Reconciling divergent estimates of oil and gas methane emissions

Daniel Zavala-Araiza[a], David R. Lyon[a], Ramón A. Alvarez[a], Kenneth J. Davis[b], Robert Harriss[a], Scott C. Herndon[c], Anna Karion[d,e,1], Eric Adam Kort[f], Brian K. Lamb[g], Xin Lan[h,2], Anthony J. Marchese[i], Stephen W. Pacala[j,3], Allen L. Robinson[k], Paul B. Shepson[l], Colm Sweeney[d,e], Robert Talbot[h], Amy Townsend-Small[m], Tara I. Yacovitch[c], Daniel J. Zimmerle[i], and Steven P. Hamburg[a,3]

[a]Environmental Defense Fund, Austin, TX 78701; [b]Carbon Now Cast, LLC, State College, PA 16802; [c]Aerodyne Research, Inc., Billerica, MA 01821; [d]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309; [e]Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO 80305; [f]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109; [g]Department of Civil and Environmental Engineering, Washington State University, Pullman, WA 99163; [h]Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX 77004; [i]Department of Mechanical Engineering, Colorado State University, Fort Collins, CO 80523; [j]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; [k]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213; [l]Department of Chemistry, Purdue University, West Lafayette, IN 47907; and [m]Department of Geology, University of Cincinnati, Cincinnati, OH 45221

Contributed by Stephen W. Pacala, November 10, 2015 (sent for review August 19, 2015; reviewed by Gregg Marland, Steven C. Wofsy, and Mark D. Zoback)

Published estimates of methane emissions from atmospheric data (top-down approaches) exceed those from source-based inventories (bottom-up approaches), leading to conflicting claims about the climate implications of fuel switching from coal or petroleum to natural gas. Based on data from a coordinated campaign in the Barnett Shale oil and gas-producing region of Texas, we find that top-down and bottom-up estimates of both total and fossil methane emissions agree within statistical confidence intervals (relative differences are 10% for fossil methane and 0.1% for total methane). We reduced uncertainty in top-down estimates by using repeated mass balance measurements, as well as ethane as a fingerprint for source attribution. Similarly, our bottom-up estimate incorporates a more complete count of facilities than past inventories, which omitted a significant number of major sources, and more effectively accounts for the influence of large emission sources using a statistical estimator that integrates observations from multiple ground-based measurement datasets. Two percent of oil and gas facilities in the Barnett accounts for half of methane emissions at any given time, and high-emitting facilities appear to be spatiotemporally variable. Measured oil and gas methane emissions are 90% larger than estimates based on the US Environmental Protection Agency's Greenhouse Gas Inventory and correspond to 1.5% of natural gas production. This rate of methane loss increases the 20-y climate impacts of natural gas consumed in the region by roughly 50%.

methane emissions | oil and gas emissions | greenhouse gas footprint | natural gas supply chain | Barnett Shale

Methane (CH_4), the principal component of natural gas, is a powerful greenhouse gas. Although natural gas emits less carbon dioxide (CO_2) per unit of energy than coal or oil when burned, CH_4 losses during the production, processing, transportation, and use of natural gas reduce its climate advantage compared with other fossil fuels. For example, if CH_4 losses are large enough (e.g., ~3% of production), new natural gas power plants can cause greater climate damage than new coal plants for decades or longer (~1% when comparing natural gas to diesel freight trucks) (1).

The lack of current data on CH_4 emissions, magnified by intense public concern over the broader environmental implications of shale gas development, has stimulated significant research to improve estimates of CH_4 emissions (2–18). A recurring theme in recent literature is that "top-down" (TD) approaches produce estimates that are significantly higher than those from "bottom-up" (BU) approaches. Concerns about available inventories and divergent TD and BU estimates create confusion regarding policy formulation and leave room for conflicting claims about the greenhouse gas implications of increased use of natural gas.

TD approaches for estimating total CH_4 emissions at the regional or larger scale include airborne mass balance (2–4, 19), atmospheric transport models (5, 6, 20–23), and enhancement ratios with well-constrained pollutants (17, 18, 24, 25). Apportionment of TD CH_4 emissions to oil and gas infrastructure and other fossil sources has been accomplished using several approaches, including: subtracting BU estimates of biogenic sources (3, 4), using isotope and hydrocarbon ratios, and through inverse modeling (6). This apportionment contributes to the uncertainty of TD estimates of fossil CH_4 emissions. Regardless of the methods used, TD estimates of total as well as oil and gas CH_4 emissions regularly exceed BU estimates based on emission inventories, which often rely on outdated emission factors (EFs), inadequate sampling (not representative of the

Significance

Past studies reporting divergent estimates of methane emissions from the natural gas supply chain have generated conflicting claims about the full greenhouse gas footprint of natural gas. Top-down estimates based on large-scale atmospheric sampling often exceed bottom-up estimates based on source-based emission inventories. In this work, we reconcile top-down and bottom-up methane emissions estimates in one of the country's major natural gas production basins using easily replicable measurement and data integration techniques. These convergent emissions estimates provide greater confidence that we can accurately characterize the sources of emissions, including the large impact that a small proportion of high-emitters have on total emissions and determine the implications for mitigation.

