

NeuPlume: Generative inversion of atmospheric point-source emissions from sparse observations

Lei Wang¹ and Xin Ma²

¹School of Earth and Space Science and Technology, Wuhan University, Wuhan 430079, China

²State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Correspondence: Xin Ma (maxinwhu@whu.edu.cn)

Abstract. Inverting point-source emissions from sparse atmospheric observations is difficult because emission rate, release height, wind speed, turbulence, and plume morphology can compensate for one another. We present NeuPlume, a neural-physical framework that formulates inversion as observation-guided generation and selection of complete three-dimensional concentration fields. A Lagrangian stochastic model builds a scenario-specific forward library, a conditional neural field compresses the simulated fields, and a latent diffusion model learns their conditional distribution. During inversion, diffusion posterior sampling generates candidate fields on a coarse-to-fine parameter grid, and NeuPlume returns the parameters and field that best match the observations. Across 30 synthetic cases, the selected fields achieve a mean spatial correlation of 0.942 and a mean emission-rate error of 13.6%. Matched stress tests show that both observation geometry and transport-model fidelity govern inversion quality: trajectory-concentrated sampling weakens full-field reconstruction, while unrepresented transport alters parameter identification and, for effective plume rise, also degrades field reconstruction. Applied to four UAV methane transects above a coal-mine ventilation shaft, NeuPlume yields known- U minimum-loss emission estimates of 5.27–15.85 kt yr⁻¹ and exposes wind-speed and height-boundary sensitivities under model mismatch. NeuPlume thus provides an extensible methodology for jointly estimating source parameters and concentration fields: additional transport regimes can be incorporated through regime-specific simulation ensembles and neural-model retraining.

1 Introduction

Quantifying greenhouse-gas emissions from individual point sources supports emission inventories, regulatory assessment, and verification of mitigation commitments. Gaussian plume (GP) models estimate emission rate by fitting analytical dispersion kernels to downwind concentration measurements Shah et al. (2019). Mass-balance (MB) methods integrate tracer flux across one or more cross-wind planes Conley et al. (2017); Varon et al. (2018). Both approaches are used with ground, airborne, satellite, and unmanned aerial vehicle (UAV) observations Shaw et al. (2021).

Sparse near-field measurements challenge these methods because source parameters, meteorology, observation geometry, and plume morphology remain coupled. GP methods commonly assume steady conditions, a spatially homogeneous wind field, and Gaussian cross-wind and vertical profiles. Near-field turbulence and plume meandering produce non-Gaussian structures Yuvaraj et al. (2026), while rapidly varying winds violate the steady-state assumption Jia et al. (2025). MB methods avoid

25 prescribing a Gaussian plume shape but remain sensitive to incomplete plume interception, background subtraction, and the
wind speed used to convert concentration into flux. UAV and controlled-release experiments show that observation coverage,
cross-section geometry, and wind inputs substantially affect methane quantification Mohammadloo et al. (2025); Liu et al.
(2024).

Wind-speed representativeness is especially important because source strength and effective advective speed are coupled in
30 steady passive-tracer transport. In an AVIRIS-4 campaign, wind-speed uncertainty contributed up to 99.4% of the variance in
cross-sectional-flux methane inversion Meier et al. (2026). Controlled-release experiments using sequential airborne imagery
found that ERA5 winds underestimated effective plume transport velocity by approximately 28% Eastwood et al. (2025).
Airborne methane detection and coal-mine remote-sensing studies likewise identify wind information as a major source of
emission-quantification error Conrad et al. (2023); He et al. (2024). Jointly searching wind speed with source parameters
35 provides a direct diagnostic of this coupling, although sparse observations may leave several parameter combinations with
similar mismatch.

Bayesian inversion provides one established treatment of non-uniqueness by combining a parameter prior and an observa-
tional likelihood. Markov chain Monte Carlo Xu et al. (2025), variational data assimilation Aghdasi et al. (2025), regional-
network inversion Whites et al. (2019), and ensemble source reconstruction De Meutter et al. (2021) have all been applied to
40 atmospheric source estimation; recent reviews compare these methods with optimization-based alternatives Xie et al. (2026).
Their computational cost grows with parameter dimension and state complexity, particularly when source terms, meteorologi-
cal parameters, and spatial concentration fields are inferred jointly.

Deep generative models offer a complementary route by learning low-dimensional representations from physical simulations
and using measurements to guide candidate generation. Conditional diffusion models have been used for inverse reconstruction
45 of permeability and seismic wavefields Liu et al. (2025b); Meng et al. (2024). Conditional neural field (CNF) latent diffusion
frameworks combine field compression with diffusion posterior sampling (DPS) to reconstruct turbulent fields from incomplete
observations Du et al. (2024); Liu et al. (2025a). In atmospheric point-source applications, learned plume representations can
preserve non-Gaussian structure that analytical plume models omit. Zhao et al. (2026), for example, combined a U-Net encoder
with inverse Gaussian fitting to reduce UAV emission-rate error.

50 This paper presents NeuPlume, a three-stage generative inversion framework for atmospheric point-source emissions. First,
a Lagrangian stochastic model generates a library of concentration fields indexed by source and meteorological parameters.
Second, a CNF compresses each field into a latent representation, and a denoising diffusion probabilistic model learns the
conditional distribution of those representations. Third, NeuPlume runs observation-guided DPS on a coarse-to-fine parameter
grid, estimates the emission rate at each node, and returns the candidate with the lowest observation mismatch. The main
55 contributions are:

1. We formulate near-field source estimation as generative candidate-field inversion, jointly returning a minimum-loss
parameter vector and its complete three-dimensional concentration field.

2. We define a simulator-independent implementation path in which a forward model maps scenario parameters to the concentration-field library used to train a new CNF and diffusion model.
- 60 3. We evaluate minimum-loss parameter and field accuracy on 30 held-out synthetic cases and use six representative off-grid cases within the trained parameter support for controlled baseline and component comparisons.
4. We apply the framework to public UAV methane observations above a coal-mine ventilation shaft and use matched stress tests to quantify sensitivity to observation geometry and unrepresented transport mechanisms.

2 Methods

65 2.1 NeuPlume model architecture

NeuPlume learns plume morphologies from forward simulations and uses sparse measurements to guide the generation and selection of candidate concentration fields.

The framework has three stages (Fig. 1). Stage 1 uses a Lagrangian stochastic model (LSM) to generate concentration fields at a unit emission rate $Q = 1 \text{ kg s}^{-1}$. Each field is indexed by $\theta = (H, U, I)$, where H is effective release height, U reference
70 wind speed, and I turbulence intensity. Stage 2 uses Hash-INR to compress each field into a 192-dimensional latent code $\mathbf{z}$, and a denoising diffusion probabilistic model (DDPM) learns $p(\mathbf{z} | \theta)$. In Stage 3, sparse observations guide diffusion posterior sampling (DPS) at each candidate θ_k . NeuPlume decodes a candidate unit-emission field, fits Q by linear scaling, evaluates the observation mismatch, refines the search around the lowest-loss coarse-grid node, and returns $(\hat{\theta}, \hat{Q}, \hat{C})$ from the minimum-loss fine-grid candidate.

75 Let $C(\mathbf{x} | \theta)$ denote a unit-emission concentration field and let $\mathcal{H}$ map the complete field to the measurement locations. The observations satisfy

$$\mathbf{y} = Q \mathcal{H}[C(\mathbf{x} | \theta)] + \epsilon, \quad (1)$$

where ϵ collects measurement and representation residuals, and linear scaling by Q follows from unit-rate normalization. NeuPlume evaluates a finite candidate set produced by the conditional generative model and coarse-to-fine grid search. Its
80 formal output is

$$(\hat{\theta}, \hat{Q}, \hat{C}) = \arg \min_{(\theta_k, Q_k, C_k) \in \mathcal{G}} \mathcal{L}_{\text{obs}}(\theta_k, Q_k, C_k), \quad (2)$$

where $\mathcal{G}$ is the generated candidate set.

2.2 Lagrangian stochastic forward simulation

The forward simulator uses a Lagrangian stochastic model (LSM) to generate concentration fields from source and meteo-
85 rological parameters. LSMs represent turbulent dispersion with stochastic particle trajectories and must satisfy a stationary

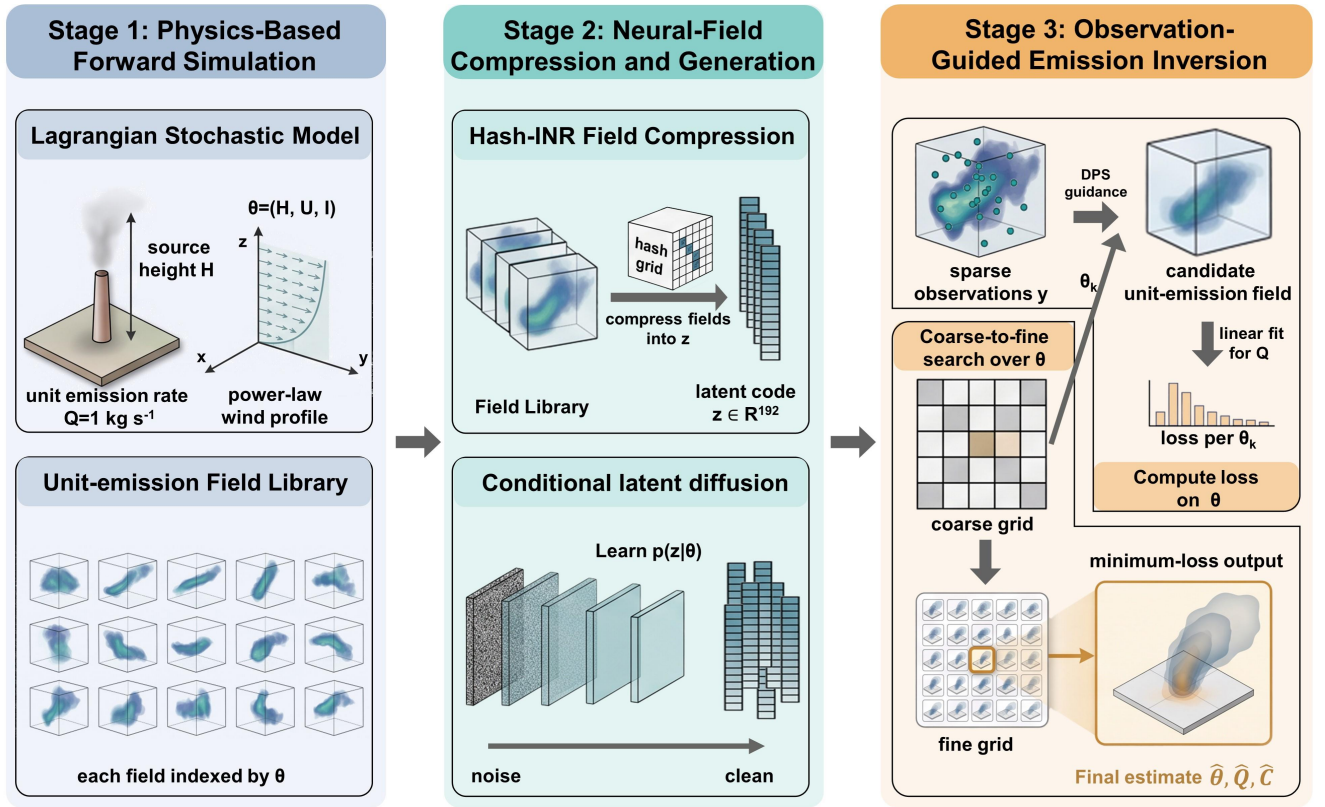


Figure 1. Overview of the three-stage NeuPlume framework. Stage 1 constructs a unit-emission field library with a Lagrangian stochastic model. Stage 2 compresses the fields with Hash-INR and learns their conditional latent distribution. Stage 3 combines observation-guided diffusion posterior sampling with coarse-to-fine parameter search and emission-rate fitting to obtain the minimum-loss reconstruction.

