Supplementary Materials: Evaluating development of empirical estimates using two top-down methods at midstream natural gas facilities

Jenna A. Brown [1], Matthew R. Harrison[2], Tecle Rufael [2], Selina A. Roman-White [3], Gregory B. Ross [3], Fiji C. George [3] and Daniel Zimmerle [1*]

S1. Methods

A typical measurement day included:

- The two top-down (TD) methods (aerial Light Detection and Ranging (LiDAR) (Solution 1) and drone-mounted flux plane mass balance (Solution 2)) were instructed to focus on producing multiple whole-facility estimates, rather than focusing on identification of emission sources from subsets of the facility. These instructions represent an intentional change from the baseline phase [32]. In the baseline phase, TD methods typically performed one or two whole-facility estimates, with substantial measurement time and analysis focused on isolating emitters to subsets of the facility or performing estimates at other facilities. In this study, the TD methods made as many whole-facility estimates as possible throughout the measurement day.
- The facility operator was instructed to:
  - Provide a contemporaneous bottom-up (BU) inventory estimate that considers the facility's operational state, employing Greenhouse Gas Reporting Program (GHGRP) methodologies plus supplemental emissions, including any available direct measurements.
  - Provide a log of when compressor units were operating and/or pressurized.
  - Estimate of the total emissions from each episodic emission event, typically maintenance activities like blowdowns (depressurization of equipment).
- An independent observer from the study team noted compressor states, state changes, and the timing of episodic events, such as blowdowns, compressor starts, etc. The observer also noted any unusual aspects of the environmental conditions, TD methods' flights, etc.

Since neither method can accurately estimate emissions from large episodic events, such as compressor blowdowns (depressurization of a compressor), the study team noted the timing of these events and discarded estimates made during large episodic events, if possible. Corrections at specific facilities are discussed in SI Section 2.

S1.1. Solution 1

Data was delivered as a set of 'flight swaths' across the facility. A flight swath represented the area covered by the aircraft during a single pass, with each flyover generating a corresponding swath. To create "whole-site" estimates, study team reviewers examined the flight swath coverage of the facility. The facility was deemed fully covered when all of its premises were encompassed by a set of sequential swaths (SI Figure S1 provides an example), termed here a 'flyover.' A typical whole-site estimate was usually complete within approximately 10 minutes. During each flyover, Solution 1 detected multiple emission sources, which were summed together to calculate the total facility emission rate. Each whole-site flyover was treated as an independent estimation of the entire facility's emissions.

S1.2. Method Uncertainty

This analysis uses the relative error in controlled release data from recent peer reviewed studies, [22] and [23] to estimate method uncertainties for Solution 1 and Solution 2 respectively. In both scenarios, the controlled tests were performed under near-ideal condi-

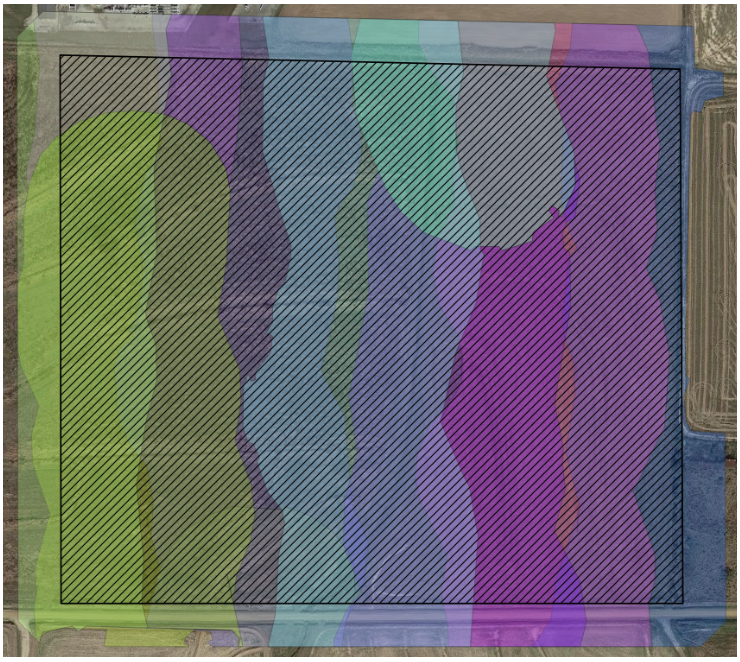


Figure S1. The facility boundary is the striped area with the black outline. Each swath pass represents a different color, this would represent one whole-site estimate.

tions and employed a single blind test methodology where the release location was known, but the release rate, including zero releases, were unknown.

For Solution 1 there were sufficient (857) controlled release points to divide controlled releases into two ranges (0 to 10 kg/h and 10 - 2500 kg/h) and develop one relative uncertainty model for each range. Uncertainty is higher for the <10kg/h bin in comparison to the >10kg/h bin. The error model was applied to each Solution 1 plume.

For Solution 2, there were 12 data points in the controlled test, and one relative error model was developed for all observed emission rates. The error model was applied to each Solution 2 observation.

For both methods, test conditions during controlled release testing were substantially different from conditions encountered on midstream facilities: All controlled releases consisted of a single emission point, isolated from other sources, and distant from structures that would disrupt the near-field wind transport. Additionally, weather conditions were constant (Solution 2) to near-constant (Solution 1) during controlled testing. This differs from midstream facilities, which often have complex configurations, including large structures that disrupt or distort the wind field, as well as overlapping and intermittent sources of emissions. Therefore, the uncertainty model used here should be considered a lower bound for uncertainty of each plume or observation.

