This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2019, 53, 4747−4754

pubs.acs.org/est

Importance of Superemitter Natural Gas Well Pads in the Marcellus Shale

Dana R. Caulton,[†,⊥] Jessica M. Lu,[†,#] Haley M. Lane,[†] Bernhard Buchholz,[‡,▽] Jeffrey P. Fitts,[†] Levi M. Golston,[†] Xuehui Guo,[†] Qi Li,[§,□] James McSpiritt,[†] Da Pan,[†] Lars Wendt,[∥] Elie Bou-Zeid,[†] and Mark A. Zondlo*[,†]

[†]Department of Civil and Environmental Engineering, Princeton University, 59 Olden St., Princeton, New Jersey 08540, United States

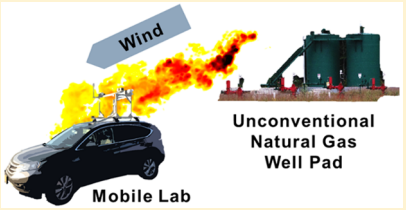
[‡]German National Metrology Institut, PTB-Braunschweig, Braunschweig 38116, Germany

[§]Department of Earth and Environmental Engineering, Columbia University, 500 W 120th St., New York, New York 10027, United States

[∥]Hunterdon Central Regional High School, Flemington, New Jersey 08822, United States

Supporting Information

ABSTRACT: A large-scale study of methane emissions from well pads was conducted in the Marcellus shale (Pennsylvania), the largest producing natural gas shale play in the United States, to better identify the prevalence and characteristics of superemitters. Roughly 2100 measurements were taken from 673 unique unconventional well pads corresponding to ∼18% of the total population of active sites and ∼32% of the total statewide unconventional natural gas production. A log-normal distribution with a geometric mean of 2.0 kg h^{-1} and arithmetic mean of 5.5 kg h^{-1} was observed, which agrees with other independent observations in this region. The geometric standard deviation (4.4 kg h^{-1}) compared well to other studies in the region, but the top 10% of emitters observed in this study contributed 77% of the total emissions, indicating an extremely skewed distribution. The integrated proportional loss of this representative sample was equal to 0.53% with a 95% confidence interval of 0.45−0.64% of the total production of the sites, which is greater than the U.S. Environmental Protection Agency inventory estimate (0.29%), but in the lower range of other mobile observations (0.09−3.3%). These results emphasize the need for a sufficiently large sample size when characterizing emissions distributions that contain superemitters.



INTRODUCTION

Many recent campaigns have undertaken the goal of sampling methane (CH_4) emissions from natural gas production facilities to understand the impact of skewed distributions and thus "superemitters" on total emissions. Brandt et al.[1] reported the disparity between bottom-up and top-down methods of CH_4 emissions where top-down methods consistently indicated higher emissions than bottom-up studies, which were largely in line with inventory estimates. More recently, Alvarez et al.[2] estimated the total emissions from the United States oil and natural gas supply to be ∼60% higher than the U.S. Environmental Protection Agency (EPA) inventory. Both works suggested that the difference could be due to the presence of "superemitters": infrequent sites with high emissions that contribute an outsize proportion to the total emissions relative to their frequency. This superemitter phenomenon was observed and reported in several recent papers,[3,4] including one study in the Barnett Shale (Texas) that managed to reconcile bottom-up and top-down emissions with extensive ground-based sampling.[5,6] However, most studies have been based on relatively small sample sizes (∼0.1−1% of total sites) and on basins with a small contribution to total U.S. production. This has created uncertainty in reported distributions because small sample sizes are insufficient to capture and characterize potentially very rare (e.g., events with frequency <1/100), but important events. For example, Brandt et al.[7] compared results from several studies that reported log-normal distributions and also simulated distributions with different assumptions. The true distributions followed even more extreme distributions than log-normal, suggesting large and very infrequent emissions where the top 5% of sites contribute ∼50% of emissions.[7] For the purposes of this work, we define "superemitters" as the top 10% of sites and also compare to the top 5% threshold.

The Marcellus shale play is the largest producing shale play in the United States and accounts for ∼38% of U.S. dry shale gas production. It outputs about three times as much as the next most productive shale play (Permian, Texas/New Mexico) and almost 1.5 times as much as the Barnett (Texas), Fayetteville (Arkansas), Eagle Ford (Texas), and Haynesville (Louisiana/Texas) shale plays combined (EIA,

Received: December 12, 2018
Revised: February 28, 2019
Accepted: March 11, 2019
Published: March 11, 2019

Downloaded via NOAA CENTRAL LIBRARY on January 30, 2024 at 15:57:22 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2019 American Chemical Society

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

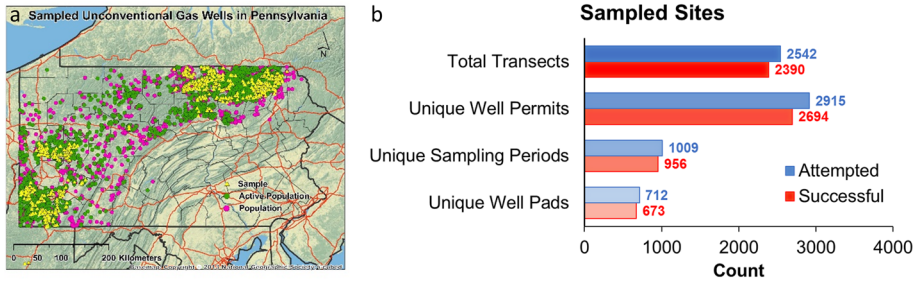


Figure 1. (a) Topographic map of Pennsylvania with major highways in orange, the population of active unconventional gas wells in green, nonactive unconventional gas wells in pink, and sampled sites in yellow. (b) Counts for each sampling type showing sites attempted and sites that successfully passed the criteria of being within 300 m of a public road, <50 m elevation difference, >1 m s^{-1} wind speed, and with an emission rate above the detection limit as described in Sample Quality Control.

