Supplementary Information to Accompany:

Home is where the pipeline ends: Characterization of volatile organic compounds present in natural gas at the point of the residential end-user

Drew R. Michanowicz[1,2 ,§,*], Archana Dayalu[3,§], Curtis L. Nordgaard[2], Jonathan J. Buonocore[1], Molly W. Fairchild[4], Robert Ackley[5], Jessica E. Schiff[6], Abbie Liu[6], Nathan G. Phillips[7], Audrey Schulman[4], Zeyneb Magavi[4], John D. Spengler[6]

[1]Harvard T.H. Chan School of Public Health, C-CHANGE, Boston, Massachusetts, 02215 United States
[2]PSE Healthy Energy, Oakland, California, 94612 United States
[3]Atmospheric and Environmental Research (AER), Lexington, Massachusetts, 02421 United States
[4]Home Energy Efficiency Team (HEET), Cambridge, Massachusetts, 02139 United States
[5]Gas Safety Inc., Southborough, Massachusetts, 01772 United States
[6]Harvard T.H. Chan School of Public Health, Boston, Massachusetts, 02215 United States
[7]Boston University, Boston, Massachusetts, 02215 United States

*Corresponding author, email: Drew@PSEHealthyEnergy.org / Michanow@HSPH.Harvard.edu
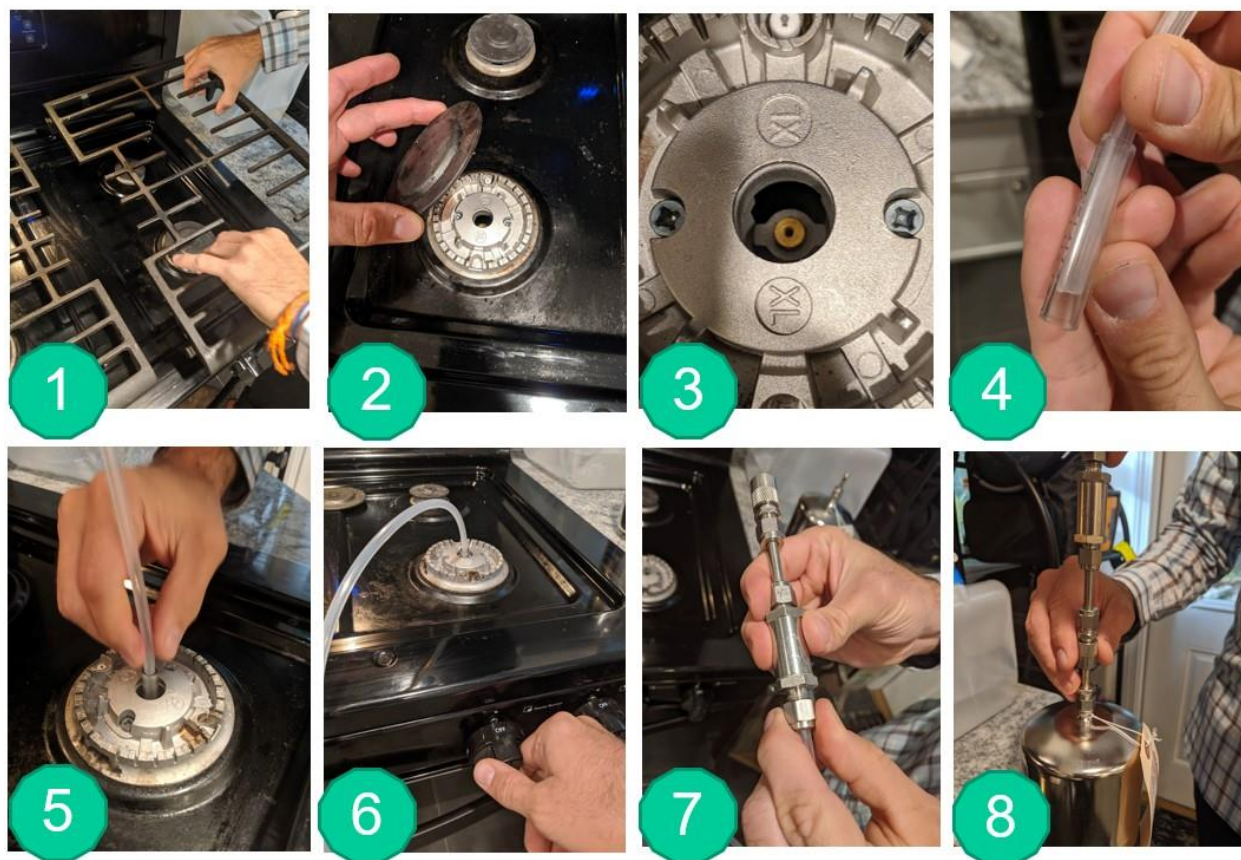§Contributed equally

Table of Contents

1.   Methods

1.1 Sample collection and processing

Whole-NG samples were collected from indoor NG-fired stovetop appliances and outdoor gas service lines across three local natural gas distribution territories. Compounds of interest included over 70 unique VOCs collected via grab samples in evacuated 1.4L Entech Silonite-lined canisters prepared by commercial environmental testing labs - Phoenix Environmental Laboratories (Phoenix Environmental Laboratories, Manchester, CT) and tested according to United States Environmental Protection Agency (USEPA) method TO-15. VOCs were subsequently separated by gas chromatography and measured by mass-selective detector techniques. Tentatively identified compounds (TICs) were also reported per sample. TICs were reported based on an evaluation of non-target compounds that had a peak area count > 10% of the nearest internal standard and met all mass spectral criteria. To further verify sufficient sample capture of NG, an aliquot from a subset of samples was tested for CH_4 and C_2H_6 using EPA method 3C performed by New England Testing Laboratory, West Warwick, RI (managed internally by Phoenix Environmental Laboratories). Analytical methods for ASTM D1945, ASTM D1946, and EPA 3C are identical, though differ by QC criteria and compounds specified for each method.

Stovetop sample collection entailed a direct in-line connection between stove natural gas outlet and sample canister via flexible Teflon-lined tubing (Figure S1). For all stoves sampled, the tubing diameter fully encompassed the gas outlet orifice (Figure S1, panel 4). The in-line connection effectively bypassed the stovetop ignition source (where present) and ensured that the sample was not diluted by ambient air and that minimal unburned NG was released. Once a secure connection was made, the gas flow knob was turned to medium-high. In most instances, the tubing was held in place by the second researcher. Once NG flow was detected from the open end via smell, the tubing was attached to an Entech Micro QT flow valve using a screw nut and cone washer (Figure S1, panel 7). The Micro QT valve was then engaged to the canister for sample collection that typically required 30 seconds for sample collection. Sample collection was completed once the cannister stopped producing a hissing sound. Stovetop models with known automatic shut-off devices or aeration devices were excluded at the outset, though not all could be identified as evidenced post-hoc by very low canister CH_4 concentrations.

Building NG service line sampling was identical to stovetop sampling, except samples were collected from service lines that were typically connected to outdoor NG grills or firepits. Typically, building NG service line sample collection entailed disconnecting an existing grill hose at the service line riser and connecting a new hose (⅜ inch or ½ inch Dozyant) via an auto-coupler quick connect. An in-line sample collection flow devoid of ambient intrusion was created by inserting the Teflon-lined tubing at least two feet into the new grill hose. Once inserted and NG flow was detected, the line was purged for 10 seconds, and the sample collection proceeded similar to stovetop sample collection. Canisters were sampled within five days of arrival and returned within three days of sample collection. All lab analyses were completed within three weeks of sampling and chain of custody forms were maintained for all samples.

Figure S1. Stovetop sample collection process

Samples were considered to be donated by homeowners since they were collected after NG was delivered to the home and were not taken directly from utility supplies. For safety, researchers were equipped with a real-time Bascom Turner methane Gas-Rover and were each trained by a natural gas safety expert. All sampling procedures and safety protocols were IRB approved with additional approval from Harvard's Chief Research Compliance Officer. This project received IRB approval on October 29th, 2019 from the Harvard T.H. Chan School of Public Health, ref: IRB19-1587 titled "Contaminants in the Kitchen."