It's worth emphasizing that uncertainty is considered for each estimate. In the case of Solution 1, uncertainty is applied to each plume, whereas for Solution 2, uncertainty is applied to each whole-site estimate (or each zonal estimate in the case of multiple flights). In Solution 1's methodology, all detection's identified in one whole-site flyover are combined (with associated uncertainty) to calculate the total facility emissions. Subsequently, the mean of all whole-site estimates was computed with Monte Carlo methods, propagating uncertainty through each calculation. Controlled test results showed these aerial methods have some bias in the mean. Therefore, when the uncertainty model is applied to the estimate reported by the solution, the mean value is shifted by that bias.

All confidence intervals (CIs) were computed from the empirical distributions using the 2.5 and 97.5 percentile values. The emission rates below are shown as a range *x [l to u]* where *x* is the mean of the distribution, *l* is the lower CI, and *u* is the upper CI.

S1.3. Statistical Approach

Comparisons between methods utilized three statistical tests with α = 0.05: the Kolmogorov-Smirnov (ks) test, the t-test, and the Wilcoxon signed-rank test. The two-sided KS test serves as a pivotal tool in gauging whether the two datasets originate from the same underlying distribution or if significant differences are present. Importantly, this test operates without requiring predefined assumptions about the underlying distribution's shape, ensuring that potential variations are not concealed.

The well-known two-sided t-test was employed to compare the means of emissions distributions obtained from Solution 1 and Solution 2. This test assumes that the data follows a normal distribution and facilitates an assessment of whether the observed mean values exhibit statistically significant disparities. Note that neither of the uncertainty models developed from controlled testing were normally distributed. The Wilcoxon sign-rank test was utilized to compare the medians of the emissions datasets. This test is particularly advantageous for small sample sizes and does not rely on specific assumptions about distribution shape.

S2. Facility Information

Exceptions to facilities during the baseline deployment:

- *Facility A and B:* Solution 2 performed one downwind flux of the whole facility. The study team analyzed all Solution 2 data and decided this estimate was the best estimate of the facility emissions.
- *Facility C:* Solution 1 did only one overflight, in the morning.
- *Facility G:* Solution 1 missed a portion of the site on the initial measurement day. The missing area was overflown at a later date.
- *Facility J:* Solution 1 flew over three times during the measurement period. All three were used in the measurement emissions check (MEC) calculation, but for comparison purposes the first two are used.
- *Facility L:* Solution 1 detected a large blowdown that was logged as a episodic event. This estimate was taken out of Solution 1's facility estimate.
- *Facility M:* Solution 1's morning flight captured the facility in a different operating state than other estimates and was not used for comparison.

With those exceptions, there are two Solution 1 measurements and one Solution 2 measurement for each facility. These adjusted estimates were averaged to produced one MEC for the facility. In the initial deployment, there were only three estimates available. In a previous study (Brown et al. [32]), all three estimates were treated as independent, introducing some bias to the Solution 1 results. For the analysis in this paper, we adjusted the methodology. We first averaged the two Solution 1 estimates and then combined this average with the Solution 2 estimate. This adjustment aligns with the End-of-Project (EOP) methodology, introducing equal bias to both methods.

Exceptions to facilities during the EOP deployment:

- *Facility A:* Solution 1 unable to fly on the EOP day and flew the next day - the compressor configurations was consistent between the two days.
- *Facility A and B:* Solution 2 flew the downwind measurement in two flux plane flights.
- *Facility C:* Solution 1 deployed a helicopter-mounted Light Detection and Ranging (LiDAR) system for EOP estimates rather than the fixed-wing platform used at all other facilities. Comparing the two platforms, the emission rate detection sensitivity of the helicopter platform is lower (flying lower and slower). However, as there is no controlled release study for the helicopter platform, quantification uncertainty estimates could not be adjusted.
- *Facility D:* A compressor start was noted in the morning, Solution 1's estimates during this time was not included in the analysis. Solution 2 was not flying during this time.

- *Facility E:* Solution 1, Solution 2 measured over two days. The same number of compressors were operating on both days. Solution 2 flew this facility in three different flux plane flights.
- *Facility G:* A compressor unit blew down - Solution 2 stopped measuring during this event (approx. 15 minutes), Solution 1's estimates made during this time were not included in the analysis.
- *Facility I:* The wind direction shifted during Solution 2 last flight causing this estimate to include enrolled and non-enrolled emissions. To correct for non-enrolled emissions, the emissions detected by Solution 1 in the non-enrolled section were subtracted to adjust the estimate.
- *Facility K:* Solution 1 unable to fly on the EOP day and flew the next day - the compressor configurations was consistent between the two days.
- *Facility N:* Solution 2 flew the downwind measurement in two flux plane flights.

Sources below Solution 1's detection sensitivity per facility:

- Facility A: Exhaust from heaters and flares
- Facility B: Exhaust from heaters and flares, fugitive leaks
- Facility C: Combustion exhaust from turbines and pneumatic emissions
- Facility D: Combustion exhaust from turbines
- Facility E: Fugitive leaks
- Facility G: Combustion exhaust from turbines
- Facility H: Fugitive leaks and exhaust from reboilers
- Facility I: Fugitive leaks
- Facility L: Fugitive leaks
- Facility M: Fugitive leaks and pneumatic emissions
- Facility N: Fugitive leaks, pneumatic emissions, combustion exhaust from turbines
- Facility O: Fugitive leaks and pneumatic emissions

S3. Statistical Tests

Tables S1 and S2 provide an overview of statistical tests for mean and paired estimates, respectively.

An approach commonly utilized for comparison of estimates is to ask whether the 95% CI of the estimates overlaps. To estimate the strength of this comparison, we use the empirical distribution of estimates to calculate the probability that the distribution with a smaller mean could be larger than the distribution with the larger mean – i.e. the overlap in the distributions. A symmetric (non-skewed) distribution compared to itself would have a score of 50%; this represents the maximum possible score. This metric is included in the columns labeled 'smaller mean > larger mean'. Scores near 50% represent similarity; lower sores represent less similarity.

