



NeuPlume: Probabilistic 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 probabilistic inversion framework that returns an ensemble of data-compatible plume fields rather than a single point estimate. A Lagrangian stochastic model builds a scenario-specific forward library, a conditional neural field compresses the concentration fields, and a latent diffusion model learns the feasible field prior. During inversion, diffusion posterior sampling combines this prior with sparse observations to infer emission rate, effective release height, wind speed, turbulence intensity, and the full concentration field. The present implementation targets passive, low-height, near-neutral releases over flat terrain. Across six synthetic cases within this scope, NeuPlume achieves mean errors of 10.0% for emission rate and 1.4% for effective release height, outperforming Gaussian plume and mass-balance baselines under the same volumetric observation protocol. On 30 holdout cases, the nominal 68% credible interval attains 73.3% empirical coverage. As an illustrative field-transfer check, NeuPlume is applied to four UAV methane transects above a coal-mine ventilation shaft; known- U posterior intervals from three flights overlap the same-shaft hourly inventory, with diagnostics identifying wind-speed and height-boundary sensitivities under model mismatch. NeuPlume provides uncertainty-aware source-parameter constraints within the physical scope of its forward library and can be re-instantiated for other regimes by rebuilding the scenario-specific simulation ensemble and retraining the neural prior.

1 Introduction

Quantifying greenhouse-gas emissions from individual point sources is a central task in atmospheric monitoring, because such estimates support national emission inventories, regulatory compliance, and the verification of climate-mitigation commitments. Current approaches fall mainly into two classes. Gaussian plume (GP) models infer emission rate by fitting analytical dispersion kernels to downwind concentration measurements Shah et al. (2019). The mass-balance (MB) method estimates emission rate by integrating the advective flux of a tracer across one or more cross-wind planes Conley et al. (2017); Varon et al. (2018). Both paradigms have been deployed on platforms ranging from ground-based sensors to satellites and, increasingly, unmanned aerial vehicle (UAV) platforms, as comprehensively reviewed by Shaw et al. (2021).



However, under sparse near-field observations, point estimates of emission rate convey more certainty than the data can support. When the wind field, dispersion morphology, or observation cross-section is prescribed externally, sources of uncertainty that the data could otherwise constrain are removed from the inversion. GP methods also rely on far-field assumptions, including steady conditions, a spatially homogeneous wind field, and Gaussian cross-wind and vertical concentration profiles. These assumptions break down in the near-field region within approximately 1 km of the source, where turbulence and plume meandering produce non-Gaussian concentration structures Yuvaraj et al. (2026) and rapidly varying winds violate the steady-state assumption Jia et al. (2025). The MB method relaxes the Gaussian-shape assumption by directly integrating observed concentrations, but it remains sensitive to incomplete plume interception, nonuniform background subtraction, and the representativeness of the wind speed used to convert column concentrations into flux. UAV and controlled-release experiments both show that observation coverage, sampling cross-sections, and wind-field inputs can substantially affect methane quantification Mohammadloo et al. (2025); Liu et al. (2024). In this setting, a point estimate of emission rate merely summarizes the sparse concentration information available and cannot characterize the structural uncertainty jointly determined by plume morphology, observation coverage, and wind-speed representativeness.

Beyond the modeling assumptions embedded in GP and MB frameworks, the treatment of wind speed forms an equally important source of error. Because the advective flux of a passive tracer is linearly proportional to local wind speed, wind-speed bias is transmitted directly to the inferred emission rate. A campaign using the AVIRIS-4 airborne imaging spectrometer demonstrated this coupling through a formal uncertainty decomposition. Wind-speed uncertainty contributed up to 99.4% of the total variance in cross-sectional-flux inversion of methane emission rate, far exceeding spectral-retrieval noise and plume-segmentation uncertainty Meier et al. (2026). Eastwood et al. (2025) directly measured plume advection velocity from sequential airborne imaging flights using optical flow; in their controlled-release experiments, ERA5 reanalysis winds underestimated the effective plume transport velocity by approximately 28%. Crucially, existing GP, MB, and related inversion frameworks usually specify wind speed from meteorological observations, reanalysis products, or in situ anemometers, and wind speed is rarely estimated jointly from the concentration data. Even if the forward dispersion model is accurate, emission estimates can inherit systematic biases from external wind-speed inputs. Airborne methane detection and remote-sensing studies of coal-mine point sources likewise identify wind-field error as a key source of quantification uncertainty Conrad et al. (2023); He et al. (2024).

Point-source emission inversion under limited measurements is intrinsically underdetermined. Because sparse concentration samples constrain only local features of the plume field, emission rate, release height, wind speed, turbulence intensity, and field morphology remain coupled. Bayesian inversion provides a natural framework for handling this non-uniqueness by combining physical prior information with observational likelihoods to characterize a family of data-compatible solutions rather than returning a single optimum. In point-source monitoring, this probabilistic description matches the information content of the observations: the same UAV trajectory may support multiple source-height, wind-speed, and source-strength combinations, and the posterior distribution gives their relative credibility and correlations. Markov chain Monte Carlo Xu et al. (2025), variational data assimilation Aghdasi et al. (2025), regional-network atmospheric inversion Whites et al. (2019), and ensemble-based source reconstruction De Meutter et al. (2021) have all been applied to emission parameter inversion. Recent reviews



compare these methods with optimization-based alternatives Xie et al. (2026). However, classical Bayesian implementations usually depend on manually specified priors, and their computational cost rises rapidly with parameter dimension and state-space complexity, limiting applicability to joint inversion of source terms, meteorological parameters, and concentration fields.

Deep generative models provide another route: learning a low-dimensional prior from physical simulations and introducing limited measurements as likelihood constraints during sampling. Conditional diffusion models can generate calibrated posterior distributions of spatially heterogeneous permeability fields in Darcy-flow inverse problems Liu et al. (2025b), and score-based posterior sampling can reconstruct seismic wavefields from incomplete and noisy measurements Meng et al. (2024). More closely related to this work, conditional neural field (CNF) latent diffusion frameworks use CNF compression, a latent diffusion prior, and diffusion posterior sampling (DPS) to generate and reconstruct spatiotemporal turbulent fields Du et al. (2024); Liu et al. (2025a). These studies show that complex physical fields can be encoded in a compact latent space and recovered under observational constraints. In the atmospheric point-source setting, Zhao et al. (2026) combined a U-Net encoder with inverse Gaussian fitting to reduce single-source emission rate error from UAV data, but that framework still targets a point estimate and does not provide a joint posterior distribution over source terms, wind speed, and concentration fields. These advances suggest that point-source inversion can learn a plume prior from high-fidelity physical simulations, rather than prescribing plume morphology as Gaussian. During inversion, incomplete concentration data concentrates the posterior on control parameters and concentration fields compatible with both the learned prior and the measurements, characterizing not only the emission rate but also the compatible meteorological conditions, plume structure, and their joint uncertainty.

This paper presents NeuPlume, a three-stage probabilistic inversion framework for atmospheric point-source emissions. In the first stage, a Lagrangian stochastic model generates a library of concentration fields covering physically relevant ranges and establishes the forward mapping from source and meteorological parameters to concentration fields. In the second stage, a conditional neural field (CNF) compresses each field into a low-dimensional latent representation, and a denoising diffusion probabilistic model learns a field prior conditioned on the source and meteorological parameter vector. In the third stage, NeuPlume performs observation-guided diffusion posterior sampling (DPS) on a coarse-to-fine parameter grid. It generates concentration-field samples for each grid node, analytically estimates the emission rate, and weights the nodes by reconstruction loss to form a discrete approximate posterior over source parameters. The main contributions of this paper are:

1. We formulate near-field point-source emission inversion as joint posterior reconstruction of physically feasible plumes and their controlling parameters, thereby incorporating non-uniqueness under sparse observations into the inversion result.
2. We provide a neural-physical implementation path that is not tied to a specific forward simulator: as long as the target scenario can be represented by a forward model mapping parameters to concentration fields, the resulting simulation ensemble can be used to learn the corresponding field prior and infer the parameter posterior under observational constraints. The transferable object is the workflow for rebuilding a scenario-specific physical library and neural prior, not direct reuse of the LSM weights trained here.



3. We evaluate inversion accuracy on six synthetic cases, showing that NeuPlume outperforms Gaussian plume and mass-balance baselines under the tested near-neutral conditions, and assess posterior calibration on a holdout set of 30 cases.
4. We use public UAV methane observations above a coal-mine ventilation shaft as an illustrative field-transfer case, compare posterior emission estimates with the same-shaft hourly inventory as a consistency reference, and use ablation experiments to identify the observation-coverage conditions under which inversion quality degrades.

The real-data case probes robustness under model mismatch rather than validating ventilation-shaft-specific plume-rise physics. The limitation lies in the present LSM training library, not in the posterior-inference architecture itself. For a different physical regime, NeuPlume must be re-instantiated by regenerating the scenario-specific forward-simulation ensemble and retraining the CNF and diffusion prior before inference. The framework and its evaluation are presented in Sections 2–6.