Participant questionnaires were completed in-person during the pre-COVID initial site visits (Winter 2020). The survey included basic demographics, home characteristics that relate to in-home appliances, fuel types, ventilation, air exchange rates and other potential confounding factors related to housing conditions. A respondent booklet, which lists multiple answer choices for the baseline questionnaire, was handed to participants while administering the questionnaire to facilitate answer selection. Study team members administering survey questions prompted participants to the appropriate page of the booklet, which is coded as needed at the beginning of each question.

In response to the COVID-19 pandemic, sampling was paused in March 2020 and resumed in June 2020 with a redesigned sample protocol. To limit the time spent in the field in participant's homes, we resumed sampling at a subset of locations to take advantage of the familiarity between participants and researchers while still hoping to achieve a subset of our original study goals. Recruitment was also resumed to identify outdoor in-line connected natural gas grills (not propane tanks) as an alternate sampling method for the home's natural gas supply that did not require entering the home during the COVID-19 pandemic. All scheduling and communication on the day of sampling was done remotely, to eliminate the need for any face-to-face communication, and provide real-time communication for safety and questions.
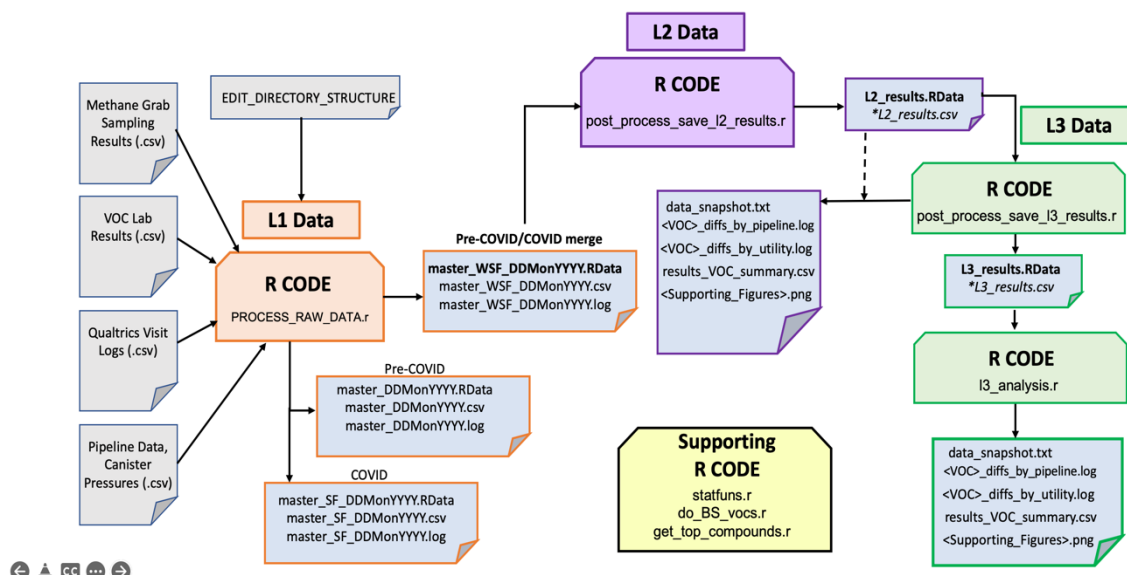
1.2 Data processing

1.2.1 Field data analysis

We constructed a semi-automated R-based (R Core Team, 2022) data management system for field data processing, quality control, and analysis. Data was merged across multiple sources, from in-field Qualtricssurvey (Qualtrics, Provo, UT) responses to natural gas pipeline data, to field sample results from commercial laboratories. The data management system was developed to ensure minimal user interaction, incorporating automated Quality Assurance/Control (QA/QC) checkpoints throughout. A flow chart for the data processing and analysis system is displayed in Figure S2. All supporting code and files displayed in Figure S2 are in a publication-ready final state. Supporting data (including L3 data set) and associated documentation can be downloaded from https://doi.org/10.7910/DVN/XGNGEO

As displayed in Figure S2, the analysis output includes supporting figures and breakdown of Volatile Organic Compound (VOC) concentration differences by pipeline and utility (note that pipeline designation is approximate). As justified in the manuscript text, we further focused our analysis on the VOCs Benzene, Toluene, Ethylbenzene, and Xylene (collectively, BTEX) due to their broad relevance to atmospheric chemistry and human health. A snapshot of the final analysis-ready L3 data set reveals the following:

- Sample Retention Percentage relative to L2: 83.9
- Samples (non-blank): 234
- No. Unique Locations: 69
- Date Range: 2019-12-17 18:50:00 2021-05-23 13:15:00
- Frequency Summary: Monthly 68 Summer 41 Winter 125
- Season: Fall-SON 61 Spring-MAM 35 Summer-JJA 11 Winter-DJF 127
- Sample Types: Grill 126 Stove 108
- Utility (Unique Loc): Columbia 20 Eversource 20 NationalGrid 29
- Utility (All Samples): Columbia 47 Eversource 50 NationalGrid 137
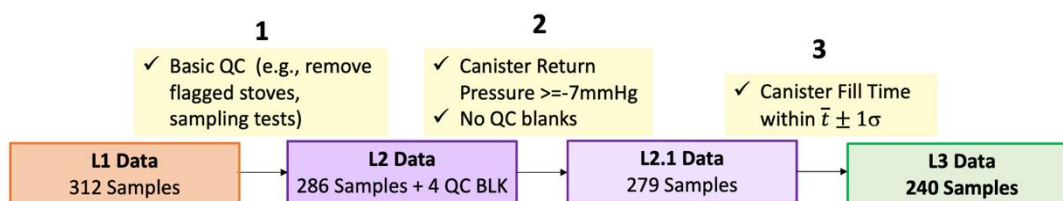- Pipeline (Unique Loc): Algonquin 43 Tennessee 26

Figure S2. Data Processing and Analysis Schematic. Naming convention of master files (DD, Mon, YYYY) reflect date of processing, and processed data files incorporate all raw data available up to that date.

