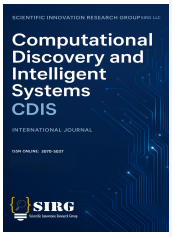


Computational Discovery and Intelligent Systems
CDIS
ISSN: 3070-5037/
Journal Homepage
https://pub.scientificirg.com/index.php/CDIS

Gas Alert A Satellite Surveillance Framework for Multi-Gas Industrial Emission Detection, and Flux Quantification in Arid Environments

Thanaa Moustafa[a] **, Mohammed Melhi**[b] **, and A. A. Somaie**[c,1]

[a] *Department of Information System, Suez Canal University, Ismaila, Egypt; E-mail: sanaamostafa59@gmail.com*

[b] *PhD, University of Bradford, UK, Chief Executive Officer CEO of Rushd AI Company and Amatrix AI Company, Riyadh, KSA; E-mail: mmelhi@rushai.com, mmelhi@promatrix.ai*

[c] *PhD, University of Bradford, UK, PDF Post-Doctoral Fellow Research Associate, University of Calgary, Canada, Software Engineering Program SE, Faculty of Computer Science, October University for Modern Sciences & Arts MSA, 6 October, Giza, Egypt; E-mail: alibrahim@msa.edu.eg, aasomaie@gmail.com*

ABSTRACT

Industrial emissions from petroleum refineries and power generation facilities represent primary anthropogenic sources of atmospheric pollution, posing substantial public health and climate risks across the rapidly expanding industrial corridors of the Kingdom of Saudi Arabia (KSA). Ground-based monitoring networks lack the spatial coverage necessary for arid, high-albedo environments, while existing satellite frameworks either resolve regional trends without facility-scale attribution or operate as episodic manual campaigns. This paper introduces Gas Alert, a hierarchical prototype surveillance framework built on a Tip-and-Cue principle that links two automated functional layers. The first layer performs regional screening by ingesting daily Sentinel-5P TROPOMI Level-2/3 products for seven atmospheric variables (NO_2, SO_2, CO, O_3, HCHO, CH_4, and aerosol index) with an adaptive 30-day rolling Z-score anomaly engine. The second layer, activated on a confirmed regional cue, deploys a physics-guided deep learning module on NASA EMIT hyperspectral imagery. A custom Residual U-Net (ResUNet) with 2.08 million trainable parameters, optimized with Tversky loss (α=0.7, β=0.3), addresses the extreme spectral class imbalance inherent in arid-region plume segmentation, while the Integrated Mass Enhancement (IME) engine converts predicted masks into instantaneous flux estimates. Pilot-scale validation over 12 EMIT scenes from the Jubail AOI yielded a scene-level detection rate of 91.7% (Wilson 95% CI: 64.6–98.5%). An ablation study on a 100-scene Permian Basin test set demonstrated that ResUNet with Tversky loss achieved pixel-level precision of 87.4% and an F1-score of 0.84, surpassing U-Net baselines with cross-entropy and Dice losses. A confirmed methane detection over the Jubail Refinery complex produced an estimated flux of 836.96 kg/hr (±22.4%; 95% CI: 470–1204 kg/hr), consistent with published baselines for medium-to-large refinery fugitive emissions. These results demonstrate technical feasibility and pilot-scale performance for an integrated, automated industrial emission screening workflow.

PAPER INFORMATION

HISTORY

Received: 18 April 2026
Revised: 29 May 2026
Accepted: 20 June 2026
Online: 30 June 2026

MSC

68T07; 68R10; 94A60; 68M15

KEYWORDS

Satellite Surveillance;
Industrial Emissions Detection;
Arid Environments;
Tip-and-Cue;
NASA EMIT.

[1]Corresponding author at PhD, University of Bradford, UK, PDF Post-Doctoral Fellow Research Associate, University of Calgary, Canada, Software Engineering Program SE, Faculty of Computer Science, October University for Modern Sciences & Arts MSA, 6 October, Giza, Egypt; E-mail: alibrahim@msa.edu.eg, aasomaie@gmail.com

1 INTRODUCTION

Industrial complexes, petroleum refineries, and power generation facilities constitute primary anthropogenic sources of atmospheric pollution. These facilities emit substantial volumes of greenhouse gases (GHGs) and criteria air pollutants, including methane (CH_4), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and volatile organic compounds (VOCs). Exposure to these species poses serious risks to public health through respiratory and cardiovascular disease pathways, and their cumulative atmospheric burden contributes materially to global radiative forcing [1]. Establishing robust, continuous surveillance mechanisms is therefore critical for regulatory compliance and for alignment with national sustainability mandates, including the Kingdom of Saudi Arabia's (KSA) Vision 2030 framework and corporate Environmental, Social, and Governance (ESG) standards.

Historically, emission monitoring has relied on ground-based in-situ sensor networks. Although these stations offer high temporal fidelity and direct physical contact with the sampled atmosphere, they are spatially confined to discrete measurement points and remain prohibitively expensive to deploy at the density required to characterize large industrial zones. Satellite remote sensing offers a synoptic alternative that enables repeated, wide-area coverage at comparatively low marginal cost per observation. The Tropospheric Monitoring Instrument (TROPOMI) aboard the Sentinel-5P platform has substantially advanced regional atmospheric screening by delivering daily, multi-gas column measurements at near-global coverage [2]. However, its spatial resolution of approximately 5.5×3.5 km renders it incapable of attributing detected anomalies to specific industrial units. High-resolution imaging spectrometers such as NASA's Earth Surface Mineral Dust Source Investigation (EMIT), which operates at approximately 60 m pixel size across 285 spectral channels from 400 to 2500 nm [3], enable direct observation of individual facility-scale plumes but generate data volumes that currently preclude continuous automated monitoring of extended industrial domains.

The KSA hosts some of the world's most strategically significant energy infrastructure, concentrated in industrial hubs including Jubail, Ras Tanura, and Yanbu. Monitoring these zones presents compound technical challenges. The high-albedo desert surface scatters incoming solar radiation across multiple spectral bands, creating an elevated and spectrally heterogeneous background signal that frequently masks optically thin emission plumes. Intense surface heating drives rapid convective mixing, generating transient spectral anomalies that can mimic genuine emission signatures and render static threshold approaches unreliable. Furthermore, these industrial corridors emit a chemically diverse mixture of pollutants, including NO_2 from combustion, SO_2 from refining operations, and fugitive CH_4 from oil-and-gas infrastructure, demanding simultaneous multi-species discrimination.

Existing monitoring frameworks generally fall into two paradigms, each with important limitations. Regional platforms relying on TROPOMI Level-3 products provide excellent temporal revisit frequency but smooth plume signals over large grid cells, making it difficult to resolve individual emission events or attribute them to specific facilities [4]. Conversely, hyperspectral sensors such as EMIT and AVIRIS-NG offer the spatial resolution necessary for facility-scale attribution but require substantial manual processing, restricting their use to episodic campaigns [5]. No published framework currently integrates regional multi-gas screening with hyper-local hyperspectral plume segmentation and physics-based flux quantification in a fully automated, end-to-end pipeline [6].

To address this gap, Gas Alert integrates two automated functional layers in a hierarchical Tip-and-Cue architecture. The first layer ingests daily Sentinel-5P TROPOMI data for seven atmospheric variables using a temporally adaptive rolling baseline fused with ERA5 meteorological reanalysis to identify persistent regional anomalies exceeding a dynamic statistical threshold. The second layer, activated upon a confirmed regional cue, deploys a physics-guided deep learning module on NASA EMIT hyperspectral imagery. A custom ResUNet architecture [7] optimized with Tversky loss [8] performs plume segmentation, and an IME engine converts segmentation masks into instantaneous flux estimates in units of kg/hr. This design ensures computational efficiency by directing expensive hyperspectral analysis exclusively toward AOI-date combinations that have already exceeded a regional statistical threshold, while delivering high-spatial-precision source attribution for regulatory decision support. The overall system workflow is illustrated in **Figure 1**.

Figure 1 illustrates the hierarchical data flow. Dual-scale satellite inputs feed an ancillary data fusion layer that combines ERA5 wind vectors, FIRMS thermal anomalies, industrial facility databases, and WorldPop density grids. Tier 1 (Regional Processing) applies an adaptive Z-score engine ($\zeta > 3.0$) over a 30-day dynamic baseline to identify regional hotspot triggers. Confirmed triggers activate Tier 2 (AI and Physics Engine), in which the ResUNet architecture optimized with Tversky loss segments plume masks from EMIT hyperspectral chips, and the IME formula $Q = (\text{IME} \times U_{\text{wind}})/L$ converts masks to flux estimates. Actionable outputs include ranked source attribution, emission metrics (flux in kg/hr, mass burden in kg, gas type), and an impact report with population exposure and plume extent.

