CASE STUDY 01 · PERMIAN BASIN ATTRIBUTION EVALUATION

Independent Evaluation of Methane Source Attribution

Issued by	Advect Labs
Worked scope	Delaware Basin · 27 Oct 2019 · 11 plumes (9 scored)
Ground truth	Cusworth et al. 2021 (ES&T Letters)
Contact	info@advectlabs.com

ADVECT LABS
METHANE ATTRIBUTION

A NOTE ON OPERATOR NAMING

Operators are referenced by masked label (Operator A, B, ...). Actual operators were identified from the public New Mexico EMNRD OCD wellbore database and the Texas Railroad Commission GIS well layers (§6); the mapping is held on file at Advect Labs. Masking is a courtesy in this public methodology release. In a client engagement, the evaluation is delivered with operators named.

Executive Summary

Public, operator-linked methane attributions are now routine, and they arrive from third parties an operator does not control: satellite portals, UN notifications, and journalists working the same public feeds (§2). The Oil and Gas Climate Initiative's response playbook, the de facto reference, records that no best practices exist for judging whether such an attribution is correct.

The product is a scorable confidence on a published attribution. Given a plume attributed to a specific asset, Advect Labs re-derives the attribution from public inputs the accuser did not use (NOAA wind reanalysis, public well registries, and a forward dispersion model) and reports a confidence over the candidate set that has been checked against ground truth. The independent inputs are what make that confidence meaningful: a number re-derived from the accuser's own pipeline could only echo it. At airborne resolution that independence does not show up as overturning the location (geometry already fixes the source under the plume), it shows up as a confidence that is scorable, and here mean-unbiased at the 150 m band (§5.1).

WHAT THIS CASE STUDY DEMONSTRATES AND DOES NOT

It demonstrates that dispersion physics can assign a scorable confidence to an existing attribution at airborne resolution: a posterior over the candidate set that, checked against published ground truth, is mean-unbiased at the 150 m correctness band on this sample (calibration-in-the-large, not reliability; §5.1). It does not refute the attributions it evaluates (at airborne resolution the location is not in contention), and refutation is not the claim here. The case where a verdict overturns a published location belongs to the transported-plume and satellite regimes, which this near-field dataset cannot exercise; that is named as the next step in §8, not demonstrated here.

The re-derivation is built as two layers that answer two different questions. **Location is geometry**: on near-field data the source sits under the plume, so the nearest mapped pad is the candidate, and across all nine scored plumes the nearest pad is the truth-nearest pad. No model improves on that. **Uncertainty is dispersion**: the forward dispersion surface does not pick the pad. It characterises the question around the geometry-chosen pad, scoring how much physical support the wind cone places on it versus the competing pads, enumerating those alternatives, and abstaining when the data is too coarse to name any single asset. The distribution, not the pad, is the model's defensible contribution, and it is the only part that can be checked against truth and recalibrated.

We work the method against Cusworth et al. 2021 because it is the rare public dataset where the answer can be checked: its supporting information publishes both the observed plumes and independently derived true source coordinates. We treat each plume as the observation, re-derive its attribution from independent public data, and score the result against the published source, which the algorithm never sees. Eleven plumes; nine scored, two excluded under a documented low-wind cutoff.

The worked example establishes what an operator must know before trusting a verdict. Location is recovered exactly (the geometry layer lands the right pad for all nine scored plumes, median 104 m from the published source), but at airborne resolution location is easy, and the method does not rest on it. What it rests on is a scorable confidence: because the verdict carries a posterior over the wind-cone candidate set, that posterior can be checked against ground truth, and here the reported confidence is mean-unbiased at the 150 m band on this sample (stated 0.88 against an empirical 0.89 over the nine scored plumes, a −0.01 gap). That is calibration-in-the-large, not reliability: it is sensitive to the 150 m threshold and, on the three contested plumes that carry it, holds by error cancellation rather than by confidence tracking correctness (§5.1). For the other six the wind cone admits only the one pad, so confidence is trivially complete. A point estimate could neither be scored this way nor disclose where it fails.

A calibrated number matters most where nothing else can confirm the claim. Most of the Cusworth sources are intermittent (detected on a single overpass and gone before a crew can reach them), so a field survey that finds nothing neither confirms nor refutes the attribution. When the source has vanished, an analytical, calibrated read of what the original detection and wind actually support is the only thing that speaks to the claim (§7).

This is a methodology demonstration on a small sample, not a basin-wide claim. At nine scored plumes, a single reclassification shifts an empirical rate by approximately 0.11. The figures are directional and are read that way throughout.

The rest of this document covers the attribution landscape (§2), the re-derivation method (§3), the results (§4) and their interpretation (§5), the per-claim evaluation record (§6), the operator response workflow it serves (§7), and limitations (§8).

02 The Public-Attribution Landscape

Three independent systems now publish operator-linked methane attributions on a rolling cadence.

The **Carbon Mapper portal** (data.carbonmapper.org) hosts plume detections from the Tanager-1 satellite, airborne campaigns, and other instruments: geolocated, rate-quantified, and overlaid on visible imagery, often clearly attributable to facilities in the basemap. More than 2,800 Tanager-1 methane sources had been published as of September 2025 (3,563 sources including CO_2).

The **UN's Methane Alert and Response System (MARS)**, operational since January 2024, aggregates detections from thirty-plus satellite instruments, notifies government focal points (and OGMP 2.0 member operators directly), and posts data publicly. More than 3,500 satellite methane alerts had issued across 33 countries as of 2025, the bulk in the oil-and-gas sector the system launched on.

The **EPA Super Emitter Response Program (SERP)**, finalized March 2024 under 40 CFR 60 Subpart OOOOb, creates a domestic notification channel for events at or above 100 kg/hr, with mandatory operator investigation timelines. A July 2025 Interim Final Rule delayed implementation to no earlier than 22 January 2027 and signalled substantive revision, so SERP's eventual scope is less settled than the standing international systems, but its notification machinery has already operated, and prior responses remain posted on EPA's ECHO explorer.