Author contributions: D.Z.-A., D.R.L., R.A.A., R.H., S.W.P., C.S., and S.P.H. designed research; D.Z.-A., D.R.L., R.A.A., R.H., S.C.H., A.K., E.A.K., X.L., A.J.M., A.L.R., P.B.S., C.S., R.T., A.T.-S., T.I.Y., and D.J.Z. performed research; D.Z.-A., D.R.L., R.A.A., A.K., E.A.K., X.L., A.J.M., S.W.P., A.L.R., P.B.S., C.S., R.T., A.T.-S., T.I.Y., and D.J.Z. analyzed data; and D.Z.-A., D.R.L., R.A.A., K.J.D., A.K., E.A.K., B.K.L., A.J.M., S.W.P., A.L.R., P.B.S., C.S., A.T.-S., D.J.Z., and S.P.H. wrote the paper.

Reviewers: G.M., Appalachian State University; S.C.W., Harvard University; and M.D.Z., Stanford University.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]Present address: National Institute of Standards and Technology, Gaithersburg, MD 20899.

[2]Present address: Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO 80305.

[3]To whom correspondence may be addressed. Email: pacala@princeton.edu or shamburg@edf.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522126112/-/DCSupplemental.

ENVIRONMENTAL SCIENCES

population of sites), and inaccurate counts and location of sites, facilities, and equipment (26) (*SI Appendix*, Fig. S1).

We examine these issues as they apply to the Barnett Shale region in north Texas, the first basin to see widespread use of horizontal drilling and hydraulic fracturing and that, today, produces 7% of US marketed natural gas (see the *SI Appendix* for maps and additional background about the Barnett Shale). BU estimates of CH_4 emissions in the Barnett, based on publically available data, vary by a factor of 3, exclude some sources, and are partially based on assumptions that are not locally calibrated and predate shale gas development (27).

The goals of this work were (*i*) to construct a custom BU CH_4 inventory that accurately represents the entire population of facilities within discrete source regions sampled by TD approaches, (*ii*) to quantify emissions from the major facility types in the region using methods that explicitly account for the contribution of high-emitters, and (*iii*) to identify the data collection and analytical elements necessary to reconcile TD and BU approaches.

This paper integrates seven datasets primarily drawn from a 2013 emission measurement campaign in the Barnett region (28) and supplemented with other recently published data: seven TD estimates from aircraft CH_4 measurements (19) (dataset 1); six TD estimates from aircraft ethane measurements (29) (dataset 2); a spatially explicit list of all oil and gas infrastructure in the region created by combining all available data (e.g., production sites, processing facilities, and compression stations) (27) (dataset 3); BU CH_4 measurements of a systematic sample of 186 Barnett production sites (30) (dataset 4); BU CH_4 measurements from a systematic national sample of 125 compressor stations and processing plants (12) (dataset 5); BU CH_4 and ethane measurements of 102 Barnett production sites, compressor stations, and processing plants that are biased toward high-emitters (31, 32) (dataset 6); and published EFs (27) (dataset 7), which provide emissions per type of emitter, for biogenic sources (e.g., landfills and feedlots) and minor fossil sources that are collectively responsible for <5% of total fossil emissions (well completions, losses from gathering and transmission pipelines, and losses during local distribution). BU datasets 4–6 are based on downwind measurements that characterize facility-wide emissions.

The primary TD estimates (dataset 1) were obtained from CH_4 concentration data collected by aircraft along transects bordering the core Barnett gas-producing region during seven flights in March and October 2013. In contrast, previous TD studies have relied on only one to two flights (2–4). The aircraft also measured ethane on six of the seven flights, and TD estimates of ethane emissions (dataset 2) are available for these days (see the *SI Appendix* for analysis of the effect of the number of flights on TD uncertainty and results from one additional flight considered to be an outlier).

Because CH_4 is emitted by both fossil and biogenic sources (e.g., wetlands, feedlots, and landfills), TD CH_4 measurements estimate the combined fossil and biogenic flux of a source region. In contrast, ethane has no biogenic sources and is the second most prevalent hydrocarbon in natural gas. Smith et al. (29) and Karion et al. (19) separated fossil from biogenic sources in the Barnett region by aircraft-based CH_4 and ethane sampling downwind of individual sources. We use their result that 79.5% [73.5–84%; 95% confidence interval (CI)] of Barnett CH_4 emissions comes from fossil sources to partition fossil CH_4 from total CH_4 emissions. An alternative, independent TD fossil CH_4 estimate was produced using dataset 2 and reanalysis of dataset 6, which agrees with the estimate reported below (*SI Appendix*, *Results and Discussion*).

Because directly measuring emissions from representative samples of every CH_4 source type in every region is costly, BU estimates often rely on published EFs combined with infrastructure counts analogous to dataset 3, known as "activity factors." In contrast, our BU estimates are based on extensive local measurements of facility-wide emissions from production sites, compressor stations, and processing plants in datasets 4 and 6, with two exceptions. First, although dataset 5 is from a national study of compressor stations and processing plants (12), the installed compression capacity of sites in the Barnett region, based on data published in ref. 27, is similar to the national sample in ref. 12. Second, dataset 7 uses published EFs or self-reported data to characterize the minor fossil sources and biogenic emissions (27).