2 Methods

2.1 NeuPlume model architecture

NeuPlume learns plume morphologies from forward simulations and constrains the resulting physical prior with field measurements to produce a joint posterior over the three-dimensional concentration field and its controlling parameters.

Computationally, NeuPlume organizes point-source emission inversion as a three-stage probabilistic framework that jointly estimates source parameters, meteorology, emission rate, and the complete concentration field from incomplete observations (Fig. 1). Stage 1 uses a Lagrangian stochastic model (LSM) to generate concentration fields. Each field is simulated at a unit emission rate $Q = 1 \text{ kg s}^{-1}$ and indexed by $\theta = (H, U, I)$, where H is the release height, U the reference wind speed, and I the turbulence intensity; together, these simulations define the physically plausible fields for the target scenario. Stage 2 trains a conditional neural field (CNF) to represent each concentration field as $f_{\psi}(\mathbf{x}; \theta)$. This implicit-field representation provides a low-dimensional state space for the generative model. A denoising diffusion probabilistic model learns the distribution of these representations under training conditions and supplies a generative field prior. Stage 3 performs observation-guided diffusion sampling over candidate θ values and forms a discrete approximate posterior through analytical Q estimation and reconstruction-loss weighting.

The corresponding inversion problem can be written as the following observation model.

Let $C(\mathbf{x} | \theta)$ denote the steady-state concentration field produced by the LSM under condition θ , and let $\mathcal{H}$ be the observation operator mapping the complete field to measurement locations, such as points along a UAV flight path. The observed concentration vector $\mathbf{y}$ follows

$$\mathbf{y} = Q \mathcal{H}[C(\mathbf{x} | \theta)] + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (1)$$

where the linear scaling by Q follows from unit-rate normalization. The goal is to recover the posterior jointly determined by θ , Q , and the underlying concentration field. The current implementation samples concentration fields from the generative prior at a given θ and represents parameter-level inversion with source-parameter marginal posterior $p(\theta, Q | \mathbf{y})$. By Bayes'

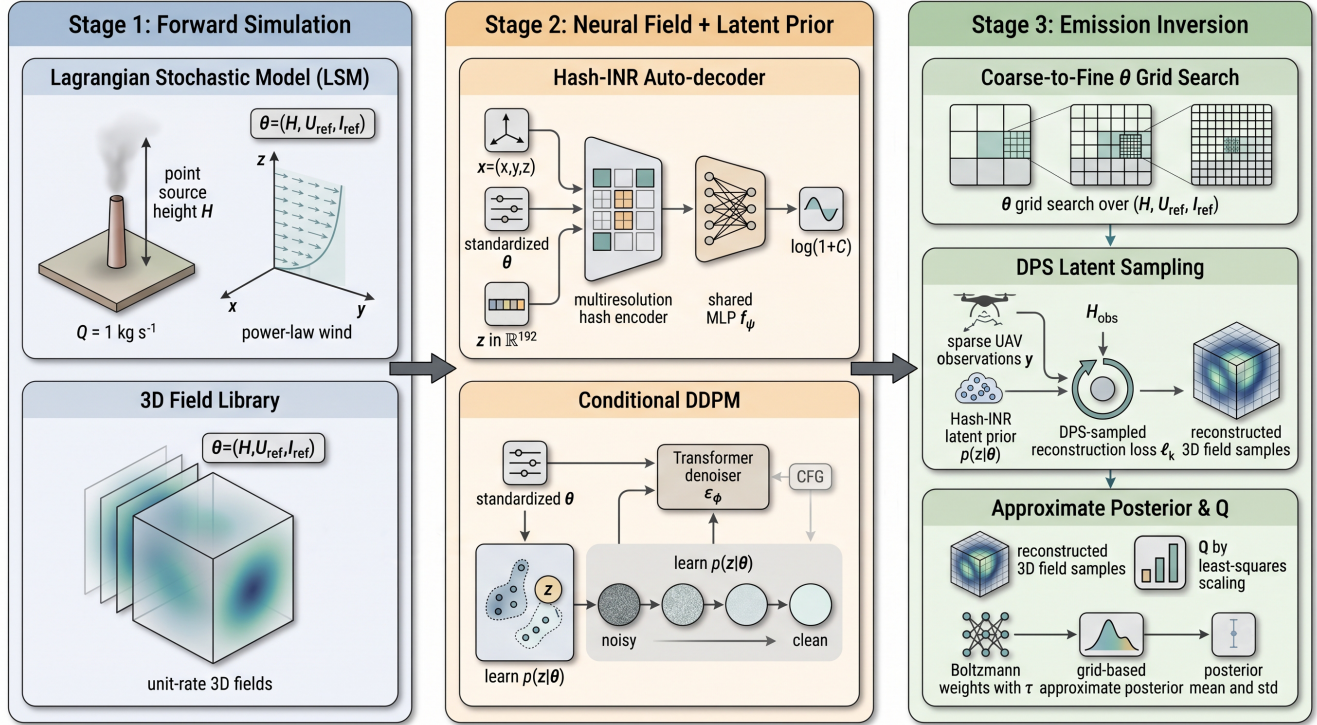


Figure 1. Overview of the three-stage NeuPlume framework. Stage 1: the Lagrangian stochastic model generates training concentration fields for 1000 parameter combinations (H, U, I) under a unit emission rate. Stage 2: the Hash-INR conditional neural field compresses each field into a latent code, and the diffusion model learns the latent-code distribution. Stage 3: diffusion posterior sampling combines the learned prior with sparse point observations to invert source parameters and reconstruct the complete three-dimensional concentration field.

theorem,

$$p(\theta, Q | y) \propto p(y | \theta, Q) p(\theta, Q). \quad (2)$$

125 NeuPlume implements this decomposition in two components: the learned implicit-field model defines a generative prior over physically consistent concentration fields, and the observation operator and noise model define the likelihood $p(y | \theta, Q)$. Computationally, DPS injects the observation-likelihood gradient into the reverse diffusion process, so generated samples match both the learned field distribution and the measurement data. The following subsections describe the LSM forward simulator (Section 2.2), the CNF and diffusion training procedure (Section 2.3), and the inversion algorithm (Section 2.4).

130 2.2 Lagrangian stochastic forward simulation

NeuPlume first takes source and meteorological parameters as inputs to generate concentration fields for neural field compression and diffusion prior learning. The forward simulator uses a Lagrangian stochastic model (LSM). LSMs represent turbulent dispersion with stochastic particle trajectories and must satisfy a stationary phase-space distribution, Eulerian-equation com-



patibility, 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 further confirm their suitability for local-scale dispersion and source-receptor sensitivity problems Li and Du (2020). Relative to the Gaussian plume formula, an LSM can represent 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 $\mathbf{x}$ and turbulent velocity fluctuation $\mathbf{u}'$, 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 $\bar{U}(z) = U_{\text{ref}}(z/z_{\text{ref}})^\alpha$; under neutral atmospheric stability, $\alpha = 1/7 \approx 0.143$, where U_{ref} is the reference wind speed at $z_{\text{ref}} = 10$ m. The $1/7$ exponent is often used as a neutral-condition approximation, but the actual wind-shear exponent varies with atmospheric stability and surface roughness. The exponent usually increases under stable conditions and decreases under unstable conditions Newman and Klein (2014). Non-neutral boundary layers also alter turbulence statistics, drift terms, and implementation of the well-mixed constraint Cassiani et al. (2014). Thus, extending NeuPlume to stable or unstable stratification requires more than replacing the power-law exponent; the LSM training library must be regenerated under the corresponding stability conditions and the CNF retrained. 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 $c(\mathbf{x})$.

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). 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.

The current LSM instance is therefore defined by a specific set of physical assumptions: a passive inert tracer, no initial source momentum, no explicit buoyant plume rise, no wind-direction shear, no convergence or divergence, 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. This open-boundary treatment is consistent with single-pass unidirectional transport,



but it cannot represent particle re-entry in recirculating flow, building wakes, or terrain-induced closed circulations. In the six synthetic test cases used for the manuscript, a diagnostic rerun with $\Delta t = 0.5$ s, 250 emitted particles per step, and one wash-out cycle shows that 92.5% of emitted particles leave the domain by the end of the simulation. Most exits occur at the downwind boundary (mean share of exits 91.8%), while side and top exits account for 8.2% of exits on average, with a maximum of 20.7% in the low-wind, high-turbulence cases. These diagnostics support treating downwind exits as open-domain outflow under the present flow regime, while also identifying finite side and top boundaries as a source of truncation when turbulence is strong or wind speed is weak.