Journalists run their own analyses against the same data: BloombergNEF named West Texas Gas, Enterprise Products Partners, Energy Transfer, and Targa Resources in an August 2024 analysis, and AP has run comparable investigations.

When an attribution lands on an operator's facilities, the operator faces an immediate question about ownership and appropriate response. OGCI's **Satellite Methane Detection Response Playbook** (November 2025) is the de facto reference. Step 1 asks whether the operator owns facilities within roughly 100 m of the plume and notes that while the source usually sits within the plume's observed extent, plumes can "detach" and transport as a coherent unit before dispersing. It then states: **"there are no best-practices defined for evaluating the accuracy of the origin of an observed plume."** That admission appears in the context of transported plumes (the case for which back-projected dispersion modelling is most applicable, since recovering an upwind origin from a displaced plume is a transport-inversion problem).

A defensible answer to that question of asset ownership must:

1. Place the source at a specific identifiable asset, not a facility aggregate.

2. Carry calibrated posterior uncertainty over the candidate set, not point estimates with confidence as decoration.
3. Be reproducible from independent public inputs (satellite data, public wind reanalyses, public asset registries).
4. Document its failure modes explicitly, so the reader can see what would invalidate the attribution.

Those four properties are the criteria a defensible re-derivation must meet. They convert a third-party notification into a defensible internal record and, if the operator chooses, a defensible public reply.

03 The Re-Derivation Method

The method re-derives a source attribution from independent public inputs, and is worked here against Cusworth et al. 2021 (ES&T Letters), an AVIRIS-NG airborne campaign over the Permian in late 2019: 3,067 plume observations across 1,756 georeferenced sources, with rates, uncertainty, and source-type labels. Its supporting information publishes both plume coordinates and independently derived source coordinates, a ground truth against which the re-derivation can be scored. Without ground truth attribution-quality claims are unfalsifiable; this dataset makes a rigorous check possible.

We selected one flight line as the anchor: `GAO20191027t163712p0000`, acquired 16:37 UTC on 27 October 2019 over the Delaware Basin (Eddy County NM, with small overlap into Loving and Culberson Counties TX). It yielded eleven plumes: by Cusworth's labels, nine tank batteries, one well, one pipeline. HRRR winds at the acquisition hour ranged 0.45–3.11 m/s (mean 1.7 m/s) in a steady westerly regime.

The attribution is built as **two layers**, by design. Location and uncertainty are different questions, and the elaborate model is needed for only one of them. Treating them together (letting the dispersion surface both choose the pad and state the confidence) gives the model a location job that bare geometry does on this data at least as well, and buries the model's real contribution inside that step.

Layer 1 — Location (geometry). The candidate pad is the one whose centroid is nearest the plume centroid; neighbours are ranked by the same distance. On near-field airborne data the emission sits under the plume, so geometry is the cheapest and the most accurate locator available, and no model improves on it.

Layer 2 — Uncertainty (dispersion). The forward dispersion surface characterises the question around the geometry-chosen pad: how much physical support the wind places on it, which other pads remain plausible, and whether the surface is too diffuse to name any-one. This is a distribution, and a distribution is what supports a checkable, recalibratable confidence, the thing a bare pad cannot carry.

Inputs. Plume coordinates and emission rates from Cusworth 2021; independently derived source coordinates from the same paper, used only as ground truth (the algorithm never sees them); the NOAA HRRR wind field at the acquisition hour, via Herbie; and active-well surface locations from the NM OCD and TX RRC public registries, filtered to the acquisition epoch.

Stage 1 — Ingestion. Filter the Cusworth plume and source lists to the target flight line, yielding eleven plumes.

Stage 2 — Wind field. Sample HRRR at each plume centroid (UTM 13N) for 10-m wind and boundary-layer height, and assign atmospheric stability. We ran a **{Class A, Class B} stability ensemble centered on A**, consistent with daytime convective conditions over low-albedo desert and with the standard Turner stability table at the prevailing low winds.[1]

Stage 3 — Asset overlay and location. Wellbore locations are clustered into inferred pads (150 m agglomeration), each inheriting its wellbores' operator. The pad whose centroid is nearest the plume centroid is selected as the attributed pad (Layer 1); neighbouring pads are ranked by the same centroid distance.

Stage 4 — Forward dispersion surface. A steady-state Gaussian plume model is run forward from each candidate location and evaluated at the observed plume centroid: a forward-model likelihood over candidate sources, on a 5 km × 6 km grid at 100 m resolution, using Briggs (1973) Pasquill-Gifford σ formulae (Seinfeld & Pandis 2016). This is equivalent to back-projecting likelihood onto the candidate sources only under the uniform, steady-wind assumption used here; it is not an adjoint transport inversion. Uncertainty is bracketed across two stability classes and three wind-direction perturbations (±10°), averaging the six resulting probability surfaces.

Stage 5 — Confidence over the wind-cone candidate set. Confidence is a relative statement: how much the physics supports the geometry pad against the pads that genuinely compete with it. Those competitors are the pads the wind could plausibly have carried the plume from: upwind of the centroid, inside the dispersion cone (within 1000 m along-wind, the near-field source proximity, and $2\cdot\sigma_y$ cross-wind, the conventional ~95% plume envelope). The dispersion support is read on the geometry pad and normalised over that local set, optionally weighted by a Gaussian centroid-proximity prior ($\sigma = 120$ m, reflecting AVIRIS-NG source localization per Frankenberg et al. 2016). Normalising over the local set, rather than over every pad in the basin, is what makes the confidence a statement about the candidates actually in contention.[2]

Stage 6 — Scoring. The ranked candidates become an attribution tier: **SINGLE** (geometry pad holds > 80% of the cone-local confidence), **AMBIGUOUS** (the geometry pad plus one or two cone neighbours jointly exceed it), or **UNRESOLVED**. A `-PROXY` modifier flags sources whose published type lies outside the wellbore candidate set (e.g. a pipeline). An `UNRESOLVED-CALM` modifier excludes plumes at or below **1.0 m/s** wind, where the steady-state Gaussian transport assumption degrades; the threshold follows the EPA AERSCREEN/AERMOD calm convention and is applied inclusively as a conservative, documented cutoff. When the cone-confidence argmax falls on a pad other than the geometry pad, that disagreement is recorded and surfaced as an ambiguity flag rather than silently overriding geometry.