www.eia.gov). The Marcellus represents a unique challenge relative to other locations due to its mountainous and forested terrain and large areal extent. Traditional techniques that employ Gaussian dispersion models to derive emission estimates are assumed to be invalid under complex topography as the assumptions necessary to derive the Gaussian equation are not met (e.g., stationary wind field, constant dispersion coefficients relationships).[8] For this reason, the Marcellus[9−12] has been understudied relative to other shale plays such as the Barnett, Pinedale, Uintah, and Denver-Julesburg basins.[3,4,13,14] The largest study in the Marcellus from Allen et al. used measurements from 47 unique sites; however, these measurements were collected at the component level and did not measure tank "flashing" emissions, which may not lead to representative emissions from a site as a whole.[15] A more recent campaign in the Marcellus from Omara et al. collected measurements from 45 sites.[12]

In order to obtain a large sample size in this relatively understudied region, we used a mobile lab to sample downwind of well pads in the Marcellus and developed a more robust sampling protocol that has been previously reported.[16] At the time of this study in Pennsylvania, there were ∼18,000 permitted unconventional gas wells with ∼10,000 of these categorized as active. Single well pads can have multiple well permits with different statuses. Wells can also be considered regulatory inactive, which corresponds to a well that can still produce natural gas but has been sealed in a temporary way such that the seal can be removed and the well returned to an active state in the future. Additionally, wells can be plugged (permanently rendered inactive) or abandoned/orphaned (not maintained, but not necessarily plugged). The Pennsylvania Department of Environmental Protection (PADEP) maintains a separate inventory of orphaned, abandoned, and plugged wells that were not targeted for sampling in this work. The breakdown of unconventional gas permits in Pennsylvania is shown in Figure S1. There is a total of ∼3000 unique unconventional well pads with at least one active well.

METHODS

Site Selection. A pseudorandom sampling method was used to sample efficiently as the large spatial extent of the Marcellus and the inaccessibility of many sites would make a random sample challenging. Unconventional natural gas sites were first screened by the minimum distance from the road to remove sites farther than 300 m from a road. Initial tests had shown that it is difficult to visually confirm a well pad beyond this distance and additional isotopic or hydrocarbon measurements were not available to confirm unseen emission sources. Additionally, high emitters can be detected at farther distances leading to potential oversampling of these sites. Sites were then grouped into nominal wind directions (N, E, S, W) and sites with elevation differences >50 m or with large physical obstructions between the site and the road were excluded. The final sites were grouped into routes (15−25 well pads per route) based on proximity and wind direction. Further details of the site selection process can be found in the Supporting Information (SI) and Figure S2.

Campaigns. In total, field campaigns were deployed in Pennsylvania to sample unconventional wells from 7/6/2015 to 7/21/2015, 10/31/2015 to 11/7/2015, 6/20/2016 to 7/2/2016, and 7/11/2016 to 7/15/2016. These field campaigns spanned 36 days, ∼200 h and covered ∼16,000 km. Typically, two outings were scheduled each day to maximize data collection under conditions with near-neutral static stability (∼0600−1100 and ∼1700−2200 local time). During each outing, most sites were sampled with two downwind transects. Two sites were selected near the beginning and end of the outing and sampled with 10 repeat downwind transects. Our previous work has shown that at least 10 transects are needed to provide a robust estimate of the influence of atmospheric variability, the greatest contributor to total uncertainty.[16] A unique "sampling period" is defined as any amount of repeat transects at a site within a 1 h interval (e.g., if a site was visited with 5 transects at 0800 local time and 20 transects at 2000 local time, these would count as two unique sampling periods). In total, 712 unique well pads were sampled. Of these, 137 well pads were resampled multiple times to investigate temporal variability as has been observed by other recent works,[10,17] to produce a total of 1009 unique sampling periods with a combined 2542 downwind transects. After measurements, some well pads were rejected as they could not be visually confirmed, had not been completed (per state database), had inaccessible roads, or were measured during low wind speeds (<1 m s^{-1}). These 712 unique unconventional well pads correspond to 2915 unique wells (1731 unique active wells after quality screening) representing ∼18% of the total active unconventional well population. A map of the sampled well pads is shown in Figure 1a with sampling counts in Figure 1b.

Mobile Lab and Data Processing. A compact sport utility vehicle was equipped with chemical and meteorological sensors on a platform raised ∼1 m above the car to reduce flow effects from the car.[18] Open-path sensors (CH_4 by LI-COR 7700 and H_2O/CO_2 by LI-COR 7500A) measured at a total height of 2.44 m above the ground and were calibrated prior to each campaign with either a NOAA standard (1.8724 ± 0.0030

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

ppmv CH_4 and 394.51 ± 0.07 ppmv CO_2) or, for higher amounts, with a commercial gas standard (2.12 ± 0.01 ppmv CH_4 in air, Air Liquide). The LI-COR 7700 and LI-COR 7500A have expected RMS noise levels of 1.6 ppbv and 0.04 ppmv at 1 Hz, respectively.

Data were processed according to the Gaussian plume method described in previous work.[4,16] Briefly, a Gaussian plume model was used to simulate concentrations along the measurement path assuming a constant reference emission rate (1 kg s^{-1} for simplicity), and modeled wind speed and stability from the NOAA ready archive meteorology (EDAS 40 km, 3 hly output available at www.ready.noaa.gov) at the time and location of measurement. The ratio between observed and modeled concentration is thus the inferred emission rate in kg s^{-1}. Vertical and horizontal dispersion coefficients were calculated according to Briggs[19] for rural sites. As noted previously, there is difficulty in applying this technique in complex terrain. To overcome this, we removed highly complex sites based on the sensor-well pad elevation differences (>50 m) and minimum road distances (>300 m) as criteria noted earlier. A detailed discussion of uncertainty analysis for this method is available in Caulton et al.[16] This work details the systematic analyses of the major uncertainties in this approach including wind speed, stability, source location, and diffusion assumptions. The Gaussian method described here was also compared to more complex emission estimate calculations including an estimate resulting from a high-resolution large eddy simulation that does not rely on diffusion parameters. In addition, three controlled releases, performed as an independent assessment of the Gaussian method by comparing the actual release rate and Gaussian estimate with on-site winds, showed an average difference <20%. Uncertainty was determined to be mainly driven by atmospheric variability, which is described more generally as transect to transect variability, but other sources, such as the NOAA winds typically used, also contribute to the overall uncertainty. For sites with ∼10 repeat transects, an average relative standard deviation of 77% (ranging from 12 to 260%) was observed for repeat emissions estimates. Using a Monte Carlo approach the sources of uncertainty including an expected 100% standard deviation (75% of repeat observations were equal or less than this value) due to random atmospheric variability, lead to an asymmetric 95% confidence interval of $0.05x$–$6.5x$, where x is the emission rate for sites with less than three transects. Uncertainty at sites with multiple transects of the plume is estimated to be $0.5x - 2.7x$ at the 95% confidence interval. Sample Quality Control presents evidence that the elevation and distance quality screening is justified.

