Supplementary Information

Quantifying regional methane emissions in the New Mexico Permian Basin with a comprehensive aerial survey

Yuanlei Chen*,†,§, Evan D. Sherwin†,§, Elena S.F. Berman‡, Brian B. Jones‡, Matthew P. Gordon‡, Erin B. Wetherley‡, Eric A. Kort¶, and Adam R. Brandt†

† Energy Resources Engineering, Stanford University, Stanford, CA 94305, USA

‡ Kairos Aerospace, Mountain View, CA 94040, USA

¶ Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA

§ Denotes equal contribution

∗ Email: yuliac@stanford.edu

Contents

S1 Controlled releases

S1.1 Hyperspectral methane detection

Kairos Aerospace (henceforth "Kairos") produces LeakSurveyor, a methane imaging system typically mounted on a light aircraft flown at standard general aviation altitudes of 900 meters (3,000 feet) above ground. The LeakSurveyor system produces a geo-referenced, false-colored plume image based on the detected excess methane concentration between the airplane and the ground (Figure 1a). The system contains a hyperspectral infrared spectrometer to detect methane, an optical imaging system to create an optical surface map of the surveyed region, and a GPS and inertial measurement unit to record the instrument's precise position and orientation.[18]

The infrared spectrometer measures the absorption of wavelengths at which methane molecules absorb strongly and other molecules do not.[18] As indicated in Kairos' technical white paper, absorption of methane molecules can be distinctive in the spectral range of 1500 nm to 3750 nm.[18] The spectral resolution of LeakSurveyor's hyperspectral infrared spectrometer is "typically around 0.5 nm or better/finer," suggesting a spectral resolution in this range for the Kairos system.[19]

The emission quantification method is described in Kairos technical paper on quantification and summarized below.[20] Kairos first uses its proprietary algorithm to compute pixel-level estimates of methane enhancement column density between the airplane and the ground. Then, Kairos designates plume pixels to be within a spatially contiguous region with methane enhancement statistically distinguishable from the background concentration. Next, Kairos estimates the wind direction based on the orientation of the vector between the plume pixel of highest methane enhancement and the furthest designated plume pixel from this pixel with maximum methane concentration. Then, a core segment of the plume is defined to contain the maximum concentration pixel and roughly 50% of plume's full length along the estimated wind direction. Next, Kairos estimates total excess methane in the

plume using Equation S1.

$$M = \sum_i D_i \times A_i \tag{S1}$$

Where M (unit: kg) is the total excess methane in the plume, D_i (unit: kg/m^2) is the excess column methane density of plume pixel i, and A_i (unit: m^2) is the area of plume pixel i.

The total excess methane mass (M) is then used to compute wind-independent emission rate using Equation S2, assuming constant emission, constant wind speed, and slow methane diffusion compared to wind speed.

$$R_{WI} = \frac{M}{L} \times \frac{3600\ seconds}{1\ hour} \tag{S2}$$

Where R_{WI} (unit: kg of methane per hour per meter per second of wind, kgh/mps) is the wind-independent methane emission rate, and L (unit: m) is the length of the core segment of the plume.

Thus, Kairos calculates wind-independent methane rates in kgh/mps by integrating the measured methane concentration enhancement over a cross section of a detected plume in the downwind direction.[20] Measured methane rates can then be calculated by multiplying this value by surface wind speed (Equation S3).

$$R_{Methane} = R_{WI} \times W \tag{S3}$$

Where $R_{Methane}$ (unit: kg of methane per hour, kg/h) is the methane emission rate, and W (unit: meter per second of wind, m/s) is the surface wind speed. More details of the selection of W is available in Section S1.3.2.

As detailed in Sherwin, Chen et al., methane plume identification "involves both automated signal processing and human review."[22] Sometimes multiple plumes can be seen from a single image,[22] in which case the plume closest to the potential emission source is used

for quantification purposes. More technical details of the technology can be found in Kairos technical white papers and patent documentation.[18–20]

S1.2 Controlled release experiments

See Sherwin, Chen et al.[22] for more details on the controlled release experiment.

Sherwin, Chen et al. conducted a single-blind controlled methane release in San Joaquin County, California to test the LeakSurveyor detection limit and quantification accuracy in October 2019.[22] The controlled release trials were partially motivated by the 2018 Mobile Monitoring Challenge (MMC) organized by the Stanford Natural Gas Initiative and the Environmental Defense Fund (EDF). For the MMC, ten methane detection technologies were tested through single-blind controlled releases. Kairos was one of the 12 teams originally selected for the MMC, but was unable to participate because rice fields surrounding the MMC test site were flooded at the time of the MMC, which blocked the Kairos spectrometer's ability to sense infrared radiation in the spectral region of interest, rendering the testing scenario out of design scope of the technology. The 2019 controlled release study by Sherwin, Chen et al. used a different test site than the 2018 MMC study, but used the same releasing apparatus and the same gas operator, Rawhide Leasing, and was conducted specifically to assess Kairos' LeakSurveyor technology.

Figure S1 shows (a) confidence in methane presence and (b) methane enhancement of one plume from the 2019 single-blind controlled releases.

Unlike the MMC, whose largest release was approximately 29 kg/h(CH_4), the 2019 Sherwin, Chen et al. controlled release had its largest release reach over 1,000 kg/h(CH_4).[22,25] Results of the 2019 single-blind controlled release are reproduced in Figure S2(a-d) with adjustments described in Section S1.3. In the Sherwin, Chen et al. controlled release test, the wind-independent methane rates reported by Kairos were multiplied by (a) cup wind meter-measured 1-minute gust wind speeds, (b) ultrasonic anemometer-measured 1-minute gust wind speeds, (c) The Dark Sky Company, LLC (henceforth "Dark Sky")-modeled 1-

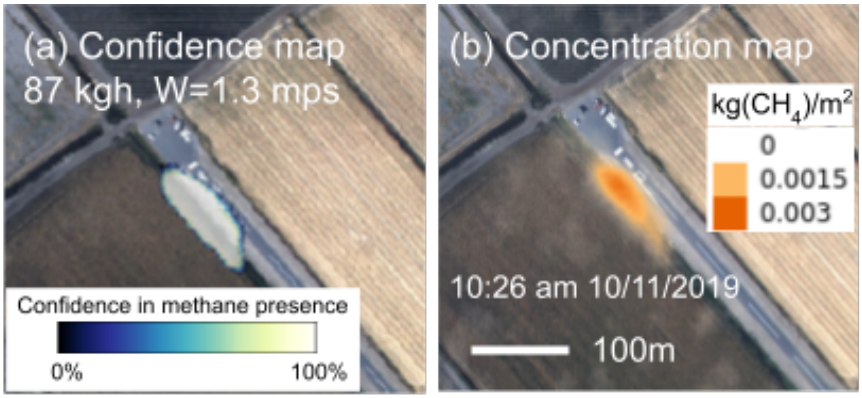

Figure S1: Confidence and enhancement map of a methane plume from the 2019 controlled release test. The methane release rate is 87 kg/h and the ground-measured wind is 1.3 m/s. Graphs adapted from Sherwin, Chen et al.[22]

minute gust wind speeds, and (d) NOAA's High Resolution Refresh (HRRR) modeled hourly surface gust speeds averaged across time and space (see Section S1.3.2 for details) in Figure S2a-d. Note that the number of data points are different in these panels because of the data exclusion criteria described in[22] and the difference in wind data availability across the wind data sources.

In addition to the Sherwin, Chen et al. single-blind tests, Kairos has conducted internal controlled release trials and collected another 312 data points, including 276 releases and 36 negative controls (see Figure S2e-f). In this paper, these data points were added to the single-blind test data set for a more comprehensive representation of performance.

S1.3 Data

S1.3.1 Releases

The flow-meter-measured release rate is treated as the actual methane release rate in the controlled release trials (AMR_{CR}) and the Kairos-estimated methane rate is denoted as EMR_{CR}. Note that Kairos reports wind-independent emission rate in the unit of kgh/mps. The wind-independent rates are multiplied with the hourly surface wind speed in mps to derive the estimated methane rate of the controlled releases, EMR_{CR}.

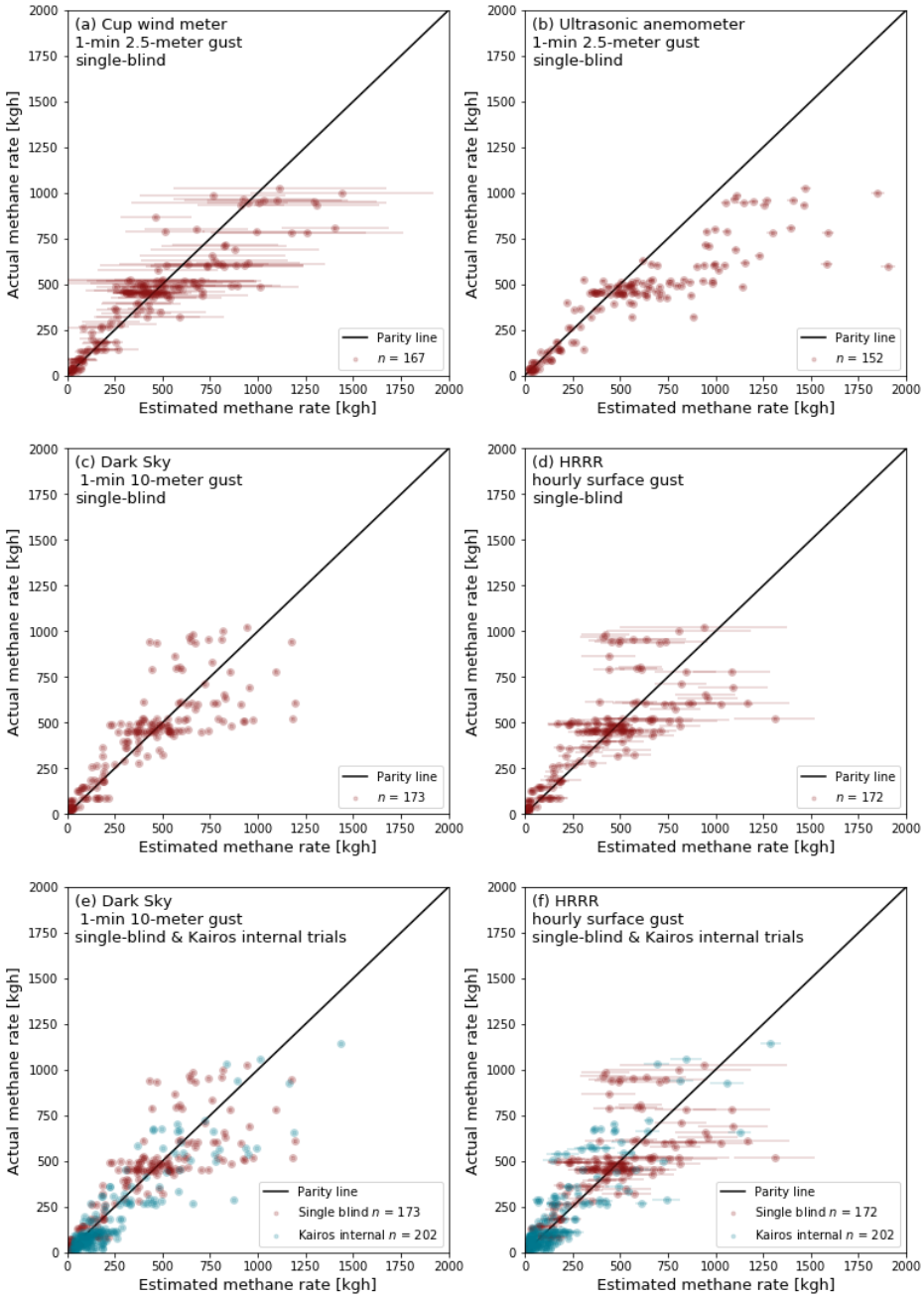

Figure S2: Scatter plots of actual and estimated methane rates in the single-blind controlled release trials based on (a) cup wind meter-measured wind, (b) ultrasonic anemometer-measured wind, (c) Dark Sky 1-minute gust wind, and (d) 27-average HRRR surface gust wind (see description in Section S1.3.2). The two bottom panels (e-f) add data points from Kairos internal field trials to the two middle panels (c-d). The false negatives are excluded from the plots for reasons detailed in Section S1.3. X error bars are based on wind uncertainties, described in Sherwin, Chen et al.[22] Note that wind measurement uncertainty in the ultrasonic anemometer is smaller than point size, while Dark Sky does not report uncertainty. Y error bars, not visible, are based on observed flow variability and flow meter error.[22]

Data in Figure S2a-d differs from the previously published version in Sherwin, Chen et al. in four respects. First, the axes are inverted because in this work our aim is to find the actual methane rates (y-axis values in Figure S2) based on Kairos-estimated methane rates (x-axis values), whereas Sherwin, Chen et al. controlled the actual flow rate data and verified the accuracy of the technology. In other words, the calibration step in this work (described in Section S1.5 and S5.1) is a inverse process to the verification performed by Sherwin, Chen et al.[22]

Second, we exclude the false negatives from our calibration chart. The Kairos-estimated wind-independent methane rates multiplied with cup wind meter, Dark Sky wind, HRRR wind (Figure S2a, c, and d) have 18 false negative controlled release trials, and the ultrasonic anemometer case (Figure S2b) has 5 false negative data points because the ultrasonic anemometer was not available for part of the field work.[22] We exclude false negatives in the quantification adjustment to avoid bias because false negatives cannot be recognized in the Kairos New Mexico Permian campaigns.

Third, we do not introduce height adjustment to the wind retrieved from Dark Sky and HRRR. In the field trial, the cup wind meter and the ultrasonic anemometer measured wind at 2.5 meters above ground, whereas Dark Sky and HRRR report a 10-meter wind by default. For a fair comparison, Sherwin, Chen et al. applied a height adjustment factor of 0.81 to the Dark Sky and HRRR wind, assuming a vertical wind profile following a power law and a surface roughness coefficient of 0.15.[26] Given the height of the release stack at 2.5 meters, this correction was appropriate. The vertical wind profile varies by topography and other weather factors, and is therefore not likely to be the exact same at the controlled release test site in San Joaquin County, California and at the emission sources found in the New Mexico Permian Basin. In this work, the height adjustment is done implicitly in the curve fitting process (Section S1.5).

The fourth difference is in the Dark Sky 1-minute gust wind speed. Our February 2020 and October 2020 retrievals of wind data from Dark Sky both differ from the Dark Sky wind

data in Sherwin, Chen et al., which was retrieved in October 2019.[22] Figure S4 compares the two retrievals of this paper (February and October 2020). Dark Sky is a private company, and their weather models and data sources are not extensively accessible. Two probable reasons for the discrepancy in the two retrievals 9-month apart are that 1) Dark Sky made changes to their weather models between the two retrievals, and 2) Dark Sky downgrades the time resolution of historic data due to memory concerns. Therefore, we refer to the February 2020 retrieval as "high time resolution retrieval" and the October 2020 retrieval as "low time resolution retrieval" to reflect this change. Moreover, Dark Sky was acquired by Apple Inc. in March 2020 and announced a planned termination of their API service by the end of 2021.[27] Considering Dark Sky's lack of documentation and its approaching expiration of data availability, we use NOAA'S HRRR hourly product for the base case here, despite the resulting downgrade in spatial and temporal granularity. Sensitivity cases based on Dark Sky winds are presented in the main text and Section S7.

Additionally, Figure S2e and S2f includes 202 data points of Kairos' internal controlled release trials, which Kairos conducted on eight distinct days, four of which focused exclusively on smaller release rate ranges (<250 kg/h). 38 data points were collected at a test site in Alberta, Canada on one test day, and the rest were collected at test sites in California on seven test days. We do not include data from Kairos' internal trials for the curve fitting process. Instead, we use data from the single-blind trials due to the more independent data generation process and the larger release rates tested.

S1.3.2 Wind

HRRR We use HRRRv3 surface gust wind, available as an hourly, 3 km grid product from the HRRR archives at the University of Utah.[28] We process the hourly data based on a method developed by Duren et al. to account for the spatial and temporal variability in HRRR wind fields. Duren et al. applied the processed HRRR wind to their Airborne Visible-Infrared Imaging Spectrometer - Next Generation (AVIRIS-NG) measurements for

identification of California's methane super-emitters.[23] The type of wind speed used to multiply with wind-independent measurements is the average of 27 HRRR winds from the spatially 3×3 box centered on the plume source and from 3 time-steps (plume detection time ±1 hour).

Figure S3 shows four parity charts using different HRRR hourly winds. Figure S3a uses the instant (at detection hour) wind at the HRRR grid point closest to the controlled release site. Figure S3b represents the average of three hourly measurements (±1 hour) at the closest grid point. Figure S3c uses the average of nine instant winds in the 3×3 box centered on the plume source. Figure S3d uses the average of 27 HRRR wind from 9 grid points and 3 time steps. The 95% confidence interval error bar length of each data point is 1.96 times the standard deviation of the measurements from the temporally and spatially neighboring wind estimates.

As Figure S3 shows, the R^2 values in the four panels are, respectively, 0.64, 0.64, 0.68, and 0.70. We choose to use the 27-average HRRR wind as the base case in this study due to its best performance in linear correlation, and also because it is an established method in Duren et al.[23]

Dark Sky Proprietary wind reanalysis product from Dark Sky aggregates weather data from a wide range of sources, including the NOAA's Meteorological Assimilation Data Ingest System (MADIS) and a network of proprietary wind stations.[24] Dark Sky models hyper-local weather conditions for each minute. As of this writing, Dark Sky has provided the best temporal and spatial resolution for modeled wind speed among available weather reanalysis products. However, when acquired by Apple Inc. in March 2020, Dark Sky announced a planned termination of their API service by the end of 2021.[27]

Figure S4 compares the two retrievals of Dark Sky wind data in February 2020 and October 2020. Both retrieve the 2019 California controlled release field trial wind data and the 2018-2020 New Mexico Permian campaign wind data. Two probable reasons for the

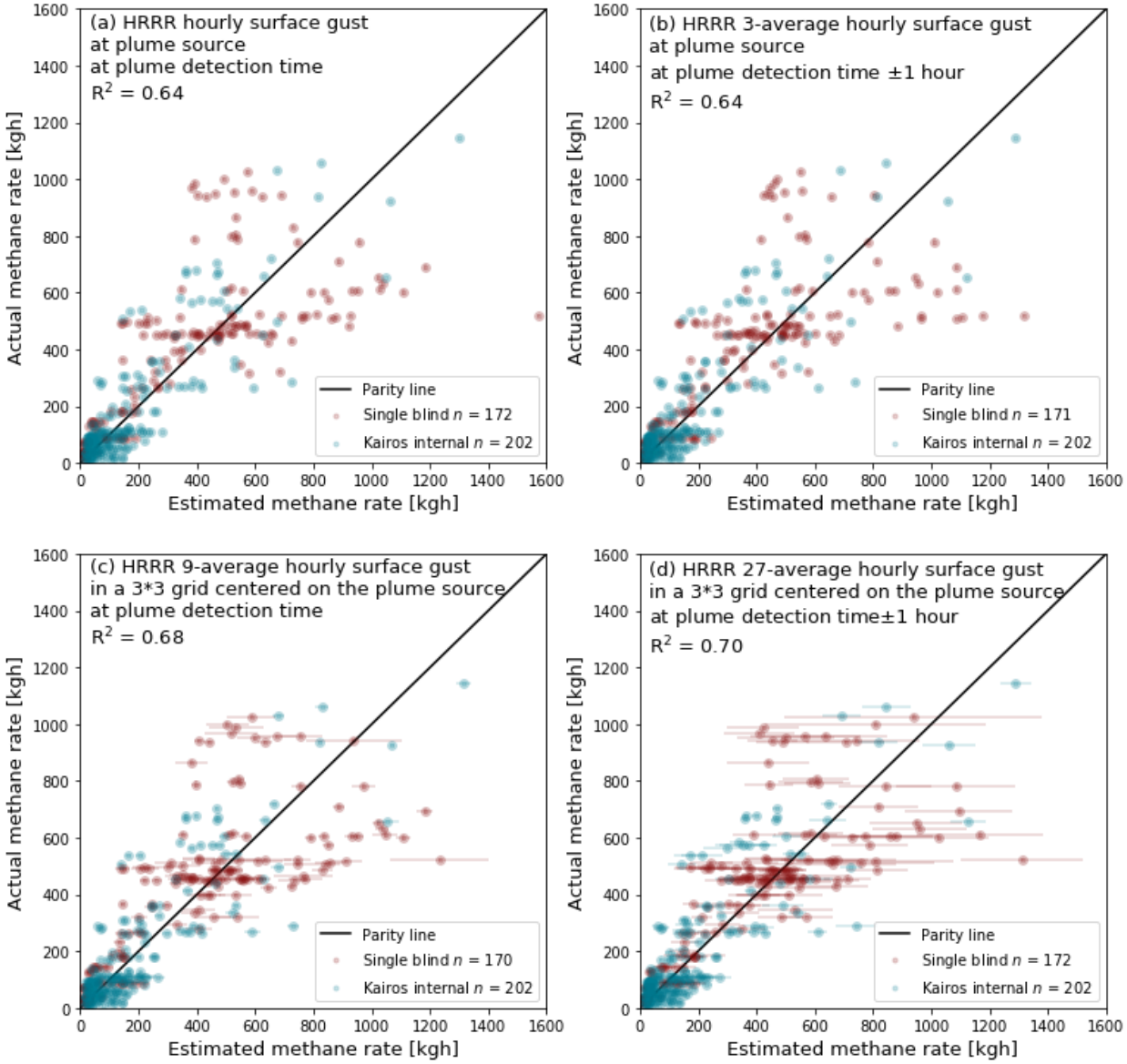

Figure S3: Choice of HRRR wind affects the evaluation of Kairos technology's quantification accuracy. (a) uses the instant (at detection hour) wind at the HRRR grid point closest to the plume source. (b) uses the average of three (±1 hour) hourly measurements at the grid point. (c) uses the average of nine instant winds in the 3×3 box centered on the plume source. (d) uses the average of 27 HRRR wind from 9 grid points and 3 time-steps. The difference in the number of data points is due to the "plume formation time sufficiency" data exclusion criteria detailed in the SI of.[22] The lengths of the error bars on the x-axes represent Kairos-reported wind-independent methane rates multiplied by 1.96 times the standard deviation of wind estimates from temporally and spatially neighboring HRRR wind estimates. The point estimates in (a) do not have associated error bars because the HRRR archive at the University of Utah does not provide estimates of uncertainty in point estimates.[28]

discrepancy in the two retrievals 9-month apart are that 1) Dark Sky made changes to their weather models between the two retrievals, and 2) Dark Sky downgrades the time resolution of historic data due to memory concerns. Therefore, we refer to the February 2020 retrieval as "high time resolution retrieval" and the October 2020 retrieval as "low time resolution retrieval" to reflect this change.

As noted above, we use HRRR for our base case in favor of HRRR's replicability and availability. Nevertheless, we present two sensitivity cases based on two retrievals of Dark Sky winds in Section S7. Note that each of the two sensitivity cases are based on their own fitted curves from the controlled release results.

