AI for Carbon Accounting

Dan Cusworth

CCAI Virtual Summer School

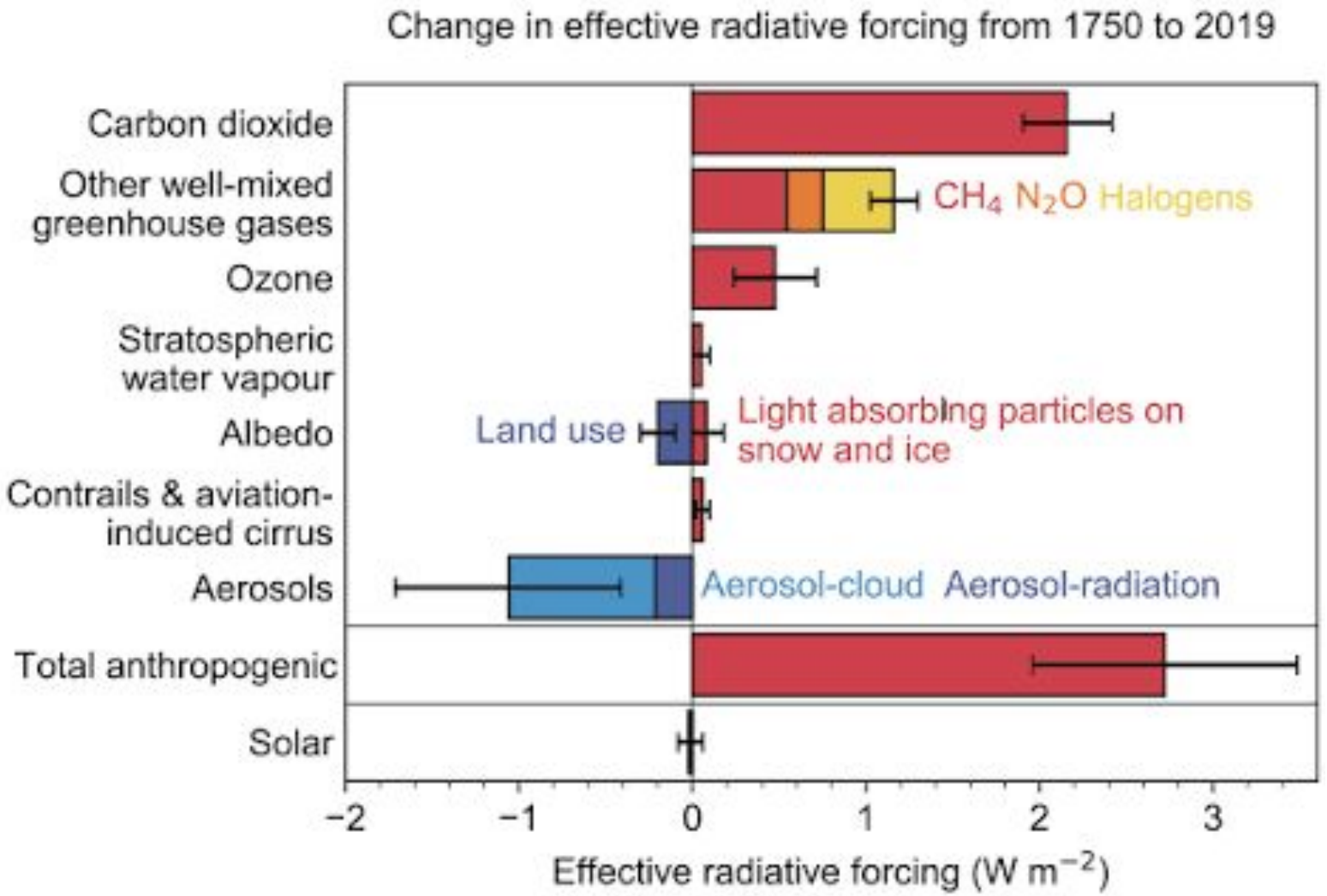
Aug, 2026

Lecture Roadmap

0:00–0:30	Part 1 Carbon Accounting Foundations
0:30–0:35	Engagement Moment
0:35–1:05	Part 2 Remote Sensing for Carbon Accounting
1:05–1:15	Break + Engagement moment
1:15–1:45	Part 3 Using AI to Improve Carbon Accounting
1:45–2:00	Discussion & Wrap-Up

Part 1 Carbon Accounting Foundations

Why do people want to account for carbon?



Emissions from two carbon molecules

Carbon Dioxide (CO2)

Methane (CH4)

Responsible for the bulk of warming since pre-industrial era

Source: IPCC

Why do people want to account for carbon:

- Estimating Greenhouse Gas (GHG) emissions
- Monitoring carbon "stocks" and other non-CO2 emissions
- Quantifying uncertainties

Focus of today's lecture

- Improving corporate accounting
- Strengthening lifecycle assessment
- Supporting verification and audit processes

A Taxonomy: Where Does AI Enter Carbon Accounting?

Use this map as a guide — each node is addressed later in the lecture

Activity Data	Emission Estimation	Reporting & Aggregation	Verification & Audit
What activities are happening & at what scale?	How much GHG is emitted per unit of activity?	Combining sources into inventories	Independent cross-checks of self-reported data
Imagery: Object detection (farms, oil wells, steel plants, trees)	Plume detection & segmentation	Automated homogenization / organization of activity datasets	Satellite-based cross-validation
Imagery: Deforestation / land-use change detection	Atmospheric inversion (top-down flux)	Lifecycle assessment scaling across product types	Anomaly / super-emitter flagging
Proxies: Scraping publicly available datasets (e.g., operator leases)	Emission factor improvement	Supply-chain Scope 3 estimation	Uncertainty quantification

This lecture deep-dives on emission estimation (plume detection) and activity data with connections to verification throughout.

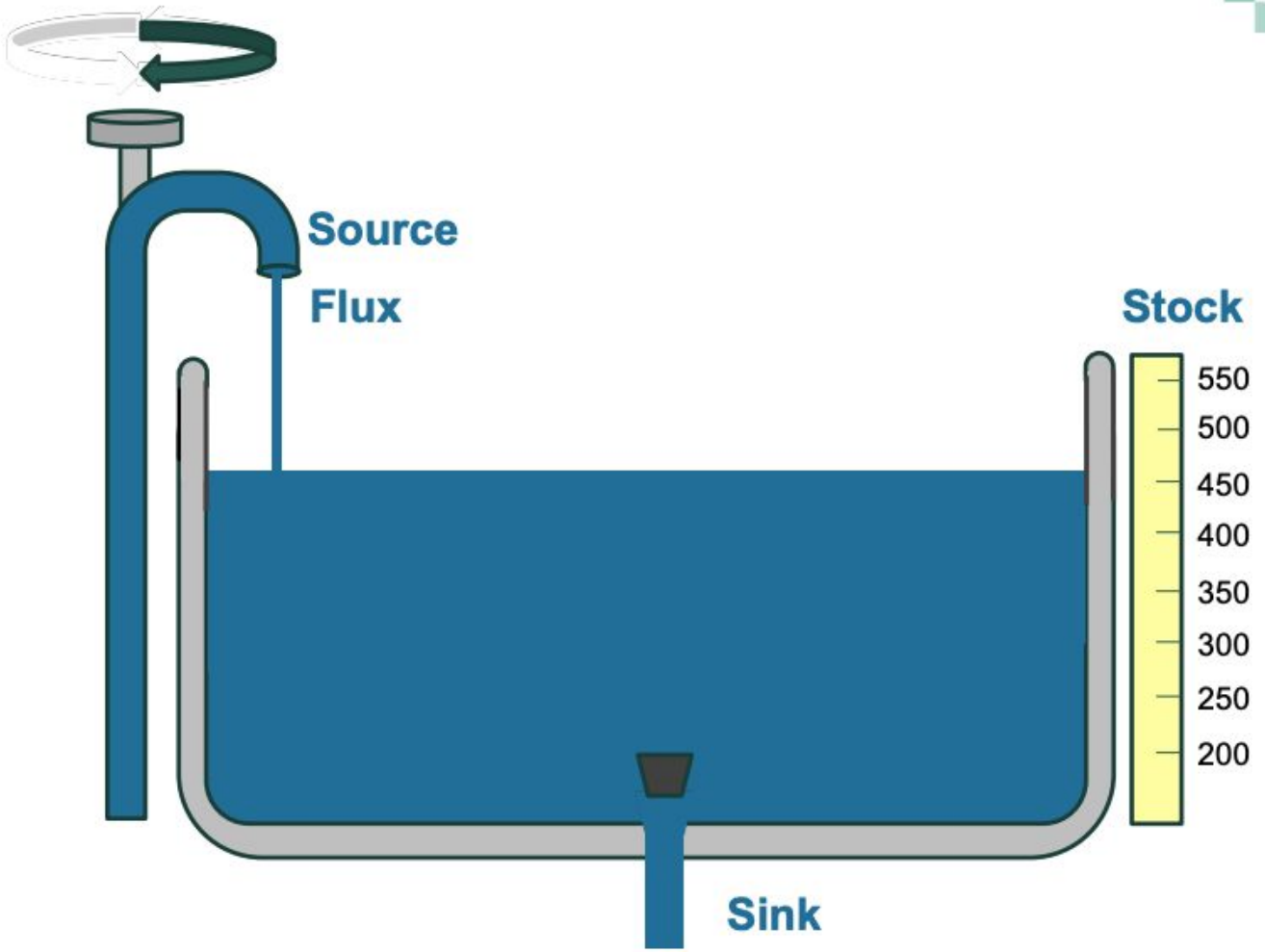
Relationship between sources, sinks, and stocks

Consider a basin with faucet and a plug at its bottom.

- **Stock**: water in basin
- **Source**: adding water to basin
- **Sink**: removing waver from basin

- The rate of increase of the stock in the basin is called the flux.

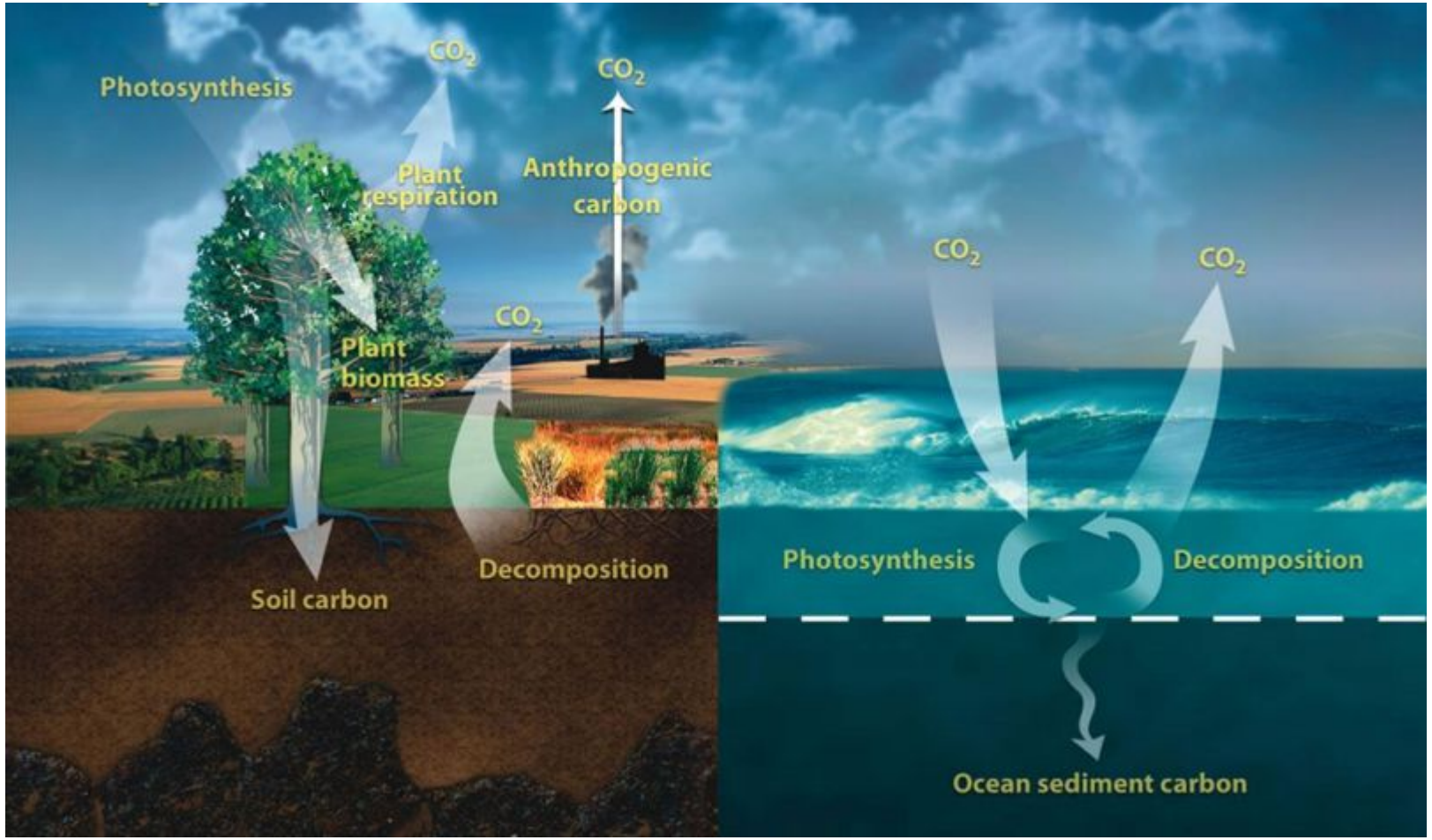
– Sources yield positive fluxes.
– Sinks yield negative fluxes.



Source: NASA ARSET

What are the major sources/sinks of carbon?

Sources and Sinks of Carbon



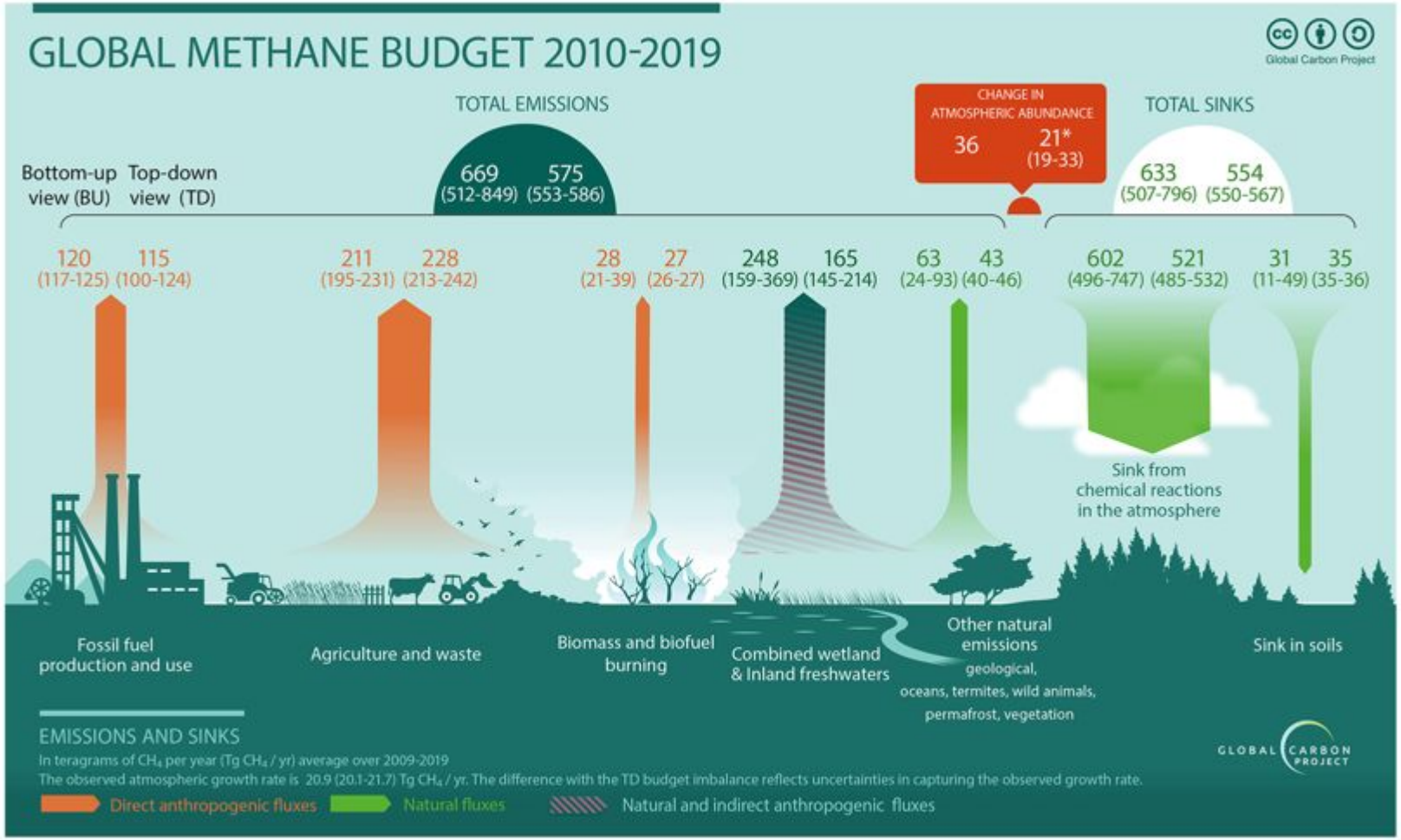
Accounting for carbon/CO2 requires understanding of

- Size of carbon stock across various “pools”
- Processes that result in changes to stocks (e.g., cycling and emissions)

Source: ORNL

What are the major sources/sinks of carbon?

Sources and Sinks of Methane (CH4)



Methane is different from CO2:

- Main sink is the atmosphere via oxidation
- Short lifetime (~10 years vs 100s years)
- 80x more powerful GHG (per unit emission) than CO2 on 20 year time horizon

Source: Sanuois et al. (2020)

Paris Agreement: Legally binding international treaty with goal to hold global warming well below 2°C

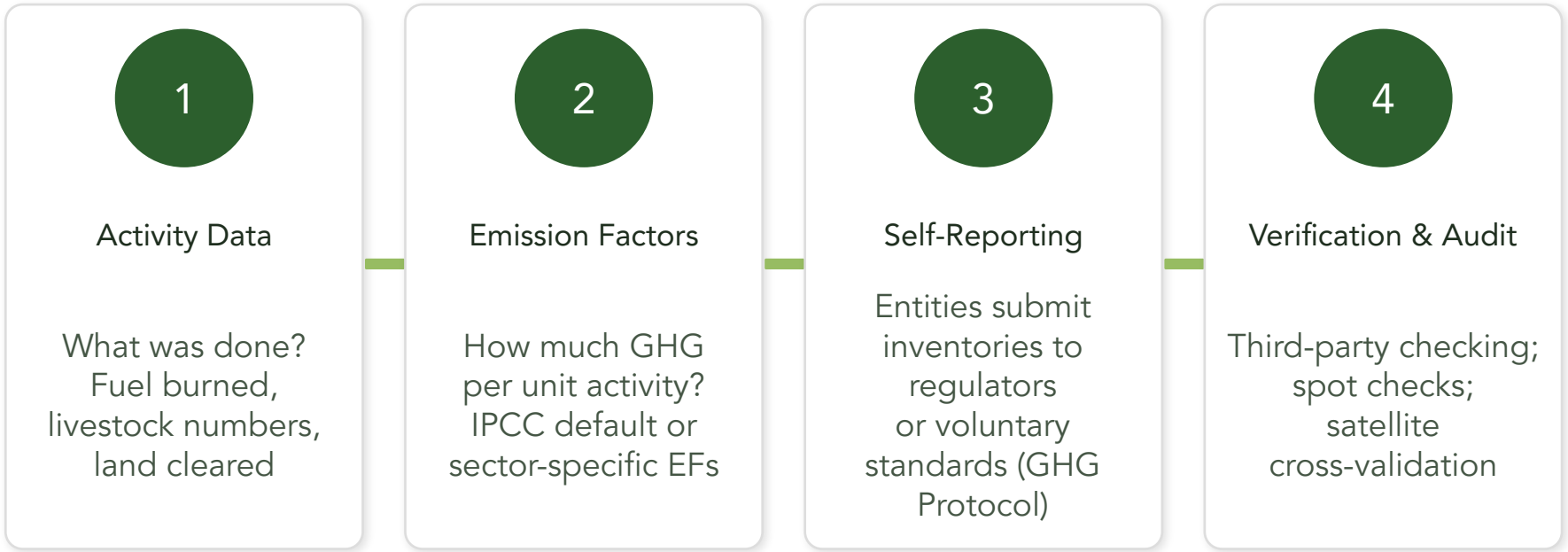


- Nationally Determined Contributions (NDCs): Every country is required to submit and progressively update their own emissions-reduction plans every five years.

- Transparency & Accountability: Members track and report their progress toward their NDCs using standardized reporting and international review.

- Periodic Global Stocktake, where countries evaluate collective progress toward goals like keeping the global temperature rise well below 2° C.