Each plume's output is a structured record (attributed pad, operator, cone-local confidence, tier, truth distance, truth-pad rank, candidate-set proximity floor, and the alternative set) rendered into the evaluation record shown in §6.

[1] Solar elevation at acquisition is ~36°, in the "moderate" insolation band; at the flight-line mean of 1.7 m/s and the sub-2 m/s winds covering most plumes, the literal Turner Table 2.1 reading is A–B, so the {A, B} ensemble is consistent with the table rather than a departure from it. Moderate insolation admits a Class C bracket only at the single ~3.1 m/s plume; that bracket is a scheduled sensitivity for future work.

[2] An earlier revision integrated the dispersion surface across all ~4,200 basin pads by Voronoi partition and normalised globally. A pad's share of a basin-wide partition scales with its Voronoi cell area, an artifact of local pad density rather than physical support, so every pad's mass collapsed toward zero and the method abstained on everything. Restricting normalisation to the wind cone (here a median of one and at most four pads) corrects this; the dispersion surface itself is unchanged.

04 Evaluation Results

Of the eleven plumes, **nine were scored**; two (0.50 and 0.45 m/s) were excluded as `UNRESOLVED-CALM` and enter no statistic. The nine scored plumes resolve to **five SINGLE, two AMBIGUOUS, one UNRESOLVED, and one UNRESOLVED-PROXY** (the pipeline source, outside the wellbore candidate set). Location accuracy is reported first and briefly, because it is the part the method does not depend on: at airborne resolution the input is already near-resolved, so geometry alone lands the right pad and the result confirms little beyond that.

Top-1 truth distance

Geometry-selected pad to Cusworth source, nine scored plumes.

THRESHOLD	PLUMES WITHIN
≤ 250 m	9 / 9
Median	104 m
Maximum	161 m

Truth-pad recall

Where the geometrically nearest pad lands in the ranked candidate order.

With pure nearest-pad location, the geometry-selected pad is the truth-nearest pad for every scored plume: median and worst truth-pad rank are both 1. The candidate-set proximity floor (median 104 m, max 161 m) equals the top-1 distances exactly: the attributed pad is already the closest pad to the truth, so the residual distance is the wellbore-to-source offset, not a ranking error. Cusworth's scored set is eight tank-battery plumes plus one pipeline plume against a wellbore-only candidate set, so for the tank plumes the algorithm finds the nearest wellbore on the same lease, a valid operator- and pad-level attribution, but a colocation claim rather than a source-type-matched measurement. The interpretation of these numbers, not the numbers themselves, is the subject of §5.

05 Interpretation

The results bear on whether a verdict from this method can be trusted. Two things matter that the attribution output cannot state on its own: whether the reported confidence is honest (§5.1), and whether that confidence stays honest as the input coarsens to satellite resolution (§5.2). Because location is geometry and geometry is exact on this data, there is no location contest to interpret. The dispersion model earns its place entirely in the uncertainty layer.

5.1 · Mean Calibration, Not Reliability

The §4 numbers describe geometric distance. They say nothing about whether the confidence the model reports is trustworthy. That question has two strengths, and this sample reaches only the weaker one. **Calibration-in-the-large** asks whether the mean stated confidence matches the mean empirical correctness — a single aggregate number. **Reliability** asks the harder question: are the candidates called 80%-likely correct about 80% of the time, the 50%-likely ones about 50%, and so on across the confidence range. The first needs only a mean; the second needs a spread of stated confidences with enough cases in each band to score. Nine plumes cannot supply that spread, and six of them carry a trivially saturated confidence — the wind cone admits a single pad, so the readout is ~1.0 by construction and tests nothing. What this section establishes is calibration-in-the-large: mean-unbiased at the 150 m band on this sample. Reliability across a confidence range is not demonstrable at this N and is not claimed.

The cone-local confidence is computed four ways (a Voronoi cell-integral and a pointwise-kernel read of the dispersion surface, each with and without the centroid prior) and all four are scored against ground truth. The framework's job is to let an evaluator select and check a readout, not to assert one a priori; on this sample the prior-weighted Voronoi read is the one that comes out mean-unbiased. Scoring the geometry top-1 pad as correct within 150 m of the Cusworth source (the pad-clustering radius), across all nine scored plumes (abstention is a tier outcome and does not suppress the nearest-pad call, so the two tier-abstaining plumes (§4) remain in this denominator), that variant gives:

QUANTITY	VALUE
Mean stated confidence (selected variant)	`0.88`
Empirical top-1 correctness	`0.89 (8 / 9)`
Mean-calibration gap (stated minus empirical)	**`−0.01 — mean-unbiased at the 150 m band on this sample`**

The denominator is all nine scored plumes; the two tier abstentions are reported separately in §4 and are not dropped here. The single miss at the 150 m band is **P01616**, the §6 worked example: its geometry top-1 pad is the truth-nearest pad (rank 1) but sits 161 m from the published source, rank-correct but band-incorrect, making it the one plume the 8/9 counts against.