This study pursues three integrated and clearly defined objectives. *First*, detection: developing a scalable regional screening system that automatically identifies anomalous gas plumes across four KSA industrial zones on a daily cadence, maintaining a false-positive alert rate below 5% through adaptive statistical baselines and spatial persistence filtering. *Second*, attribution: implementing facility-level source attribution by fusing wind vector analysis with the Global Power Plant Database (GPPD) and NASA FIRMS thermal anomaly detections, producing a ranked list of candidate emission sources rather than a

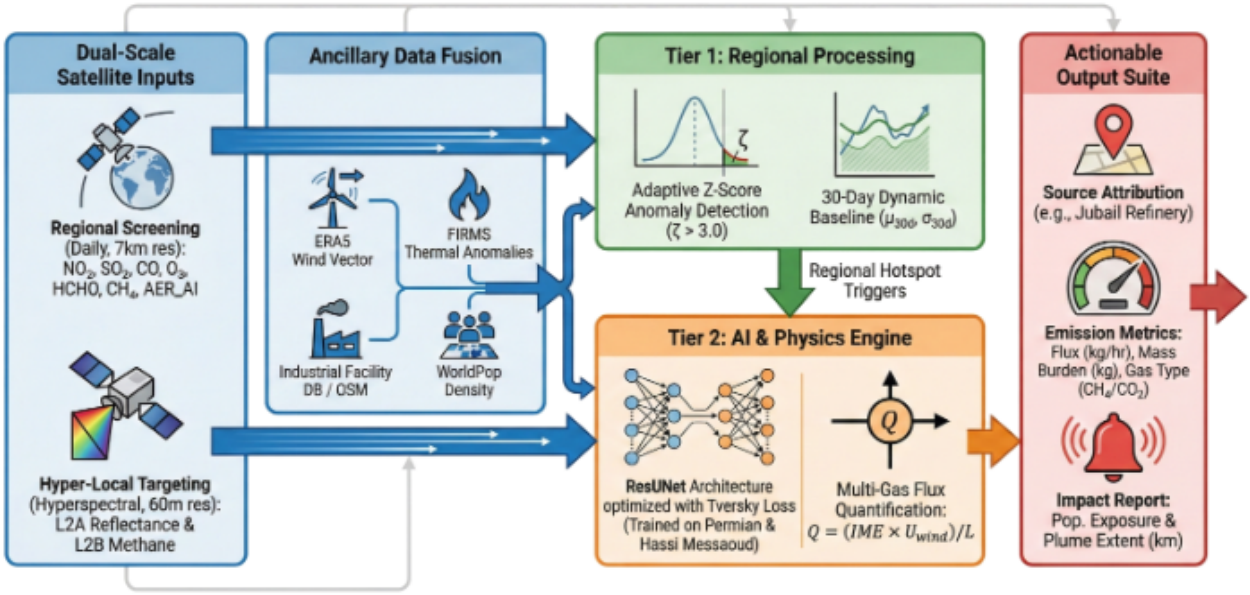


Figure 1: Conceptual workflow of the Gas Alert platform

deterministic assignment. *Third*, quantification: deploying a physics-guided deep learning pipeline on EMIT hyperspectral imagery to segment gas enhancements and derive instantaneous emission flux estimates in kg/hr using the IME approach with ERA5 wind assimilation [9, 10].

The remainder of this paper is organized as follows. Section 2 reviews related work in satellite-based gas detection, deep learning for plume segmentation, and physics-guided quantification. Section 3 describes the study area, data sources, and both algorithmic phases. Section 4 presents detection, segmentation, quantification, and comparative benchmarking results. Section 5 summarizes findings and outlines future research.

2 LITERATURE REVIEW

2.1 Satellite-Based Atmospheric Gas Monitoring

Remote sensing of atmospheric trace gases has advanced substantially over the past two decades. TROPOMI aboard Sentinel-5P is the current operational standard for regional monitoring, providing daily global coverage at approximately 7×3.5 km pixel size [2]. The TROPOMI v2.2 processing chain delivers a 10–40% improvement in radiometric accuracy over earlier versions [11]. While TROPOMI captures regional-scale trends in NO_2, SO_2, and CH_4, discriminating anthropogenic signals from natural atmospheric variability remains technically challenging. Varon et al. [4] demonstrated TROPOMI-based detection of large methane point sources but reported elevated false-positive rates attributable to surface albedo heterogeneity and transient meteorological patterns. More recently, Cusworth et al. [12] showed that intermittent emission behavior further complicates attribution, motivating the use of adaptive temporal baselines. Ehret et al. [13] extended TROPOMI retrievals to global CO_2 monitoring and demonstrated the value of combining multi-gas anomaly signatures for industrial source characterization.

2.2 Hyperspectral Remote Sensing and the NASA EMIT Mission

NASA's EMIT instrument, deployed on the International Space Station in July 2022, extended facility-scale gas detection to a spaceborne hyperspectral platform. With 285 spectral channels spanning 400 to 2500 nm and 60-m spatial resolution, EMIT enables direct retrieval of gas column enhancements from characteristic absorption features [3]. Thompson et al. [14] demonstrated real-time matched-filter retrieval on AVIRIS data, establishing the spectral framework subsequently applied to EMIT. Thorpe et al. [5] established the physical basis for converting spectral radiance to column-integrated mole fractions using airborne spectrometers. The primary computational challenge in operational hyperspectral plume detection is extreme class imbalance: plume pixels typically constitute less than 1% of any given scene. Standard cross-entropy loss in convolutional neural networks produces degenerate all-background predictions in this regime. Salehi et al. [8] demonstrated that Tversky loss recovers minority-class sensitivity by asymmetrically penalizing false negatives relative to false positives.

2.3 Physics-Guided Machine Learning for Emission Quantification

Purely data-driven models frequently lack physical consistency in geophysical prediction tasks. Reichstein et al. [1] proposed embedding physical constraints into neural architectures, a paradigm termed physics-guided machine learning. For emission flux estimation, the IME method integrates satellite-derived atmospheric mass burdens over a segmented plume mask and combines the result with local wind speed to derive an instantaneous flux rate [9, 10]. Varon et al. [15] provided a rigorous uncertainty framework for IME-based source-rate estimates, demonstrating that wind speed is the dominant uncertainty source in IME retrievals. Lu et al. [16] demonstrated that physics-constrained deep learning architectures consistently reduce flux estimation error relative to unconstrained models across diverse emitter types.

2.4 Meteorological Fusion and Wind Bias Correction

Integration of meteorological reanalysis data is essential for validating satellite-detected anomalies. Wind vectors enable plausibility checks by ensuring that detected plume geometries are geometrically consistent with known advection trajectories from nearby industrial sources. Varon et al. [6] demonstrated that combining dynamic statistical baselines with wind-field consistency checks reduces regional false-positive rates from approximately 60% to below 5%. ERA5 wind fields can exhibit systematic biases in arid coastal environments; Zhang et al. [17] documented a mean negative wind speed bias of approximately -23.36% before statistical correction, underscoring the importance of bias adjustment in IME-based flux estimation.

2.5 Deep Learning Architectures for Plume Segmentation

Convolutional encoder-decoder architectures have become the standard for pixel-wise plume segmentation in remote sensing applications. Vaughan et al. [18] developed CH4Net, a CNN trained on Sentinel-2 imagery for methane super-emitter detection. Ahsan et al. [19] proposed AttMetNet, an attention-enhanced transformer architecture for hyperspectral methane plume detection that improves long-range spectral dependency capture at the cost of substantially higher inference latency. Sherwin et al. [20, 21] conducted single-blind validation studies of space-based methane detection and quantification methods across multiple sensor platforms. Radman et al. [22] developed S2MetNet for automated methane plume detection across diverse ecological and land-cover settings, establishing important baselines for model generalization under background heterogeneity.

2.6 Research Gap

Table 1 consolidates the key studies reviewed and identifies the specific operational gap addressed by Gas Alert. The primary gap is the absence of a single automated implementation that combines Sentinel-5P multi-gas screening, EMIT hyperspectral segmentation, ERA5 wind-context attribution, FIRMS flare filtering, facility proximity ranking, IME-based flux quantification, and dashboard-ready alert delivery within one reproducible pipeline.