This setting applies to many tracer-release activities under moderate to strong winds, but the framework has not yet been evaluated under stable or unstable stratification, plume-rise-dominated ventilation shafts, complex terrain, or recirculating meteorology (Section 5.4). A total of 1 000 parameter combinations are uniformly sampled from this space. This sample size covers the three-dimensional parameter space while keeping the LSM generation cost manageable. Smaller training sets were not systematically evaluated. For each combination, the LSM is integrated over a $520 \times 200 \times 100$ m³ computational domain ($x \in [-20, 500]$, $y \in [-100, 100]$, $z \in [0, 100]$ m), with time step $\Delta t = 0.5$ s, 250 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 of NeuPlume compresses the discrete concentration fields generated by the LSM into a continuous, conditionable, low-dimensional representation to reduce storage and interpolation cost in diffusion modeling and inversion. For each source-meteorological condition $\theta = (H, U_{\text{ref}}, I_{\text{ref}})$, an implicit neural function $f_{\psi}(\mathbf{x}; \theta)$ maps any spatial coordinate $\mathbf{x} \in \mathbb{R}^3$ to concentration. This representation preserves spatial continuity and conditional dependence while compressing each field for the diffusion model. Conditional neural fields have been used for spatial dimension reduction and continuous-field representation of turbulence data Guo et al. (2026). We implement this compression with *Hash-INR*.

Hash-INR contains a multiresolution hash encoder and a condition-concatenated MLP. Spatial coordinates pass through a 16-level hash encoder. Each level has a learnable hash table with two features per entry, and resolutions span geometrically from 16^3 to 512^3 voxels. Trilinear interpolation within levels and concatenation across levels produce a 32-dimensional spatial feature vector. Before concatenation, the conditional parameters θ are standardized to zero mean and unit variance with training-set statistics, then joined with the spatial features and a sample-specific 192-dimensional latent code $\mathbf{z}$. This 227-dimensional vector enters a four-hidden-layer MLP, with 256 units per layer and ReLU activations, to predict $\log(1 + C)$ at each query point. We also evaluated a SIREN architecture with periodic activations, but it was unstable in the current setting; Appendix A gives the comparison.

Hash-INR is jointly optimized over all N training conditions for 300 epochs. Network weights and per-sample latent codes use Adam with learning rates of 10^{-3} and 5×10^{-4} , respectively. The objective combines log-space mean squared error, a linear-space penalty for high-concentration regions, and ℓ_2 regularization of latent codes. The penalty is applied to values



above the 90th percentile of the concentration distribution. For stability, the gradient norm is clipped to 1, and the learning rate is halved after a validation plateau. Key hyperparameters are summarized in Table 1.

Table 1. Key hyperparameters of the NeuPlume framework.

Component	Parameter	Value
Hash-INR	Levels	16
	Features/level	2
	Layers	4
	Hidden units	256
	Latent dimension	192
	Epochs	300
Diffusion	Blocks	6
	Heads	8
	Hidden dim.	512
	Steps	1000
	Epochs	100
Inversion	Coarse points	88
	Fine points	75
	DPS	1000
	Scale	2.5
	τ	1.0

After Hash-INR converges, each training condition θ_i has a latent code $\mathbf{z}_i \in \mathbb{R}^{192}$ that, with the shared network weights, encodes its concentration field. The set $\{\mathbf{z}_i\}_{i=1}^N$ is a low-dimensional representation of the diffusion fields. We train a denoising diffusion probabilistic model (DDPM) to learn $p(\mathbf{z} | \theta)$ and use it as the latent space prior for posterior sampling in Section 2.4. Inspired by the CoNFILd latent space diffusion framework Du et al. (2024), the denoiser uses a Transformer-based DiT architecture. The 192 latent dimensions are sequence tokens, projected to a 512-dimensional hidden space and combined with learnable positional encodings. The denoiser has six Transformer blocks, each with 8-head self-attention, cross-attention, and a feed-forward network with expansion ratio 4. The diffusion timestep is embedded by a 256-dimensional sinusoidal encoding and a two-layer MLP, and θ by an independent MLP. Their sum is used as a single-token context. The forward process uses a 1000-timestep linear noise schedule, with β increasing from 10^{-4} to 2×10^{-2} . The diffusion model is trained for 100 epochs with batch size 64, Adam, a 10^{-4} learning rate, and cosine annealing. Latent codes and condition vectors are z-score normalized before training.

The diffusion model uses classifier-free guidance (CFG) to improve conditional consistency during inference. During training, the condition vector is zeroed with probability p_{drop} , so a single denoiser learns both conditional and unconditional noise



prediction. During inference, the two predictions are combined with guidance scale $s > 1$:

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

where $\epsilon_{\emptyset}$ and ϵ_{θ} denote unconditional and conditional predictions, respectively. Together, the trained Hash-INR decoder and latent diffusion model form a conditional generative model: for any θ , the reverse diffusion process draws latent samples from $p(\mathbf{z} | \theta)$, and Hash-INR decodes them into complete concentration fields, which Section 2.4 combines with observational constraints. The Hash-INR and diffusion architectures and training hyperparameters are listed in Table 1.

2.4 Point-source emission inversion

NeuPlume performs inversion in the candidate source and meteorological parameter space. For each candidate θ , the conditional diffusion model generates concentration-field samples in the Hash-INR latent space. The emission rate Q is estimated analytically from the linear relationship between the observed concentration and the predicted unit field. The reconstruction loss is then compared across candidate parameters to construct a discrete approximate posterior over source parameters. The result is a weighted set of physically plausible solutions that explain the observations.

The generative backbone of NeuPlume follows the idea of latent space diffusion for conditional neural fields, running the reverse diffusion process in the latent space of INR weights to generate conditional concentration-field samples Du et al. (2024); Liu et al. (2025a). Observational constraints are introduced through DPS, which incorporates the likelihood gradient of a known forward observation operator into the reverse diffusion process. This step balances the generative prior and measurement consistency Chung et al. (2023). Starting from Gaussian noise $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the reverse process uses the noise-prediction network $\epsilon_{\phi}(\mathbf{z}_t, t, \theta)$ to iteratively denoise the latent vector. Classifier-free guidance combines conditional and unconditional predictions:

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

where s is the guidance scale. The predicted clean latent vector $\hat{\mathbf{z}}_0$ is obtained through the standard DDPM relation $\hat{\mathbf{z}}_0 = (\mathbf{z}_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}_t) / \sqrt{\bar{\alpha}_t}$ and is decoded by the INR to produce a concentration field, which is then compared with the observations. When observation guidance is enabled, the observation-likelihood gradient is injected into the reverse process to guide sampling. Following the DPS framework and score-based posterior sampling in related geophysical inverse problems Meng et al. (2024), the noise update is modified as

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

where ζ controls the strength of observation guidance, and $\mathcal{L}_{\text{obs}}$ is the observation loss defined below. Because each stochastic reverse trajectory produces a distinct ψ sample, running multiple trajectories from independent noise realizations yields an ensemble of concentration fields, providing candidate samples for subsequent grid-node loss evaluation and approximate posterior construction. This ensemble represents possible plumes compatible with the observations under the given candidate parameters.



The observation operator maps the generated concentration field to measurement space. Given a candidate parameter vector $\theta = (H, U_{\text{ref}}, I_{\text{ref}})$ and emission rate Q , the CNF-reconstructed field is sampled at M observation locations and scaled by Q , giving the predicted observations $\hat{y} = Q \cdot \mathcal{H}[c(\theta)]$, where $\mathcal{H}$ denotes spatial interpolation at the sensor locations. The observation loss is defined in smoothed L_1 (Huber) form:

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

This loss behaves as mean squared error (MSE) for small residuals and transitions to L_1 loss for large errors, reducing sensitivity to outlying observations. Because the concentration field of an inert tracer in steady transport scales linearly with source strength, for each candidate θ , the emission rate Q is analytically estimated as the least-squares scaling factor between predicted and observed concentrations, rather than being included in the grid search. The remaining parameters are estimated with a coarse-to-fine grid search strategy. In the coarse-search stage, the algorithm evaluates a broad grid covering physically reasonable ranges of H , U_{ref} , and I_{ref} . For each grid node, one or more DPS reverse trajectories are executed, and the lowest reconstruction loss is recorded. The fine-search stage refines the search around the maximum a posteriori (MAP) estimate from the coarse-search stage, using a denser grid and additional DPS samples at each node, thereby improving the precision and reliability of subsequent uncertainty estimation.

NeuPlume supports two complementary inversion modes. In the *known wind speed* mode, the reference wind speed is provided by meteorological measurements and kept fixed during inversion. The search space is reduced to (H, I_{ref}) . Because a major source of degeneracy is removed, the grid search can converge efficiently. This mode is appropriate when reliable wind-speed observations are available at or near the measurement points. In the *blind* mode, the conditional parameters $(H, U_{\text{ref}}, I_{\text{ref}})$ are jointly searched without external wind-speed data, while Q is still analytically recovered at each candidate node. The blind mode is more broadly applicable, but it faces the inherent $H-U-Q$ degeneracy: different combinations of source height, wind speed, and emission rate can produce similar downwind concentration patterns. Coarse-to-fine grid search partially mitigates this degeneracy by systematically exploring the parameter space. Compared with the known-wind-speed case, the residual ambiguity appears as a wider approximate posterior distribution. Both modes share the same generative backbone (DDPM + CFG), and only the dimension and range of the parameter grid differ.