1.2.2 Data QC and filtering procedure

The samples went through two main phases of QC and filtration to ensure that the final analysis data represented pure NG. The first level of filtering removed samples flagged in the field as questionable, including stoves that were later found to have safety shut-off features or aeration devices that significantly attenuated sampling of the NG stream. In addition, the L1 data set contained a series of test samples as part of sampling methodology development; these were also removed in the first stage of filtering. The first stage of filtering resulted in an intermediate (L2) data set that contained 290 samples (including four QC blanks). The second stage of filtering (L2 to final L3 data) involved removal of the four QC blanks and removal of insufficiently filled sample canisters or canisters with potential ambient air intrusion.
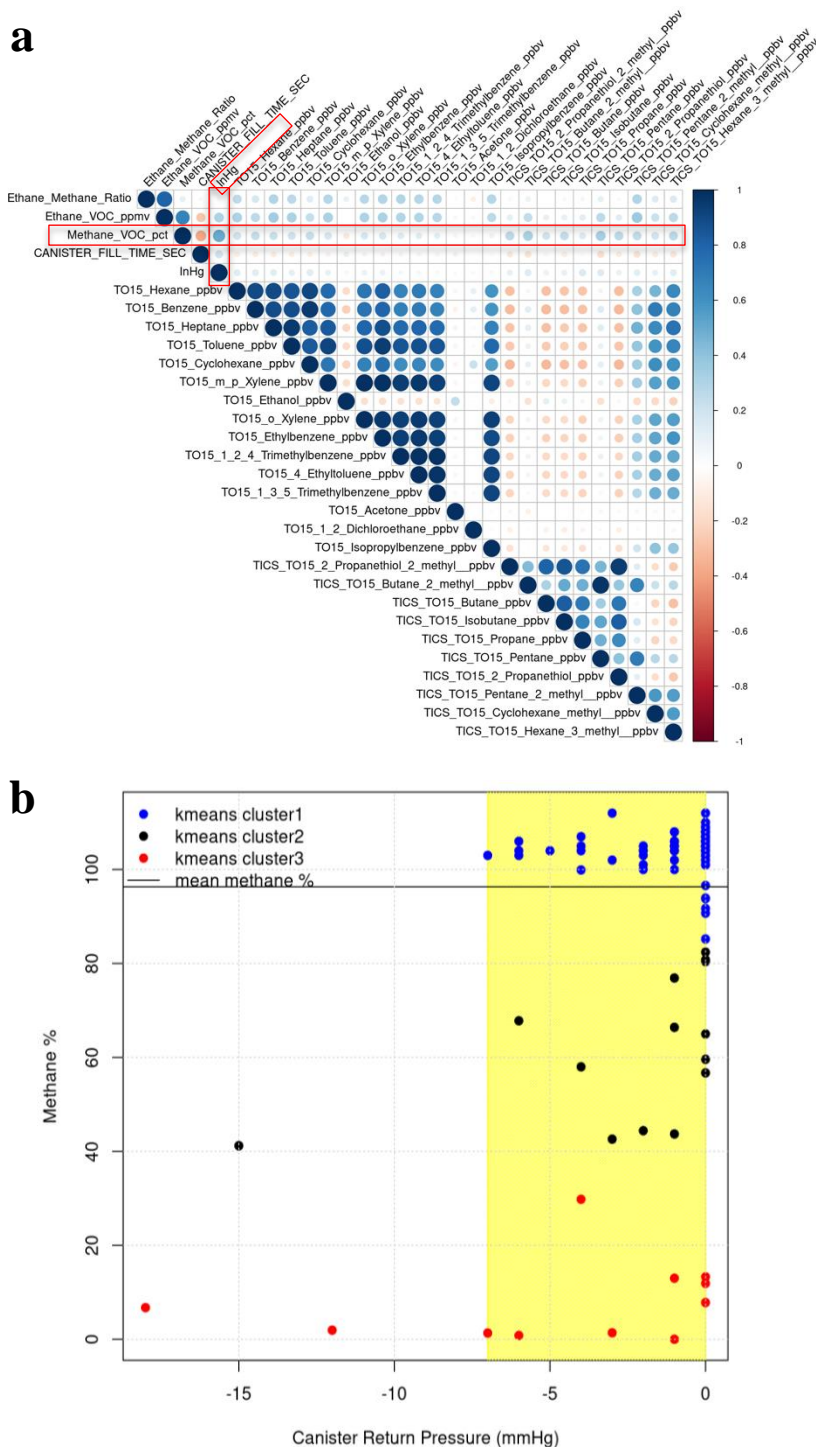
In March 2020, a 19-sample subset of field samples was tested for methane (CH_4) percentage prior to the COVID-19 sampling pause (pre-COVID). We found that the amount of CH_4 captured in each stove could be highly variable, ranging from 2% to 90% with a median of 43% (Figure S3). For reference, typical CH_4 composition in pure natural gas is 78%-90% (1-3).With suspension of field work in March 2020, we could not enact a field plan that immediately incorporated CH_4 testing. Sampling resumed in July 2020 and continued through May 2021; all future samples came with methane % and ethane:methane ratio. Our goal was to salvage as many pre-COVID data points as possible, while maintaining confidence that all samples selected for our final analysis-ready data were representative of the pure end-use natural gas stream in the absence of methane percentage information. In addition to selecting pre-COVID data points that were likely representative of the pure natural gas stream, a sample filtering scheme was established to ensure only selection of high-quality data points for the final analysis product.

Figure S3. Diagram of the data filtering process from raw (L1) to intermediate (L2) to analysis-ready (L3). Sample size in each compartment indicates the number of samples retained from the previous filter.
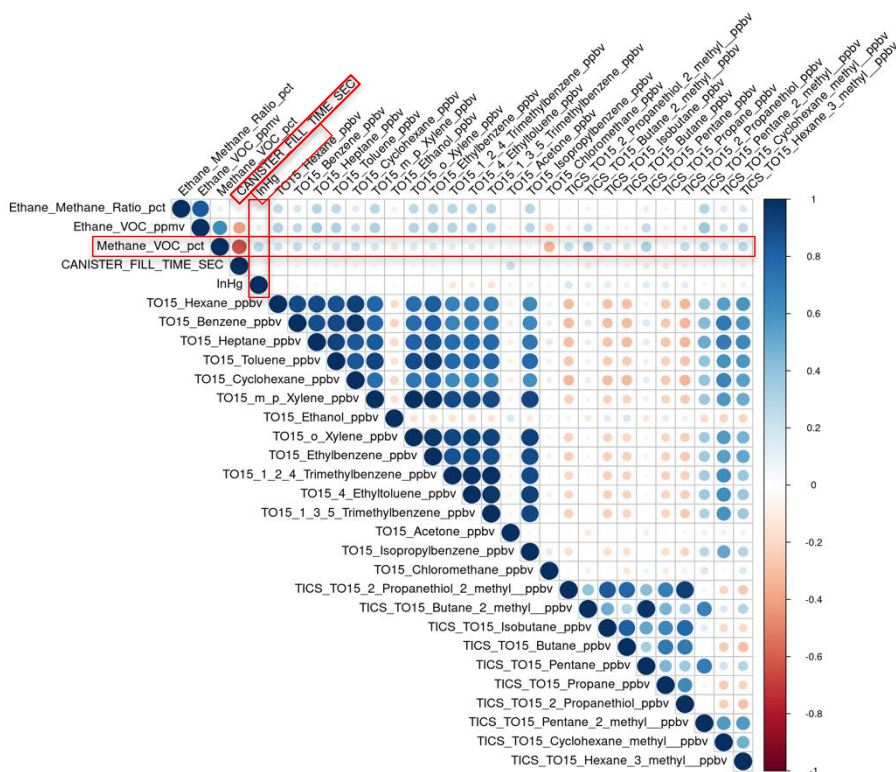
Not removing low sample capture samples would result in trace gas underestimation. Thus, we established a 3-level QC/filter system that is summarized in Figure S3. Based on a correlation plot of key variables in the L2 data set, we found that the canister return pressure was strongly correlated with methane capture %. This likely reflected the fact that safety features preventing continued gas flow out of the burner gas valve prevented complete filling of the evacuated -30 millimeters mercury (mmHg) canister, while also entraining ambient air. Figure S4a displays the correlation plot corresponding to the 286-sample L2 data set; Figure S3b displays

Figure S4. L2 Filtering Scheme, Phase 1, using Canister Fill Pressure relative to the -30mmHg evacuated sample canister. (a) Correlation plot for key variables with Methane % and Canister Fill Pressure highlighted. (b) Methane % plotted as a function of Canister Fill Pressure and clustered using 3 k-means clusters. The minimum allowed pressure selection criteria of -7mmHg was established by selecting the minimum pressure represented in Cluster 1.