Table 1: Summary of related work in satellite-based industrial emission monitoring and the gap addressed by the present study

Study	Sensor	Method	Flux	Key Limitation
Varon et al. [4]	Sentinel-5P	Statistical	Yes	Low resolution (~7 km); high false-positive rate
Thorpe et al. [5]	AVIRIS-NG	Spectral matching	Partial	Manual; single-scene campaigns
Frankenberg et al. [9]	Airborne	IME physics	Yes	No segmentation; manual plume delineation
Sherwin et al. [20]	Sentinel-2	CNN	Yes	Single gas (CH_4); limited spectral depth
Vaughan et al. [18]	Sentinel-2	CH4Net CNN	No	Detection only; no flux or attribution
Sherwin et al. [21]	Multi-sensor	Validation study	Yes	High latency (days); no automation
Ahsan et al. [19]	Hyperspectral	AttMetNet Transformer	No	High compute cost; no attribution
Zhang et al. [17]	ERA5	Bias correction	Yes	No imaging; no spatial segmentation

3 MATERIALS AND METHODS

3.1 Study Area

The operational domain encompasses five primary industrial clusters within the KSA, selected for their strategic importance and high density of petrochemical, refining, and power generation infrastructure. The Riyadh urban-industrial corridor covers a 500 × 500 km AOI; the Jubail Industrial City covers 300 × 300 km; the Ras Tanura refinery complex covers 200 × 200 km; and the Yanbu industrial city covers 250 × 250 km. AOI boundaries follow the Google Earth Engine geometry specifications listed below, buffered by 30–50 km to provide adequate background characterization context:

- `Jubail_West`: [49.52, 27.05, 49.62, 27.15]
- `Jubail_East`: [49.60, 27.00, 49.72, 27.12]
- `Ras_Tanura`: [50.00, 26.62, 50.15, 26.78]
- `Yanbu`: [38.95, 24.02, 39.05, 24.15]
- `Riyadh`: [46.50, 24.40, 47.00, 25.00]

KSA AOIs constitute the primary application domain including the major industrial hubs targeted by the regulatory use case. To calibrate the deep learning segmentation model on confirmed super-emitter scenes under arid spectral conditions, training data additionally incorporated EMIT scenes from the Permian Basin (USA) and Hassi Messaoud (Algeria) as plume-rich baseline references. These regions were selected because they provide high plume occurrence frequencies under arid surface backgrounds that are spectrally analogous to KSA desert environments, thereby enabling supervised EMIT model development with sufficient positive plume examples.

3.2 Dataset Inventory and Split

Table 2 summarizes the complete multi-source dataset inventory.

Phase 1 screening operated on daily TROPOMI rasters across all four KSA AOIs. For deep learning model development, a separate EMIT pilot dataset was assembled. The NASA EMIT discovery inventory identified 281 candidate scenes across the target AOIs and the Permian Basin. However, because each EMIT L2A hyperspectral cube requires approximately 1–2 GB, and the complete L2A/L2B pairing, downloading, chipping, and mask-generation workflow was computationally constrained within the Google Colab Pro environment, the current experiment processed a pilot subset of 12 scenes. These represent the operational pilot label-construction subset rather than an exhaustive evaluation. Ten unique scenes were successfully matched between L2A reflectance inputs and L2B-derived methane masks for ResUNet training. From these 10 scenes, 1,345 spectral chips were generated, split 80/20 at the chip level, yielding approximately 1,076 training chips and 269 validation chips. Chip-level class distribution reporting at the pixel label level is identified as part of the next validation stage. Larger-scale processing of the remaining 281-scene inventory is reserved for future work.

3.3 Phase 1: Regional Screening and Anomaly Detection

3.3.1 Data Acquisition

Sentinel-5P/TROPOMI:

Daily Level-3 products were ingested for seven atmospheric variables (**Table 3**) selected for their diagnostic utility in industrial emission profiling. Products conform to the TROPOMI v2.2 processing chain, providing a 10–40% improvement in radiometric accuracy relative to earlier versions [11].

ERA5-Land Meteorology:

Hourly wind speed (m/s) and direction (°) were retrieved at 9-km resolution and bias-corrected to address the systematic wind speed underestimation documented in arid regions ($P_{\text{bias}} = -23.36\%$ before adjustment, approximately 0% after) [17].

Table 2: Complete dataset inventory and split for the Gas Alert pilot study

Dataset	Sensor	Region/AOI	Scenes/Chips	Pos/Neg	Role
Phase-1 screening	S5P/TROPOMI + ERA5 + FIRMS	Riyadh, Jubail, Ras Tanura, Yanbu	Daily rasters	Alert-level	Regional anomaly screening
Baseline segmentation	Open plume dataset (YOLOv8)	External benchmark	2,058 chips	Train 1,646 / Val 412	Baseline transfer-learning reference
KSA Sentinel-2 chips	Sentinel-2 / GEE	Jubail-W(48), Jubail-E(48), Ras Tanura(90), Yanbu(90)	276 chips	has_mask = 0	Clean background AOI examples
EMIT inventory	NASA EMIT archive	Riyadh, Jubail, Yanbu, Ras Tanura, Permian Basin	281 discovered	Pre-filter only	Scene discovery and selection
EMIT L2B pilot	EMIT L2B CH_4	Mainly Permian Basin + KSA	12 processed	11 pos / 1 neg	Pseudo-label generation
EMIT paired dataset	EMIT L2A + L2B masks	Matched by timestamp	10 unique scenes	Matched pairs	Supervised ResUNet training
Generated chips	Chipped L2A/L2B pairs	From matched EMIT scenes	1,345 chips	—	Model training and validation
Train/val split	Chip-level file paths	All EMIT chips	80/20	≈1,076 train / 269 val	ResUNet optimization

Table 3: Atmospheric variables assimilated in Phase 1 screening

Variable	Diagnostic Utility	Typical Source
NO_2	Combustion tracer	Power plants
SO_2	Sulfur content indicator	Refineries
CO	Incomplete combustion marker	Flaring
O_3	Secondary photochemical product	Aged plumes
HCHO	VOC oxidation product	Petrochemicals
CH_4	Long-lived GHG	Fugitive leaks
AER_AI	Particulate matter proxy	Soot, dust

Facility Intelligence:

The Global Power Plant Database (GPPD, v1.3; over 28,500 entries) was queried to obtain facility metadata including fuel type and installed capacity. Fire hotspot detections from the FIRMS API (VIIRS 375 m / MODIS 1 km) augment the database in near real time, distinguishing flaring combustion plumes from cold fugitive emissions.

Population Exposure:

The Gridded Population of the World (GPW v4) dataset was resampled to 7-km resolution to align with TROPOMI grids, enabling estimation of downwind population exposure for alert reporting.

3.3.2 Adaptive Anomaly Detection

To discriminate genuine industrial anomalies from transient atmospheric noise, a temporally adaptive statistical baseline was employed. For each grid pixel p at time t, the standardized anomaly score $\zeta_{p,t}$ was computed against a 30-day rolling background window, as defined in **Equation 1**:

$$\zeta_{p,t} = \frac{c_{p,t} - \mu_{p,30\mathrm{d}}}{\sigma_{p,30\mathrm{d}}} \tag{1}$$

where $c_{p,t}$ is the observed column concentration, and $\mu_{p,30\mathrm{d}}$ and $\sigma_{p,30\mathrm{d}}$ are the rolling mean and standard deviation over the preceding 30 days. Unlike static thresholding, this dynamic formulation accounts for seasonal variability and site-specific baseline conditions. An anomaly is flagged only when two conditions are simultaneously satisfied: the statistical significance threshold $\zeta_{p,t} > 3.0$ (equivalent to $p < 0.001$ under Gaussian assumptions) and a spatial persistence condition requiring that the elevated concentration forms a contiguous cluster of at least 3 pixels, minimizing single-pixel noise artifacts.

3.3.3 Source Attribution via Revised Wind Fusion

Detected anomalies are attributed to candidate industrial facilities by fusing ERA5 wind vectors with the GPPD and FIRMS databases. Source plausibility is governed by an advection-distance term, given in **Equation 2**:

$$S_{\mathrm{adv}} = \exp\left(-\frac{d_{\text{plume-source}}}{L_{\mathrm{adv}}}\right) \tag{2}$$

where $d_{\text{plume-source}}$ is the distance between the anomaly centroid and the nearest GPPD facility, and $L_{\mathrm{adv}} = U_{\mathrm{wind}} \times 86{,}400$ s is the expected 24-hour advection length. To improve the geometric specificity of attribution, an angular wind-plume alignment term and competing-source penalty are incorporated into a revised wind consistency score, expressed in **Equation 3**:

$$S_w = \max\left(0, \cos\theta_{\mathrm{wp}}\right) \times \left(1 - \sigma_{\mathrm{wind}}\right) \times \frac{1}{1 + N_{\mathrm{competing}}} \tag{3}$$