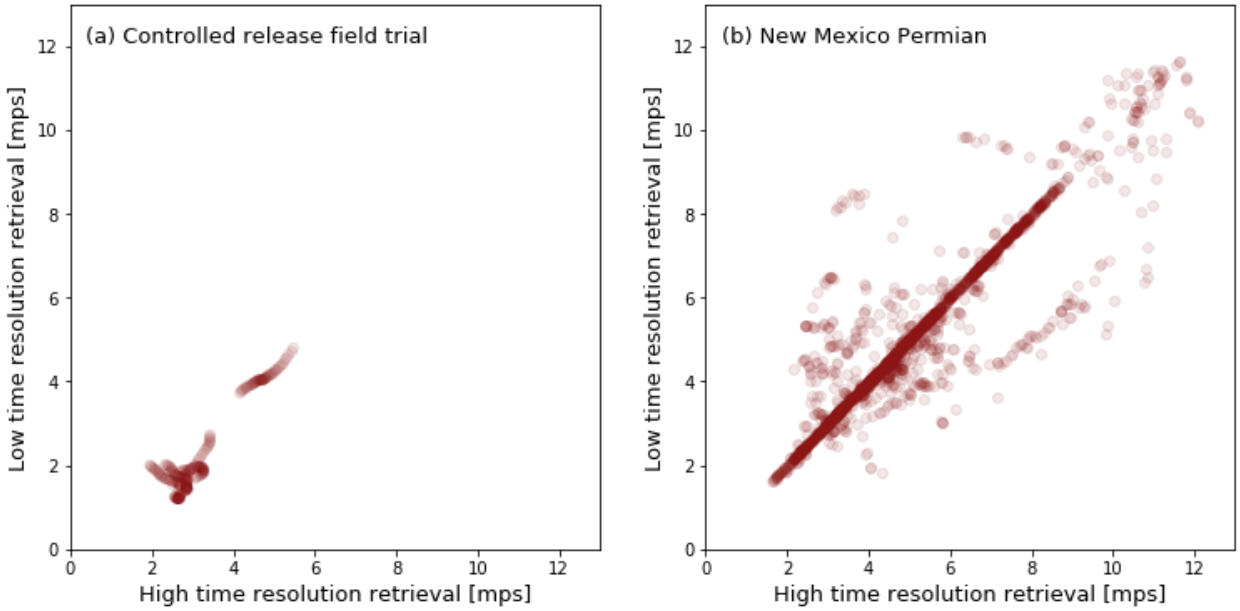

Figure S4: Comparison of Dark Sky 1-minute gust wind speed of (a) the controlled release field trial and (b) the plume incidences detected in New Mexico Permian from high-time-resolution (February 2020) and low-time-resolution (October 2020) data retrievals.

Modeled vs measured wind Figure S5 reveals that Dark Sky-modeled wind shows a better linear correlation with the cup wind meter-measured wind at the controlled release sites (the San Joaquin site in the single-blind trials as well as other sites for Kairos internal trials) due to its advantage in spatial and temporal resolution over HRRR. Regardless, we use HRRR wind for our base case in favor of HRRR's accessibility and replicability.

Figure S5 also demonstrates the challenge of accurately modeling wind speed with reanalysis products such as HRRR and Dark Sky. The errors in release rate quantification depends both on the uncertainties in the Kairos-reported wind-independent release rates and on the uncertainties in the modeled wind speeds in mps. In fact, the release rate quantification error can be largely attributable to wind errors. Use of improved wind reanalysis products in future studies has a large potential to improve emission quantification accuracy.

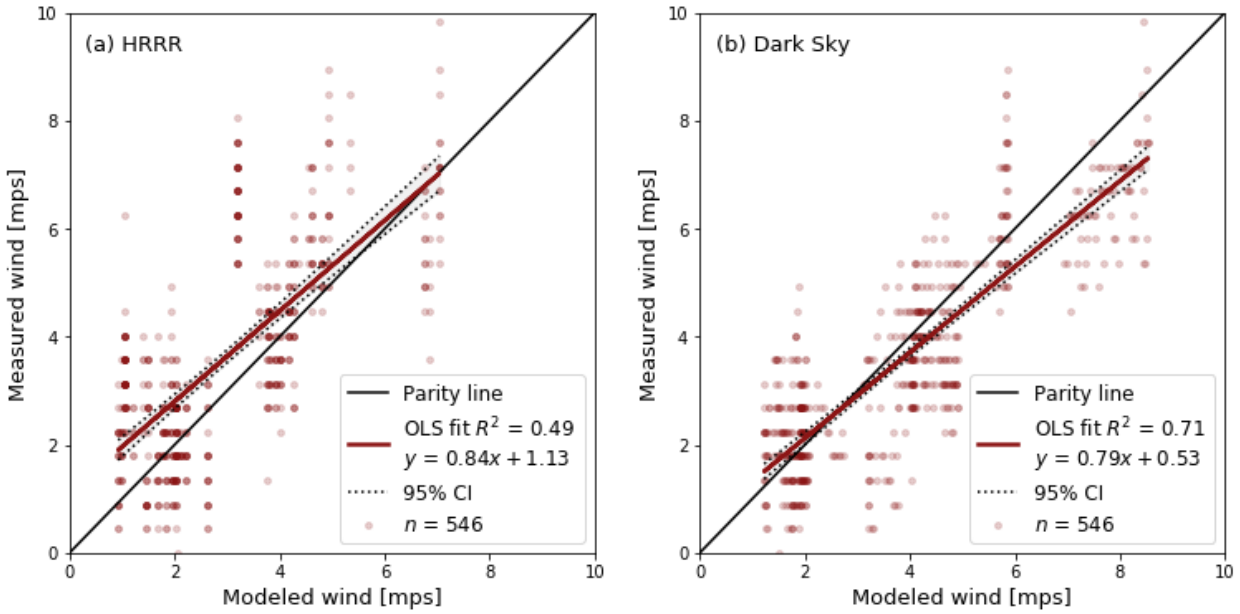

Figure S5: Cup wind meter-measured minutely gust wind from controlled release trials (38 data points from Alberta, Canada and 508 from California test sites) plotted against (a) HRRR 27-average wind gust and (b) Dark Sky-modeled 1-minute wind gust. The black dashed lines show 95% confidence intervals of the regression line fits. Due to its finer spatial and temporal resolution, Dark Sky-modeled wind shows a smaller spread around the parity line and the regression fit.

S1.4 Detection limit

We find the partial detection range (PDR) of the technology by finding the minimum release rate that the technology can detect with some nonzero probability and the rate above which all controlled releases are detected. The PDR is assessed with wind-independent release rate in the unit of kgh/mps, because wind dissipates the released methane and is a common

factor to incorporate in PDR characterization of many methane detection technologies.[25,29]

The single-blind field trial was primarily set up to test the accuracy of the technology at medium- to large- methane rates, with limited resources devoted to characterizing the PDR of the Kairos technology.[22] Therefore, data collected from both single-blind trials and Kairos internal trials are combined in this study to better characterize Kairos instrument's detection limit. Single-blind trials contribute 221 valid data points, including 200 positive releases and 21 negative control trials. No false positives were detected for the 21 negative control trials.

Kairos internal trials, which were conducted on eight different days ranging from October 2017 to September 2020, include 312 valid data points, including 276 positive releases and 36 negative controls. No false positives were detected for these 36 Kairos-internal negative control trials.

Therefore, in this study, we assume that all detected plumes in the New Mexico Permian are valid emissions. The observed distribution of emissions falls off rapidly in the partial detection range of (4,14] kgh/mps (see Section S1.4). Given that false positives would likely fall in this smaller range, this suggests that to the extent they exist in the dataset, false positives likely play a small role.

We define the cup wind meter-measured wind-independent release rates as the actual wind-independent methane rates ($AMR_{CR}^{WN} = \frac{AMR_{CR}}{Wind^{cup}}$).

Among the 476 valid releases from both the single-blind trials and the Kairos internal trials, 192 releases have $AMR_{CR}^{WN} < 20$ kgh/mps. Figure S6a shows binary detection results as black circles on the upper and lower end of the graphs. These binary results are grouped into bins with widths of 2 kgh/mps and the detection probability of each bin is represented as a fraction x/y, where x is the number of detected releases and y is the number of all releases in the bin. The 5%, 50%, and 95% detection probabilities according to the sigmoid fit in Figure S6a occur at 4.5, 9.5, and 14.6 kgh/mps. For simplicity, the wind-independent PDR is determined to be (5,15] kgh/mps. The estimated 50% detection threshold of 9.5 kgh/mps

is close to Kairos' self-reported a 50% probability detection threshold at 9.8 kgh/mps.[18] The PDR identified using data from both the single-blind and the Kairos internal controlled releases is the same as the PDR of (5,15] kgh/mps identified using only the single-blind trial data in Sherwin, Chen et al.[22]

In real field applications, actual cup-based wind measurements on the ground are not available for each plume. For this reason, modeled wind sources such as HRRR have to be used. Modeled wind introduces errors to the analysis (Figure S6b). When we switch the denominator in calculating AMR_{CR}^{WN} from cup wind meter-measured wind speed to HRRR modeled-wind, the proportion detected does not grow smoothly with release rates due to the errors in modeled wind.

Using HRRR synthetic wind speed ($Wind^{HRRR}$) as the denominator, the number of data points with wind-independent release rates below 30 kgh/mps is 206. Figure S6c shows 10 incidences of detection failure when the wind-independent release rate is above 15 kgh/mps. This is caused by the noise in the synthetic HRRR wind (Figure S6b). Of these 10 data points, 5 are from the single-blind trials and 5 are from Kairos internal trials. In all these 10 cases, the HRRR winds are significantly smaller than the cup wind meter measurements, driving the wind-independent release rates larger.

Due to the challenge of modeling wind, the synthetic HRRR wind introduces noise to characterizing the PDR. The 50% detection probabilities on the sigmoid fit is 10.8 kgh/mps, which is close to the value based on measured winds and indicates that the synthetic wind introduces noise but does not strongly affect the estimate of where 50% detection probability occurs.

For the base case of study-area emission quantification presented in the main text, we use 2 and 30 kgh/mps as the minimum and full detection limits (MDL and FDL) and the detection probability of each bin in Figure S6c to characterize the partial detection thresholds. Despite the noise introduced by using synthetic wind, the results based on HRRR wind are more generalizable for interpreting New Mexico data given the absence of

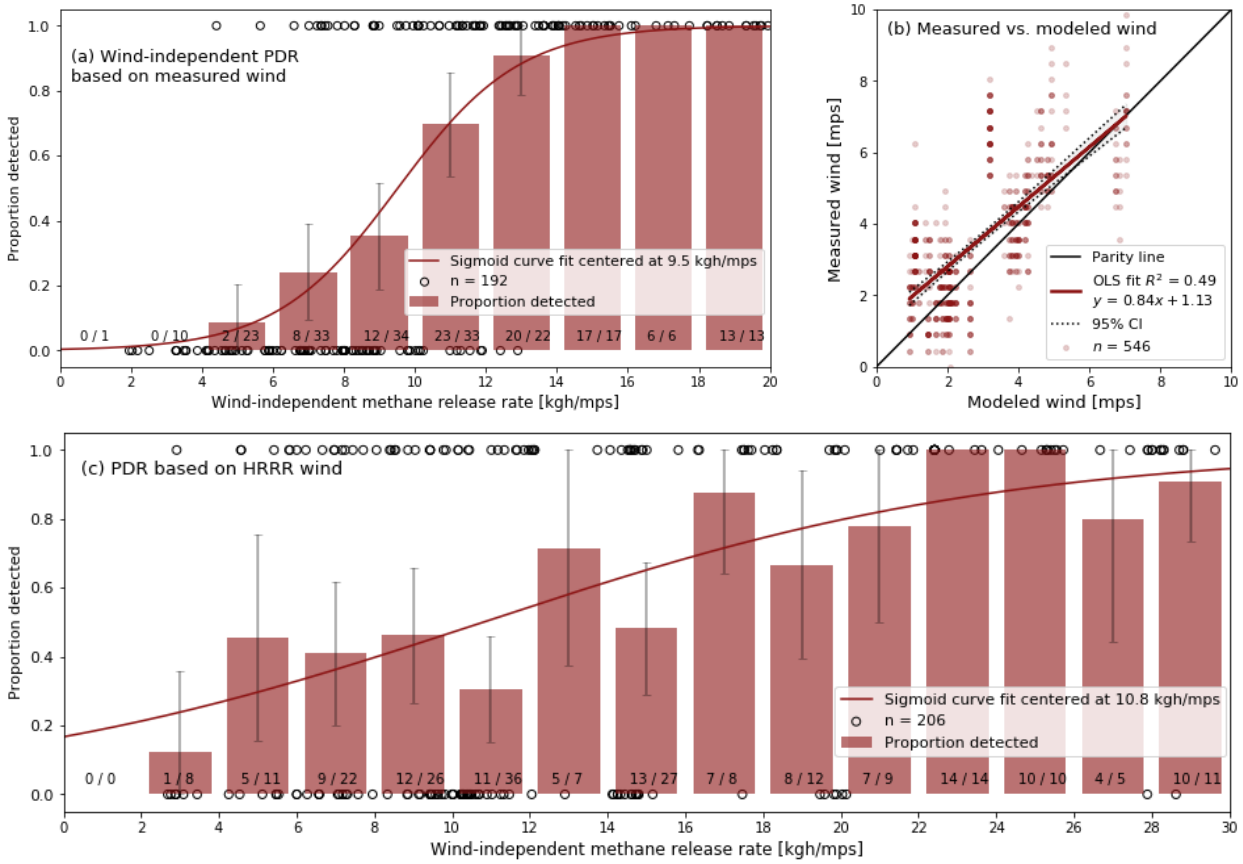

Figure S6: (a) Wind-independent partial detection range (PDR) is (4,14] kgh/mps based on actual release rates and cup wind meter-measured wind speed. Binary results are plotted as black circles on the upper and lower ends of the graph. The binary results are grouped into bins with widths of 2 kgh/mps and the detection probability of each bin is represented as a fraction x/y, where x is the number of detected releases and y is the number of all releases in the bin. The height of each bar indicates the value of x/y. The error bars of each bin indicate 95% confidence interval of detection probability assuming a binomial distribution. (b) Figure S5a reproduced to demonstrate the errors introduced by modeled wind here. (c) PDR characterization based on HRRR wind. Modeled wind introduces noise into the detection probability of each bin of wind-independent methane release rate.

ground wind measurements in the New Mexico emissions data set.

In the New Mexico survey, none of the 1985 plumes are associated with wind-independent emission rates below 2 kgh/mps. 1112 are within the PDR of 2 to 30 kgh/mps. The remaining 873 plumes are above 30 kgh/mps. Despite the fact that over half of the plumes are found below the FDL, due to the heavy-tailed emission size distribution, the contribution of total emissions from these detected plumes within the PDR is less than 5% of total emissions of the entire study area and therefore does not greatly affect the final quantification results. Section S5.1 describes the contribution of emissions from the partial and the full detection ranges in detail. We do not present a sensitivity case based on the partial detection characteristics in Figure S6a because of the insignificant contribution to the total emissions from the smaller leaks.

S1.5 Curve fitting

Data selection We exclude data points from Kairos' internal trials in the quantification curve fitting process, and preferentially use the data from the single-blind study in favor of the data independence and the larger release rates tested in the single-blind trials.

We exclude undetected releases (EMR_{CR}=0) in the curve fitting process, considering that the New Mexico Permian campaign data does not indicate the probability of leakage based on null observations and thus the inclusion of the undetected controlled releases would introduce bias in quantification.

Curve type selection Sherwin, Chen et al. find a good linear fit of data when setting actual methane rate (AMR_{CR}) on the x-axis and estimated methane rate (EMR_{CR}) on the y-axis, despite some heteroscedasticity and noise introduced by HRRR wind.[22] Therefore, as a starting point, we fit a conventional linear regression model (Figure S7) to (EMR_{CR}, AMR_{CR}). The fitted relationship is as follows:

$$\widehat{AMR}_{CR} = 0.72 EMR_{CR} + 121.55 \tag{S4}$$

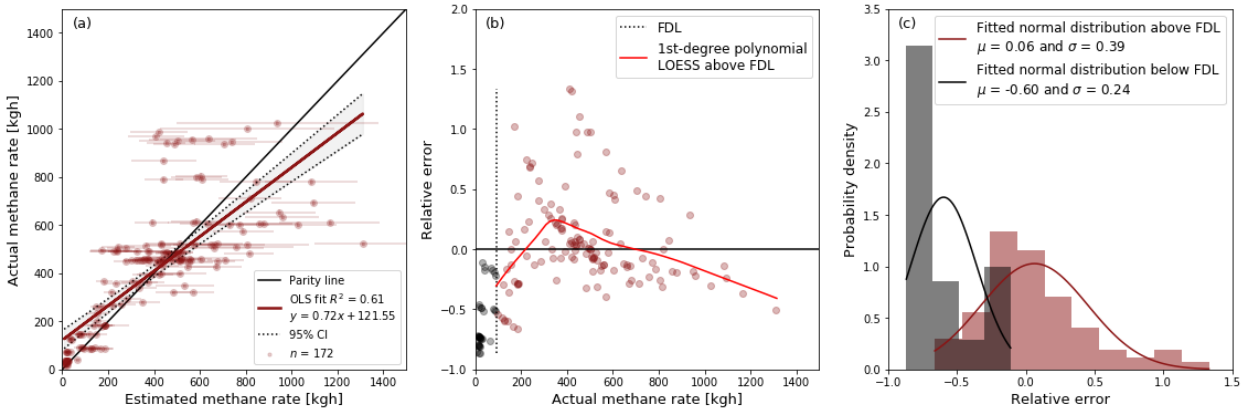

Figure S7: (a) Conventional linear fit of single-blind controlled release trial data. The black dashed lines show the edges of the 95% confidence intervals of the fit. The R^2 value here differs from the value presented in Figure S3d due to the exclusion of data from Kairos internal trials. (b) Relative error (defined in Equation S6) first grows at small actual release rates and declines at larger release rates. The black dotted line indicates the full detection limit (FDL) assuming New Mexico Permian median 27-average HRRR wind speed from HRRR at plume detections (see Section S1.3.2). (c) Histograms of relative errors for releases above and below FDL. The two distributions differ and the distribution for releases below FDL does not resemble a normal distribution.

The intercept of Equation S4 is 121.55±40.63 (95% CI), which is significantly larger than the expected value of zero. The nonzero intercept will artificially increase low-end emissions, while the fact that the line's slope is below 1 will reduce the magnitude of emissions above the intersection with the parity line at ∼500 kg/h. To address the possibility that the nonzero intercept could introduce upward bias into the calibration, we fit another regression line with the intercept fixed at the origin, as presented in Figure S8 and Equation S5.

$$\widehat{AMR}_{CR} = 0.93 EMR_{CR} \tag{S5}$$

Compared to Sherwin Chen et al., flipping the axes to place estimated methane emissions on the x-axis also exacerbates the non-uniformity in residuals.[22] This is due in part to boundary bias issues, described in Section S1.6. In Figure S7b, we plot the relative error δ

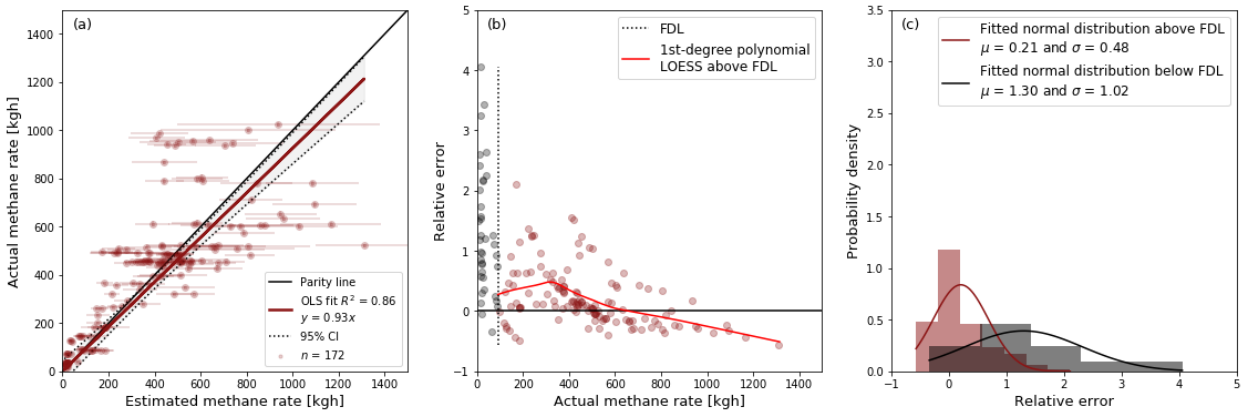

Figure S8: (a) Linear regression with intercept fixed at the origin for single-blind controlled release trial data. The black dashed lines show the edges of the confidence intervals of the fit. Note that the R^2 defined for linear regression with intercept fixed at the origin differs from the R^2 defined for ordinary least square fit presented in Figure S7. (b) Relative error (defined in Equation S6) first grows at small actual release rates and declines at larger release rates, likely due in part to boundary bias issues described in the Section S1.6. The black dotted line indicates the full detection limit (FDL). (c) Histograms of relative errors for releases above and below FDL. The two distributions differ and the distribution for releases below FDL does not resemble a normal distribution, in part because it is not possible to have relative error below -100%.

(defined in Equation S6) with respect to AMR_{CR} and observe a downward trend in δ.

$$\delta = \frac{AMR_{CR} - \widehat{AMR}_{CR}}{\widehat{AMR}_{CR}} \tag{S6}$$

The nonzero intercept and the non-uniform residual spread illustrate potential limitations of the linear model. The downward trend in δ, in particular, suggests that the linear fit may be over-estimating the leakage rate at high volumes. There are two main hypotheses for why this may occur:

1. There is an underlying nonlinearity in the relationship between gas release volume and Kairos' quantification algorithm.

2. There is an experimental issue with boundary effects in the sample size driving errors to be increasingly one-sided as we approach the largest release rates.

In the base case, we apply the first hypothesis and fit a sublinear curve to the data

prompted by the inspection of downward-bending trend at large AMR_{CR} values. The power curve (Equation S7) is found by fitting a regression line to the log-transformed data. As Figure S9b shows, the relative error δ of the sublinear fit stays relatively constant with respect to AMR_{CR}. The relative errors for releases below and above Kairos full detection limit are not statistically different (Figure S9c). The relative error roughly follows a normal distribution of mean 0.07 and standard deviation 0.40 (Equation S8).

$$\widehat{AMR}_{CR} = 4.08 EMR_{CR}^{0.77} \tag{S7}$$

$$\delta = \frac{AMR_{CR} - \widehat{AMR}_{CR}}{\widehat{AMR}_{CR}} \sim N(\mu_\delta = 0.07, \sigma_\delta = 0.4) \tag{S8}$$

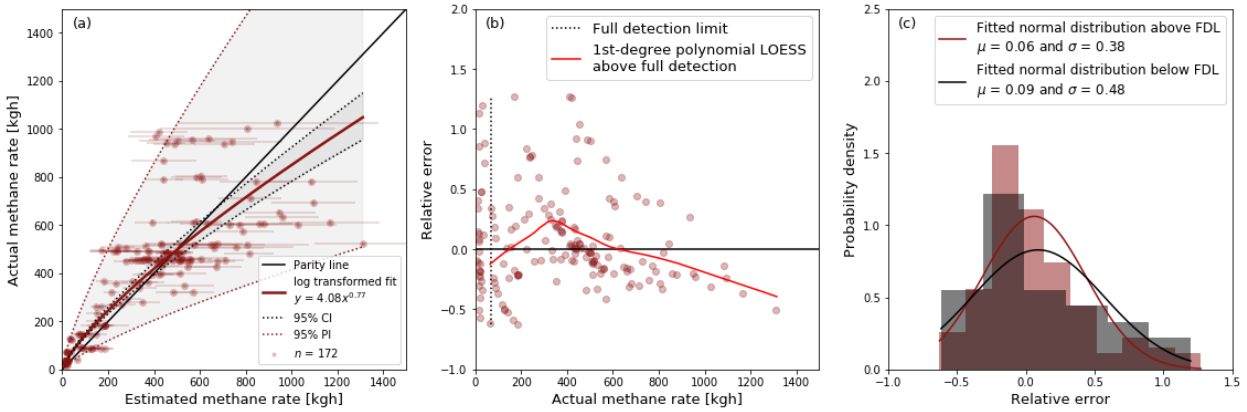

Figure S9: (a) Sublinear fit of single-blind controlled release trial data. The black dashed lines show the edges of the confidence intervals of the fits and the red dashed lines show the edges of the prediction intervals of the fits. (b) Relative error does not show distinct trends with respect to actual methane rate. (c) Distributions of relative error for data above and below full detection limit (MDL) are not statistically different.