a

b
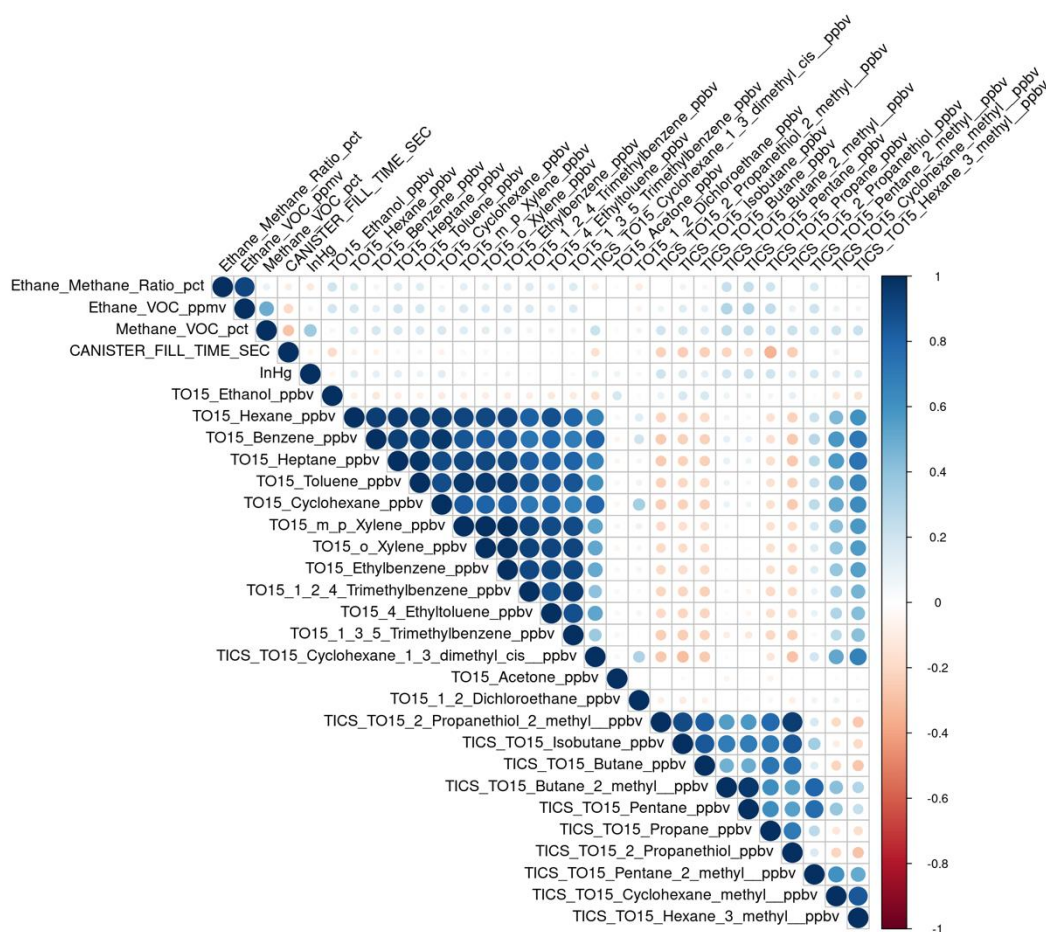
Figure S5. L2 Filtering Scheme, Phase 2, using threshold of Canister Fill Time outside of ± 1σ of mean fill time. (a) Correlation plot for key L2.1 variables with Methane % and Canister Fill Time highlighted, and (b) Methane percent plotted as a function of Canister Fill Time and colored according to original Phase 1 fill pressure clusters for direct comparison across filter phases.

the clustering of points corresponding to methane % with the lower bound of acceptable canister pressure being -7mmHg. We then examined the correlations in the subsequent L2.1 data set (after removing seven samples) and found that canister fill time was secondarily revealed as a strong negative correlate with methane capture. That is, the longer the time taken to fill the canister, the higher the likelihood of a lower-pressure ambient (and/or contaminated headspace) air intrusion either due to a poor seal or a stove safety feature. Figure S5 displays the L2.1 correlation plot (Figure S5a) and the relationship between canister fill time and % methane capture (Figure S5b).

Figure S6 displays the correlation plot for the final L3 dataset. Unless explicitly specified, the L3 data set is used for all analysis pertaining to the end-user natural gas stream. We note that there are weak-to-no correlations with methane % across variables, relative to previous filtering levels (Figures S4a and S5a). This provides additional confidence in a final analysis data set: the lack of strong correlations with Methane % indicates consistency of methane percentage across final data set and broad representation of NG methane with other natural gas components.
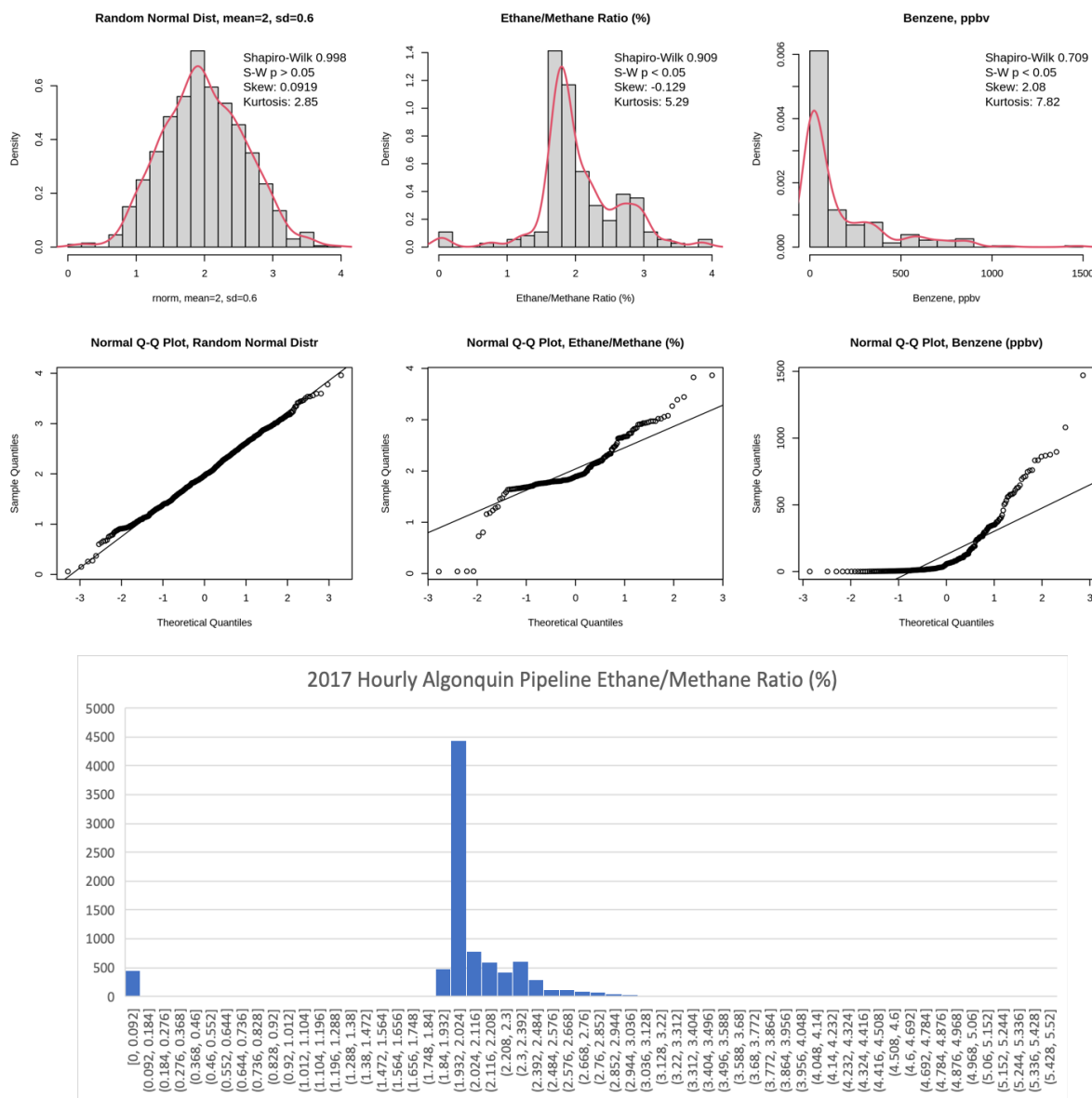

Figure S6. Correlation plot for main variables present in L3 final analysis-ready data set. Note the weak-to-no correlations with Methane % across variables, relative to previous filtering levels. The lack of strong correlations of variables with Methane % indicates consistency of Methane % across final dataset, providing additional confidence in the filtering procedure.

2. Results

2.1 Ethane to methane ratio

Figure S7 displays the ethane:methane ratio and relevant statistics and comparisons (including comparisons with pipeline data), demonstrating robust natural gas capture.