2.4.1 Uncertainty quantification

The multi-sample grid search provides the discrete loss set needed to construct an approximate posterior over source parameters. Limited measurements may simultaneously support multiple source-term and environmental-parameter combinations, so the posterior distribution itself is part of the inversion result rather than only an error bar after point estimation. At each grid node θ_k , the best reconstruction loss ℓ_k from N independent DPS trajectories serves as a surrogate for the negative log-likelihood. A set of Boltzmann-type weights is then computed:

$$w(\theta_k) \propto \exp\left(-\frac{\ell_k - \ell_{\min}}{\tau}\right), \quad (8)$$



where $\ell_{\min} = \min_k \ell_k$ improves numerical stability, and τ is the energy-scale parameter controlling the sharpness of the resulting distribution. The weighted mean and standard deviation of each marginal parameter distribution are obtained directly from these weights. The shape of the joint distribution is then used to diagnose parameter correlations, multimodality, and identifiability differences caused by the observation geometry. This construction is a discretized Gibbs distribution on the evaluated grid, not a strict Bayesian posterior. We accordingly term it an *approximate posterior*.

The default energy scale $\tau = 1.0$ was selected to balance posterior sharpness and coverage reliability. To verify the calibration of the approximate posterior, we adopted a holdout protocol: 30 synthetic test conditions that were independently drawn from the training distribution and were not used in any parameter tuning were inverted, and the empirical coverage of nominal credible intervals was evaluated. A well-calibrated approximate posterior should produce coverage close to the nominal level. Detailed calibration results are reported in Section 4.3.

3 Experimental setup

The experimental setup includes synthetic cases, real UAV observations, method configurations, and evaluation metrics. These experiments test inversion accuracy against known truth within the current LSM training domain, posterior calibration across conditions, and emission-rate consistency in field observations. For synthetic validation, we constructed six test cases whose source and meteorological parameters cover limiting conditions in the Lagrangian stochastic model (LSM) training distribution ($H \in [0, 50]$ m, $U_{\text{ref}} \in [1, 15]$ m s⁻¹, $I_{\text{ref}} \in [0.05, 0.30]$). Under neutral atmospheric stability, the LSM simulates particle dispersion in a $520 \times 200 \times 100$ m³ domain and discretizes the time-averaged concentration field onto 10^5 grid points. Table 2 lists the true parameters of each case. Case 1 represents typical mid-range conditions ($H = 26.8$ m, $U = 7.7$ m s⁻¹, $I = 0.132$). The other cases represent high turbulence intensity (Case 2, $I = 0.251$, and Case 6, $I = 0.294$), a low source with weak wind (Case 3, $H = 9.4$ m, $U = 4.0$ m s⁻¹), a high source under strong wind (Case 4, $H = 47.9$ m, $U = 14.1$ m s⁻¹), and extremely low wind speed (Case 5, $U = 1.3$ m s⁻¹). The emission rate is fixed at $Q = 1.0$ kg s⁻¹ in all cases, so performance differences mainly arise from differences in meteorological parameters and source-height conditions. For each case, the LSM generates the full concentration field, and a subset of grid points is randomly sampled as virtual observations to mimic sparse, irregular data in UAV measurements; the corresponding full-field morphologies are shown in Figure 2. This three-dimensional random sampling protocol tests NeuPlume’s ability to recover source terms from volumetric concentration data. It differs from cross-sectional flux flight geometries, so the synthetic baseline results should be interpreted in light of the observation geometry. All synthetic inversions use a fixed random seed (seed = 42) for reproducibility. Cross-condition robustness is verified with an independent 30-case uncertainty quantification calibration experiment (Section 4.3).

The real-data experiment uses a methane measurement campaign at the Pniówek coal mine (shaft V) in the Upper Silesian Coal Basin, Poland. We selected four AirCore-UAV flights above shaft V (#11, #12, #14, and #15), each covering different meteorological conditions and plume-interception quality. All four flights were conducted on 21 May 2018 under moderate wind conditions (ground-station $U = 6\text{--}7$ m s⁻¹). This wind-speed range falls within the current LSM training range, but the Pniówek shaft may involve initial momentum, buoyant plume rise, non-neutral boundary-layer structure, or local flow



effects that are outside the passive, neutral training library. Atmospheric stability on the measurement day was not directly measured. The real-data experiment is therefore designed as an illustrative consistency check and diagnostic stress test under model mismatch, not as an absolute validation of ventilation-shaft emission accuracy. In preprocessing, we reversed the raw analyzer concentration sequence and time-aligned it with the GPS trajectory, assigning each methane observation to a spatial location along the flight path. External comparisons include the same-shaft hourly emission inventory ($13.3 \pm 4.1 \text{ kt yr}^{-1}$) and the published inverse Gaussian (IG) plume-fitting result ($9.5 \pm 3.5 \text{ kt yr}^{-1}$, $n = 13$) and mass-balance (MB) integration result ($9.8 \pm 5.0 \text{ kt yr}^{-1}$, $n = 14$) from the same campaign Andersen et al. (2023).

NeuPlume is compared with two baseline methods using the same observation data and meteorological inputs. The GP model assumes an analytical steady-state plume solution and inverts Q , H , and U by minimizing the residual sum of squares between predicted and observed concentrations. GP inversion uses differential evolution for global search, followed by gradient-based local refinement (L-BFGS-B) initialized at the global optimum. The MB method interpolates observed concentrations onto a two-dimensional cross-section perpendicular to the mean wind direction and estimates the emission rate by integrating the resulting concentration flux. Q is obtained directly from the integrated flux without an explicit plume model. GP inversion is run in two modes. In blind-estimation mode, all parameters are estimated only from observations. In known U mode, wind speed is specified from meteorological-station data to isolate the effect of wind-speed uncertainty on emission estimation. The MB method is run only in known U mode because it requires wind speed to calculate advective flux through the cross-section. NeuPlume also supports both modes, and its known U variant fixes wind speed during diffusion posterior sampling. Thus, the comparison covers both fully data-driven inversion and inversion constrained by a given wind speed.

Evaluation metrics are defined by data source. For synthetic cases, parameter-estimation error is measured as the relative error (in %) of each inverted physical parameter (Q , H , U) against the true value. For synthetic cases with a known full concentration field, Supplementary Fig. S2 provides a qualitative visual comparison of orthogonal concentration slices between true and reconstructed fields. Uncertainty quantification (UQ) calibration is evaluated by empirical coverage probability. A well-calibrated approximate posterior distribution should contain the true value within its 68%, 90%, and 95% credible intervals at the corresponding nominal probabilities. Coverage evaluation uses 30 cases independently drawn from the training distribution, avoiding conflation of calibration quality with single-case performance. For real-data cases, the true Q is unknown, so parameter errors relative to ground truth are not reported. NeuPlume emission rate estimates are compared with the same-shaft hourly emission inventory and with the IG and MB estimates reported by Andersen et al. (2023) only as consistency references.

4 Results

4.1 Synthetic data validation

Table 2 summarizes inversion results for the six synthetic test cases. All cases assume neutral atmospheric stability and span a range of source and meteorological conditions: wind speed $U = 1.3\text{--}14.1 \text{ ms}^{-1}$, effective release height $H = 9.4\text{--}47.9 \text{ m}$, and turbulence intensity $I = 0.08\text{--}0.29$. The design includes a near-calm low-source release (Case 3), a high-source high-wind-speed release (Case 4), and strong-turbulence conditions (Cases 2 and 6), testing how the learned field prior constrains



parameter posterior distributions under different morphologies. NeuPlume's coarse-to-fine (c2f) inversion jointly estimates all
345 source parameters and achieves a six-case mean relative error of 10.0

Figure 2 shows the Lagrangian stochastic model ground-truth concentration fields for the six cases as maximum-intensity
projections (MIPs). The fields show asymmetric, non-Gaussian near-source dispersion, revealing spatial structures that are
poorly captured by a Gaussian plume assumption. Figure 3 reports the grid search loss surfaces from diffusion posterior
sampling. In each case, the loss in the (U, H) plane has a single, clear minimum aligned with the true parameter location. This
350 unimodal loss structure indicates that, within the investigated range, the posterior mode provides a stable point estimate for the
source parameters. Together with the error statistics in Table 2, these results show that NeuPlume constrains source parameters
by preserving non-Gaussian plume structures in the learned field prior.