The systematic sampling schemes behind datasets 4 and 5 were designed to characterize emissions from a representative distribution of facilities. For example, to produce dataset 4, measured facilities were distributed approximately evenly throughout the core of the Barnett production region. However, systematic sampling can also yield relatively few of the infrequent high-emitters that dominate total emissions (see *Results and Discussion* and *Methods*). The under-sampling of high-emitters is another likely reason for the gap between published BU and TD estimates, underscoring the need to characterize the extremes of emission rate distributions by deliberately targeting high-emitters. To obtain the relatively large sample of high-emitters in dataset 6, two teams drove around the Barnett region and estimated the emission rate of the source of each plume encountered (31, 32). Because high-emission sources produce plumes that remain above an instrument's detection limit over longer distances than low-emission sources, this method tends to favor identifying high-emitters (*SI Appendix*).

We used a statistical method to estimate emissions probability density functions (pdfs) from datasets 4–6, which are then used to derive EFs (*Methods* and *SI Appendix*). Results in the main text were determined using a statistical estimator that integrates the systematic samples (datasets 4 and 5) with the samples biased toward high-emitters (dataset 6), using a power law to estimate the bias (*Methods*). Two variations of the statistical estimator were examined; the first relies only on the systematic samples, and the second characterizes the high-emitter–biased samples based on Gaussian plume theory (see the *SI Appendix* for results and description of these two variations). Because the three estimators produced similar results, we only present results from the power law estimator in the main body of this paper.

Finally, the new BU estimate is enhanced by updated facility counts (dataset 3, assembled from diverse public records and remote imagery to locate infrastructure, notably compressor stations) not clearly identified in available records (27). To most directly compare with TD estimates, we created separate BU estimates for each source region corresponding to the seven mass balance flights (*SI Appendix*). There is a separate BU estimate for each TD estimate because each flight sampled a somewhat different source region, and the corresponding BU estimate includes only the facilities inside that region (see Fig. 1 and the *SI Appendix* for details about BU uncertainty). The source regions for the flights differ slightly from the 25-county Barnett production region (*SI Appendix*, Fig. S4). Consequently, to facilitate comparison with previous inventories, we also created a BU estimate that is spatially resolved for the Barnett 25-county region.

Results and Discussion

Convergence of TD and BU Estimates. The excellent agreement of the TD and BU estimates shows that it is possible to constrain regional CH_4 emissions relatively tightly with either approach (Fig. 1). The mean difference between the TD and BU estimates for total CH_4 emissions, expressed as a percentage of the average TD estimate is 0.1% ± 21% (95% CI) (*SI Appendix*); the corresponding difference for fossil methane is 10% ± 32%, expressed as a percentage of the mean TD fossil estimate. This implies that the BU datasets captured the effect of skewed emissions distributions (Fig. 2) and did not miss any major source of CH_4 in the region. Averaging data from multiple flights substantially reduced the CI for the mean TD estimate of total CH_4 emissions from all sources (mean with 95% CI is 71 ± 12 Mg CH_4/h, with a daily range of 41–88 Mg CH_4/h; Fig. 1, left column). We hypothesize that some of the

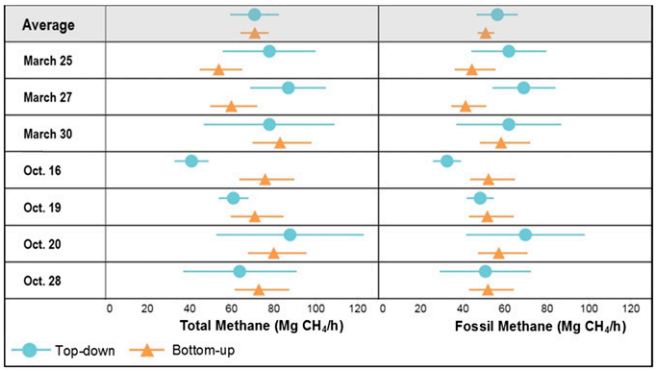

Fig. 1. Summary of aircraft-based TD (blue circles) and custom BU (orange triangles) emission estimates with 95% CIs from a 2013 measurement campaign in the Barnett Shale. The first row shows the seven-flight average, with subsequent rows showing estimates for each individual flight. The middle column shows estimated CH_4 from all sources, averaging 71 ± 12 Mg CH_4/h (TD) and 71 ± 7 Mg CH_4/h (BU). The right column shows estimates from fossil sources, averaging 56 ± 10 Mg CH_4/h (TD) and 51 ± 4 Mg CH_4/h (BU).

residual daily variation in total CH_4 emissions was caused by biogenic sources (33–35). The agreement also suggests well-designed TD or BU surveys can characterize emissions with sufficient accuracy to verify overall regional effectiveness of any future regulation of CH_4 emissions.

Methane Emissions Are Significantly Higher than Estimates Based on Public Inventories. Our spatially resolved BU oil and gas CH_4 emission estimate for the 25-county Barnett region is 59 Mg CH_4/h (48–73 Mg CH_4/h; 95% CI), with the three main sources being production sites (53%), compressor stations (31%), and processing plants (13%).

For context, emissions of 59 Mg CH_4/h represent a loss of 1.5% (1.2–1.9%) of total Barnett production (resolved for the Barnett 25-county region) during the study period, with a value of ~$100 million/y (at $200/Mg). Our BU estimate is 1.9 times the estimated emissions based on the US Environmental Protection Agency (EPA) Greenhouse Gas Inventory (36), 3.5 times that using the EPA Greenhouse Gas Reporting Program (37), and 5.5 times that using the Emissions Database for Global Atmospheric Research (EDGAR, Version 4.2) (38) (*SI Appendix*). The ratios of TD/BU reported for other basins using these three BU inventories show similar ratios (*SI Appendix*), suggesting that the BU approach used here could be used elsewhere to effectively characterize CH_4 emissions.