What are some traditional methods used for carbon accounting?

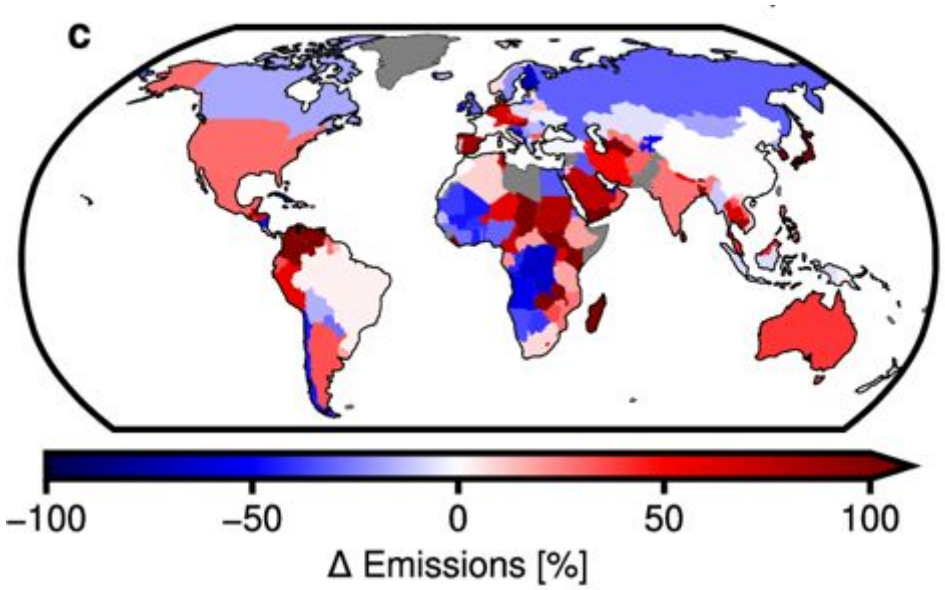


Where can this become complicated?

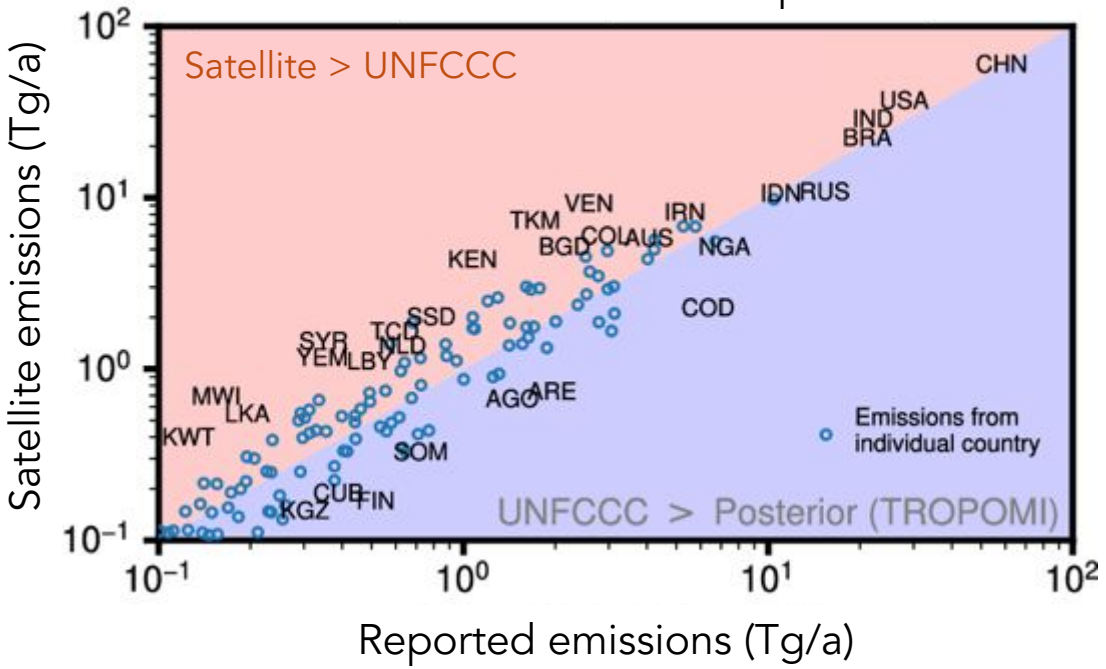
- Self-reported data may have incentive biases
- Emission factors may not represent local conditions
- Verification is resource-intensive and often incomplete.

Example: “Top-down” satellite study compared to IPCC/UNFCCC CH4 reporting

Relative difference of satellite estimate vs UNFCCC reports



Satellite vs UNFCC reports



Source: East et al. (2025)

Questions for the class

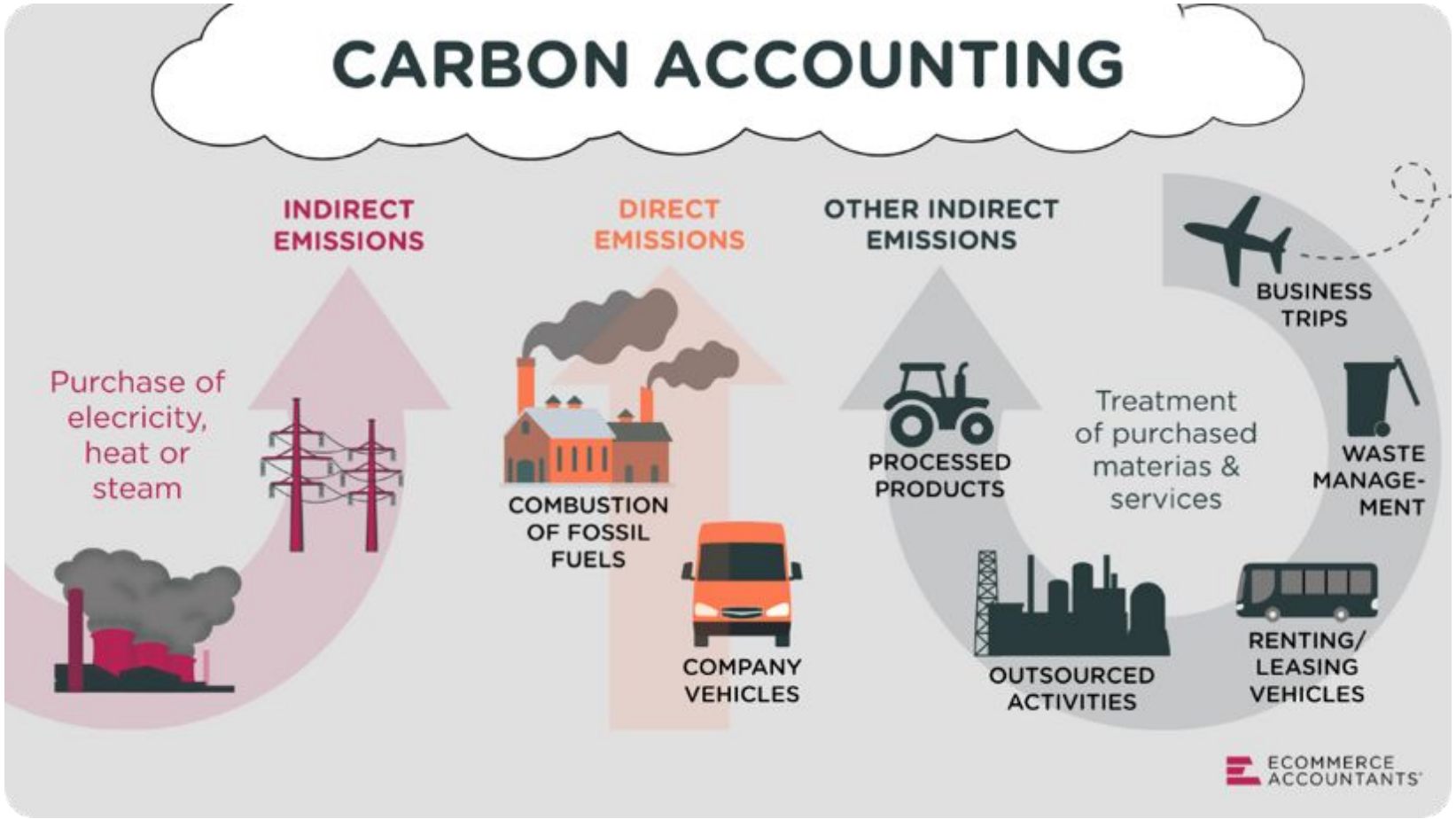
- We'll get deeper into "top-down" methods (atmospheric-measurement based) for carbon accounting in a moment. But based on what you know and have seen on top-down vs. bottom-up measurements

  - Live poll: What are you more inclined to trust? Why?
    - Vote: What do you trust more: Top-down vs bottom-up?

  - How would you build trust in one method, or both?

  - How good is "good enough"?

Quickly, overview of some other carbon-accounting use cases: Corporate carbon accounting



Accounting:

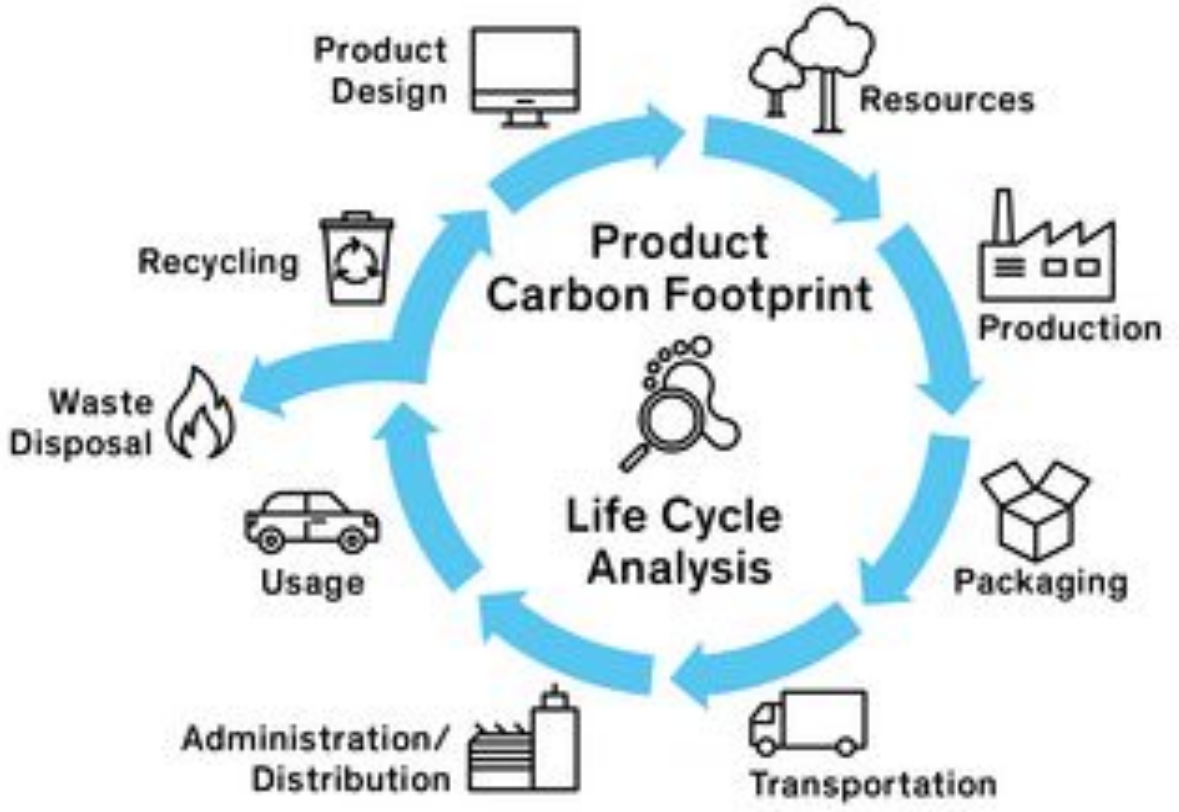
- Quantification of all emission sources, direct and indirect

Markets:

- Purchasing of credits or emission allowances if/when emissions in excess:
  - Cap-and-trade (compliance): purchasing “surplus” emissions from low-emitting entity
  - Carbon credits (voluntary): purchasing credits from verified projects that remove or mitigate emissions

Source: ecommerceaccountants.co.uk

Lifecycle assessment: Breaking down carbon from individual products or processes



General follows standard procedures (e.g., ISO 14040 14044) to account for all materials, energy, and water used to generate product/process

Potential AI applications:

- Automated data extraction
- Predictive modeling (e.g., if supplier data is missing)
- Predictive modeling of potential carbon benefits if parts of process are updated
- Scaling to all products/processes in a portfolio (instead of just flagship products)

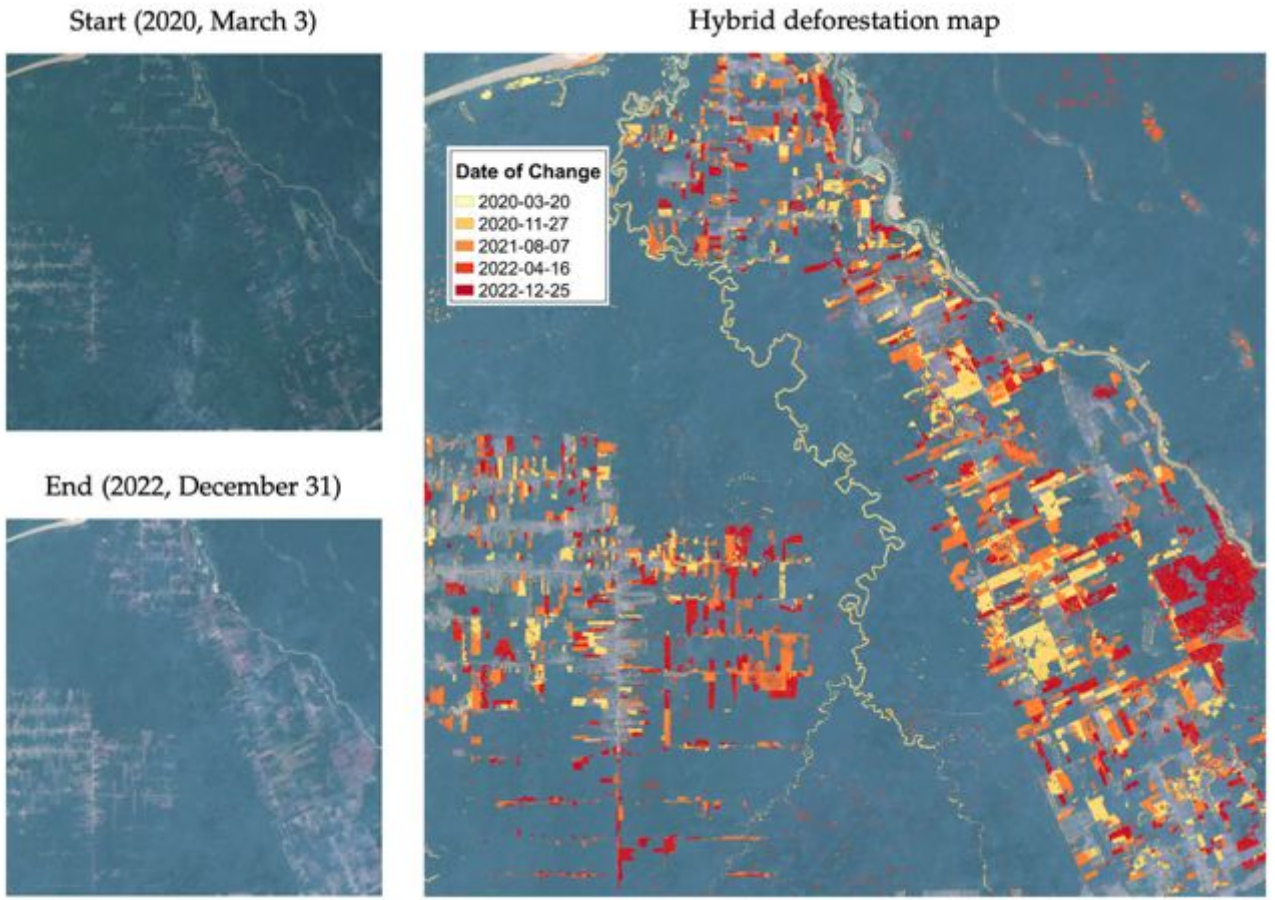
Source: myclimate.org

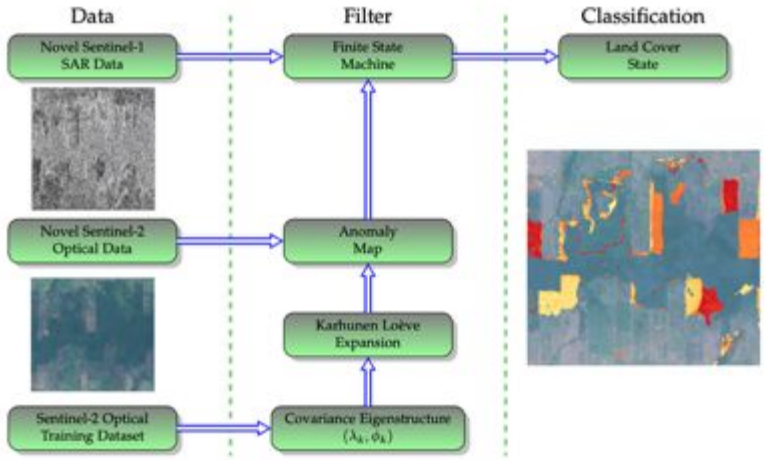
Part 2 Remote Sensing for Carbon Accounting

Returning to emission quantification: Role of satellite remote sensing

Activity data: Using high resolution imagery (visible, infrared, radar) for identification of equipment, processes, and change

Example: using a combination of visible (RGB) and synthetic aperture radar (SAR) to track deforestation



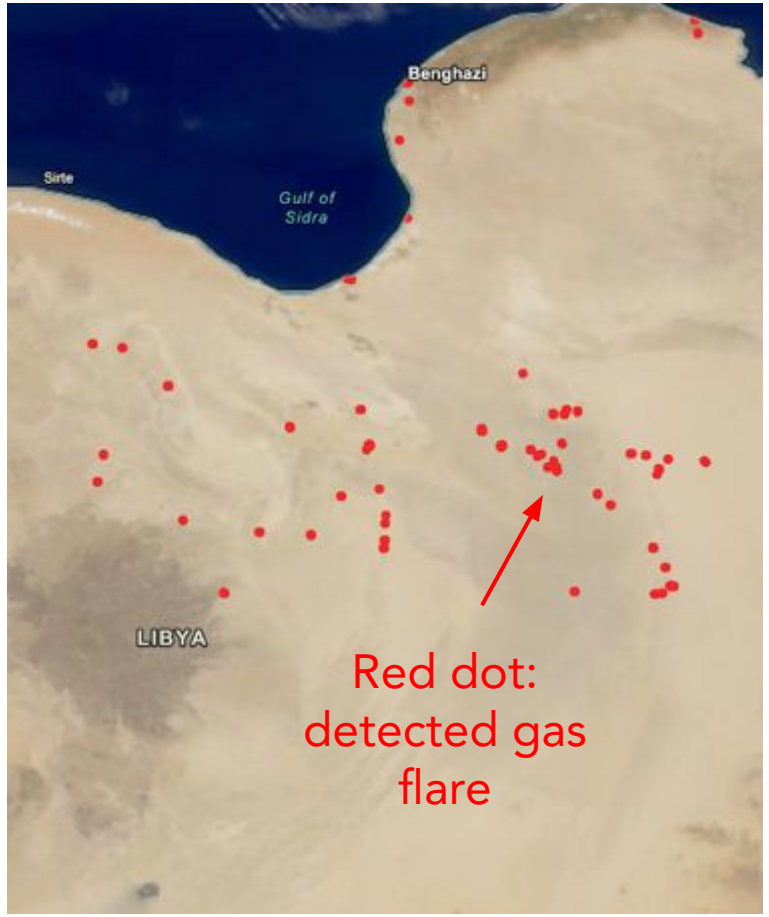


Source: Castrillon-Candas et al. 2026: https://arxiv.org/pdf/2510.14092´

Returning to emission quantification: Role of satellite remote sensing

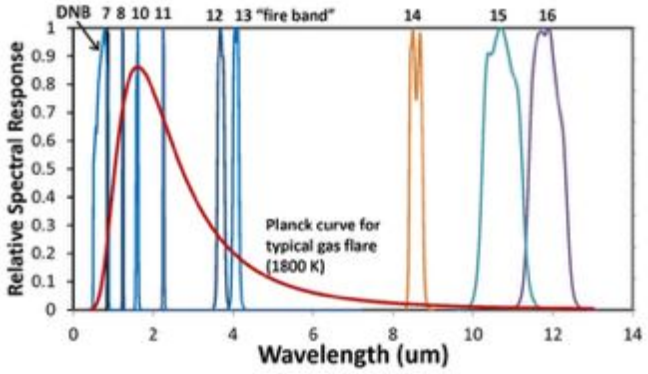
Activity data: Detection of thermal combustion both in natural and anthropogenic sectors

