates of dry natural gas production and hydrocarbon composition measurements from a variety of basins (Table S4 and Table S5) to arrive at a global and U.S. C_3/C_1 emission ratio. We refer to this value as a "literature" emission ratio. For our global literature emission ratio, we use hydrocarbon and dry natural gas production data from the top 5 producing natural gas basins around the world that made up 50% of the total natural gas production in 2019. For our U.S. emission ratio, we include the top 7 natural gas producing basins that account for 86% of total U.S. natural gas production. The literature emission ratio of C_2/C_1 is calculated similarly. We arrive at a single literature emission estimate for both the US and globe (Figure 13, main text). Due to the limited published data on hydrocarbon composition, we compile data we found for a range of years: U.S. (2003-2020), Russia (1995-2018), Qatar (2005), Iran (2006), Canada (2004-2020). We calculate

confidence intervals for each basin by performing a pairs bootstrap with co-measurements of hydrocarbon composition using the same bootstrap methods used with the NOAA and FRAPPE data as described in the methods section of the main text.

Separately, we calculate an "implied" emission ratio by taking the ratio between our revised C_3H_8 emissions with several literature estimates of CH_4 emissions from oil and natural gas processes. The results are shown in Figure S63. The implied emission ratio for our revised C_2H_6 is calculated similarly. For the EPA U.S. CH_4 emissions estimate,[63] we sum the categories "Natural Gas Systems", "Petroleum Systems", and "Abandoned Oil and Gas Wells" from Table 3-1 using the 2017 estimates (to match NOAA/ATom time period during 2016-2018).

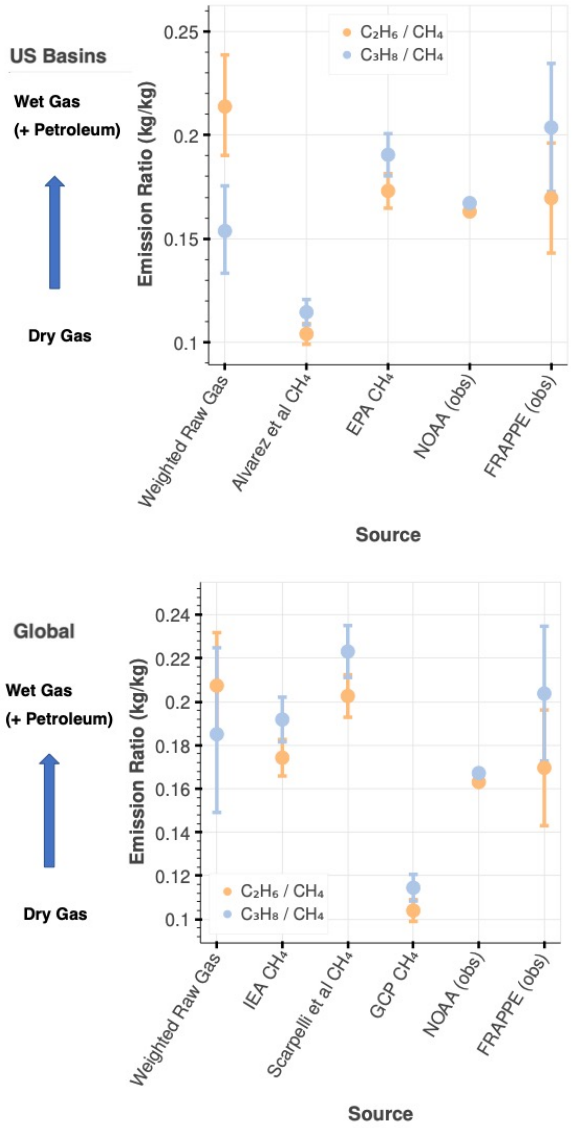


Figure S63. Literature and Observationally-Informed Emission ratios. Top: U.S. basins; Bottom: Global basins. The weighted raw gas ratio represents the "literature ratio" described in the main text. OIER, ratios between our revised C_2H_6 and C_3H_8 emissions and literature CH_4 emission estimates, are shown for several literature CH_4 estimates, including Alvarez et al. 2018 (13 Tg/yr)[64] and EPA 2017 estimate (7.8 Tg/yr, 2021 report)[63] for U.S. basins, and IEA 2021 (76.4 Tg/yr),[65] Scarpelli et al. 2020 (65.7 Tg/yr),[66] and Global Carbon Project 2020 bottom-up estimate[67] (128 Tg/yr, 2008-2017 average) for global basins. The variability in the literature ratio is attributed to the 95% CI of pairs bootstrap samples of hydrocarbon composition measurements