Lyon et al.'s (27) BU estimate of total oil and gas emissions for the 25-county Barnett region (46 Mg CH_4/h) is 20% lower than our BU estimate, with most of the difference attributed to production sites (see the *SI Appendix* for detailed comparison). Because both BU estimates used dataset 3 to generate the activity factors, the main cause for the difference is the EFs. In turn, these EFs differ because of the manner in which the systematic samples are integrated with the high-emitter–biased samples (fat-tail sites). The statistical estimator we developed produces a single, characteristic emission distribution representative of each source type, including the influence of high-emitters, thereby eliminating the need to assume a probability of fat-tail sites as done by Lyon et al. Our statistical framework predicts that the fat tail of production sites has a higher probability than assumed by Lyon et al. (Further details are in the *SI Appendix*.)

Using the Technology Warming Potential (TWP) framework of Alvarez et al. (1), we quantified the increase in cumulative radiative forcing (a metric indicative of climate damage) attributable to leakage of Barnett Shale gas. Each percentage of natural gas lost to the atmosphere before combustion adds roughly 30% to the 20-y radiative forcing due to the fuel-cycle CO_2 emissions alone (i.e., CO_2 emissions from end-use natural gas combustion plus upstream combustion, venting, and leakage) or 10–15% on a 100-y time horizon (*SI Appendix*). Therefore, CH_4 leakage of 1.5% from the Barnett natural gas supply chain increases the radiative forcing from fuel-cycle CO_2 emissions alone by about 50% over a 20-y basis (~20% on a 100-y basis). The measured Barnett methane leakage is low enough that gas fired electricity in this region causes less climate forcing than coal-fired electricity (see the *SI Appendix* for details, including effect of CH_4 mitigation). By contrast, use of compressed natural gas sourced in the Barnett instead

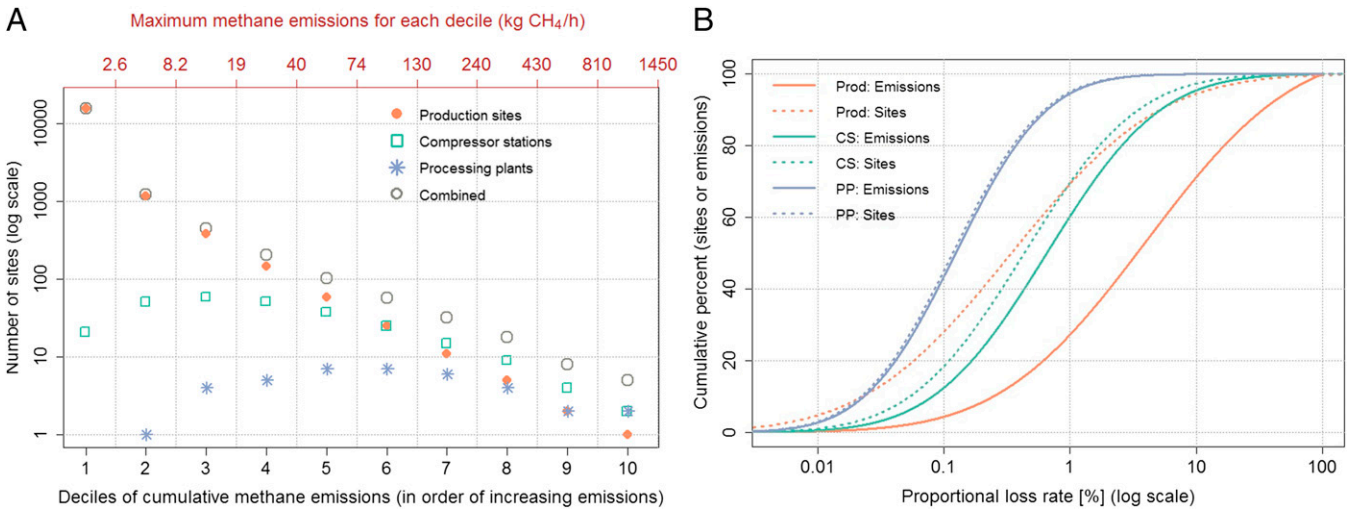

Fig. 2. Methane emissions are dominated at any one time by a few high-emitters. (*A*) Number of sites, by type, in each decile of cumulative emissions from production sites, processing plants, and compressor stations in the 25-county Barnett region. These three site types account for over 95% of regional oil and gas CH_4 emissions (*SI Appendix*). Upper *x* axis shows the maximum emission rate for each decile. (*B*) Cumulative percent of emissions (solid lines) and cumulative percent of sites (dotted lines) as a function of proportional loss rate (emissions divided by gas production or throughput) for production sites (Prod), compressor stations (CS), and processing plants (PP).

Table 1. EFs derived from the statistical estimator

Source	μ	σ	EF, kg CH_4/h
Production sites	−1.79 (−2.13, −1.45)	2.17 (1.96, 2.39)	1.76 (1.27, 2.45)
Compressor stations	3.05 (2.77, 3.32)	1.49 (1.32, 1.67)	64.2 (48.8, 84.4)
Processing plants	4.41 (3.92, 4.91)	1.31 (1.00, 1.62)	195 (121, 315)

The 95% CI appears between parentheses. $EF = e^{\mu+\frac{1}{2}\sigma^2}$.

of diesel for freight trucks would cause greater forcing for several decades (1, 39, 40).