The −0.01 is a property of the 150 m band, not of the method. "Correct" here means the geometry top-1 pad lands within 150 m of the Cusworth source — a defensible threshold, but a chosen one, and the gap is sensitive to it. The mean stated confidence (0.88) does not move when the band moves; only what counts as a hit does. Sweeping the band across values a reviewer could equally defend:

CORRECTNESS BAND	EMPIRICAL TOP-1	MEAN-CALIBRATION GAP
100 m	`0.44 (4 / 9)`	`+0.44 (overconfident)`
150 m	`0.89 (8 / 9)`	`−0.01 (mean-unbiased)`
200 m	`1.00 (9 / 9)`	`−0.12 (underconfident)`

The gap flips sign across the range: tighten the band to 100 m and the method reads overconfident, loosen it to 200 m and it reads underconfident, and the −0.01 sits at the crossing. This is the §5.2 move applied here — the headline number is contingent on a threshold, and the honest disclosure is to show that contingency rather than report the single band at which the gap happens to vanish.

The aggregate match is error cancellation, not tracking. The −0.01 is carried entirely by the three plumes whose cone holds two to four pads (P01615, P01616, P01617); the other six are saturated and carry no calibration information. Across that contested set the three stated confidences are 0.986, 0.937, and 0.000, and they do not order the outcomes. P01616 at 0.937 is the one band miss — high confidence on a wrong call (overconfident). P01617 is stated at 0.000 yet is correct: there the dispersion argmax fell on a competing pad, the geometry-vs-physics disagreement Stage 6 flags, leaving the geometry pad — the truth-nearest one — at zero cone-local mass (underconfident). P01615 at 0.986 is correct. The aggregate near-match is the overconfident wrong call and the underconfident right call netting out, not the confidence ranking the cases correctly. Three contested plumes, ordered this way, cannot establish that the confidence tracks correctness; the contested-set ordering is uninformative at this N.

The other three variants are underconfident on this scope (gaps of −0.21 to −0.41), and the calibration check is what tells them apart. Three cautions, all consequences of the near-field data and the procedure, fix what this is. First, **the variant is selected in-sample**: the readout is chosen against the same nine plumes it is then scored on, so the −0.01 is an in-sample fit, not an out-of-sample guarantee; confirming it requires a held-out benchmark (§8). Second, **the wind cone is small**: for six of the nine scored plumes it admits only the geometry pad, so the confidence is trivially complete and carries no calibration information, and the gap rests on the three contested plumes above. Third, **N is small**: a single reclassification moves the empirical rate ~0.11, so every figure here carries a wide error bar.

This is what the method exists to provide. A confidence built as a distribution over the candidate set can be checked against ground truth and recalibrated when a gap appears. A point estimate offers no equivalent: it carries no physically grounded, recalibratable distribution against which empirical correctness can be measured. A heuristic baseline can emit a number (one can always impose an inverse-distance score), but that score has no transport-physics basis and no principled mechanism to widen as the measurement coarsens (§5.2).

5.2 · Calibration Under Degradation

The calibration above is validated only at airborne (AVIRIS-NG, ~100–150 m localization) resolution. At satellite resolution the confidence is **unvalidated**: this study carries no satellite ground-truth benchmark, so nothing here certifies a satellite-resolution verdict. To characterise the direction of the effect without leaving the Cusworth ground truth, we degrade the airborne centroids with controlled, isotropic Gaussian noise ($\sigma_{noise} \in \{0, 250, 500, 1000, 2000, 3500\}$ m, 30 realizations per plume per level) and re-run.[3] This is a stress test of the architecture, not a substitute for validation against real satellite plumes.

To place the noise levels against real instruments, name them by what they actually deliver. EMIT images at roughly 60 m ground sampling distance and Carbon Mapper's Tanager-1 at roughly 30 m, but ground sampling distance is pixel size, not source-localization accuracy, and the two differ by a large factor. This study's own airborne anchor sets the scale: AVIRIS-NG has a ground sampling distance of ~1–8 m, yet its source localization is taken at ~100–150 m (Frankenberg et al. 2016), localization 20–100× coarser than the pixel. Applying that same gap to the spaceborne imagers places their source localization well above 100–150 m: coarser than AVIRIS-NG, not finer, and squarely inside the degraded levels swept here (250 m and beyond). The coarser, area-flux feeds (TROPOMI and the km-scale instruments aggregated by MARS) fall further still. So the operational question (can this be trusted on the satellite notification that just arrived?) lands in the degraded, unvalidated regime, not near the airborne one: a real Tanager-1, EMIT, or MARS detection is exactly the input the sweep shows the confidence is not yet calibrated for. The one public-portal product that does sit at the calibrated resolution is Carbon Mapper's airborne (AVIRIS-NG) plumes, a genuine deployment surface that carries no public ground truth, which is precisely where a calibrated read earns its keep.

Location stays nearest-pad on the noisy centroid; the candidate set and ground truth are unchanged. Location accuracy degrades with noise (the pad nearest a badly-mislocated centroid is increasingly not the truth pad) but that is backdrop. The headline is the uncertainty layer: does it abstain when it should, and does its confidence stay honest?

Abstention behaves as intended. The fraction of plumes the tier logic marks UNRESOLVED climbs from 0.22 at native resolution to roughly half by satellite-scale noise, because the cone widens and no local cluster reaches the 0.80 threshold. "The data is too coarse to name a single asset" is the honest response, and the method gives it.