In addition to the emission rate, a limit of detection (LOD) was calculated at every site by calculating the emission rate corresponding to a theoretical observed peak of 50 ppbv. The 50 ppbv threshold was determined by roughly three times the average observed standard deviation of the background (∼15 ppbv) in a ∼10 min "clean" area. The limit of detection combines all factors that may reduce the detection ability of the method, including high wind speeds, unstable conditions, long distances and disparity between source emission altitude and measurement height. Sites were determined to have no emissions if retrieved emission rates were less than the LOD. The median calculated LOD was 0.12 kg h^{-1}.

Additionally, the data set used here including the emission rates, date/time, LOD, uncertainty estimate, meteorology, site locations, and traits including spud date, operator, production, and status is available from the DataSpace archive at Princeton University (http://arks.princeton.edu/ark:/88435/dsp01wh246v90d).[34] This archive is free and open to the public.

RESULTS AND DISCUSSION

Sample Representativeness. The sampled well pads were compared to the overall population of unconventional gas wells in Pennsylvania to assess differences in spud date (well age), total gas production volume, operator, and region as obtaining a representative sample is necessary to produce accurate results. Lifetime gas volume was used to calculate average daily production (cumulative from 2003 to 2016 divided by the number of days of production) and is used in this section because some sites were sampled multiple times. While emissions measurements were made at the well pad (i.e., site) level, statewide data are reported at the individual well level. Individual wells of the sites measured were extracted so that the data could be directly compared on the same level. Similar distributions of spud date were obtained for the sampled sites and population shown in Figure 2a with the sampled distribution being four months older than the overall population. No significant differences were found for the total production distribution as shown in Figure 2b. Comparing operator distribution is more challenging as there are 74 unique operators with active well permits in this region, some with only a few well pads that are geographically clustered. Instead, we separated the operators by company size as defined by the number of wells they operated. In the overall population, 54% of active wells are operated by "large" companies, here defined as those with >500 wells. These companies were slightly overrepresented in the study (65% of the sample population) as they had the most well pads and were geographically diverse. The geographic distribution of the overall population of wells in the northeast, southwest and elsewhere is 51%, 46%, and 3%, respectively. In our sample, these regions, respectively, represented 58%, 42%, and <1% of the sites. Overall, the sampled wells compare appropriately to the population, with no clear bias toward a category that would affect the results.

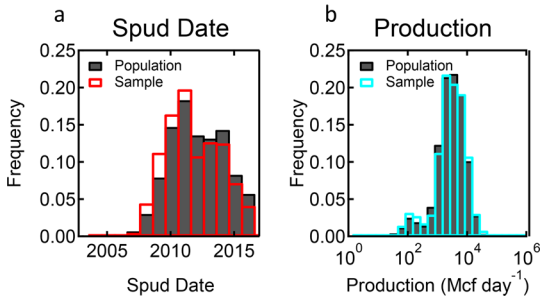


Figure 2. (a) Distribution of the spud dates of sampled wells and the population of active wells in Pennsylvania. (b) Distribution of the average daily production volume (cumulative from 2003 to 2016 divided by the number of days of production) of sampled wells and the population of active wells in Pennsylvania.

Sample Quality Control. The complex topography of the Marcellus makes applying Gaussian dispersion approaches challenging as the assumptions necessary to derive the Gaussian plume equation are not valid. A more detailed description of this analysis is provided in the SI and Figures

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

S3−S6. To summarize, the analysis yielded results that suggest our treatment of the data was not affected by topography. In addition, several screening criterion were examined and the final selection criterion removed sites with LOD > 3.6 kg h^{-1}. This criterion was selected as it combined the effects of all factors that affect data quality while simultaneously retaining the most data, screening out only ∼5% of the calculated emissions as shown in Figure 1. A more detailed discussion of the selection of this threshold is available in the SI and Figure S7.

Sample Averaging Schemes. As there is substantial uncertainty in any single observation, it is advantageous to average to produce a more robust data set. In Figure 3 the distributions for all emission rates are shown along with observations averaged to unique sampling periods (aggregated to those with measurements <1 h apart) and observations averaged to unique well pad. These scenarios reduce sample size to 1009 sampling periods and 712 well pads, which are further reduced to 956 and 673 after quality screening. While the distribution geometric means and standard deviations are not significantly different, the contribution of the top 10% of emissions does decrease for emissions that have been averaged. The unique sampling periods correspond to additional experiments that resampled sites at different time scales to observe changes in emissions over time. As real variability was observed, the sample period mean (with 956 observations) is used as the best statistic for reducing the effect of random error while preserving the potential influence of actual emissions variability. The sample geometric mean emission rate is 2.0 ± 0.2 kg h^{-1} with a geometric standard deviation of 4.4 kg h^{-1} and an arithmetic mean of 5.5 kg h^{-1} (±0.9 kg h^{-1} standard error). The contribution of the top 10% of emissions, (emissions > ∼9 kg h^{-1}) account for 77% of the total emissions. This number includes nonemitting sites: 31% of sites sampled had no observed emissions. The geometric mean is the best estimate of the expected emission at a single site or the most frequent emission expected and would be expected to be consistent across studies with a small number of samples. The arithmetic mean is suitable as an emission factor for calculating expected emissions from a large sample; the arithmetic mean and contribution of the top 10% would not be expected to be consistent across studies with different sample sizes and the large sample analyzed here is expected to better constrain these estimates. However, with a log-normal population, a single emission rate estimate will not always work if the inventory area of interest is small and should be used with caution.