Figure S7. Ethane:Methane Ratio (%) in L3 data. *Top panel*: histograms with probability density, normalcy test statistics, skew, and kurtosis for (left) a reference random normal distribution with similar spread to the ethane:methane ratios; (middle) ethane:methane ratios; (right) typical NMVOC distribution in our NG samples, here Benzene. Skew of 0 and kurtosis of 3 represent perfectly normal data (with associated Shapiro-Wilk p-value > 0.05). *Middle*: Q-Q normal plots for each of the distributions in the top panel. *Bottom:* Hourly pipeline ethane:methane ratios (%) reported as a range for the Algonquin pipeline (4).

2.2 Sensitivity tests

Our early sample design involved conducting sensitivity tests on stovetop sampling including (1) varying burner location across the cooktop, and (2) with and without pre-boiling to clear out old gas and/or dead space in the gas lines. In the case of the preboiling test, we varied the sample start time from cold start to 3-minute boil to 10-minute boil. The purpose of the pre-boil test was to assess the sensitivity of NG composition to age of gas in the riser. There were no significant differences observed among samples based on burner location or extent to which gas line was purged prior to the sampling. Note that we did not conduct sensitivity tests for situations where gas was sitting in the riser for longer than 48 hours.

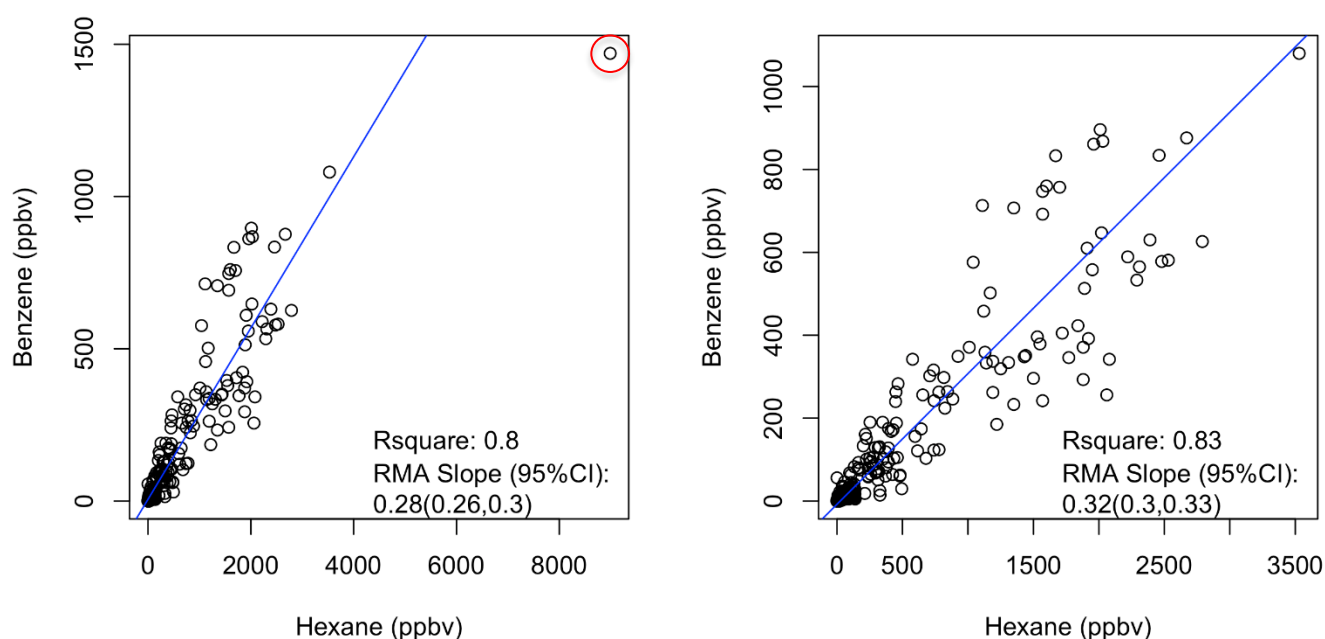## 2.3 Sample spatial and temporal variability

We formally tested the differential variability of space (σ_{sp}^2) vs. time (σ_t^2). σ_t^2 was calculated by pooling variances from each unique location having two or more samples across multiple dates (i.e., repeat samples over time). Similarly, σ_{sp}^2 was calculated by pooling variances from each sampling day having two or more samples across multiple locations. Sample size n was the total number of samples used in each temporal and spatial variance calculation. The F-statistic F ($\sigma_t^2/ \sigma_{sp}^2$) provides the ratio of the two pooled variances; the greater the deviation from 1 (in this case, at the 95% significance level) the more likely the variances are unequal (i.e., alternative hypothesis H_1) given H_0 that assumes equal variances. As shown in Table S1, all BTEX constituents have significantly higher temporal variance than spatial variance.

Table S1. Comparison of pooled variances (s²) of key NG constituents, pooled over time (σ_t^2) and space (σ_{sp}^2).

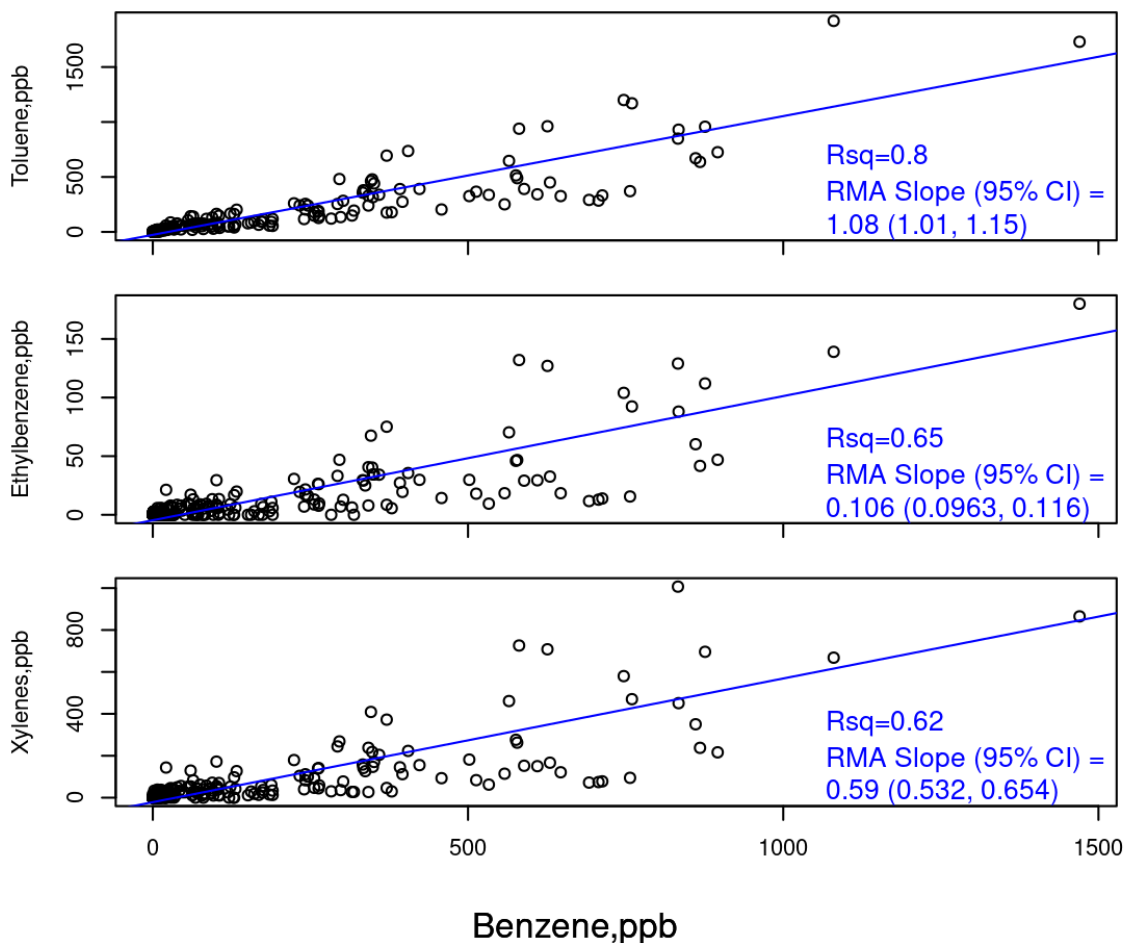|	Benzene	Toluene	Ethyl-benzene	Xylenes	Ethane/Methane	Methane
σ_t^2	43731	44309	464.10	14628	0.23746	18.5008
n(t)	160	160	160	160	142	142
σ_{sp}^2	21222	21824	202.52	5781	0.20729	51.7547
n(sp)	175	175	175	175	148	148
F($\sigma_t^2/ \sigma_{sp}^2$)	2.06	2.03	2.29	2.53	1.15	0.357
F(95%LCI)	1.52	1.50	1.69	1.86	0.823	0.258
F(95% UCI)	2.79	2.75	3.10	3.43	1.59	0.495
H_0 or H_1?	H1 ($\sigma_t^2 > \sigma_{sp}^2$)	H1 ($\sigma_t^2 > \sigma_{sp}^2$)	H1 ($\sigma_t^2 > \sigma_{sp}^2$)	H1 ($\sigma_t^2 > \sigma_{sp}^2$)	H0 ($\sigma_t^2 = \sigma_{sp}^2$)	H1 ($\sigma_t^2 < \sigma_{sp}^2$)