phase-space distribution, Eulerian-equation compatibility, and the well-mixed condition Thomson (1987). Established codes such as FLEXPART and STILT apply Lagrangian particle models to point-source dispersion, near-field source-receptor relationships, and flux inversion Stohl et al. (2005); Lin et al. (2003). Microscale comparisons of forward-in-time and backward-in-time LSMs support their use for local-scale dispersion and source-receptor sensitivity problems Li and Du (2020). Relative to the Gaussian plume formula, an LSM represents near-field dispersion and moderate-turbulence non-Gaussian structures at a computational cost several orders of magnitude below large-eddy simulation Katata et al. (2015).

The model continuously releases particles from a point source at height H above the ground. Each particle carries position x and turbulent velocity fluctuation u' , which evolves by a Langevin-type stochastic differential equation:

$$du'_i = -\frac{u'_i}{T_L} dt + \sqrt{\frac{2\sigma_i^2}{T_L}} dW_i, \quad i \in \{x, y, z\}, \quad (3)$$

where T_L is the Lagrangian integral time scale, σ_i is the directional velocity standard deviation, and dW_i denotes an independent Wiener increment. The mean wind field follows the power-law profile $\bar{U}(z) = U_{\text{ref}}(z/z_{\text{ref}})^\alpha$. We set $\alpha = 1/7 \approx 0.143$ for

neutral atmospheric stability and define U_{ref} at $z_{\text{ref}} = 10$ m. An elastic ground boundary reverses the vertical velocity component, and particles leaving the domain are removed. After a flushing period for statistical steadiness, we bin particle positions onto a regular three-dimensional grid and divide by grid-cell volume to obtain the concentration field $c(\mathbf{x})$.

100 Because the steady-state concentration field of an inert tracer is linear in source strength, all training simulations set a unit emission rate $Q = 1 \text{ kg s}^{-1}$. Thus, the LSM, CNF, and diffusion model learn only the unit-rate concentration distribution determined by $(H, U_{\text{ref}}, I_{\text{ref}})$. During inversion, a candidate emission rate Q linearly scales this unit field to give the corresponding concentration prediction.

With the unit-rate convention above, we constructed the training dataset required by the conditional neural field (Section 2.3).
 105 The training set covers the three-dimensional conditional parameter space spanned by effective release height $H \in [0, 50]$ m, reference wind speed $U_{\text{ref}} \in [1, 15] \text{ m s}^{-1}$, and turbulence intensity $I_{\text{ref}} \in [0.05, 0.3]$. Here, H is the effective release height represented in the training library; it is not necessarily identical to the physical outlet height when initial momentum or buoyancy produces plume rise. These ranges cover near-surface passive point-source monitoring conditions, from calm and weak-turbulence scenarios to strong-wind and high-turbulence cases. Atmospheric stability is neutral in all simulations.

110 The current LSM instance represents a passive inert tracer without initial source momentum, explicit buoyant plume rise, wind-direction shear, convergence, or divergence. It uses flat terrain, neutral stability, and an open single-source domain. Particles that cross the lateral, top, upwind, or downwind boundaries are removed, whereas the ground boundary is reflective.

We uniformly sample 1 000 parameter combinations from this space. For each combination, the LSM is integrated over a $520 \times 200 \times 100 \text{ m}^3$ computational domain ($x \in [-20, 500]$, $y \in [-100, 100]$, $z \in [0, 100]$ m), with time step $\Delta t = 0.5$ s, 250
 115 particles emitted per step, and approximately 10^5 active particles at steady state. The resulting 1 000 concentration fields are paired with their conditional parameter triplets $(H, U_{\text{ref}}, I_{\text{ref}})$ to form the supervised training set for neural field compression.

2.3 Conditional neural field training

Stage 2 compresses the discrete LSM concentration fields into a continuous, conditionable representation. For each $\theta = (H, U_{\text{ref}}, I_{\text{ref}})$, an implicit function $f_{\psi}(\mathbf{x}; \theta)$ maps $\mathbf{x} \in \mathbb{R}^3$ to concentration. We implement this conditional neural field with
 120 Hash-INR; conditional neural fields have also been used for continuous representations of turbulence data Guo et al. (2026).

Hash-INR combines a 16-level multiresolution hash encoder with a condition-concatenated multilayer perceptron. Each hash level contains two features per entry, with resolutions spanning 16^3 to 512^3 voxels. Trilinear interpolation and concatenation yield 32 spatial features. The standardized condition vector is concatenated with these features and a sample-specific 192-dimensional latent code $\mathbf{z}$. A four-hidden-layer network with 256 ReLU units per layer predicts $\log(1 + C)$.

125 Hash-INR is optimized jointly over all training conditions for 300 epochs. Network weights and latent codes use Adam learning rates of 10^{-3} and 5×10^{-4} , respectively. The objective combines log-space mean squared error, a linear-space penalty above the 90th concentration percentile, and ℓ_2 regularization of the latent codes. Gradients are clipped to 1, and the learning rate is halved after a validation plateau.

The complete training and inversion settings are listed in Appendix A (Table A1).

130 After Hash-INR training, each condition has a latent code that encodes its concentration field with the shared decoder weights. A DDPM learns $p(\mathbf{z} | \boldsymbol{\theta})$ over these codes. Inspired by CoNFILd Du et al. (2024), the denoiser uses a DiT architecture with six Transformer blocks, eight-head self-attention, cross-attention, and a feed-forward expansion ratio of 4. The 192 latent dimensions are projected into a 512-dimensional hidden space. Diffusion timestep and condition embeddings form the context. Training uses a 1000-step linear noise schedule, $\beta \in [10^{-4}, 2 \times 10^{-2}]$, 100 epochs, batch size 64, Adam with a 10^{-4} learning
 135 rate, and cosine annealing.

Classifier-free guidance is trained by zeroing the condition vector with probability p_{drop} . During generation, conditional and unconditional predictions are combined as

$$\hat{\epsilon} = \epsilon_{\emptyset} + s(\epsilon_{\boldsymbol{\theta}} - \epsilon_{\emptyset}), \quad (4)$$

where s is the guidance scale. The trained DDPM generates latent candidates at a specified $\boldsymbol{\theta}$, and Hash-INR decodes them
 140 into complete concentration fields for the inversion procedure in Section 2.4.

2.4 Point-source emission inversion

NeuPlume searches a finite source and meteorological parameter grid. For each candidate $\boldsymbol{\theta}$, the conditional diffusion model generates a concentration field in the Hash-INR latent space. The algorithm estimates the emission rate Q from the linear relation between the predicted unit field and the observations, evaluates the observation mismatch, and returns the lowest-loss
 145 candidate field and parameter vector.

The generative backbone follows latent-space diffusion for conditional neural fields Du et al. (2024); Liu et al. (2025a). Starting from Gaussian noise $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the reverse process uses the noise-prediction network $\epsilon_{\phi}(\mathbf{z}_t, t, \boldsymbol{\theta})$. Classifier-free guidance combines conditional and unconditional predictions:

$$\hat{\epsilon}_t = \epsilon_{\phi}(\mathbf{z}_t, t, \mathbf{0}) + s[\epsilon_{\phi}(\mathbf{z}_t, t, \boldsymbol{\theta}) - \epsilon_{\phi}(\mathbf{z}_t, t, \mathbf{0})], \quad (5)$$

150 where s is the guidance scale. The standard DDPM relation gives the predicted clean latent vector, which the INR decodes into a concentration field.

NeuPlume uses diffusion posterior sampling (DPS) Chung et al. (2023), which introduces the gradient of the observation-mismatch objective into the reverse process:

$$\epsilon_t^* = \hat{\epsilon}_t - \zeta \nabla_{\mathbf{z}_t} \mathcal{L}_{\text{obs}} \sqrt{1 - \bar{\alpha}_t}, \quad (6)$$

155 where ζ controls observation guidance. DPS supplies observation-guided candidate fields for grid-node evaluation, and the minimum-loss rule in Eq. (2) produces a point estimate.

For a generated unit field $\mathbf{c}(\boldsymbol{\theta})$ sampled at M observation locations, the predicted observations are $\hat{\mathbf{y}} = Q \mathcal{H}[\mathbf{c}(\boldsymbol{\theta})]$. Candidate quality is measured with a smoothed L_1 observation-mismatch objective:

$$\mathcal{L}_{\text{obs}}(\boldsymbol{\theta}, Q) = \frac{1}{M} \sum_{j=1}^M h(\hat{y}_j - y_j^{\text{obs}}), \quad h(e) = \begin{cases} \frac{1}{2}e^2 & |e| \leq 1, \\ |e| - \frac{1}{2} & \text{otherwise.} \end{cases} \quad (7)$$

160 The Huber form limits the influence of large residuals. The field of a passive tracer scales linearly with source strength, so Q is initialized by the least-squares scaling between the generated unit field and the observations and may be refined during the reverse process.

The coarse stage evaluates a broad grid of H , U_{ref} , and I_{ref} . At each node, NeuPlume generates a candidate field, fits Q , and evaluates the observation mismatch. The node with the smallest loss defines the center of the fine grid. The fine stage evaluates
165 local offsets and returns the lowest-loss fine-grid candidate with its fitted Q and decoded concentration field.

NeuPlume supports two inversion modes. In known- U mode, meteorological measurements fix the reference wind speed and the search covers (H, I_{ref}) . In blind mode, the search covers $(H, U_{\text{ref}}, I_{\text{ref}})$. Blind inversion exposes the H - U - Q degeneracy because different parameter combinations can yield similar downwind concentrations. The observation-mismatch surface and the distance between competing low-loss candidates describe this ambiguity.