Table S1. Statistical tests comparing mean estimates for the baseline and EOP deployments.

Facility ID	Statistical Test - Baseline (%) Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean	Statistical Test - EOP (%) Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean
A	0	0	0	6	0	0	0	17
B	1	1	1	43	0	0	1	35
C	0	0	0	4	0	0	0	0
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	1
G	0	0	0	29	0	0	1	33
H	0	0	0	0	0	0	0	8
I	0	0	0	30	0	0	0	1
J	0	0	0	0	0	0	0	0
K	0	0	0	6	0	0	0	0
L	0	0	0	0	1	1	1	42
M	0	0	0	6	0	0	0	0
N	0	0	0	32	0	0	0	0
O	0	0	0	20	0	0	0	0

If Solution 1 produced 10 estimates at one facility and Solution 2 produced 5, this would result in a total of 10 * 5 = 50 pairwise comparisons for that specific facility. In total, across all facilities, there were 773 such pairwise comparisons for the EOP.

Table S2 presents the comparison of individual estimates at each facility. Highlights of comparisons are:

- There is no observable correlation between the number of comparisons and the degree of agreement or disagreement between the methods.
- The highest agreement at one facility, based on the 2-sided ks test (α = 0.05), shows the methods agree in 40% (31 of 78) of comparisons.
- At another facility, based on the 2-sided t-test (α = 0.05) the methods agree in 29% of comparisons (20 of 68). At the same facility, the Wilcoxon test (α = 0.05) exhibits the highest agreement between the methods with 68% of comparisons (46 of 68).
- On average between all 773 pairwise comparisons, the methods display an overlap within the 95% CI 81% of the time. However, this does not imply equality, as the smaller *mean* estimate is only larger than the lager *mean* estimate 15% of the time.

Table S2. Statistical tests on all EOP individual comparisons

Facility ID	Num. of Comparisons	Statistical Test (%) Avg ks-test True	Avg t-test True	Avg Wilcoxon test True	smaller mean > larger mean (%)
A	24	0	4	33	29
B	68	37	29	68	37
C	12	0	0	0	4
D	44	0	0	0	0
E	72	10	3	11	17
G	45	27	20	49	28
H	78	40	19	42	26
I	69	12	12	14	17
J	64	5	3	3	12
K	40	0	0	0	0
L	24	38	33	58	36
M	52	0	0	0	3
N	57	0	0	0	0
O	124	0	0	0	3

S4. Baseline vs. EOP

Comparing the mean's of the methods between the two deployments, in aggregate, Solution 1 estimated a reduction of 740 [414 to 1260] kg/h in emissions from baseline to EOP, while Solution 2 estimated a reduction of only 40.4 [-783 to 836] kg/h. On a per-facility basis, both Solution 1 and Solution 2 estimated lower emissions at a total of 8 facilities from baseline to EOP.

Table S3. Facility estimates and variability between methods

Facility ID	Baseline Estimates (kg/h)		EOP Estimates (kg/h)		EOP Operating State Change ‡	Uncertainty Reduction *	
	Solution 1	Solution 2	Solution 1	Solution 2		Solution 1	Solution 2
A	203 [+18%/-14%]	376 [+46%/-49%]	284 [+6.3%/-5.7%]	355 [+43%/-37%]	0	1	1
B	529 [+47%/-26%]	578 [+46%/-49%]	211 [+7.8%/-6.7%]	226 [+29%/-30%]	0	1	1
C	122 [+36%/-24%]	192 [+32%/-33%]	42 [+19%/-15%]	103 [+25%/-24%]	0	1	1
D	79 [+31%/-19%]	33 [+37%/-44%]	103 [+15%/-11%]	657 [+42%/-43%]	0	1	0
E	393 [+20%/-14%]	1,291 [+25%/-25%]	764 [+8.3%/-6.8%]	1,178 [+30%/-29%]	0	1	0
G	63 [+32%/-22%]	47 [+34%/-24%]	73 [+6.8%/-6.6%]	80 [+36%/-34%]	1	1	0
H	43 [+7.9%/-4.2%]	125 [+42%/-42%]	59 [+7.9%/-6.9%]	79 [+33%/-32%]	1	0	1
I	44 [+47%/-26%]	36 [+53%/-52%]	41 [+8.3%/-7.3%]	28 [+35%/-33%]	0	1	1
J	14 [+36%/-23%]	27 [+30%/-34%]	5 [+22%/-17%]	9 [+29%/-26%]	0	1	1
K	68 [+17%/-13%]	102 [+39%/-35%]	17 [+11%/-11%]	102 [+30%/-30%]	0	1	1
L	1,068 [+40%/-22%]	459 [+43%/-48%]	512 [+18%/-13%]	496 [+30%/-27%]	0	1	1
M	61 [+36%/-26%]	90 [+23%/-26%]	63 [+6.6%/-5.9%]	148 [+42%/-38%]	1	1	0
N	38 [+41%/-25%]	45 [+46%/-39%]	16 [+11%/-11%]	130 [+35%/-31%]	0	1	1
O	270 [+33%/-22%]	364 [+36%/-37%]	66 [+4.6%/-4.1%]	136 [+29%/-27%]	0	1	1

‡ A 1 indicates that the operational state changed during the EOP measurement period, a 0 means the operational state was constant during the EOP measurement period.
* A 1 indicates uncertainty decreased from baseline to EOP, a 0 indicates uncertainty increased.

S5. Case Studies

S5.1. Facility N

Major equipment at this facility is listed below.

- Two (2) Natural gas fired turbines (16,000 horsepower [hp] each)

S5.2. Facility E

Major equipment at this facility is listed below.