(see main text for more detail). The variability in the OIER is attributed to the 95% CI of our revised C_3H_8 and C_2H_6 emission estimates. We also compare C_3H_8/CH_4 and C_2H_6/CH_4 correlations from in-situ observations, including NOAA observations from Northern Oklahoma (2017 average from Figure S21, units of kg/kg) and FRAPPE observations from Northern Colorado (2014 from Figure S9, units of kg/kg). The variability in the NOAA ratio is relatively low because it is calculated from a multi-year average slope, and the error in the slope is low (see Figure S21, left). The variability in the FRAPPE ratio is relatively high because we use the 95% CI derived directly from our bootstrap samples described in the main text.

6.2. Impact of reallocation of CH_4 emissions on the transportation sector footprint

There has been much debate about the greenhouse gas mitigation impact of switching from coal to natural gas energy in electricity production. The impact depends on how much CH_4 is lost during natural gas production, processing, and transport as these losses will offset some of the benefits of the lower CO_2 emissions. A study by the Environmental Defense Fund suggested that a CH_4 leak rate greater than 3% would negate the climate benefits of switching from coal to gas in the near term (the current leak rate is estimated at 2.3%).[68] As described here and in other studies, although CH_4 emissions associated with dry natural gas production likely remain underestimated, flared and vented associated gas from petroleum exploration contributes significantly to the total emissions. Since global oil production has continually increased over the past 3 decades (Figure S2), and our findings suggest that CH_4 losses are at least proportional to production (but likely greater and biased towards oil-producing sites), a significant fraction of the estimated CH_4 emissions may be misallocated to dry CH_4 production and should instead be included with the oil production sector.

Correctly attributing CH_4 emissions from oil production to the transportation sector, rather than the power sector, increases the greenhouse gas footprint of petroleum-based transportation, while decreasing the greenhouse gas emissions ascribed to natural gas-powered power plants. As an example, if we assume that 20% of natural gas losses are associated with petroleum exploration (associated natural gas makes up 20% of total natural gas marketed production in the U.S.[69]), the CO_2 equivalent footprint of the global transportation sector would increase by roughly 5%, using IEA's estimate of 76 Tg/year CH_4 emissions from oil and natural gas and recent transportation CO_2 emission estimates.[70,71] Equivalently, the U.S. transportation sector CO_2 footprint would increase by 2%, using the EPA's U.S. 2017 estimate (2021 report)[63] of 7.8 Tg/year CH_4 emissions and the U.S. petroleum/transportation sector CO_2 (equivalent) footprint as reported by the EPA (2017 estimate, 2021 report)[63]. Both of our estimates are lower bounds that will only increase when accounting for vented and flared losses of associated natural gas that is not accounted for in marketed associated gas.

Methods:

For a global estimate, we estimate the transportation footprint from Figure 4 from Liu et al. 2020[71] and multiply it by 365 to get 7,300 Mmt (million metric tones) CO_2 equivalents/year. We use 76 Tg/year of CH_4 from (IEA)[65] as a global estimate from oil and natural gas and find that 20% of that number results in 15.2 Tg, multiplied by 25 GWP, yields 6% of the transportation sector. Using the same process for the U.S., the EPA estimates the transportation sector at 1,740.2 Mmt CO_2/year (Table 3-5, Petroleum fuel/Transportation sector for 2017).[63] We use the EPA's 7.8 Tg/year[64] of CH_4 for the oil and gas estimate and find that 20% of 7.8 Tg/year multiplied by 25 GWP results in 2.2% of the U.S. transportation sector.