Note that the sublinear power law fit (Equation S7) is conservative in estimating large emissions and is even more so for estimating some emissions found in the New Mexico Permian Basin that are larger than the largest controlled release (1025 kg/h).

Note that in both the linear and the sublinear cases, there is potential upward bias from emissions below ~400 kg/h. However, on average we would expect this power curve

to introduce a downward bias into the results due to scale bias issues related to the fact that the mean of a lognormal distribution is greater than the exponentiated mean of the logarithm of the underlying data points. This is discussed further in a report by Eastern Research Group.[30]

In Section S5.1, for the base case we first apply Equation S7 to the estimated emission rates and then use a Monte Carlo approach to incorporate the measurement error. We do this by rearranging Equation S6 and assuming that the distribution of δ is the same for the controlled releases and in the New Mexico Permian Basin:

$$AMR_{CR} = (1+\delta)\widehat{AMR}_{CR}, \text{ where } \delta \sim N(\mu_\delta = 0.07, \sigma_\delta = 0.4) \tag{S9}$$

S1.6 Boundary bias correction

Alternatively, hypothesis 2 asserts that the nonlinearity results from an experimental issue with boundary effects. The largest actual methane rate (AMR_{CR}) in the single-blind field trial was 1025 kg/h, which indicates that all Kairos-estimated methane rates (EMR_{CR}) above that value can only be overestimates of AMR_{CR}. The inclusion of these estimates with asymmetric error bars in the curve fitting process could introduce an downward bias in the slope coefficient.

To demonstrate this phenomenon, we carry out a simulation using the following steps:

1. Suppose that the true (EMR_{CR}, AMR_{CR}) relationship can be described by the parity line ($\widehat{AMR}_{CR} = EMR_{CR}$) in Figure S10.

2. Apply a normally distributed relative error δ with μ_δ=0 and σ_δ=0.39 (39%) to $\widehat{AMR}_{CR}$ values on the parity line to generate the data for simulation. Assume these are the true values of AMR_{CR}. In this simulation practice, the value of σ_δ=0.39 is taken as that of the relative errors from fitting the controlled releases data with a linear line (Equation S4). This assumption is made for the purpose of demonstration.

3. Leave out the data points with $AMR_{CR} > AMR_{CR}^{max}$.

4. Refit a regression line to EMR_{CR} and the remaining $AMR_{CR} <= AMR_{CR}^{max}$. Denoted the new line fit as $\widehat{AMR}'_{CR} = \alpha'$ EMR + β'.

5. Repeated steps 2-4 one hundred times.

Figure S10 shows the results of the simulation. The black scattered points show (EMR_{CR}, AMR_{CR}) from one of the 100 simulation realizations. The data points with simulated $AMR_{CR} > AMR_{CR}^{max}$ are shown as crosses and are excluded in the refitting processes. These data points are the ones "missing" from the field trial because the actual release rate never exceeded 1025 kg/h. Without these data points, linear regression produces a biased estimate of the relationship. We refer to this as the "boundary bias effect."

This problem does not occur in Sherwin, Chen et al., where the plots are presented with EMR_{CR} on the y-axis and AMR_{CR} on the x-axis, because a vertical cutoff threshold of data points does not bias the linear fit. However, in this study we present AMR_{CR} on the y-axis and EMR_{CR} on the x-axis for calibration purposes. Consequently, a horizontal cutoff threshold would bias the linear fit slope low, as the simulations in Figure S10 show. The refit slopes α' from the 100 simulations have a mean of 0.81±0.14 (95% CI), which is significantly lower than the true slope of 1.

The simulated slope of 0.81±0.14 by taking into account the boundary effect demonstrates that Hypothesis 2 can be one major cause of the non-linearity in the data and drives the slope (0.72±0.09) in Equation S4 below 1, as shown in Figure S10. Note that 0.72 is within the confidence interval of the simulated mean (0.81±0.14).

To address the boundary bias problem, we leave out data points with large EMR_{CR} values from the curve fitting process, as we believe that smaller EMR_{CR} are less affected by the boundary bias effect. By setting a vertical cutoff threshold and fit curves exclusively to data points left of the threshold, we can reduce the boundary bias. The criterion for determining the cutoff threshold (EMR_{CR}^{cutoff}) is:

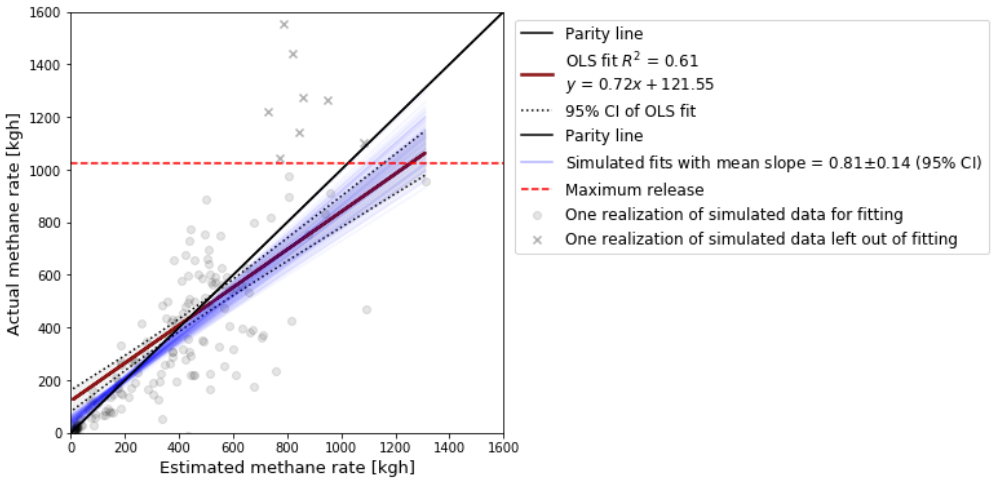

Figure S10: Simulations for demonstrating the boundary bias effect. Light blue lines represent linear fits to 100 simulated datasets assuming perfect 1:1 agreement between estimated and actual emissions with normally distributed percent error. The light blue lines are fit only to all simulated data points below the maximum emission rate of 1025 kg/h observed in controlled releases (dashed red line).[22] This simulates the effect on the linear fit slope of limiting the size of the maximum methane release used for calibration. Gray dots represent one realization of simulated data, with high-emission points excluded from the fit marked with an "x." The simulations assume a percent error standard deviation of 39% based on the observed error distribution to the linear fit for observed data (solid red line). Note that the slope of the red line, fit to actual data, is within the confidence interval of the simulated datasets.

$$(1 + \sigma_\delta) EMR_{CR}^{cutoff} <= AMR_{CR}^{max} \tag{S10}$$

We also show a more stringent case of the following criteria:

$$(1 + 2\sigma_\delta) EMR_{CR}^{cutoff} <= AMR_{CR}^{max} \tag{S11}$$

In Equation S10 and S11, the values of σ_δ are based on the distributions of relative errors from linear fittings to data with $EMR_{CR} < EMR_{CR}^{cutoff}$. Therefore, the values of EMR_{CR}^{cutoff} cannot be obtained explicitly. Hence, we test EMR_{CR}^{cutoff} at 10 kg/h increments (Table S1) and find EMR_{CR}^{cutoff} of 750 kg/h that is one σ_δ below AMR_{CR}^{max} and 590 kg/h to be $2\sigma_\delta$ below AMR_{CR}^{max}. Note that the δ used here is the δ (fitted normal distribution

of methane emissions percent errors) based on emissions larger than the full detection limit (FDL), because releases size above FDL are more relevant for finding the cutoff threshold and the prospect of partial detection introduces non-normal errors.

Table S1: Selection of the upper boundary cutoff threshold for estimated methane rate using Equation S10 and S11. By testing EMR_{CR}^{cutoff} at 10 kg/h increments, the one σ_δ cutoff threshold is 750 kg/h and the $2\sigma_\delta$ cutoff threshold is 590 kg/h. Note that by a cutoff of 750 kg/h, the slope is within 10% of perfect 1:1 agreement.

EMR_{CR}^{cutoff}	N	Slope	Intercept	μ_δ	σ_δ	$(1+\sigma_\delta)EMR_{CR}^{cutoff}$	$(1+2\sigma_\delta)EMR_{CR}^{cutoff}$
1320	172	0.719	121.545	4.5%	39.0%	1834	2348
1310	171	0.759	109.714	4.0%	38.3%	1812	2314
					.....		
760	155	0.939	64.976	2.4%	36.9%	1040	1320
750	154	0.955	61.488	2.2%	36.7%	1025	1300
740	153	0.945	63.727	2.4%	36.9%	1013	1286
					.....		
600	139	1.031	47.328	1.6%	36.4%	819	1037
590	139	1.031	47.328	1.6%	36.4%	805	1020
580	137	1.023	48.505	1.7%	36.8%	794	1007
					.....		

Figure S11 shows three sets of regression results based on (a) all single-blind controlled release data, (b) estimated releases below 750 kg/h, and (c) estimated releases below 590 kg/h. Cases b and c, respectively, correspond to cutting off thresholds of one and two σ_δ below the maximum release rate. The exclusion of data points of large EMR_{CR} values brings the slope up, as the data points used in the line fitting processes are less affected by the boundary bias effect.

In the base case presented in the paper, we apply hypothesis 1 and do not apply this data exclusion criterion. Section S7 presents a case assuming hypothesis 2 that employs Equation S10 as the cutoff threshold criteria. The resulting linear fit following hypothesis 2 is:

$$\widehat{AMR}_{CR} = 0.96 EMR_{CR} + 61.49 \tag{S12}$$

Note that this slope of 0.96 is approaching perfect agreement (1.00), indicating that

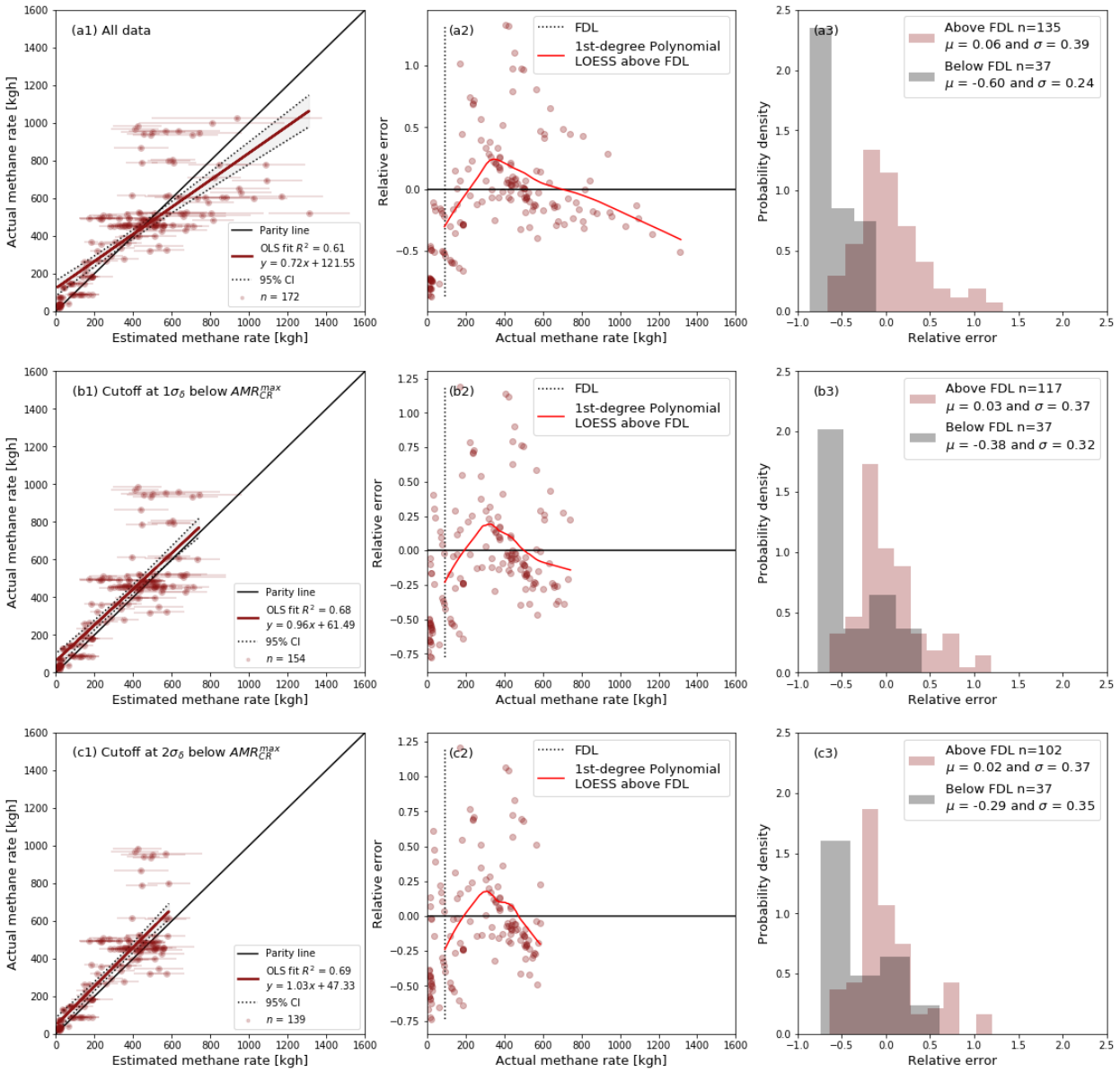

Figure S11: Linear regression based on data points vertically cutoff at different thresholds. The top panels present a case with no cut off. The middle panels present the base case of excluding data points that are within σ_δ below AMR_{CR}^{max}. The bottom panels show the other sensitivity case when data points within $2\sigma_\delta$ below AMR_{CR}^{max} were excluded. Note that the σ_δ values here are individually computed for each regression case (see the three panels on the right) and the σ_δ for data points above full detection threshold is applied.

boundary bias effects may be a major cause of the lower-than-parity slope in the fit using the full dataset.

The intercept in Equation S12 is 61.49 (±40.06), which is still larger than zero. This is expected due to the exclusion of undetected controlled releases (EMR_{CR}=0) in the fitting process.

S2 New Mexico Permian survey

S2.1 Survey

Kairos' technology is capable of surveying approximately 390 km^2 (150 sq. mi.) of land per day.[18] At this speed, Kairos was able to survey the entire population of oil and gas (O&G) assets multiple times in the New Mexico Permian Basin from October 2018 to January 2020. The campaign covered 35,923 km^2 (13,870 sq. mi.) and 26,292 active wells, or 91.2% of all active wells in the covered region.

Table S3 details of the numbers of wells, pipelines, compressor stations, storage tanks, gas processing plants, and other/ambiguous sources in the region. Note that there are no comprehensive datasets for storage tanks, compressor stations, and gathering lines, and thus the counts of these assets in Table S3 are not as reliable as the well and the gas processing plant counts. The compressor station count includes both some large compressor stations documented by the EIA and some smaller-scale compressors that were found with emissions during the campaign. Kairos do not provide a storage tank count, and the Stanford-determined tank counts is the number of storage tank sites found with emissions. See Section S4.2 for the asset type determination process.

Each covered facility could be seen on more than one flight because repeated measurements were taken on different days, and overlap of areas surveyed also occurred during different passes on the same day. Each facility was imaged an average of four times over the course of the campaign.

Table S2: Number of production assets grouped by well status.

Well status	$N_{regional\ total}$	$N_{covered}$	%Covered
Active	28,839	26,292	91.2%
Cancelled	11	6	54.5%
Completed	85	81	95.3%
Drilled	125	108	86.4%
Drilling	1	1	100.0%
Drilled and uncompleted	22	19	86.4%
Inactive	2,430	1,952	80.3%
Plugged and abandoned	598	528	88.3%
Permitted	1	1	100%
Temporarily abandoned	73	65	89.0%
Unknown*	650	630	96.9%
Total	32835	29683	90.4%

Note: * The well status "unknown" is a category in Enverus well dataset. Detected emissions in the New Mexico Permian from a site with an "unknown" well is defined as emissions from asset type "well site" (see Section S4.2).

Table S3: Number of covered assets by asset type.

Asset type	$N_{covered}^{Kairos-provided}$	$N_{covered}^{Stanford-decided}$
Well	29,683	29,683
Gas processing plant	39	39
Compressor station	187	263
Storage tank*	N/A	59
Pipeline	18,614	18,614
Other/ambiguous	382	53

Note: * Storage tank sites found with emissions during the campaign.

Over roughly the same period of time, the Environmental Defense Fund (EDF) initiated the Permian Methane Analysis Project (PermianMAP) to publish emission data collected from sensors on satellites, helicopters, aircraft, vehicles, and towers.[31] PermianMAP's study area straddles New Mexico and Texas and overlaps partially with Kairos' study area. Using the aircraft- and tower-based measurements in PermianMap along with inverse modeling, Lyon et al. estimated methane emissions for a 100 km $\times$ 100 km study area within the Permian Basin.[8] Note that the aerial measurements used in Lyon et al. are regional atmospheric boundary layer methane concentration for inverse modeling, and are different from

the hyperspectral imaging data of this study.

In addition, NASA's Jet Propulsion Lab (JPL) also flew over the Permian Basin for methane leakage detection with aerial surveys using the Airborne Visible-Infrared Imaging Spectrometer - Next Generation (AVIRIS-NG) and Global Airborne Observatory (GAO) systems.[12] Satellite systems such as the TROPOspheric Monitoring Instrument (TROPOMI), GHGSat, and the Greenhouse Gases Observing Satellite (GOSAT) were also able to monitor the same region for methane concentration changes.[32–34] Robertson et al. carried out ground campaigns to measure site-level emissions in the Permian Basin using OTM-33A method.[9] Table S4 compiles studies of Permian Basin methane emissions with other instruments.

Note that the temporal scope of these studies vary. Permian Basin natural gas and oil production follows a growth trend (Figure S12). Oil stays as the primary product in the region, with its product value far exceeding the value of gas. Regardless of the natural gas production growth, the total product value based on Henry Hub spot price stays relatively stationary.[4]

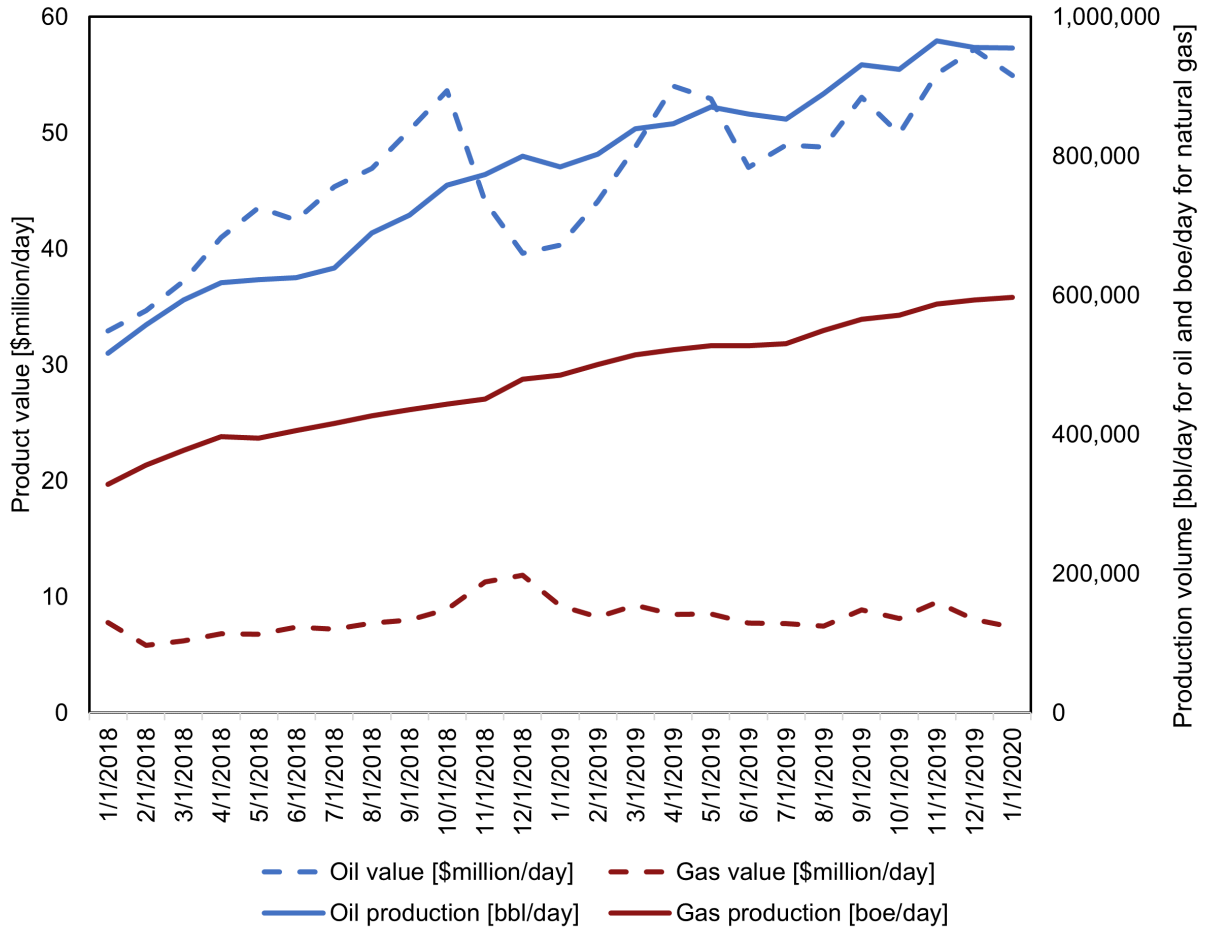

Figure S12: Oil value far exceeds gas value in the Kairos survey area, making oil the primary product in oil and gas (O&G) production. The product values of oil and gas are respectively calculated with monthly average WTI crude oil spot prices and Henry Hub natural gas spot prices.[4] O&G production in the Kairos survey area grew from January 2018 to January 2021.[35] Gas production is converted to barrel-of-oil equivalent (boe) per day.

Table S4: Permian Basin oil and gas methane emission studies.