Gas flaring (VIIRS satellite)

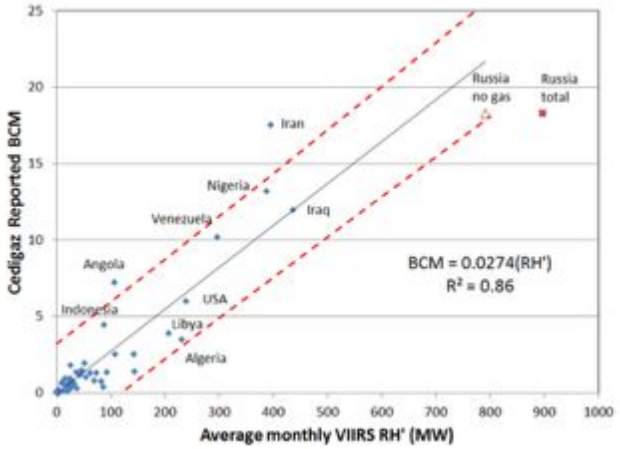


Source: Elvidge et al. (2016), NASA Worldview

Fit a Planck curve to observed spectrum to estimate heat from flare



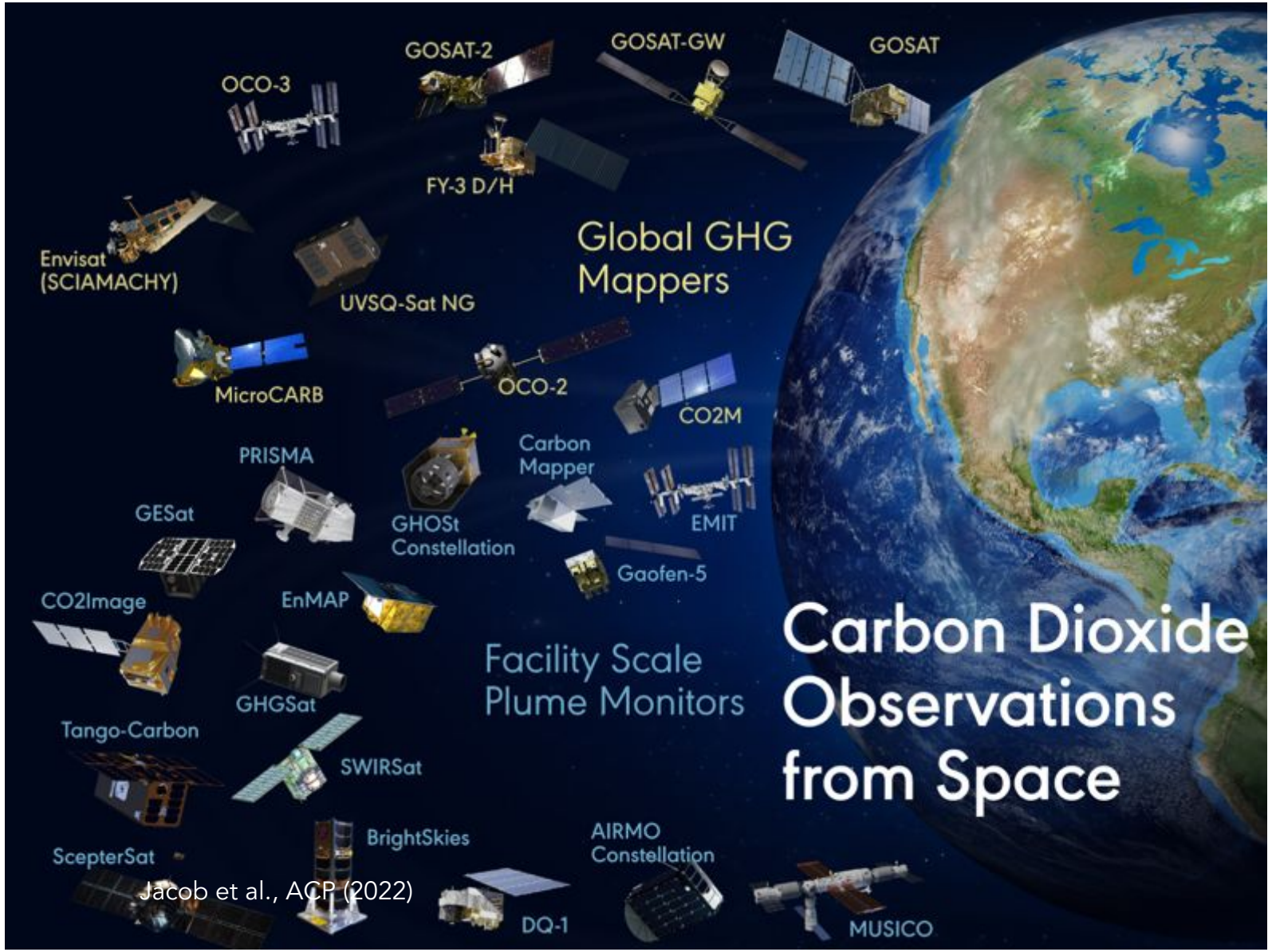
Relate heat signature to volume of gas flared



Wildfire tracking (VIIRS satellite)



Returning to emission quantification: Role of satellite remote sensing



Jacob et al., ACP (2022)

Two broad categories:

Global GHG Mappers: Wide-swath, coarse spatial resolution instruments designed to constrain regional to global net flux.

Facility Scale Plume Monitors: Provide emission detection and quantification of super-emitters at fine spatial resolution

Returning to emission quantification: Role of satellite remote sensing



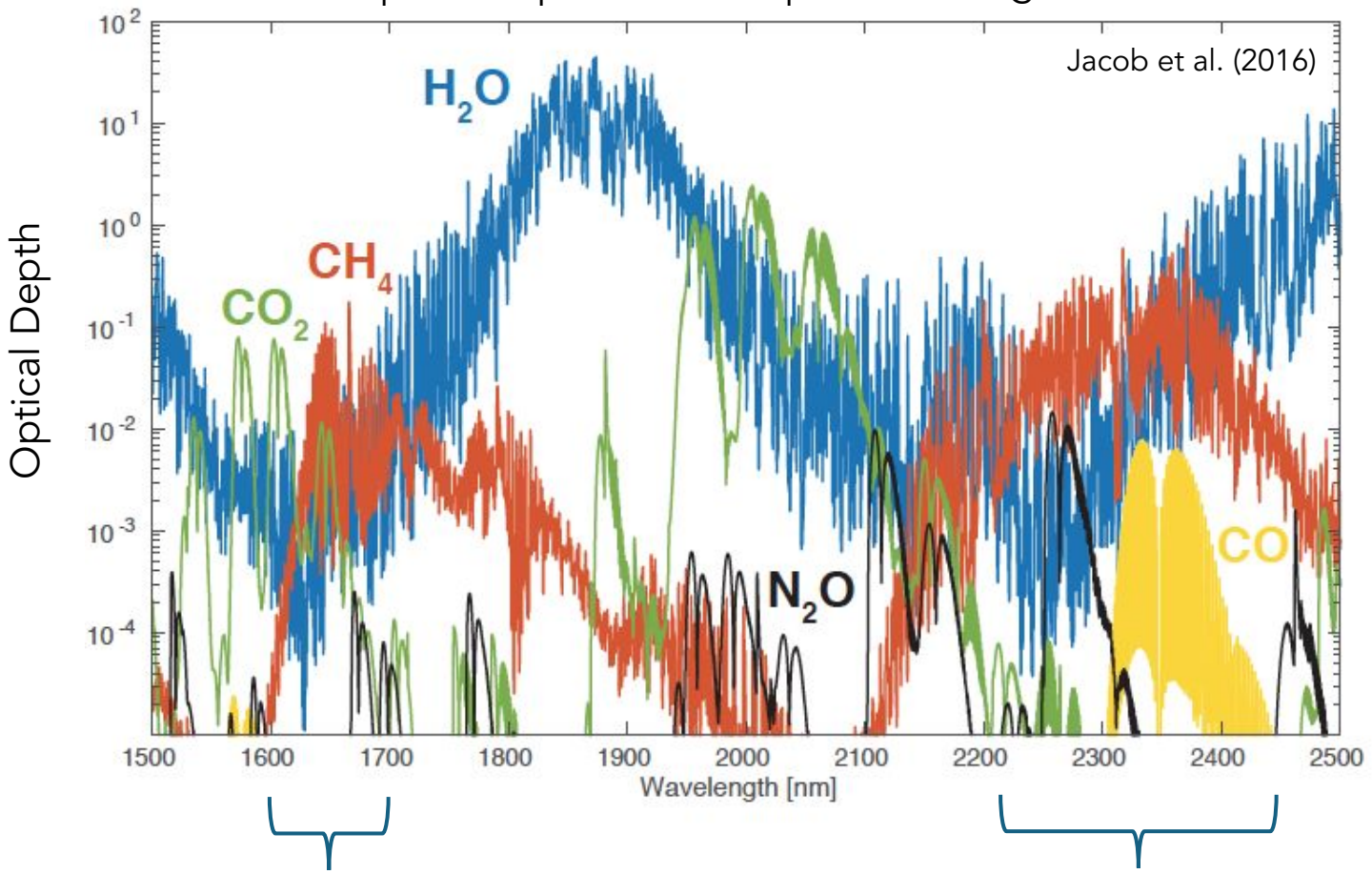
Two broad categories:

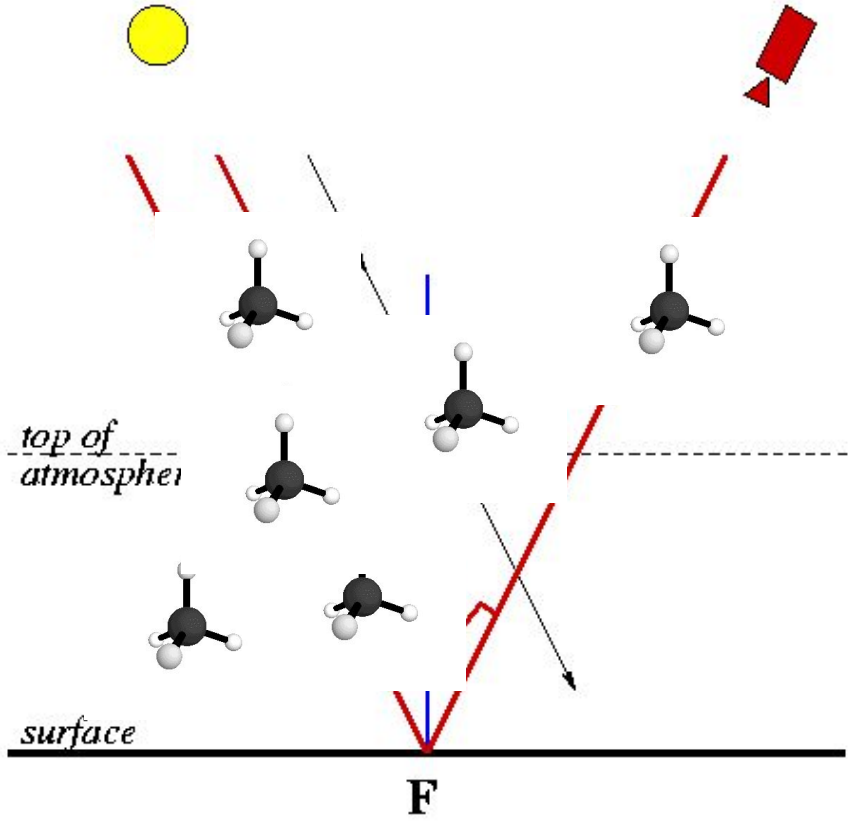
Global GHG Mappers: Wide-swath, coarse spatial resolution instruments designed to constrain regional to global net flux.

Facility Scale Plume Monitors: Provide emission detection and quantification of super-emitters at fine spatial resolution

Remote sensing of CH4 relies on observing radiance near CH4 rovibrational absorption features.

Optical depths of atmospheric trace gases





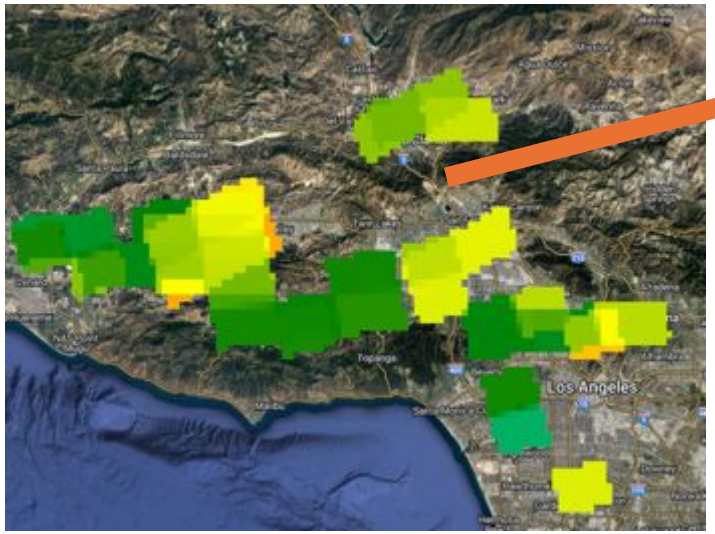
Atmospheric transmittance:

$$T = \frac{\Phi_e^t}{\Phi_e^i} = e^{-\tau} = 10^{-A}$$

Optical depth (depends on gas absorption)

The separation between global GHG mapper and facility scale monitor is determined by the spatial vs. spectral resolution trade space.

Snapshot TROPOMI over LA Basin / Aliso Canyon (Sep 15, 2020)

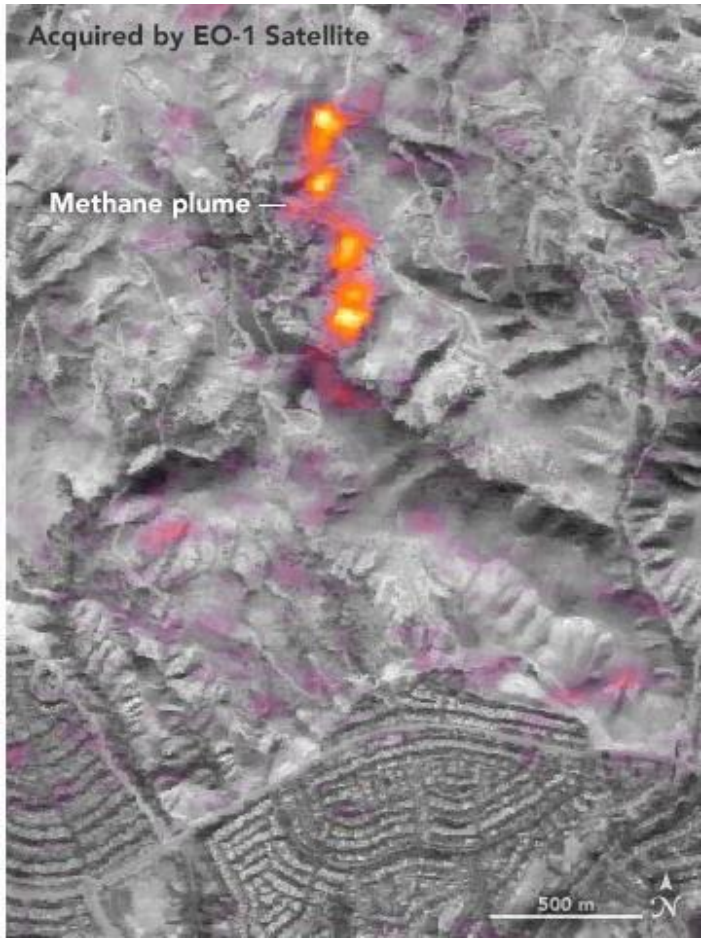


Google Earth Engine





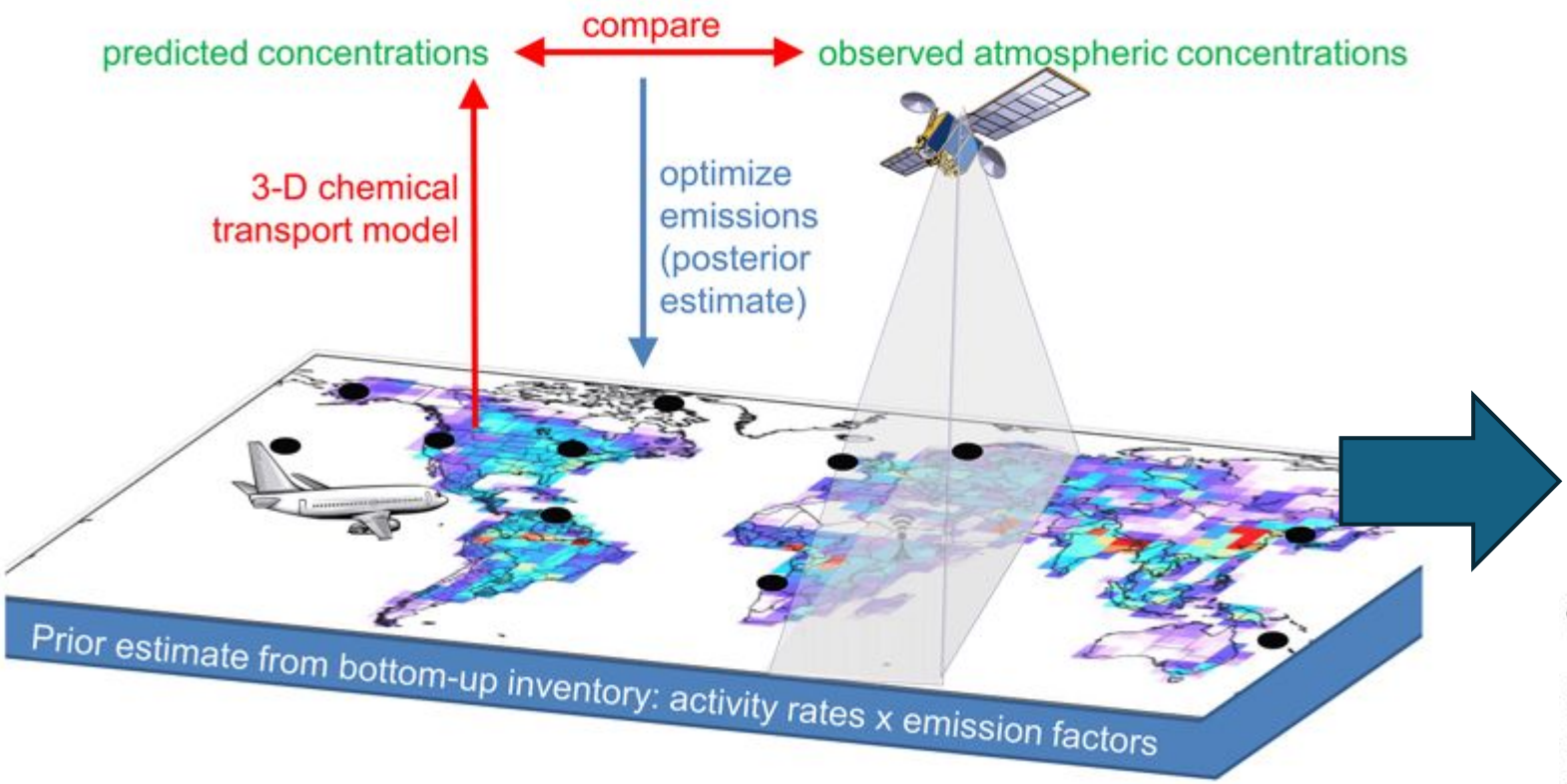
EO-1 Hyperion instrument: Aliso Canyon blowout (2016)

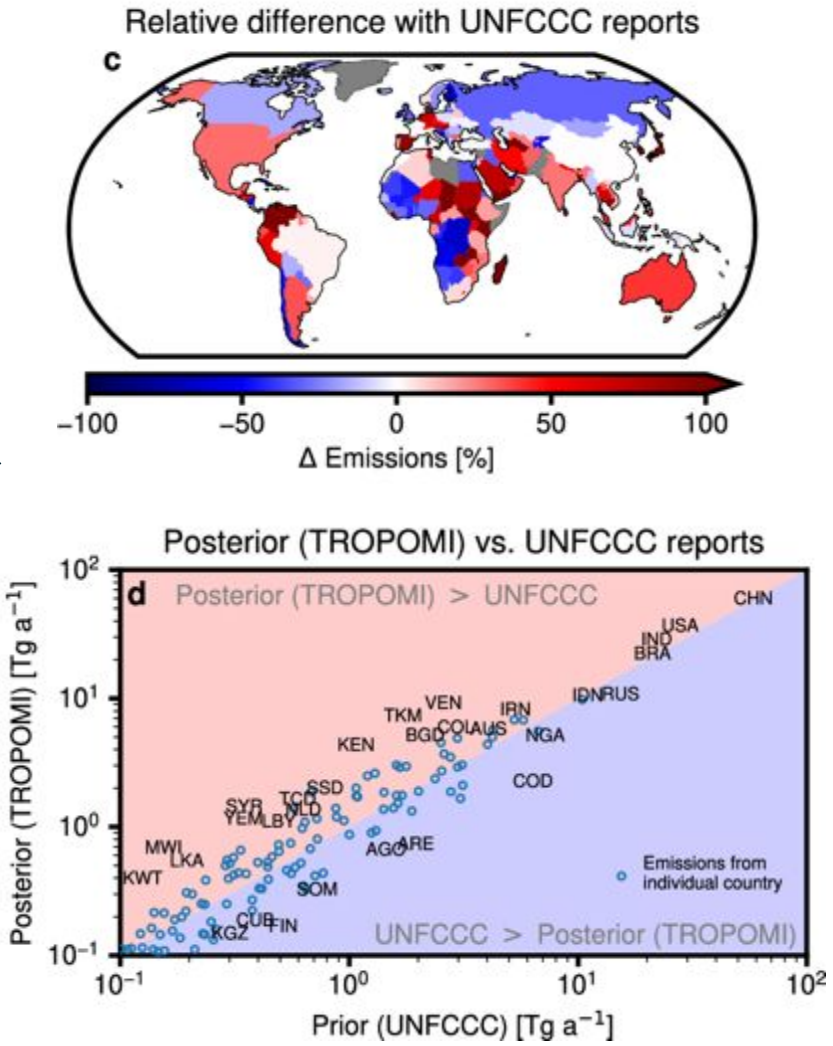


Thompson et al., GRL (2016)

Coarser spectral resolution -> finer spatial resolution. Ability to resolve large plumes at facility scales.