References

(1) U.S. Energy Information Administration. Where our natural gas comes from https://www.eia.gov/energyexplained/natural-gas/where-our-natural-gas-comes-from.php (accessed 2022 -03 -18).
(2) U.S. Energy Information Administration. Natural gas explained https://www.eia.gov/energyexplained/natural-gas/.
(3) Keller, A. NGL 101- The Basics, 2012.
(4) Xiao, Y.; Logan, J. A.; Jacob, D. J.; Hudman, R. C.; Yantosca, R.; Blake, D. R. Global Budget of Ethane and Regional Constraints on U.S. Sources. *J. Geophys. Res.* **2008**, *113* (D21). https://doi.org/10.1029/2007JD009415.
(5) MacIntyre, S., Wilczewski, W. U.S. Ethane Consumption, Exports to Increase as New Petrochemical Plants Come Online. U.S. Energy Information Administration April 2018.
(6) U.S. Energy Information Administration. Natural Gas Weekly Update. July 7, 2016.
(7) U.S. Energy Information Administration. Natural Gas Weekly Update. March 1, 2018.
(8) U.S. Energy Information Administration. Drilling Productivity Report for Key Tight Oil and Shale Gas Regions, 2022.
(9) Worldwide Power Products. 10 Biggest Shale Plays in the US - New & Used Generators, Ends and Engines | Houston, TX | Worldwide Power Products.
(10) International Energy Agency. IEA Atlas of Energy: Natural Gas Production http://energyatlas.iea.org/#!/tellmap/-1165808390.
(11) International Energy Agency. IEA Atlas of Energy: Crude Oil Production http://energyatlas.iea.org/#!/tellmap/-1920537974.
(12) SoCalGas. Playa Del Rey Withdrawn Natural Gas Composition, 2021.
(13) U.S. Energy Information Administration. Natural Gas Gross Withdrawals and Production (U.S.), 2020.
(14) U.S. Energy Information Administration. Natural Gas Plant Field Production, 2020.
(15) Parrish, D. D.; Stohl, A.; Forster, C.; Atlas, E. L.; Blake, D. R.; Goldan, P. D.; Kuster, W. C.; de Gouw, J. A. Effects of Mixing on Evolution of Hydrocarbon Ratios in the Troposphere: MIXING EFFECTS ON NMHC RATIO EVOLUTION. *J. Geophys. Res. Atmospheres* **2007**, *112* (D10). https://doi.org/10.1029/2006JD007583.
(16) Boyd, D. T.; Mankin, C. J. Map of Oklahoma Oil and Gas Fields - Oklahoma Geological Survey, 2002.
(17) Parker, J.; Wertz, J. Oklahoma Oil Production in the Last Five Years: Has It Changed?, 2017.
(18) Best, A. Water Education Colorado: Do Oil and Water Mix?, 2013.
(19) toddkarin. global-land-mask (1.0.0) https://pypi.org/project/global-land-mask/ (accessed 2022 -08 -22).
(20) NASA Global Modeling and Assimilation Office. GEOS5124 FP-IT 2d Assimilated State.
(21) Seinfeld, J. H.; Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Third edition.; John Wiley & Sons: Hoboken, New Jersey, 2016.

(22) Gelaro, R.; McCarty, W.; Suárez, M. J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C. A.; Darmenov, A.; Bosilovich, M. G.; Reichle, R.; Wargan, K.; Coy, L.; Cullather, R.; Draper, C.; Akella, S.; Buchard, V.; Conaty, A.; da Silva, A. M.; Gu, W.; Kim, G.-K.; Koster, R.; Lucchesi, R.; Merkova, D.; Nielsen, J. E.; Partyka, G.; Pawson, S.; Putman, W.; Rienecker, M.; Schubert, S. D.; Sienkiewicz, M.; Zhao, B. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30* (14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

(23) Lin, H.; Jacob, D. J.; Lundgren, E. W.; Sulprizio, M. P.; Keller, C. A.; Fritz, T. M.; Eastham, S. D.; Emmons, L. K.; Campbell, P. C.; Baker, B.; Saylor, R. D.; Montuoro, R. Harmonized Emissions Component (HEMCO) 3.0 as a Versatile Emissions Component for Atmospheric Models: Application in the GEOS-Chem, NASA GEOS, WRF-GC, CESM2, NOAA GEFS-Aerosol, and NOAA UFS Models. *Geosci. Model Dev.* **2021**, *14* (9), 5487–5506. https://doi.org/10.5194/gmd-14-5487-2021.

(24) Community, T. I. G.-C. U. *Geoschem/HEMCO: HEMCO 3.0.0-Rc (HEMCO 3.0.0 Release Candidate)*; Zenodo, 2021. https://doi.org/10.5281/ZENODO.4429214.