170 3 Experimental setup

The primary synthetic validation uses 30 concentration fields generated with the same neutral-stability LSM procedure as the training library and held out from conditional neural field and diffusion training. The validation conditions cover $H = 10$ – 40 m, $U_{\text{ref}} = 3$ – 12 m s^{-1} , $I_{\text{ref}} = 0.08$ – 0.27 , and five emission rates from 0.75 to 1.25 kg s^{-1} . Each time-averaged field contains 10^5 evaluation points in the $520 \times 200 \times 100$ m^3 domain, and 1% of these points (1 000 observations) are selected by a case-specific
175 random volumetric mask.

Six representative validation cases span low and high release heights, wind speeds, and turbulence intensities within the validation design; their emission rates range from 0.75 to 1.25 kg s^{-1} (Table B2). The baseline and component comparisons use the same 1% random-volume mask for each case. This geometry differs from plume-crossing transects and defines the interpretation of the GP and MB comparisons.

180 The UAV trajectories and methane measurements used in the observation-geometry stress test and field experiment come from the publicly available AirCore-UAV data for flights #11, #12, #14, and #15 above shaft V of the Pniówek coal mine in Poland Andersen et al. (2023). For the observation-geometry stress test, each simple-transport truth field is sampled with four equal-budget protocols. The random-volume protocol selects 1 000 of the 10^5 evaluation points. The complete, partial, and off-center protocols transform trajectories from flight #11 or #12, #14, and #15 into the synthetic domain and select the 1 000
185 evaluation points nearest each trajectory. The partial protocol retains one crosswind portion of its source trajectory; the off-center protocol applies a 60 m crosswind shift and domain truncation. These compound protocols test three realistic trajectory patterns rather than individual geometric factors.

The transport-mismatch stress test uses the same six off-grid parameter settings, 1% random-volume observation masks, and inversion protocol. New truth fields are generated with the simple transport model, a strong-shear model with power-law
190 exponent $\alpha = 0.30$, and an effective-plume-rise model with prescribed vertical mean velocity $w(x) = 0.6 \exp(-x/80)$ m s^{-1} for $x \geq 0$. All fields are inverted with the CNF and diffusion models trained on the simple transport fields. The parameter

settings remain within the trained support, whereas the strong-shear and plume-rise fields contain transport mechanisms absent from the training library.

The field experiment applies NeuPlume to the methane measurements from these four flights. Ground-station wind speed was $6\text{--}7\text{ ms}^{-1}$. Preprocessing reverses the analyzer sequence, aligns it with GPS time, and assigns each methane observation to the corresponding flight location. From the same study, descriptive external references are the same-shaft hourly inventory ($13.3 \pm 4.1\text{ kt yr}^{-1}$) and the published campaign-level integrated Gaussian (IG, $9.5 \pm 3.5\text{ kt yr}^{-1}$, $n = 13$) and published campaign-level mass-balance (MB, $9.8 \pm 5.0\text{ kt yr}^{-1}$, $n = 14$) estimates. These published campaign-level values are distinct from our flight-level MB baseline described below.

NeuPlume is compared with GP and our flight-level MB baseline using the same observations and meteorological inputs. GP minimizes residual sum of squares with differential evolution followed by L-BFGS-B. It is evaluated in blind and known- U modes. MB interpolates observations onto a cross-wind section and integrates concentration flux using the specified wind speed, so it is evaluated only in known- U mode. NeuPlume also uses blind and known- U modes; the latter fixes wind speed during candidate generation and grid search.

For synthetic cases, parameter accuracy is evaluated using the absolute relative error of the selected Q , H , U , and I . Full-field reconstruction is evaluated using normalized root-mean-square error (RMSE divided by the root-mean-square truth concentration), Pearson spatial correlation, concentration-weighted centerline displacement, and horizontal and vertical width bias. The same metrics are also computed after excluding the 1% observation mask. Stress-test degradation is defined as the tested value minus the matched baseline for parameter errors and normalized RMSE, and as the baseline minus the tested value for spatial correlation; positive values therefore denote worse performance for every displayed metric. For field cases, ground-truth parameter errors are unavailable. We report the minimum-loss NeuPlume estimates and their descriptive differences from the inventory center and range and from the published IG and MB references.

4 Results

4.1 Synthetic data validation

Across the 30 synthetic validation cases, the mean absolute relative errors were 2.39% for effective release height H , 23.13% for wind speed U , 5.74% for turbulence intensity I , and 13.59% for emission rate Q (Table B1). The corresponding medians were 0.75%, 16.96%, 5.41%, and 10.72%, respectively.

The selected three-dimensional fields had a mean spatial correlation of 0.942 and a mean normalized RMSE of 0.345 against the LSM truth. Their concentration-weighted centerlines were displaced by 0.78 m on average. The reconstructed crosswind and vertical widths were narrower than the truth by 1.20 and 0.76 m on average. Metrics computed only at the 99% unobserved points were nearly identical: the mean spatial correlation was 0.942, normalized RMSE was 0.345, and centerline displacement was 0.78 m. Figure 2 relates each estimate directly to its truth value and shows the joint full-field agreement and width biases.

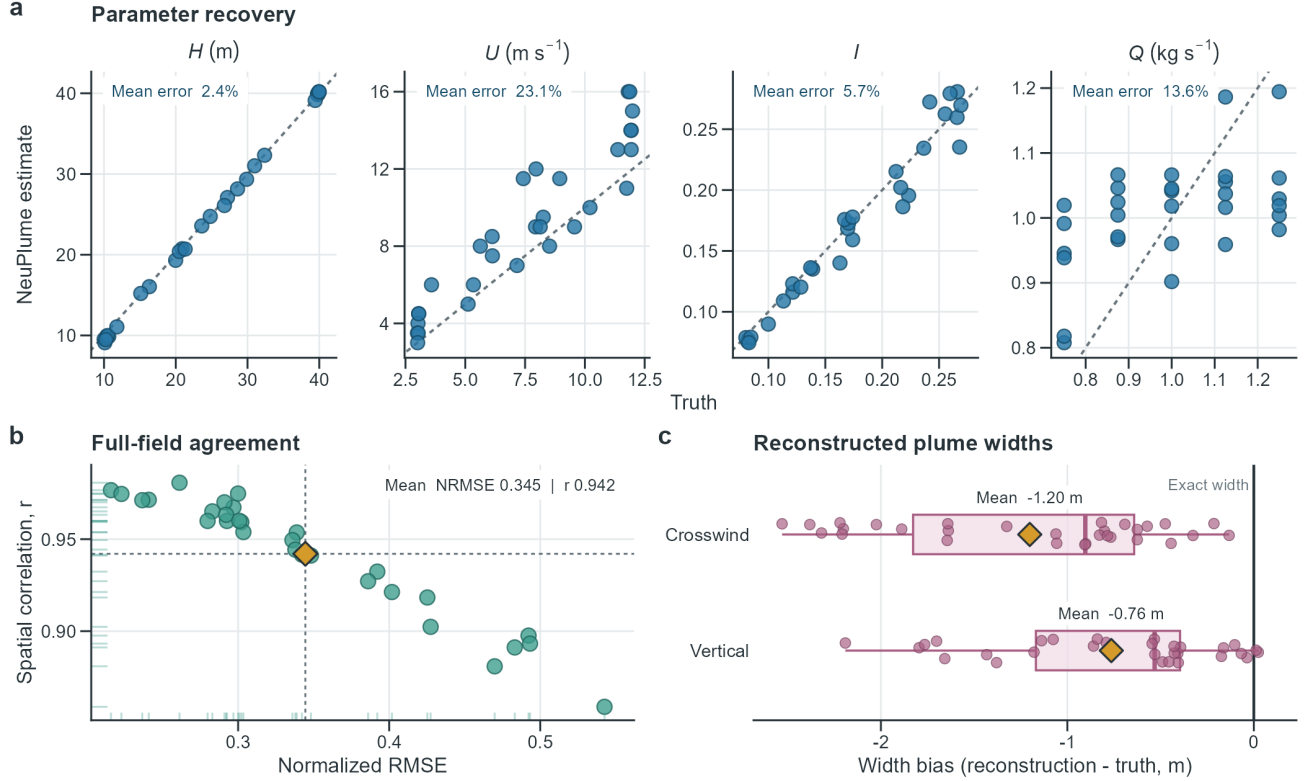


Figure 2. Synthetic validation across 30 held-out cases. (a) Each point is one synthetic case. Estimated values are plotted against the true H , U , I , and Q ; dashed diagonal lines denote exact recovery and inset labels give mean absolute relative errors. (b) Joint full-field normalized RMSE and spatial correlation, for which the upper-left direction indicates lower error and stronger spatial agreement; dashed crosshairs and the diamond mark the two sample means. (c) Crosswind and vertical width biases; negative values indicate a narrower reconstruction. Points are cases, boxes show medians and interquartile ranges, and diamonds mark means. Quantitative field metrics use all 100 000 evaluation points.

Figure 3 provides the corresponding three-dimensional comparison for all six preselected representative cases. Across these cases, spatial correlation ranges from 0.918 to 0.981 and centerline displacement ranges from 0.18 to 1.68 m. The complete 225 30-case summary and the physical conditions of the six cases are reported in Appendix B.