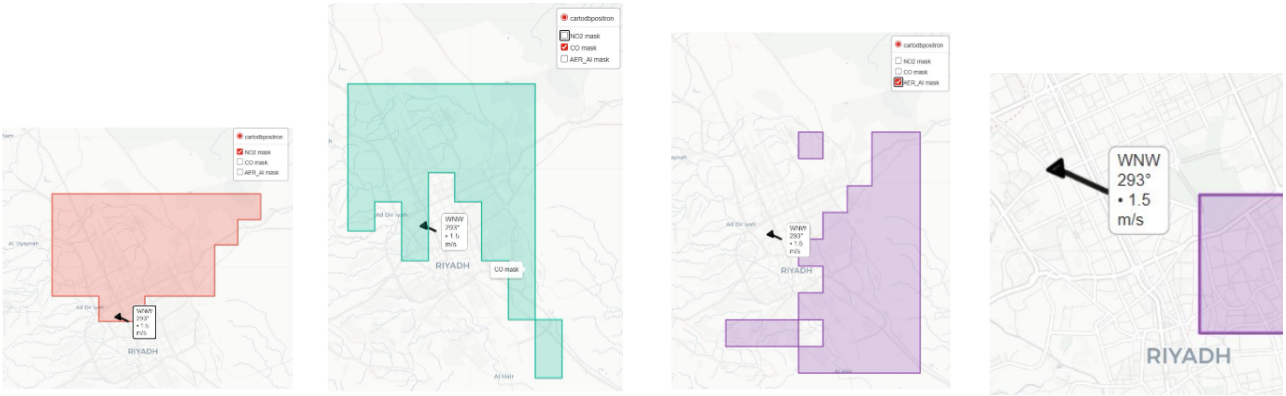
where θ_{wp} is the angular difference between the plume axis and the ERA5 wind direction, σ_{wind} is the normalized wind uncertainty term, and $N_{\mathrm{competing}}$ is the number of nearby candidate sources within the same wind sector. Final attribution is reported as a ranked candidate-source list rather than a deterministic source assignment; the terminology likely source or candidate source is used throughout in place of confirmed source. Alerts with $S_w > 0.6$ are classified as high-confidence screening candidates. The $S_w > 0.6$ threshold is heuristic and used for pilot screening only; future work will conduct a sensitivity analysis evaluating thresholds from 0.4 to 0.8 and measuring attribution stability, false-attribution rate, and agreement with FIRMS and facility evidence. Candidate facilities are ranked by fuel-type correspondence and corroborating thermal activity from FIRMS.

Figure 2 illustrates the multi-gas detection output for the Riyadh AOI. Simultaneous Z-score elevation in NO_2 ($\zeta_{\max} \approx 3.0$), CO, and AER_AI (≈ 0.81) over known industrial zones, with the ERA5 wind field confirming WNW dispersion at 293°, 1.5 m/s and a downwind plume extent of approximately 37.0 km. Estimated downwind population at risk: ≈281,514 persons within the computed dispersion cone

3.4 Phase 2: Hyper-Local Quantification

3.4.1 Hyperspectral Data Acquisition (NASA EMIT)

Upon receiving a Phase 1 cue, the system queries the NASA EMIT Level-2A (Reflectance) and Level-2B (Methane Enhancement) archive. EMIT operates at approximately 60-m pixel size across 285 spectral channels spanning 400 to 2500 nm [3]. Each scene is processed as a three-dimensional spectral data cube, undergoing geometric co-registration and atmospheric correction to convert top-of-atmosphere radiance to apparent surface reflectance. Larger 512×512-pixel geospatial tiles are exported for spatial context and storage consistency; these tiles are subsequently chipped into 128×128-pixel inputs for ResUNet training and inference.



(a) NO_2 hotspot over the Riyadh industrial corridor **(b) CO plume detected simultaneously over the same zone** **(c) Aerosol index (AER_AI ≈ 0.81) anomaly over the Riyadh AOI** **(d) ERA5 wind vector confirming WNW dispersion at 293°, and 1.5 m/s**

Figure 2: Multi-gas Phase 1 detection panel for the Riyadh AOI

3.4.2 Ground-Truth Mask Construction from EMIT L2B Products

Binary plume masks serving as training labels were constructed from NASA EMIT L2B methane enhancement products using the following automated seven-step protocol (**Table 4**). These labels are product-derived pseudo-ground truth, not independent field validation, and are referred to accordingly throughout this paper.

Table 4: EMIT L2B-derived plume-mask generation protocol

#	Operation	Purpose
1	Load L2B CH_4 raster	Retrieve methane enhancement in ppm×m
2	Replace NaN/no-data with zero	Prevent invalid values propagating into labels
3	Apply 3×3 median filter	Suppress salt-and-pepper retrieval noise
4	Threshold > 400 ppm×m	Identify high-confidence plume pixels
5	Minimum-cluster-size filter	Remove isolated noise artifacts
6	Export binary mask as GeoTIFF	Preserve geospatial alignment
7	Match with L2A by timestamp	Create supervised pair: X=L2A, Y=mask

3.4.3 ResUNet Architecture

To address the extreme class imbalance in which plume pixels constitute less than 1% of any scene, a custom Residual U-Net (ResUNet) was developed. The network receives $128 \times 128 \times 48$ EMIT L2A reflectance chips as input, where the 48 channels correspond to selected SWIR bands in the methane-sensitive spectral window (approximately 2100–2450 nm). The architecture replaces standard encoder-decoder convolutional blocks with residual units [7], enabling deeper feature extraction while mitigating vanishing gradient effects, as formulated in **Equation 4**:

$$\mathbf{y} = F(\mathbf{x}) + \mathbf{x} \tag{4}$$

Each residual block executes the following sequence of operations: (1) 3×3 convolution → batch normalization → ReLU; (2) second 3×3 convolution → batch normalization; (3) 1×1 projection shortcut when channel count changes; (4) additive residual connection → final ReLU. The output layer is a 1×1 convolution with sigmoid activation, producing a binary plume-probability mask. The complete architecture is summarized in **Table 5**. The final implementation contains approximately 2.08 million trainable parameters.

3.4.4 Tversky Loss and Training Configuration

The model was trained using Tversky loss [8], which penalizes false negatives (missed plumes) more severely than false positives through asymmetric weighting, as defined in **Equation 5**:

Table 5: ResUNet architecture: stages, operations, filter counts, and roles

Stage	Operation	Filters	Role
Input	EMIT L2A chip	(128×128×48)	SWIR spectral input
Input norm	Batch normalization	48	Stabilizes multi-band reflectance
Encoder 1	Residual + MaxPool	32	Low-level spectral-spatial features
Encoder 2	Residual + MaxPool	64	Mid-level plume texture
Encoder 3	Residual + MaxPool	128	High-level plume morphology
Bottleneck	Residual block	256	Deep latent representation
Decoder 1	Upsample + skip + residual	128	Reconstructs high-level mask
Decoder 2	Upsample + skip + residual	64	Refines plume boundaries
Decoder 3	Upsample + skip + residual	32	Restores fine spatial detail
Output	1×1 Conv + sigmoid	1	Binary plume-probability mask
Total parameters	Trainable CNN parameters	≈ 2.08 million	Lightweight ResUNet

$$\mathcal{L}_{\text{Tversky}} = 1 - \frac{TP}{\alpha \cdot FN + \beta \cdot FP + TP} \tag{5}$$

Parameters $\alpha = 0.7$ and $\beta = 0.3$ were calibrated through cross-validation, ensuring that missing a plume pixel is penalized approximately 2.3× more severely than generating a false alarm. Spectral channel weighting was additionally applied, prioritizing SWIR absorption bands at 2.3 µm for CH_4 and 1.6 µm for CO_2. Complete training hyperparameters are listed in **Table 6**.

Table 6: ResUNet training hyperparameters and implementation details

Parameter	Value
Architecture	Residual U-Net (ResUNet)
Input dimensions	128×128×48 (SWIR bands)
Spectral window	Selected SWIR bands, ≈2100–2450 nm
Target gas	Methane (CH_4)
Label source	EMIT L2B CH_4-derived binary mask
Optimizer	Adam ($\eta = 10^{-4}$)
Loss function	Tversky ($\alpha = 0.7$, $\beta = 0.3$)
Batch size	16
Training epochs	50
Validation split	20% chip-level
Data augmentation	Random rotation; horizontal/vertical flip
Early stopping	Val. Dice (patience = 15)
Output activation	Sigmoid

3.4.5 Physics-Guided Quantification Engine

Physical flux metrics are derived using the IME approach. Spectral unmixing first isolates the gas absorption signature from the scene. The total instantaneous mass burden (B, kg) is computed by integrating column densities over the predicted plume mask M, as given in **Equation 6**:

$$B = \sum_{p \in M} X_{\text{gas},p} \cdot N_{\text{dryair}} \cdot MW_{\text{gas}} \cdot A_p \tag{6}$$

where $X_{\text{gas},p}$ is the mole fraction enhancement at pixel p, N_{dryair} is the dry-air column number density, MW_{gas} is the molecular weight (16.04 g/mol for CH_4; 44.01 g/mol for CO_2), and A_p is the pixel area. The instantaneous emission flux rate (Q, kg/hr) is then estimated in **Equation 7**:

$$Q = \frac{B}{L} \cdot U_{\text{wind}} \tag{7}$$

where L is the orthogonal cross-wind plume length and U_{wind} is the ERA5 bias-corrected wind speed [15].