- Two (2) Amine Treaters controlled by Thermal Oxidizer
- Fifteen (15) Compressor Driver Engines (four-stroke lean burn [4SLB], natural gas-fired)
- Two (2) Electric motor driven regen. gas compressors
- Two (2) Electric motor driven compressors to stabilize overhead gas stream
- Three (3) Generators
- Six (6) Dessicant Dehydrators
- One (1) Process/Emergency Flare
- Four (4) Heaters
- Six (6) Condensate and Water Tanks

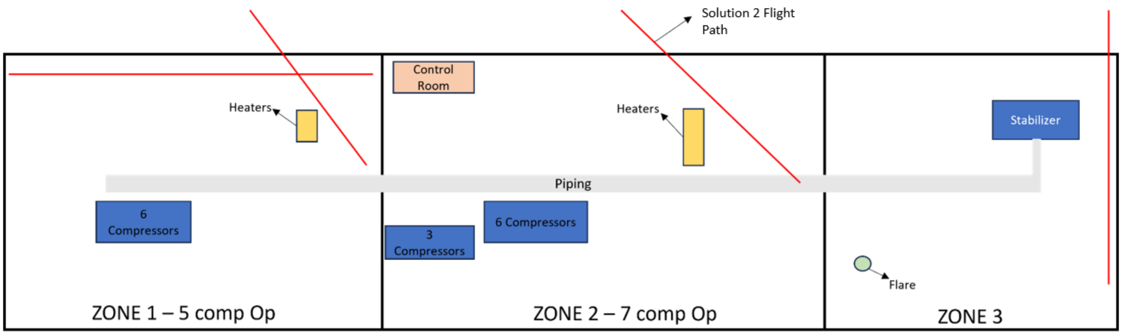


Figure S2. The black outlined areas represent the zones by which Solution 2 divided the facility and the red lines are the flux plane for each zone. Major equipment is laid out.

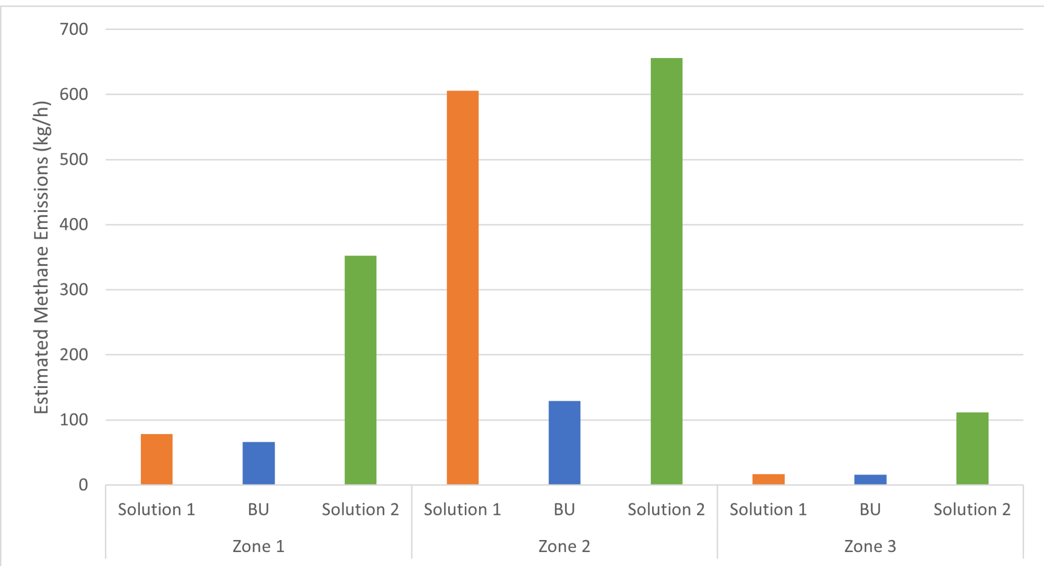


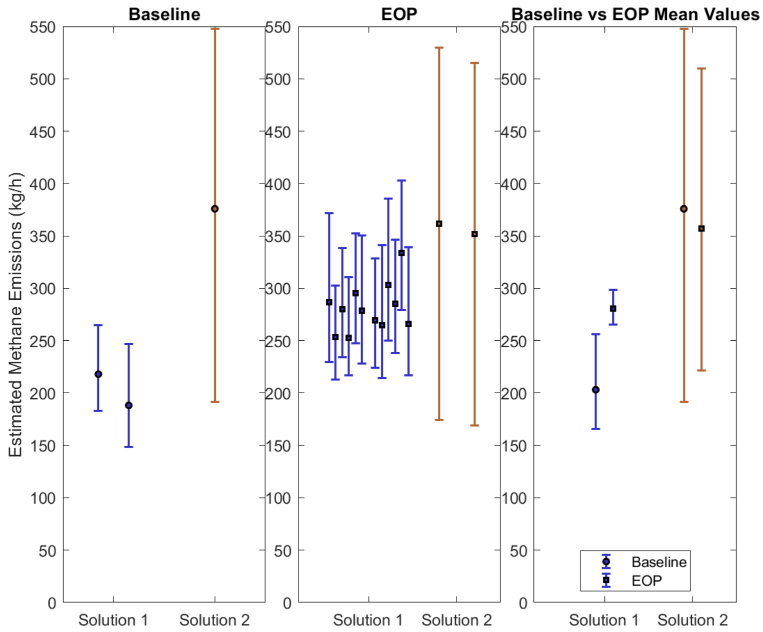
Figure S3. The Solution 1, BU, and Solution 2 estimate's are compared by zone.

Table S4. Solution 1 estimates per compressor compared to stack testing, crankcase vent (CCV) and rod packing vent (RPV) estimates.