Global GHG mapping satellites estimate regional emissions by "inverting" observed atmospheric concentrations to emission fluxes



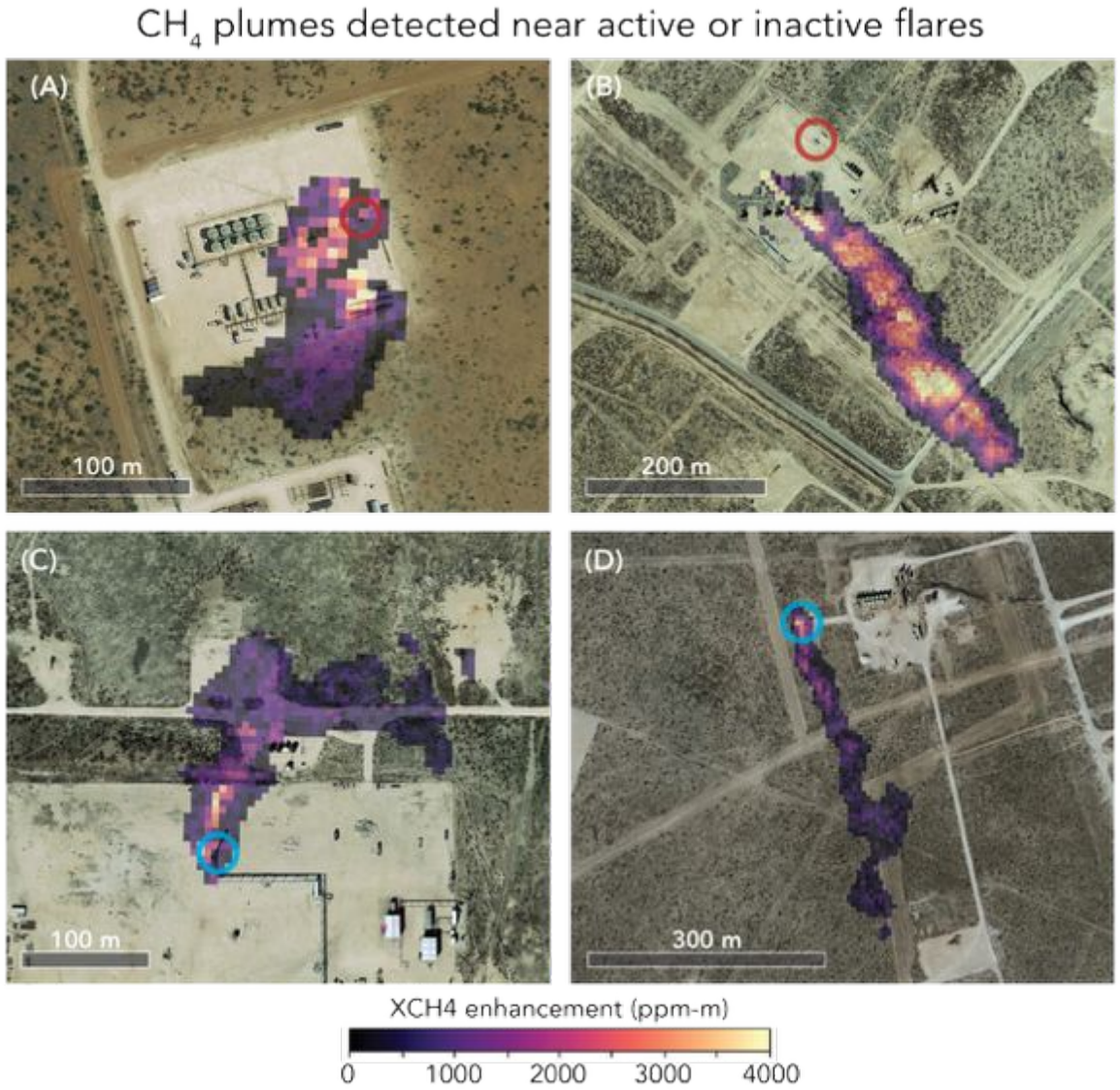


$$\hat{\mathbf{x}} = \arg\min \left\{ \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|_{\mathbf{S}_o} + \|\boldsymbol{x} - \boldsymbol{x}_A\|_{\mathbf{S}_A} \right\}$$

Here, x represents emissions, y is concentrations, and H is a transport model.

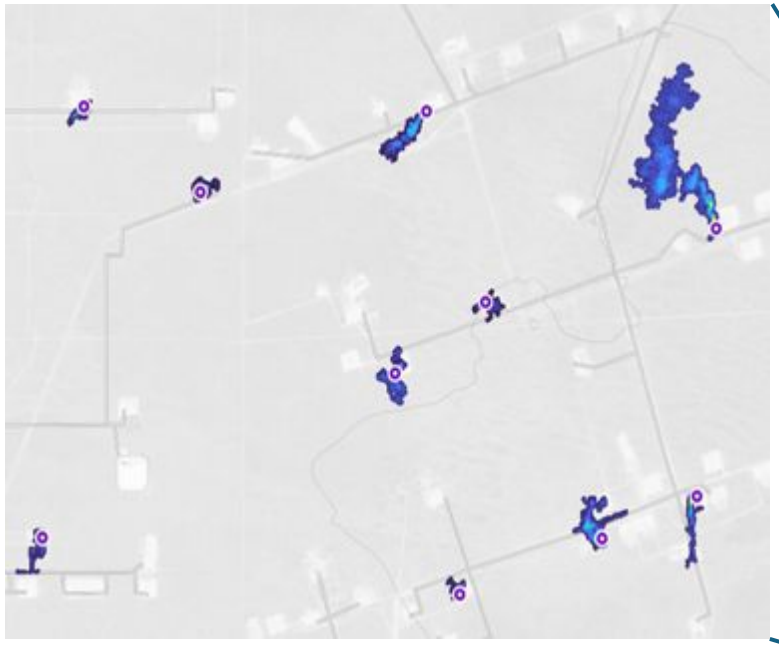
Worden et al., ACP (2022); East et al. (2025)

Facility scale mapping satellites are sensitive to only a fraction of all emission sources, but provide facility scale attribution

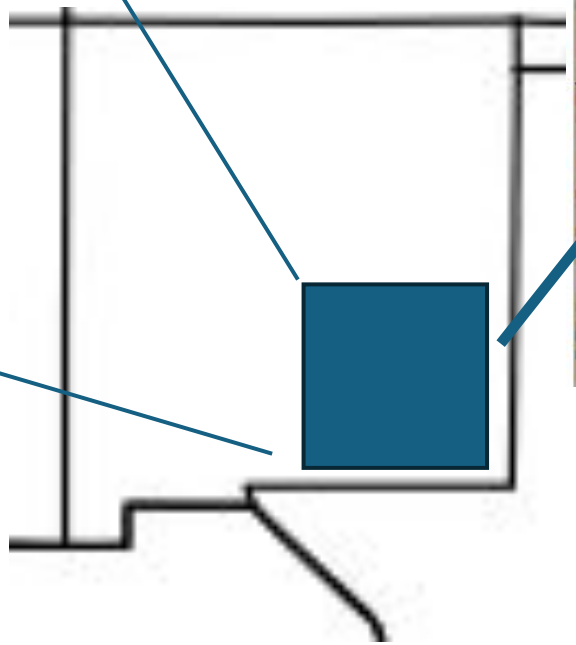


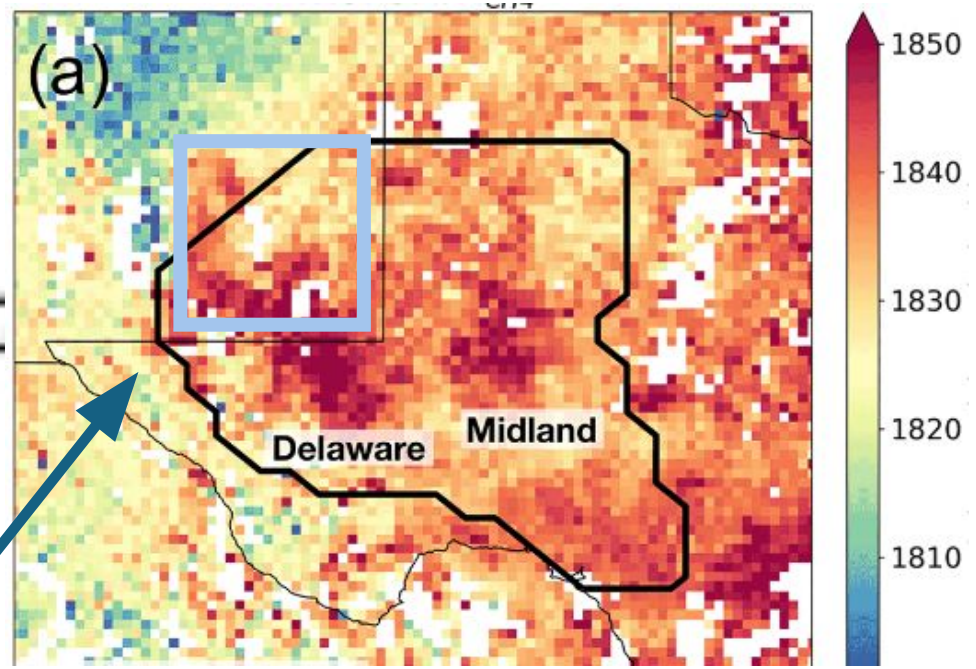
Understanding emissions at basin scales:
total flux, super-emitters, attributed datasets

2024 airborne + satellite study in New Mexico Permian



Wall-to-wall aerial mapping of all infrastructure in NM Permian – over 200,000 site visits in 3 weeks

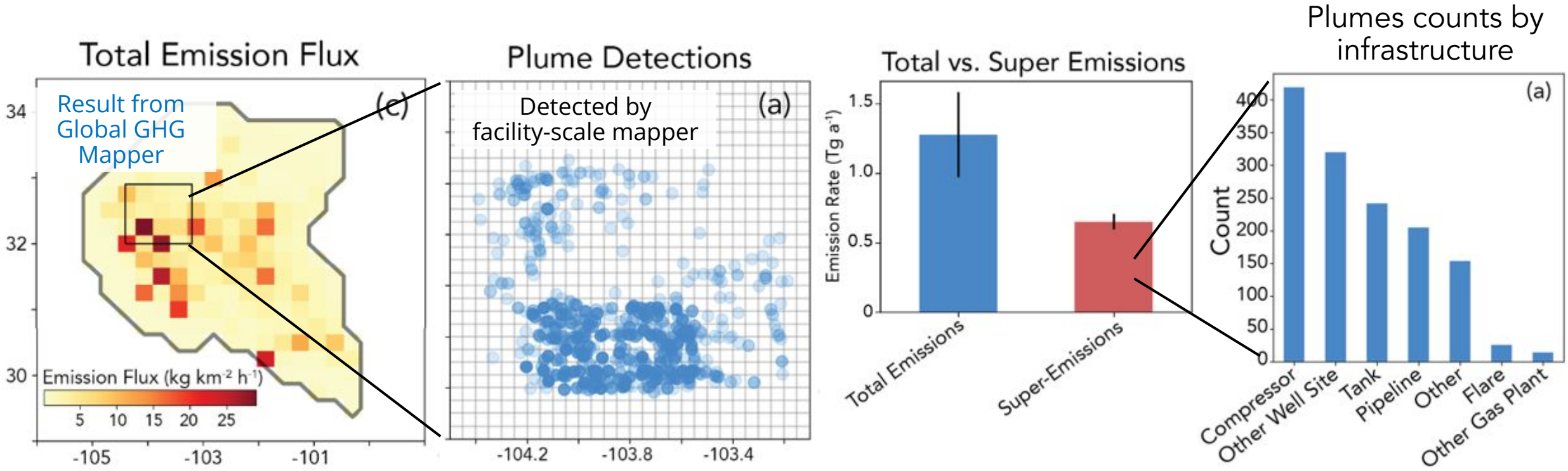




Basin flux estimation using TROPOMI satellite observations

Understanding emissions at basin scales:
total flux, super-emitters, attributed datasets

2024 airborne + satellite study in New Mexico Permian

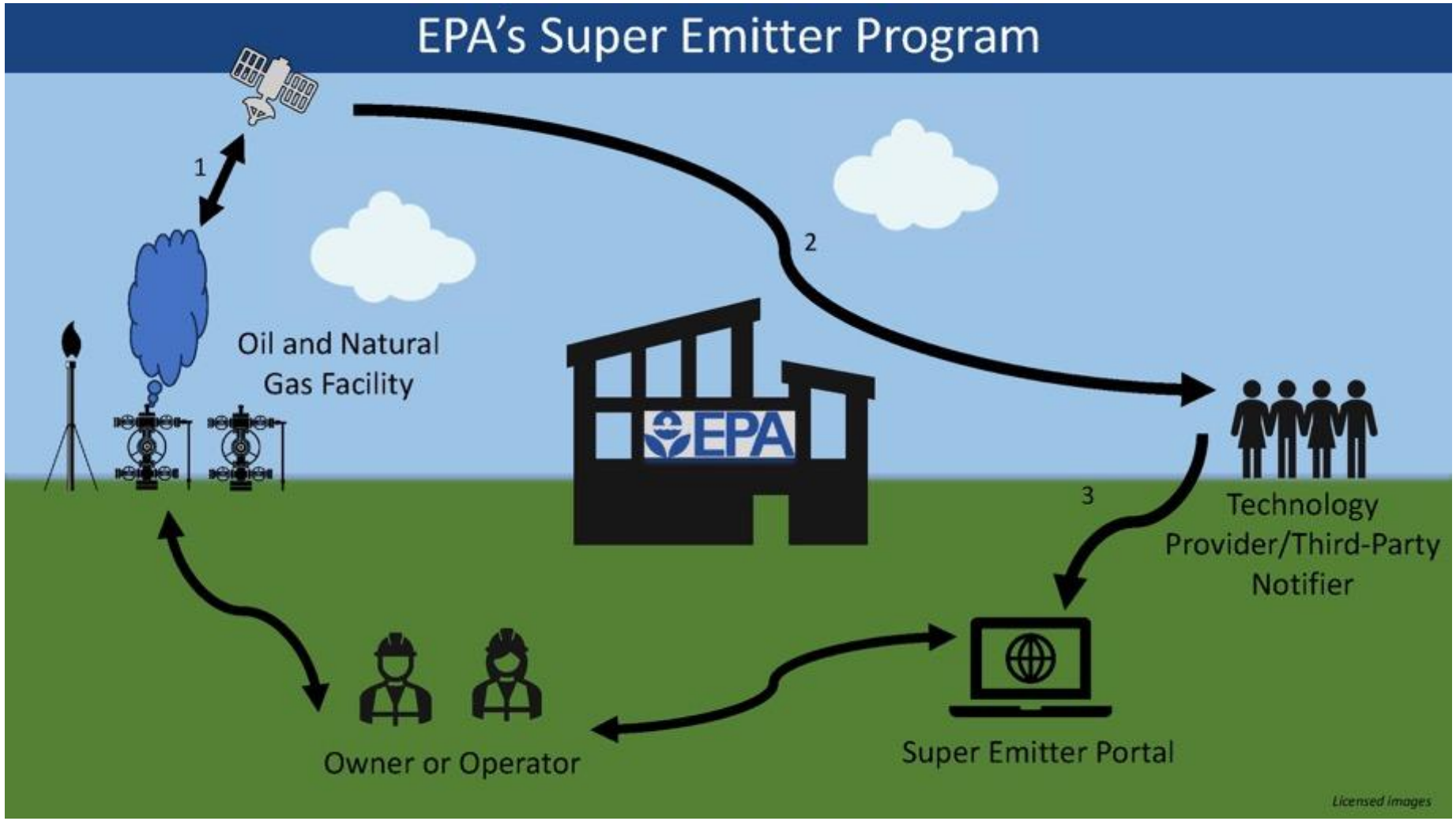


Cusworth et al. (in review)

Intensity + infrastructure + operator attribution provides a pathway for mitigation

Cusworth et al. (2026)

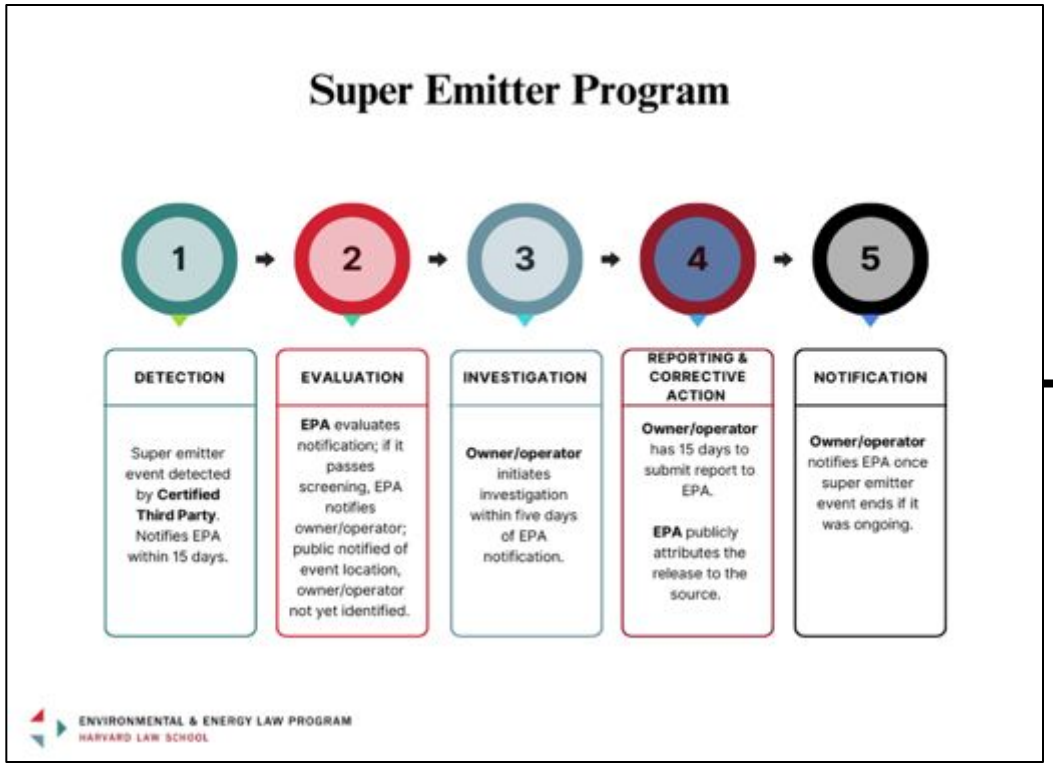
Connecting back to reporting programs:



New rule in USA that satellite detections of super-emitters by third parties could be reported to EPA (by approved reporters) requiring operator responses

Source:epa.gov

Any super-emitter detections were expected to be then reported by operators in required operator reporting program (Greenhouse Gas Reporting Program)



Annually: operator must report any emission > 100 kg/h regardless of duration

Annually: operators pay Waste Emission Charge for leakage over 1000 tCH4/y

EPA GHGRP (subpart W)* total methane emissions for any operation over 1000 tCH4/yr