Study	Spatial scope	Time of data collection	Measurements	NG production loss
Zhang et al.	Permian Basin	05/2018 - 03/2019	TROPOMI	3.7%(±0.7%, scenarios)[7]
Zhang et al.	Delaware sub-basin	05/2018 - 03/2019	TROPOMI	4.1%[7]
Zhang et al.	Midland sub-basin	05/2018 - 03/2019	TROPOMI	3.4%[7]
Schneising et al.	Permian Basin	01/2018 - 12/2019	TROPOMI	3.7%(±1.4%, 1σ)[10] [a]
Lyon et al.	100km×100km study area that straddles New Mexico and Texas in the Delaware sub-basin responsible for 40% of Permian Basin-wide production	01/2020 - 08/2020	Continuous methane concentration measurements from four tower-based monitors and six flights of aerial measurements of atmospheric boundary layer methane concentration along the study area perimeter	3.3% (+0.7%/-0.6%, 95% CI)[8]
Cusworth et al.	Permian Basin	09/2019 - 11/2019	AVIRIS-NG	N/A[12]
Robertson et al.[b]	Delaware sub-basin	08/2018	OTM 33A	0.89 (0.42–1.83)%[9]
Robertson et al.[b]	New Mexico part of the Permian Basin	08/2018	OTM 33A	1.03 (0.42–2.69)%[9]
Robertson et al.[b]	Texas part of the Delaware sub-basin	08/2018	OTM 33A	0.61 (0.22–1.30)%[9]

Notes: [a] Lyon et al. made monthly production loss estimates. The 3.3% (+0.7%/-0.6%, 95% CI) value presented here is their March 2020 estimate based on continuous tower-based monitoring and two flight days of aerial measurements in March 2020.

[b] Robertson et al. focused on well site emissions. Their emission estimates are for production emissions, and not midstream emissions.

The emissions data provided by Kairos provides unique empirical insights due to its extensive coverage. Surveying over 90% of production assets in a large region is challenging and costly, and is viable only through aerial surveys if the goal is to complete the effort within a reasonably short time frame.

Unlike conventional bottom-up studies which cover a small subset of all assets in the study area, our study is better characterized as a basin-wide study that covers >90% of production assets in the study region. With a much larger dataset, more super-emitter observations are available to us, enabling us to better quantify the emissions that follow a heavy-tailed distribution.

Kairos conducted the vast majority of the surveys from 9 am to 4 pm local time to ensure adequate illumination for hyperspectral remote sensing. On each flight day, Kairos flew to a predetermined polygon. The sizes of these polygons vary, but cover on the order of 400 km^2 imaged per day. These polygons can overlap with each other, resulting in revisits to some assets on different days.

Figure S13 shows the flight passes over a predetermined polygon. At a flight height of roughly 900 meters above ground, a single flight pass can image a strip of area along the flight line. After covering one stripe of the polygon, Kairos turns around the airplane and covers a partially overlapping area, with design overlap of 50%. Due to this coverage pattern, same-day revisits are common. If an asset is located at point A in Figure S13, the time between consecutive visits to point A is on the magnitude of one hour, depending on the length of the stripe. For example, an asset located at point B in Figure S13 may experience revisits that are minutes apart.

S2.2 Data anonymization measures

To comply with the confidential terms with their clients, Kairos anonymized the emissions data prior to sharing it with the Stanford team. Kairos provided 2076 measured methane plumes from 968 distinct emission sources. Each plume is associated with location infor-

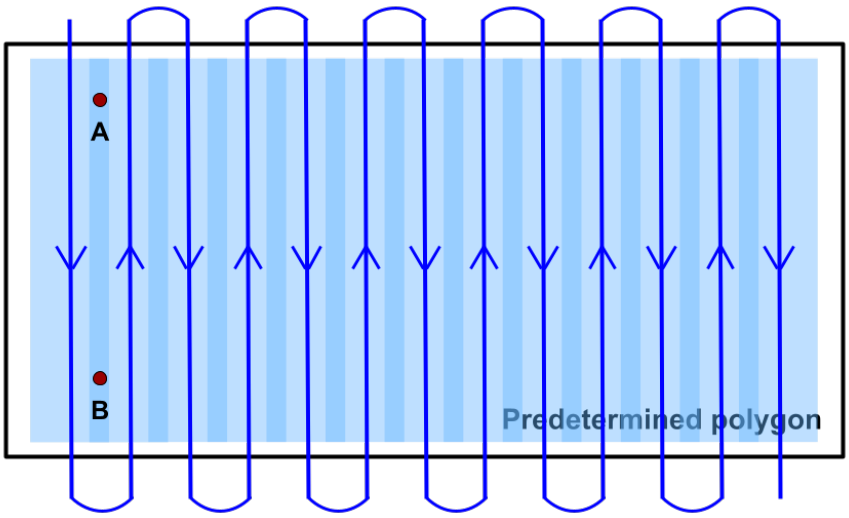

Figure S13: Kairos flight coverage pattern. The black rectangle indicates the predetermined polygon for a certain flight day. The lines with arrows are the flight lines. The blue semi-transparent rectangles are the stripes of survey area covered with the flights on the center lines of the rectangles. Strips covered by forward and backward flights may partially overlap, resulting in a darker shade of blue indicating double coverage in the figure. Point A and B are the locations of two production assets that are covered twice on a single day because they are within overlapping stripes covered by forward and backward overflights.

mation truncated to the second decimal place of longitude and latitude, as well as a flight overpass timestamp accurate to the second. 91 of the 2076 plumes are left out of the analysis due to either being designated as a "blowoff" of a larger plume found concurrently or due to having been found outside of New Mexico. See Section S2.3 for details of the plume screening process. The 1985 remaining plumes are associated with 958 emission sources.

Kairos associates plume detections to assets identified in optical aerial images and labeled asset data to allow identification of the sources of the plumes. More details on the association process are described in Section S4.2.

We retrieved well characteristics and production data from Enverus. To reduce the potential for de-anonymizing the Kairos data by creating one-to-one association of assets in Kairos dataset with assets documented by Enverus, we took the measures summarized in Table S5 when sharing data. Spatial coordinates of assets, detected plumes, and emission sources were truncated to two decimal places. Each distinct truncated latitude and longitude

pair is associated with 4.8 assets on average. Operator IDs in the Kairos dataset were replaced with pseudo IDs to mask operator information. Well characteristics, including depth, production type, well status, producing entity count, completion date, drill type, peak oil, peak gas, cumulative oil, cumulative gas, and cumulative water, were binned to lower precision levels so that each characteristic level could be associated with over 100 assets, with the sole exception that there are only 87 wells with a production type identified as "water" in the study area, according to Enverus. Monthly well production of oil, gas, and water are encoded with a reversible monotonic nonlinear transformation as an additional anonymization measure.

Table S5: Data shared with Stanford team by Kairos and measures of data anonymization

Data	Anonymization measure
Location	Longitudes and latitudes truncated to two decimal places
Operator ID	Pseudo-ID created to mask operator information
Well characteristics	Each characteristic was binned so that each bin has >100 wells*
Well production	Monotonic transformation applied to monthly production data

Note: * One exception is that there are only 87 wells with deemed production type "water" in the study area.

S2.3 Data screening

Blowoff plumes During the field campaign, when two or more plumes from the same emission source were recorded from the same flight pass, it was either because there were multiple emission sources (at the same site) emitting at the same time, or there was only one leak source but non-uniform dissipation had caused small "blowoff" plumes to detach from the main body of the plume.

In the first case, Kairos analysts treated the multiple emission sources as one, since the spatial precision of Kairos technology renders its emission attribution capacity to be the closest to site-level. There are also cases when plumes from multiple leak points at one site merged into one large plume. Under this circumstance, Kairos recorded one plume incidence

and performed quantification for the combined plume.

In the second case, blowoff plumes are excluded from the analysis because blowoff plumes are in most cases an order of magnitude smaller than the main plume in terms of wind-independent emission rate (ER^{WN}_{plume}). The inclusion of blowoff plumes in the quantification process would bias the estimate low because the small blowoff plumes would enter the plume selection process (see Section S5.1) for the corresponding emission sources as candidates and reduce the chance of selecting the larger, main-body plumes for these emission sources.

To exclude the blowoff plumes, we identify concurrently observed plumes associated with the same emission source and keep the largest plume in terms of ER^{WN}_{plume} in the dataset. We identify 36 blowoff plumes from the 2076 plume records provided by Kairos and exclude them from the analysis.

Plumes outside of the study area The plume records shared by Kairos contain 56 plumes located in Texas. Some of these plumes are on the Texas side of the New Mexico/Texas border and the rest are close to the Midland Airport, where Kairos airplanes were based during the campaign. The locations of these plumes are shown in Figure S14. These plumes are excluded from the analysis, leaving the final 1985 plumes from 958 distinct emission sources for analysis.

S3 Emission intermittency

Our method estimates total emissions from the New Mexico Permian Basin based on measurements from an aerial campaign conducted over more than a year. As shown in Figure 1c in the main text, some assets were covered more times than others. In many cases, emissions from a given emission source are intermittent, with the airplane detecting emissions at some times and not others. Thus, our statistical approach to estimating basin-wide emissions must produce an unbiased estimate from a dataset with different levels of coverage for different assets.

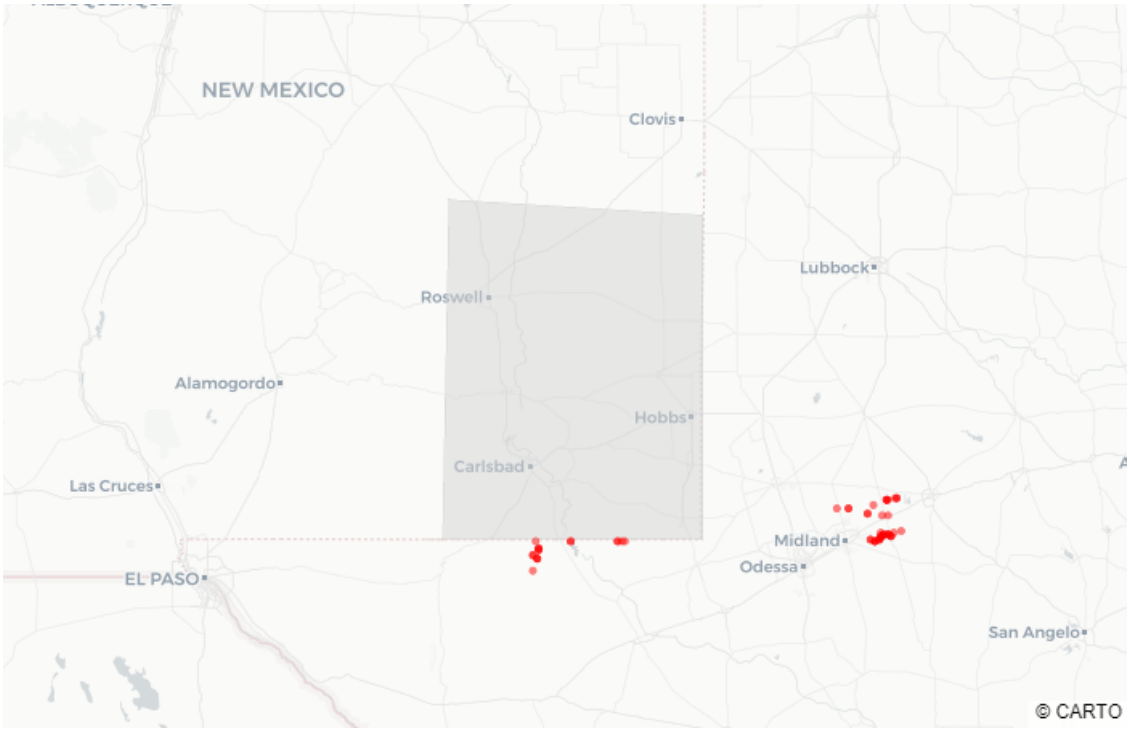

Figure S14: Plumes outside of the study area.

Fortunately, the sampling approach taken by Kairos during the aerial survey allows us to produce such an unbiased estimate even with an uneven coverage profile. The mathematical framework is described below.

S3.1 Treatment of intermittency

In Section S5.1, we define the number of plumes associated with emission source i as J_i and the number of coverages for the emission source as C_i. The vast majority of the emission sources have multiple coverages ($C_i > 1$), and some have multiple plume observations ($J_i > 1$) during different coverages. We define $P_i = \frac{J_i}{C_i}$ as the probability of emission detection at emission source i. Whether emission source i is active at any given time thus follows a Bernoulli distribution of parameter P_i. We then model intermittency explicitly, using P_i as the probability with which a given emission source will emit in a given Monte Carlo iteration (choosing at random from plumes identified at that emission source in instances in which that source is emitting).

As discussed in Section S2.1, time intervals between revisits vary from minutes to months. We treat each visit as independent regardless of the varying time period since the previous visit.

S3.2 Notation and key concepts

The following illustrates why the above approach produces an unbiased estimate of basin-wide emissions.

Suppose assets are divided into three categories (subsets):

- X = Always-leakers
- Y = Intermittent leakers
- Z = Never-leakers

with N_x, N_y, and N_z being the number of assets in each category in the study area.

Then,

$$N_x + N_y + N_z = N \tag{S13}$$

where N is the total number of assets in the study area. Note that $N_x + N_y$ is roughly equal to the number of emission sources 958, with a few edge cases of long pipeline assets possibly having multiple emission sources.

Let $x \in X, y \in Y, z \in Z$ be assets within X, Y, Z. Let assets be indexed by i. By definition of P_i in Equation S19:

- $P_x = 1 \; \forall x \in X$
- $0 < P_y < 1 \; \forall y \in Y$
- $P_z = 0 \; \forall z \in Z$

Suppose the $y_i \in Y$ emits natural gas at a given time with probability $P(y_i|\ell_i = 1) \in (0, 1)$ with some emission flux $\rho(y_i)$, where ℓ_i is a binary variable indicating whether asset i is an emitter (whether or not it is emitting at a detectable level at a given point in time).

S3.3 Characterizing cases

Assume some subset of all assets in a basin are covered by an emissions detection and quantification technology. As long as the asset selection is plausibly random, quantified emissions can be scaled up to basin-wide emissions to produce an unbiased estimate. As over 90% of assets were covered in our study, assumptions surrounding statistical representativeness of the remaining 10% of assets have a relatively small effect on estimated basin-wide emissions.

S3.3.1 A single snapshot

Suppose all assets in the population are covered once , $C_i = 1 \; \forall i$. We call this a "snapshot". Assuming the intermittency profile is itself stationary over time. i.e., that intermittent sources are equally likely to be emitting at any time of day or hour of the year, and that the emissions profile does not change over the duration of the measurement campaign, this is functionally equivalent to sampling all assets simultaneously.

No intermittency If there is no intermittency (Y=$\emptyset$), then all observed plumes in this snapshot are always-leakers and the total emissions from the snapshot are thus an unbiased estimate of total emissions from the study area.

Pervasive intermittency If all leaking assets observed in the snapshot are intermittent, then the snapshot is also trivially an unbiased estimate of total emissions over time (again, assuming the intermittency profile is stationary over time. The mean of random sample from a population is an unbiased estimate of the mean of the population).

Partial intermittency If some fraction of leaking assets observed are intermittent and some are persistent, the snapshot is also an unbiased estimate of persistent emissions. If we remove these emissions from the analysis, the problem then becomes the case of pervasive intermittency for the remaining emitting assets. Thus, a single snapshot over the full study area results in an unbiased estimate of total emissions regardless of the intermittency profile of the emissions sources (unless the profile varies over time).

S3.3.2 Multiple snapshots

The above analysis generalizes trivially to the case of multiple snapshots ($C_i = k > 1 \ \forall i$). Each individual snapshot yields an unbiased estimate of total emissions from the $C_i = 1 \ \forall i$ case and the average of unbiased estimates of a number is itself an unbiased estimate of that number.

Thus, if C_i is the same for all assets, we can account for intermittency by taking the average of emissions in each snapshot across the entire population and averaging across those snapshots. For the $C_i = 1 \ \forall i$ case, this is the same as assuming all emissions observed are persistent.

For $C_i > 1 \ \forall i$, one way to compute overall emissions is by first averaging across each individual snapshot separately and then averaging over total emissions from all snapshots.

If the sample size, C_i, is the same for all assets, then we will get the same answer by first averaging over all emissions estimates for each asset i, and then summing these emissions over all assets.

S3.3.3 Uneven coverage counts

Suppose the survey sampled some assets more than others. In this case, the problem reduces to a series of cases of multiple snapshots with $C_i = F_j \ \forall i \in F_j$, where F_j is the set of emission sources, i, which have coverage count $j = 1, 2, 3, ...$

By the logic above, the sum of emissions over all assets in F_j produces an unbiased

estimate of emissions in F_j. As a result, the sum of emissions from all F_j will produce an unbiased estimate of total emissions, because the union of the F_j is equal to the surveyed population of assets.

As a result, the sum of average emissions estimated for each emission source is an unbiased estimate of basin-wide emissions even in a case with oversampling of some assets.

The above logic only holds if the number of times an asset is covered **does not** depend on whether emissions are detected there in previous coverages. In our survey, the presence or absence of emissions in prior overflights did not affect decisions related to future coverage. As a result, our analysis should produce an unbiased estimate of total basin-wide emissions over the covered areas.

To qualitatively illustrate this, if there are intermittent emission sources in a basin, a given overflight may or may not find an emission. For every emission source it finds, there is a corresponding number of intermittent sources a survey will miss in a given set of overflights. The proportion of missed intermittent emission sources will depend on the intermittency profile of the assets in question. Similarly, for every intermittent source that a given set of overflights misses, it will find others. The result is an unbiased estimate of average emissions across the survey area.

S3.4 Survey time

The above analysis assumes no variation in seasonal and diurnal emissions profiles, which may not be the case in practice. Kairos surveys were primarily conducted between 9am to 4pm (Figure S15), and more flight days are possible in summer than in winter because the summer sun angle means that more hours in the day are available for the campaign. Thus, it is possible that diurnal or seasonal variation in emissions affected our results. Figure S15 illustrates our results by time of day. The average time of point-asset coverage is 11:18am and the overall fraction of point-asset emitting is 1.42±0.07%, suggesting a 1.42% chance that a point asset is emitting at a given overpass.

We define the fraction of point assets emitting at the time of inspection as

$$Fraction\ of\ point\ assets\ emitting\ = \frac{Point\ asset\ plume\ count}{Point\ asset\ coverage\ count} \tag{S14}$$

The fraction of point assets emitting value of 9-10 am and that of 1-2 pm shown in Figure S15b are respectively 1.79±0.20% and 1.25±0.17%, indicating that a set of surveys done exclusively between 9 am and 10 am would see 26% more emissions than this study, and a study comprised only of surveys done between 1 pm and 2 pm would see 12% fewer emissions than that of this study on an incidence basis. The time-dependence of emissions may introduce bias to emission estimates based on measurements from satellites with constant overhead times.

Without emission data from the night hours, we assume that the emission rates are the same during the daytime hours during which data were collected. In addition, as shown in Figure S16, the fraction of point assets emitting varies by month but does not show a strong seasonal trend. We therefore assume a stationary emissions profile over time of day and season.

S3.5 Shrinking persistence with more frequent sampling in time

Persistence is defined as being equivalent to detection probability P_y in this study. The average persistence of all 958 emission sources is 0.51 (+0.49/-0.34). Note that this average persistence value is defined specifically for the 958 emission sources. If Kairos sampled more frequently in time, they would observe more less-frequent emitters and include them in the pool of emission sources. The newly-added emission sources will typically have persistence values lower than the current average persistence of 0.51. In other words, the mean of P_y will decrease because the newly entered $y \in Y$ have lower-than-average emission probabilities, as shown in Figure S17. Figure S18 shows average persistence as a function of coverage count and also demonstrates shrinking persistence with more frequent sampling in time.

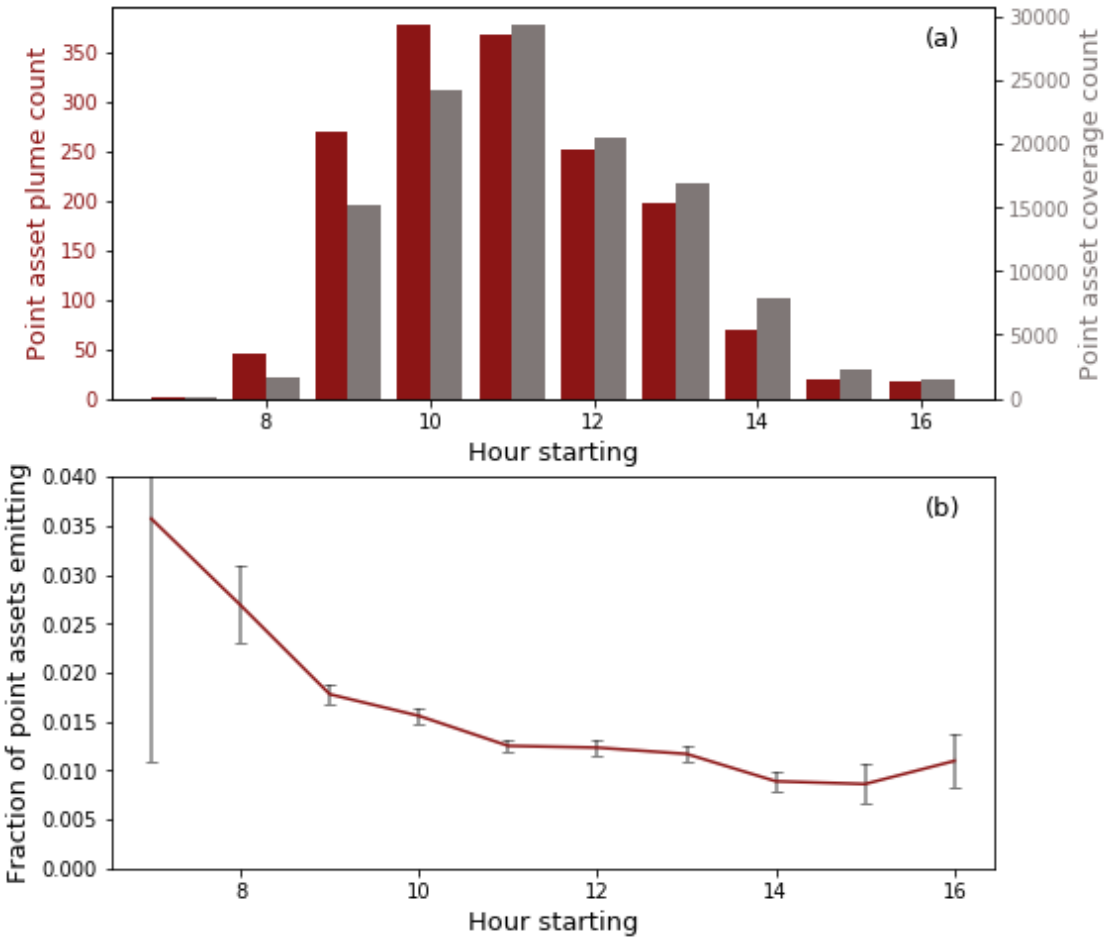

Figure S15: (a) Histogram of asset coverage records and plume detection records distributed in each hour. Note that pipeline coverage and emission records are excluded in this plot due to the low confidence in pipeline coverage counts. There exist seven point asset coverages and no plume recorded after 5pm. We do not show these seven data points considering the data scarcity of this hour. (b) Fraction of point assets emitting of each hour. The error bars show 95% confidence intervals.