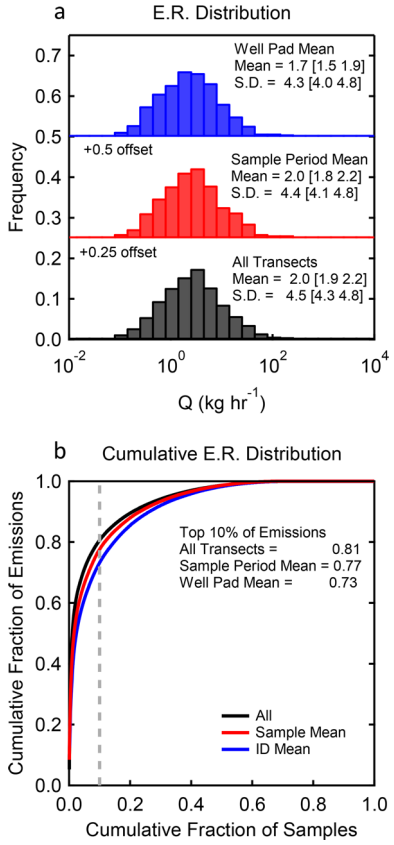


Figure 3. (a) Distribution of emissions for three scenarios: all transects, transects averaged to unique sample periods, and transects averaged to unique well pads, excluding sites with no emissions. (b) Cumulative emissions for the same three scenarios, including sites with no emissions.

Emission Comparison by Well Status Geographic Area, Operator Size and Production Class. The emissions for the well pad averaged database were sorted into classes for well status, geographic area, operator size, and production class (details of the methodology to categorize well pads with multiple wells into different classes are available in the SI). Well pad averages were used as most of the repeat sampling was done for active sites in northeastern Pennsylvania. As defined in Sample Representativeness, these classes focus on the larger differences between groups (northeast vs southwest) rather than individual locations. For instance, northeastern Pennsylvania is dominated by dry gas production, natural gas with relatively low heavier hydrocarbon ratios, and southwestern Pennsylvania transitions to wet gas production with higher hydrocarbon ratios. Operators were defined as "large" if they owned >500 wells and "small" if they had less wells than this threshold. Production volumes are reported by individual well, but were aggregated to the well pad level and correspond to the month a site was sampled. Sites were assigned a production class according to a base 10 log of their production value. Sites with no reported production were included in "<10 Mcf day^{-1}". Distributions for each class are shown in Figure 4 along with the frequency of sites with emissions > 9 kg h^{-1}. Due to the difference in the number of observations for different classes, a multicomparison using iterative one-way ANOVA testing was done using the log-transformed data to observe differences in means. At the 95% confidence interval, the mean emission rates are not statistically different by well status, region, operator size, or production class. Small differences apparent in classes are not significant using a two-sample Kolmogorov−Smirnov test in almost all cases, likely due to the small number of observations for some of these classes. However, the comparable emissions of plugged and regulatory inactive (which should also have a plug) indicate wells emit past their active lifespans and may continue to be a significant source of emissions, consistent with previous studies on plugged/abandoned well emissions.[20,21] While no difference is observed for the dry northeast and wet southwest, the occurrence of wet gas sites may not have been frequent enough to truly differentiate these regions. For instance, a recent study found that wet gas signals as denoted by high ethane ratios were only found in parts of Washington County, not the southwest region as a whole.[22] Operator distributions

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

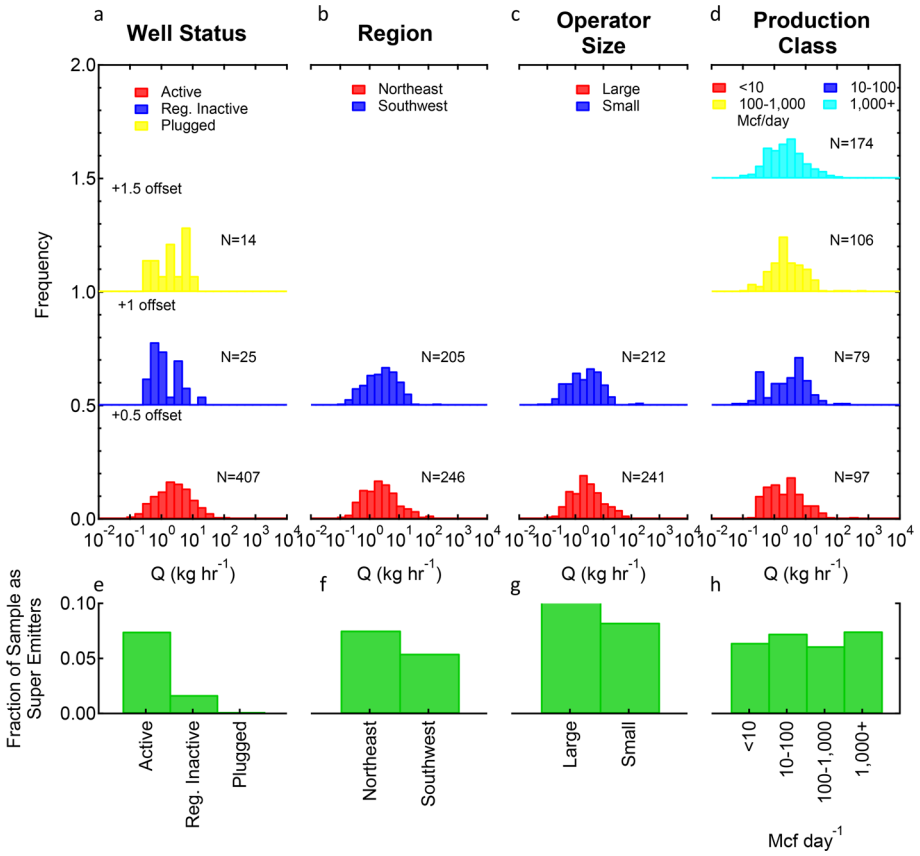


Figure 4. Comparison of emission rate distributions for different classes of (a) well status, (b) region, (c) operator type, and (d) production level. Fraction of the sample for those classes identified as superemitters for (e) well status, (f) region, (g) operator type, and (h) production level.