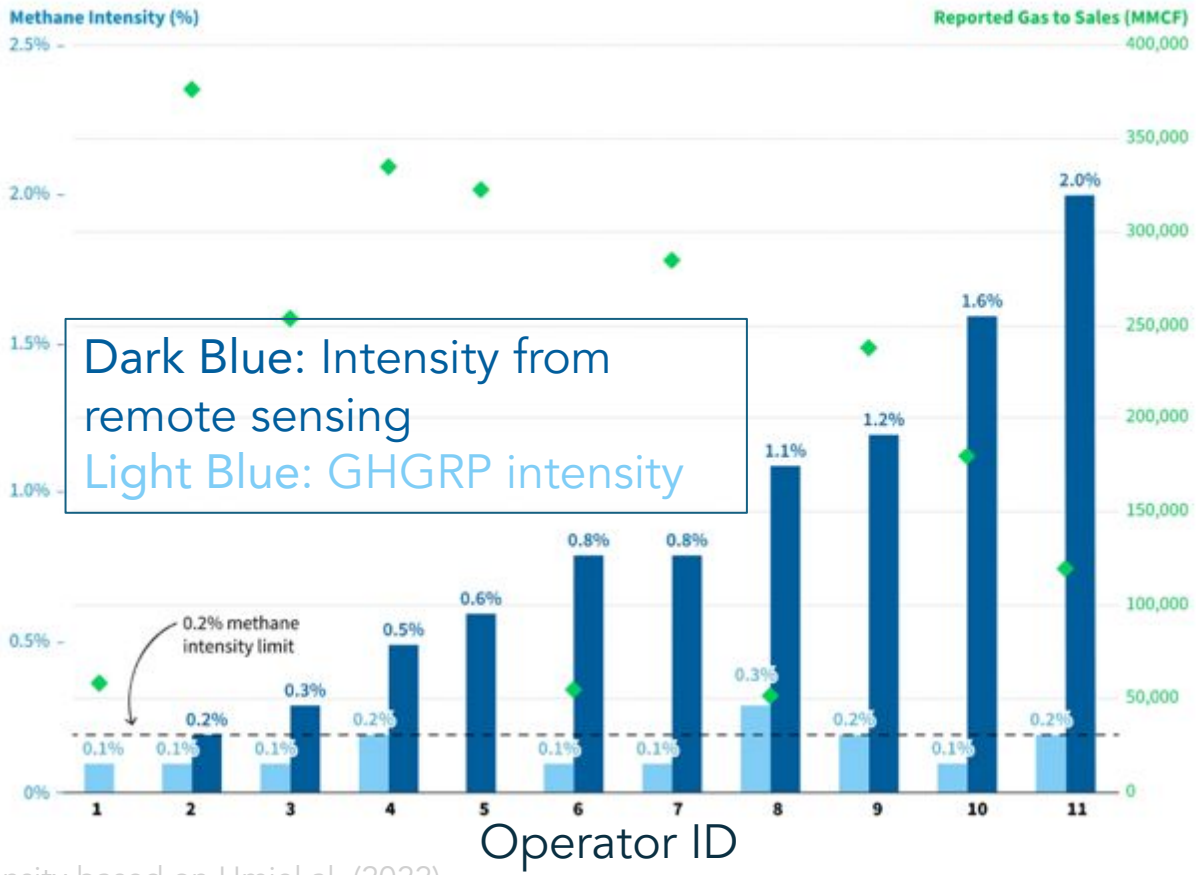
IRA Methane Emission Reduction Program (MERP)*

Other (non super-emitter) empirical methane measurements &/or bottom-up estimates

Other super-emitter single-detection reporting (e.g., operator LDAR screening)

The Inflation Reduction Act (IRA) stipulated that any wasteful emissions would result in a fee (Waste Emissions Charge – WEC). Fees based on GHGRP reporting.

Observed Intensity by operator from aircraft campaigns compared to current GHGRP protocols



Dark Blue: Intensity from remote sensing
Light Blue: GHGRP intensity

Intensity based on Hmiel al. (2022)

RMI conservative analysis of under-reporting

Waste emissions charge difference: Super-emitter observations vs. reported emissions for the Permian Basin in 2020

Operator	**Reported to EPA:** Tax payments at $900/ton	**Aerial measurements:** Tax payments at $900/ton	**Cumulative difference** between tax payments
1	$0	$0	$0
2	$0	$0	$0
3	$0	$5,000,000	$5,000,000
4	$300,000	$16,000,000	$21,000,000
5	$0	$25,000,000	$45,000,000
6	$0	$6,000,000	$51,000,000
7	$0	$30,000,000	$81,000,000
8	$800,000	$8,000,000	$89,000,000
9	$0	$41,000,000	$130,000,000
10	$0	$42,000,000	$172,000,000
11	$900,000	$37,000,000	$208,000,000

Relevance to IRA: Unexposed under-reporting across US could translate to >$800M/year underpayment of Waste Emissions Charge

Break

Discussion

How could/should we base top-down results in ground truth, standards, and validation?

Has your opinion changed regarding trustworthiness of top-down vs bottom-up accounting?

- Vote: what do you trust more? Top-down, bottom-up?

What types of experiments would you suggest/perform to build confidence in top-down/remote sensing estimates

Part 3 Using AI to Improve Carbon Accounting

Activity Data: Filling in production gaps

Example: Gary Works Steel Plant



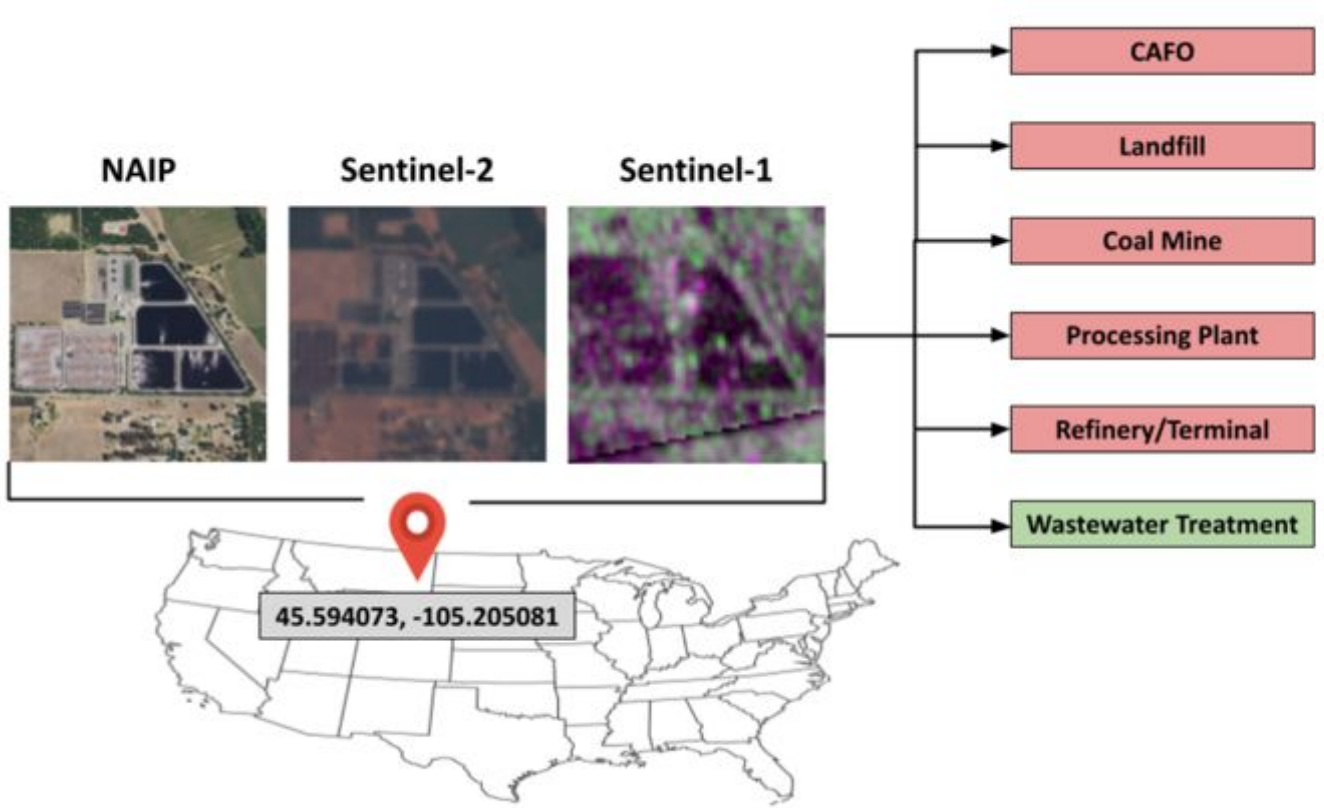
TransitionZERO

Hundreds of steel plants globally do not report emissions.

To create database of estimated emissions – train a machine learning model on heat signatures detected from Landsat-8 and Sentinel-2 infrared bands.

Source: TransitionZERO; Climate TRACE

Activity Data: Object detection

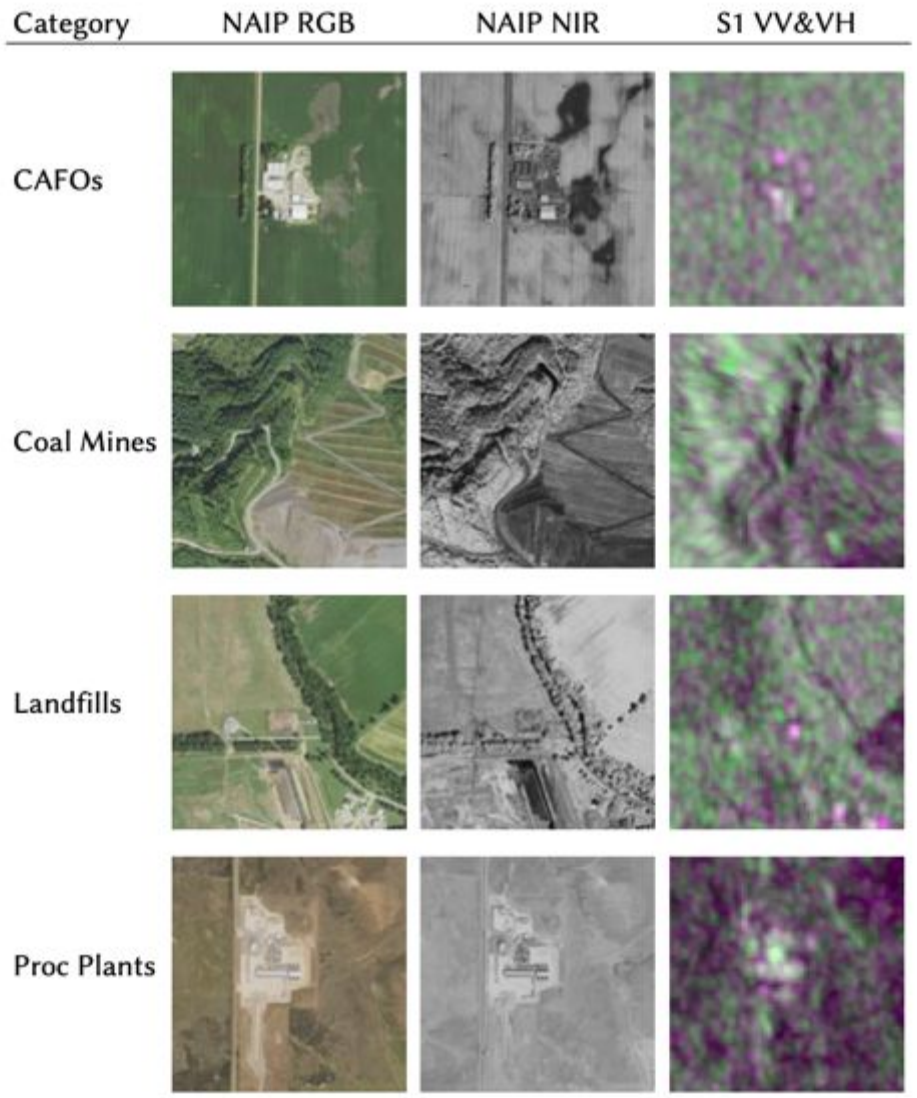


MEthane Tracking Emissions Reference (METER)

Goal: to build a complete database of anthropogenic methane emission infrastructure using variety of satellite imaging.

Source: METER-ML

Activity Data: Object detection



To build labels, researchers scraped any available Geographic Information Systems (GIS) infrastructure database and clipped satellite imagery accordingly

~50,000 labels/images for training

Source: METER-ML

Activity Data: Object detection

Classification results (AUPRC) after application of CNNs to labeled datasets

Image Product	Bands	CAFOs	Landfills	Mines	Proc Plants	R&Ts	WWTPs	Overall
S1	VH&VV	0.519	0.107	0.152	0.218	0.487	0.119	0.267
S2	RGB	0.889	0.268	0.305	0.374	0.694	0.204	0.456
S2	All	0.889	0.189	0.382	0.368	0.690	0.183	0.450
S2 & S1	All	0.923	0.152	0.379	0.391	0.612	0.231	0.448
NAIP	RGB	0.903	0.270	0.348	0.327	0.849	0.182	0.480
NAIP	All	**0.945**	**0.276**	0.401	**0.508**	**0.857**	**0.303**	**0.548**
NAIP & S2 & S1	All	0.889	0.214	**0.473**	0.457	0.796	0.272	0.517

So much satellite data, but good labels are still hard to come by and vary considerably by geography

Validating labels: 2 expert reviewers manually reviewed 1500 labels:

If the expert identified no clearly visible methane source categories in the image, the example was labeled "negative", and if the expert was uncertain about any label, the example was labeled "uncertain". The two labels per example were then resolved as follows:

1. If the experts agreed and neither was uncertain, the agreed upon label was taken as the final label.
2. If the experts disagreed, and one was uncertain but the other was not, the expert's certain label was taken as the final label.
3. If the experts disagreed, but one agreed with the original label, the original label was taken as the final label.
4. In all other scenarios, the example was reviewed jointly by the experts and a final label was assigned.

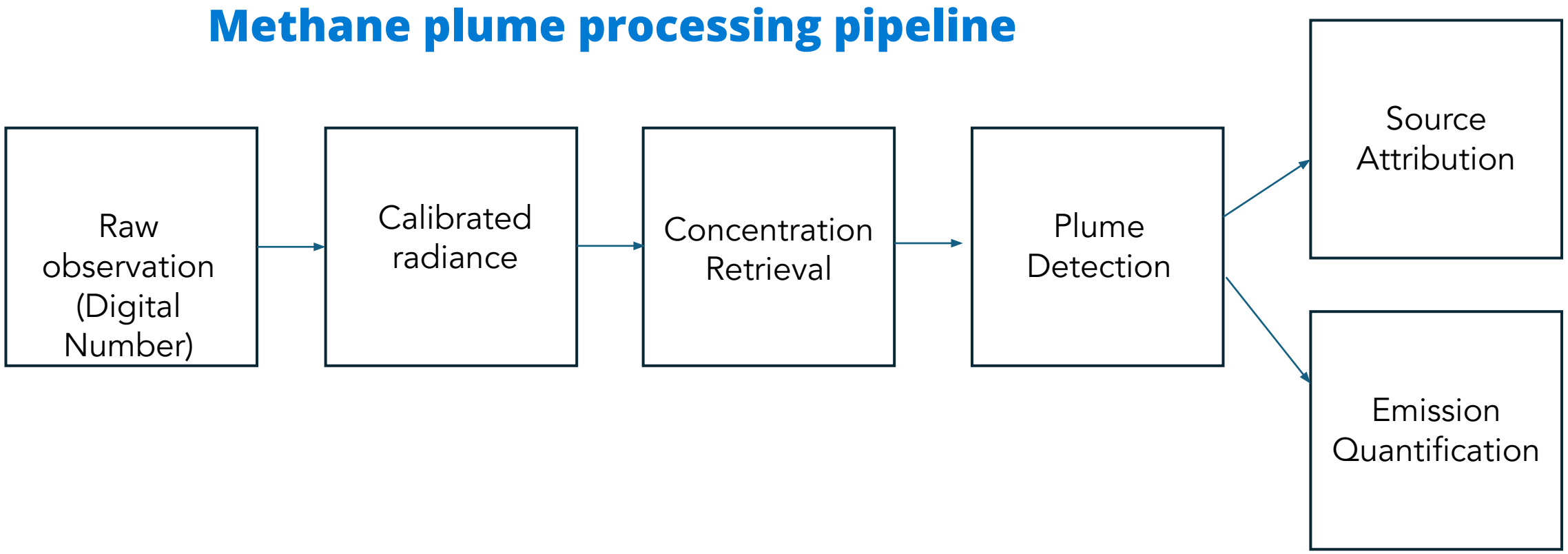
Source: METER-ML

Worked through example:

Methane plume processing pipeline

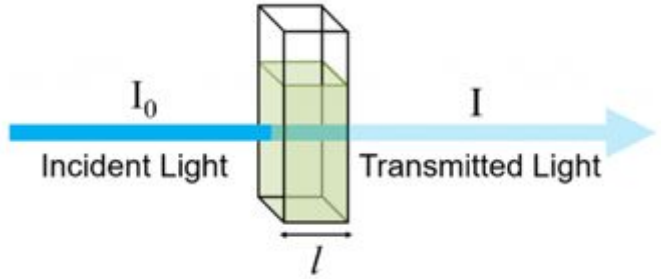
Task: As we work through this example, think of the ways AI may aid the process

Creating plume data products requires multiple processing stages.



The satellite observes top of the atmosphere radiance, from which column concentrations are estimated through a retrieval.

Columnwise "Matched Filter" algorithm – based on Beer-Lambert's law



$$x_m = x_r e^{-k\Delta c}$$

Modeled spectrum

Reference spectrum

Absorption

Concentration Enhancement

Taylor expansion (linearized case)

$$x_m = x_r(1 - k\Delta c)$$

Using a Maximum Likelihood Estimator, the "optimal" concentration enhancement that minimizes the residuals between model and observation (scaled by some covariance) is:

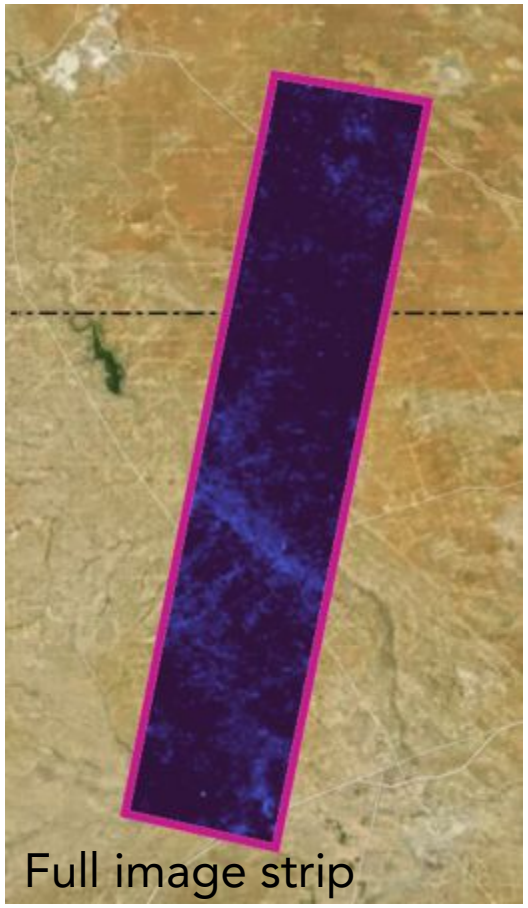
$$\Delta c = \frac{t^T \Sigma^{-1}(x_r - x_o)}{t^T \Sigma^{-1} t}$$

Column CH4 enhancement

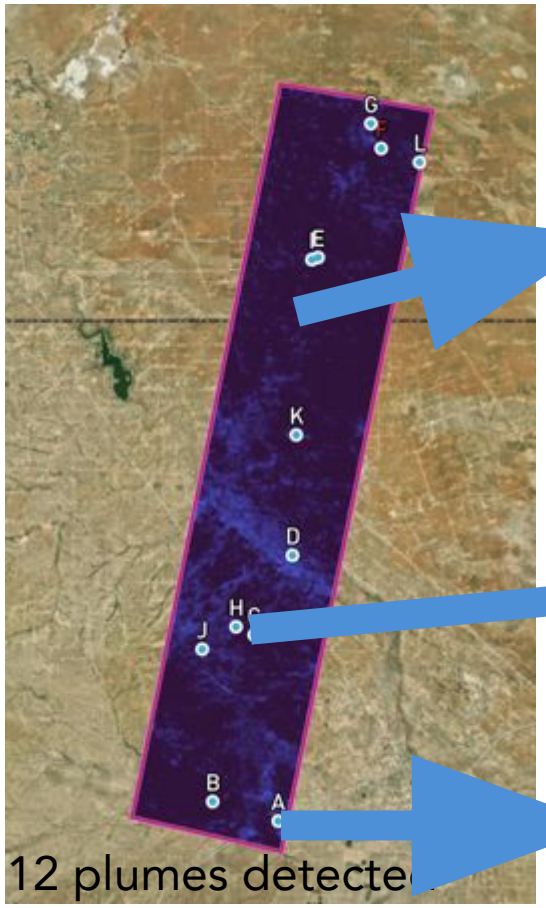
We choose x_r and Σ to be the mean and covariance of a collection of pixels in the image, and t is a mapping between radiance and concentration