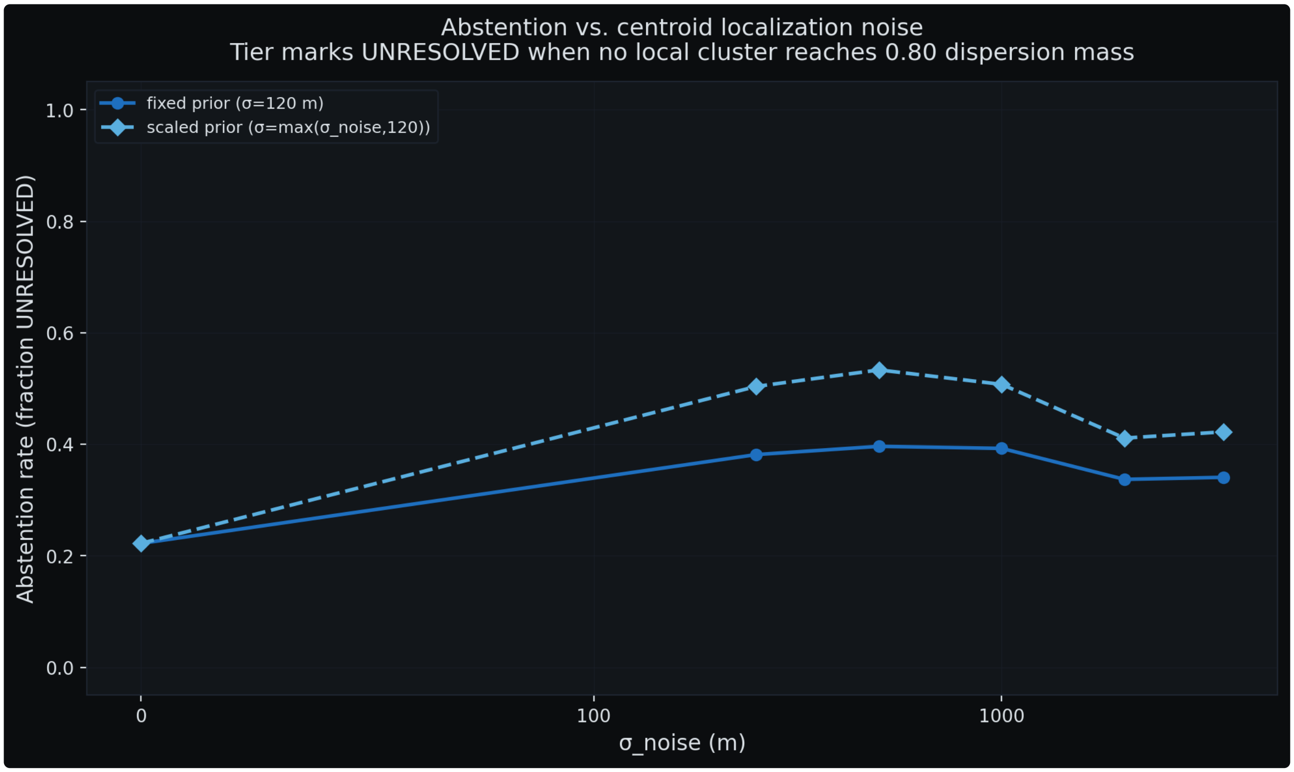

FIGURE 1 Abstention under synthetic input noise. The fraction of scored plumes the tier logic marks UNRESOLVED climbs from ~0.22 at native airborne resolution toward ~0.5 at satellite-scale noise: as the cone widens, no local cluster reaches the 0.80 confidence threshold and the method declines to name a single asset.

Calibration isolates a failure mode that distance ranking cannot have, because distance ranking carries no confidence to miscalibrate. We ran the prior-weighted confidence two ways: a **fixed prior** (σ = 120 m, tuned for airborne input) and a **scaled prior** (widening as the assumed input coarsens):

Σ_{NOISE} (M)	EMPIRICAL CORRECT	FIXED PRIOR (GAP)	SCALED PRIOR (GAP)
0	0.89	−0.01	−0.01
250	0.54	+0.28	+0.17
500	0.24	+0.63	+0.43
1000	0.15	+0.75	+0.55
2000	0.03	+0.92	+0.76

With a fixed prior, stated confidence stays high while empirical correctness collapses: by σ = 2000 m the model is asserting ~0.95 confidence on attributions correct 3% of the time. This is the failure mode of a pipeline tuned for airborne input and applied unmodified to coarser satellite data. The scaled prior is less overconfident at every degraded level (the gap is roughly a third smaller throughout) but it does **not**, on this scope, restore calibration: a substantial overconfidence gap remains all the way down. What the scaling changes is the magnitude of the drift, not its existence. The honest reading is narrow and worth stating exactly: the method's confidence is calibrated at the resolution it was built for, degrades into overconfidence as the input coarsens, and a resolution-aware prior slows that degradation without curing it. A satellite-resolution verdict from this pipeline, today, would carry a confidence that needs further widening before it can be trusted, and the value of the check is that it measures this rather than asserting calibration it does not have.

The diagnostic value is the existence of this check, independent of the specific numbers. A pipeline that emits a distribution over candidates can be validated against ground truth and its prior adapted as input characteristics change; that property is not available when the output is a single point estimate.