Reported methane emission rates in other natural gas-producing basins sometimes differ from our Barnett results (2–4, 21). Differences between basins would be expected, even if estimates for all regions were obtained with our methods, because gas reservoirs vary by age, geologic properties, ease of maintenance (accessibility), and local practice, all of which may affect emission rates. For example, intermittent activities that can result in high short-term emissions, such as well completions and liquid unloadings (clearing the well bore of liquids) (9, 11), are more common in other basins relative to the Barnett in 2013. Finally, the long-distance transmission and storage of natural gas results in a substantial increment of CH_4 emissions that needs to be considered when analyzing the climate implications of natural gas consumption in regions that are not proximate to a production area (ref. 41 and *SI Appendix*).

Barnett Methane Emissions Are Dominated at Any One Time by a Few High Emitters. The estimated emission distributions imply that, at any one time, 2% of facilities in the Barnett region are responsible for half of the emissions, and 10% are responsible for 90% of emissions (Fig. 2*A*). High-emitters are divided roughly equally among production sites, compressors, and processing plants, but an individual production site is less likely to be a high-emitter than is a compressor station or processing plant. Even though a facility's emission rate depends only weakly on the total amount of gas produced or processed, facility-level emissions as a fraction of the total CH_4 produced or processed is a more effective metric than absolute emissions to identify sites with avoidable emissions (ref. 42 and *SI Appendix*). For example, our results indicate that 30% of production sites emit >1% of the natural gas they produce; these sites account for 70% of production site emissions (Fig. 2*B*).

Future work is needed to understand the characteristics that cause an individual site to be a high-emitter. A variety of avoidable operating conditions could lead to excess emissions at those sites, including both persistent and episodic phenomena (12, 42). Our work identifies facility-level high-emitters as drivers of regional emissions; documenting the root cause of avoidable operating conditions at a component level could improve the efficacy of mitigation strategies.

The large number of facilities in the Barnett region cause high-emitters to always be present, and these high-emitters seem to be spatially and temporally dynamic. Other studies and infrared imagery from helicopter flights suggest that hydrocarbon plumes escaping from facilities in the region often represent malfunctions or other avoidable conditions, such as a stuck valve or routine flashing, that could occur at any facility, rather than permanent design flaws in a few facilities (12, 41, 42). To reduce those emissions requires operators to quickly find and fix problems that are always present at the basin scale but that appear to occur at only a subset of sites at any one time, and move from place to place over time.

Conclusions

To inform the design of future research on oil and gas CH_4 emissions, we summarize several elements that contributed to the convergence of estimates in the Barnett Shale using independent TD and BU approaches:

i) Uncertainty in estimates of oil and gas CH_4 emissions using TD approaches can be reduced when replicate mass balance measurements produce a representative central estimate and a signature compound such as ethane is used to distinguish fossil CH_4 from biogenic CH_4.

ii) BU estimates require accurate facility counts of all major sources. A major reason that previous inventories underestimated emissions is the omission of numerous facilities (e.g., relatively high-emitting compressor stations in the Barnett).

iii) EFs require effective characterization of the entire distribution of sources. Two conditions must be satisfied to accomplish this: (*i*) the sampling strategy must capture the low-probability, high-emitting sources that define the fat tail of the distribution (10, 12, 26, 41), and (*ii*) emissions datasets must be integrated in such a way that the emission distributions accurately capture the magnitude and frequency of high-emitting sources. Here, we used a statistical estimator to model the skewed distributions from the major sources, integrating several measurement datasets, and deriving EFs that capture the dominant effect of high-emitters.

Methods

BU Estimates: Statistical Estimator. We developed a statistical estimator that integrates the systematic samples (datasets 4 and 5) with the samples biased toward high-emitters (dataset 6), using a power law to estimate the bias. We use the statistical estimator to produce emissions pdfs, which were then used to derive facility-level EFs for production sites, compressor stations, and processing plants. Additionally, we tested two variations of the statistical estimator: (*i*) we use only the systematic samples, and (*ii*) we integrate the high-emitter–biased samples using Gaussian plume theory (see the *SI Appendix* for results and description of these two variations). Because the three estimators produced similar results, we report results only from the power law estimator in the main body of this paper.

A plume of gas emitted from a point source blows downwind and spreads both in the orthogonal horizontal direction and vertically, which causes the concentration of the gas to decrease with increasing downwind distance from its source until it falls beneath the detection limit of an instrument. This means that the length of the detectable portion of a plume increases with emission strength. The length of the detectable portion of the plume also obviously depends on meteorological conditions.

Now consider driving on a road while sampling the concentration of CH_4 or ethane. Under semiflat terrain conditions (such as the ones present in the Barnett region), the area "searched" for emissions sources is larger for large sources than for small because large sources can be detected further downwind than small sources. We thus write this area as a function of the emissions strength *E: A(E)*. For a given source strength, one can imagine a line segment extending upwind from the road and terminating at the distance that would make the concentration on the road fall to the detection limit if there were a source of strength *E* at the end of the line segment. As the sampling vehicle moves on the road, the upwind line segment will change in length and direction as meteorological conditions change. Collectively, the area swept by the segment as it moves and changes length and direction is *A(E)*. This increases with *E* because the line segments increase in length with *E*.