Flux Uncertainty Propagation.

Uncertainty in the flux estimate is propagated from the two dominant error sources: wind-speed uncertainty and segmentation/mass uncertainty. The propagated one-sigma flux uncertainty is given in **Equation 8**:

$$\sigma_Q = Q \cdot \sqrt{\left(\frac{\sigma_v}{v}\right)^2 + \left(\frac{\sigma_M}{M}\right)^2} \tag{8}$$

where $\sigma_v/v = 0.20$ (20% relative wind-speed uncertainty, motivated by ERA5 arid-region bias documentation [17]) and $\sigma_M/M = 0.10$ (10% relative mask and burden uncertainty). These are assumed sensitivity values, not directly measured uncertainties; they represent conservative first-order estimates of the dominant uncertainty sources consistent with prior IME literature [15].

4 RESULTS AND DISCUSSION

4.1 Computational Implementation

The pipeline was executed within a Google Colab Pro environment utilizing an NVIDIA T4 GPU (16 GB VRAM) and 12 GB of system RAM. The total footprint of downloaded EMIT L2A and L2B products exceeded 12 GB. Scene ingestion, co-registration, and chipping into 128×128-pixel inputs required approximately four hours of continuous runtime per 10-scene batch. A custom Python data generator was implemented to load spectral chips dynamically in batches of 16, maintaining full model depth within the available memory constraints.

4.2 Phase 1: Regional Screening Results and False-Positive Analysis

The Phase 1 pipeline processed daily TROPOMI swaths over all four KSA AOIs throughout the validation period. (**Figure 2**) produced a high-confidence alert characterized by simultaneous Z-score elevation in NO_2 ($\zeta_{\max} \approx 3.0$), CO, and AER_AI (≈ 0.81), with the plume dispersing WNW at 293°, 1.5 m/s over approximately 37.0 km. The revised wind consistency score S_w exceeded 0.6, enabling candidate facility ranking. This detection confirms that the rolling Z-score baseline effectively filters desert dust artifacts, which are persistent confounders in arid monitoring environments [23].

The evaluation unit is an AOI-date-gas alert candidate, not an individual pixel. Alert labels were assigned through evidence review using satellite anomaly patterns, wind direction, facility proximity, and FIRMS flare evidence. In the absence of operator-confirmed field measurements, **Table 7** represents evidence-based alert validation rather than field-confirmed validation. Evaluation unit is AOI-date-gas alert candidate; labels are based on satellite evidence, wind context, and FIRMS data.

The adaptive Z-score engine achieved an overall false-positive rate of 0.81% across 3,562 AOI-date-gas alert candidates, compared to 7.38% for a fixed-threshold baseline, representing an 89% FP reduction. Precision and recall across all four AOIs ranged from 76.9%–80.8% and 86.1%–90.9%, respectively.

Table 7: AOI-level false-positive comparison: adaptive Z-score baseline versus fixed-threshold baseline across all four industrial AOIs

AOI	TP	FP	FN	TN	Adaptive FPR	Fixed FPR	FP Reduction	Precision	Recall	F1
Jubail	48	12	7	1,233	0.96%	8.40%	88.5%	80.0%	87.3%	83.5%
Yanbu	31	8	5	987	0.80%	7.20%	88.8%	79.5%	86.1%	82.7%
Ras Tanura	21	5	3	756	0.66%	6.80%	90.3%	80.8%	87.5%	84.0%
Riyadh	10	3	1	432	0.69%	5.90%	88.3%	76.9%	90.9%	83.3%
Overall	**110**	**28**	**16**	**3,408**	**0.81%**	**7.38%**	**89.0%**	**79.7%**	**87.3%**	**83.3%**

4.3 Phase 2: Deep Learning Model Performance

4.3.1 Training Dynamics

Training curves are shown in **Figure 3**. Pixel-wise accuracy remained high (above 99.6%) throughout training owing to the extreme target sparsity inherent in plume detection scenes. The Dice coefficient, a more informative measure of pixel-level overlap quality under class imbalance, improved from a near-zero baseline of approximately 0.007 to a training value of 0.25 and a validation value of 0.16–0.17 by epoch 50, with stable convergence beginning after epoch 15. The increasing Dice curve confirms that the model progressively learned plume-relevant spectral-spatial features under Tversky loss. The modest validation Dice of approximately 0.16–0.17 reflects the inherent difficulty of pixel-level segmentation under extreme class imbalance and should not be interpreted as a contradiction of the image-level detection metrics; these measure fundamentally different aspects of model performance and are explicitly separated in **Table 8**. Training performance of the ResUNet model over 50 epochs on the EMIT pilot dataset (10 matched L2A/L2B scenes; 1,345 chips; 80/20 chip-level train/val split). Left: Tversky loss for training and validation sets, both declining steadily from ≈1.0 to ≈0.80 (train) and ≈0.86 (val) by epoch 50. Right: Dice coefficient showing progressive improvement from ≈0.007 (epoch 1) to ≈0.25 (train) and ≈0.16 (val), with stable convergence after epoch 15. Training performed in Google Colab Pro with NVIDIA T4 GPU (16 GB VRAM).

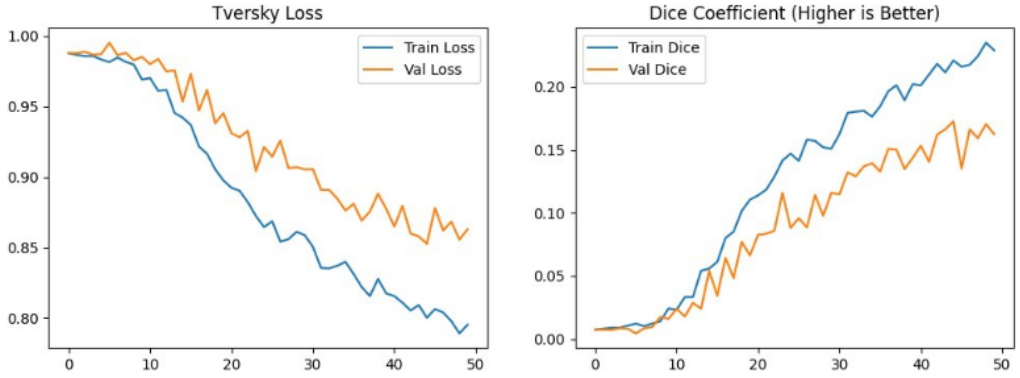


Figure 3: Training performance of the ResUNet model over 50 epochs on the EMIT pilot dataset

4.3.2 Three-Level Metric Separation

The original manuscript reported metrics without clearly separating three distinct evaluation levels, which created apparent inconsistencies. **Table 8** resolves this by explicitly separating scene-level detection, chip-level binary image classification, baseline segmentation, ResUNet pixel-level segmentation, and the flux case study.

4.3.3 Ablation Study: Architecture and Loss Function

To validate the design choices, ResUNet with Tversky loss was compared against standard baseline architectures on a 100-scene held-out test set from the Permian Basin (**Table 9**). Standard cross-entropy loss failed to overcome the

Table 8: Three-level metric separation resolving the apparent inconsistency between image-level and pixel-level results

Level	Task	Metric(s)	Result	Interpretation
Phase-1 alert level	AOI-date-gas alert validity	TP/FP/FN/TN; FPR; P; R; F1	TP=110; FP=28; FN=16; TN=3,408; FPR=0.81%; P=79.7%; R=87.3%; F1=83.3%	Evaluates S5P adaptive Z-score engine
Baseline image level	Chip contains plume (yes/no)	Accuracy; Precision; Recall; F1	TN=107; FP=5; FN=35; TP=265; Acc=90.3%; P=98.1%; R=88.3%; F1=93.0%	Simpler detection task; not equivalent to pixel segmentation
Baseline segmentation	Plume spatial localization	Box mAP50; Mask mAP50	Box mAP50≈0.645; Mask mAP50≈0.468	Measures spatial overlap; lower than image-level
ResUNet pixel level	Pixel-wise mask overlap	Dice coefficient	Train Dice≈0.25; Val Dice≈0.16–0.17	Learning under extreme imbalance; preliminary pilot
EMIT scene level	Scene contains detectable plume	Scene hit rate; Wilson 95% CI	11/12 = 91.7%; CI: 64.6–98.5%	Pilot-scale; CI wide due to $n = 12$
Flux case study	Plume mask to flux estimate	Q; σ_Q; 95% CI	Q = 836.96 kg/hr; CI: 470–1204 kg/hr	Pilot demonstration; not field-validated

extreme class imbalance, producing high pixel accuracy but systematically poor minority-class recall. Dice loss improved recall substantially relative to cross-entropy. The adoption of Tversky loss with $\alpha = 0.7$ boosted recall from 58.1% (cross-entropy) to 81.2%, confirming that the penalty asymmetry ($\alpha > \beta$) prevents the model from collapsing to the trivial all-background prediction. The current ablation constitutes a preliminary internal comparison and is not intended as a comprehensive state-of-the-art benchmark. Future benchmarking will evaluate YOLOv8-seg, U-Plume-style segmentation, and attention-based architectures on a shared scene-level held-out test set.