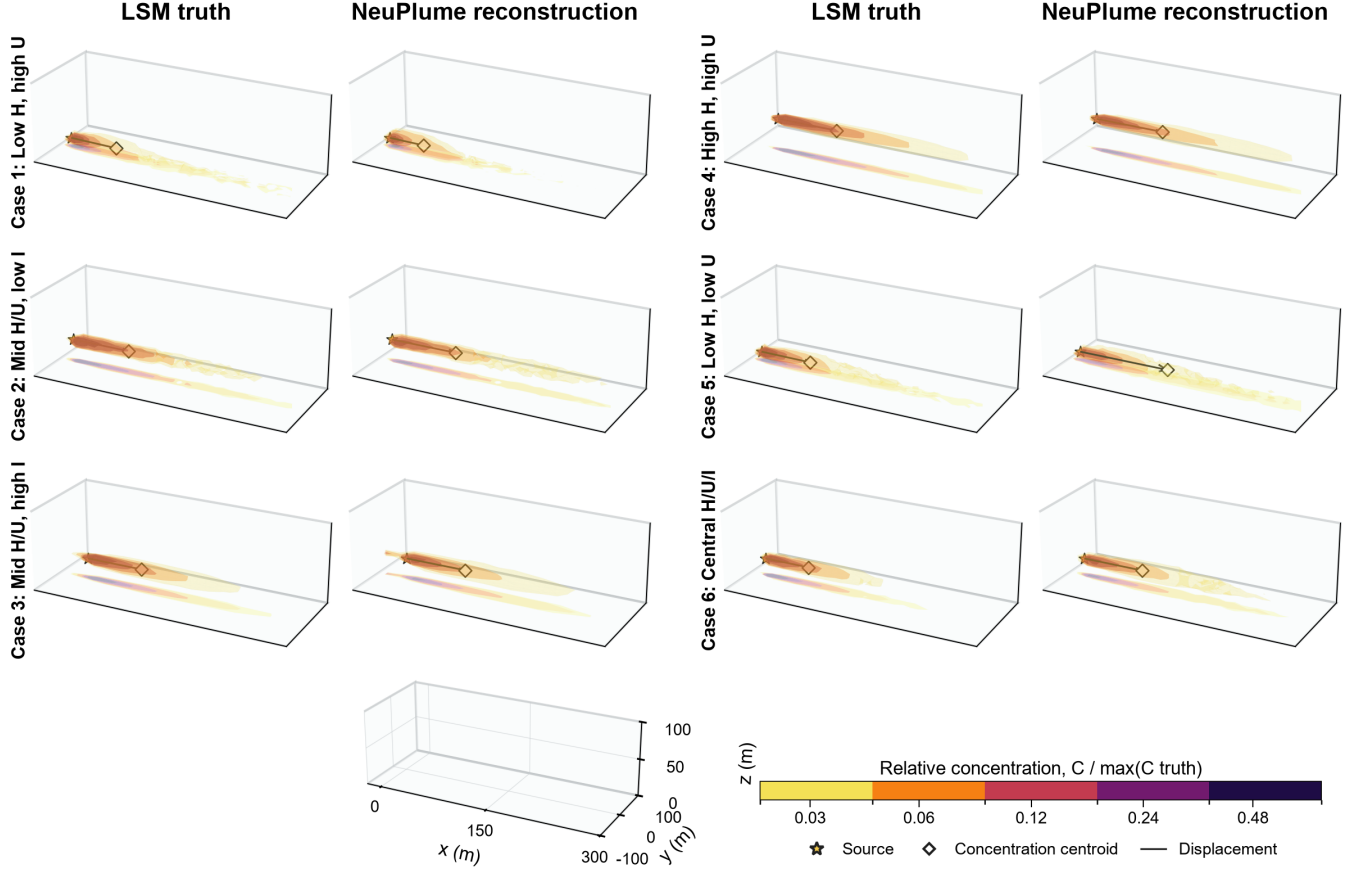


Figure 3. Three-dimensional plume comparisons for the six representative cases. The first two columns compare the LSM truth and NeuPlume reconstruction for Cases 1–3: Case 1 (low H , high U), Case 2 (mid H/U , low I), and Case 3 (mid H/U , high I). The last two columns show the same comparison for Cases 4–6: Case 4 (high H , high U), Case 5 (low H , low U), and Case 6 (central $H/U/I$). Nested isosurfaces show the three-dimensional concentration field, with maximum-concentration projections on the ground and rear planes. Color and opacity denote $C / \max(C_{\text{truth}})$ within each case; values below 0.02 are omitted for visibility. Stars mark the source location at the case-specific effective release height, diamonds mark the concentration-weighted centroid of the displayed volume bins, and lines show their displacement. The display uses $52 \times 30 \times 20$ bins reconstructed from all 100,000 evaluation points.

4.1.1 Six-case benchmark results

The six representative cases support the GP and MB comparisons, Gaussian-plume control, and coarse-to-fine analysis (Table 1). Across these cases, NeuPlume’s mean absolute relative errors are 12.5% for Q , 2.5% for H , 25.6% for U , and 7.3% for I .

230 The fine stage selects a lower minimum loss in every case. Across the six cases, mean minimum loss decreases from 0.172 to 0.142; the case-wise decrease ranges from 0.001 to 0.080. Appendix Fig. B1 reports the case-wise changes and the coarse and local fine parameter spaces for the largest- and smallest-improvement cases. Table 1 reports the corresponding parameter accuracy.

Table 1. Inversion results for the six representative cases. Values are absolute relative errors (%) against true values. NeuPlume, GP, and MB use the same 1 000 random-volume observations.

Case	NeuPlume				GP blind		GP known U		MB known U
	Q	H	U	I	Q	U	Q	H	Q
Case 1	9.8	4.9	14.3	12.6	62.6	54.4	51.9	0.4	146.4
Case 2	17.0	0.4	54.9	1.7	43.4	54.8	143.1	35.4	529.6
Case 3	6.2	0.2	15.0	7.6	81.3	88.5	30.8	7.5	299.2
Case 4	15.1	0.4	8.9	5.3	198.7	16.6	87.0	14.2	207.3
Case 5	25.1	6.0	47.3	14.0	559.4	440.9	234.9	49.3	450.1
Case 6	1.8	2.8	13.4	2.7	5861.9	48.2	278.3	1.6	343.3
Mean	12.5	2.5	25.6	7.3	1134.5	117.2	137.6	18.1	329.3
Median	12.4	1.6	14.7	6.4	140.0	54.6	115.1	10.8	321.3

4.1.2 Baseline comparison

235 NeuPlume, GP regression Shah et al. (2019), and the MB method are applied to the same observations from the six representative cases. Under this random volumetric geometry, mean Q error is 1 134.5% for blind GP, 137.6% for known- U GP, 329.3% for known- U MB, and 12.5% for NeuPlume; the corresponding medians are 140.0%, 115.1%, 321.3%, and 12.4% (Table 1). The comparison tests how methods designed for analytical plume fitting or cross-sectional flux integration transfer to sparse volumetric observations.

240 4.1.3 Gaussian plume control experiment

The Gaussian-plume control uses the same source parameters as the six representative cases and generates observations with an analytical Gaussian-plume model (Table C1). In known- U mode, GP recovers all six cases with Q and H errors below 0.001%. In blind mode, its mean Q and U errors are both 126.3%, consistent with the Q - U ambiguity when both parameters are inferred from the Gaussian-plume equation. The known- U result confirms that the GP implementation recovers parameters
245 when its field assumption matches the data generator.

4.1.4 Observation geometry and transport-model mismatch

The geometry and transport stress tests use the six representative conditions within the trained parameter support. Figure F1 shows the case-wise transport-mismatch changes, Fig. G1 shows the transformed UAV sampling geometries, and Fig. G2 shows their case-wise degradation relative to random-volume sampling.

250 With 1 000 random-volume observations per case, the mean errors in H , U , I , and Q are 2.46%, 25.63%, 7.31%, and 12.51%, respectively; mean full-field normalized RMSE is 0.322 and spatial correlation is 0.957. The complete, partial, and off-center trajectory protocols use the same point budget. Their mean normalized RMSE values are 0.744, 0.565, and 0.611, while mean spatial correlations are 0.622, 0.803, and 0.817. Mean I error exceeds 23% for all three trajectory protocols. The ordering varies by metric: the complete trajectory has the largest mean H error and normalized RMSE, the partial trajectory
255 has the largest mean U and Q errors, and the off-center trajectory has the largest mean I error.

Under strong shear, mean H and U errors increase from 2.46% and 25.63% to 3.44% and 27.41%, whereas mean I and Q errors decrease slightly to 7.20% and 9.40%. Mean full-field normalized RMSE changes from 0.322 to 0.308 and spatial correlation from 0.957 to 0.953. Effective plume rise produces a different response: mean H and U errors increase to 22.72% and 35.12%, mean I and Q errors are 8.78% and 11.83%, normalized RMSE increases to 0.518, and spatial correlation
260 decreases to 0.821.

4.2 Matched component comparison

We compare full NeuPlume with two references using the same six representative cases, 1 000 random-volume observations per case, and loss definition (Fig. 4). The CNF library search removes diffusion and evaluates learned latent fields directly on the same coarse-to-fine parameter grid. Relative to this reference, full NeuPlume reduces the mean errors in H , U , I , and
265 Q from 3.21%, 49.36%, 22.10%, and 18.54% to 2.46%, 25.63%, 7.31%, and 12.51%, respectively. Mean normalized RMSE decreases from 0.438 to 0.322, and spatial correlation increases from 0.910 to 0.957. These matched results show that diffusion improves parameter selection and full-field reconstruction relative to direct CNF library search under the tested protocol.

Direct LSM search evaluates the 1 000-member physical library against the off-grid representative truth fields and provides a finite-library lookup reference. It attains a mean normalized RMSE of 0.178 and a mean spatial correlation of 0.981. Its
270 remaining parameter errors (H : 1.62%, U : 17.79%, I : 11.41%, Q : 22.91%) quantify discretization and parameter ambiguity in this same-simulator lookup.