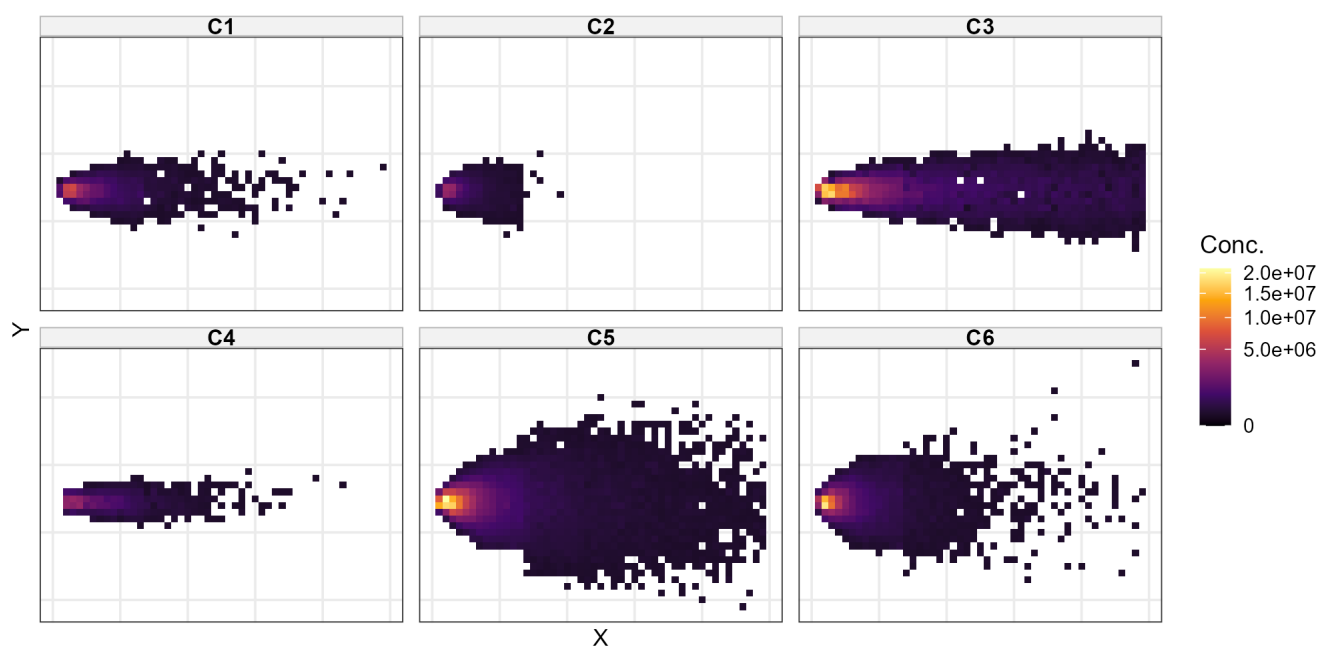


Figure 2. Lagrangian stochastic model ground-truth concentration fields for six synthetic test cases (top row: Cases 1–3; bottom row: Cases 4–6), shown as z -axis maximum-intensity projections (MIPs). The cases cover non-Gaussian plume morphologies.

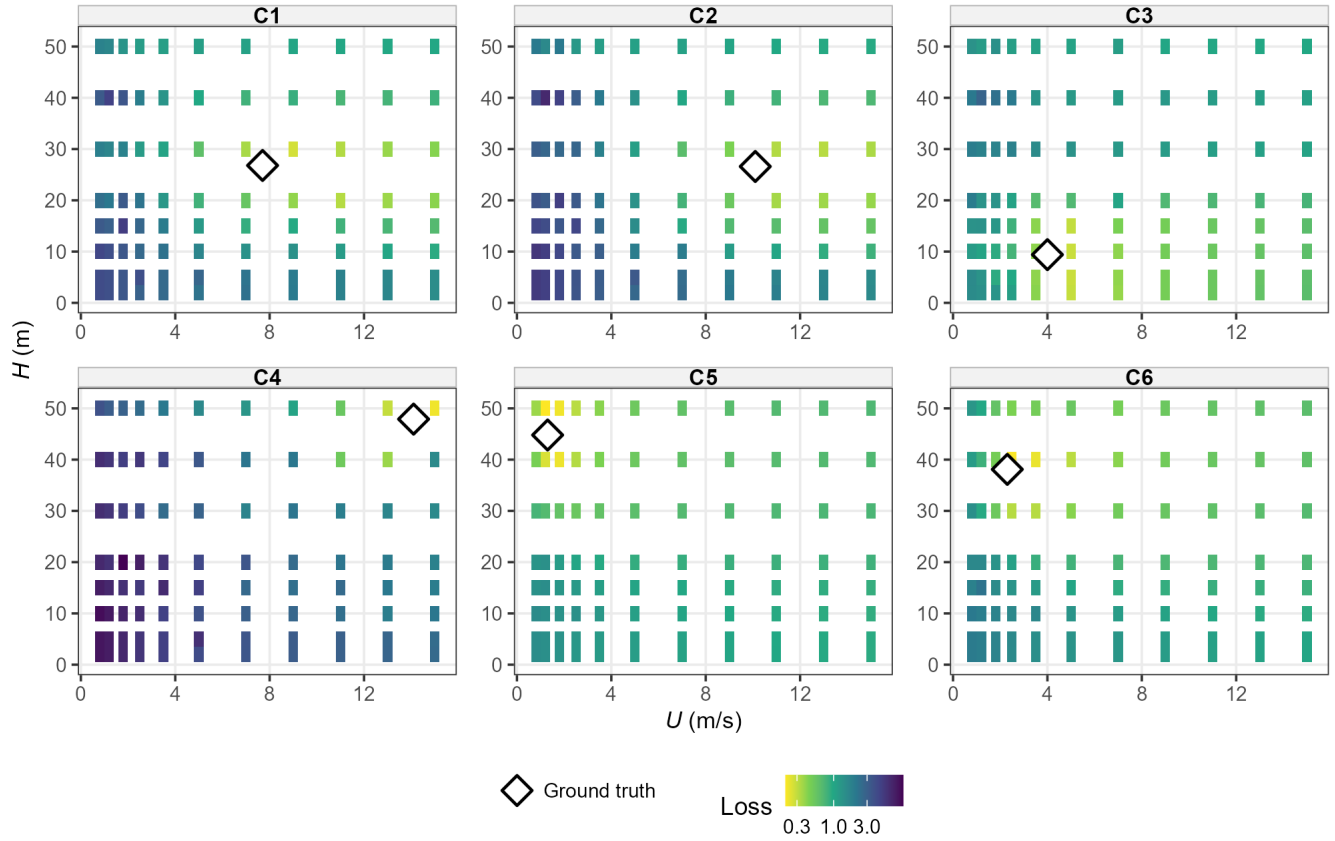


Figure 3. Coarse-to-fine grid search loss surfaces for six synthetic test cases in the (H, U) parameter plane. The cost functions from diffusion posterior sampling form a clear minimum aligned with the true value (white diamond), indicating a stable unimodal objective in the investigated range.

4.1.1 Parameter accuracy

Coarse-to-fine search improves the estimation accuracy of Q and H . With coarse grid search alone, the mean relative error of Q is 25.0

This gain is most pronounced in cases with strong turbulence or low release height. For the strong-turbulence cases (Cases 2 and 6, $I > 0.25$) and the low-release-height case (Case 3, $H = 9.4$ m), broader or near-ground plumes produce flatter coarse-resolution loss surfaces, reducing coarse search's ability to distinguish the optimum. In these cases, coarse-only Q errors all exceed 34

The parameters also differ in identifiability. H is the most stable parameter across the six cases, with a relative-error range of 0.3–4.8



Table 2. Inversion results for six synthetic test cases. Values are relative errors (%) against true values. Dashes mark parameters not reported by the baseline.

Case	NeuPlume c2f			GP blind		GP known U		MB known U
	Q	H	I	Q	U	Q	H	Q
1 – Typical	14.2	1.1	1.0	1836	—	383	—	525
2 – High I	10.9	1.3	4.0	71	—	232	—	494
3 – Low H , low U	17.3	4.8	6.8	6707	—	441	—	1656
4 – High H , high U	2.1	0.3	7.5	444	—	946	—	1998
5 – Low U	7.3	0.5	5.2	937	—	12	—	387
6 – Very high I	8.2	0.6	12.2	4896	—	58	—	408
Mean	10.0	1.4	6.1	2482	—	345	—	912

4.1.2 Baseline comparison

A baseline comparison evaluates the contribution of non-Gaussian field representation to estimating source parameters. We apply NeuPlume and two classical point-source methods to the same synthetic observations: GP regression Shah et al. (2019) and the MB method. This protocol uses random volumetric point observations rather than plume-crossing transects, which is unfavorable to GP and MB because both baselines are designed around analytical plume shapes or cross-sectional flux integration. The comparison therefore tests how each method handles observation geometry mismatch, not its performance under ideal sampling conditions. Table 2 shows that, when all source parameters are unknown in blind inversion, GP has a mean relative Q error of 2482

Even when the true wind speed is provided (known- U mode), GP still produces a mean relative Q error of 345

For MB, Q errors range from 387 to 1998% in known- U mode, with a mean of 912%. Under the observation protocol formed by 1 000 random sampling points within the computational domain, random point observations do not provide the cross-section information required by the MB method. In contrast, NeuPlume uses the learned concentration-field structure directly, keeping Q errors one to two orders of magnitude below the baseline methods.