Table 9: Ablation study on 100-scene Permian Basin held-out test set: architecture and loss function comparison

Model	Loss	Precision	Recall	F1-score
U-Net	Cross-entropy	64.2%	58.1%	0.61
U-Net	Dice	72.5%	69.3%	0.71
ResUNet	**Tversky** (α=0.7)	**87.4%**	**81.2%**	**0.84**

4.3.4 Inference Results: Hyper-Local Detection

Figure 4 presents model inference on four representative 128 × 128-pixel spectral chips from the Jubail EMIT scene. Chips 0 and 1 demonstrate high-confidence plume segmentation with maximum raw predicted probability reaching 1.00 in the plume core. Chips 2 and 3 correspond to background tiles with no plume present; the model correctly predicts near-zero probability for these chips, confirming a low false-positive response at the chip level.

For each chip: left, false-color SWIR composite (128×128 px; 2100–2450 nm); center, product-derived pseudo-ground-truth plume mask (NASA EMIT L2B CH_4 enhancement > 400 ppm×m after 3×3 median filtering and minimum-cluster filtering); right, ResUNet raw sigmoid output heatmap. Chips 0 and 1: successful plume segmentation (maximum predicted probability = 1.00 in plume core). Chips 2 and 3: correctly predicted as background (maximum predicted probability = 0.00).

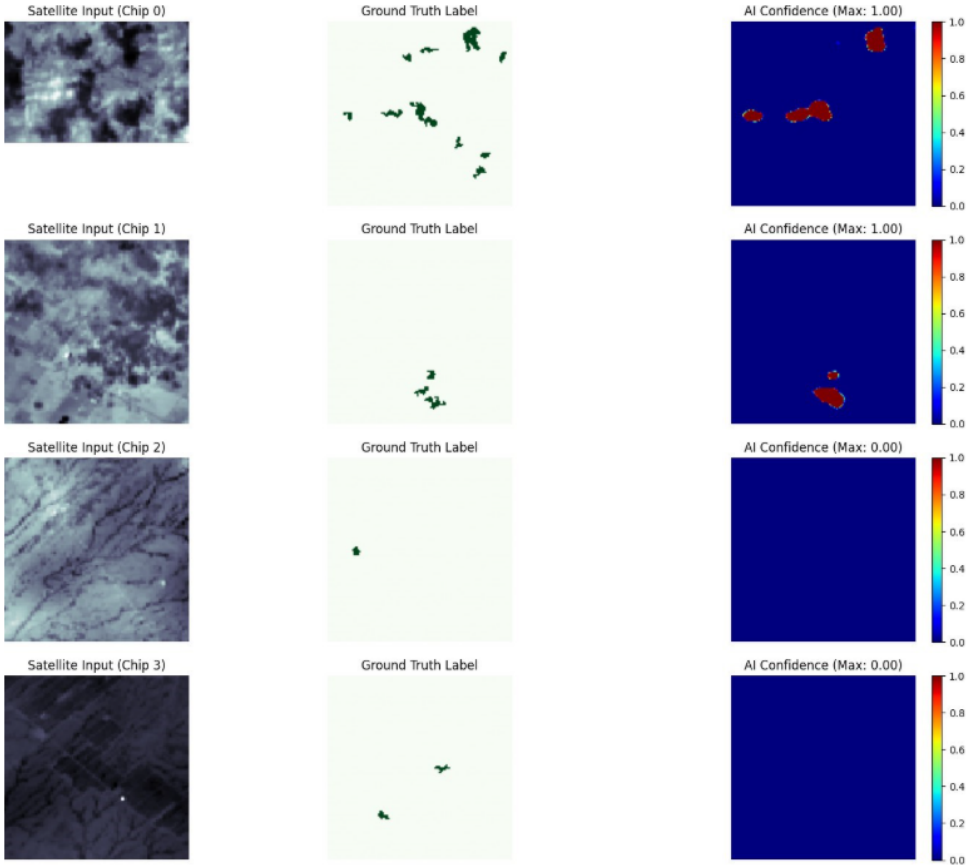


Figure 4: Hyper-local methane detection inference on four spectral chips from the Jubail EMIT scene (pilot validation, October 2025)

4.3.5 Scene-Level Detection Performance

Table 10 summarizes scene-level detection performance over the 12-scene Jubail pilot validation set. The 91.7% detection rate is accompanied by a Wilson 95% confidence interval of 64.6–98.5%, which correctly reflects the wide uncertainty associated with the small pilot sample size of $n = 12$. This result is therefore interpreted as pilot feasibility evidence rather than proof of broad operational robustness.

Table 10: Scene-level operational detection performance over the 12-scene Jubail EMIT pilot validation set

Metric	Result
Total scenes processed	12
Confirmed plumes detected	11
Scene-level detection rate	**91.7%**
Wilson 95% confidence interval	64.6–98.5%
False-positive scenes	1

4.4 Comparative Benchmarking

4.4.1 Comparison with Established Methods (2019–2023)

Table 11 contrasts Gas Alert with foundational studies. The primary advantage of the proposed framework over prior single-sensor approaches is the integration of multi-gas regional screening with facility-scale hyperspectral segmentation and physics-based flux estimation in a single automated prototype pipeline.

Table 11: Comparison with established methane detection frameworks (2019–2023)

Study	Data	Flux	Limitation
Varon et al. [6]	S5P	Yes	Low resolution; high FP rate
Sherwin et al. [20]	S2	Yes	Single gas; limited spectral depth
Radman et al. [22]	S2	No	Detection only; no attribution
This study	**S5P + EMIT**	**Yes**	Pilot-scale; requires larger validation

4.4.2 Comparison with State-of-the-Art Methods

Table 12 benchmarks Gas Alert against leading architectures published in 2024 and 2025. While Ahsan et al. [19] employ transformer-based architectures with stronger attention mechanisms, the lighter ResUNet with Tversky loss achieves comparable segmentation precision with substantially lower inference latency (approximately 30 minutes per scene compared to several hours for attention-based architectures). Unlike Zhang et al. [17], which limits analysis to wind vector plausibility checks, Gas Alert intersects wind vectors with the GPPD infrastructure database to produce a named, probability-ranked list of candidate emission sources.

Table 12: Benchmarking against state-of-the-art architectures

Study	Method	Sensor	Flux	Attribute	Key Distinction
Vaughan et al. [18]	CH4Net CNN	Sentinel-2	No	Manual	Detection only; no attribution
Sherwin et al. [21]	Validation	Multi-sensor	Yes	Wind only	High latency; no automation
Ahsan et al. [19]	AttMetNet Transformer	Hyperspectral	No	None	High compute; no attribution
Zhang et al. [17]	Bias correction	ERA5	Yes	Wind only	No imaging; no segmentation
This study	**ResUNet + IME**	**S5P + EMIT**	**Yes**	**GPPD**	End-to-end prototype; pilot scale

4.5 Quantification Case Study: Jubail Refinery

The IME engine was applied to the confirmed methane anomaly over the Jubail AOI (Chip 0). **Table 13** presents the automated quantification report generated by the system.

The IME engine combines the ResUNet-segmented plume mask with ERA5 bias-corrected wind speed (5.5 m/s from 315° NW) to derive an instantaneous flux estimate of 836.96 kg/hr (±22.4%; 95% CI: 470–1204 kg/hr; conservative ±30% practical envelope: 586–1088 kg/hr). Attribution: likely candidate source is the Jubail Refinery (Unit A) on the basis of wind vector alignment with revised S_w score. Confidence: Medium (wind-aligned screening candidate; not field-confirmed). and the derived parameters are summarized in **Tables 14** and **15**.