2.4 Co-variance of benzene and hexane in natural gas

We found that hexane – which is commonly reported in hourly NG quality postings data – was highly correlated with benzene within our dataset of 234 whole-NG samples. Correlation between hexane and benzene was highly significant with an R^2 of 0.80 and $p < 0.00001$. The 1:1 plots and statistics are included in Figure S8. Notably, one outlier sample reported an extremely high hexane value > 8000ppbv highlighted in red in Figure S8. Figure S8 shows correlation plots and statistics with (right panel) and without this outlier (left panel). These abundance ratios provide a first-order approximation of benzene content in consumer-grade natural gas where hexane is reported in hourly transmission gas postings data.

Figure S8. Correlation plots of Benzene to Hexane. *Left panel:* 1:1 plot with all 234 samples included. *Right panel*: 1:1 plot with one hexane outlier removed (n = 233).

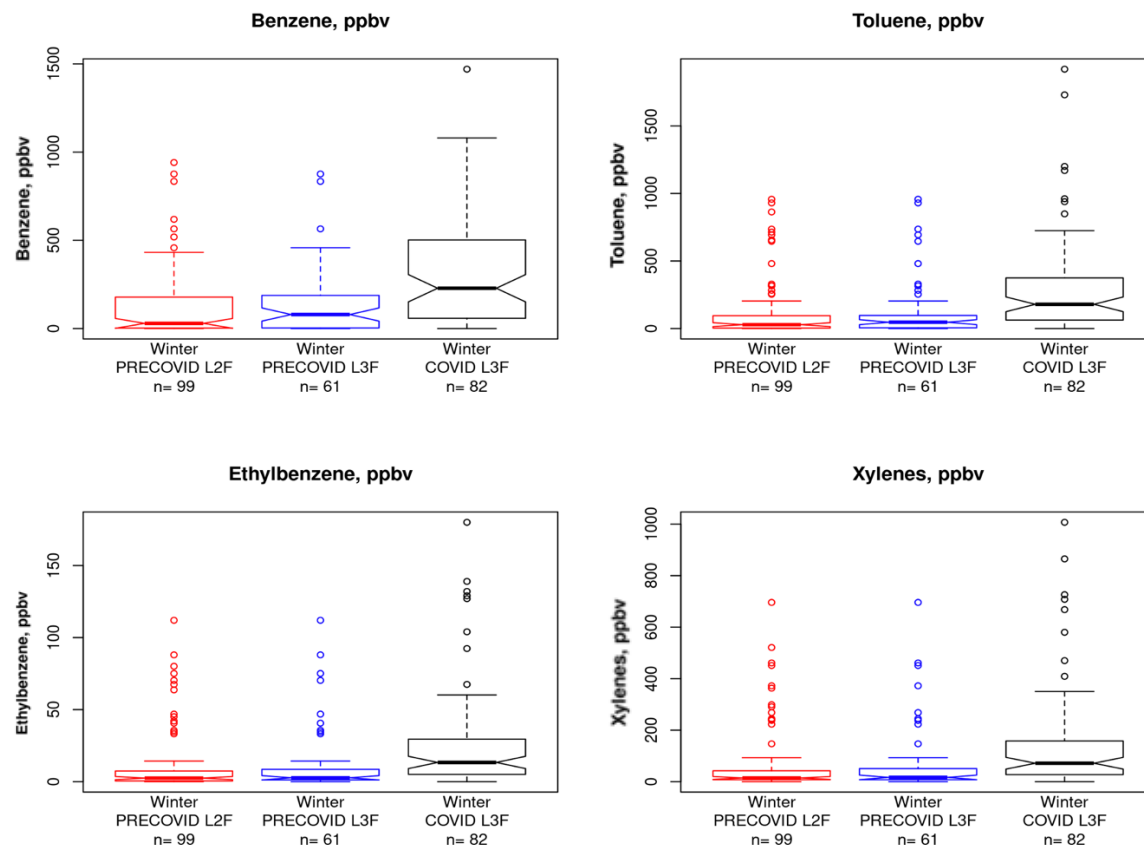## 2.5 Co-variance of BTEX components in natural gas

Figure S9. Correlation plots of Benzene with Toluene, Ethylbenzene, and Xylenes.

2.6 Comparison of winter BTEX sample profiles: pre-COVID and COVID sample phases

Due to the COVID sampling pause, the only season represented fully across both sampling years was winter (December, January, and February); all other seasons had one sampling phase. Therefore, to assess variability within the winter season, we separately examined the two years of winter data. We further examined the pre-COVID data across filtering levels (namely, L2 vs L3) to additionally gauge the impacts of the methane filtering process on the overall signal. Figure S10 summarizes the results for BTEX. As expected, we note that removing samples with ambient intrusion through the filtering process skews the filtered samples to a higher BTEX concentration. But – with the possible exception of Benzene – the overall impact of removing the <30 pre-covid winter samples on pre-COVID BTEX is not statistically significant. Meanwhile, the COVID-era (i.e., winter 2020) L3 filtered data BTEX subset is significantly higher than pre-COVID (i.e.,

winter 2019). Given the strong seasonality in BTEX content of natural gas (Section 3.3), we hypothesize that the higher BTEX in the winter 2020 is linked to the lower regional air temperatures (See main manuscript text) that impact the source, processing, storage, supply and, therefore, the content of VOCs in the gas stream.
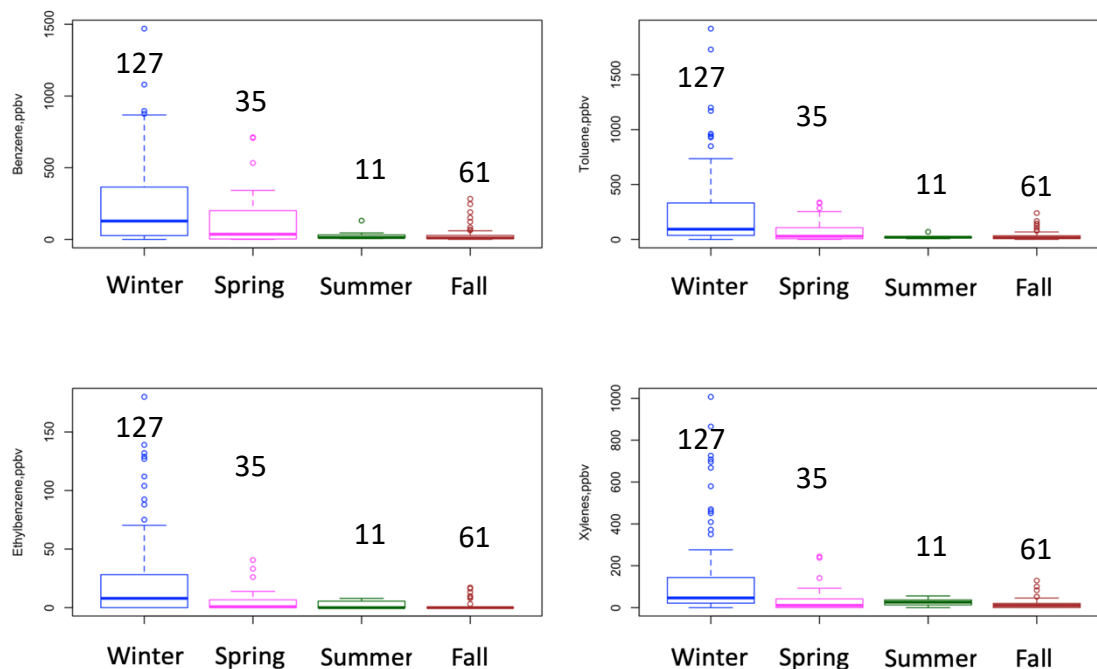