In an ideal world where constant monitoring instruments with equivalent sensitivity levels are installed for all production assets in the study area, the persistence value would be significantly lower than 0.51 due to the chances of observing less-frequent emitters.

When quantifying basin-total emissions, the decrease in P_y will be offset by the increase in N_y, and the quantification remains unbiased regardless of the sampling frequency.

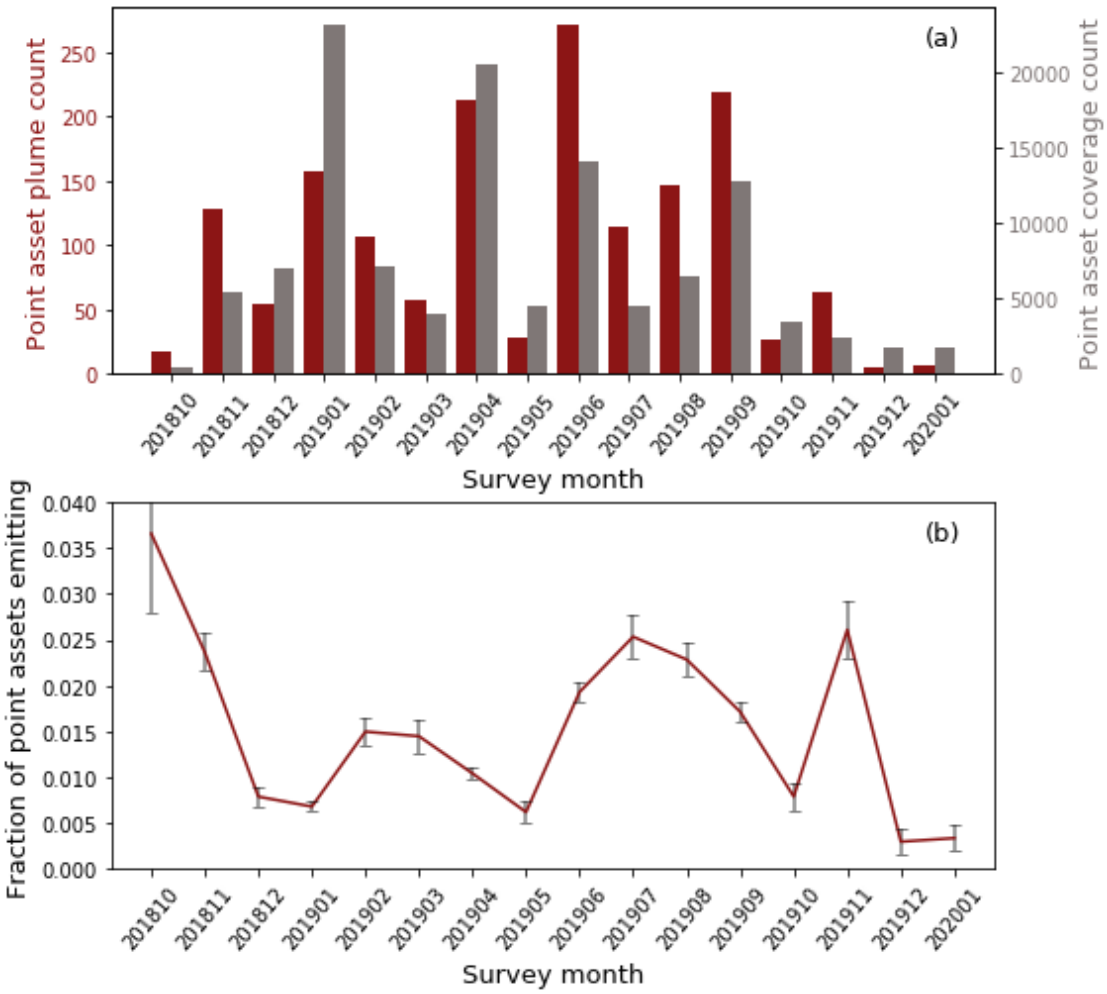

Figure S16: (a) Histogram of asset coverage records and plume detection records distributed in each survey month. Note that pipeline coverage and emission records are excluded in this plot due to the low confidence in pipeline coverage counts. (b) Monthly variation in fraction of point assets emitting of each month. The error bars show 95% confidence intervals.

S3.6 An unbiased estimate of total emissions

Thus, our method gives an unbiased estimate of overall emissions in our study area for emissions larger than the instrument's 100% detection threshold under the following assumptions:

1. The decision to collect more samples or discontinue data collection at any given site is not dependent on whether emissions were previously detected or not detected at that site.

2. The assets surveyed are representative of the assets in the full survey area.

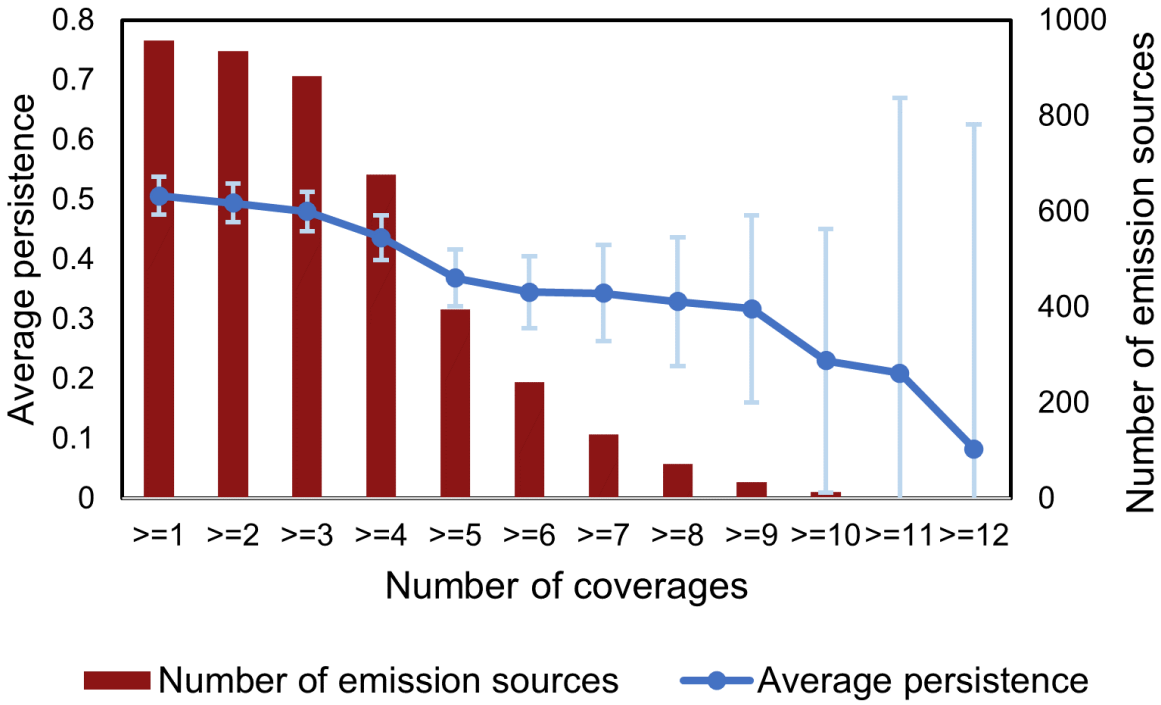

Figure S17: Average persistence decreases with exclusion of emission sources with low coverage counts.

3. The distribution of emissions is stationary over time of day and season.

Condition 1 holds. Kairos' sampling strategy was not informed by whether or not emissions had been seen in a given area.

Given that over 90% of all production assets were surveyed, any potential violations of Condition 2 are unlikely to have a large effect on our results.

For Condition 3, the emissions distribution does change somewhat over time and space. As a result, our estimate likely holds for emissions during the survey time frame of roughly 10am-4pm. Further research is needed to more closely characterize the distribution of emissions at night and in the early morning and evening, as well as seasonal variation.

Note that our method gives an unbiased estimate of overall emissions under the three assumptions above. For a single intermittent emission source, it is very likely that its temporal variation follows a heavy-tailed distribution, in which case as discussed in the main text, a small number of observations cannot fully characterize its variation in time, thus potentially

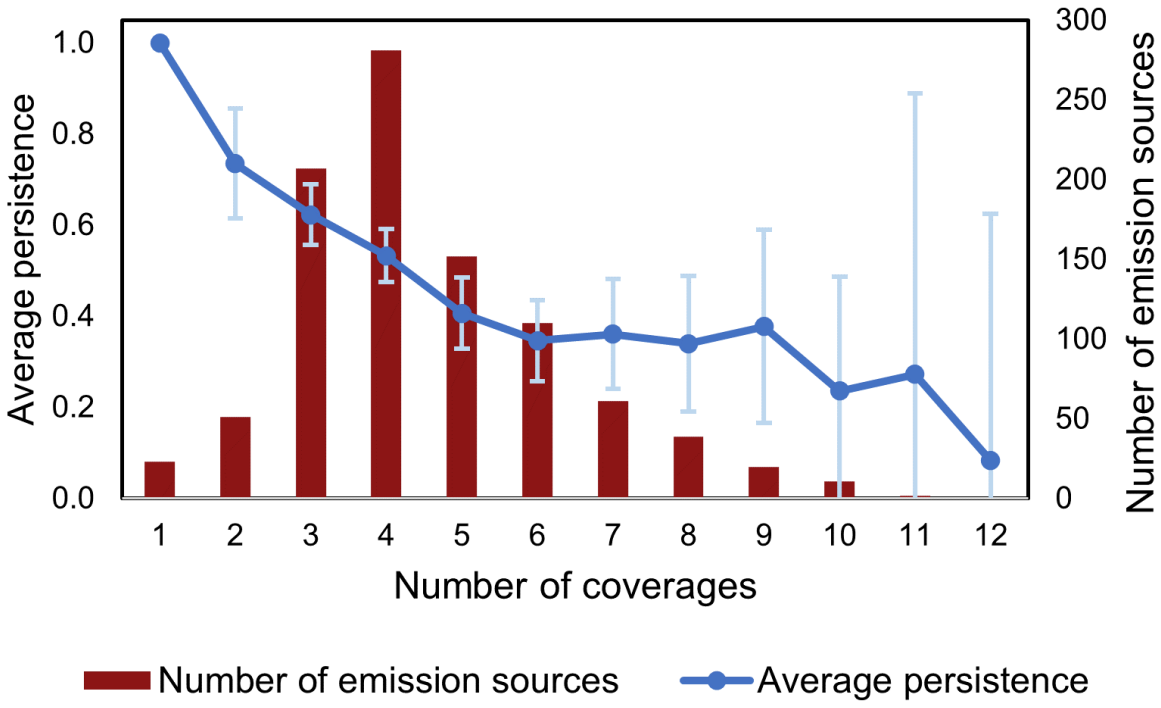

Figure S18: Average persistence decreases as number of coverages increases.

biasing the estimate for that single source. For basin-wide emission estimates, however, we take advantage of the ergodicity demonstrated in the dataset (assuming the distribution is stationary over time). A dataset is termed "ergodic" if it is representative of the population from which it was collected, which is likely the case for a dataset such as ours, which samples 90% of the population.

Thus, the distribution of methane emissions from a single source over time is assumed to be the same as the distribution of emissions from a large number of emission sources measured at one time (or in this study's case, four times on average). Therefore, the ensemble result is an unbiased estimate of average emissions across the survey area during the daytime hours suitable for aerial observations.

S4 Emission attribution

S4.1 Asset types

In Figure 3, we present emissions from six asset types: well site, gas processing plant, compressor station, storage tank, pipeline, and other/ambiguous. In this paper, these asset types are defined as:

- *Well site* is defined as the ensemble of all assets found on a congruent gravel or concrete ground with at least one well, which can include gathering lines, storage tanks, and compressor stations.
- *Gas processing plant* is defined as a site with gas processing plants and supporting facilities on site such as gathering lines, storage tanks, and compressor stations.
- *Compressor station* is a site with at least one compressor and does not contain wells or gas processing plants, but may contain gathering lines and storage tanks.
- *Storage tank* refers to a site that may only contain storage tanks and gathering lines.
- *Pipeline* emissions must come from a segment of the pipeline that is at least 200 meters (typical well pad diameter) away from any well sites, gas processing plant sites, compressor station sites, and storage tank sites.
- *Other/ambiguous* sources are associated with emissions that cannot be clearly attributed to above-mentioned asset types (see Section S4.2).

We aggregate the assets at the site-level due to the challenge of sub-site emission attribution with aerial surveys, although in cases when assets are situated far apart from each other on a site, component-level attribution is possible from imagery. Many such cases suggest that well site emissions are from gathering lines, compressor stations, and storage tanks, rather than directly from the well itself. Figure 1a, for example, is a site that contains

compressors and storage tanks and is thus defined as a compressor station site in this paper. Although the methane plume most likely arises from the storage tanks in the optical image, we characterize this emission incidence as a compressor station emission.

The site-level data are actionable for the operators who subscribe to Kairos' service to find and fix leaks. In some cases, companies derive significant revenues from fixing identified leaks. Triple Crown Resources, one of Kairos' clients, published a preprint that states the deployment of Kairos aerial surveys for their assets resulted in a payout period of 18.7 days and a cumulative cash-on-cash profit over $90,000 over 4 months.[36]

Figure 3 displays emissions by asset type. In Figure 3a, 79±46 of the 153 t/h of measured emissions ($\hat{E}^{Measured}$) comes from well sites; 4±2 t/h are from gas processing plants; 26±16 t/h are from compressor stations; 9±6 are from stand-alone storage tanks; 29±20 are from pipelines, and the rest 7±4 are from other/ambiguous assets. See section S4.2 for the asset attribution method.

Although even the smallest methane plume detected by the airborne sensor (13 kg/h after calibration) could be considered a super-emitter in a typical ground-based study (Section S6), the emissions detected in this study follow a heavy-tailed distribution, as shown in Figure 3b. Regardless of the asset type of the emission source, the size distribution of the detected super-emitters is as heavy or heavier than a log-normal distribution. Overall, the top 5% of the observed sources account for 32% of $\hat{E}^{Measured}$.

Table S6: Number of emission sources and emissions by asset type. The emissions reported here are measured source-averaged emission rates accounting for intermittency ($\hat{E}^{Measured}$). N_{ES} denotes number of emission sources.

Asset type	N_{ES}	$\%N_{ES}$	Total emissions (t/h)	Emission contribution
Well site	543	57%	79	52%
Gas processing plant	15	2%	4	2%
Compressor station	113	12%	26	17%
Storage tank	59	6%	9	6%
Pipeline	175	18%	29	19%
Other/ambiguous	53	6%	7	4%
Total	958	100%	153	100%

S4.2 Plume-asset association

Associating plumes to emitting assets is a two-step process. Multiple inspections of an emitting asset can record multiple plumes from that asset. Therefore, the first step of the association process is to find all plumes that appear to be from the same source and generate an emission source for that group of plumes. Regularly during the campaign, Kairos compiles plume observations and uses a clustering algorithm to group together plumes that are likely to be from the same emission source.

The second step is to associate the emission sources to emitting assets. This is done internally by Kairos, using simultaneously-collected optical imagery to identify the asset from which a detected plume is emitting. With the rapid development of O&G infrastructure throughout the basin, having this up-to-date optical image is critical for determining the source of a detected plume. Assets in the optical imagery are labeled using data from external O&G asset datasets ingested from Enverus and IHS Markit. These datasets are updated regularly and document the locations of wells and pipelines. Kairos also obtains compressor station and gas processing plant location data from Energy Information Administration (EIA).

Of the 1985 eligible plumes, the Kairos-internal association process (henceforth "Kairos association") is able to associate 1184 plumes to 576 assets in the datasets mentioned above. The remaining 801 plumes are reported to be from 382 distinct "unknown" sources that cannot be associated to the documented assets. The association failure can occur under the following circumstances:

1. Newly built assets are not included in the datasets mentioned above at the time of the association process.

2. Datasets mentioned above do not contain certain assets, such as small-scale compressors, small gathering stations, and stand-alone storage facilities.

3. The plume is roughly equally close to two distinct assets, making the association am-

biguous.

To address the unknown plume-asset associations, the Stanford authors conducted an independent manual association (henceforth "Stanford association") for the 382 "unknown" emission sources with the most up-to-date asset data.

The Stanford association process is presented in Figure S19. The stopping criteria of the Stanford association process is "unique asset type," meaning that if the plume can be associated with a group of assets of the same asset type, then the association stops, because the main goal of this association exercise is to attribute emissions to a certain asset type, not to an exact asset.

As shown in Figure S19, we associate emissions to the close-by assets based on the following:

- Proximity: Assets within 200 meters of the plume cluster center are candidates for association. We also tested 100 meters and 500 meters as the radius, finding that 200 meters presents the best balance between plume migration potential (many plumes can have lengths exceeding 100 meters) and attribution efficacy (too many assets entering as candidates).

- Wind direction: Dark Sky wind direction is extracted by Kairos for each detection. If the proximity test fails, then we use the wind direction test to screen out assets that are not within a 120° cone upwind of the plume cluster center.

- Prioritization of point sources over pipelines: Pipelines are connected to point sources such as wells and compressor stations. If a pipeline asset is contiguous with, for example, a compressor station, and both assets are potentially responsible for the methane plume observed, then we chose compressor station as the source because it is a more important feature of the site compared to the pipeline.

This Stanford association process for the 382 unknown emission sources shows that circumstance 1 is the most common association failure and can often be resolved with the

Figure S19: Flow chart of Stanford association of emission sources to assets.

updated asset dataset. However, there are 53 emission sources without any candidate asset in the 1) up-to-date asset dataset, 2) optical imagery taken by Kairos simultaneous to plume detections, or 3) satellite images. The associations for these sources therefore remain as unknown. Possible explanations for these remaining unknown emissions include false positives, leaks from underground structures, and plumes from equipment not associated with gas production. Of these 53 unknown emission sources, 9 are from visible structures that appear unrelated to gas operations and 33 are not associated with any visible structures. Note that these unassociated plumes have small emission rates on the lower extreme of the emission size distribution of all observations, and therefore have very little impact on the quantification of the regional total. The final 11 emission sources are "ambiguous" sources with multiple asset types possibly contributing to the associated plumes detected.

Total persistence-averaged emissions from these other/ambiguous-typed emission sources are respectively 1.0, 3.5, and 2.0 t/h for other, unassociated, and ambiguous sources, or 0.6%, 2.3%, and 1.3% of 153 t/h of $\hat{E}^{Measured}$.

Table S7: Number of emission sources by asset type.

Asset type	Kairos-determined	Stanford-determined
Well site	409	543
Gas processing plant	15	15
Compressor station	37	113
Storage tank	N/A	59
Pipeline	115	175
Other/ambiguous	382	53

When circumstance 2 occurs, we use visual inspection to select the asset associated with the plume following the process shown in Figure S19. The vast majority of these undocumented assets are determined to be small gathering and compressor stations, gathering lines, or storage tanks. These types of infrastructure are in most cases sited in close proximity to producing wells. As described in Section S4.1, the set of possible asset types in the Stanford association is well site, gas processing plant, compressor station, storage tank, pipeline,

other/ambiguous.

When circumstance 3 occurs, and if all possible responsible assets fall in the same asset type (e.g., plume between two well sites), we randomly select one of the assets to be the emitting asset, as this will not affect our emission quantification or attribution. If the possible responsible assets have more than one asset type (e.g., compressor station site and well site), then we keep the association as "ambiguous". See Figure S19 for the plume-asset association criteria and process.

After the Stanford association of the 382 "unknown" sources, 134 more sources are categorized as well sites, 76 more as compressor stations, 59 more as storage tanks, 60 more as pipelines, and 53 emission sources remain "other/ambiguous" sources.

For the 576 emission sources with Kairos-provided association to known asset types, the Stanford team randomly select 50 of them (responsible for 106 methane plumes detected) and conduct the Stanford association process to verify source attribution. Our independent verification confirms that 1) the clustering algorithm is valid, and 2) Kairos associations are in general agreement with Stanford associations, with the exceptions that two Kairos-determined pipeline emissions are categorized as well site emissions in Stanford association process. These two pipeline segments are sited on the edges of the two well sites and are closer to the methane plumes than the wells on the well sites.

S5 Basin-wide emission quantification method

Plumes detected by Kairos are analyzed and aggregated using the diagram shown in Figure 1e in the main text (reproduced as Figure S20) to determine the total emissions in the survey area. The following sections describe each node in the diagram in detail.

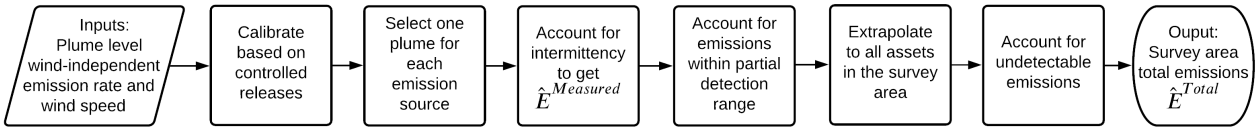

Figure S20: Figure 2 reproduced. Process flow for study area total upstream and midstream oil and gas methane emissions quantification.

S5.1 Quantification method

S5.1.1 Inputs

The Kairos plume dataset contains 1985 eligible plumes from 958 distinct emission sources. For each plume, Kairos reports a wind-independent emission rate (ER^{WN}_{plume}) in kgh/mps. The plume-level emission rate was then computed with Equation S15, where ER_{plume} is the methane emission rate of each plume, and $Wind^{HRRR}_{plume}$ is the HRRR wind speed in mps at the time and place of the plume imaging.

$$ER_{plume} = ER^{WN}_{plume} \times Wind^{HRRR}_{plume} \tag{S15}$$

In absence of hyper-localized wind measurements at each methane plume detected in New Mexico, we use HRRR surface wind ($Wind^{HRRR}_{plume}$). HRRR provides hourly weather data of 3 km $\times$3 km spatial resolution.[37]

We pull $Wind^{HRRR}_{plume}$ values at plume source coordinates and at plume observation timestamps. The HRRR wind used is the average value from three time steps ($\pm$1 hour) and nine surrounding locations (3$\times$3 grid points of HRRR). We select the 27-average HRRR wind (3$\times$3 grid $\times$ 3 time steps) as $Wind^{HRRR}_{plume}$ because this is an established method used in quantifying super-emitters using data from AVIRIS-NG.[23] Note that the plume coordinates were truncated to two decimal places for anonymization purposes. However, in New Mexico, the ground distance between two locations of 0.01 degree difference in either longitude and latitude is close to 1 km, which indicates that the truncation will not affect the retrieval accuracy of wind data at the HRRR spatial resolution of 3 km.

Although the relevant timescale for wind variations in plume formation is on the order of tens of minutes or less (see[22] supplementary information for more discussion on "full plume development"), we select the wind averaged over three hours due to the challenge of accurately modeling sub-hourly level wind speed with reanalysis products such as HRRR.

In contrast, the private company Dark Sky provides surface wind speed ($Wind_{plume}^{Dark\ Sky}$) at a much finer temporal and spatial resolution than HRRR. More importantly,[22] shows that the one-minute Dark Sky wind gust speed gives rise to an R^2 value closer to 1 than HRRR 27-average wind does. However, as noted above, due to Dark Sky's expiring availability and nontransparent model details, we choose HRRR as the default wind data source and presented a sensitivity case using Dark Sky data in Table 1 and Section S7.