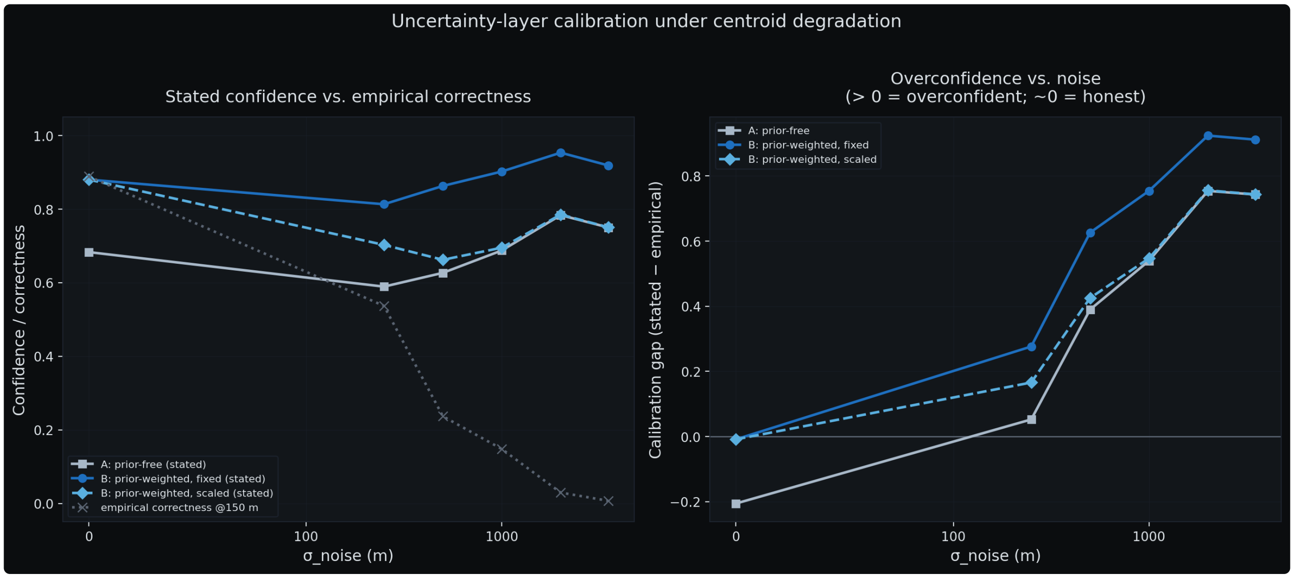

FIGURE 2 Calibration under degradation. Left: stated confidence (prior-free, and prior-weighted under fixed and scaled priors) against empirical correctness as input coarsens. Right: the calibration gap. The fixed prior drifts further into overconfidence than the scaled prior at every noise level, but neither restores calibration. The scaled prior slows the drift without curing it.

[3] Noise levels are illustrative regime spans, not instrument-specific localization claims. Real retrievals carry anisotropic, wind-dependent, spatially correlated error that an isotropic Gaussian model does not capture; the sweep shows the architecture responds to the kind of degradation those instruments produce, and is not a substitute for validation against real satellite plumes.

06 Per-Claim Evaluation Record

The pipeline renders each plume into a structured evaluation record: the artifact an operator's regulatory, ESG, or legal lead files alongside their response to a notification. It addresses six standard fields: source identification, quantification methodology, uncertainty analysis, attribution rationale, calibration note, and data provenance.

Below is the record for **plume P01616**, chosen because it exercises the uncertainty layer rather than a trivial case: it is a SINGLE-tier attribution where the wind cone admitted four competing pads, the dispersion physics placed 93.7% of its cone-local confidence on the geometry-selected pad, and that pad is the truth-nearest pad. It sits 161 m from the published centroid (the wellbore-to-tank-battery offset), so it is also a clean illustration of the residual geometric floor.

EVALUATION RECORD · METHANE SOURCE ATTRIBUTION

Plume P01616

Issued by: Advect Labs (methodology artifact, Case Study 01)
Source flight: `GAO20191027t163712p0000`
Acquisition: 2019-10-27 16:37:12 UTC
Detection rate: 142 ± 74 kg/hr (Cusworth Q_{plume} retrieval; 1σ as reported)

---

01 Source Identification

Cusworth source identifier	`P01616`
Cusworth source-type label	tank
Plume centroid (WGS84)	`32.2548°N, 104.0703°W`
Cusworth published source coord	(masked; 161 m from attributed pad)
Attributed asset	Operator C, Pad C-007 (API redacted)
Attributed operator	Operator C
Jurisdiction	New Mexico OCD (Eddy County, FIPS 015)

02 Quantification Methodology

Emission rate is taken directly from Cusworth et al. 2021 (Q_{plume} = 142 kg/hr, σ as published); Advect Labs does not modify the published rate, as this record addresses attribution rather than flux quantification. Wind: NOAA HRRR 17Z analysis (2019-10-27), 3 km grid, interpolated to centroid in UTM 13N at 1.98 m/s. Location is nearest-pad geometry. Confidence is the forward-model dispersion likelihood on the attributed pad (Briggs 1973

PG σ formulae, Seinfeld & Pandis 2016, {A, B} ensemble centered on A, on a 5 km × 6 km grid at 100 m, averaged over six scenarios of three wind directions × two stability classes) normalised over the wind-cone candidate set (four pads here) and weighted by the centroid prior (σ = 120 m).

03 Uncertainty Analysis

COMPONENT	TREATMENT	MAGNITUDE
Wind direction	Ensemble bracket	`±10°`
Atmospheric stability	PG-class ensemble	`A / B`
Plume centroid localization	Gaussian prior	`σ = 120 m`
Source coordinate (Cusworth)	Reported (no published σ)	—

DIAGNOSTIC		VALUE
Cone-local confidence (selected variant)		`0.937`
Pads in wind-cone candidate set		`4`
Attribution tier		`SINGLE`
Top-1 truth distance		`161 m`
Truth pad rank		`1`
Candidate-set proximity floor		`161 m`
Failure-mode classification	Rank-correct (truth pad ranked 1); outside the 150 m band (161 m), the single band miss in §5.1's 8/9	

The candidate-set floor (161 m) matches the top-1 distance exactly: geometric attribution is at the limit of the wellbore-location data for this plume. Refinement below 161 m would require an added asset layer (operator tank-battery polygons or pipeline rights-of-way), relevant here because the published source type is `tank`, which the wellbore-only candidate set does not contain directly.

04 Attribution Rationale

P01616 was detected under a 1.98 m/s wind. The wind cone admitted four candidate pads; the forward dispersion surface placed 93.7% of its cone-local confidence on Pad C-007 (Operator C), a single-pad attribution by tier criterion, with the truth pad ranked first. The Cusworth source type is `tank`; the candidate set is wellbore-derived, so the attributed wellbore is colocated with (not identical to) the responsible tank battery. This is the colocation regime described in §4: correct at the operator and pad level, at the limit of geometric precision from a wellbore-only set.