Plume detection process: picking needles from haystacks

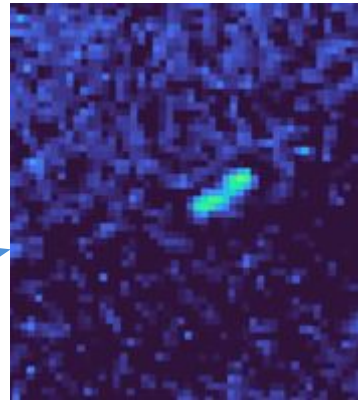
Example: 20x100 km2 image (Tanager satellite) taken over dense oil&gas infrastructure in Texas



Full image strip



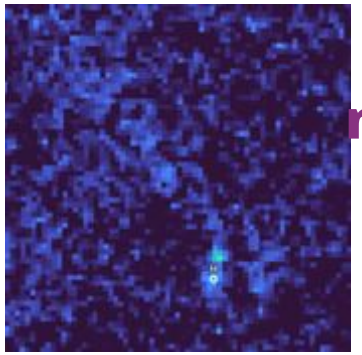
12 plumes detecte





False enhancement – white roofs

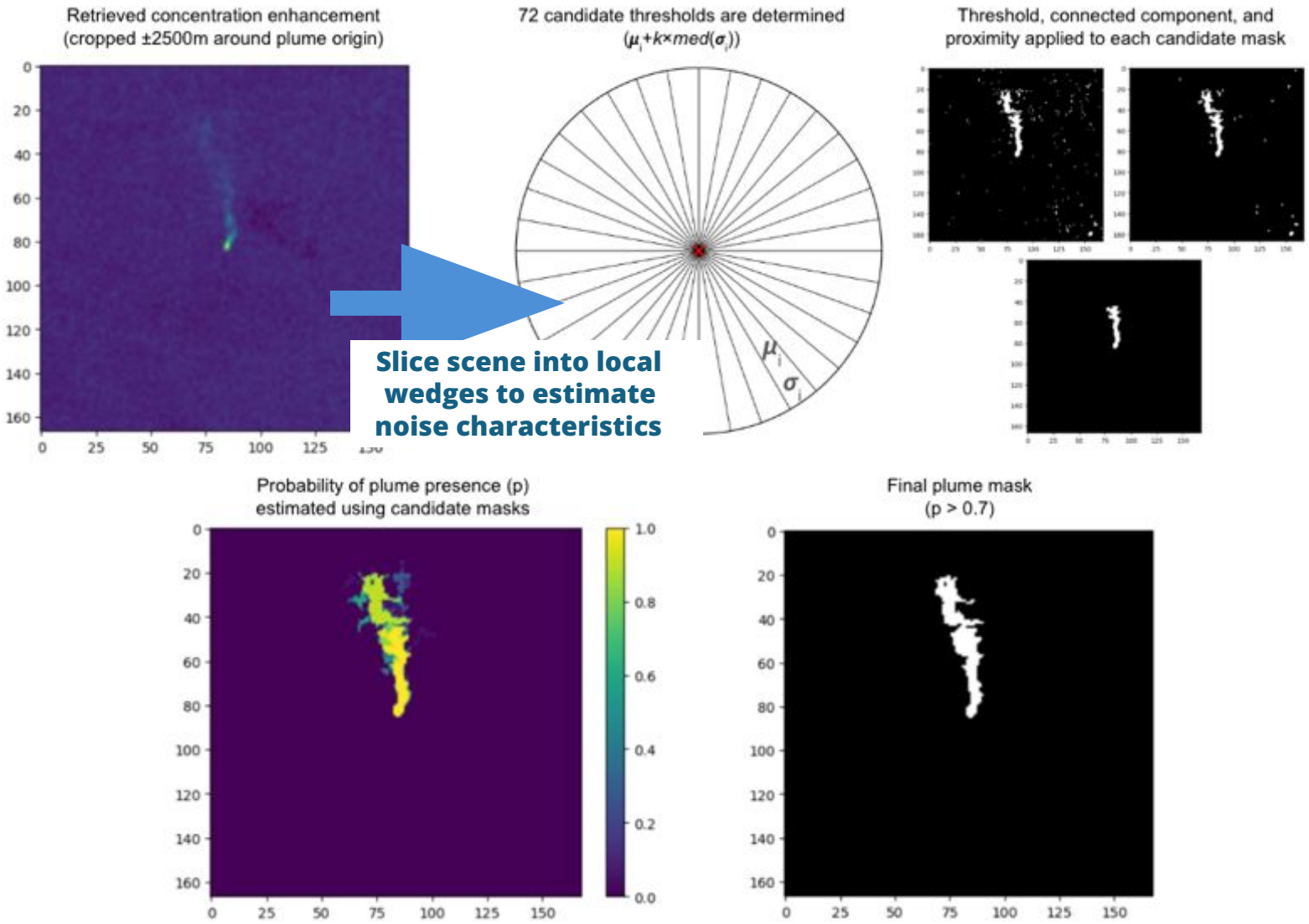




Small marginal plume

Question: What's the risk of a single false positive (think back to earlier in the lecture)?

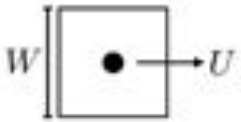
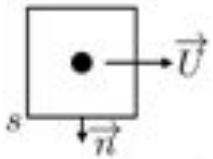
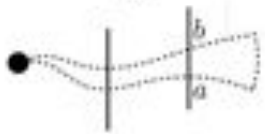
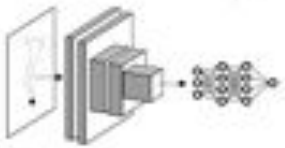
Segmenting (“masking”) plume detections



Critical steps towards quantification

Segmentation defines plume boundaries, enabling quantification of mass and length

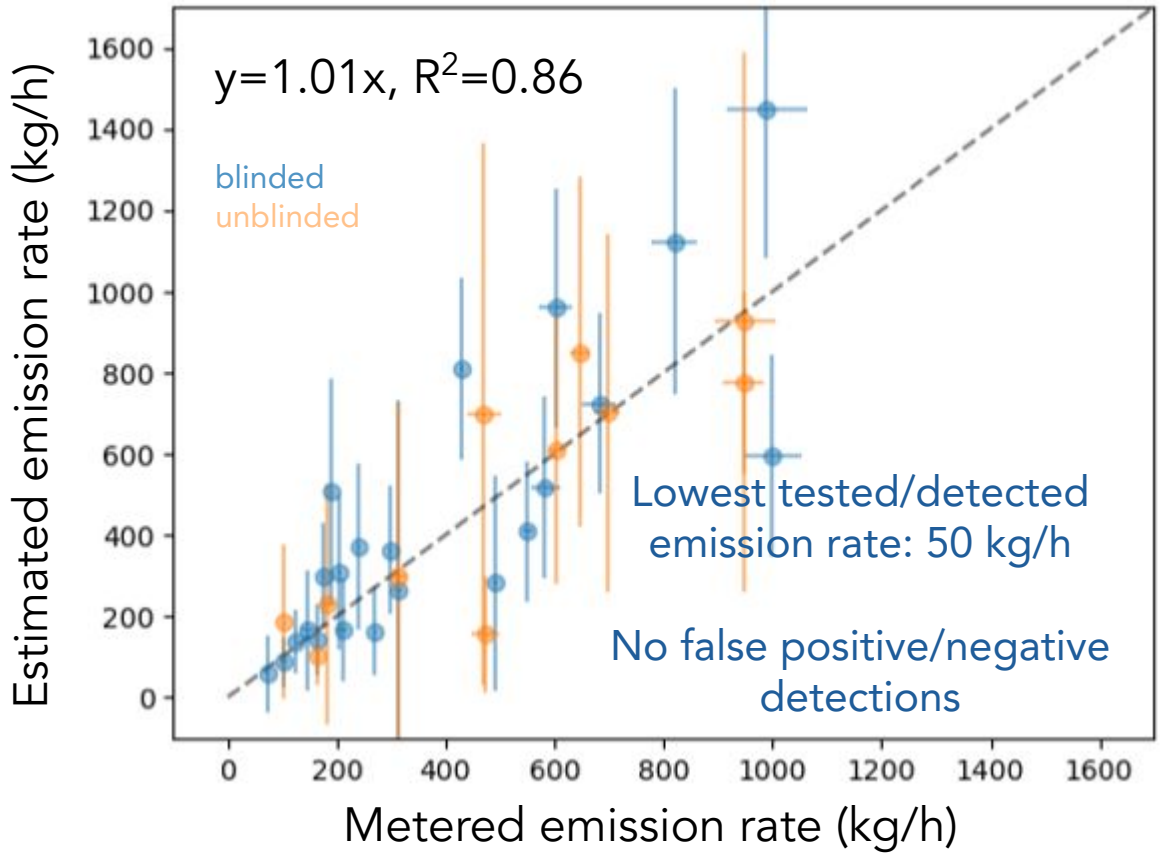
Quantifying emission rate from a plume

Gaussian Plume		$Q = U\Delta\Omega(x,y)\left(\sqrt{2\pi}\sigma_y(x)e^{\frac{y^2}{2\sigma_y(x)^2}}\right)$
Local mass balance		$Q = UW\Delta\Omega$
Gauss theorem		$Q = \oint_s \Omega(s)\vec{U}\cdot\vec{n}\,ds$
Cross-sectional flux (CSF)		$Q = U\int_a^b \Delta\Omega(x,y)dy$
Integrated mass enhancement (IME)		$Q = \frac{U_{\text{eff}}\ \text{IME}}{L}$
Angular width		$Q = f(\text{IME},\theta)$
Machine learning		$Q = \text{CNN}(\text{Plume image})$

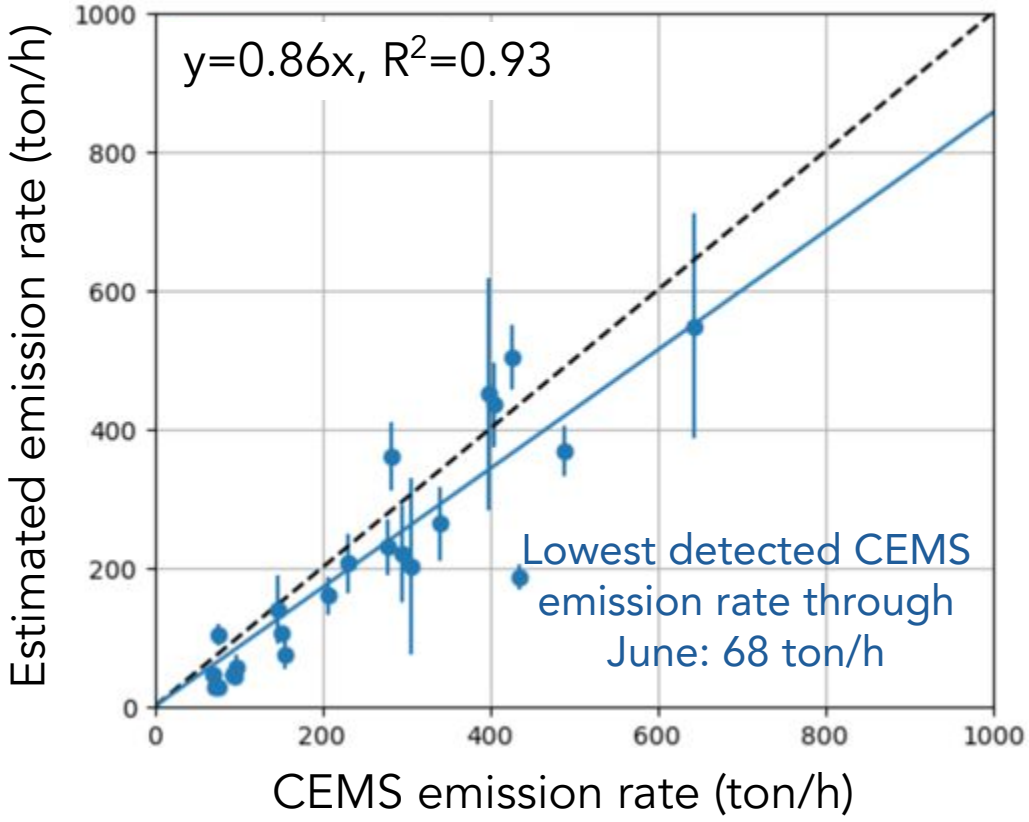
Most common approaches

Controlled release experiments can be performed to assess the bias and uncertainty in quantification methods

Tanager CH4 Controlled Releases (Stanford; Casa Grande, AZ)

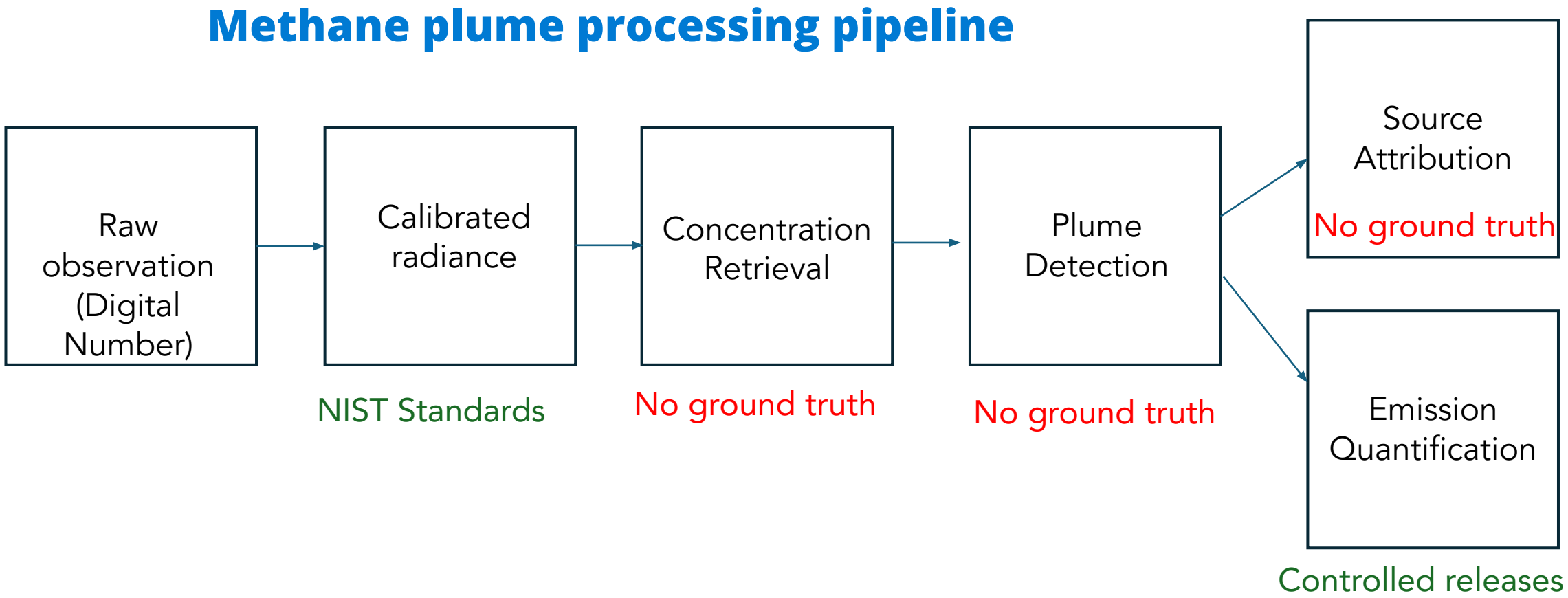


Tanager satellite CO2 Continuous Emissions Monitoring (CEMS, US EPA, Regionally distributed)

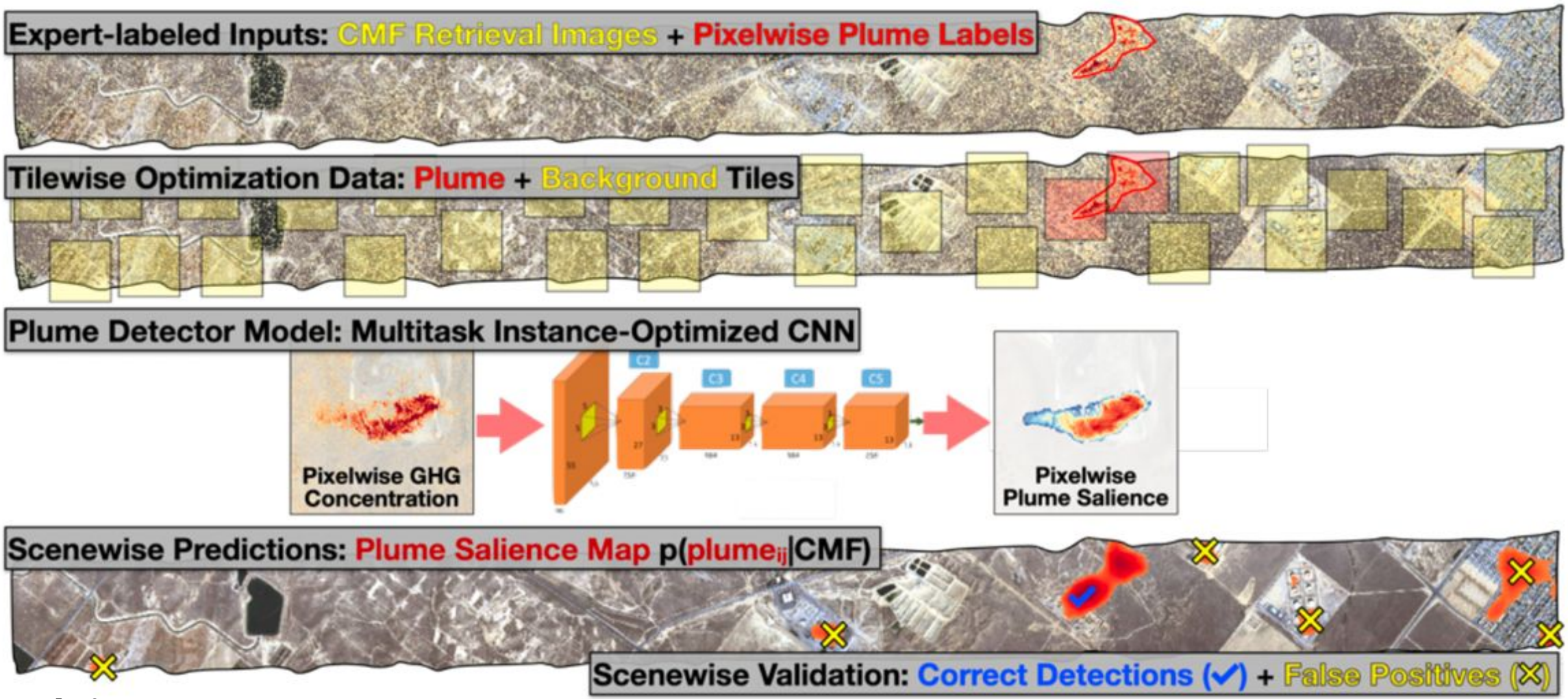


Reuland et al. (submitted)

Limited ground truth

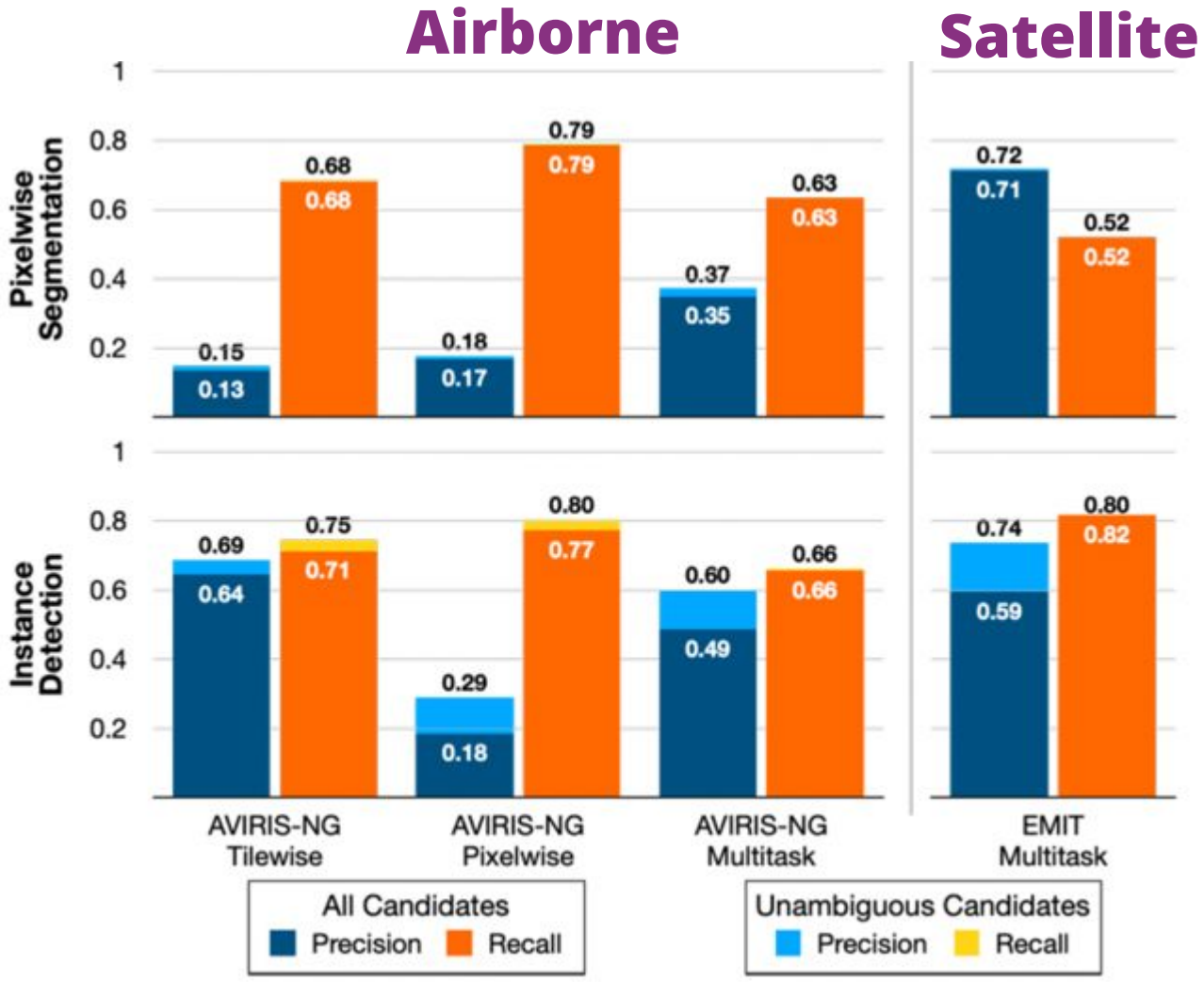


Plume detection is the bottleneck for the plume processing pipeline. Explore methods to reduce latency / need for manual annotation



Bue et al. (in review)

What's the learning objective? Finding enhanced methane pixels (dense segmentation) or finding plumes (instance detection)?