4.1.3 Gaussian plume control experiment

The Gaussian plume control experiment separates model-assumption mismatch from implementation error. We use the same source parameters as the six synthetic cases, generate synthetic data with an analytical Gaussian plume model, and repeat the baseline comparison (Supplementary Table S2). On data satisfying the Gaussian assumption, Gaussian plume regression in known- U mode recovers all six cases, with Q and H errors close to zero ($< 10^{-6}$)

By contrast, GP in blind-inversion mode still overestimates Q on Gaussian data, with a mean error of 718

In the fully blind-inversion configuration, NeuPlume also treats wind speed U as unknown and jointly estimates Q , H , U , and I . Compared with known- U , errors in Q and H remain similar, but U error increases, with a six-case mean of 13.2



4.2 Real-data case study

The real-data experiment is interpreted under model mismatch. The current LSM library does not include initial momentum, buoyant plume rise, non-neutral stability, or local recirculation, all of which may occur near a coal-mine ventilation shaft. We therefore use the Pniówek V data as an illustrative field-transfer consistency check rather than an absolute emission-validation experiment. We analyze four flights introduced in the experimental setup: #11 ($N_{\text{obs}} = 362$), #12 ($N_{\text{obs}} = 214$), #14 ($N_{\text{obs}} = 108$), and #15 ($N_{\text{obs}} = 130$). Each flight's observations are inverted in two settings. Known- U mode fixes wind speed to the co-located meteorological station, whereas blind inversion estimates U . As the main external comparator, we use the same-shaft hourly emission inventory ($13.3 \pm 4.1 \text{ kt yr}^{-1}$). The integrated Gaussian result reported by Andersen et al. (2023) ($Q_{\text{IG}} = 9.5 \pm 3.5 \text{ kt yr}^{-1}$, $N = 13$ flights) and the mass-balance result ($Q_{\text{MB}} = 9.8 \pm 5.0 \text{ kt yr}^{-1}$, $N = 14$ flights) are supplementary method references from the same campaign. Known- U is used for the main consistency comparison, and blind inversion diagnoses wind-input accuracy and parameter identifiability. Table 3 summarizes the results; Figure 4 shows the observation geometry and posterior distributions.

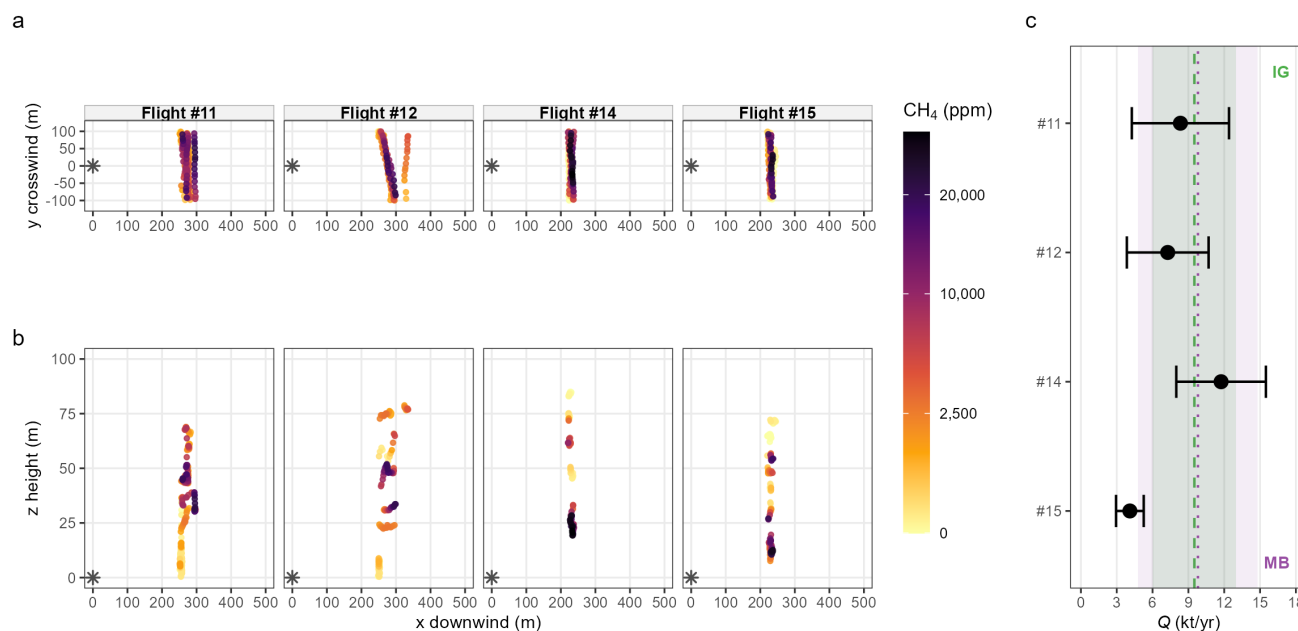


Figure 4. Illustrative real-data inversion results for the Pniówek V coal-mine ventilation shaft. (a) xy -plane and (b) xz -plane views show UAV observations for flights #11, #12, #14, and #15; color denotes measured methane enhancement. (c) Known- U posterior emission rate estimates (Q) and credible intervals. The main external comparator is the same-shaft hourly inventory ($13.3 \pm 4.1 \text{ kt yr}^{-1}$); integrated Gaussian (IG) and mass-balance (MB) estimates are supplementary references. These estimates are consistency checks under model mismatch, not validated absolute fluxes.



Table 3. Illustrative real-data inversion results for the Pniówek V ventilation shaft. The main external comparator is the same-shaft hourly inventory 13.3 ± 4.1 kt yr⁻¹. As supplementary references, Andersen et al. (2023) reported IG = 9.5 ± 3.5 kt yr⁻¹ and MB = 9.8 ± 5.0 kt yr⁻¹. Known- U is used for the main consistency comparison. Blind inversion jointly estimates Q and U to diagnose wind-speed representativeness and parameter degeneracy.

Flight	N_{obs}	U_{meteo} (m s ⁻¹)	Known U : $\bar{Q} \pm \sigma_Q$ (kt yr ⁻¹)	Blind: $\bar{Q} \pm \sigma_Q$ (kt yr ⁻¹)	Blind: $\bar{U} \pm \sigma_U$ (m s ⁻¹)
#11	362	7.0	8.33 ± 4.07	3.59 ± 0.82	2.57 ± 0.69
#12	214	7.0	7.27 ± 3.42	3.68 ± 1.02	2.55 ± 0.70
#14	108	6.0	11.74 ± 3.75	8.01 ± 23.81	1.81 ± 0.70
#15	130	6.0	4.10 ± 1.16	13.97 ± 5.87	13.00 ± 0.71

395 Flights #11 and #12 show that, with sufficient coverage, NeuPlume can infer emission rate estimates from real UAV data that are compatible with the hourly inventory. In known- U mode, #11 gives $\hat{Q} = 8.33 \pm 4.07$ kt yr⁻¹ and #12 gives $\hat{Q} = 7.27 \pm 3.42$ kt yr⁻¹. Both posterior intervals overlap the inventory range, although their centers are low. Measurement points cover downwind distances of about 252–297 m (#11) and 248–335 m (#12), with 362 and 214 observations, respectively. This spatial coverage constrains the likelihood and separates the emission rate signal from background noise. Blind inversion gives
400 lower wind speeds ($\hat{U} \approx 2.6$ m s⁻¹) and lower emission rates, indicating that Q – U coupling still shifts the posterior when wind speed is weakly constrained. Thus, known- U results are used for the main consistency comparison.

Flight #14 reflects the identifiability limit caused by limited near-source sampling. Known- U gives $\hat{Q} = 11.74 \pm 3.75$ kt yr⁻¹, with a posterior interval overlapping the hourly inventory. However, blind inversion yields a very broad emission rate posterior ($\hat{Q} = 8.01 \pm 23.81$ kt yr⁻¹), and the maximum-a-posteriori source height is only $H = 2.24$ m. This flight has the fewest in-
405 domain observations ($N_{\text{obs}} = 108$), concentrated at short downwind distances (221–238 m). Under this geometry, near-source concentration gradients allow multiple parameter combinations to produce similar concentrations at the observation locations. The main limitation is therefore parameter identifiability under limited near-source coverage.

Flight #15 shows that wind-input accuracy strongly affects inversion. Known- U gives $\hat{Q} = 4.10 \pm 1.16$ kt yr⁻¹, below the hourly inventory range. Blind inversion returns $\hat{Q} = 13.97 \pm 5.87$ kt yr⁻¹ and estimates $\hat{U} = 13.0$ m s⁻¹, above the
410 meteorological-station value $U = 6.0$ m s⁻¹. This agrees with the Q – U scaling in steady transport: the blind-mode wind speed is 2.17 times the station value, and the blind to known- U emission-rate ratio is 3.41. If station wind speed does not represent the effective transport speed at the UAV sampling height, fixed-wind inversion compensates through a lower Q .