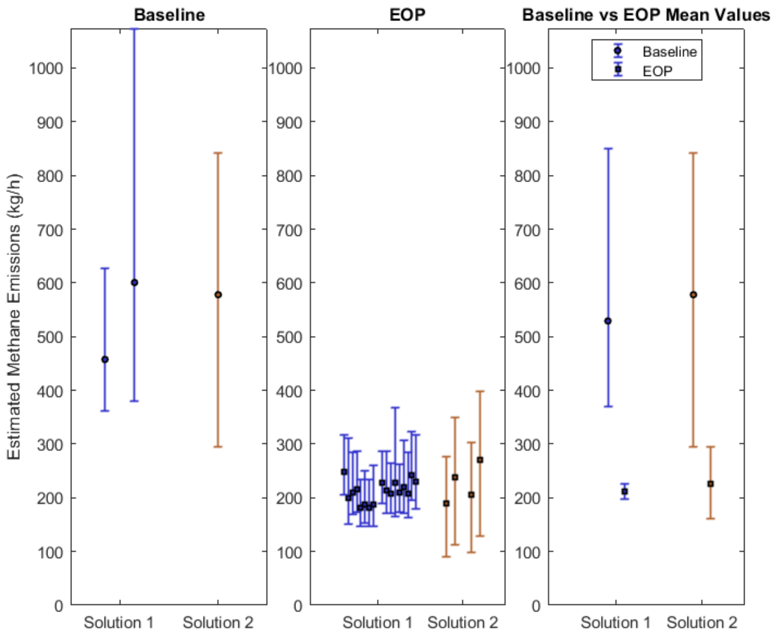
Compressor ID	Solution 1 Avg. Estimate (kg/h)	Stack Test Avg. Estimate (kg/h)	CCV Estimate (kg/h)	RPV Estimate (kg/h)
C-1	18.0	14.1	4.6	2.8
C-2	18.3	7.3	3.0	2.6
C-3	–	–	–	–
C-4	12.0	6.2	0.8	1.3
C-5	11.7	6.7	1.7	1.6
C-6	12.6	6.7	2.1	1.0
C-7	24.6	23.0	0.8	2.5
C-8	–	–	–	–
C-9	119.5	8.7	3.6	3.4
C-10	119.5	9.4	3.8	0.1
C-11	119.5	15.7	1.5	0.0
C-12	119.5	7.2	2.0	2.0
C-13	9.0	5.3	5.0	13.9
C-14	11.9	6.2	5.8	9.6
C-15	–	–	–	–

C-3, C-8 and C-15 were not operational on the EOP day

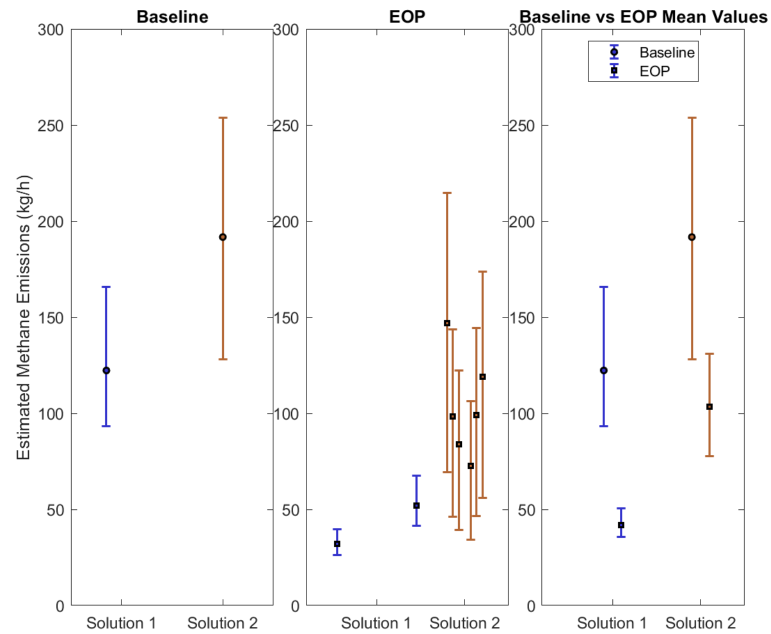
S6. All EOP Plots



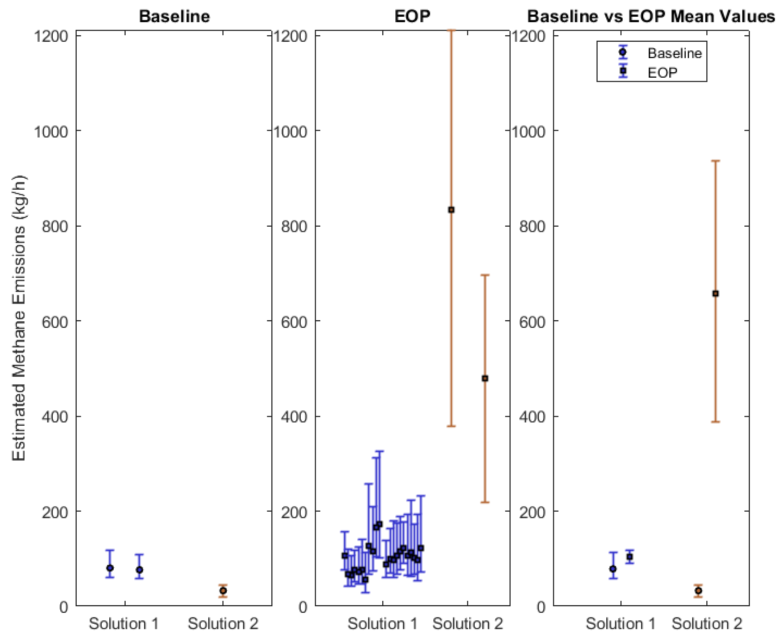
(a) Facility A baseline estimates versus EOP estimates by Solution 1 and Solution 2.