05 Calibration Note

The 0.937 confidence reflects the model's support under the documented uncertainty inputs. P01616 is rank-correct (the truth pad ranks first) but at 161 m it falls just outside the 150 m correctness band, making it the single band miss among the nine scored plumes (§5.1). A high stated confidence on a band miss is exactly the kind of local overconfidence the calibration check exists to surface; in aggregate the contested cases still balance to a −0.01 gap (mean stated 0.88 against 0.89 empirical), calibrated on this sample but bounded by a wide small-N error bar and selected in-sample (§5.1). **A consumer should read 0.937 as a high-confidence attribution at the operator and pad level whose residual 161 m sits at the wellbore-geometry limit, not as a sub-150 m guarantee, and not yet validated at satellite resolution (§5.2).**

06 Data Provenance

SOURCE	IDENTIFIER / VERSION	USE
Cusworth 2021 plume list	`DOI 10.1021/acs.estlett.1c00173, SI XLSX`	Plume coords, rate, σ
Cusworth 2021 source list	`Same DOI, SI XLSX`	Published source location, type
NOAA HRRR	`2019-10-27 17Z analysis (F00), via Herbie / NCEI`	Wind field, PBLH
NM EMNRD OCD wells	`Statewide shapefile, retrieved 2026-04-24, filtered to first-production ≤ 2019-10-27`	Candidate pad set
TX RRC GIS well layers	`Loving / Culberson county shapefiles, retrieved 2026-04-24, filtered identically`	Candidate pad set (TX)
EPSG:32613 (UTM 13N)	–	All spatial operations

Note: the candidate set is reconstructed by filtering on first-production date to match the acquisition epoch, since the retrieved shapefiles include later-permitted wells.

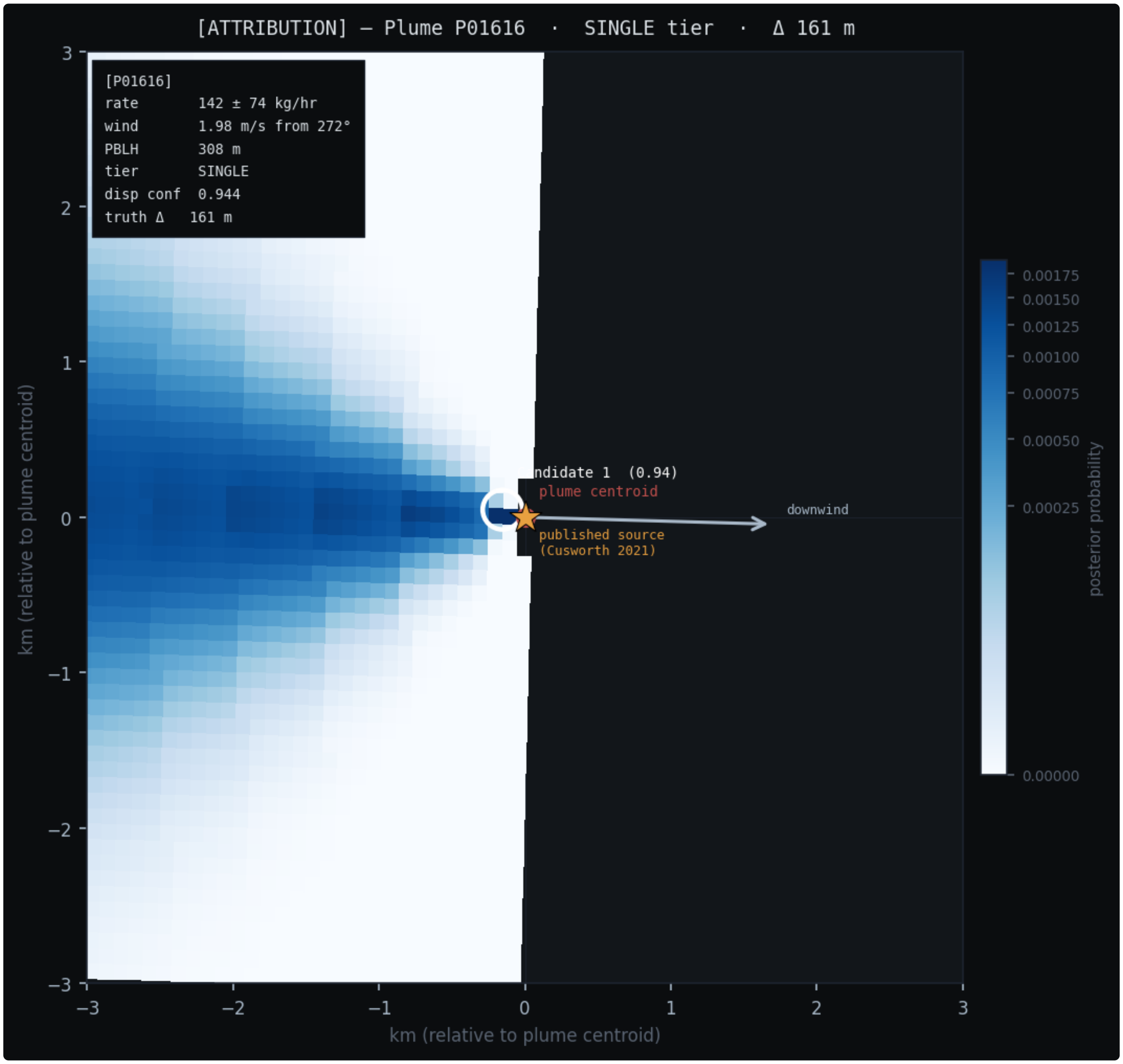

FIGURE 3 Spatial detail for plume P01616. The dispersion posterior surface (blue intensity) with the attributed pad (Operator C, Pad C-007) and nearby candidate wellbore locations. The truth coordinate sits 161 m from the attributed pad, at the limit of wellbore-location precision for a `tank`-type source matched against a wellbore-only candidate set.

07 Integration With the Operator Response Workflow

This evaluation is one component of an operator's response, not the whole of it. OGCI's playbook defines six steps: (1) initial assessment and location review, (2) source investigation and identification, (3) repair prioritization, (4) post-mitigation validation, (5) response preparation and documentation, (6) continuous improvement.

Advect Labs delivers the independent re-derivation behind **Steps 1 and 2**. The record answers Step 1's first decision point ("do you own facilities within ~100 m of the plume?") with a specific attributed asset, a documented confidence over the candidate set, and an explicit calibration statement. Step 2's order-of-magnitude analysis is a rate-capacity screen; the attribution confidence is spatial and complementary, directing the desktop and field team to the asset cluster most likely responsible.