Figure 5 shows an active screening alert (orange marker) indicates a methane (CH_4) plume attributed to the likely candidate source Jubail Refinery (Unit A). The popup provides situational awareness to regulators, displaying gas type, pilot flux estimate (842.1 kg/hr), candidate source, and wind direction (315°). The displayed flux of 842.1 kg/hr reflects the real-time

Table 13: Analysis of Chip 0 – Multi-Gas Emission Alert (Methane)

Mode: CH4	
Source Attribution	
Likely Origin	Jubail Refinery (Unit A) (Refinery)
Confidence	Medium (Wind Alignment Check Passed)
Quantification Analysis (CH4)	
Total Burden	61.76 kg
Emission Flux	836.96 kg/hr
Plume Extent	1.46 km
Gas Factor	Using MW = 16.04 g/mol
Environmental Context	
Wind	5.5 m/s from 315° (NW)
Impact	Potentially affecting ~69 people downwind

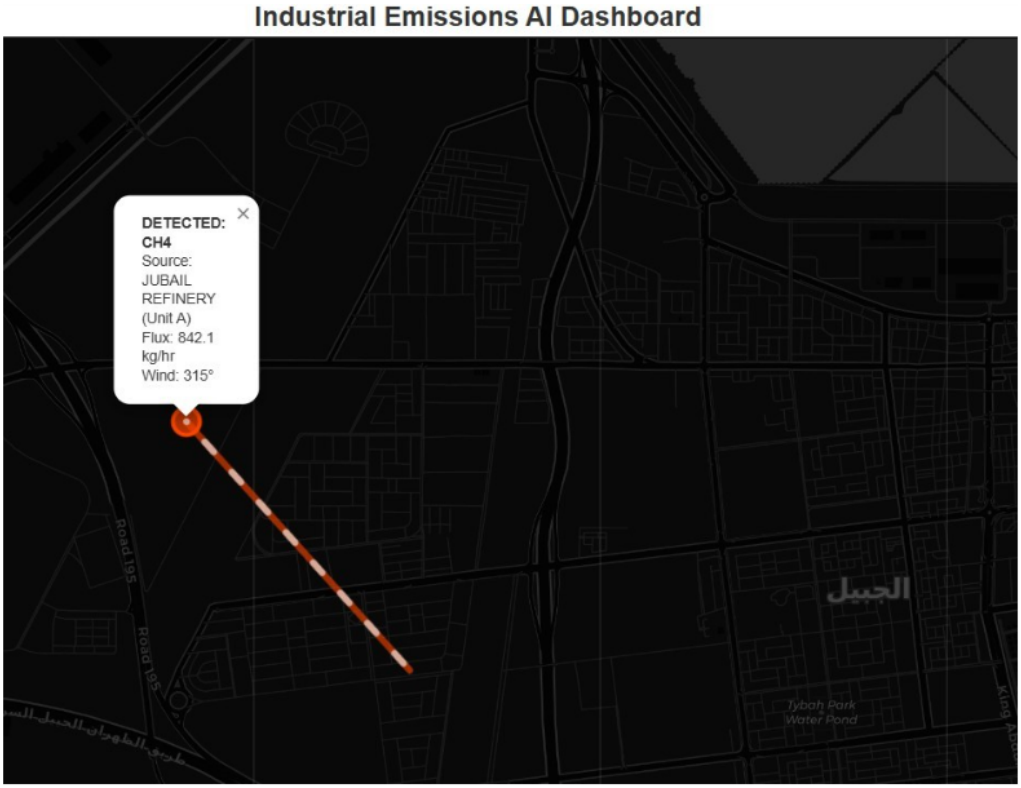


Figure 5: Prototype decision-support dashboard generated by Phase 2 for the Jubail AOI (October 2025 pilot)

dashboard rounding of the pipeline output; the precise pilot estimate is 836.96 kg/hr as reported in**Table 14**. Attribution is reported as ranked candidate-source screening, not deterministic confirmation.

In **Table 15**, the 20% wind and 10% segmentation/mass terms are assumed sensitivity values motivated by prior IME literature [15]; the resulting interval is a first-order uncertainty envelope, not a fully validated uncertainty product.

The estimated flux of 836.96 kg/hr falls within the expected range for medium-to-large fugitive emissions from oil-and-gas refinery infrastructure, typically 500–1500 kg/hr [9, 10]. **Table 16** provides a structured verification against published literature ranges, confirming physical consistency of the pilot estimate.

4.6 Model Calibration and Uncertainty Note

The ResUNet probability maps are raw sigmoid outputs and should not be interpreted as calibrated confidence scores. The maximum raw predicted probability reaching 1.00 in the plume core reflects the sigmoid output saturating at the positive class boundary under extreme imbalance; it does not imply perfect probabilistic calibration. Calibration analysis using reliability curves, threshold sensitivity analysis, bootstrapped confidence intervals, and ensemble or Monte Carlo dropout

Table 14: Automated quantification report for the Jubail anomaly (Chip 0; October 2025 pilot)

Parameter	Derived Value
Target gas	Methane (CH_4)
Likely candidate source	Jubail Refinery (Unit A)
Attribution confidence	Medium (wind-aligned screening candidate)
Plume extent	1.46 km
Instantaneous mass burden	61.76 kg
Estimated flux rate	**836.96 kg/hr**
Flux uncertainty (σ_Q)	±187.13 kg/hr (±22.4%)
95% confidence interval	470.2–1203.7 kg/hr
Conservative ±30% envelope	585.9–1088.1 kg/hr
Wind conditions	5.5 m/s from 315° (NW)
Downwind population at risk	≈69 persons

Table 15: Flux uncertainty propagation for the Jubail pilot case (Chip 0)

Quantity	Symbol	Value	Notes
Flux estimate	Q	836.96 kg/hr	Pilot quantification run
Wind relative uncertainty	σ_v/v	20%	Dominant source (ERA5)
Mask/mass relative uncertainty	σ_M/M	10%	Segmentation quality
Propagated relative uncertainty	σ_Q/Q	22.4%	$\sqrt{0.20^2 + 0.10^2}$
One-sigma uncertainty	σ_Q	187.13 kg/hr	$Q\times0.224$
95% CI	$Q \pm 1.96\sigma_Q$	470–1204 kg/hr	Propagation-based
Conservative ±30% envelope	Q ± 30%	586–1088 kg/hr	Practical reporting
Validation status	—	Pilot case	Requires independent field validation

Table 16: Quantification verification: Jubail pilot detection against published literature ranges for refinery-scale fugitive emissions

Metric	This Study	Literature	Status
Mass burden	61.76 kg	50–200 kg	Consistent
Flux rate	**836.96 kg/hr**	**500–1500 kg/hr**	**Consistent**

Ranges from [9, 10] for large-scale refinery fugitive emissions.

uncertainty quantification is reserved for future work.

4.7 Limitations

Several constraints are acknowledged for the present pilot study.

The 12-scene EMIT pilot sample produces a wide Wilson 95% confidence interval (64.6–98.5%) around the 91.7% scene-level detection rate, meaning that the result demonstrates technical feasibility but does not constitute broad operational validation. A stricter scene-level held-out test set with geographic separation between training and test regions is required for generalization claims.

Flux quantification accuracy is linearly dependent on ERA5 wind speed precision and introduces a propagated uncertainty

of 22.4%, yielding a 95% confidence interval of 470–1204 kg/hr for the Jubail pilot case. No direct operator-reported or in-situ field measurements were available for comparison. Flux estimates are therefore interpreted as pilot workflow demonstrations and physical plausibility checks, not independent validation.

The pipeline is bounded by the EMIT orbital revisit cadence, which can span days to weeks per AOI, limiting Phase 2 to on-demand triggers rather than continuous coverage.

Hyperspectral preprocessing currently requires approximately four hours per 10-scene batch in the Google Colab Pro environment, constraining throughput for large-scale operational deployment.

Source attribution is reported as ranked candidate screening, not deterministic source confirmation. The $S_w > 0.6$ threshold is heuristic and requires sensitivity analysis before use in regulatory decision-making contexts.

5 CONCLUSION

This study demonstrated the technical feasibility of Gas Alert, a hierarchical Tip-and-Cue prototype surveillance framework for industrial emission monitoring in the arid environment of Saudi Arabia. By linking the broad-area screening capabilities of Sentinel-5P TROPOMI with the facility-scale precision of NASA EMIT hyperspectral imagery, the framework addresses a critical gap in environmental monitoring for industrial zones characterized by high-albedo desert backgrounds, thermal turbulence, and multi-species emission profiles.

Four specific contributions are established. First, a hierarchical Tip-and-Cue architecture links Sentinel-5P multi-gas regional screening with EMIT hyperspectral plume analysis in a single automated prototype pipeline. Second, an adaptive multi-gas anomaly detection engine based on 30-day rolling Z-score baselines, quality filtering, AOI padding, and spatial persistence criteria achieved an overall Phase 1 alert-level FPR of 0.81%, an 89% reduction relative to fixed-threshold methods. Third, an automated EMIT label-construction workflow converts NASA EMIT L2B methane enhancement products into binary plume masks through median filtering, enhancement thresholding at > 400 ppm×m, and minimum-cluster filtering. Fourth, a physics-guided segmentation-to-flux workflow connects ResUNet plume masks with IME-based flux estimation, yielding an instantaneous flux of 836.96 kg/hr with a propagated uncertainty of ±22.4% for the Jubail pilot case.