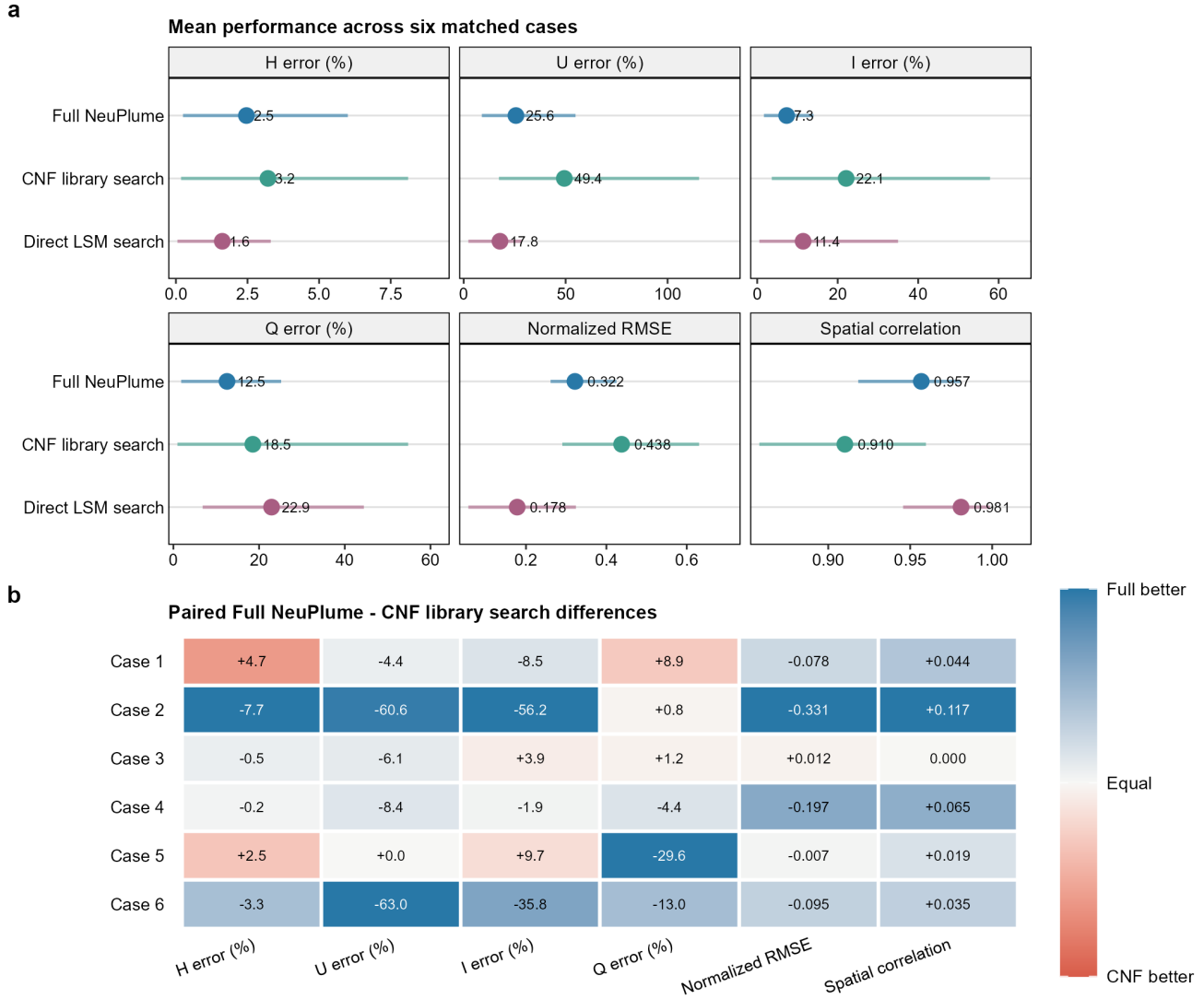


Figure 4. Matched accuracy comparison for the six representative cases. (a) Mean absolute relative errors for effective release height H , wind speed U , turbulence intensity I , and emission rate Q , together with mean full-field normalized RMSE and spatial correlation; horizontal intervals span the minimum to maximum across the six cases. (b) Case-wise paired differences between Full NeuPlume and CNF library search. Cell labels report the raw Full NeuPlume minus CNF library search difference in the units of each metric. Colour direction is oriented so that blue denotes better Full NeuPlume performance and orange denotes better CNF library-search performance, with intensity scaled separately within each metric. Lower values indicate better performance for the four parameter errors and normalized RMSE; higher values indicate better spatial correlation. Direct LSM search is a finite-library lookup reference whose 1 000 candidates do not contain the six truth conditions.

4.3 Real-data case study

We evaluate the Pniówek V flights in known- U and blind modes. The same-shaft hourly inventory (13.3 ± 4.1 kt yr $^{-1}$) provides the primary descriptive reference. Campaign-level IG (9.5 ± 3.5 kt yr $^{-1}$) and MB (9.8 ± 5.0 kt yr $^{-1}$) estimates provide additional references Andersen et al. (2023). Table 2 and Fig. 5 report NeuPlume’s selected minimum-loss estimates.

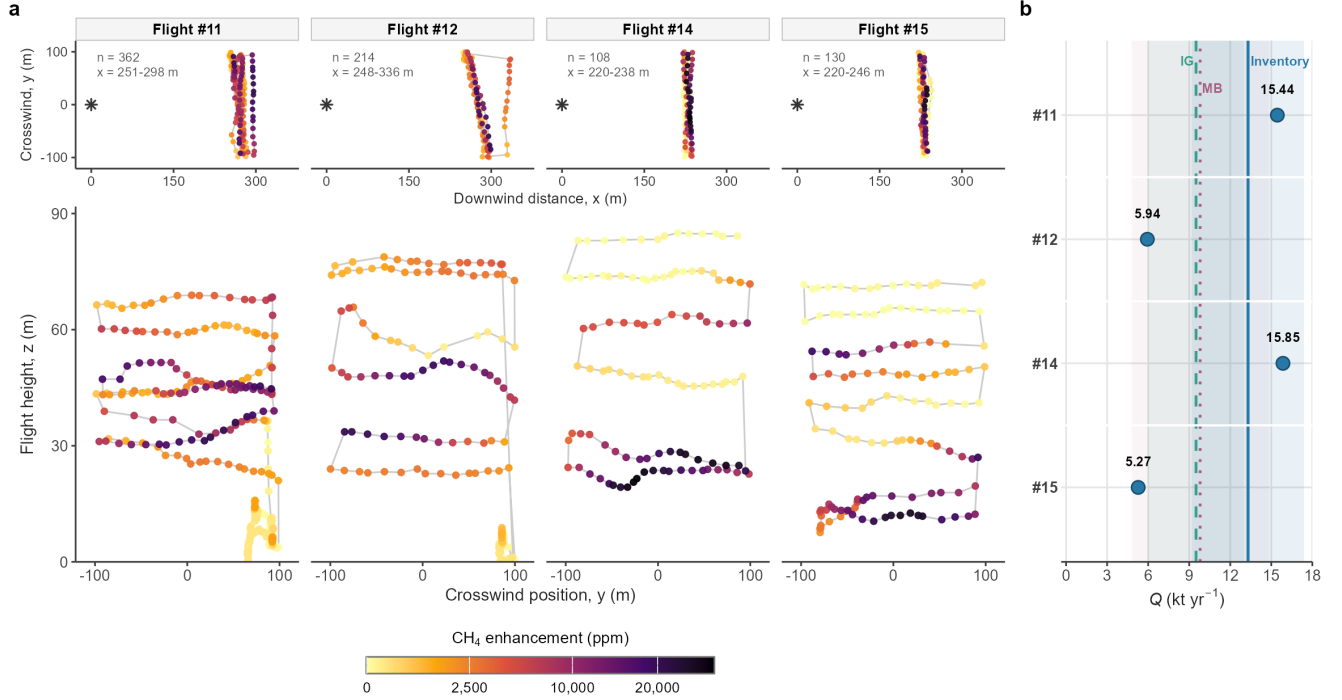


Figure 5. Real-data inversion for the Pniówek V ventilation shaft. (a) UAV sampling geometry for each flight. The upper strips locate the observations in the downwind–crosswind plane; stars mark the shaft. The lower panels show the crosswind–height sampling curtains, with grey lines tracing the flight paths and color denoting methane enhancement. Labels report the number of observations and sampled downwind range. (b) NeuPlume known- U minimum-loss Q estimates (points). The same-shaft inventory and published IG and MB values retain their reported external ranges.

Known- U estimates for flights #11 and #14 are 2.14 and 2.55 kt yr $^{-1}$ above the inventory center and fall within its reported 9.2–17.4 kt yr $^{-1}$ range. Flights #12 and #15 are 7.36 and 8.03 kt yr $^{-1}$ below the center and lie 3.26 and 3.93 kt yr $^{-1}$ below the lower end of that range.

Blind inversion selects lower wind speeds for flights #11, #12, and #14 and correspondingly lower Q values. Flight #15 selects $U = 14.0$ m s $^{-1}$, more than twice the station value, and $Q = 13.26$ kt yr $^{-1}$. Across all four flights, the selected U and Q move in the same direction relative to the known- U solution.

Table 2. Minimum-loss NeuPlume estimates for the Pniówek V flights. Known- U fixes wind speed to the meteorological-station value; blind inversion selects Q and U jointly.

Flight	N_{obs}	U_{meteo} (m s ⁻¹)	Known- U : $\hat{Q}$ (kt yr ⁻¹)	Blind: $\hat{Q}$ (kt yr ⁻¹)	Blind: $\hat{U}$ (m s ⁻¹)
#11	362	7.0	15.44	5.17	3.0
#12	214	7.0	5.94	3.49	2.5
#14	108	6.0	15.85	2.84	1.3
#15	130	6.0	5.27	13.26	14.0

Known- U inversion selects effective heights of 52.2, 43.8, 3.3, and 9.2 m for flights #11, #12, #14, and #15, respectively. The corresponding blind-mode heights are 61.4, 47.9, 2.2, and 10.6 m.

The GP and our flight-level MB baseline produce substantially larger or smaller point estimates under these sparse tra-
285 jectories (Table D1). GP blind estimates span 25–1792 kt yr⁻¹ and GP known- U estimates span 31–49 kt yr⁻¹, while MB estimates span 0.009–0.055 kt yr⁻¹. NeuPlume known- U estimates span 5.27–15.85 kt yr⁻¹.

5 Discussion

The 30-case validation shows that NeuPlume preserves plume location and large-scale spatial organization more reliably than cross-sectional spread. Strong positional agreement together with negative mean width bias identifies cross-sectional spread as
290 the systematic field-reconstruction error. In the six representative cases, NeuPlume also yields lower mean emission-rate error than GP and MB under random volumetric sampling. This comparison establishes NeuPlume’s advantage for emission-rate estimation under the tested sparse three-dimensional observation geometry.

5.1 Measurement geometry, wind representativeness, and identifiability

The field inversions expose strong coupling between wind-speed representativeness and the selected emission rate. Because
295 the CNF represents unit-emission fields, the fitted source strength scales with effective transport speed. Across all four flights, the blind solutions move U and Q in the same direction relative to the known- U solutions. This behavior is consistent with wind sensitivities reported in cross-sectional-flux and controlled-release studies Meier et al. (2026); Eastwood et al. (2025); Conrad et al. (2023); He et al. (2024).

The controlled geometry test shows that point count alone does not determine performance. Concentrating the matched point
300 budget near transformed UAV trajectories degrades parameter and full-field recovery on average relative to random-volume sampling (Figs. G1–G2). Each trajectory protocol represents a compound sampling pattern, but their common contrast with the matched random-volume protocol supports a clear measurement objective: sample multiple downwind positions, obtain crosswind and vertical contrast, and record wind over the plume-transport scale. UAV studies likewise identify plume coverage,

platform configuration, and wind sampling as major determinants of emission accuracy Liu et al. (2024); Mohammadloo et al. (2025); Vollrath et al. (2026); Scheutz et al. (2025).

5.2 Physical scope and boundary effects

The present implementation and validation cover passive, low-height, neutral releases over flat terrain. The current forward model uses open boundaries. Within this parameterization, H is an effective release height that can absorb unrepresented vertical transport. Flights #11 and #12 select heights near or above the 50 m training-grid ceiling after continuous refinement, marking the edge of the trained height support. The present observations do not resolve which omitted transport process produces these boundary solutions.

The two transport stress tests show why field agreement and parameter identification must be assessed separately. Strong shear changes the selected simple-model parameters without a corresponding mean loss of full-field similarity, whereas effective plume rise degrades both parameter and field recovery. In the plume-rise test, H absorbs part of the missing vertical transport and functions as an effective height rather than a physical outlet elevation.

Open-boundary diagnostics in Table E1 show that most particle exits occur through the downwind boundary and quantify the smaller side and top losses. Measurements and forward domains for weak-wind or high-turbulence regimes require sufficient crosswind and vertical extent to capture the broader plume. The current LSM also excludes flow regimes in which material exits and later re-enters through recirculation or building wakes.