(25) Tzompa-Sosa, Z. A.; Mahieu, E.; Franco, B.; Keller, C. A.; Turner, A. J.; Helmig, D.; Fried, A.; Richter, D.; Weibring, P.; Walega, J.; Yacovitch, T. I.; Herndon, S. C.; Blake, D. R.; Hase, F.; Hannigan, J. W.; Conway, S.; Strong, K.; Schneider, M.; Fischer, E. V. Revisiting Global Fossil Fuel and Biofuel Emissions of Ethane. *J. Geophys. Res. Atmospheres* **2017**, *122* (4), 2493–2512. https://doi.org/10.1002/2016JD025767.

(26) Gelman, A. *Bayesian Data Analysis*, Third edition.; Chapman & Hall/CRC texts in statistical science; CRC Press: Boca Raton, 2014.

(27) Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2021.

(28) Stan Dev Team. CmdStanPy (0.9.76), 2021.

(29) Kumar, R.; Carroll, C.; Hartikainen, A.; Martin, O. ArviZ a Unified Library for Exploratory Analysis of Bayesian Models in Python. *J. Open Source Softw.* **2019**, *4* (33), 1143. https://doi.org/10.21105/joss.01143.

(30) Talts, S.; Betancourt, M.; Simpson, D.; Vehtari, A.; Gelman, A. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *ArXiv180406788 Stat* **2020**.

(31) Bois, J. *Justinbois/Bebi103: 0.1.0*; CaltechDATA, 2020. https://doi.org/10.22002/D1.1615.

(32) Bois, J. S. *Justinbois/Iqplot: 0.1.6*; CaltechDATA, 2020. https://doi.org/10.22002/D1.1614.

(33) Rudiger, P.; Stevens, J.-L.; Bednar, J. A.; Nijholt, B.; Mease, J.; Andrew; B, C.; Randelhoff, A.; Tenner, V.; Maxalbert; Kaiser, M.; Ea42gh; Stonebig; Hoxbro; Samuels, J.; Pevey, K.; LB, F.; Tolmie, A.; Stephan, D.; Bois, J.; Lowe, S.; Bampton, J.; Henriqueribeiro; Ruoyu0088; Lustig, I.; Klein, A.; Van De Ven, B.; Raillard, D.; Signell, J.; Talirz, L. *Holoviz/Holoviews: Version 1.14.5*; Zenodo, 2021. https://doi.org/10.5281/ZENODO.5114034.

(34) Brendan Collins, B. V. D. V. Bokeh: Essential Open Source Tools for Science. **2020**. https://doi.org/10.5281/ZENODO.4317717.

(35) Reback, J.; Jbrockmendel; McKinney, W.; Van Den Bossche, J.; Augspurger, T.; Cloud, P.; Hawkins, S.; Gfyoung; Sinhrks; Roeschke, M.; Klein, A.; Terji Petersen; Tratner, J.; She, C.; Ayd, W.; Hoefler, P.; Naveh, S.; Garcia, M.; Schendel, J.; Hayden, A.; Saxton, D.; Shadrach, R.; Gorelli, M. E.; Jancauskas, V.; Fangchen Li; Attack68; McMaster, A.;

Battiston, P.; Skipper Seabold; Kaiqi Dong. *Pandas-Dev/Pandas: Pandas 1.3.1*; Zenodo, 2021. https://doi.org/10.5281/ZENODO.5136416.

(36) Virtanen, P.; Gommers, R.; Burovski, E.; Oliphant, T. E.; Weckesser, W.; Cournapeau, D.; Alexbrc; Reddy, T.; Peterson, P.; Haberland, M.; Wilson, J.; Nelson, A.; Endolith; Mayorov, N.; Walt, S. V. D.; Ilhan Polat; Laxalde, D.; Brett, M.; Larson, E.; Millman, J.; Lars; Mulbregt, P. V.; Eric-Jones; CJ Carey; Moore, E.; Kern, R.; Peterbell10; Leslie, T.; Perktold, J.; Striega, K. *Scipy/Scipy: SciPy 1.6.2*; Zenodo, 2021. https://doi.org/10.5281/ZENODO.4635380.

(37) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(38) Höglund-Isaksson, L. Bottom-up Simulations of Methane and Ethane Emissions from Global Oil and Gas Systems 1980 to 2012. *Environ. Res. Lett.* **2017**, *12* (2), 024007. https://doi.org/10.1088/1748-9326/aa583e.