Let *q(x)* be the lognormal probability density of the natural logarithm of emissions but biased by the fact that the area sampled grows with the

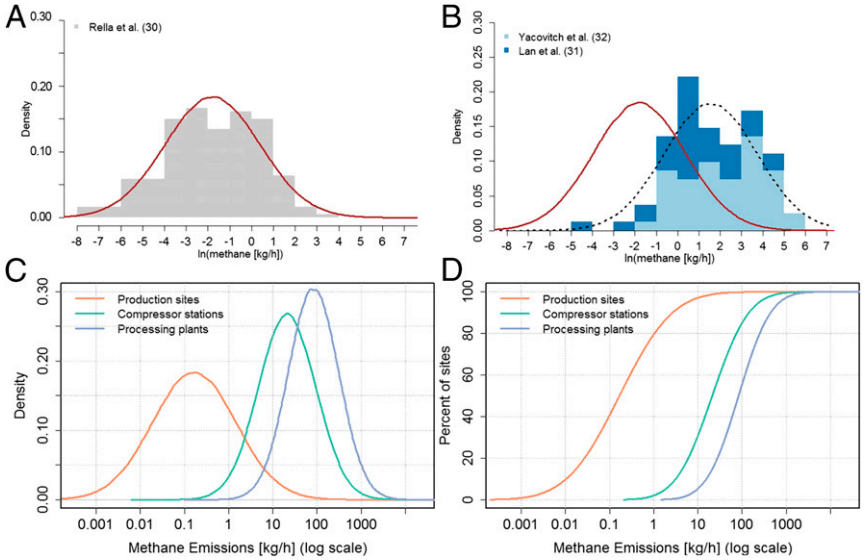

Fig. 3. Summary of results from the statistical estimator. (*A*) Systematic sample for production sites from Rella et al. (30). (*B*) Production sites from the high-emitter–biased samples of production sites [Lan et al. (31) and Yacovitch et al. (32)]. The red and dotted normal distributions are the fitted pdfs. Notice the mean shift that preserves the variance. (*C*) Fitted pdf under the statistical estimator, for each source type. (*D*) Fitted cumulative distribution function under the statistical estimator, for each source type.

emission (see the *SI Appendix* for detailed discussion about selection of lognormal distribution for the pdfs):

$$q(x) = \frac{A(x)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\int_0^\infty A(x)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(v-\mu)^2}{2\sigma^2}}dv}. \quad [1]$$

The sample of production sites obtained by Lan et al. (31) and Yacovitch et al. (32) looks remarkably like the systematic sample from Rella et al. (30), except that it is shifted 3–4 powers of *e* to the right by the high-emitter bias (Fig. 3*B*). Note that the Yacovitch et al. and Lan et al. sample appears approximately normal and that the sample's variance is close to that of the systematic sample, despite the very large shift in the mean (Fig. 3*B*).

What functional form for $A(x)$ would preserve normality, conserve the variance, and shift the mean? The answer is a power function in the arithmetic space: $A(E) = cE^\theta$, which is an exponential function of x: $A(x) = ce^{x\theta}$. With exponential $A(x)$:

$$q(x) = \frac{ce^{x\theta}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\int_0^\infty ce^{v\theta}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(v-\mu)^2}{2\sigma^2}}dv}; \quad [2]$$

and by completing the square, one can show that this high-emitter–biased distribution is simply the normal distribution with mean $\mu + \theta\sigma^2$:

$$q(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu-\theta\sigma^2)^2}{2\sigma^2}}. \quad [3]$$

To recap, if the swept area function $A(E)$ is cE^θ, then the high-emitter–biased samples should be lognormal if the unbiased distribution is lognormal. The mean of the logs should increase by $\theta\sigma^2$, and the variance should stay the same, just as it appears to in Fig. 3*B*.

We now argue that a power law is a physically natural simple choice for the bias function $A(E)$. In Gaussian dispersion theory, the concentration ϕ at distance D downwind of a source with emissions rate E along the plume's centerline is equal to

$$\phi = \frac{E}{W\sigma_y(D)\sigma_z(D)}, \quad [4]$$

where B is a constant, W is wind speed, $\sigma_y(D)$ is the SD in the horizontal direction orthogonal to the wind, and the $\sigma_z(D)$ is the SD in the vertical (43). Log–log plots of $\sigma_y(D)$ and $\sigma_z(D)$ are usually close to linear (35), and thus the product $\sigma_y(D)\sigma_z(D)$ is close to a power function: $\sigma_y(D)\sigma_z(D) \approx aD^b$. If ϕ_0 is the concentration detection limit and $D^*(E)$ is the maximum distance for detection of a source of strength E, then

$$\phi_0 = \frac{E}{WaD^*(E)^b}, \quad [5]$$

which rearranges to $D^*(E) =$ a constant times $E^{1/b}$, which implies that $A(E)$ is also a power function.

This heuristic argument should be viewed as secondary to the observation that only a power function for the bias $A(E)$ will transform a normal distribution of the logs into another normal distribution with the same variance but a different mean. Thus, the data provide a strong empirical argument for the functional form of $A(E)$, which is also reasonable given the physics.