The case where this matters most is the one field work cannot close on its own. The Cusworth sources are predominantly intermittent: nine of eleven have published persistence f ≤ 0.143, detected on a single overpass. An intermittent source is gone by the time a crew arrives, so a field survey that finds nothing neither confirms nor refutes the attribution. Reconciling a public claim with an empty field result then requires an analytical evaluation of what the original detection and wind actually support, which is what this layer provides. Steps 3–6 sit downstream and are the operator's, drawing on their maintenance records, field teams, and reporting systems; the record slots in as a documentation input, not a substitute.

The playbook's own framing that no best practices exist for evaluating plume origin accuracy is the gap this case study addresses. The calibration analysis in §5 is the basis on which a regulatory or ESG lead can judge whether the approach is rigorous enough for their use.

08 Limitations

Established. A two-layer re-derivation pipeline (location by nearest-pad geometry, uncertainty by forward-model dispersion over a wind-cone candidate set) worked against published ground truth on one well-characterised Delaware Basin flight line. The geometry layer recovers the true source for all nine scored plumes (median 104 m, max 161 m); the uncertainty layer attaches a confidence that can be scored against truth and is calibrated on this sample (−0.01 gap), abstains when it should, and degrades into measurable, disclosed overconfidence as the input coarsens. None of those outputs can come from a point estimate.

Not established. Nine scored plumes is not a basin-wide claim, and the calibration gap carries an error bar wide enough that one reclassification moves the empirical rate ~0.11; with the wind cone admitting a single pad for six of the nine, the calibration rests on just three contested cases whose stated confidence does not track their correctness, so the aggregate match is error cancellation rather than ranking (§5.1). The wind regime was a steady ~1.7 m/s westerly; variable or stagnant regimes are not characterised. The degradation sweep uses synthetic isotropic Gaussian noise, which does not capture real instrument error structure, and it shows the confidence is not yet calibrated at satellite resolution; a resolution-aware prior slows the overconfidence but does not cure it. The re-derivation here also shares the accuser's plume observation, bringing independence only in the wind, asset, and transport inputs; a fully independent second opinion would re-detect the plume from separate data as well. And the check runs against a benchmark where ground truth exists (near-field airborne plumes sitting on their sources), so it does not yet demonstrate a verdict on a real public attribution where no ground truth is available, nor the transported-plume regime (plume displaced from source) in which dispersion back-projection would be expected to contribute to location as well as uncertainty, and which this near-field dataset does not contain. A resolution-aware abstention policy (sweeping the SINGLE threshold as a function of input localization, separately from the prior) is the natural next check: it could convert overconfident satellite calls into honest abstentions and turn graceful degradation into a demonstrated property, but it is not run here.

Scope of delivery. Independent, calibrated assurance on a published methane attribution: from a plume and the asset it has been attributed to, to a defensible confidence over the candidate set, rendered into the per-claim record of §6. On near-field airborne detections the verdict is a calibrated confidence on the attributed location; the stronger verdict that overturns a published location is a transported-plume and satellite capability, not demonstrated here. Advect Labs does not own or replicate the detection layer (TROPOMI, EMIT, Carbon Mapper / Tanager-1, GHGSat, Kayrros, AVIRIS-NG), and does not, in this scope, originate attributions of its own: it re-derives and tests the attributions others publish.

Engagement. The intended reader is the regulatory affairs, ESG, sustainability, or measurement lead at an operator whose assets have been, or could be, named in a public methane attribution. The conversation starts at `info@advectlabs.com`.

References

PRIMARY LITERATURE

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W., … Miller, C. E. (2021). Intermittency of large methane emitters in the Permian Basin. ES&T Letters, 8(7), 567–573. DOI: 10.1021/acs.estlett.1c00173. AVIRIS-NG plume/source lists; primary input and ground truth.

Cusworth, D. H., et al. (2022). Multiscale methane measurements at oil and gas facilities reveal necessary framework for emissions accounting. PNAS, 119(38). 150 m pad-clustering convention.

Briggs, G. A. (1973). Diffusion Estimation for Small Emissions. ATDL Contribution 79, NOAA. Pasquill-Gifford σ_y formulae.

Pasquill, F. (1961). The estimation of the dispersion of windborne material. Meteorological Magazine, 90.

Turner, D. B. (1970). Workbook of Atmospheric Dispersion Estimates. EPA. Table 2.1 atmospheric stability assignment.

Seinfeld, J. H., & Pandis, S. N. (2016). Atmospheric Chemistry and Physics (3rd ed.). Table 18.2 σ formulae.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., … Green, R. O. (2016). Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region. PNAS, 113(35). AVIRIS-NG centroid localisation.

DATA SOURCES

NOAA HRRR. 3 km hourly CONUS analysis, 2019-10-27 17Z, via Herbie / NCEI. Wind field and planetary boundary-layer height.

New Mexico EMNRD Oil Conservation Division. Statewide wells shapefile. emnrd.nm.gov/ocd. Candidate pad set (NM).

Texas Railroad Commission. GIS well layers, Loving and Culberson counties. rrc.texas.gov. Candidate pad set (TX).

RESPONSE-WORKFLOW AND PUBLIC-ATTRIBUTION REFERENCES

OGCI (2025). Satellite Methane Detection Response Playbook. Oil and Gas Climate Initiative, November 2025.

Carbon Mapper data portal. data.carbonmapper.org. Tanager-1 and airborne plume detections.

UNEP IMEO. Methane Alert and Response System (MARS). methanedata.unep.org.

EPA. Methane Super Emitter Program, 40 CFR 60 Subpart OOOOb. Implementation extended to 2027-01-22 per 90 FR 35966. EPA AERSCREEN User's Guide, EPA-454/B-15-005. 1.0 m/s calm-wind convention.