The forward-library design defines a direct extension path for additional regimes. Stable and unstable boundary layers require stability-conditioned simulations because they alter wind shear, vertical exchange, turbulent-velocity distributions, and the LSM drift needed to satisfy the well-mixed condition Newman and Klein (2014); Cassiani et al. (2014). Time-varying releases require a temporally resolved library and field representation, for which sequential Bayesian source-term methods provide an established reference Xu et al. (2025).

5.3 Method positioning and validation

NeuPlume targets near-field settings where a simulation-trained representation can describe non-Gaussian structure and several source and environmental parameters must be searched jointly. GP and MB remain efficient choices when their plume and sampling assumptions hold Shah et al. (2019); Krings et al. (2013); Conley et al. (2017). The field comparison here is therefore descriptive: without a flight-specific certified emission rate, the spread among NeuPlume, GP, and MB estimates cannot establish absolute accuracy. A controlled-release experiment with a known reference emission rate for each flight would provide the decisive test of NeuPlume's absolute emission-rate accuracy under field conditions Morales et al. (2022); Vollrath et al. (2026). Such validation should span wind regimes and observation geometries while retaining the minimum-loss selection rule used here.

6 Conclusions

335 NeuPlume estimates point-source parameters and complete three-dimensional concentration fields from sparse observations by searching a generative representation of LSM simulations. Across 30 held-out synthetic cases, fields retained plume-scale spatial organization (mean spatial correlation, 0.942), and emission-rate error averaged 13.6%; reconstructed plumes remained systematically narrow. In six matched cases, diffusion improved parameter and field metrics over direct CNF library search, and NeuPlume reduced emission-rate error relative to GP and MB under random-volume sampling.

340 Stress tests identified observation geometry and forward-model fidelity as distinct controls. UAV-like trajectories weakened field reconstruction at the same observation budget. Under transport mismatch, shear shifted parameters while preserving mean field similarity, whereas effective plume rise degraded parameter and field recovery. The Pniówek inversions exposed a positive wind-speed–emission-rate coupling: higher selected wind speeds required larger emission-rate estimates to reproduce the observed concentration enhancements, whereas lower selected wind speeds produced lower estimates. NeuPlume thus provides an

345 extensible framework for jointly estimating source parameters and concentration fields. The framework was validated here for passive, low-height, neutral releases over flat terrain; additional transport regimes can be incorporated through regime-specific simulation ensembles and neural-model retraining.

Appendix A: Training and inversion hyperparameters

Table A1 lists the neural representation, diffusion training, candidate-generation, coarse-to-fine search, and emission-rate re-
350 finement settings used in the reported experiments. The real-data experiments use the same DPS settings, search I_{ref} on a
discrete grid, and set $n_{\text{fine}} = 3$.

Table A1. Training and inversion hyperparameters of the NeuPlume framework.

Component	Parameter	Value
<i>Hash-INR</i>		
	Hash levels	16
	Features per level	2
	Spatial resolutions	16^3 – 512^3
	Hidden layers	4
	Hidden units	256
	Latent dimension	192
	Training epochs	300
	Network learning rate	10^{-3}
	Latent-code learning rate	5×10^{-4}
<i>Conditional diffusion model</i>		
	Transformer blocks	6
	Attention heads	8
	Hidden dimension	512
	Diffusion steps	1000
	Training epochs	100
	Batch size	64
	Learning rate	10^{-4}
	Noise schedule	Linear, $\beta \in [10^{-4}, 2 \times 10^{-2}]$
<i>Coarse grid search</i>		
	H grid	2, 5, 10, 15, 20, 30, 40, 50 m
	U_{ref} grid	0.8–15.0 m s^{-1} (11 levels)
	I_{ref} grid	0.05–0.30 (coarse-node values)
	Candidate nodes	88

Appendix B: Detailed synthetic validation

355 Table B1 gives the complete 30-case parameter and field summary. Table B2 lists the physical conditions of the six preselected representative cases used for method comparisons and controlled stress tests. Figure B1 reports the coarse-to-fine observation-mismatch results.

Table B1. Performance across the 30-case synthetic validation set. Parameter rows report absolute relative error. Field rows compare the selected concentration field with the LSM truth. Width bias is reconstruction minus truth, so negative values indicate a narrower reconstructed plume.

Metric	Mean	Median	Range	Unobserved-point mean
H error (%)	2.390	0.752	0.075–10.032	n/a
U error (%)	23.128	16.962	0.296–67.262	n/a
I error (%)	5.742	5.414	0.196–14.597	n/a
Q error (%)	13.590	10.716	1.819–35.923	n/a
Spatial correlation	0.942	0.954	0.859–0.981	0.942
Normalized RMSE	0.345	0.320	0.216–0.543	0.345
Centerline displacement (m)	0.782	0.680	0.144–1.954	0.783
Crosswind width bias (m)	-1.202	-0.904	–2.530—–0.133	-1.202
Vertical width bias (m)	-0.764	-0.532	–2.191–0.024	-0.765

Table B2. Physical conditions for the six representative cases.

Case	Parameter-space description	H (m)	U (ms ^{–1})	I	Q (kgs ^{–1})
Case 1	Low H , high U	10.02	11.38	0.242	1.000
Case 2	Mid H/U , low I	27.21	7.43	0.080	0.875
Case 3	Mid H/U , high I	24.81	8.26	0.260	1.125
Case 4	High H , high U	39.76	11.94	0.167	1.250
Case 5	Low H , low U	10.36	3.05	0.163	0.750
Case 6	Central $H/U/I$	21.31	7.94	0.139	1.000

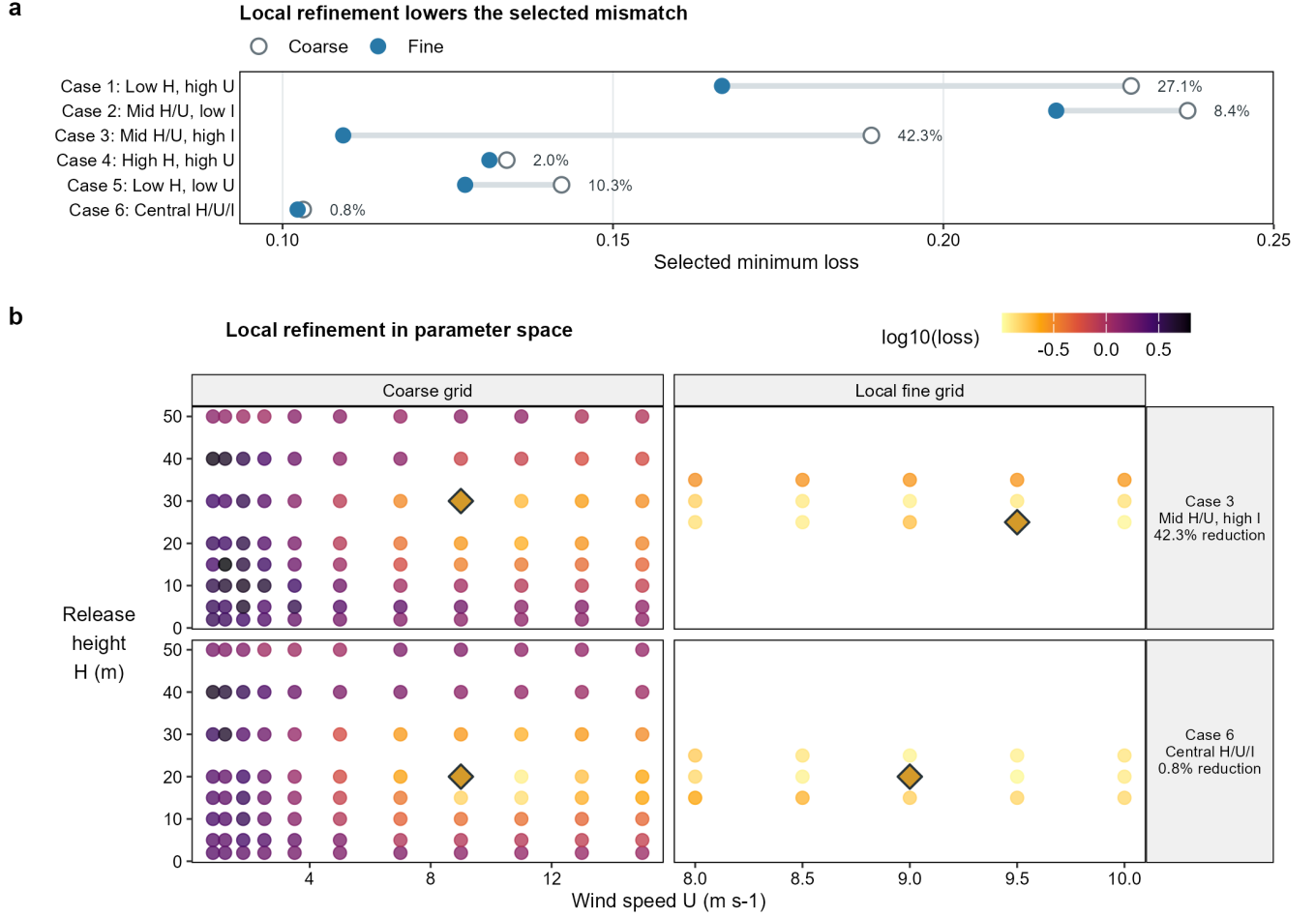


Figure B1. Coarse-to-fine optimization of the observation-mismatch objective. (a) Selected minimum losses for all six representative cases, ordered by relative reduction. The fine search lowers the selected loss in every case, with reductions from 0.8% to 42.3%. Descriptive labels summarize each case's location in the H – U – I parameter domain. (b) Coarse and local fine H – U search spaces for Case 3 (mid H/U , high I) and Case 6 (central $H/U/I$), which have the largest and smallest reductions, respectively. Coarse points show node losses; fine-grid color shows the minimum loss across candidate I values at each H – U location. Yellow diamonds mark the selected node at each stage. Each case has 88 coarse and 75 fine nodes.

Appendix C: Gaussian-plume control

Table C1 reports the Gaussian-plume control for the six representative source conditions. The analytical Gaussian-plume model generated the observations used in this control.

Table C1. Gaussian-plume control on analytical GP synthetic data. Values are absolute relative errors. Known- U errors below 0.001% show parameter recovery when the plume model and wind input match the data generator.

Case	GP known- U Q error (%)	GP known- U H error (%)	GP blind inversion Q error (%)	GP blind inversion U error (%)
Case 1	<0.001	<0.001	43.49	43.49
Case 2	<0.001	<0.001	145.04	145.04
Case 3	<0.001	<0.001	137.83	137.83
Case 4	<0.001	<0.001	66.93	66.93
Case 5	<0.001	<0.001	228.09	228.09
Case 6	<0.001	<0.001	136.44	136.44
Mean	<0.001	<0.001	126.30	126.30

Appendix D: Field-data baseline estimates

360 Table D1 reports the flight-level GP and our flight-level MB baseline estimates used to bound the field-data comparison in Sect. 4.3. These values are separate from the published campaign-level MB reference reported in the main text and Fig. 5.