(39) Dalsøren, S. B.; Myhre, G.; Hodnebrog, Ø.; Myhre, C. L.; Stohl, A.; Pisso, I.; Schwietzke, S.; Höglund-Isaksson, L.; Helmig, D.; Reimann, S.; Sauvage, S.; Schmidbauer, N.; Read, K. A.; Carpenter, L. J.; Lewis, A. C.; Punjabi, S.; Wallasch, M. Discrepancy between Simulated and Observed Ethane and Propane Levels Explained by Underestimated Fossil Emissions. *Nat. Geosci.* **2018**, *11* (3), 178–184. https://doi.org/10.1038/s41561-018-0073-0.

(40) Franco, B.; Mahieu, E.; Emmons, L. K.; Tzompa-Sosa, Z. A.; Fischer, E. V.; Sudo, K.; Bovy, B.; Conway, S.; Griffin, D.; Hannigan, J. W.; Strong, K.; Walker, K. A. Evaluating Ethane and Methane Emissions Associated with the Development of Oil and Natural Gas Extraction in North America. *Environ. Res. Lett.* **2016**, *11* (4), 044010. https://doi.org/10.1088/1748-9326/11/4/044010.

(41) Simpson, I. J.; Sulbaek Andersen, M. P.; Meinardi, S.; Bruhwiler, L.; Blake, N. J.; Helmig, D.; Rowland, F. S.; Blake, D. R. Long-Term Decline of Global Atmospheric Ethane Concentrations and Implications for Methane. *Nature* **2012**, *488* (7412), 490–494. https://doi.org/10.1038/nature11342.

(42) Howard, T.; Ferrara, T. W.; Townsend-Small, A. Sensor Transition Failure in the High Flow Sampler: Implications for Methane Emission Inventories of Natural Gas Infrastructure. *J. Air Waste Manag. Assoc.* **2015**, *65* (7), 856–862. https://doi.org/10.1080/10962247.2015.1025925.

(43) Thiagarajan, N.; Xie, H.; Ponton, C.; Kitchen, N.; Peterson, B.; Lawson, M.; Formolo, M.; Xiao, Y.; Eiler, J. Isotopic Evidence for Quasi-Equilibrium Chemistry in Thermally Mature Natural Gases. *Proc. Natl. Acad. Sci.* **2020**, *117* (8), 3989–3995. https://doi.org/10.1073/pnas.1906507117.

(44) Burruss, R. C.; Ryder, R. T. *Open-File Report*; Composition of crude oil and natural gas produced from 14 wells in the Lower Silurian "Clinton" sandstone and Medina Group, northeastern Ohio and northwestern Pennsylvania; Open-File Report; U.S. Geological Survey: Reston, Virginia 20192, 2003.