With $A(x) = ce^{x\theta}$ and the normal distribution $q(x)$ above, the log likelihood function for the systematic and high-emitter–biased samples is

$$\ell_2(\underline{\mu}, \underline{\sigma}, \theta) = \sum_{j=1}^{3}\left\{S_{0j}ln\Phi\left(\frac{x_j^* - \mu_j}{\sigma_j}\right) - (S_{rj} + S_{bj})ln\sigma_j - \sum_{i=1}^{S_{rj}}\frac{(x_i - \mu_j)^2}{2\sigma_j^2} + \sum_{k=1}^{S_{bj}}\frac{\left(x_k - \mu_j - \theta\sigma_j^2\right)^2}{2\sigma_j^2}\right\}, \quad [6]$$

where $\underline{\mu}$ and $\underline{\sigma}$ are the vectors $\underline{\mu} = (\mu_1, \mu_2, \mu_3)$ and $\underline{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, S_{bj} is the sample size for facility type j in the high-emitter–biased dataset, and k indexes the sites in the high-emitter–biased datasets. Note that although this log-likelihood function has separate means and SDs for each type of facility (indexed by j), it has only a single value of the bias exponent θ. A single θ value is used because the bias is a function of emissions strength, not the type of facility, whereas the means and SDs of the emissions pdfs do vary with the type of facility. We estimated the three means, three SDs, and single bias parameter by maximizing the likelihood function and calculated confidence limits by inverting the information matrix. Because Lan et al. (31) and Yacovitch et al. (32) used marginally different Gaussian plume models in their estimation procedures, we also considered a variant of the above log-likelihood function with two separate values of θ, one for each of the two studies. However, the hypothesis of a single value of θ could not be rejected by a likelihood ratio test (the log-likelihood was improved by only ~1.0 because of the addition of a second θ, and the 0.05 significance threshold is ~1.9).

The normal densities in Fig. 3 *A* and *B* are maximum likelihood estimates. Notice how well they fit the data for both the systematic and high-emitter–biased samples. It is also reassuring that the estimated value of θ (~0.7) is within a narrow range of values that we expect, given the exponents in the Gaussian plume literature (44) under the meteorological conditions that dominated the periods of sampling (relatively stable categories). On the

other hand, the Gaussian plumes literature also shows that *A(E)* is not constant. *A(E)* changes with stability class and, for non-neutral stability classes, with distance downwind [because plots of $\log(\sigma_y(D)\sigma_z(D))$ vs. $\log D$ are mildly concave down rather than perfectly linear]. These concerns motivated the variation of integrating the systematic and high-emitter–biased samples, which produces similar results (*SI Appendix*).

Fig. 3 *C* and *D* show the pdf and cumulative distribution function for each of the modeled sources. Similarly, Table 1 summarizes the parameters and EFs derived from the statistical estimator and used to produce the BU estimate.

ACKNOWLEDGMENTS. We thank the project's Scientific Advisory Panel: D. Blewitt, S. Hanna, D. Jacob and F. O'Sullivan. Funding for this work was provided by the Alfred P. Sloan Foundation, Fiona and Stan Druckenmiller, Heising-Simons Foundation, Bill and Susan Oberndorf, Betsy and Sam Reeves, the Robertson Foundation, TomKat Charitable Trust, and the Walton Family Foundation.