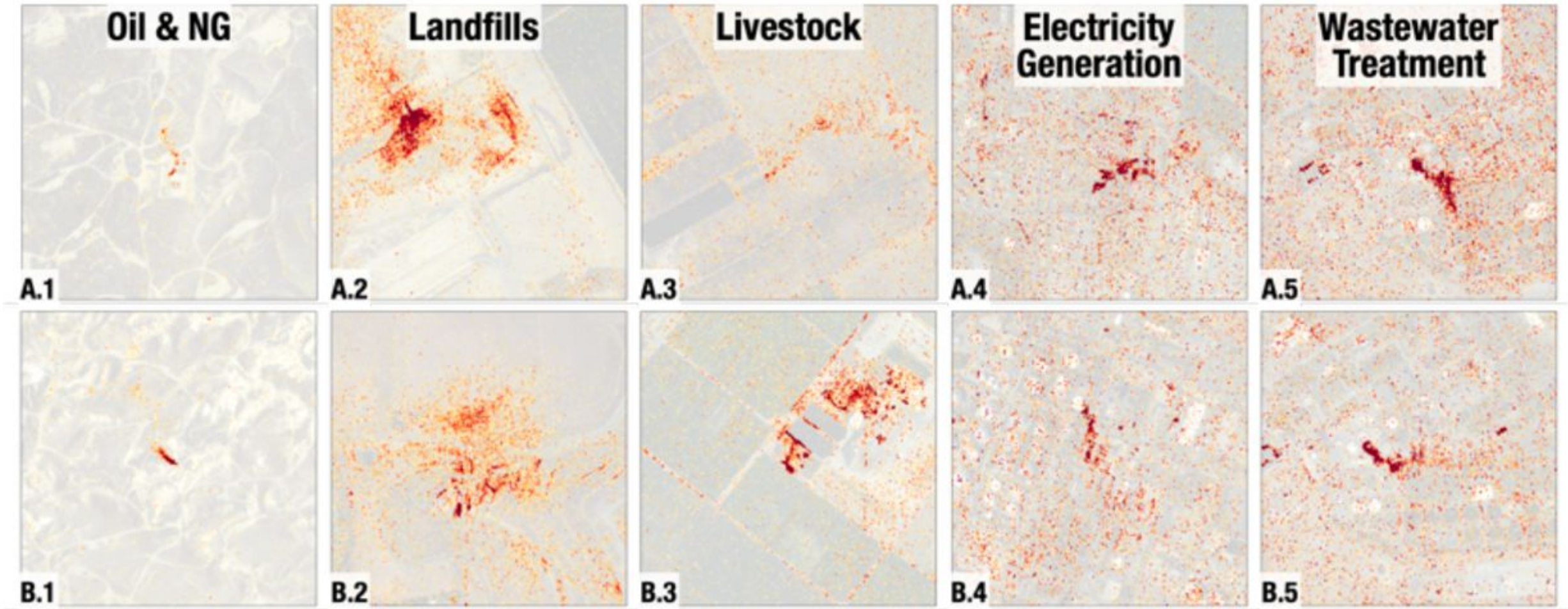


Problem: methane super-emitters are rare: 1000s of training samples

Model architecture and loss function makes a huge difference depending on task

Bue et al. (in review)

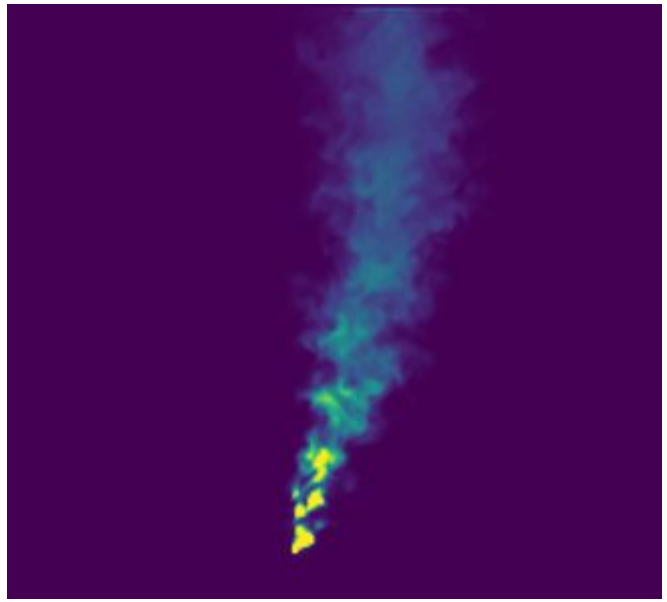
Methane plumes have a wide diversity of shapes depending on the emission sector.
Instrument noise is also regionally and environmentally dependent.

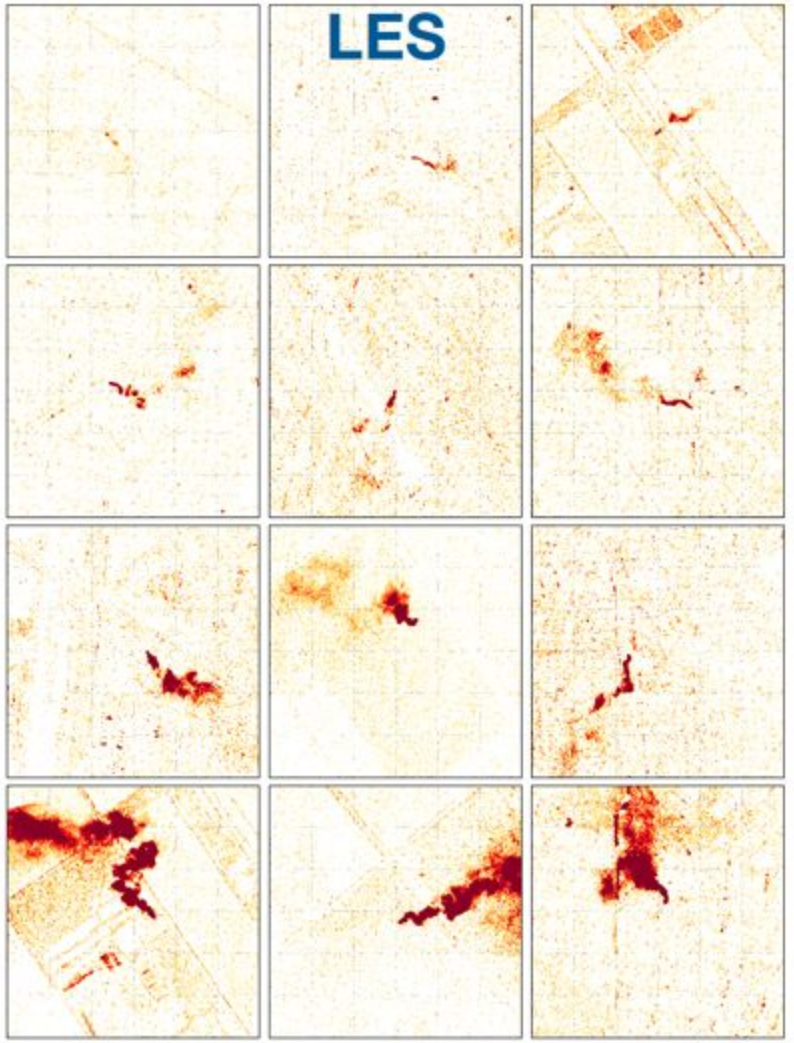


Bue et al. (in review)

Data augmentation: injecting synthetic plumes into realistic imagery to create more samples

Generate fake plumes using large-eddy simulations (LES)



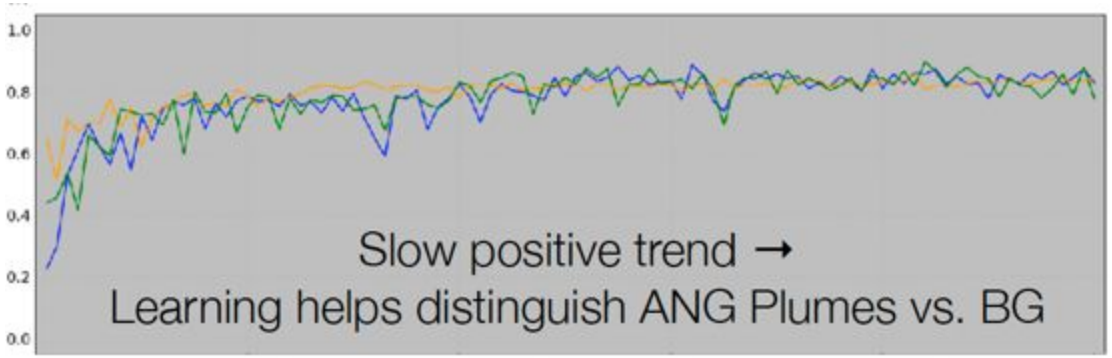


Insert plumes into tiles with realistic noise for training

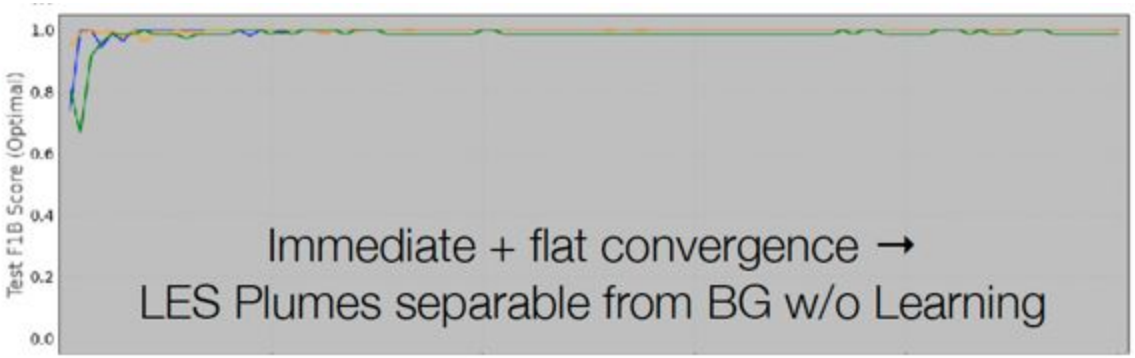
Augmentation has a generalization issue with CNN architectures

F1 Score

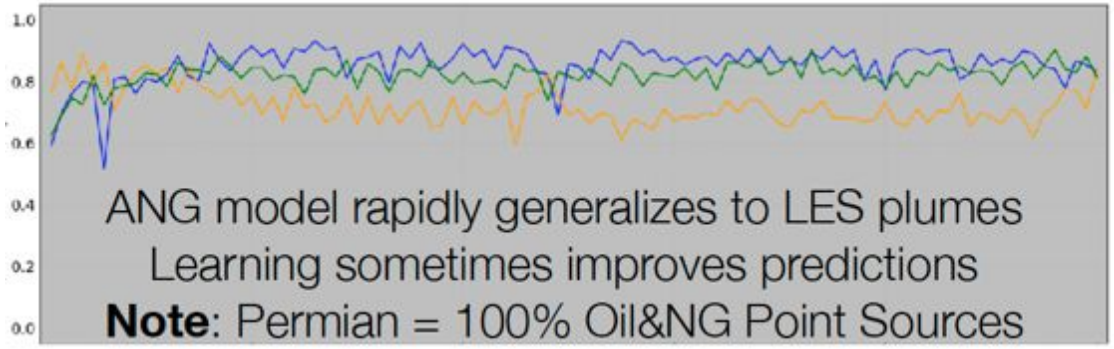
Train Airborne -> Test Airborne



Train LES -> Test LES



Train Airborne -> Test LES



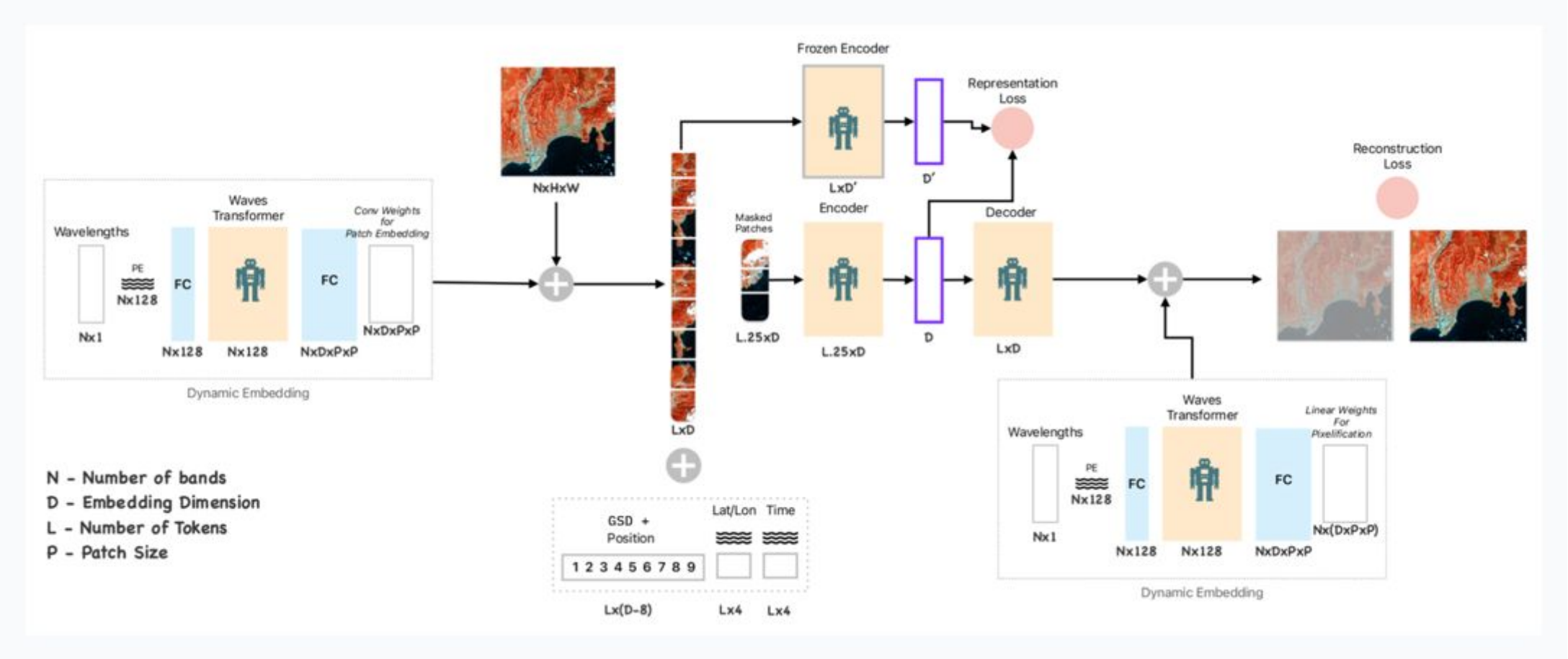
Train LES -> Test Airborne



Epoch

Test of Earth Foundation models on this application

Clay Embeddings



Tested a variety of existing and new embeddings using Tanager, Sentinel-2, and varying chip size.

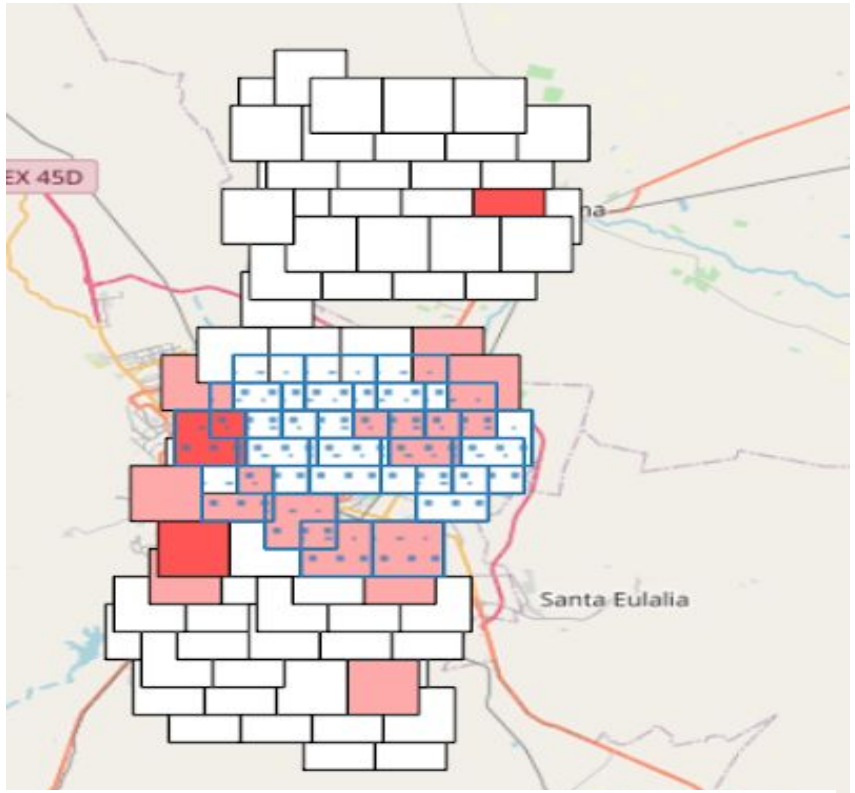
Embeddings suggest ~10 Tanager bands + Sentinel-2 are all that are needed for best classification

Source(s)	Chip Size [meters][1]	# Tanager Bands	Positive F1[2]	Class-weighted F1	Negative F1
Tanager	3840	10	0.64	0.61	0.19
Tanager	**3840**	**All**	**0.74**	**0.69**	0.19
Tanager	2560	10	0.43	0.40	0.17
Tanager	2560	All	0.69	0.64	**0.21**
Tanager, S2	3840	10	**0.74**	**0.69**	0.15
Tanager, S2	3840	All	0.69	0.64	0.12

Poor performance on predicting non-plume class.
Worse performance than CNN-based tests

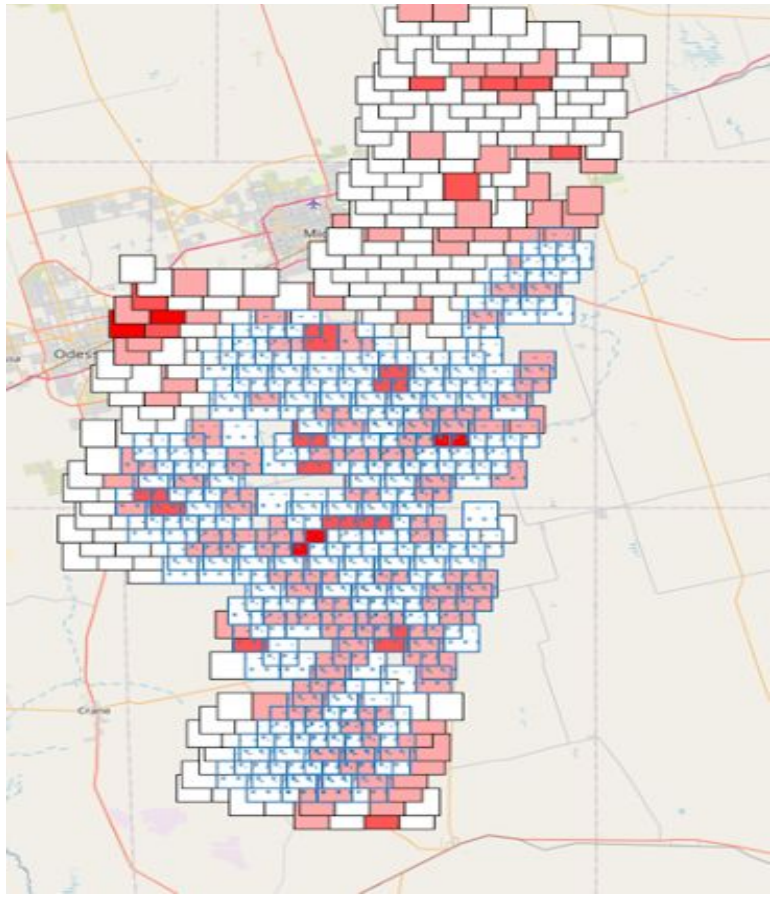
Looking visually at output shows no clear patterns in mispredictions