Table D1. GP and MB estimates for the four Pniówek V flights. The same-campaign published IG estimate is $9.5 \pm 3.5 \text{ kt yr}^{-1}$.

Flight	GP blind	GP known U	MB known U
	$\hat{Q}$ (kt yr $^{-1}$)	$\hat{Q}$ (kt yr $^{-1}$)	$\hat{Q}$ (kt yr $^{-1}$)
#11	831	46	0.009
#12	31	41	0.013
#14	1792	49	0.055
#15	25	32	0.040

Appendix E: Boundary-loss diagnostics

The LSM open-boundary accounting provides the physical-domain diagnostic discussed in Sect. 5.2.

Table E1. Boundary-loss diagnostics for the six representative cases using the LSM configuration specified in Sect. 2.2 and a unit emission rate.

Metric	Mean	Minimum	Maximum
Particle exit fraction	0.920	0.913	0.928
Retained particle fraction	0.080	0.072	0.087
Side/top share of exits	0.032	1.4×10^{-5}	0.102
Downwind share of exits	0.968	0.898	1.000
Side/top share of emitted particles	0.029	1.3×10^{-5}	0.095

Appendix F: Transport-model mismatch

365 Figure F1 reports the case-wise changes produced by strong shear and effective plume rise relative to the matched simple-transport baseline.



Figure F1. Controlled transport-mismatch stress tests across six representative conditions within the trained parameter support. Rows show changes in normalized RMSE, correlation loss, and the absolute relative errors of H , U , I , and Q ; columns identify Cases 1–6 for each mismatch condition. Changes are relative to each case’s simple-transport baseline under the same 1% random-volume observation mask. Positive values denote worse performance, with correlation loss defined as baseline minus tested spatial correlation. Cell labels report un-truncated changes; parameter-error colors are truncated at ± 40 percentage points to preserve contrast.

Appendix G: Observation-geometry stress test

Figure G1 defines the three transformed UAV sampling protocols, and Fig. G2 reports their case-wise changes relative to matched random-volume sampling.

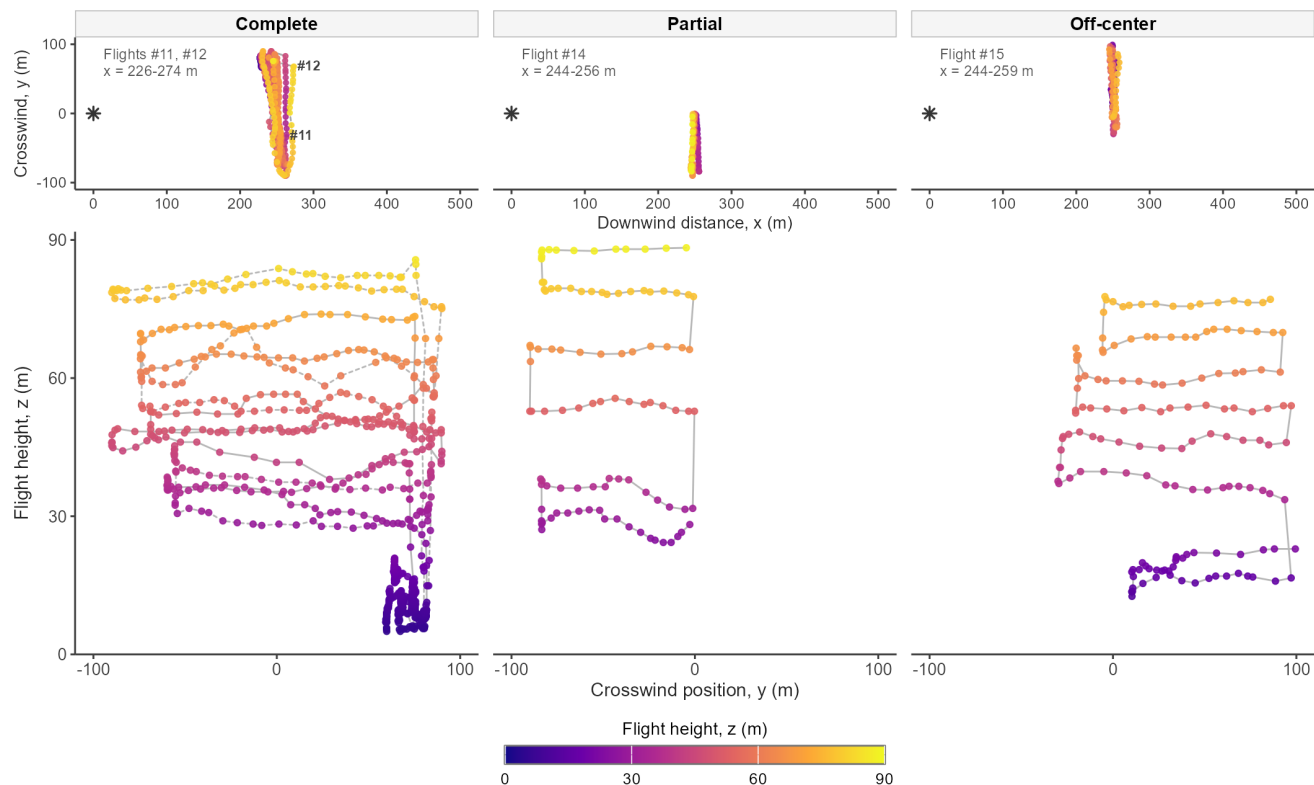
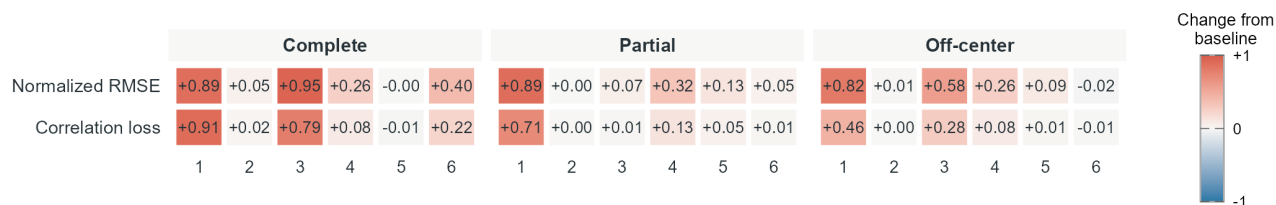


Figure G1. Complete, partial, and off-center transformed UAV sampling trajectories used in the controlled observation-geometry stress test. The upper strips locate the trajectories in the downwind-crosswind plane; stars mark the source and annotations identify the source flights and sampled downwind ranges. The lower panels show the crosswind-height sampling curtains. Color denotes flight height. Each protocol uses a different source flight and transformation and therefore represents a compound sampling geometry.

a



b

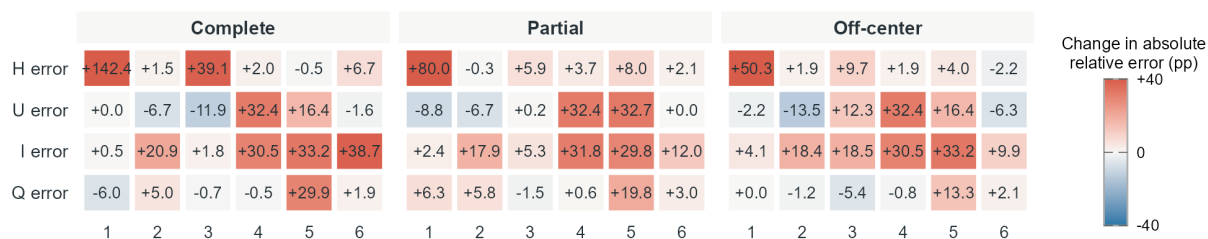


Figure G2. Case-wise degradation under three transformed UAV sampling protocols across six representative conditions. Rows show changes in normalized RMSE, correlation loss, and the absolute relative errors of H , U , I , and Q ; columns identify Cases 1–6 for each protocol. Changes are relative to each case’s matched 1,000-point random-volume baseline. Positive values denote worse performance, with correlation loss defined as baseline minus tested spatial correlation. Cell labels report untruncated changes; parameter-error colors are truncated at ± 40 percentage points to preserve contrast. Each protocol combines a different source flight and transformation and therefore represents a compound geometry.

370 *Code and data availability.* The code and data supporting this study are available from the first author upon reasonable request.

Author contributions. Lei Wang: conceptualization, methodology, software, validation, formal analysis, investigation, visualization, writing—original draft, and writing—review and editing. Xin Ma: conceptualization, supervision, project administration, funding acquisition, resources, and writing—review and editing.

Competing interests. The authors declare that they have no competing interests.

375 *Financial support.* This research was supported by the National Natural Science Foundation of China (grant nos. U25D9004 and 42171464), the Fundamental Research Funds for the Central Universities (grant nos. ZNJC202415 and 2042026kf0075), the National Key Research and Development Program of China (grant no. 2024YFC3015600), and the Science and Technology Program of Hubei Province (grant nos. 2025BEB017 and 2026BEB014).

Acknowledgements. The numerical calculations reported in this paper were performed on the supercomputing system at the Supercomputing
380 Center of Wuhan University.