The baseline comparison further shows that the physical latent prior prevents extreme estimates under sparse field observations, although physical model error remains unresolved. In blind inversion mode, the Gaussian plume baseline gives
415 $\hat{Q}$ values ranging from 25 to 1792 kt yr⁻¹, indicating that an unconstrained analytical plume surface can produce physically implausible concentration fields under incomplete data. Even with fixed wind speed, Gaussian plume estimates reach $\hat{Q} \in [31, 49]$ kt yr⁻¹. MB deviates in the opposite direction, with near-zero estimates ($\hat{Q} \approx 0.009$ – 0.055 kt yr⁻¹), because sparse and irregular UAV sampling does not satisfy its dense cross-section requirement. In contrast, NeuPlume known- U posterior means are $\hat{Q} \in [4.1, 11.7]$ kt yr⁻¹ across the four flights, and three posterior intervals overlap the hourly inventory (full



420 comparison in Supplementary Table S3). This comparison supports consistency under model mismatch, not absolute validation of shaft-specific emission rate.

4.3 Uncertainty quantification

After evaluating emission rate point estimates, uncertainty quantification tests whether NeuPlume’s approximate posterior intervals have reasonable empirical coverage. In point-source inversion with limited measurements, credible intervals bound the interpretation of point estimates under underdetermined observations. We evaluate calibration on 30 holdout samples independently drawn from the training distribution. Each sample uses the default temperature $\tau = 1.0$ and runs the full diffusion posterior sampling and importance-sampling workflow (Section 2.4). For each sample, the approximate posterior distributions of (H, U, I) are summarized by symmetric credible intervals at 50%, 68%, 90%, and 95% nominal levels, with coverage defined as the fraction of samples whose true parameter lies inside the interval.

430 The overall coverage indicates that NeuPlume’s posterior intervals are empirically calibrated on the current holdout set, with conservative behavior at high nominal levels. Figure 5a shows that, after pooling parameters, empirical coverage is 73.3% for the 68% interval, 97.8% for the 90% interval, and 100% for the 95% interval. These values are slightly above the nominal levels but remain close to the ideal calibration line. Parameter-level coverage identifies the source of this pattern: at the 68% nominal level, wind speed U (70.0%) and turbulence intensity I (66.7%) are close to nominal coverage, whereas release height H (83.3%) is overcovered (Figure 5b).

Coverage for wind speed U deviates further from nominal levels at narrower intervals, suggesting that the $H-U-Q$ degeneracy mainly affects the marginal posterior of wind speed. At the 50% nominal level, empirical coverage for U decreases to 26.7%. When the nominal level increases to 90% and 95%, U coverage returns to 100%. This undercoverage at narrow intervals is consistent with transport physics: wind-speed changes can be partly compensated by effective release height and source-strength adjustments, making the marginal posterior of U more sensitive to interval width. Therefore, applications should not interpret the U point estimate alone; they should report posterior intervals and, where possible, cross-check them against independent wind-speed observations.

The temperature parameter τ in the diffusion posterior sampling likelihood (Eq. 8) controls the sharpness of the approximate posterior: smaller τ concentrates samples in high-likelihood regions, whereas larger τ flattens the distribution. A fixed default temperature is useful only if coverage and effective sample size are insensitive to reasonable perturbations. We scan $\tau \in \{0.01, 0.05, 0.1, 0.5, 1.0, 5.0, 10.0\}$ over 18 parameter-case pairs (six synthetic cases $\times$ three parameters) and compute coverage at each nominal level (Figure 6a).

The temperature-sensitivity results support using $\tau = 1.0$ as the default. For $\tau \in [0.1, 10]$, both 90% and 95% coverage remain stable at 94.4%, showing that high-confidence intervals are insensitive to temperature changes across two orders of magnitude. Only at $\tau = 0.01$ do 90% and 95% coverage fall to 66.7% and 72.2%, respectively, indicating that excessive sharpening overconcentrates the posterior in high-likelihood regions. The 50% and 68% intervals are more sensitive to posterior shape and therefore vary more with τ . However, for $\tau \geq 0.05$, no collapse similar to $\tau = 0.01$ occurs. At $\tau = 1.0$, the mean effective sample size across the six cases is 70.7 out of 75 diffusion samples, indicating that the importance weights are not

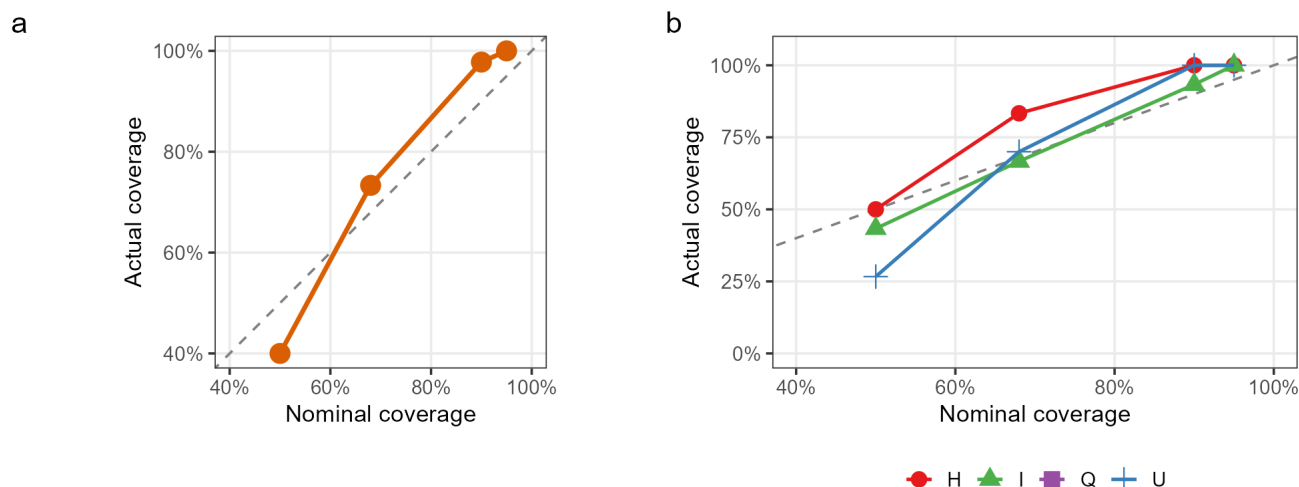


Figure 5. Uncertainty quantification calibration on 30 independent holdout cases ($\tau = 1.0$). (a) Overall empirical coverage after pooling parameters versus the nominal credible interval level. The dashed diagonal denotes ideal calibration. (b) Per-parameter coverage at the 68% nominal level: U and I are close to nominal coverage (70.0% and 66.7%), while H is overcovered (83.3%).

dominated by a few samples. With only 18 parameter-case pairs, these results support $\tau = 1.0$ as a stable default within the tested range but do not guarantee universal calibration.

Beyond coverage, the geometry of the posterior diagnoses parameter identifiability. Figure 6b and c show the joint (H, U) posteriors for two synthetic cases. Importance weights are obtained by applying the Boltzmann transform to the diffusion posterior sampling loss (Eq. 8), and the complete marginal and joint distributions for Case 1 are shown in Supplementary Fig. S1. In Case 1, a typical mid-range scenario, the joint posterior forms a compact unimodal cluster near the truth with weak $H-U$ correlation (Figure 6b).

By contrast, the high-release-height, high-wind-speed conditions of Case 4 ($H = 47.9$ m, $U = 14.1$ m s⁻¹) produce an elongated joint posterior with negative $H-U$ correlation (Figure 6c), reflecting physical degeneracy: a higher release height can partly offset lower wind speed in the downwind concentration pattern. Traditional point estimates cannot express this uncertainty structure. Because observations in Case 4 cover only 1% of the downwind cross-section, the data cannot independently constrain H and U , and the emission estimate must be interpreted together with the posterior shape. The contrast between Cases 1 and 4 illustrates how posterior geometry diagnoses the strength of observational constraints.