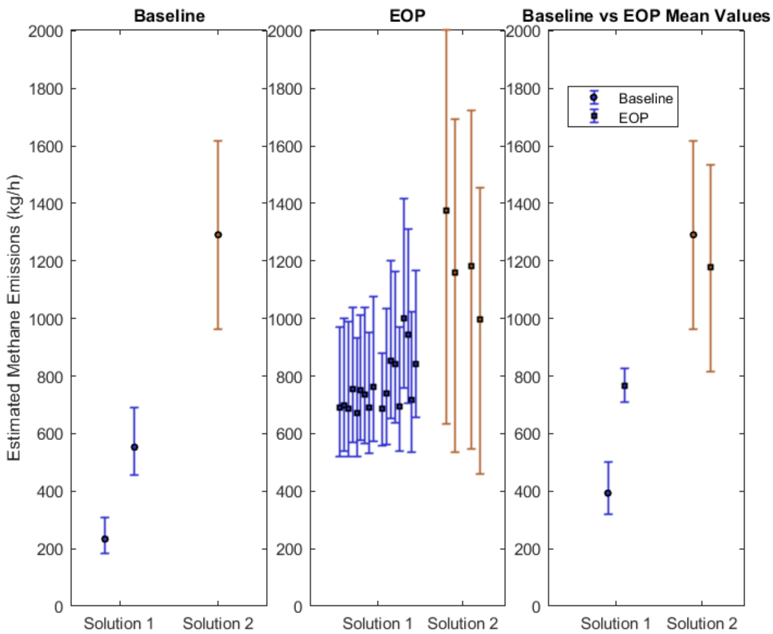
(b) Facility B baseline estimates versus EOP estimates by Solution 1 and Solution 2.



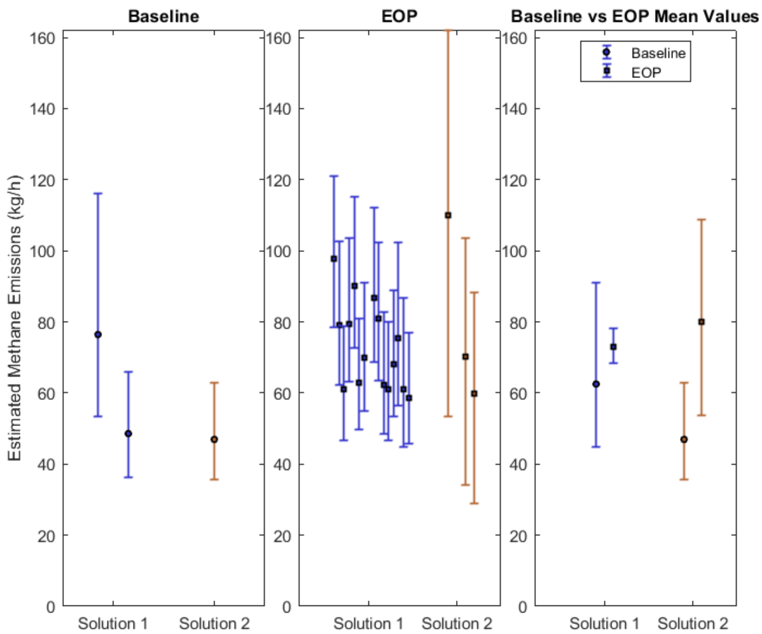
(a) Facility C baseline estimates versus EOP estimates by Solution 1 and Solution 2.



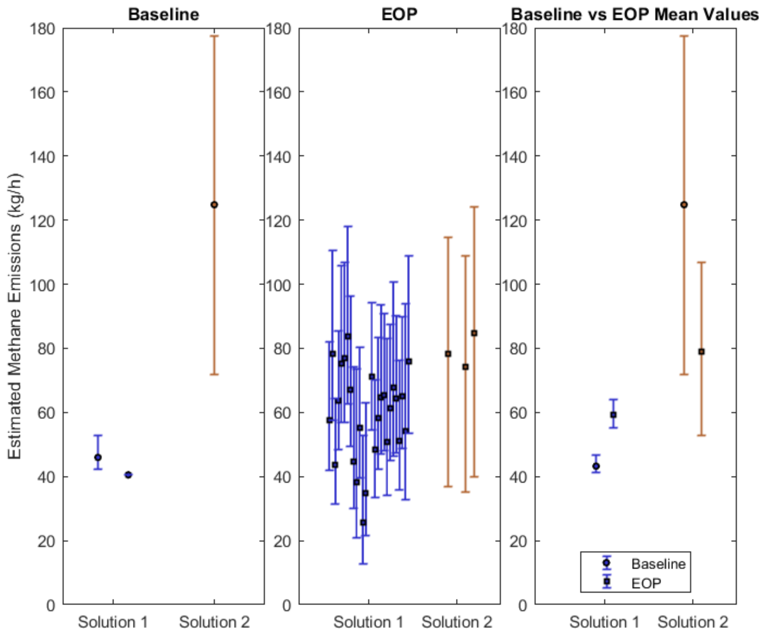
(b) Facility D baseline estimates versus EOP estimates by Solution 1 and Solution 2.