1. Alvarez RA, Pacala SW, Winebrake JJ, Chameides WL, Hamburg SP (2012) Greater focus needed on methane leakage from natural gas infrastructure. *Proc Natl Acad Sci USA* 109(17):6435–6440.
2. Karion A, et al. (2013) Methane emissions estimate from airborne measurements over a western United States natural gas field. *Geophys Res Lett* 40(16):4393–4397.
3. Pétron G, et al. (2014) A new look at methane and non-methane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. *J Geophys Res Atmos* 119(11):6836–6852.
4. Peischl J, et al. (2015) Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions. *J Geophys Res Atmos* 120(5):2119–2139.
5. Jeong S, et al. (2013) A multitower measurement network estimate of California's methane emissions. *J Geophys Res Atmos* 118(19):11339–11351.
6. Miller SM, et al. (2013) Anthropogenic emissions of methane in the United States. *Proc Natl Acad Sci USA* 110(50):20018–20022.
7. Schwietzke S, Griffin WM, Matthews HS, Bruhwiler LMP (2014) Natural gas fugitive emissions rates constrained by global atmospheric methane and ethane. *Environ Sci Technol* 48(14):7714–7722.
8. Caulton DR, et al. (2014) Toward a better understanding and quantification of methane emissions from shale gas development. *Proc Natl Acad Sci USA* 111(17): 6237–6242.
9. Allen DT, et al. (2013) Measurements of methane emissions at natural gas production sites in the United States. *Proc Natl Acad Sci USA* 110(44):17768–17773.
10. Allen DT, et al. (2015) Methane emissions from process equipment at natural gas production sites in the United States: Pneumatic controllers. *Environ Sci Technol* 49(1): 633–640.
11. Allen DT, et al. (2015) Methane emissions from process equipment at natural gas production sites in the United States: Liquid unloadings. *Environ Sci Technol* 49(1): 641–648.
12. Mitchell AL, et al. (2015) Measurements of methane emissions from natural gas gathering facilities and processing plants: Measurement results. *Environ Sci Technol* 49(5):3219–3227.
13. Subramanian R, et al. (2015) Methane emissions from natural gas compressor stations in the transmission and storage sector: Measurements and comparisons with the EPA greenhouse gas reporting program protocol. *Environ Sci Technol* 49(5):3252–3261.
14. Lamb BK, et al. (2015) Direct measurements show decreasing methane emissions from natural gas local distribution systems in the United States. *Environ Sci Technol* 49(8): 5161–5169.
15. Turner AJ, et al. (2015) Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data. *Atmos Chem Phys Discuss* 15(4):4495–4536.
16. Townsend-Small A, Tyler SC, Pataki DE, Xu X, Christensen LE (2012) Isotopic measurements of atmospheric methane in Los Angeles, California, USA: Influence of "fugitive" fossil fuel emissions. *J Geophys Res* 117(D7):D07308.
17. Wennberg PO, et al. (2012) On the sources of methane to the Los Angeles atmosphere. *Environ Sci Technol* 46(17):9282–9289.
18. Peischl J, et al. (2013) Quantifying sources of methane using light alkanes in the Los Angeles basin, California. *J Geophys Res Atmos* 118(10):4974–4990.
19. Karion A, et al. (2015) Aircraft-based estimate of total methane emissions from the Barnett Shale region. *Environ Sci Technol* 49(13):8124–8131.
20. Petron G, et al. (2012) Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study. *J Geophys Res Atmos* 117(D4):D04304.
21. Kort EA, et al. (2014) Four Corners: The largest US methane anomaly viewed from space. *Geophys Res Lett* 41(19):6898–6903.
22. McKain K, et al. (2015) Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts. *Proc Natl Acad Sci USA* 112(7):1941–1946.
23. Katzenstein AS, Doezema LA, Simpson IJ, Blake DR, Rowland FS (2003) Extensive regional atmospheric hydrocarbon pollution in the southwestern United States. *Proc Natl Acad Sci USA* 100(21):11975–11979.
24. Wunch D, et al. (2009) Emissions of greenhouse gases from a North American megacity. *Geophys Res Lett* 36(15):L15810.
25. Hsu Y, et al. (2010) Methane emissions inventory verification in southern California. *Atmos Environ* 44(1):1–7.
26. Brandt AR, et al. (2014) Methane leaks from North American natural gas systems. *Science* 343(6172):733–735.
27. Lyon DR, et al. (2015) Constructing a spatially resolved methane emission inventory for the Barnett Shale region. *Environ Sci Technol* 49(13):8147–8157.
28. Harriss R, et al. (2015) Using multi-scale measurements to improve methane emission estimates from oil and gas operations in the Barnett Shale region, Texas. *Environ Sci Technol* 49(13):7524–7526.
29. Smith ML, et al. (2015) Airborne ethane observations in the Barnett shale: Quantification of ethane flux and attribution of methane emissions. *Environ Sci Technol* 49(13):8158–8166.
30. Rella CW, Tsai TR, Botkin CG, Crosson ER, Steele D (2015) Measuring emissions from oil and natural gas well pads using the mobile flux plane technique. *Environ Sci Technol* 49(7):4742–4748.
31. Lan X, Talbot R, Laine P, Torres A (2015) Characterizing fugitive methane emissions in the Barnett Shale area using a mobile laboratory. *Environ Sci Technol* 49(13): 8139–8146.
32. Yacovitch TI, et al. (2015) Mobile laboratory observations of methane emissions in the Barnett Shale region. *Environ Sci Technol* 49(13):7889–7895.
33. Xu L, Lin X, Amen J, Welding K, McDermitt D (2014) Impact of changes in barometric pressure on landfill methane emission. *Global Biogeochem Cycles* 28(7):679–695.
34. Czepiel PM, et al. (2003) The influence of atmospheric pressure on landfill methane emissions. *Waste Manag* 23(7):593–598.
35. Lavoie TN, et al. (2015) Aircraft-based measurements of point source methane emissions in the Barnett Shale basin. *Environ Sci Technol* 49(13):7904–7913.
36. US Environmental Protection Agency (2014) Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2012 (US Environmental Protection Agency, Washington, DC). Available at www2.epa.gov/climatechange/ghgemissions/usinventoryreport.html.
37. US Environmental Protection Agency (2014) Greenhouse Gas Reporting Program, (US Environmental Protection Agency, Washington, DC). Available at ghgdata.epa.gov/ghgp/main.do.
38. European Commission Joint Research Centre/Netherlands Environmental Assessment Agency (2013) Emission Database for Global Atmospheric Research, Release Version 4.2 (JRC Joint Research Centre, Ispra, Italy). Available at edgar.jrc.ec.europa.eu.
39. Camuzeaux JR, Alvarez RA, Brooks SA, Browne JB, Sterner T (2015) Influence of methane emissions and vehicle efficiency on the climate implications of heavy-duty natural gas trucks. *Environ Sci Technol* 49(11):6402–6410.
40. Tong F, Jaramillo P, Azevedo IML (2015) Comparison of life cycle greenhouse gases from natural gas pathways for medium and heavy-duty vehicles. *Environ Sci Technol* 49(12):7123–7133.
41. Zimmerle DJ, et al. (2015) Methane emissions from the natural gas transmission and storage system in the United States. *Environ Sci Technol* 49(15):9374–9383.
42. Zavala-Araiza D, et al. (2015) Toward a functional definition of methane superemitters: Application to natural gas production sites. *Environ Sci Technol* 49(13): 8167–8174.
43. Seinfeld JH, Pandis SN (2012) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (Wiley, Hoboken, NJ), pp 749–752, pp 862–866.
44. Gifford F (1959) Statistical properties of a fluctuating plume dispersion model. *Adv Geophys* 6:117–137.