Example output from Clay model over Oklahoma



Blue dots: known plume locations
White squares: Negative class
Red squares: Positive class

Example output from Clay model over Texas

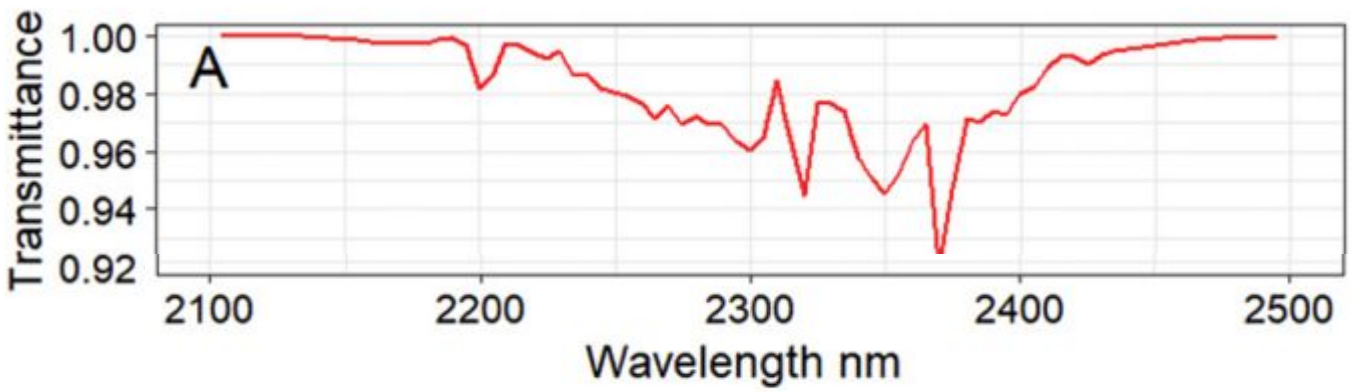


Hypothesis:

Methane absorption is a minor feature in observed solar backscatter

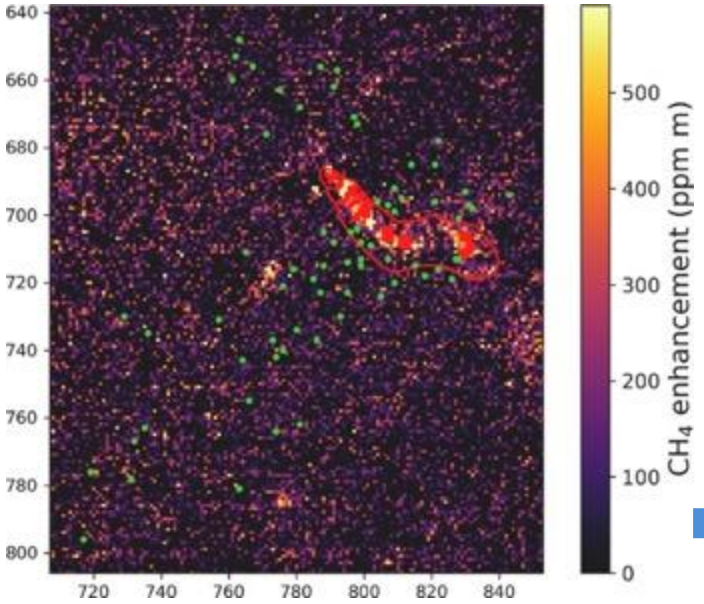
Generic earth foundation models may lose this subtle signature during the embedding creation process

Go back to the spectra, physics, and CNNs: post-hoc assessment

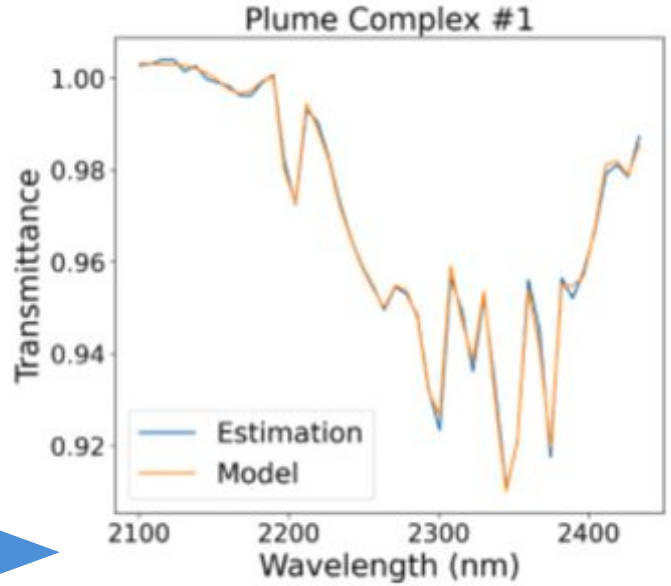


This is what methane absorption looks like spectrally at 5nm band-spacing (i.e., a satellite like Tanager)

So if I have a real plume at take spectra in the plume and out of the plume…



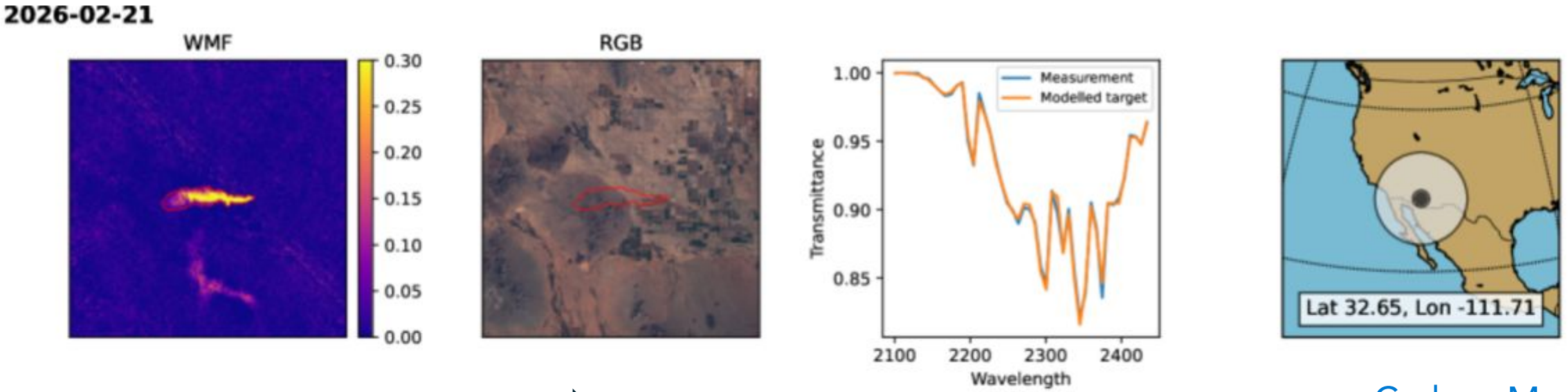
I should get something that looks like absorption



Xiang et al. (2025)

Putting ML + physics intuition together for low latency plume notification

Example of a fully automated plume alert in operations



CNN detection of plume → Passed through automated (non-ML) physics check → Carbon Mapper alerted local authority

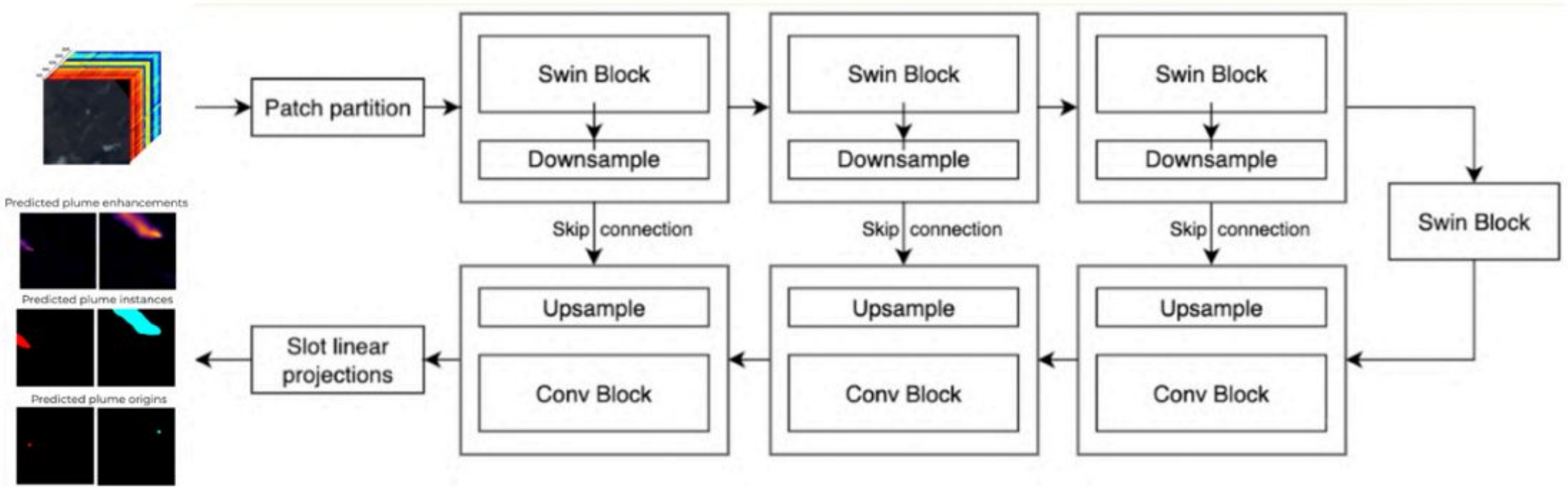
Here – a high false positive ML model is acceptable when combined with deep physics intuition to produce a detection

Question: do AI or ML systems need to be truly end-to-end?

Růžička et al. (2026)

Moving into the world of transformers and massively augmented datasets

MAPL-EMIT: Use a U-net architecture with Swin Transformer encoder



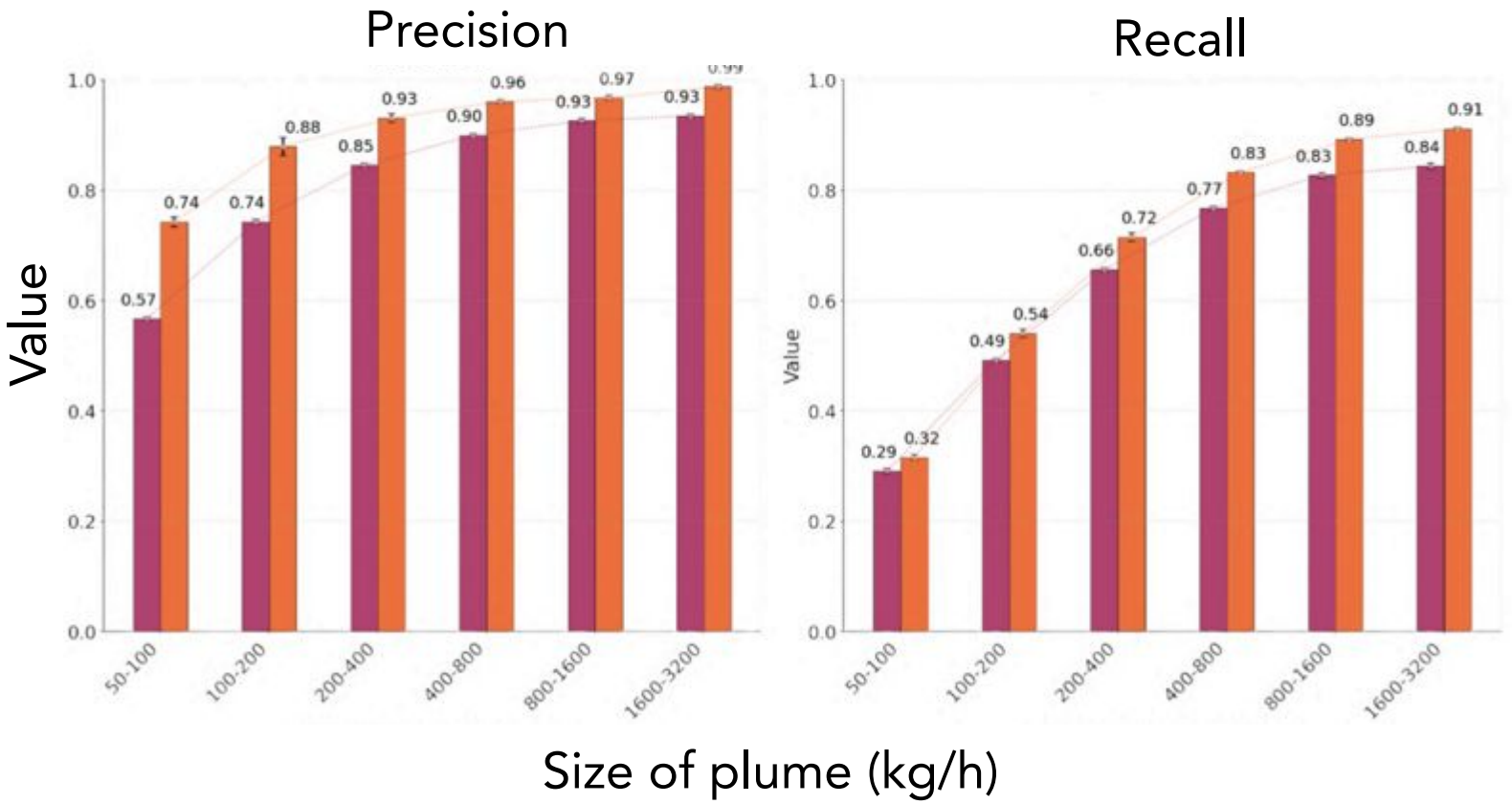
Augmentation via synthetic plumes: 3.6 million synthetic plumes

Training performed on all radiance channels and pixels in an image tile

Batchu et al. (2026)

Moving into the world of transformers and massively augmented datasets

MAPL-EMIT: Use a U-net architecture with Swin Transformer encoder



Encouraging, but still have a "ground truth" problem:

"…only 54% of the MAPL-EMIT plumes were captured in the EMIT L2B plume collection indicating that MAPL-EMIT identified twice as many plumes as human analysts did."

How do we evaluate precision confidently if there is potentially this level of error in human labels?

Batchu et al. (2026)

Discussion

Going back to previous questions, but now through an AI lens:

How could/should we base top-down [or AI] results in ground truth, standards, and validation?

Live Poll: Has your opinion changed regarding trustworthiness of top-down vs AI-driven vs bottom-up accounting?

- Vote: what do you trust more? Top-down (manual); Top-down (AI), bottom-up?

What types of experiments would you suggest/perform to build confidence in top-down/remote sensing estimates

Key Considerations for AI-Based Carbon Accounting

Themes that cut across all the examples in this lecture

Ground-truth & Label Quality

Often, we lack reliable ground truth (retrieval, detection, quantification). Labels vary in quality across geographies. Design your evaluation to account for this.

Geographic & Domain Generalization

Models trained on one region or sensor often fail on another. Evaluate out-of-distribution performance explicitly before deployment.

False Positive vs. False Negative Costs Are Asymmetric

A false positive may trigger a regulatory action or operator fee; a false negative lets emissions go unaccounted. The right operating point on the precision-recall curve depends on the downstream decision, not just the F1 score.

Physics/Explainability as a Guard Rail

Can you explain the outputs of an AI model? Physics keeps us grounded, especially for applications where trust is critical.

Deployment Pathway & Stakeholder Buy-in

Who accepts the output — regulators, operators, auditors? Technical performance alone is insufficient without institutional trust.

Data Availability Drives Scope

Richest data: North America (e.g., GHGRP), EU (e.g., ETS). Sparse data: most of the Global South. Your model's coverage claims may not hold where they matter most for global accounting.

Ethical & Responsible AI in Carbon Accounting

Surveillance implications

- Third-party satellite detection of facility emissions creates an independent monitoring layer. However, it also raises questions:
  - Who should be allowed to report? The EPA's approval pathway for Super Emitter reporters is one model for governing access.
  - Who controls the data?

Uneven Performance Across Geographies

- Models trained on data-rich regions (US, EU) often degrade in regions with sparse labels (South Asia, Africa, Latin America) where accurate accounting may matter most for global progress.
- Benchmark performance on held-out domestic data is not sufficient evidence of deployment readiness globally.

Who Bears the Consequences of Errors?

- False positives: an operator is flagged for a regulatory action they didn't trigger. If the model is wrong, who is liable?
- False negatives: emissions go undetected, undermining the integrity of carbon markets or allowing illegal dumping.
- In both cases, the affected parties are rarely the model developers.

Transparency & Contestability

- A satellite-derived emissions estimate that triggers a regulatory fee must be explainable. Black-box models create challenges for due process.
- What recourse does an operator have to contest a detection? Standards bodies (ISO, WMO, Verra) are beginning to address this for carbon markets.
- Open-source models and public benchmark datasets (METER-ML, EMIT L2B) support scrutiny. Proprietary systems without disclosure undermine trust.

How to Get Started: Datasets, Tools & Open Problems

Open Datasets & Benchmarks

METER-ML

Multi-sensor benchmark for methane infrastructure detection (CAFOs, landfills, wells, processing plants).
https://stanfordmlgroup.github.io/projects/meter-ml/

Climate TRACE

Open facility-level emissions inventory for 352M+ assets across all major sectors. Free API + CSV download.
Climatetrace.org

TROPOMI / Sentinel-5P

ESA's daily global methane, CO column products. Free via Copernicus Open Access Hub.
https://scihub.copernicus.eu/

Carbon Mapper Data Portal

Methane and Carbon Dioxide retrievals from the Tanager, EMIT, and various airborne instruments:
data.carbonmapper.org

Open Research Problems

- Better UQ for end-to-end plume quantification pipelines
- Cross-sensor / cross-geography generalization benchmarks
- Scalable ground-truth collection for hard-to-label sectors (agriculture, cement)
- Integrating AI detections into regulatory reporting frameworks
- Applying foundation models with physics-informed pre-training to spectral RS data
- Closing the gap between atmospheric inversion and facility-level attribution at scale

Wrap-Up: What We Covered & What's Open

✅ What We Covered

- Why & how carbon accounting works
- Bottom-up vs. top-down estimation
- Satellite-based CH_4/CO_2 monitoring
- Atmospheric inversion methods
- AI for activity data
- AI for plume detection
- Using AI to streamline decision making
- Uncertainty quantification

🔭 Open Questions & Beyond This Lecture

- Corporate & supply-chain accounting (lightly touched on)
- Lifecycle assessment with AI
- AI for strengthening audit processes
- Sensor fusion (satellite + ground + aircraft)
- Model governance, standards & certification

Time for Additional Questions/Discussion

Quiz

Q1 *Multiple Choice*

Which of the following best describes the key limitation of bottom-up carbon accounting?

A. It requires satellite observations, which are expensive
B. It relies on self-reported activity data and emission factors, which may not reflect actual local emissions
C. It can only be used for CO_2, not methane
D. It is only applicable to developed countries with strong regulatory frameworks

Q2 *Multiple Choice*

In the high-resolution methane plume processing pipeline, which stage currently has the LEAST available ground truth for training or evaluating AI models?

A. Calibration of raw satellite observations
B. Concentration retrieval using matched filter methods
C. Plume detection
D. Emission quantification

Q3 *True / False*

True or False: Earth foundation models (e.g., Clay) outperformed CNN-based approaches for methane plume detection in the Tanager satellite experiments described in this lecture.

Q4 *Multiple Choice*

A company wants to deploy an AI-based methane plume detector to support a Super Emitter Program reporting initiative. Which consideration should most influence the choice of decision threshold (precision vs. recall tradeoff)?

A. Maximizing overall F1 score on the held-out test set
B. The relative cost of false positives vs. false negatives, given the downstream regulatory and financial consequences
C. Using the same threshold as the best published paper on plume detection
D. Minimizing inference time to enable real-time alerts

Pre-readings

Joyce et al. (2023) — METER-ML: A Multi-Sensor Earth Observation Benchmark for Automated Methane Source Mapping

Jacob et al. (2022) — Quantifying methane emissions from the global scale down to point sources using satellite observations