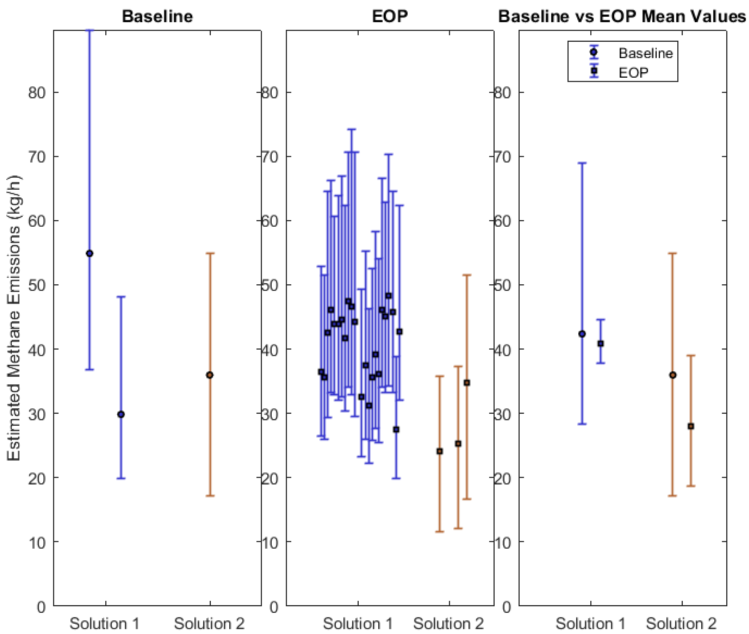
(a) Facility E baseline estimates versus EOP estimates by Solution 1 and Solution 2.



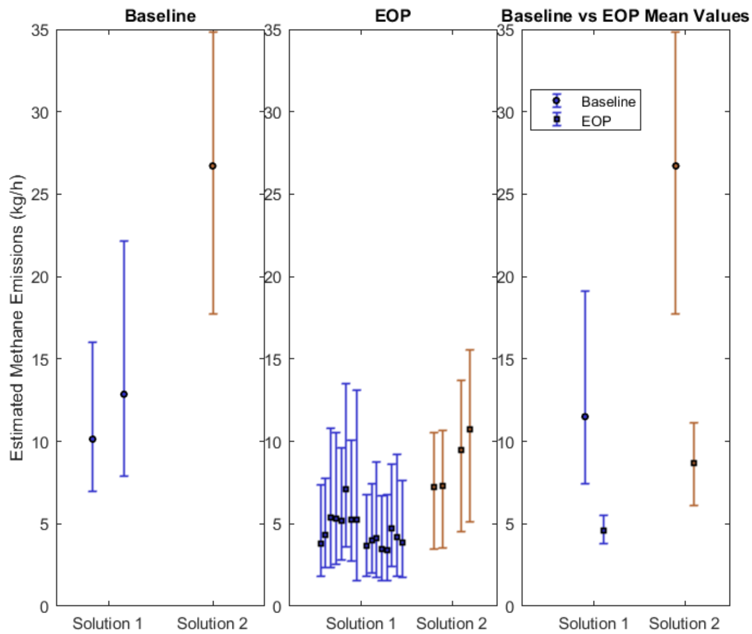
(b) Facility G baseline estimates versus EOP estimates by Solution 1 and Solution 2.



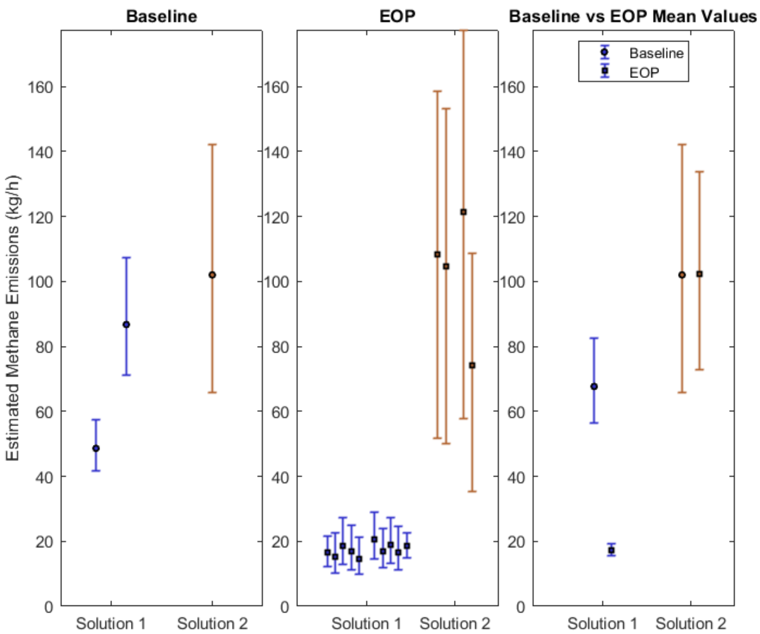
(a) Facility H baseline estimates versus EOP estimates by Solution 1 and Solution 2.



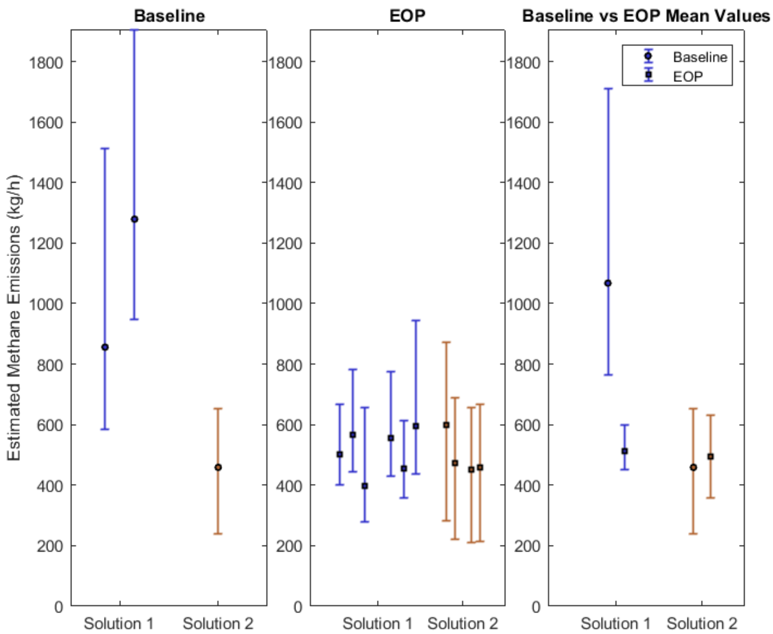
(b) Facility I baseline estimates versus EOP estimates by Solution 1 and Solution 2.