References

- Aghdasi, S., Rayner, P., Deutscher, N., and Silver, J.: Exploring uncertainty reduction in high-resolution methane emissions in Gippsland through in-situ data: A Bayesian inverse modeling and variational assimilation method, *ATMOSPHERIC RESEARCH*, <https://doi.org/10.1016/j.atmosres.2025.107911>, 2025.
- 385 Andersen, T., Zhao, Z., de Vries, M., Necki, J., Swolkien, J., Menoud, M., Röckmann, T., Roiger, A., Fix, A., Peters, W., and Chen, H.: Local-to-regional methane emissions from the Upper Silesian Coal Basin (USCB) quantified using UAV-based atmospheric measurements, *ATMOSPHERIC CHEMISTRY AND PHYSICS*, <https://doi.org/10.5194/acp-23-5191-2023>, 2023.
- Cassiani, M., Stohl, A., and Brioude, J.: Lagrangian Stochastic Modelling of Dispersion in the Convective Boundary Layer with Skewed Turbulence Conditions and a Vertical Density Gradient: Formulation and Implementation in the FLEXPART Model, *Boundary-Layer Meteorology*, <https://doi.org/10.1007/s10546-014-9976-5>, 2014.
- 390 Chung, H., Kim, J., McCann, M. T., Klasky, M. L., and Ye, J. C.: Diffusion Posterior Sampling for General Noisy Inverse Problems, in: International Conference on Learning Representations, <https://doi.org/10.48550/arXiv.2209.14687>, 2023.
- Conley, S., Faloon, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R.: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-10-3345-2017>, 2017.
- 395 Conrad, B. M., Tyner, D. R., and Johnson, M. R.: Robust probabilities of detection and quantification uncertainty for aerial methane detection: Examples for three airborne technologies, *Remote Sensing of Environment*, <https://doi.org/10.1016/j.rse.2023.113499>, 2023.
- De Meutter, P., Hoffman, I., and Ungar, K.: On the model uncertainties in Bayesian source reconstruction using an ensemble of weather predictions, the emission inverse modelling system FREAR v1.0, and the Lagrangian transport and dispersion model Flexpart v9.0.2, *GEOSCIENTIFIC MODEL DEVELOPMENT*, <https://doi.org/10.5194/gmd-14-1237-2021>, 2021.
- 400 Du, P., Parikh, M., Fan, X., Liu, X., and Wang, J.: Conditional neural field latent diffusion model for generating spatiotemporal turbulence, *NATURE COMMUNICATIONS*, <https://doi.org/10.1038/s41467-024-54712-1>, 2024.
- Eastwood, M., Thompson, D. R., Green, R. O., Fahlen, J., Adams, T., Brandt, A. R., Brodrick, P., Chlus, A., Kort, E. A., Reuland, F., and Thorpe, A. K.: Direct measurement of plume velocity to characterize point source emissions, *Proceedings of the National Academy of Sciences of the United States of America*, <https://doi.org/10.1073/pnas.2507350122>, 2025.
- 405 Guo, J., Du, P., Fan, X., Li, Y., and Wang, J.: Conditional neural field for spatial dimension reduction of turbulence data: A comparison study, *PHYSICS OF FLUIDS*, <https://doi.org/10.1063/5.0310238>, 2026.
- He, Z., Gao, L., Liang, M., and Zeng, Z.-C.: A survey of methane point source emissions from coal mines in Shanxi province of China using AHSI on board Gaofen-5B, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-17-2937-2024>, 2024.
- 410 Jia, M., Fish, R., Daniels, W., Sprinkle, B., and Hammerling, D.: A fast and lightweight implementation of the Gaussian puff model for near-field atmospheric transport of trace gasses, *SCIENTIFIC REPORTS*, <https://doi.org/10.1038/s41598-025-99491-x>, 2025.
- Katata, G., Chino, M., Kobayashi, T., Terada, H., Ota, M., Nagai, H., Kajino, M., Draxler, R., Hort, M., Malo, A., Torii, T., and Sanada, Y.: Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and oceanic dispersion model, *ATMOSPHERIC CHEMISTRY AND PHYSICS*, <https://doi.org/10.5194/acp-15-1029-2015>, 2015.
- 415

- Krings, T., Gerilowski, K., Buchwitz, M., Hartmann, J., Sachs, T., Erzinger, J., Burrows, J. P., and Bovensmann, H.: Quantification of methane emission rates from coal mine ventilation shafts using airborne remote sensing data, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-6-151-2013>, 2013.
- Li, S. and Du, K.: Comparisons of forward-in-time and backward-in-time Lagrangian stochastic dispersion models for micro-scale atmospheric dispersion, *JOURNAL OF THE AIR & WASTE MANAGEMENT ASSOCIATION*, <https://doi.org/10.1080/10962247.2020.1728424>, 2020.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research: Atmospheres*, <https://doi.org/10.1029/2002JD003161>, 2003.
- 425 Liu, X., Parikh, M., Fan, X., Du, P., Wang, Q., Chen, Y., and Wang, J.: CoNFILd-inlet: Synthetic turbulence inflow using generative latent diffusion models with neural fields, *PHYSICAL REVIEW FLUIDS*, <https://doi.org/10.1103/PhysRevFluids.10.054901>, 2025a.
- Liu, Y., Paris, J.-D., Broquet, G., Roy, V. B., Fernandez, T. M., Andersen, R., Berlanga, A. R., Christensen, E., Courtois, Y., Dominok, S., Dussenne, C., Eckert, T., Finlayson, A., de la Fuente, A. F., Gunn, C., Hashmonay, R., Hayashi, J. G., Helmore, J., Honsel, S., Innocenti, F., Irjala, M., Log, T., López, C., Martínez, F. C., Martinez, J., Massardier, A., Nygaard, H. G., Reboredo, P. A., Rousset, É., Scherello, A., 430 Ulbricht, M., Weidmann, D., Williams, O., Yarrow, N., Zaréa, M., Ziegler, R., Sciare, J., Vrekoussis, M., and Bousquet, P.: Assessment of current methane emission quantification techniques for natural gas midstream applications, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-17-1633-2024>, 2024.
- Liu, Y., Zhang, Z., Yu, D., Liu, P., Yan, X., and Zhang, K.: Conditional diffusion inversion for Darcy flows in heterogeneous porous media with random parameters, *PHYSICS OF FLUIDS*, <https://doi.org/10.1063/5.0281578>, 2025b.
- 435 Meier, S., Vögtli, M., Hueni, A., McManemin, A., Brandt, A. R., Juéry, C., Blandin, V., Brunner, D., and Kuhlmann, G.: Quantifying CH₄ point source emissions with airborne remote sensing: first results from AVIRIS-4, *Atmospheric Measurement Techniques*, 19, 333–358, <https://doi.org/10.5194/amt-19-333-2026>, 2026.
- Meng, C., Gao, J., Tian, Y., Chen, H., Zhang, W., and Luo, R.: Stochastic Solutions for Simultaneous Seismic Data Denoising and Reconstruction via Score-Based Generative Models, *IEEE Transactions on Geoscience and Remote Sensing*, <https://doi.org/10.1109/TGRS.2024.3421597>, 2024.
- 440 Mohammadloo, T. H., Jones, M. C., van de Kerkhof, B., Dawson, K., Smith, B. J., Conley, S. A., Corbett, A., and Ijzermans, R.: Quantitative estimate of several sources of uncertainty in drone-based methane emission measurements, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-18-1301-2025>, 2025.
- Morales, R., Ravelid, J., Vinkovic, K., Korbe, P., Tuzson, B., Emmenegger, L., Chen, H., Schmidt, M., Humbel, S., and Brunner, D.: 445 Controlled-release experiment to investigate uncertainties in UAV-based emission quantification for methane point sources, *ATMOSPHERIC MEASUREMENT TECHNIQUES*, <https://doi.org/10.5194/amt-15-2177-2022>, 2022.
- Newman, J. F. and Klein, P.: The Impacts of Atmospheric Stability on the Accuracy of Wind Speed Extrapolation Methods, *Resources*, <https://doi.org/10.3390/resources3010081>, 2014.
- Scheutz, C., Knudsen, J., Vecchi, N., and Knudsen, J.: Validation and demonstration of a drone-based method for quantifying fugitive methane emissions, *JOURNAL OF ENVIRONMENTAL MANAGEMENT*, <https://doi.org/10.1016/j.jenvman.2024.123467>, 2025.
- 450 Shah, A., Allen, G., Pitt, J., Ricketts, H., Williams, P., Helmore, J., Finlayson, A., Robinson, R., Kabbabe, K., Hollingsworth, P., Rees-White, T., Beaven, R., Scheutz, C., and Bourn, M.: A Near-Field Gaussian Plume Inversion Flux Quantification Method, Applied to Unmanned Aerial Vehicle Sampling, *ATMOSPHERE*, <https://doi.org/10.3390/atmos10070396>, 2019.

Shaw, J., Shah, A., Yong, H., and Allen, G.: Methods for quantifying methane emissions using unmanned aerial vehicles: a review, *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES*,
455 <https://doi.org/10.1098/rsta.2020.0450>, 2021.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmospheric Chemistry and Physics*, <https://doi.org/10.5194/acp-5-2461-2005>, 2005.

Thomson, D. J.: Criteria for the Selection of Stochastic Models of Particle Trajectories in Turbulent Flows, *Journal of Fluid Mechanics*,
460 <https://doi.org/10.1017/S0022112087001940>, 1987.

Varon, D., Jacob, D., McKeever, J., Jervis, D., Durak, B., Xia, Y., and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, *ATMOSPHERIC MEASUREMENT TECHNIQUES*, <https://doi.org/10.5194/amt-11-5673-2018>, 2018.

Vollrath, C., Barchyn, T., Munn, A., Wearmouth, C., and Hugenholtz, C.: A human-portable mass flux method for methane
465 emissions quantification: controlled release testing performance evaluation, *ATMOSPHERIC MEASUREMENT TECHNIQUES*,
<https://doi.org/10.5194/amt-19-583-2026>, 2026.

Whites, E., Rigby, M., Lunt, M., Smallman, T., Comyn-Platt, E., Manning, A., Ganesan, A., O'Doherty, S., Stavert, A., Stanley, K.,
Williams, M., Levy, P., Ramonet, M., Forster, G., Manning, A., and Palmer, P.: Quantifying the UK's carbon dioxide flux: an
atmospheric inverse modelling approach using a regional measurement network, *ATMOSPHERIC CHEMISTRY AND PHYSICS*,
470 <https://doi.org/10.5194/acp-19-4345-2019>, 2019.

Xie, Z., Tang, J., Li, R., Tian, G., and Ma, R.: A Review of Source-Term Estimation for Continuous Methane Monitoring: From Data
Acquisition to Modeling and Estimation, *ACS OMEGA*, <https://doi.org/10.1021/acsomega.5c13215>, 2026.

Xu, Y., Dong, X., and Fang, S.: Efficient Bayesian source reconstruction and uncertainty quantification of atmospheric radionuclide releases
by replacing release rate sampling with Maximum-A-Posteriori estimation of time-varying release rates, *JOURNAL OF HAZARDOUS*
475 *MATERIALS*, <https://doi.org/10.1016/j.jhazmat.2025.138171>, 2025.

Yuvaraj, R., Lauvaux, T., Abdallah, C., Ciais, P., Guery, J. A., Bonne, J., Groshenry, A., Hoang, N. M., and Joly, L.: High-Resolution
Modeling of Methane Plumes: Validation and Sensitivity Experiments to Explore Emission Quantification Approaches., *Environmental*
science & technology, <https://doi.org/10.1021/acs.est.5c15941>, 2026.

Zhao, Z., Sun, S., Zhou, B., Sun, J., Ciais, P., Houweling, S., Krol, M., and Chen, H.: Advancing Point Source Methane Emissions Quan-
480 tification With UAV Sampling and Deep Learning: A Large-Eddy Simulation Study, *JOURNAL OF GEOPHYSICAL RESEARCH-*
ATMOSPHERES, <https://doi.org/10.1029/2025JD045952>, 2026.