Note that this wind speed may be an overestimate of the relevant winds because $Wind_{plume}^{HRRR}$ in Equation S15 is wind speed measured at 10 meters above ground. The height of methane leak sources can vary, from 1-4 meters for pumpjacks or compressors, to >8 meters for a standard 15,000 bbl oilfield tank. Given that wind speed diminishes closer to the ground surface due to friction,[26] we may expect the 10-meter HRRR wind estimate to be on the high side (resulting in high CH_4 emissions estimates) before calibration. However, because we calibrate the proportionality between estimated and actual releases based on the same dataset, these effects should be "built in" to the resulting calibration.

S5.1.2 Calibration based on controlled releases

To adjust for the bias and uncertainty in ER_{plume}, both from wind speed deviation as well as other systematic problems, we use Equation S16 to compute a calibrated estimate for emission rate ($\widehat{ER}_{plume}$).

$$\widehat{ER}_{plume} = \alpha(ER_{plume})^{\beta} \tag{S16}$$

To account for the uncertainties in the fit, we apply an error term to the calibrated emission rate ($\widehat{ER}_{plume}$) using Equation S17, noting that that error is observed to grow

linearly with the magnitude of ER_{plume} as shown in Figure S16. We discuss the fitting process and the residuals in Section S1.5.

$$\widetilde{ER}_{plume} = (1+\delta)\widehat{ER}_{plume} \tag{S17}$$

δ in Equation S17 is defined in Equation S6. δ is essentially a relative error term and is modeled to follow a normal distribution. See Equation S8 and Figure S9 for the distribution of δ.

For the base case, we find fitted values of α=4.08 and β=0.77 from Equation S7 and estimate from the residuals that δ follows a normal distribution with mean 0.07 and standard deviation 0.4 (Equation S8).

Using a Monte Carlo (MC) approach, we draw from the δ distribution 1000 times to evaluate the resulting uncertainties in $\widetilde{ER}_{plume}$. This uncertainty is carried over in subsequent steps.

S5.1.3 Plume selection for each emission source

Because of the Kairos survey pattern described in Section S2, an emission source leaking at different times during repeated visits can have more than one methane plume observation associated with it. Note that an emission source is defined as any point in an asset that is determined to be responsible for emissions. A long pipeline, for example, can potentially have multiple emission sources.

We define ES_i as the set of all plumes associated with emission source i. For each emission source i, the number of plumes associated is J_i and the associated plumes are denoted as $plume_{i,j} \in ES_i$.

If there are multiple plume records associated with a given emission source i ($J_i > 1$), we randomly draw one plume from the set ES_i for each MC realization and use the emission rate of the selected plume as that of the emission source i. Thus, each plume in the set has equal probability $1/J_i$ of being selected. Note that we do this before accounting for

intermittency of each emission source.

$$ER_{ES_i} \xleftarrow{uniformly\ random\ draw} \widetilde{ER}_{plume_{i,j}},\ where\ plume_{i,j} \in ES_i \tag{S18}$$

We follow an MC approach and repeat the random selection process 1000 times for each emission source to quantify the uncertainties of emissions. The selected plumes in the MC realizations have calibrated emission rates applied with a normally distributed relative error ($\widetilde{ER}_{plume}$).

S5.1.4 Accounting for intermittency

If all emissions were persistent, then once an emitting site was first identified and the emission rates stays above the detection limit, Kairos would always detect plumes at this site in subsequent visits. However, this is not true in the context of methane emissions from the O&G sector for several reasons: 1) some leaks can be detected and fixed, 2) leaks can be sporadic due to intermittent events such as separator dump events, and 3) intentional venting due to emergency conditions is usually of short duration.

Hence, Kairos found that most emission sources are intermittent, emitting at some times but not others. The probability of emissions detection for emission source i is computed with Equation S19, where J_i denotes the number of plumes found at emission source i, and C_i is the number of flight passes over emission source i.

$$P_i = \frac{J_i}{C_i} \tag{S19}$$

At each instant, whether emission source i is actively emitting is assumed to follow a Bernoulli distribution of probability P_i. The emission rate of emission source i after accounting for intermittency ($ER_{ES_i}^{ITMT}$) is thus:

$$ER_{ES_i}^{ITMT} = X_i ER_{ES_i},\ where\ X_i \sim Bernoulli(P_i) \tag{S20}$$

Again, X_i is drawn 1000 times for each emission source i following an MC approach. Uncertainties from previous steps are carried over in this MC step. Note that the previous step, plume selection for each emission source, does not take into account the intermittency information. For example, if an emission source was covered 10 times with three detected emissions, the plume selection step would randomly choose one of the three emissions each iteration. The intermittency step would choose whether that emission source was emitting (probability 3/10) or not (probability 7/10) in that iteration.

The sum of the emission rates of the intermittent emission sources ($\sum_i ER_{ES_i}^{ITMT}$) is an unbiased estimate of the total measured emissions from a snapshot of all covered assets for reasons detailed in Section S3. In the main text, the total measured emissions is defined as:

$$\hat{E}^{Measured} = \sum_i ER_{ES_i}^{ITMT} \tag{S21}$$

The probability of emission detection at emission source i (P_i) can also be interpreted as the emission persistence of emission source i.

S5.1.5 Extrapolation to partial detection range

Figure S6 shows that within the PDR of 2 and 30 kgh/mps, plumes have a less-than-one probability of being detected by Kairos. For example, if the detection probability within (10, 12] kgh/mps is 50%, then for each plume detected in this range in the survey, we would expect that there is one other similar-sized plume that is missed. The detection probability grows with the wind-independent emission rate. Figure S6 presents the detection probabilities for emissions within the PDR using a bin size of 2 kgh/mps.

The values of $ER_{ES_i}^{ITMT}$ from the previous step vary in the 1000 MC realizations and hence the number of plumes within the PDR varies. For each MC realization, we identified the undetected emissions in the PDR with a bootstrapping approach:

1. Find $ER_{ES_i}^{ITMT} \in$ (2,30] kgh/mps $\forall i$ and put them into 2 kgh/mps-sized bins indexed

by k.

2. For each bin k, denote the number of plumes detected within that partial detection bin as N_k^D and the probability of detection of that bin as P_k^D. See Figure S6 for the values of P_k^D found in the controlled release trials.

3. The number of plumes undetected in that bin is thus $N_k^{UD} = \frac{(1-P_k^D)N_k^D}{P_k^D}$.

4. For each bin k, sample with replacement N_k^{UD} times from the N_k^D detected plumes of known emission sizes in that partial detection bin.

5. The undetected emissions in the partial detection bin k is the sum of sampled emissions. Denote this sum as ER_k^{UD}.

6. Sum up the undetected emissions of all partial detection bins to derive the total emissions in the PDR that are missing from the Kairos observations.

Total emissions after extrapolation to the PDR is then as follows:

$$ER^{>MDL} = \sum_i ER_{ES_i}^{ITMT} + \sum_k ER_k^{UD} \tag{S22}$$

S5.1.6 Extrapolation to all assets in the region

In the absence of whole-population data for pipelines and compressor stations, well assets are used to estimate the extent of coverage in the survey region. The Enverus dataset has 32,835 wells in the region. Kairos' survey covers 29,683 of the wells, or 90.4% of the total.

Figure S21 shows the locations and the types of covered and uncovered wells. Many of the uncovered wells are in the Northern part of the study area. Despite the differences in spatial distribution, the distributions of the types of wells covered and uncovered are similar.

For this reason, the uncovered wells were assumed to have the same leak characteristics and magnitude as the covered assets. Total emissions were thus multiplied by a factor of 32,835/29,683 = 1.106 to spatially extrapolate to all assets in the study region. By doing

this, we effectively assume that midstream emissions scale linearly as a function of upstream well counts.

(a) 29,683 covered wells (b) 3,152 uncovered wells

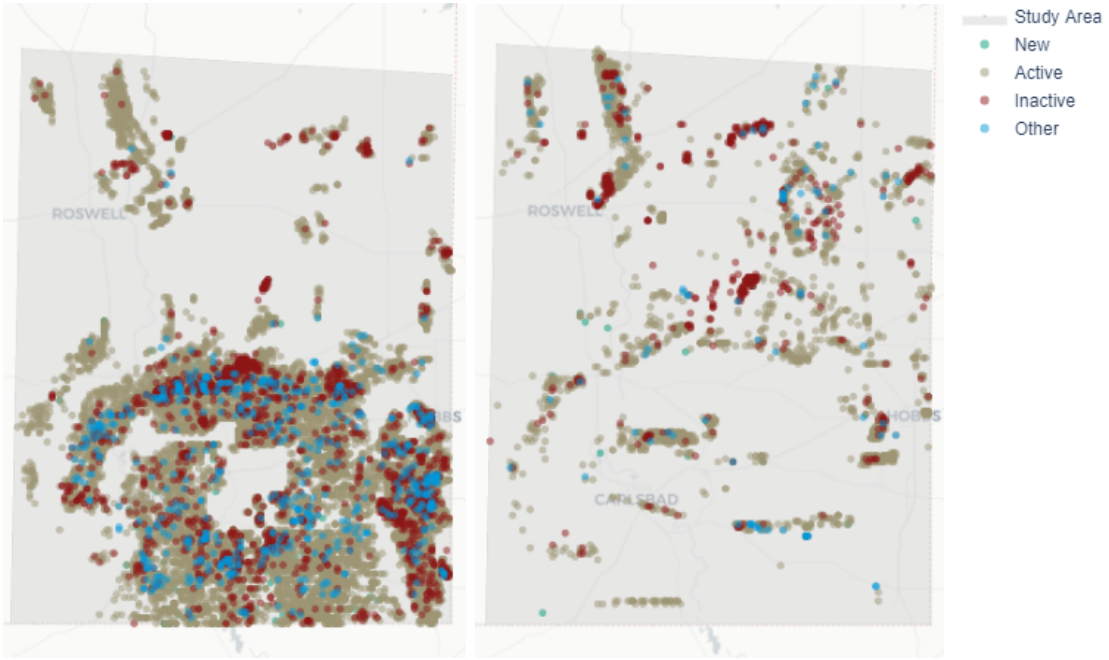

Figure S21: (a) Covered and (b) uncovered wells colored by status. The "new" wells refer to all drilling, drilled, drilled but uncompleted, and completed (but not yet producing as of March 2020) wells by Enverus category.[35] "Other" wells include cancelled, plugged and abandoned, permitted, temporarily abandoned, and unknown wells by Enverus category.

After spatial extrapolation, the total emission estimate is $1.106(\sum_i ER_{ES_i}^{ITMT} + \sum_k ER_k^{UD})$,

.

$$ER^{Spacial\ scale-up} = 1.106 ER^{>MDL} \tag{S23}$$

S5.1.7 Extrapolation to leaks below minimum detection

Kairos technology is designed to find medium- to large-sized leaks over a large geographic area at a fast pace. Section S5.1.5 describes procedures we followed to add to our emission estimate the undetected emissions within the PDR. To reach a total emission estimate, we need to include emissions that were below MDL as well.

Without data about the Permian emissions below the Kairos MDL, we first assume that this portion of the emissions is some fraction of the national-average NG loss rate of 2.2%

(+0.4%,-0.3%, 95% CI), which is the for upstream and midstream NG loss rate estimated by Alvarez et al.[13] To determine what fraction of the national loss rate would be entirely undetectable, we assess what fraction of leaks from prior national-scale studies would have – in theory – been undetectable by a Kairos survey under New Mexico-like wind conditions.

To do this, we use the Omara et al. dataset of 1009 site-level emission measurements from 9 field campaigns across the US.[15] This dataset formed the basis for the Alvarez et al. meta-analysis. We synthetically place these emissions in our study region by first generating wind-independent emission rates using the HRRR wind data from our New Mexico Permian survey area. The resulting distribution of these wind-independent leaks is shown in Figure S22c, showing skew in the leak sizes.

To simulate the test environment in alignment with the actual New Mexico campaign, we pull 10,000 wind records from HRRR with 100 timestamps and 100 asset locations randomly drawn from Kairos' asset coverage records. Figure S22b shows the distribution of these 10,000 wind records. Figure S23 shows the 100 randomly selected asset locations and Figure S24 shows the seasonal and diurnal distributions of the 100 selected timestamps, as well as those of all point asset visits (pipeline visits excluded).

To address the uncertainties in wind, 1000 MC realizations are performed to associate Omara site emissions with New Mexico winds, as shown in Figure S22c. Approximately 79 of the 1009 emission sites would be detectable by Kairos at MDL of 2 kgh/mps. Because of the skewed distribution of emission size in Figure S22a, these 8% of Omara et al. emissions are responsible for 67±3% of total emissions. Figure S22d shows the range of the fraction of emissions in the Omara et al. dataset that could possibly be detected by Kairos in this simulation.

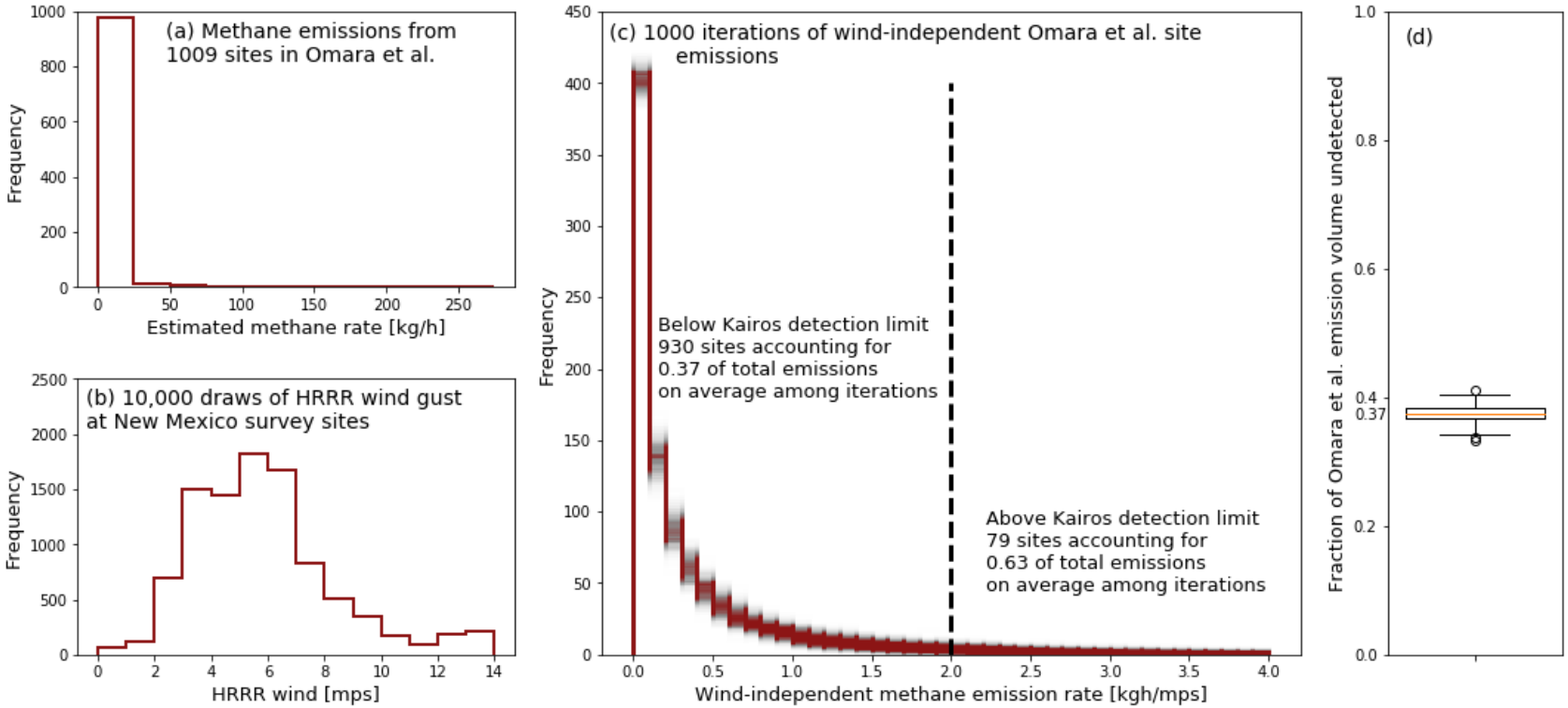

Figure S22: Simulation of Kairos survey of emissions from the Omara et al. sites if located in the New Mexico Permian. (a) is a histogram of the 1009 emission sizes taken directly from.[15] (b) is a histogram of 10,000 HRRR winds drawn from 100 New Mexico asset locations and 100 survey timestamps. (c) shows 1000 iterations of associating each emission rate with one of the New Mexico winds drawn in (b) to derive the wind-independent emission rate of the Omara et al. sites if they are located in the New Mexico Permian. The dashed line indicates the minimum detection threshold of Kairos technology, below which an emission cannot be detected by Kairos. (d) is a box plot showing the mean (orange tick) of 37%, the standard deviation (length of box), the 95% CI (top and bottom ticks), and the outliers (dots outside of the range of the 95% CI) of the fraction of the volume of methane emissions in the Omara et al. dataset that cannot be detected by Kairos in simulated flights over these sites.

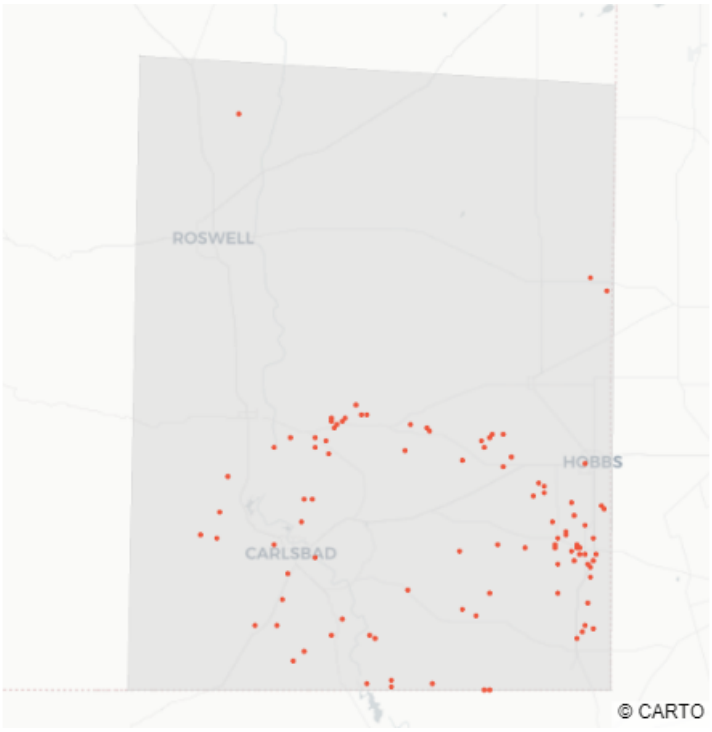

Figure S23: Locations of the 100 selected sites for drawing New Mexico wind for generating synthetic wind-independent Omara et al. emissions.

Given the estimated sensitivity, 37±3% of emissions are smaller than the MDL and are expected to be missed by Kairos under New Mexico Permian conditions. Assuming the national average loss rate of 2.2%, these 37% undetectable emissions correspond to $37\% \times 2.2\% = 0.8\%(\pm 0.1\%)$ of the additional loss rate below Kairos MDL. Note that this assumes that the Permian emissions below Kairos MDL are distributed similarly to those from the Omara et al. dataset (this in principle is unknowable with our measurements).

We denote this loss rate below MDL as $\%Loss^{<MDL}$ and add to the sum of emissions from the last step $\%Loss^{<MDL}$ multiplied with the methane production rate (PR_{CH_4}) in kg/h of the survey area to reach a regional total methane emissions of:

$$ER^{Total} = ER^{Spacial\ \ scale-up} + \%Loss^{<MDL} PR_{CH_4} \tag{S24}$$

Note that the ER^{Total} here is equivalent to $\hat{E}^{Total}$ in the main text. Calculation of PR_{CH_4} is described in Section S5.1.8.

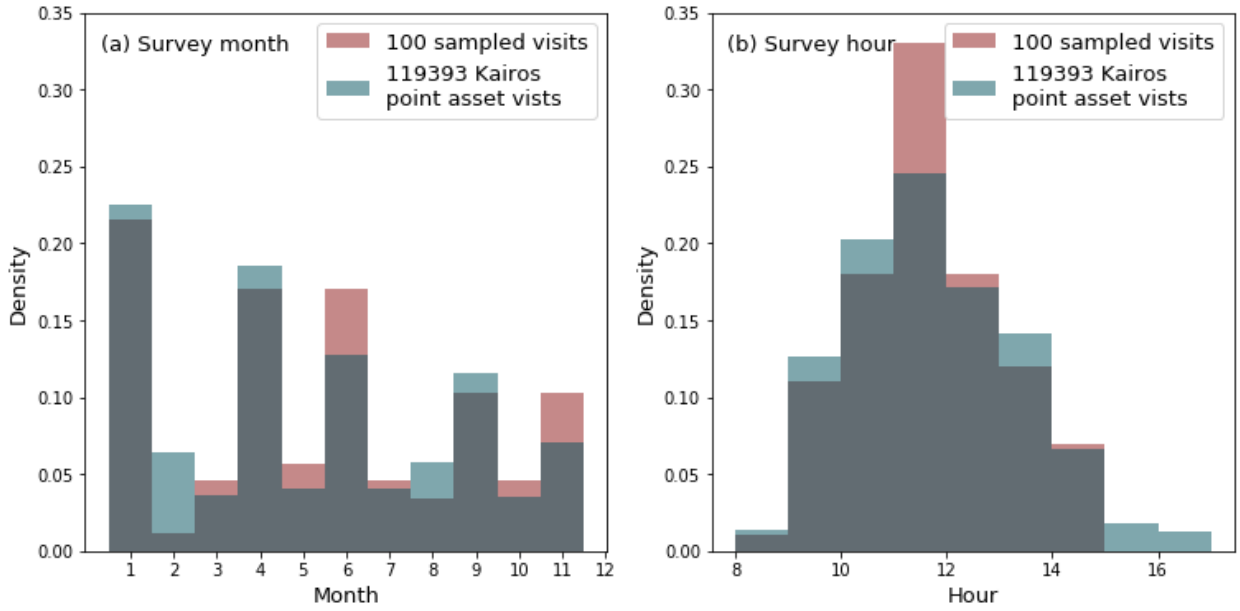

Figure S24: (a) Month and (b) hour distributions of the 100 sampled site visit times for drawing New Mexico wind for generating synthetic wind-independent Omara et al. emissions are plotted in red. The overall month and hour distributions of all 119,393 point asset visits (pipeline visits excluded) are plotted in teal.

S5.1.8 Loss rate calculation

We convert the CH_4 emission rates into NG emission rates based on the methane content in NG within different segments of the NG supply chain. NG from well sites, compressor stations, and storage tanks is assumed to have a mean volumetric methane content of 78.8%.[38] NG in pipelines and gas processing plants is assumed to have a mean methane content of 90.3%.[38] Other/ambiguous emission sources are also assumed to have a methane content of 78.8%. For an MC run indexed by m, let U_m be the set of active emission sources associated to assets that are well site, compressor station, storage tank, or other/ambiguous types. Let V_m be the set of active emission sources associated with pipelines or gas processing plants. $W_m = U_m \cup V_m$ is the set of all active emission sources of the m^{th} MC run. The mean the asset-type weighted mean methane content from emissions of the m^{th} MC run is:

$$\overline{(\%CH_4)_m} = \frac{78.8\% \sum_{i \in U_m} ER_{ES_i}^{ITMT} + 90.3\% \sum_{i \in V_m} ER_{ES_i}^{ITMT}}{\sum_{i \in W_m} ER_{ES_i}^{ITMT}} \tag{S25}$$

The asset-type weighted mean methane content from active emission sources of the 1000 MC runs is 81.2% (+0.9%/-0.5%).