Table 1. Unconventional Natural Gas Emission Rates from Different Campaigns in the Appalachian Region

study	sample size	sample unit	mean (kg h^{-1})[a]	geo. μ (kg h^{-1})	geo. σ (kg h^{-1})	top 10% contribution	top 5% contribution
this work	956	sample period	5.5	2.0	4.4	77	66
this work	673	well pad	4.3	1.7	4.3	73	61
Omara et al.[12]	45	well pad	7.8	4.3	3.2	47	25
Omara et al.,[11] unconventional	13	well pad	18.8	5.7	4	50	<40%
Omara et al.,[11] conventional	18	well pad	0.82	0.55	4.2	36	20%
Allen et al.,[9] tracer only	5	well pad	1.2				
Goetz et al.[10]	3	well pad	8.7				

[a]Mean values here are author reported values. All other statistics are calculated from publicly available data sets and correspond to all reported emissions; not all studies reported 0 emissions.

showed no differences and most production classes were not different. The "<10 Mcf day^{-1}" class distribution was significantly different from the "10−100 Mcf day^{-1}" and "100−1000 Mcf day^{-1}" distributions; however, the differences appear to be small as all production classes have a similar frequency of superemitters.

Comparison to Other Studies. Statistics for the results presented here and other studies in the same region are shown in Table 1.[9−12] Note that only emissions computed from tracer releases were compared for Allen et al.[9] The geometric mean emission rate observed is within the range of previously observed estimates, but wide variation has been reported across the Marcellus and most studies have had very small sample sizes making this comparison ambiguous. For instance, the geometric and arithmetic mean emission rates reported in this work fall between Omara et al.'s[11] comparison between conventional and unconventional sites despite the fact that our sample is only composed of unconventional sites and both statistics are less than the results from Omara et al.'s[12] more recent and larger unconventional sample from the same region. Notably, the geometric mean reported in this work (2.0 kg h^{-1}) is similar to other mobile laboratory derived geometric means or medians across Texas, Colorado, Arkansas, Wyoming, and British Columbia with mostly much smaller samples sizes (0.63−3.7 kg h^{-1}).[3,13,14,23−26] The arithmetic mean reported here (5.5 kg h^{-1}) is, however, on the high end of estimates for other regions as directly reported or calculated from their available data (1.7−9.7 kg h^{-1}) where most observed mean emission rates < 3 kg h^{-1}.[3,13,14,23,24] As mentioned, the larger sample size in this work is expected to increase the accuracy of the arithmetic mean, so the higher value observed here indicates the necessity of large samples to calculate this metric from a log-normal distribution. The contribution of the top 10% of sites (corresponding to 77% of

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

emissions) and the contribution of the top 5% of sites (corresponding to 66%) are at the high range of estimate from many studies, suggesting the distribution observed here is actually skewed to be more extreme than log-normal.[7,12,25] This is confirmed by looking at the qq-plot that compares the observed quantiles of the data to the predicted quantiles of the distribution fit. A qq-plot for the sample mean emission rate data set is shown in Figure S8. The emission rate quantiles deviate from the expected 1:1 relationship at higher quantiles. The increased contribution of the higher emitters in this study highlights that there may be a few very significant and very infrequent emissions that drive the overall emissions inventory. Emissions less frequent than possible to capture in this study may have even higher contributions to the total emission rate. Additionally, the geometric mean loss rate observed (0.41%) is in the lower range of previous work (0.09−3.3%) for geometric mean or median), but the arithmetic mean loss rate (6.4%) falls in the middle of the reported arithmetic mean loss rates (0.57−12.6%).[3,13,14,23−26] The overall comparison to other work shows that mean emission rates may not vary much by basin, but large data sets are still needed to capture the full nature of the distribution as deviations in the "tail" could still lead to large differences in the number and impact of superemitters.

Functional Superemitters. As discussed in Zavala-Araiza et al., the term "superemitter" is not well-defined. In addition to the distribution of emission rates, loss rates have also been utilized to examine the occurrence of "functional superemitters".[27] Proportional loss rates are defined as the emission rate normalized by site natural gas production (available monthly through the PADEP), as shown in eq 1. Produced natural gas is assumed to be 80% methane.

$$\text{loss rate} = \frac{\text{emission rate}}{0.8(\text{natural gas production})} \quad (1)$$

The distinction designates sites as superemitters if they are outside the range of normal proportional loss rates, which can more easily be attributed to defective equipment. Comparatively, sites with properly functioning equipment but large production rates would give large signals, but would be more difficult to mitigate. The comparison between emission rates and proportional loss rates is shown in Figure 5. The proportional loss distribution shown here also corresponds to the sample period average. A comparison between proportional loss distributions using all transects, sample period averages and well pad averages is shown in Figure S9. Note that proportional loss rates can only be calculated for sites with reported production and so has a slightly different population than the emission rates. The geometric mean loss rate was 0.41% with a geometric standard deviation of 10% and an arithmetic mean of 6.4%. Also note that there are 11 observations with a > 100% proportional loss. All of these sites were determined to be active, but had low production rates of <150 Mcf day^{-1} compared to a sample median of ∼1000 Mcf day^{-1}. Variation in day to day production that is masked by the monthly summaries available or contribution of leaks from storage units may be responsible for these results. In addition, sites report the volume of gas sold, not necessarily the total volume produced making >100% rates possible, particularly if leaks or venting occurs before the gas metering point and the site has low production. The proportional loss distribution is significantly more skewed than the emission rates, with a larger geometric standard deviation and a higher