(45) Fairhurstm, B.; Hanson, M. L. Evolution and Development of the WolfBone Play, Southern Delaware Basin, West Texas: An Emerging Frontier, An Oil-Rich Unconventional Resource; #10411 (2012). **2012**, 64.
(46) Bullin, K.; Krouskop, P. *Composition Variety Complicates Processing Plans from US Shale Gas*; Bryan Research and Engineering Inc., 2008.
(47) Higley, D. K.; Cox, D. O. *Oil and Gas Exploration and Development along the Front Range in the Denver Basin of Colorado, Nebraska, and Wyoming*; Petroleum Systems and Assessment of Undiscovered Oil and Gas in the Denver Basin Province, Colorado, Kansas, Nebraska, South Dakota, and Wyoming - USGS Province 39; Digital Data Series DDS-69-P; U.S. Geological Survey, 2007; p Ch.2.
(48) Laughrey, C. D. *Comparative Natural Gas Geochemistry of the Hanson 31-5054z Water Well and Select Upper Jurassic - Lower Cretaceous Oil and Gas Wells, Northwest Louisiana*; OilTracers Updated Report No. 18-2422; Louisiana Department of Natural Resources, 2018.
(49) Ghandi, A.; Yeh, S.; Brandt, A.; Vafi, K.; Cai, H.; Wang, M. Q.; Scanlon, B. R.; Reedy, R. C. *Energy Intensity and Greenhouse Gas Emissions from Crude Oil Production in the Eagle Ford Region: Input Data and Analysis Methods*; Institute of Transportation Studies, University of California, Davis, 2015.
(50) Kurchatova, A.; Rogov, V.; Taratunina, N. Geochemical Anomalies of Frozen Ground Due to Hydrocarbon Migration in West Siberian Cryolithozone. *Geosciences* **2018**, *8* (12), 430. https://doi.org/10.3390/geosciences8120430.
(51) Klyusov, V. A.; Ershov, A. A. Experience in Operation of Processing Equipment for Gas Pretreating Systems at the Yamburg Gas-Condensate Field. *Chem. Pet. Eng.* **1995**, *31* (12), 759–761. https://doi.org/10.1007/BF01155172.
(52) Andreev, O. P.; Arabsky, A. K.; Bashkin, V. N.; Priputina, I. V.; Barsukov, P. A. *Assessment of the Life Cycle of Nitrogen and Carbon Compounds Emitted in the Arctic: The Example of Natural Gas Fields of Gazprom Dobycha Yamburg*; International Gas Union Research Conference, 2011.
(53) Seligman, B. J. Key Factors Influencing the Reliability of Trunk Gas Pipelines in the West Siberian North, Pembroke College, 1998.
(54) Lechtenbömer, S.; Assonov, S. S.; Dienst, C.; Brenninkmeijer, C.; Fischedick, M.; Hanke, T.; Langrock, T. *Greenhouse Gas Emissions from the Russian Natural Gas Export Pipeline System*; Wuppertal Institute for Climate, Enivornment, Energy/Max-Planck-Institute for Chemistry.
(55) Batalin, O.; Vafina, N. Condensation Mechanism of Hydrocarbon Field Formation. *Sci. Rep.* **2017**, *7* (1), 10253. https://doi.org/10.1038/s41598-017-10585-7.
(56) *Biogeochemical Technologies for Managing Pollution in Polar Ecosystems*, 1st ed. 2017.; Bashkin, V. N., Ed.; Environmental Pollution; Springer International Publishing : Imprint: Springer: Cham, 2017. https://doi.org/10.1007/978-3-319-41805-6.
(57) Department of Energy. *Liquefied Natural Gas: Understanding the Basic Facts*; DOE/FE-0489; United States of America, 2005.
(58) Aali, J.; Rahimpour-Bonab, H.; Kamali, M. R. Geochemistry and Origin of the World's Largest Gas Field from Persian Gulf, Iran. *J. Pet. Sci. Eng.* **2006**, *50* (3–4), 161–175. https://doi.org/10.1016/j.petrol.2005.12.004.
(59) Murillo, C. A. *Natural Gas Liquids in North America: Overview and Outlook to 2035*; 130; Canadian Energy Research Institute, 2012.