Figure S10. BTEX distributions across two winter seasons and filtering levels. Sample sizes provided for each category. Each level of filtered data is noted by the L<#>F convention.

2.7 Comparison of BTEX across seasons

Figure S11 summarizes the variability of BTEX constituents across seasons. As detailed previously, the unforeseen interruption of our original sampling strategy resulted in a disproportionately higher sampling during winter season, relative to other seasons (particularly the Summer). As noted in the main manuscript text, BTEX constituents in aggregate are ~3x Spring levels and ~8x levels in Summer and Fall.

Figure S11. Seasonal comparison of BTEX constituents. Sample sizes in each seasonal bin indicated above each box. Note that the winter season had higher representation relative to the remaining seasons due to COVID-19 pandemic-related changes in the original sample design.

2.8 Natural gas components and range of composition: This study and others

Table S2. Top 15 TICS (Tentatively Identified VOCs) present in L3 dataset. Due to data skew, statistical quantities are derived from a 1000-sample bootstrap of data. Non-detects were treated as zero in calculating descriptive statistics.
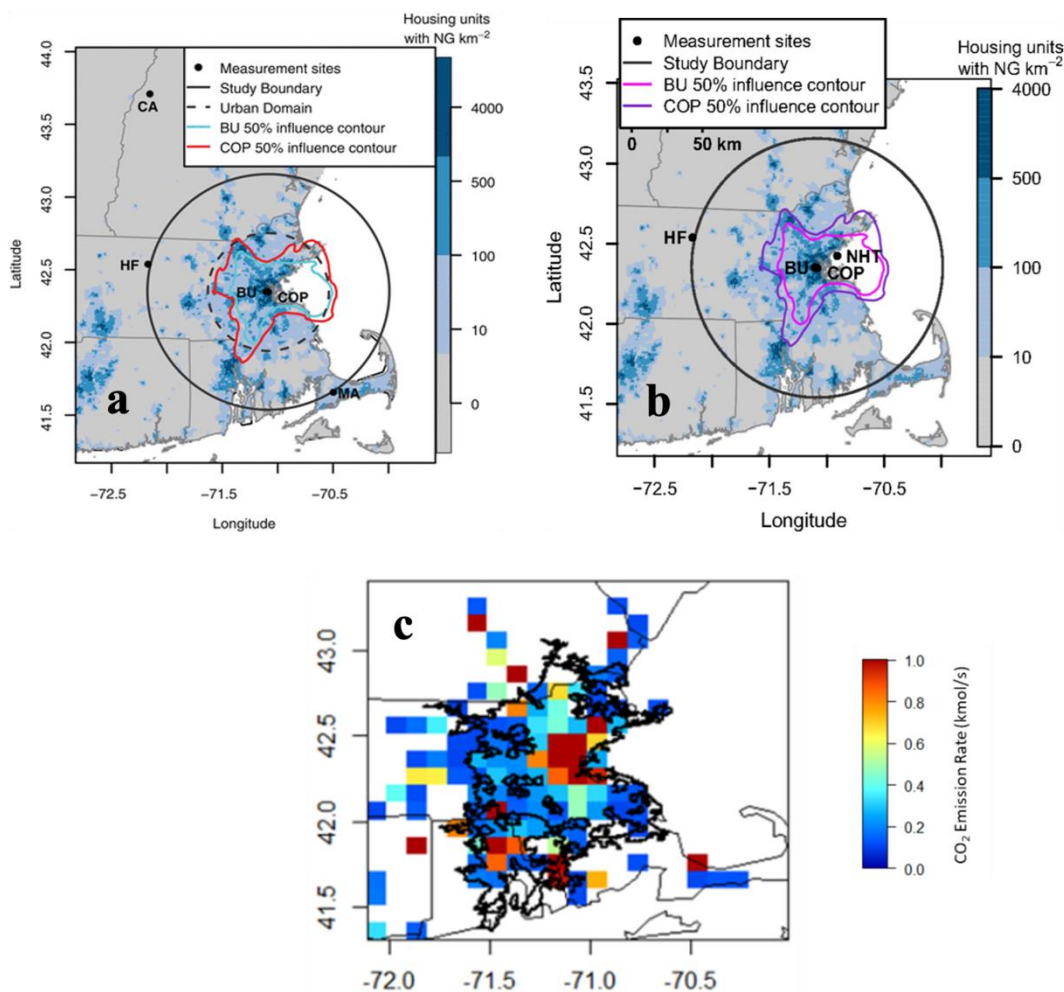
VOC (TICS)	n (% records)	Mean (ppbv)	SD (ppbv)	95% LCI (ppbv)	95% UCI (ppbv)
TICS_TO15_2_Propanethiol_2_methyl__ppbv	226 (95)	579	41	499	661
TICS_TO15_Butane_2_methyl__ppbv	224 (93)	304	19	267	341
TICS_TO15_Butane_ppbv	215 (90)	694	56	592	811
TICS_TO15_Pentane_ppbv	214 (89)	203	13	178	228
TICS_TO15_Isobutane_ppbv	212 (88)	638	43	553	719
TICS_TO15_Propane_ppbv	204 (85)	761	56	661	877
TICS_TO15_2_Propanethiol_ppbv	186 (78)	22.2	3	16.8	28.3
TICS_TO15_Pentane_2_methyl__ppbv	171 (71)	101	7.8	86.2	116
TICS_TO15_Cyclohexane_methyl__ppbv	142 (59)	132	16	102	164
TICS_TO15_Hexane_3_methyl__ppbv	122 (51)	34.8	3	29.3	40.8
TICS_TO15_Octane_ppbv	111 (46)	64.9	7.9	49.9	79.9
TICS_TO15_Pentane_3_methyl__ppbv	108 (45)	45	4.1	37	52.8
TICS_TO15_Neopentane_ppbv	106 (44)	21.6	2.1	17.5	25.8
TICS_TO15_Cyclopentane_methyl__ppbv	103 (43)	20.1	2.1	16	24.3
TICS_TO15_Butane_2_2_dimethyl__ppbv	90 (38)	15	1.6	12	18

Table S3. Names and content (mol %) of NG components covered by ASTM D1945-14a.Reproduced from Burger et al. (5).

component	mol %
helium	0.01–10
hydrogen	0.01–10
oxygen	0.01–20
nitrogen	0.01–100
carbon dioxide	0.01–20
methane	0.01–100
ethane	0.01–100
hydrogen sulfide	0.3–30
propane	0.01–100
isobutane	0.01–10
n-butane	0.01–10
neopentane	0.01–2
isopentane	0.01–2
n-pentane	0.01–2
hexane isomers	0.01–2
heptanes+	0.01–1

[a]The ranges listed are for the method and not the acceptable ranges for commercial natural gas. Components are listed in order of retention times.