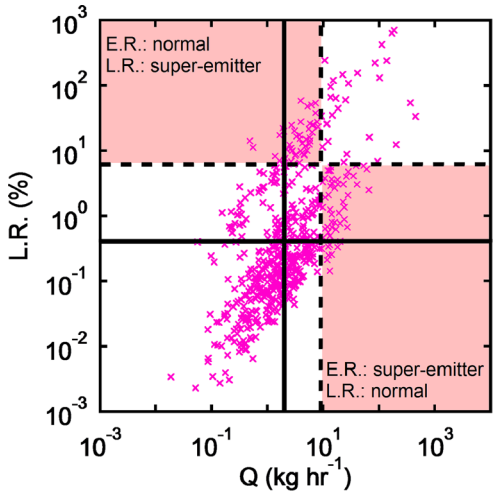


Figure 5. Correlation between emission rate and proportional loss. The red shading shows regions where emission rate and proportional loss definitions of superemitter differ. The black solid lines correspond to the geometric mean emission rate and proportional loss, while the black dashed lines correspond to the emission rate and proportional loss superemitter thresholds.

contribution from the top 10% of emissions (93% vs 77%). The proportional loss and emission rates also do not identify the same sites as superemitters, with emission rate superemitters being >9 kg h^{-1} and proportional loss superemitters >7%. The total contributions from the emission rate superemitters correspond to a combined emission rate of 8,715 kg h^{-1} while the proportional loss superemitters isolate a combined emission rate of 6,943 kg h^{-1}.

Total Emissions. The combined proportional loss of the representative sample of sites was 0.53%. This is calculated by summing the emissions and production for this sample as shown in eq 2:

$$\text{combined loss rate} = \frac{\sum \text{emission rates}}{0.8(\sum \text{natural gas production})} \quad (2)$$

Our combined proportional loss of 0.53% has a 95% confidence interval of 0.45−0.64%. The confidence interval was calculated by simulating 1000 distributions using our calculated distribution statistics (Figure 3) and summing the resulting emission. We do not have data on production variability, so production is assumed to be the same for each of these distributions. This is higher than the range reported by Allen et al.[9] 0.14−0.41% reflecting different assumptions of "normal production" and including high and low estimates for specific sources. This value is considerably higher than the most recent EPA estimates of onshore natural gas production leaks of 0.29%, corresponding to production and exploration which includes nonroutine events like well completions that we have no evidence were measured in this study.[28] Aircraft derived proportional loss rates in northeastern Pennsylvania for unconventional sites including gathering stations span 0.08−0.72%.[29] Recent syntheses of observational data also points to larger emission rates from production: 0.9% for the Appalachia region (or 0.6% for PA specifically) and 1.5−2.3% for the entire United States, though we note that the number of observational measurements in the current Marcellus study is comparable to[12] or larger[2] than the number of sites used for the entire U.S. national estimates. The combined unconven-

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

tional natural gas production across Pennsylvania was ~5.1 × 10^9 Mcf in 2016.[30] At the combined proportional loss measured in this study (0.53%), ~27 × 10^6 Mcf yr^{-1} of natural gas (~0.41 [0.35−0.49] $TgCH_4$ yr^{-1}) would be emitted from normal unconventional gas extraction in Pennsylvania alone. This is much higher that the PA DEP reported 0.10 $TgCH_4$ yr^{-1} for 2015, which includes drilling and completion/workover events that again we do not have evidence we measured.[31] This estimate is in line with recent work to upscale emissions specifically from unconventional emissions in all of Pennsylvania that resulted in an emission range of 0.25−0.56 $TgCH_4$ yr^{-1}.[11] This estimate is also in line with observed top down flux estimates of natural gas emissions for all production and transmission sectors in different portions of Pennsylvania (0.13−0.41 $TgCH_4$ yr^{-1}).[29,32,33]

IMPLICATIONS FOR EMISSION MONITORING

The results of the largest, most representative data set of emissions from well pads in the Marcellus region are presented here and show a geometric mean emission rate of 2.0 ± 0.2 kg h^{-1}. These emissions are within the range of previously reported emissions in many basins including the Marcellus. The emissions rates were highly skewed, with 77% of emissions coming from the top 10% of sites. The proportional loss rates were even more skewed with 93% of emissions coming from the top 10% of sites. The presence of a very skewed data set reiterates the importance of identifying superemitters for efficient emission reduction strategies. The integrated proportional loss of 0.53% for all sites is significantly higher than the median or geometric mean value and inventory estimates, further underscoring the need for quantification of these sources as results from small samples may fail to appropriately upscale results. Aggregate comparison of mean emission rates and distribution by well status, region, operator size, and production class showed almost no significant differences. As the significance of the infrequent, but consequential superemitters becomes clear, more detailed analysis of the characteristics and causes of emissions will be needed to design effective emissions monitoring strategies and regulations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.8b06965.

Descriptions of site selection, effect of topography, and LOD threshold (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: mzondlo@princeton.edu.

ORCID

Mark A. Zondlo: 0000-0003-2302-9554

Present Addresses

[⊥]D.R.C.: Department of Atmospheric Science, University of Wyoming, 1000 E. University Ave, Laramie, WY, 82071, United States.

[#]J.M.L.: Draper Richards Kaplan Foundation, 1600 El Camino Real, Suite 155, Menlo Park, CA 94025, United States.

[▽]B.B.: Department of Civil and Environmental Engineering, Princeton University, 59 Olden St., Princeton, New Jersey 08540, United States.Cornell University, School of Civil and Environmental Engineering, Ithaca, NY 14853, United States.

[□]Q.L.: Cornell University, School of Civil and Environmental Engineering, Ithaca, NY 14853, United States.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank all members of the fieldwork team including Stephany Paredes-Mesa, Tanvir Mangat, and Kira Olander. We thank LI-COR Biosciences for lending instrumentation used in this campaign. This work was funded by NOAA CPO/AC4, #NA14OAR4310134.