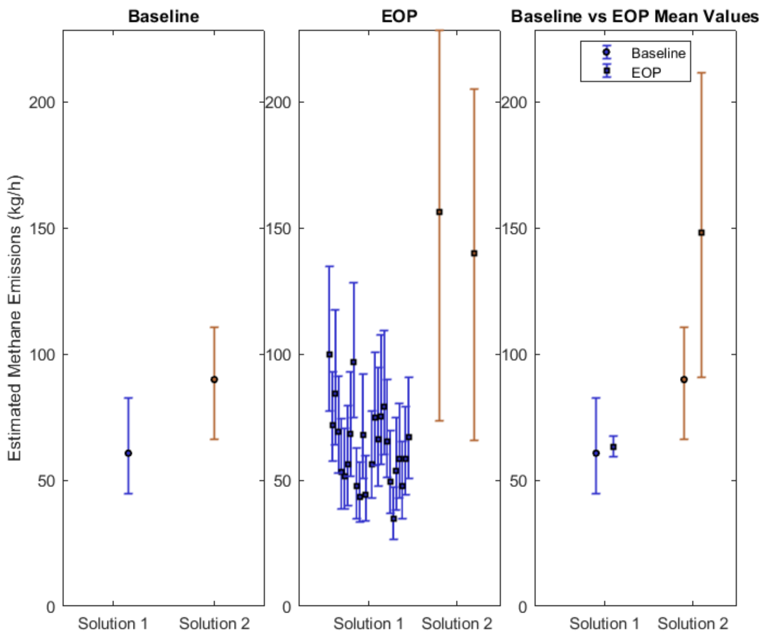
(a) Facility J baseline estimates versus EOP estimates by Solution 1 and Solution 2.



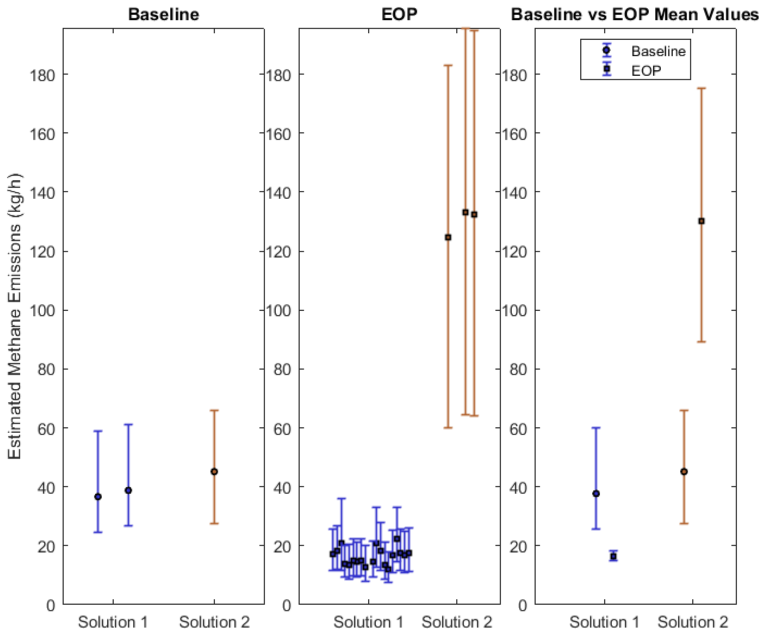
(b) Facility K baseline estimates versus EOP estimates by Solution 1 and Solution 2.



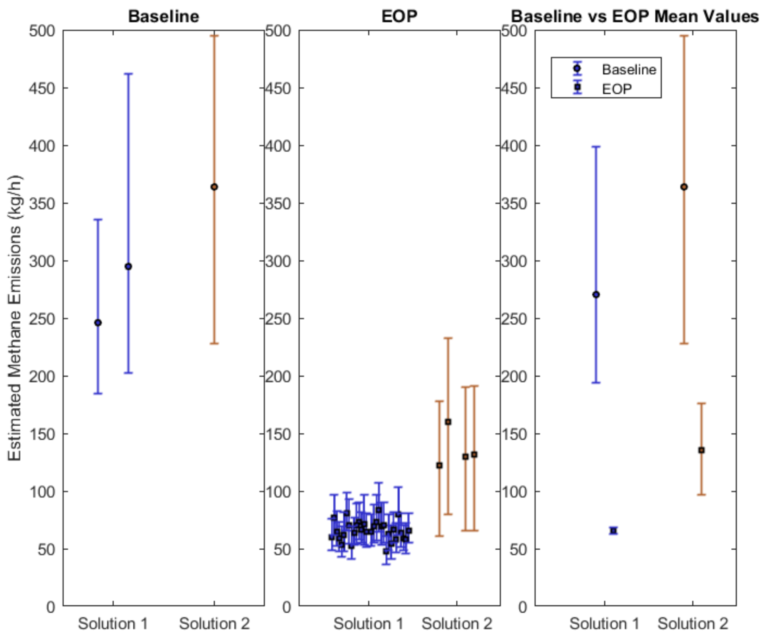
(a) Facility L baseline estimates versus EOP estimates by Solution 1 and Solution 2.



(b) Facility M baseline estimates versus EOP estimates by Solution 1 and Solution 2.



(a) Facility N baseline estimates versus EOP estimates by Solution 1 and Solution 2.



(b) Facility O baseline estimates versus EOP estimates by Solution 1 and Solution 2.