2.9 NG Methane emissions and estimated corresponding BTEX fluxes: domain variations

Figure S12. Variations in spatial domains used for total NG-BTEX emissions estimates. (a) Sargent et al. (6) study boundaries; (b) McKain et al. (7) study boundary; and (c) Plant et al. (8) spatial domain.

We estimate BTEX fluxes associated with regional NG leaks based on observations from three studies (6-8)). The three studies have slight variations in their spatial domain as shown in Figure S11. Sargent et al. (6) use the same study radius as McKain et al. (7) but the relevant area within the radius is accounted for differently – Sargent et al. (6) used an area of ~12,000km^2 while McKain et al. (7) used an area of ~18,000km^2. Plant et al. (8) based their Boston NG emissions estimates on the 18,000km^2 area of McKain et al. (7).

We calculate BTEX annual emissions associated with natural gas using Eq (S1) below (reproduced from Eq 1 in the main text) following methodology similar to (9)Deighton et al. (9) :

$$BTEX_{kg\ yr^{-1}} = CH_{4_{kg\ yr^{-1}}} \times \frac{MW\ BTEX_{g\ mol^{-1}}}{MW\ CH_{4_{g\ mol^{-1}}}} \times \frac{[BTEX]_{ppbv}}{[CH_4]_{ppbv}} \qquad \text{(Eq S1)}$$

McKain et al. (7) and Sargent et al. (6) provided their NG CH_4 fluxes as g m^{-2} yr^{-1}. These were converted to Tg CH_4 yr^{-1} by multiplying by the area of each domain (Table S4). Plant et al. (8) only provided CH_4 emissions as Tg yr^{-1}; for reference, we have also provided Plant et al. (8) CH_4 fluxes in Table S4 as g m^{-2} yr^{-1}. The final reported BTEX emissions estimates for each of the three CH_4 emissions studies (Main Text, Table 5) incorporated uncertainty in the observed BTEX/CH_4 ratios and the CH_4 emissions. We did this by first obtaining a 1000-sample bootstrap of the BTEX/CH_4 ratios. Next, for each of the 1000 BTEX/CH_4 ratios we used Eq (S1) and data from Table S4 below to calculate an associated mean, lower, and upper, estimate of annual BTEX fluxes. Finally, we pooled all these annual BTEX flux estimates (ie., as kg yr^{-1}) and obtained the mean and quantiles (0.025, 0.975). These latter values are reported in Table 5 of the main text.

Table S4. CH_4 fluxes and domain areas from prior studies used in the BTEX annual flux estimates for the Greater Boston Region.

Study	NG CH_4 flux, g m^{-2} yr^{-1}	Greater Boston Study Domain Area (km^2)
McKain et al. (2015)	15.3 ± 3.5	18,064
Plant et al. (2018)	8.89 ± 2.2	18,064
Sargent et al. (2021)	14.0 ± 2.7	12,351

Table S5. Ratios of [BTEX]/[CH_4] (ppbv/ppbv) observed in this study's dataset. Statistics are derived from a 1000-sample bootstrap of BTEX/CH_4 data points.

ppbv/ppbv	Benzene/CH_4	Toluene/CH_4	Ethylbenzene/CH_4	Xylenes/CH_4
Mean	1.6E-7	1.5E-7	1.2E-8	7.2E-8
95% Lower CI	9.9E-8	8.8E-8	6.7E-9	4.1E-8
95% Upper CI	2.4E-7	2.3E-7	1.9E-8	1.1E-7

2.10 NG odorant content

Given the high detection frequencies of NG odorant content in whole-NG samples (see SI Table S2), two additional analyses were performed related to odorant content to 1) determine what proportion of samples met the federal NG odorant regulation (49 CFR 192.625), and to 2) determine what CH_4 concentration in NG equates to the odor detection thresholds of TBM and IPM. From 49 CFR 192.625, distribution-grade NG must contain odorants at a detectable concentration in air that is one-fifth of the lower explosive limit (LEL) for NG. Methane's conventional LEL is 5%, resulting in a one-fifth value of ~1% NG in air by volume. Therefore, to determine if NG odorant content is above or below odor detection thresholds, we refer to reported odor detection and recognition thresholds as reported elsewhere (see SI Table S6). Next, to determine the CH_4 equivalent at the odor detection threshold in NG, we first scaled all odorant concentrations to 100% CH_4 minus C_2H_6 %. This provided a scaled whole-NG odorant concentration equivalency to account for any samples that did not capture near 100% CH_4 + C_2H_6. Given the similar odor detection thresholds for both TBM and IPM of ~0.006 ppbv, we added TBM and IPM content per samples. Assuming odor detection of NG is primarily driven by TBM and IPM content, we then solved for the CH_4 concentration at which each sample fell below a TBM+IPM concentration of 0.006 ppbv. This resulted in a mean concentration of 21.3 ppmv CH_4 (95% CI: 16.7 - 25.9) and a median value of 8.8 ppmv indicating that the distribution was long-tailed resulting from the wide range in odorant concentrations.

Table S6. Odor detection thresholds

Odorant	CASRN	Odor Detection Threshold (ppb)
Tert-butyl mercaptan (TBM)	75-66-1	**0.029 [a]; 0.08[b]; 0.00626[c]**
Isopropyl mercaptan (IPM)	75-33-2	**0.0008[a]; 0.006[b]**

a Nagata and Takeuchi (10)
b Clanton and Schmidt (11)
c Wise, Rowe (12)

3. References

1.   ERG. Condensate tank oil and gas activities Prepared for: Texas Commission on Environmental Quality Air Quality Division. 2012;Accessed March 11, 2021. https://www.tceq.texas.gov/assets/public/implementation/air/sip/hgb/RS_Report_Attachment_C.pdf.

2.   Liss W, Thrasher W, Steinmetz G, Chowdiah P, Attari A. Variability of natural gas composition in select major metropolitan areas of the united states. final report, august 1990-february 1992. American Gas Association Labs., Cleveland, OH (United States); 1992.

3.   Kinder Morgan. Tennessee Gas Pipeline Hourly Gas Quality Data. Accessed April 22, 2022, Available from: https://pipeline2kindermorgancom/GasQuality/RepMainLinePointsaspx?code=TGP&type=5.

4.   Enbridge. Algonquin Gas Transmission Hourly Gas Quality. Accessed April, 2022, Available from: https://infopostenbridgecom/InfoPost/AGHomeasp?Pipe=AG.

5.   Burger JL, Lovestead TM, Bruno TJ. Composition of the C6+ fraction of natural gas by multiple porous layer open tubular capillaries maintained at low temperatures. Energy & Fuels. 2016;30(3):2119-26.

6.   Sargent MR, Floerchinger C, McKain K, Budney J, Gottlieb EW, Hutyra LR, et al. Majority of US urban natural gas emissions unaccounted for in inventories. Proceedings of the National Academy of Sciences. 2021;118(44).

7.   McKain K, Down A, Raciti SM, Budney J, Hutyra LR, Floerchinger C, et al. Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts. Proceedings of the National Academy of Sciences. 2015;112(7):1941-6.

8.   Plant G, Kort EA, Floerchinger C, Gvakharia A, Vimont I, Sweeney C. Large fugitive methane emissions from urban centers along the US East Coast. Geophysical research letters. 2019;46(14):8500-7.

9.   Deighton JA, Townsend-Small A, Sturmer SJ, Hoschouer J, Heldman L. Measurements show that marginal wells are a disproportionate source of methane relative to production. Journal of the Air & Waste Management Association. 2020;70(10):1030-42.

10.   Nagata Y, Takeuchi N. Measurement of odor threshold by triangle odor bag method. Odor measurement review. 2003;118:118-27.

11.   Clanton CJ, Schmidt DR. Sulfur compounds in gases emitted from stored manure. Transactions of the ASAE. 2000;43(5):1229.

12.   Wise P, Rowe S, Dalton P. Odorization of Natural Gas: What are the Challenges? bioRxiv. 2021.