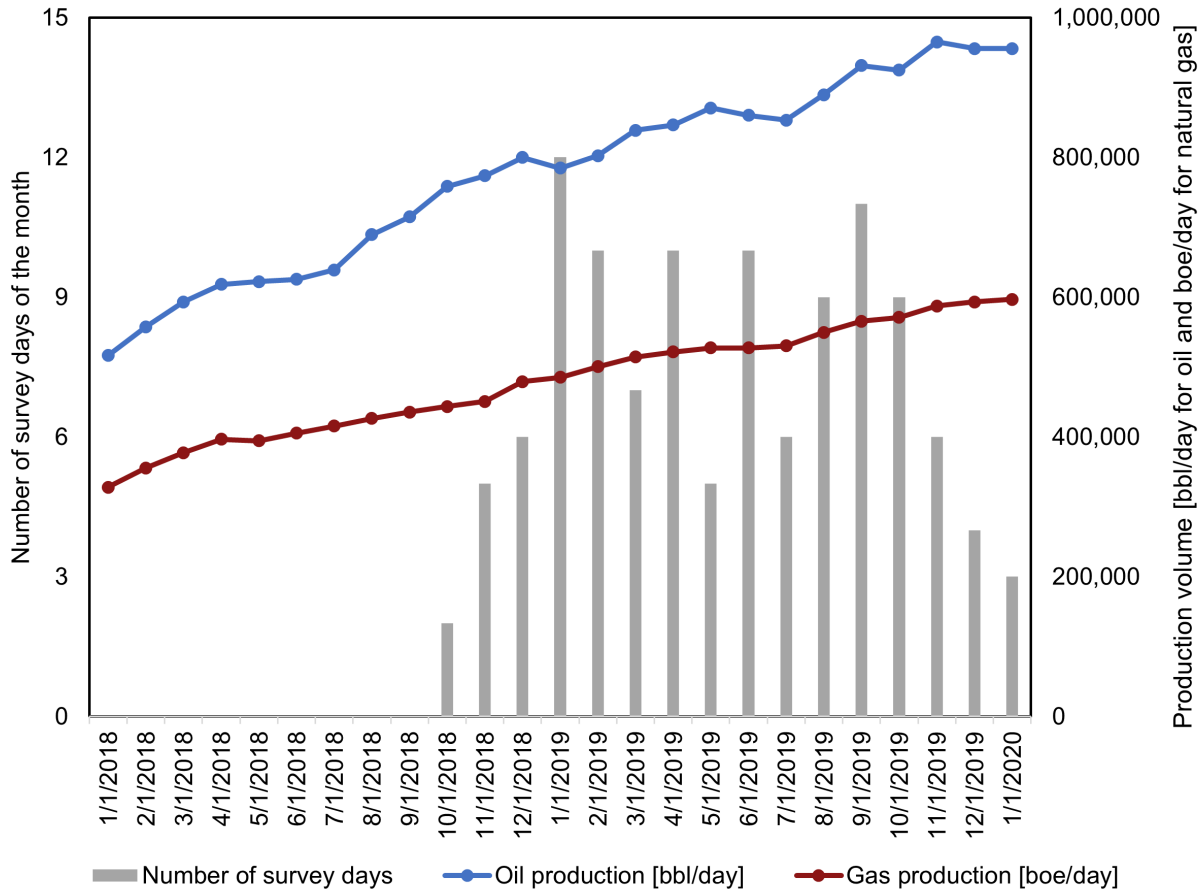

Figure S25: Number of survey days (left axis) of each month during the survey time. Monthly oil and gas production volume (right axis) in the survey region. Data from January 2018 to September 2018 are included to demonstrate the trend in production growth. Gas production is converted to barrel-of-oil equivalent (boe) per day.

In sum, our results show that O&G production in the survey region grew over time (see Figure S25). The NG production rate (PR_{NG}) of the survey area used in this study is a survey-time weighted average NG volume of 3.198 billion cubic feet per day (bcfd). The weights are the number of survey days of each month, as shown by the bars in Figure S25. Using the same assumption for methane content of 78.8% from wells,[38] the methane production rate in the survey area is:

$$PR_{CH_4} = 78.8\% PR_{NG} \tag{S26}$$

The loss rate of NG above MDL is computed by Equation S27.

$$\%Loss^{>MDL} = \frac{1.106(\sum_i ER_{ES_i}^{ITMT} + \sum_k ER_k^{UD})}{\overline{\%CH_4}\ PR_{NG}} \tag{S27}$$

The method for finding undetected methane losses below MDL is described in section S5.1.7.

In addition, the NG production loss based on measured emissions ($\hat{E}^{Measured}$) is:

$$\%Loss^{Measured} = \frac{\sum_i ER_{ES_i}^{ITMT}}{\overline{\%CH_4}\ PR_{NG}} = \frac{\hat{E}^{Measured}}{\overline{\%CH_4}\ PR_{NG}} \tag{S28}$$

The total loss rate of NG in the survey area is:

$$\%Loss^{Total} = \%Loss^{>MDL} + \%Loss^{<MDL} \tag{S29}$$

S5.2 Emission estimates by stage

Table S8 shows emission estimates from each quantification stage described in Section S5.1. All 1985 plumes sum up to a point estimate of 679 metric tonnes per hour (t/h) before calibration based on controlled releases. The sublinear calibration relationship (Equation S7) reduces total plume-level emission to 640 t/h. Randomly selecting one plume for each emission source brings down the mean of the total emissions estimate to 293 (+130/-131) t/h. After accounting for emission intermittency, the total emissions fall to 153 (+71/-70) t/h, or approximately half of the emissions before accounting for intermittency. This is the estimated methane rate that Kairos is able to detect at each snapshot of time of the entire survey region and is denoted as $\hat{E}^{Measured}$ in the main text.

The extrapolated emissions are shown in Figure 1a. Extrapolation to PDR adds an

additional 7 (+14/-6) t/h to the emission-source-level total, suggesting a less significant role for the smaller sources in the PDR compared to the hyper-emitters found in the survey area. Extrapolation to the total survey area, assuming that assets that are not covered by Kairos emit at the same rate as the assets covered, yields a regional total emission of 177 (+74/-68) t/h that are above the Kairos MDL. The last extrapolation we perform is adding emissions below minimum detection limit. This is done by finding the 0.8% NG production loss rate below minimum detection ($\%Loss^{<MDL}$) and multiplying it with gross methane production in the area. The total estimated methane emissions ($\hat{E}^{Total}$) is thus 194 (+72/-68) t/h.

Using monthly production data from Enverus and flight records of each survey month (Figure S25), the survey time-weighted NG production in the study area is 3.198 bcfd, indicating a %NG production loss estimate of 7.4% (±3.4%) corresponding to measured emissions, an above MDL %NG production loss estimate of 8.5% (+3.5%/-3.3%), and an overall %NG production loss estimate of 9.4% (+3.5%/-3.3%) from all upstream and midstream O&G assets in the survey area.

According to the MC simulations, the fifth percentile of the total NG production loss rate is 6.1%, suggesting a small chance of loss rate being smaller than 6.1%. The estimated loss rate here is significantly higher than the existing estimates in the literature listed in Table S4, for reasons described in Section S6.

S6 Emission sizes

Even the largest datasets of site-level point source methane emissions from ground surveys cover at most 1009 well sites.[15]

This comparison can be done in the other direction as well. Note that the overall incidence of leaks in our study is 1083 well-site plumes from 98,000 well site visits, for a total per-survey probability of well emission detection of 1.1%. Given 1009 samples in the Omara et al. dataset, we would then expect a Kairos-like technology seeing a total of approximately

Table S8: Total survey area emission quantification by stage (unit: t/h of methane if not specified)

Stage	Notation	Mean	Stddev	$5^{th}\%$	$50^{th}\%$	$95^{th}\%$
Raw plumes,including repeats	ER_{plume}	679	0	679	679	679
Controlled release calibration	$\widetilde{ER_{plume}}$	640	170	360	639	929
One plume per source	ER_{ES}	293	78	162	293	423
Intermittency	$ER_{ES}^{ITMT}(\hat{E}^{Measured})$	153	43	83	152	224
Partial detection scale-up	$ER^{>MDL}$	160	39	98	158	225
Spatial scale-up	$ER^{Spacial\ \ scale-up}$	177	43	109	175	249
Total emissions	$ER^{Total}(\hat{E}^{Total})$	194	43	126	192	266
%NG production loss measured	$\%Loss^{Measured}$	7.4%	2.1%	4.0%	7.3%	10.8%
%NG production loss above MDL	$\%Loss^{>MDL}$	8.5%	2.1%	5.2%	8.4%	12.0%
%NG production loss below MDL	$\%Loss^{<MDL}$	0.8%	0.1%	0.7%	0.8%	0.9%
%NG production loss total	$\%Loss^{Total}$	9.4%	2.1%	6.1%	9.3%	12.9%

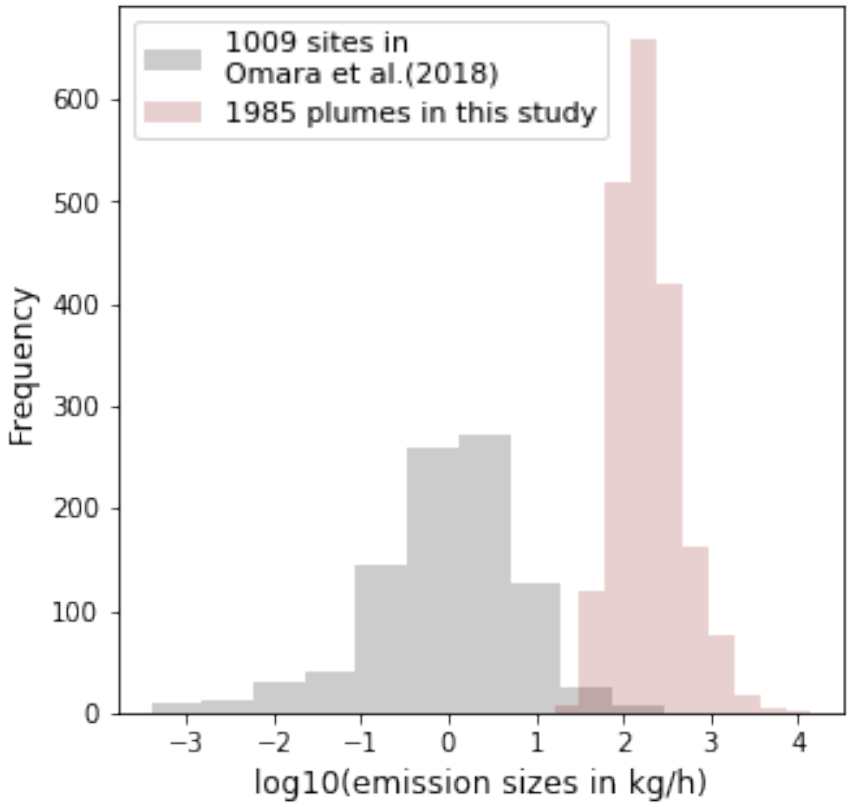

Figure S26: Size distributions of site-level emissions in this study and 9 other studies summarized in.[15]

11 plumes, assuming that each well site in the Permian Basin has on average 1.2 wells on site.[9] Given that there are most certainly null (0) measurements without plumes that were in the underlying studies but not reported into the Omara et al. dataset, we believe that this value of 11 plumes aligns with the above value of 79 detections by Omara et al. above Kairos' MDL.

The Permian Basin could be an outlier in terms of its prevalence of emissions on the scales of 100, 1000, and 10,000 kg/h. However, sample size and the detection limit likely also play a role in the different distributions. The ground-based surveys of the Omara et al. dataset are more sensitive to small-sized leaks than aerial surveys, yet the small sample size achievable by ground surveys limits the detection of low-probability and high-impact super-emitter events that are detectable by basin-wide aerial surveys.

As Figure 3c shows, 50% of total emissions are from 118 (∼12%) of the 958 sources larger than 308 kg/h, indicating the importance of finding and fixing the less-frequent super emitters.

The distribution of the plume emission rates is shifted to the left of the emission source size distribution in Figure S27, because the emission source averages take into account null observations. For both distributions, the heavy tail gets even heavier toward the high end and contains a disproportionate presence of midstream assets.

Given the importance of the heavy tail and that there are over one million active O&G wells in the US, with extensive supporting infrastructure, including gathering pipelines, compressor stations, and processing plants, a ground-based approach may not be sufficient to characterize total methane emissions in a producing region.

S7 Sensitivity cases

We perform a number of sensitivity tests in order to explore possible modeling variations that may affect our results. These include fitting alternative models to the controlled release

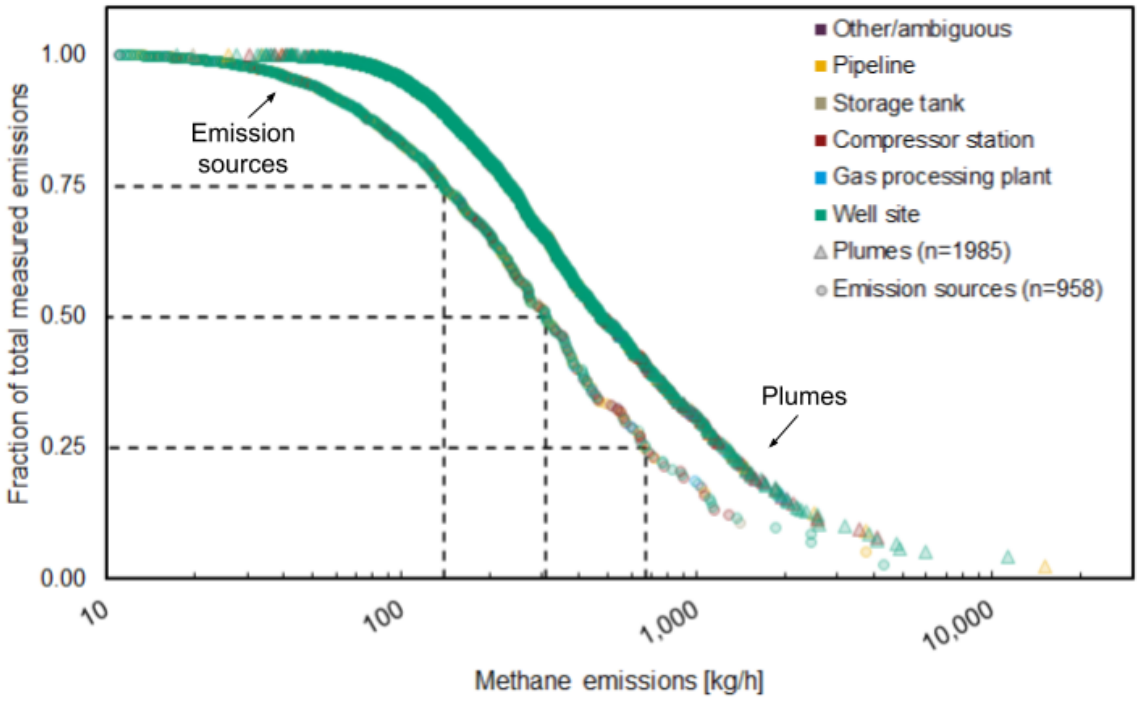

Figure S27: Size distributions of plumes and emission sources and their contribution to total emissions. The round symbols indicate the 958 emission sources and the curve is the same as the curve in Figure 4c in the main text. The triangular symbols indicate the 1985 detected methane plumes. The distribution of the plume emission rates is placed to the left of the emission source size distribution because the emission source averages take into account the null observations.

data, disabling extrapolation of controlled release results, excluding large plumes as possible outliers, excluding emissions below Kairos minimum detection threshold (MDL), and applying alternative wind datasets. The following sections describes each sensitivity case summarized in Table 1 in the main text and reproduced here as Table S9.

Table S9: Survey-area total methane emission rate and loss rate estimates, as a fraction of total methane production, for the base case and seven sensitivity cases. The two alternative calibration methods increase emissions relative to our base case. Using alternative wind data results in comparable emission estimates. The last three sensitivity cases estimate the emission lower-bound and show robustness of the base case emission estimates.

Cases	$\hat{E}^{Total}$ (t/h)			%NG production loss		
	Mean	5^{th}%	95^{th}%	Mean	5^{th}%	95^{th}%
Base case	194	126	266	9.4%	6.1%	12.9%
Linear fit for calibration	212	136	296	10.2%	6.6%	14.3%
Linear fit forced through origin for calibration	228	131	335	11.0%	6.4%	16.0%
Cutoff at 1σ below max controlled release	216	137	301	10.4%	6.9%	14.6%
Dark Sky wind high time resolution	181	124	244	8.7%	6.1%	11.8%
Dark Sky wind low time resolution	217	142	301	10.4%	6.8%	14.3%
Disable extrapolation	167	119	220	8.1%	5.7%	10.6%
Exclude top 20 plumes	173	117	233	8.3%	5.5%	11.2%
No below minimum detection emissions	177	109	249	8.5%	5.2%	12.0%

S7.1 Linear fit for calibration

As an alternative modeling approach, we fit the controlled release data with a linear regression model instead of a sublinear fit in our base case. The fitting parameters are shown in Equation S4. As discussed in S1.5, the linear model has its limitations of nonzero intercept and nonuniform residuals, as well as boundary bias concerns discussed in Section S1.6. Nevertheless, the resulting emission estimate gives insight into the impact of model selection on emission quantification.

The NG production loss rate estimate based on a linear fit is 10.2% (+4.1%/-3.6%), which is greater than the base case estimate because switching from a sublinear relationship to a linear one brings up the calibrated emissions of the top-emitting plumes. This sensitivity

case serves as an upper bound for the emission estimate.

S7.2 Linear fit forced through origin for calibration

The non-zero intercept of Equation S4 artificially increase low-end emissions and the below-one slope of Equation S4 reduce the magnitude of emissions above the intersection with the parity line at ∼500 kg/h. To address the possibility that the nonzero intercept could introduce upward bias into the calibration, we fit another regression line with the intercept fixed at the origin (Equation S5). The resulting %NG production loss estimate is 11.0% (+5.0%/-4.6%).

S7.3 Cutoff at 1σ below max controlled release

Our base case assumes the overestimation tendency for larger releases is from an underlying nonlinearity in the relationship between gas release volume and the appearance of the plume in Kairos imagery following hypothesis 1. Alternatively, the nonlinearity can be caused by the boundary effect introducing unbalanced errors in the data points of large release estimates (hypothesis 2). Section S1.6 describes the criteria we use to exclude data points in the calibration process to avoid biases caused by boundary effect.

If the cutoff threshold is chosen to be 750 kg/h, one σ_δ below AMR_{CR}^{max} (Equation S10), then the calibration relationship becomes Equation S12. The residual distribution of this fitting model is shown in Figure S11. This method brings the mean %NG production loss estimate up to 10.4%.

S7.4 Dark Sky wind

We also explore the effect of wind speeds. Dark Sky provides wind speed estimates on a finer temporal and spatial grid than NOAA's HRRR. More details of Dark Sky can be found in Section S1.3.2. As mentioned above, we retrieved Dark Sky wind data on two occasions for

the same emission locations and timestamps: the first time in February 2020 and the second time in October 2020. The resulting wind speeds from the two retrievals differ substantially from each other (Figure S4), potentially due to the reduction in time resolution and model changes between the two retrievals.

We present two sensitivity cases in Table S9 based on the two Dark Sky wind data retrievals. The high time resolution (February 2020) retrieval results in a loss rate estimate of 8.7%(+3.1%/-2.6%), whereas the low time resolution (October 2020) retrieval leads to a loss rate estimate of 10.4%(+3.9%/-3.6%). The confidence intervals of the two Dark Sky cases and that of the base case largely overlap, indicating that errors from modeled wind, although significant, do not negate the large emissions estimate of this study.

S7.5 Disallow extrapolation

The largest release tested in the field trial was 1025 kg/h.[22] However, the largest emission seen in the Permian campaign is 15,115 kg/h (after calibration with Equation S7). If we disallow extrapolation of Equation S7 and limit all observed emissions larger than 1025 kg/h to have an estimated emission size of 1025 kg/h, the resulting loss rate mean reduces to 8.1% (+2.5%/-2.4%). Note that this is a highly conservative assumption and this estimate can be interpreted as a lower-bound of the mean loss rate in the study area.

S7.6 Exclude top 20 plumes

If we treat the top 20 plumes ($\approx$ 1% of plumes) as possible outliers and suppose that Kairos did not encounter them in the survey, the mean loss rate would be 8.3% (+2.9%/-2.8%), or a 12% relative reduction from the base case. The sizable decline resulting from the removal of the top 20 plumes is further illustration of the significant contribution of a few large sources on overall basin emissions.

S7.7 No below minimum detection emissions

We add 0.8±0.1% to our loss rate estimate to account for undetected emissions below minimum detection limit MDL (Equation S29). The loss rate below MDL is based on the national average loss rate estimate by Alvarez et al. and the size distribution of site-level emissions compiled by Omara et al.[13,15] Removing $\%Loss^{<MDL}$=0.8% from our estimate gives a total mean loss rate of 8.5% (+2.5%/-3.2%), which is also a lower-bound for the mean %NG production loss in the study area.

S7.8 Conclusion

The sensitivity cases show that even the lower-bound estimates of the conservative scenarios based on our method are larger than estimates from other Permian studies summarized in Table S4.

S8 Comparison with Cusworth et al. 2021

Cusworth et al. conducted aerial surveys of the Permian Basin from September 2019 to November 2019 with two remote sensing airborne platforms: the Next-Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) and the Global Airborne Observatory (GAO).[12] With repeated, "wall-to-wall" (full coverage) surveys in their study area (purple polygons in Figure S28), Cusworth et al. detected 3067 methane plumes from 1756 emission sources.

In the overlapping domain of this study and the Cusworth et al. study (see Figure S28), Kairos detected 1874 methane plumes from 902 distinct emission sources, while Cusworth et al. detected 1263 plumes from 612 distinct emission sources.