REFERENCES

(1) Brandt, A. R.; Heath, G. A.; Kort, E. A.; O'Sullivan, F.; Petron, G.; Jordaan, S. M.; Tans, P.; Wilcox, J.; Gopstein, A. M.; Arent, D.; Wofsy, S.; Brown, N. J.; Bradley, R.; Stucky, G. D.; Eardley, D.; Harriss, R. Methane Leaks from North American Natural Gas Systems. *Science* **2014**, *343* (6172), 733−735.

(2) 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**, *361* (6398), 186−188.

(3) Rella, C. W.; Tsai, T. R.; Botkin, C. G.; Crosson, E. R.; Steele, D. Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique. *Environ. Sci. Technol.* **2015**, *49* (7), 4742−4748.

(4) Yacovitch, T. I.; Herndon, S. C.; Pétron, G.; Kofler, J.; Lyon, D.; Zahniser, M. S.; Kolb, C. E. Mobile Laboratory Observations of Methane Emissions in the Barnett Shale Region. *Environ. Sci. Technol.* **2015**, *49* (13), 7889−7895.

(5) Zavala-Araiza, D.; Lyon, D. R.; Alvarez, R. A.; Davis, K. J.; Harriss, R.; Herndon, S. C.; Karion, A.; Kort, E. A.; Lamb, B. K.; Lan, X.; Marchese, A. J.; Pacala, S. W.; Robinson, A. L.; Shepson, P. B.; Sweeney, C.; Talbot, R.; Townsend-Small, A.; Yacovitch, T. I.; Zimmerle, D. J.; Hamburg, S. P. Reconciling Divergent Estimates of Oil and Gas Methane Emissions. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (51), 15597−155602.

(6) Lyon, D. R.; Zavala-Araiza, D.; Alvarez, R. A.; Harriss, R.; Palacios, V.; Lan, X.; Talbot, R.; Lavoie, T.; Shepson, P.; Yacovitch, T. I.; Herndon, S. C.; Marchese, A. J.; Zimmerle, D.; Robinson, A. L.; Hamburg, S. P. Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region. *Environ. Sci. Technol.* **2015**, *49* (13), 8147−8157.

(7) Brandt, A. R.; Heath, G. A.; Cooley, D. Methane Leaks from Natural Gas Systems Follow Extreme Distributions. *Environ. Sci. Technol.* **2016**, *50* (22), 12512−12520.

(8) Veigele, W. J.; Head, J. H. Derivation of the Gaussian Plume Model. *J. Air Pollut. Control Assoc.* **1978**, *28* (11), 1139−1140.

(9) Allen, D. T.; Torres, V. M.; Thomas, J.; Sullivan, D. W.; Harrison, M.; Hendler, A.; Herndon, S. C.; Kolb, C. E.; Fraser, M. P.; Hill, A. D.; Lamb, B. K.; Miskimins, J.; Sawyer, R. F.; Seinfeld, J. H. Measurements of Methane Emissions at Natural Gas Production Sites in the United States. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (44), 17768−17773.

(10) Goetz, J. D.; Floerchinger, C.; Fortner, E. C.; Wormhoudt, J.; Massoli, P.; Knighton, W. B.; Herndon, S. C.; Kolb, C. E.; Knipping, E.; Shaw, S.; DeCarlo, P. Atmospheric Emission Characterization of Marcellus Shale Natural Gas Development Sites. *Environ. Sci. Technol.* **2015**, *49* (11), 7012−7020.

(11) Omara, M.; Sullivan, M. R.; Li, X.; Subramanian, R.; Robinson, A. L.; Presto, A. A. Methane Emissions from Conventional and