(60) Natural Gas Intelligence. Information about the Duvernay Shale https://www.naturalgasintel.com/information-about-the-duvernay-shale/.
(61) Dieckmann, V.; Fowler, M.; Horsfield, B. Predicting the Composition of Natural Gas Generated by the Duvernay Formation (Western Canada Sedimentary Basin) Using a Compositional Kinetic Approach. *Org. Geochem.* **2004**, *35* (7), 845–862. https://doi.org/10.1016/j.orggeochem.2004.02.010.
(62) Alberta Energy Regulator. Natural Gas, 2021.
(63) U.S. Environmental Protection Agency (EPA). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2019, 2021.
(64) Alvarez, R. A.; Zavala-Araiza, D.; Lyon, D. R.; Allen, D. T.; Barkley, Z. R.; Brandt, A. R.; Davis, K. J.; Herndon, S. C.; Jacob, D. J.; Karion, A.; Kort, E. A.; Lamb, B. K.; Lauvaux, T.; Maasakkers, J. D.; Marchese, A. J.; Omara, M.; Pacala, S. W.; Peischl, J.; Robinson, A. L.; Shepson, P. B.; Sweeney, C.; Townsend-Small, A.; Wofsy, S. C.; Hamburg, S. P. Assessment of Methane Emissions from the U.S. Oil and Gas Supply Chain. *Science* **2018**, eaar7204. https://doi.org/10.1126/science.aar7204.
(65) IEA. Methane Tracker Database: Interactive Database of Country and Regional Estimates for Methane Emissions and Abatement Options, 2021.
(66) Scarpelli, T. R.; Jacob, D. J.; Maasakkers, J. D.; Sulprizio, M. P.; Sheng, J.-X.; Rose, K.; Romeo, L.; Worden, J. R.; Janssens-Maenhout, G. A Global Gridded (0.1° × 0.1°) Inventory of Methane Emissions from Oil, Gas, and Coal Exploitation Based on National Reports to the United Nations Framework Convention on Climate Change. *Earth Syst. Sci. Data* **2020**, *12* (1), 563–575. https://doi.org/10.5194/essd-12-563-2020.
(67) Saunois, M.; Stavert, A. R.; Poulter, B.; Bousquet, P.; Canadell, J. G.; Jackson, R. B.; Raymond, P. A.; Dlugokencky, E. J.; Houweling, S.; Patra, P. K.; Ciais, P.; Arora, V. K.; Bastviken, D.; Bergamaschi, P.; Blake, D. R.; Brailsford, G.; Bruhwiler, L.; Carlson, K. M.; Carrol, M.; Castaldi, S.; Chandra, N.; Crevoisier, C.; Crill, P. M.; Covey, K.; Curry, C. L.; Etiope, G.; Frankenberg, C.; Gedney, N.; Hegglin, M. I.; Höglund-Isaksson, L.; Hugelius, G.; Ishizawa, M.; Ito, A.; Janssens-Maenhout, G.; Jensen, K. M.; Joos, F.; Kleinen, T.; Krummel, P. B.; Langenfelds, R. L.; Laruelle, G. G.; Liu, L.; Machida, T.; Maksyutov, S.; McDonald, K. C.; McNorton, J.; Miller, P. A.; Melton, J. R.; Morino, I.; Müller, J.; Murguia-Flores, F.; Naik, V.; Niwa, Y.; Noce, S.; O'Doherty, S.; Parker, R. J.; Peng, C.; Peng, S.; Peters, G. P.; Prigent, C.; Prinn, R.; Ramonet, M.; Regnier, P.; Riley, W. J.; Rosentreter, J. A.; Segers, A.; Simpson, I. J.; Shi, H.; Smith, S. J.; Steele, L. P.; Thornton, B. F.; Tian, H.; Tohjima, Y.; Tubiello, F. N.; Tsuruta, A.; Viovy, N.; Voulgarakis, A.; Weber, T. S.; van Weele, M.; van der Werf, G. R.; Weiss, R. F.; Worthy, D.; Wunch, D.; Yin, Y.; Yoshida, Y.; Zhang, W.; Zhang, Z.; Zhao, Y.; Zheng, B.; Zhu, Q.; Zhu, Q.; Zhuang, Q. The Global Methane Budget 2000–2017. *Earth Syst. Sci. Data* **2020**, *12* (3), 1561–1623. https://doi.org/10.5194/essd-12-1561-2020.
(68) Alvarez, R. A.; Pacala, S. W.; Winebrake, J. J.; Chameides, W. L.; Hamburg, S. P. Greater Focus Needed on Methane Leakage from Natural Gas Infrastructure. *Proc. Natl. Acad. Sci.* **2012**, *109* (17), 6435–6440. https://doi.org/10.1073/pnas.1202407109.
(69) Clemente, J. The Rise of U.S. Associated Natural Gas. *Forbes*. June 3, 2018.
(70) EPA (Environmental Protection Agency). Global Anthropogenic Non-CO2 Greenhouse Gas Emissions: 1990-2030., 2012.
(71) Liu, Z.; Ciais, P.; Deng, Z.; Lei, R.; Davis, S. J.; Feng, S.; Zheng, B.; Cui, D.; Dou, X.; Zhu, B.; Guo, R.; Ke, P.; Sun, T.; Lu, C.; He, P.; Wang, Y.; Yue, X.; Wang, Y.; Lei, Y.;

Zhou, H.; Cai, Z.; Wu, Y.; Guo, R.; Han, T.; Xue, J.; Boucher, O.; Boucher, E.; Chevallier, F.; Tanaka, K.; Wei, Y.; Zhong, H.; Kang, C.; Zhang, N.; Chen, B.; Xi, F.; Liu, M.; Bréon, F.-M.; Lu, Y.; Zhang, Q.; Guan, D.; Gong, P.; Kammen, D. M.; He, K.; Schellnhuber, H. J. Near-Real-Time Monitoring of Global CO2 Emissions Reveals the Effects of the COVID-19 Pandemic. *Nat. Commun.* **2020**, *11* (1), 5172. https://doi.org/10.1038/s41467-020-18922-7.