We apply the method described in Section S5.1 to the plumes from Cusworth et al. (without applying calibration based on controlled releases, as we do not have sufficient controlled release data to do so). After accounting for intermittency, we estimate measured

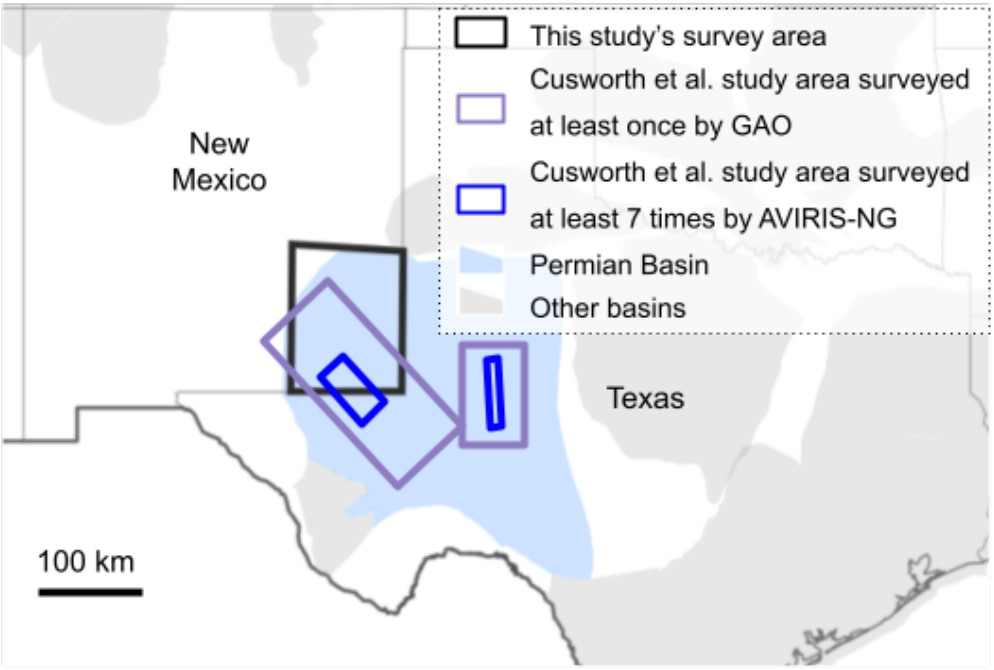

Figure S28: Figure 1b in the main text reproduced with the addition of the Cusworth et al. study area.[12] The Cusworth et al. study area boundaries are obtained by digitizing Figure 1 in Cusworth et al.[12] using WebPlotDigitizer.[39] The purple-bordered polygons indicate that the Cusworth et al. study area that was surveyed at least once by the Global Airborne Observatory (GAO) instrument. Sites in the blue polygons were surveyed at least 7 times by the Next-Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) instrument. The purple polygons contain the blue polygons.

emissions from the JPL survey at 98(±12) t/h in the overlapping area of the black and the purple polygons in Figure S28 (henceforth "overlapping domain"), which corresponds to a NG production loss rate of 4.4%(±0.5%) assuming 3.41 bcfd of NG production, average of September 2019 to November 2019 NG production in the overlapping domain reported by Enverus.[35]

As for the Kairos dataset, after accounting for intermittency, the upstream and midstream O&G sector in the overlapping domain emits methane at 145 (+19/-16) t/h (this value for the entire Kairos study area has a mean of 153 t/h). Considering that Cusworth et al. had a full coverage of O&G wells and Kairos covered 95% of the wells in the overlapping domain, we apply a spatial extrapolation factor of 1/0.95 to Kairos estimates for a more direct comparison with emission estimates based on Cusworth et al. data. The spatially-extrapolated

Kairos-measured methane emissions is thus 153 (+20/-17) t/h in the overlapping domain, corresponding to a NG production loss rate of 7.6% (+1.0%/-0.8%), based on Kairos survey time-averaged NG production of 3.09 bcfd in the overlapping domain.

Although the Cusworth et al. repeated, full-coverage aerial study of the Permian Basin is similar to the Kairos New Mexico campaign, we need to make additional assumptions to directly compare the two estimated NG production loss rates:

1. Stationary ground truth emissions distribution in the overlapping domain across the study time periods

2. Consistent definitions of coverage counts (C_i) and plume counts (J_i) for individual emission sources across both studies

3. Same detection limits and detection probabilities in the partial detection range for Kairos technology, AVIRIS-NG, and GAO

4. Same exclusion criteria for unclear plume signals and artifacts

Evaluating the extent to which these assumptions hold, we can examine the divergence in estimated NG production loss rates by comparing the distribution of methane plume sizes.

S8.1 Apparent difference in effective minimum detection threshold

The final two assumptions, identical minimum detection capabilities and the same exclusion criteria for the technologies in the two studies, appear not to entirely hold. The cumulative distribution function of total plume-level emissions as a function of emission size, shown in S29, is approximately a straight line in both cases, excluding the low and high tails. This indicates a higher frequency of smaller plume size. The straight line behavior begins to flatten on the low end between 100 and 150 kg/h for Kairos and 200 and 300 kg/h for

Cusworth et al. Above this size, plume frequency begins to fall at smaller emission sizes relative to the distribution established for larger emissions.

It is very unlikely that the decline in plume counts with smaller plume size is due to a decline in ground-truth plumes to observe. As seen in Omara et al., ground surveys indicate increasing likelihood of finding plumes with smaller plume sizes down to scales much smaller (e.g., 1-10 kgh). Thus, the decrease in plume counts with smaller size, in both cases, is almost certainly due to detection capabilities of the instruments and the methods used to filter possible false positives.

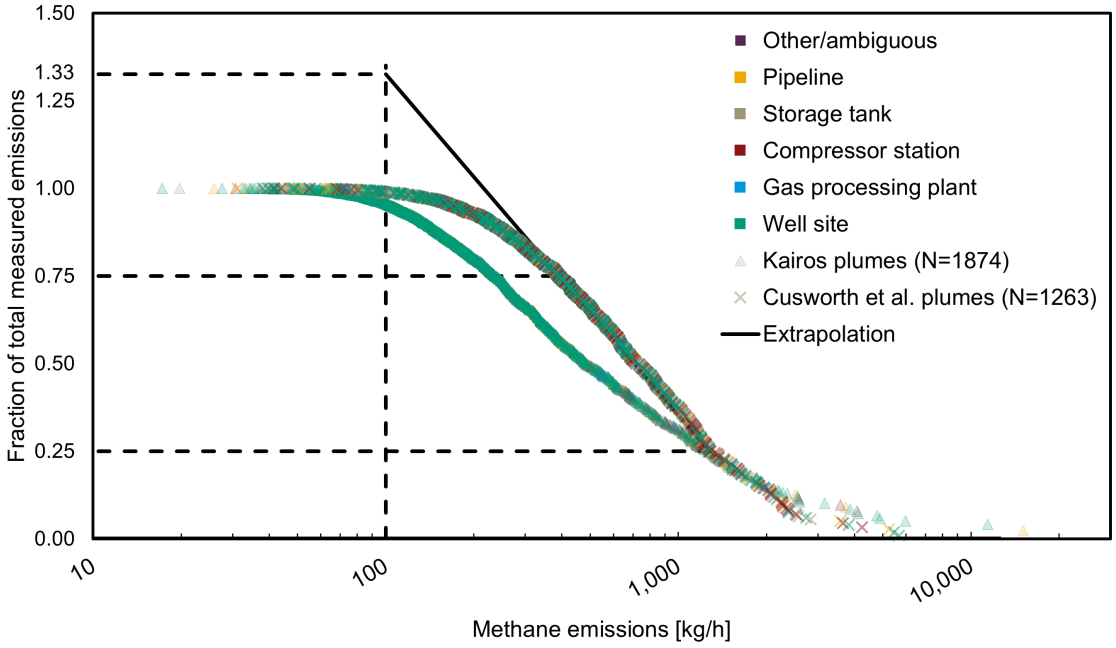

Figure S29: Comparison of plume size distributions in the overlapping domain (intersection of the black and the purple polygons in Figure S28) of this study and the Cusworth et al. 2020 Permian campaign.[12] Of the 1985 Kairos-detected plumes, 1874 are within the overlapping study area. Definitions of asset types from the two studies may vary. Note that extrapolating the Cusworth et al. distribution to the point at which the Kairos distribution approaches minimum detection capabilities increases total estimated emissions by 33%. Also, note that the difference in slopes is largely due to the higher effective minimum detection threshold in the Cusworth et al. study.

Both approaches are measuring the same area using a similar detection and quantification method and survey design. As a result, we would expect both technologies to be sampling

from the same underlying distribution. This suggests that the flattening of the Cusworth et al. emissions distribution is likely due to a higher effective minimum detection threshold for this survey.[12] The alternative hypothesis that the true underlying distribution of emissions falls off in this manner appears unlikely given that the Kairos survey continues to observe the straight line behavior until plumes that reach its documented partial or minimum detection threshold from.[22]

This higher effective minimum detection threshold is consistent with the plume selection methods used in Cusworth et al.[12] Controlled release testing of the AVIRIS-NG and GAO instruments demonstrates the ability to detect emissions of 2.3 kg/h at a flight altitude likely lower than 1 km. The altitude of the flights in the Permian survey was 8 km and 4.5 km for AVIRIS-NG and GAO, respectively. This may increase the technical minimum detection threshold of the instruments. More importantly, candidate plumes were filtered through a human quality control process with relatively conservative exclusion criteria to reduce the likelihood of false positives. As a result, the combination of instrument capabilities and human review is a possible explanation for the relatively lower number of plumes identified in the 100-300 kg/h range in Cusworth et al.

To determine the likely effects of this higher effective minimum detection threshold, we extrapolate the straight line behavior from the Cusworth et al. distribution (see Figure S29) to the point at which the Kairos distribution flattens, selected as 100 kg/h based on the analysis described below in Section S8.4. The straight line is fit through the 75th and 25th percentiles of the Cusworth et al. emissions distribution (1276 kg/h and 391 kg/h, respectively), between which the distribution appears linear. The slope of this line is -0.97 in the logarithmic space (although the slope of the distribution would of course be shallower if the minimum detection limit were lower). This extrapolation suggests that if the process used in Cusworth et al. had the same minimum detection level as the Kairos survey, overall emissions would rise by 33%. This would increase the estimated fractional loss from 4.4% to 5.9%.

S8.2 Differing loss rate estimates within and out of the AVIRIS-NG repeated sampling area

Cusworth et al. covered the 55,000 km^2 of study area in the Delaware and Midland Basin (purple polygons in Figure S28) at least once by their GAO instrument.[12] To evaluate intermittency, Cusworth et al. deployed AVIRIS-NG to sample a subregion (blue polygons in Figure S28) of their study area with higher frequency. GAO was deployed at ∼4.5 km altitude and AVIRIS-NG was deployed at 8 km altitude. Although the two systems have the identical imaging spectrometers, each of which "has the same performance",[12] the difference in altitude may result in differing minimum detection thresholds of the two systems.

The vast majority of methane plume and flight coverage records by AVIRIS-NG are within the blue polygons in Figure S28. To roughly separate the detection results from the two systems, we divide the overlapping domain into two parts: within and out of the blue polygons. The NG production loss rate estimates based on Cusworth et al. data are respectively 3.1% (+0.9%/-0.8%) and 5.6% (+0.7%/-0.6%) within and out of the blue polygons, suggesting either a larger NG production loss out of the blue polygons, or a higher sensitivity in the GAO instrument than the AVIRIS-NG instrument.

The NG production loss rate estimates in the two parts of the overlapping domain based on Kairos data set are roughly the same.

S8.3 Differing definitions of coverage and plume counts

Our analysis treats each overflight over an asset as a coverage instance and each detected plume at an emissions source as a separate plume. Cusworth et al. were focusing on emission intermittency across days. As a result, multiple coverage instances or multiple plume detections in a given day are merged into single instances in both cases. For plumes, daily emissions are the average of all detected emissions on that day. As a result, in instances in which an asset was covered twice in a day but only one plume was detected, the dataset

would present the detected emissions as that day's emissions, which would increase emissions relative to the coverage-level method used in this paper.

The survey design in Cusworth et al. did not systematically attempt to cover assets multiple times in one day, although this did occur incidentally for a small number of documented emissions sources, potentially resulting in a modest correction to the overall emissions estimate.

S8.4 Kairos plume counts by emission size

Table S10: Kairos-detected methane plumes in the domain overlapping with the Cusworth et al. survey, grouped by one-quarter order of magnitude plume size bins.

Size tranches of emissions (kg/h)	N_{plumes}	Average plume size	N_{plumes} per well visit	Emission contribution
<10	0	0	0.00%	0.00%
10 - 18	1	17	0.00%	0.00%
18 - 32	4	26	0.00%	0.02%
32 - 56	58	46	0.06%	0.44%
56 - 100	299	80	0.29%	3.90%
100 - 178	555	136	0.53%	12.38%
178 - 316	458	238	0.44%	17.88%
316 - 562	274	405	0.26%	18.21%
562 - 1,000	125	719	0.12%	14.75%
1,000 - 1,778	67	1,270	0.06%	13.96%
1,778 - 3,162	22	2,173	0.02%	7.84%
3,162 - 5,623	8	4,022	0.01%	5.28%
>5,623	3	10,833	0.00%	5.33%
Total	1,874		1.80%	100.00%

If we group the 1874 Kairos plumes in the overlapping domain by plume sizes and set the bin width to be one quarter order of a magnitude, the majority of plume-level emissions are contributed by plumes sized between 100 and 1778 kg/h (>10% emission contribution in Table S10). This trend may begin to fall off in the 100-178 kg/h tranche (12.38% of total emissions v. 13.96-18.21% of total emissions for the next four larger tranches). The trend certainly falls off in the 56-100 kgh tranche, which constitutes only 3.90% of total emissions.

As a result, we select 100 kg/h as the threshold where the Kairos distribution in Figure S29 starts to flatten. We use this 100 kg/h level at the endpoint for the extrapolation of the Cusworth et al. emissions distribution in Figure S29.

References

(1) U.S. Environmental Protection Agency (EPA), Understanding Global Warming Potentials. https://www.epa.gov/ghgemissions/understanding-global-warming-potentials, (Accessed on 02/09/2021).

(2) U.S. Energy Information Administration (EIA), US energy facts explained. `https://www.eia.gov/energyexplained/us-energy-facts/`, (Accessed on 02/09/2021).

(3) U.S. Energy Information Administration (EIA), Permian region drilling productivity report. `https://www.eia.gov/petroleum/drilling/pdf/permian.pdf`, (Accessed on 09/29/2020).

(4) U.S. Energy Information Administration (EIA), Natural gas spot and futures prices (NYMEX). `https://www.eia.gov/dnav/ng/ng_pri_fut_s1_d.htm`, (Accessed on 09/29/2020).

(5) U.S. Energy Information Administration (EIA), Permian Basin natural gas prices up as a new pipeline nears completion. `https://www.eia.gov/naturalgas/weekly/archivenew_ngwu/2020/03_19/tabs-rigs-1`, (Accessed on 05/10/2021).

(6) U.S. Department of Energy (DOE), New Mexico natural gas flaring and venting regulations. `https://www.energy.gov/sites/prod/files/2019/08/f66/New Mexico.pdf`, (Accessed on 05/09/2021).

(7) Zhang, Y.; Gautam, R.; Pandey, S.; Omara, M.; Maasakkers, J. D.; Sadavarte, P.; Lyon, D.; Nesser, H.; Sulprizio, M. P.; Varon, D. J.; Zhang, R.; Houweling, S.; Zavala-

Araiza, D.; Alvarez, R. A.; Lorente, A.; Hamburg, S. P.; Aben, I.; Jacob, D. J. Quantifying methane emissions from the largest oil-producing basin in the United States from space. *Science advances* **2020**, *6*, eaaz5120.

(8) Lyon, D. R.; Hmiel, B.; Gautam, R.; Omara, M.; Roberts, K. A.; Barkley, Z. R.; Davis, K. J.; Miles, N. L.; Monteiro, V. C.; Richardson, S. J.; Conley, S.; Smith, M. L.; Jacob, D. J.; Shen, L.; Varon, D. J.; Deng, A.; Rudelis, X.; Sharma, N.; Story, K. T.; Brandt, A. R.; Kang, M.; Kort, E. A.; Marchese, A. J.; Hamburg, S. P. Concurrent variation in oil and gas methane emissions and oil price during the COVID-19 pandemic. *Atmospheric Chemistry and Physics* **2021**, *21*, 6605–6626.

(9) Robertson, A. M.; Edie, R.; Field, R. A.; Lyon, D.; McVay, R.; Omara, M.; Zavala-Araiza, D.; Murphy, S. M. New Mexico Permian Basin measured well pad methane emissions are a factor of 5–9 times higher than US EPA estimates. *Environmental Science & Technology* **2020**, *54*, 13926–13934.

(10) Schneising, O.; Buchwitz, M.; Reuter, M.; Vanselow, S.; Bovensmann, H.; Burrows, J. P. Remote sensing of methane leakage from natural gas and petroleum systems revisited. *Atmospheric Chemistry and Physics* **2020**, *20*, 9169–9182.

(11) Irakulis-Loitxate, I.; Guanter, L.; Liu, Y.-N.; Varon, D. J.; Maasakkers, J. D.; Zhang, Y.; Chulakadabba, A.; Wofsy, S. C.; Thorpe, A. K.; Duren, R. M.; Frankenberg, C.; Lyon, D. R.; Hmiel, B.; Cusworth, D. H.; Zhang, Y.; Segl, K.; Gorrono, J.; Sanchez-Garcia, E.; Sulprizio, M. P.; Cao, K.; Zhu, H.; Liang, J.; Li, X.; Aben, I.; Jacob, D. J. Satellite-based survey of extreme methane emissions in the Permian basin. *Science Advances* **2021**, *7*, eabf4507.

(12) Cusworth, D. H.; Duren, R. M.; Thorpe, A. K.; Olson-Duvall, W.; Heckler, J.; Chapman, J. W.; Eastwood, M. L.; Helmlinger, M. C.; Green, R. O.; Asner, G. P.; Den-

nison, P. E.; Miller, C. E. Intermittency of Large Methane Emitters in the Permian Basin. *Environmental Science & Technology Letters* **2021**, *8*, 567–573.

(13) 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 US oil and gas supply chain. *Science* **2018**, *361*, 186–188.

(14) Maasakkers, J. D.; Jacob, D. J.; Sulprizio, M. P.; Turner, A. J.; Weitz, M.; Wirth, T.; Hight, C.; DeFigueiredo, M.; Desai, M.; Schmeltz, R.; Hockstad, L.; Bloom, A. A.; Bowman, K. W.; Jeong, S.; Fischer, M. L. Gridded national inventory of US methane emissions. *Environmental science & technology* **2016**, *50*, 13123–13133.

(15) 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. *Environmental science & technology* **2018**, *52*, 12915–12925.

(16) Rutherford, J. S.; Sherwin, E. D.; Ravikumar, A. P.; Heath, G. A.; Englander, J.; Cooley, D.; Lyon, D.; Omara, M.; Langfitt, Q.; Brandt, A. R. Closing the methane gap in US oil and natural gas production emissions inventories. *Nature communications* **2021**, *12*, 1–12.

(17) Brandt, A. R.; Heath, G.; Kort, E.; O'Sullivan, F.; Pétron, G.; Jordaan, S. M.; Tans, P.; Wilcox, J.; Gopstein, A.; Arent, D.; Wofsy, S.; Brown, N.; Bradley, R.; Stucky, G.; Eardley, D.; Harriss, R. Methane leaks from North American natural gas systems. *Science* **2014**, *343*, 733–735.

(18) Kairos Aerospace, Technical White Paper: Methane Detection. https://osf.io/7njpv/, 2019.

(19) Jones, B. B.; Dieker, S. W. Systems and methods for detecting gas leaks. https://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=netahtml PTO search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN 10267729, 2019.

(20) Kairos Aerospace, Methane Emissions Quantification. https://osf.io/y6w7r/, 2020.

(21) U.S. Energy Information Administration (EIA), Maps: Oil and Gas Exploration, Resources, and Production. https://www.eia.gov/maps/maps.htm, (Accessed on 04/26/2021).

(22) Sherwin, E. D.; Chen, Y.; Ravikumar, A. P.; Brandt, A. R. Single-blind test of airplane-based hyperspectral methane detection via controlled releases. *Elementa: Science of the Anthropocene* **2021**, *9*.

(23) Duren, R. M.; Thorpe, A. K.; Foster, K. T.; Rafiq, T.; Hopkins, F. M.; Yadav, V.; Bue, B. D.; Thompson, D. R.; Conley, S.; Colombi, N. K.; Frankenberg, C.; McCubbin, I. B.; Eastwood, M. L.; Falk, M.; Herner, J. D.; Croes, B. E.; Green, R. O.; Miller, C. E. California's methane super-emitters. *Nature* **2019**, *575*, 180–184.

(24) Dark Sky by Apple Inc., Dark Sky data attribution. https://darksky.net/attribution, (Accessed on 07/16/2021).

(25) Ravikumar, A. P.; Sreedhara, S.; Wang, J.; Englander, J.; Roda-Stuart, D.; Bell, C.; Zimmerle, D.; Lyon, D.; Mogstad, I.; Ratner, B.; Brandt, A. R. Single-blind inter-comparison of methane detection technologies–results from the Stanford/EDF Mobile Monitoring Challenge. *Elementa: Science of the Anthropocene* **2019**, *7*.

(26) Bañuelos-Ruedas, F.; Camacho, C. Á.; Rios-Marcuello, S. Methodologies used in the extrapolation of wind speed data at different heights and its impact in the wind energy resource assessment in a region. *Wind farm-technical regulations, potential estimation and siting assessment* **2011**, 97–114.

(27) Grossman, A. Dark Sky Has a New Home. https://blog.darksky.net/dark-sky-has-a-new-home/#:~:text=By\%20Adam\%20Grossman\%20on\%20August,be\%20receiving\%20a\%20full\%20refund.

(28) HRRR archive at the University of Utah. http://home.chpc.utah.edu/~u0553130/Brian_Blaylock/, (Accessed on 10/10/2020).

(29) Ravikumar, A. P.; Wang, J.; McGuire, M.; Bell, C. S.; Zimmerle, D.; Brandt, A. R. "Good versus good enough?" Empirical tests of methane leak detection sensitivity of a commercial infrared camera. *Environmental science & technology* **2018**, *52*, 2368–2374.

(30) Eastern Research Group, City of Fort Worth natural gas air quality study. https://www.fortworthtexas.gov/departments/development-services/gaswells/air-quality-study/final, 2011; (Accessed on 07/16/2021).

(31) Environmental Defense Fund, Permian Methane Analysis Project (PermianMAP). https://www.permianmap.org/, (Accessed on 09/29/2020).

(32) Royal Netherlands Meteorological Institute, TROPOMI methane data product. http://www.tropomi.eu/data-products/methane, (Accessed on 02/05/2021).

(33) GHGSat, GHGSat global emissions monitoring services. https://www.ghgsat.com/data-products-analytics/, (Accessed on 02/05/2021).

(34) European Space Agency, About GOSAT-2. https://earth.esa.int/eogateway/missions/gosat-2?text=methane, (Accessed on 02/05/2021).

(35) Enverus, Exploration and Production. https://www.enverus.com/industry/exploration-and-production/.

(36) Johnson, F.; Wlazlo, A.; Keys, R.; Desai, V.; Wetherley, E. B.; Calvert, R.; Berman, E. S. Airborne methane surveys pay for themselves: An economic case study of increased revenue from emissions control. https://doi.org/10.31223/X5RP7S, 2021.

(37) National Oceanic and Atmospheric Administration (NOAA), High Resolution Rapid Refresh (HRRR) CONUS 2-D Fields GRIB2 table documentation. https://rapidrefresh.noaa.gov/hrrr/HRRRv4_GRIB2_WRFTWO.txt, 2020.

(38) Environmental Defense Fund, New Data: Permian Oil & Gas Producers Releasing Methane at Three Times National Rate. https://www.edf.org/media/new-data-permian-oil-gas-producers-releasing-methane-three-times-national-rate.

(39) Rohatgi, A. Webplotdigitizer: Version 4.4. 2020; https://automeris.io/WebPlotDigitizer.