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754

Unconventional Natural Gas Production Sites in the Marcellus Shale Basin. *Environ. Sci. Technol.* **2016**, *50* (4), 2099−2107.
(12) Omara, M.; Zimmerman, N.; Sullivan, M. R.; Li, X.; Ellis, A.; Cesa, R.; Subramanian, R.; Presto, A. A.; Robinson, A. L. Methane Emissions from Natural Gas Production Sites in the United States: Data Synthesis and National Estimate. *Environ. Sci. Technol.* **2018**, *52* (21), 12915−12925.
(13) Brantley, H. L.; Thoma, E. D.; Squier, W. C.; Guven, B. B.; Lyon, D. Assessment of Methane Emissions from Oil and Gas Production Pads Using Mobile Measurements. *Environ. Sci. Technol.* **2014**, *48* (24), 14508−14515.
(14) Lan, X.; Talbot, R.; Laine, P.; Torres, A. Characterizing Fugitive Methane Emissions in the Barnett Shale Area Using a Mobile Laboratory. *Environ. Sci. Technol.* **2015**, *49* (13), 8139−8146.
(15) Zavala-Araiza, D.; Alvarez, R. A.; Lyon, D. R.; Allen, D. T.; Marchese, A. J.; Zimmerle, D. J.; Hamburg, S. P. Super-Emitters in Natural Gas Infrastructure Are Caused by Abnormal Process Conditions. *Nat. Commun.* **2017**, *8*, 14012.
(16) Caulton, D. R.; Li, Q.; Bou-Zeid, E.; Fitts, J. P.; Golston, L. M.; Pan, D.; Lu, J.; Lane, H. M.; Buchholz, B.; Guo, X.; Mcspiritt, J.; Wendt, L.; Zondlo, M. A. Quantifying Uncertainties from Mobile-Laboratory-Derived Emissions of Well Pads Using Inverse Gaussian Methods. *Atmos. Chem. Phys.* **2018**, *18* (20), 15145−15168.
(17) Lavoie, T. N.; Shepson, P. B.; Cambaliza, M. O. L.; Stirm, B. H.; Conley, S.; Mehrotra, S.; Faloona, I. C.; Lyon, D. Spatiotemporal Variability of Methane Emissions at Oil and Natural Gas Operations in the Eagle Ford Basin. *Environ. Sci. Technol.* **2017**, *51* (14), 8001−8009.
(18) Tao, L.; Sun, K.; Miller, D. J.; Pan, D.; Golston, L. M.; Zondlo, M. A. Low-Power, Open-Path Mobile Sensing Platform for High-Resolution Measurements of Greenhouse Gases and Air Pollutants. *Appl. Phys. B: Lasers Opt.* **2015**, *119* (1), 153−164.
(19) Briggs, G. A. *Diffusion Estimation for Small Emissions*; National Oceanic and Atmospheric Administration: Oak Ridge, TN, 1973.
(20) Kang, M.; Kanno, C. M.; Reid, M. C.; Zhang, X.; Mauzerall, D. L.; Celia, M. a.; Chen, Y.; Onstott, T. C. Direct Measurements of Methane Emissions from Abandoned Oil and Gas Wells in Pennsylvania. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (51), 18173−18177.
(21) Kang, M.; Christian, S.; Celia, M. A.; Mauzerall, D. L.; Bill, M.; Miller, A. R.; Chen, Y.; Conrad, M. E.; Darrah, T. H.; Jackson, R. B. Identification and Characterization of High Methane-Emitting Abandoned Oil and Gas Wells. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (48), 13636−13641.
(22) Goetz, J. D.; Avery, A.; Werden, B.; Floerchinger, C.; Fortner, E. C.; Wormhoudt, J.; Massoli, P.; Herndon, S. C.; Kolb, C. E.; Knighton, W. B.; Peischl, J.; Warneke, C.; de Gouw, J. A.; Shaw, S. L.; DeCarlo, P. F. Analysis of Local-Scale Background Concentrations of Methane and Other Gas-Phase Species in the Marcellus Shale. *Elem. Sci. Anthr.* **2017**, *5*, 1−20.
(23) Robertson, A. M.; Edie, R.; Snare, D.; Soltis, J.; Field, R. A.; Burkhart, M. D.; Bell, C. S.; Zimmerle, D.; Murphy, S. M. Variation in Methane Emission Rates from Well Pads in Four Oil and Gas Basins with Contrasting Production Volumes and Compositions. *Environ. Sci. Technol.* **2017**, *51* (15), 8832−8840.
(24) Atherton, E.; Risk, D.; Fougere, C.; Lavoie, M.; Marshall, A.; Werring, J.; Williams, J. P.; Minions, C. Mobile Measurement of Methane Emissions from Natural Gas Developments in Northeastern British Columbia, Canada. *Atmos. Chem. Phys.* **2017**, *17* (20), 12405−12420.
(25) Zavala-Araiza, D.; Herndon, S. C.; Roscioli, J. R.; Yacovitch, T. I.; Johnson, M. R.; Tyner, D. R.; Omara, M.; Knighton, B. Methane Emissions from Oil and Gas Production Sites in Alberta Canada. *Elem. Sci. Anthr.* **2018**, *6* (27), 1−13.
(26) Yacovitch, T. I.; Daube, C.; Vaughn, T. L.; Bell, C. S.; Roscioli, J. R.; Knighton, W. B.; Nelson, D. D.; Zimmerle, D.; Pétron, G.; Herndon, S. C. Natural Gas Facility Methane Emissions: Measurements by Tracer Flux Ratio in Two US Natural Gas Producing Basins. *Elem Sci. Anth* **2017**, *5*, 69.
(27) Zavala-Araiza, D.; Lyon, D.; Alvarez, R. A.; Palacios, V.; Harriss, R.; Lan, X.; Talbot, R.; Hamburg, S. P. Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites. *Environ. Sci. Technol.* **2015**, *49* (13), 8167−8174.
(28) United States Environmental Protection Agency. *Inventory of U. S. Greenhouse Gas Emissions and Sinks: 1990−2016*; 2018.
(29) Barkley, Z. R.; Lauvaux, T.; Davis, K. J.; Deng, A.; Miles, N. L.; Richardson, S. J.; Cao, Y.; Sweeney, C.; Karion, A.; Smith, M.; Kort, E. A.; Schwietzke, S.; Murphy, T.; Cervone, G.; Martins, D.; Maasakkers, J. D. Quantifying Methane Emissions from Natural Gas Production in North-Eastern Pennsylvania. *Atmos. Chem. Phys.* **2017**, *17* (22), 13941−13966.
(30) Pennsylvania Department of Environmental Protection. *Oil and Gas Inventory*; http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/OG_Well_Inventory (accessed Dec. 1, 2017).
(31) Pennsylvania Department of Environmental Protection. *Unconventional Natural Gas Emission Inventory* http://www.dep.pa.gov/Business/Air/BAQ/BusinessTopics/Emission/Pages/Marcellus-Inventory.aspx (accessed Dec. 1, 2017).
(32) Caulton, D. R.; Shepson, P. B.; Santoro, R. L.; Sparks, J. P.; Howarth, R. W.; Ingraffea, A. R.; Cambaliza, M. O. L.; Sweeney, C.; Karion, A.; Davis, K. J.; Stirm, B. H.; Montzka, S. A; Miller, B. R. Toward a Better Understanding and Quantification of Methane Emissions from Shale Gas Development. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (17), 6237−6242.
(33) Peischl, J.; Ryerson, T. B.; Aikin, K. C.; de Gouw, J. A.; Gilman, J. B.; Holloway, J. S.; Lerner, B. M.; Nadkarni, R.; Neuman, J. A.; Nowak, J. B.; Trainer, M.; Warneke, C.; Parrish, D. D. Quantifying Atmospheric Methane Emissions from the Haynesville, Fayetteville, and Northeastern Marcellus Shale Gas Production Regions. *J. Geophys. Res. Atmos.* **2015**, *120* (5), 2119−2139.
(34) Caulton, D. R.; Zondlo, M. A. *Marcellus Shale Unconventional Natural Gas Well Pad Emissions*; Princeton DataSpace: Princeton, NJ, 2018.

DOI: 10.1021/acs.est.8b06965
Environ. Sci. Technol. 2019, 53, 4747−4754