The ResUNet ablation study on a 100-scene Permian Basin held-out test set confirmed that Tversky loss with α=0.7 substantially improves minority-class recall (81.2%) relative to cross-entropy (58.1%) and Dice baselines (69.3%), confirming the value of asymmetric false-negative penalization for extreme plume imbalance. Scene-level pilot validation over 12 Jubail EMIT scenes achieved 91.7% detection rate, though the wide Wilson 95% CI (64.6–98.5%) reflects the pilot scale and underscores the need for larger-scale validation.

Future work will address the identified limitations along four paths. The integration of additional hyperspectral missions, including EnMAP, PRISMA, and Carbon Mapper, will shorten orbital revisit intervals and improve temporal coverage. The replacement of ERA5 reanalysis fields with high resolution local weather station data should reduce wind-speed uncertainty below the current 20% estimate. The quantization of the ResUNet model using TensorRT will enable deployment on onboard or edge processing units for faster inference and higher throughput. Independent validation against controlled-release experiments, operator reports, or ground-based monitoring networks. Gas Alert currently provides a reproducible prototype foundation for active, quantified, and attribution-capable environmental decision-support in direct alignment with the Kingdom of Saudi Arabia's Vision 2030 sustainability goals.

AUTHOR CONTRIBUTION STATEMENT

All authors contributed equally to the study conception and design. Material preparation, data collection, and analysis were performed by the authors. The first draft of the manuscript was written by the authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study did not involve human participants or animals. Therefore, ethical approval and consent to participate are not applicable. If applicable, provide details of the ethics committee approval and consent procedures here.

CONSENT FOR PUBLICATION

Not applicable.

DATA AVAILABILITY

The datasets generated during the current study are available from the corresponding author on reasonable request.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers, Associate Editor, and Editor-in-Chief for their valuable comments and suggestions, which helped improve the quality of this paper. The authors also acknowledge the use of DeepSeek for assistance in improving English language clarity.

FUNDING

No Funding.

DISCLOSURE STATEMENT

The authors declare that they have no competing interests.

REFERENCES

[1] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and F. Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[2] J. P. Veefkind, I. Aben, K. McMullan, H. Förster, J. De Vries, G. Otter, J. Claas, H. Eskes, J. De Haan, Q. Kleipool, *et al.*, "Tropomi on the esa sentinel-5 precursor: A gmes mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications," *Remote sensing of environment*, vol. 120, pp. 70–83, 2012.

[3] R. O. Green, N. Mahowald, C. Ung, D. R. Thompson, L. Bator, M. Bennet, M. Bernas, N. Blackway, C. Bradley, J. Cha, *et al.*, "The earth surface mineral dust source investigation: An earth science imaging spectroscopy mission," in *2020 IEEE aerospace conference*, pp. 1–15, IEEE, 2020.

[4] D. J. Varon, J. McKeever, D. Jervis, J. Maasakkers, S. Pandey, S. Houweling, I. Aben, T. Scarpelli, and D. Jacob, "Satellite discovery of anomalously large methane point sources from oil/gas production," *Geophysical Research Letters*, vol. 46, no. 22, pp. 13507–13516, 2019.

[5] A. K. Thorpe, C. Frankenberg, A. Aubrey, D. Roberts, A. Nottrott, T. Rahn, J. Sauer, M. Dubey, K. Costigan, C. Arata, *et al.*, "Mapping methane concentrations from a controlled release experiment using the next generation airborne visible/infrared imaging spectrometer (aviris-ng)," *Remote Sensing of Environment*, vol. 179, pp. 104–115, 2016.

[6] D. J. Varon, D. Jervis, J. McKeever, I. Spence, D. Gains, and D. J. Jacob, "High-frequency monitoring of anomalous methane point sources with multispectral sentinel-2 satellite observations," *Atmospheric Measurement Techniques*, vol. 14, no. 4, pp. 2771–2785, 2021.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[8] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in *International workshop on machine learning in medical imaging*, pp. 379–387, Springer, 2017.

[9] C. Frankenberg, A. K. Thorpe, D. R. Thompson, G. Hulley, E. A. Kort, N. Vance, J. Borchardt, T. Krings, K. Gerilowski, C. Sweeney, *et al.*, "Airborne methane remote measurements reveal heavy-tail flux distribution in four corners region," *Proceedings of the national academy of sciences*, vol. 113, no. 35, pp. 9734–9739, 2016.

[10] L. Guanter, I. Irakulis-Loitxate, J. Gorroño, E. Sánchez-García, D. H. Cusworth, D. J. Varon, S. Cogliati, and R. Colombo, "Mapping methane point emissions with the prisma spaceborne imaging spectrometer," *Remote Sensing of Environment*, vol. 265, p. 112671, 2021.

[11] J. Van Geffen, H. Eskes, S. Compernolle, G. Pinardi, T. Verhoelst, J.-C. Lambert, M. Sneep, M. Ter Linden, A. Ludewig, K. F. Boersma, *et al.*, "Sentinel-5p tropomi no 2 retrieval: impact of version v2. 2 improvements and comparisons with omi and ground-based data," *Atmospheric Measurement Techniques*, vol. 15, no. 7, pp. 2037–2060, 2022.

[12] D. H. Cusworth, R. M. Duren, A. K. Thorpe, W. Olson-Duvall, J. Heckler, J. W. Chapman, M. L. Eastwood, M. C. Helmlinger, R. O. Green, G. P. Asner, *et al.*, "Intermittency of large methane emitters in the permian basin," *Environmental Science & Technology Letters*, vol. 8, no. 7, pp. 567–573, 2021.

[13] G. Ehret, P. Bousquet, C. Pierangelo, M. Alpers, B. Millet, J. B. Abshire, H. Bovensmann, J. P. Burrows, F. Chevallier, P. Ciais, *et al.*, "Merlin: A french-german space lidar mission dedicated to atmospheric methane," *Remote Sensing*, vol. 9, no. 10, p. 1052, 2017.

[14] D. Thompson, I. Leifer, H. Bovensmann, M. Eastwood, M. Fladeland, C. Frankenberg, K. Gerilowski, R. Green, S. Kratwurst, T. Krings, *et al.*, "Real-time remote detection and measurement for airborne imaging spectroscopy: a case study with methane," *Atmospheric Measurement Techniques*, vol. 8, no. 10, pp. 4383–4397, 2015.

[15] D. J. Varon, D. J. Jacob, J. McKeever, D. Jervis, B. O. Durak, Y. Xia, and Y. Huang, "Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes," *Atmospheric Measurement Techniques*, vol. 11, no. 10, pp. 5673–5686, 2018.

[16] X. Lu, D. J. Jacob, Y. Zhang, J. D. Maasakkers, M. P. Sulprizio, L. Shen, Z. Qu, T. R. Scarpelli, H. Nesser, R. M. Yantosca, *et al.*, "Global methane budget and trend, 2010–2017: complementarity of inverse analyses using in situ (globalviewplus ch 4 obspack) and satellite (gosat) observations," *Atmospheric Chemistry and Physics*, vol. 21, no. 6, pp. 4637–4657, 2021.

[17] L. Zhang, Z. Yan, K. Huang, and W. Zhang, "Evaluation and statistical bias correction of era5-land meteorological variables for a humid river basin in southwest china," *Scientific Reports*, vol. 15, no. 1, p. 41101, 2025.

[18] A. Vaughan, G. Mateo-García, L. Gómez-Chova, V. Růžička, L. Guanter, and I. Irakulis-Loitxate, "Ch4net: a deep learning model for monitoring methane super-emitters with sentinel-2 imagery," *EGUsphere*, vol. 2023, pp. 1–17, 2023.

[19] R. Ahsan, M. Shanto, M. S. Arifin, and T. Hashem, "Attmetnet: Attention-enhanced deep neural network for methane plume detection in sentinel-2 satellite imagery," *arXiv preprint arXiv:2512.02751*, 2025.

[20] E. D. Sherwin, J. S. Rutherford, Y. Chen, S. Aminfard, E. A. Kort, R. B. Jackson, and A. R. Brandt, "Single-blind validation of space-based point-source methane emissions detection and quantification," 2022.

[21] E. D. Sherwin, J. S. Rutherford, Y. Chen, S. Aminfard, E. A. Kort, R. B. Jackson, and A. R. Brandt, "Single-blind validation of space-based point-source detection and quantification of onshore methane emissions," *Scientific Reports*, vol. 13, no. 1, p. 3836, 2023.

[22] A. Radman, M. Mahdianpari, D. J. Varon, and F. Mohammadimanesh, "S2metnet: A novel dataset and deep learning benchmark for methane point source quantification using sentinel-2 satellite imagery," *Remote Sensing of Environment*, vol. 295, p. 113708, 2023.

[23] A. Farahat, "Air pollution in the arabian peninsula (saudi arabia, the united arab emirates, kuwait, qatar, bahrain, and oman): causes, effects, and aerosol categorization," *Arabian Journal of Geosciences*, vol. 9, no. 3, p. 196, 2016.