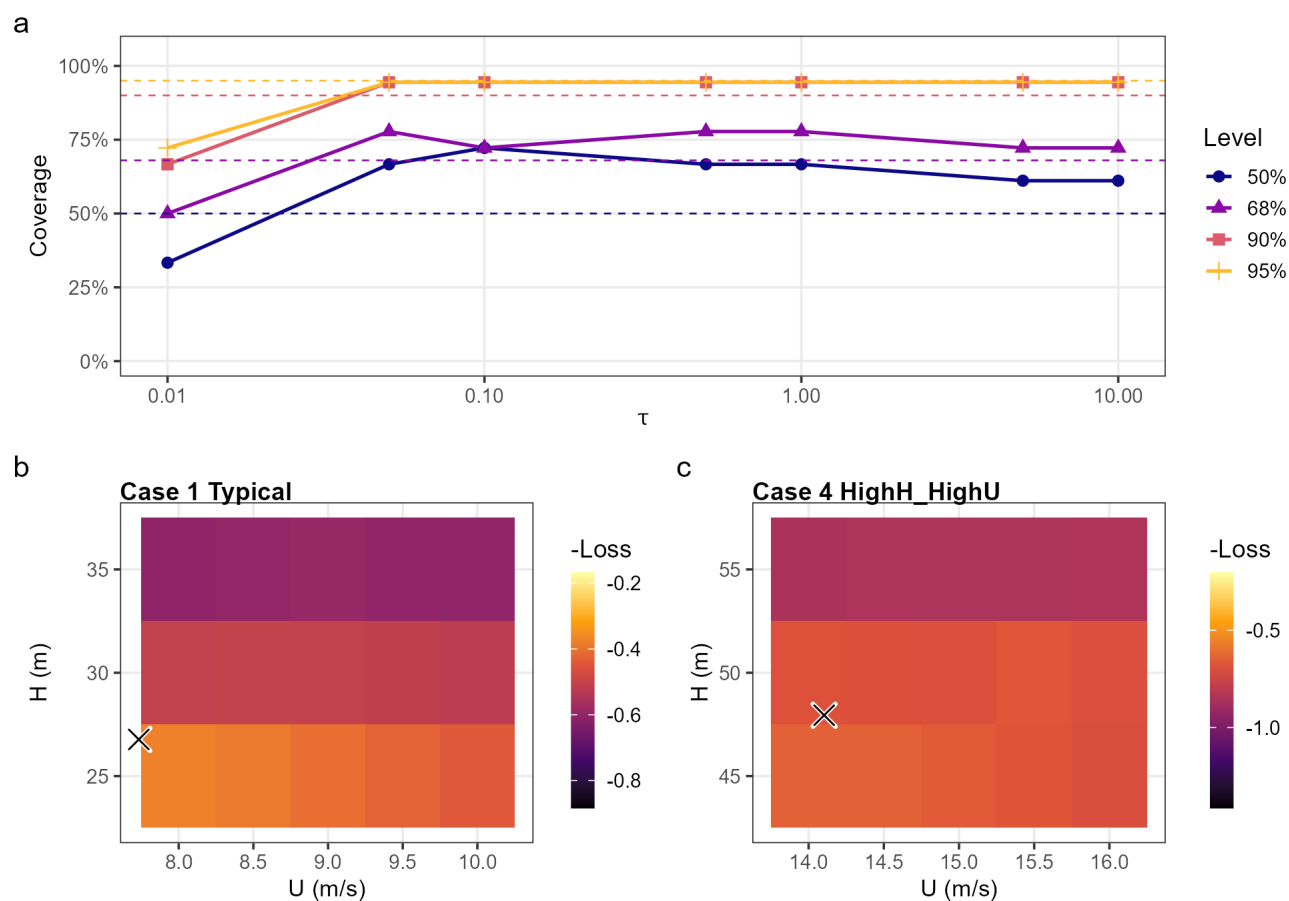


Figure 6. Approximate posterior distribution analysis. (a) Coverage sensitivity to Boltzmann temperature τ over 18 parameter-case pairs. For $\tau \in [0.1, 10]$, 90% and 95% coverage are stable at 94.4%. (b, c) Joint (H, U) posterior distributions for Case 1 (compact, low epistemic uncertainty) and Case 4 (diffuse, high epistemic uncertainty with negative $H-U$ correlation), showing different observation constraints.



4.4 Ablation study

4.4.1 Coarse-to-fine parameter search

The coarse-to-fine search described in Section 2.4 balances computational cost and parameter accuracy by first locating low-
470 loss regions on a coarse grid and then performing a fine grid search around the winning cell. To quantify the contribution of the refinement stage, we compare the full two-stage search with a coarse-only baseline across the six synthetic cases. Under coarse-only search, the mean emission rate error is $\bar{\epsilon}_Q = 25.0\%$, and the mean release-height error is $\bar{\epsilon}_H = 8.6\%$. After adding fine search, these errors decrease to $\bar{\epsilon}_Q = 10.0\%$ and $\bar{\epsilon}_H = 1.4\%$, corresponding to mean improvements of 15.0 and 7.2 percentage points.

475 Most of the refinement gain comes from improved release-height estimation. In the high- H region, the spacing of the coarse grid can reach $\Delta H = 10$ m, which is coarse relative to the sensitivity of the plume to release height. Fine search restores subgrid accuracy by locally subdividing the winning cell. Case-wise results also show that refinement is not monotonically beneficial for every metric: the Q error in Case 1 increases slightly from 13.7% to 14.2%, and that in Case 4 increases from 1.7% to 2.1% (Supplementary Table S1). However, the H error in Case 4 simultaneously decreases from 4.3% to 0.3%, and
480 the six-case mean still shows a clear benefit. Overall, the coarse-to-fine decomposition preserves the global search capability of the coarse scan while avoiding high-resolution enumeration over the entire parameter domain.

4.4.2 Observation-coverage sensitivity

The observation coverage r_{obs} is defined as the proportion of computational grid cells that contain at least one measurement point, and it determines how much information the likelihood term can use to constrain source parameters. We scan r_{obs} from
485 0.5% to 10% across the six synthetic cases while keeping all other settings unchanged (Figure 7). When coverage decreases from 1% to 0.5%, the mean emission rate error increases from 26.4% to 42.1%, a degradation of 15.7 percentage points. This jump indicates that, under extremely low coverage, the observational likelihood surface is insufficient to reliably distinguish candidate parameter combinations, making grid search more likely to select a suboptimal solution.

Once coverage exceeds 1%, the mean emission rate error enters a plateau and mainly remains within 25%–29%. Under
490 the current three-dimensional random sampling protocol, the marginal benefit of adding observation points decreases rapidly. Inversion quality is limited by observation geometry and parameter degeneracy, so the number of measurement points is not the only limiting factor. Approximately 1% coverage can serve as a practical minimum target under the current setting, below which inversion quality degrades sharply.

4.4.3 DPS observation-likelihood guidance

495 The observation-likelihood guidance ablation evaluates the marginal contribution of the gradient term in the reverse diffusion trajectory. Both configurations use the same DPS reverse diffusion process to generate conditional concentration-field samples; they differ only in whether the observation-likelihood gradient is injected into the reverse trajectory. Under the current coarse-

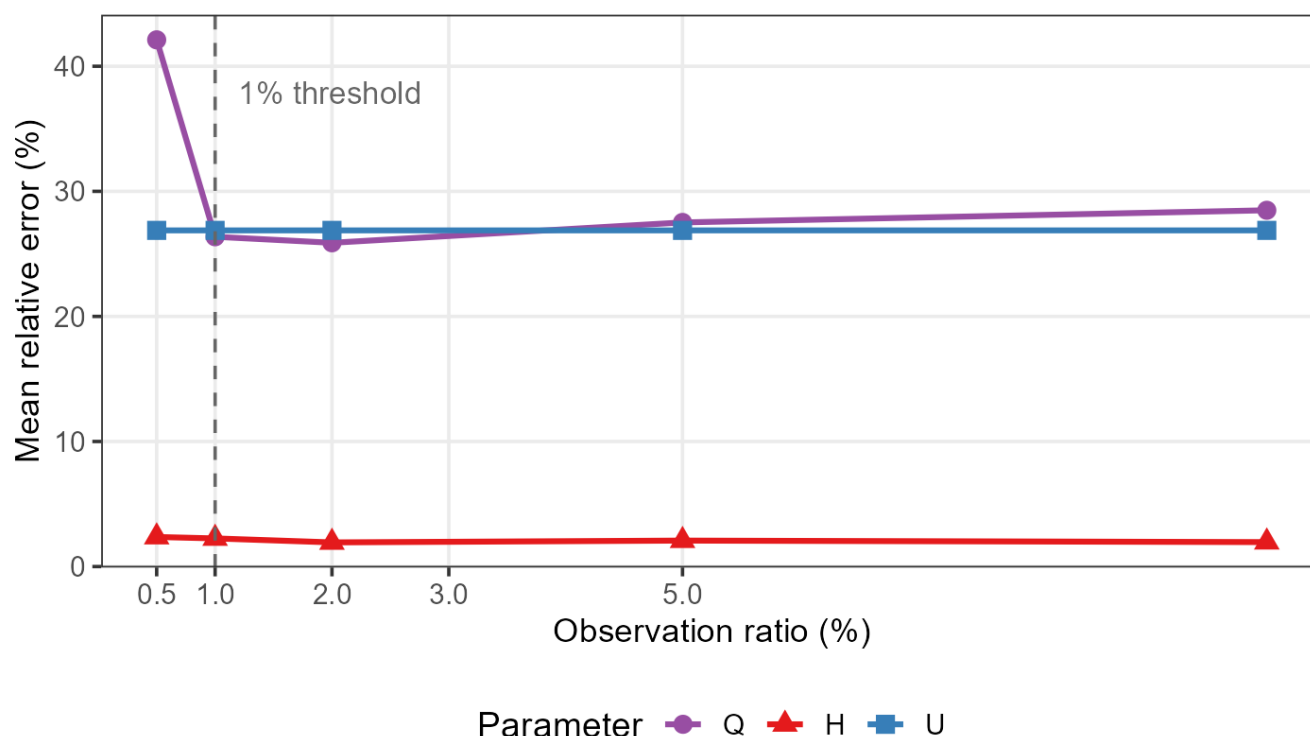


Figure 7. Sensitivity of emission rate error to spatial observation coverage r_{obs} . When coverage is below 1%, the mean Q error increases markedly (42.1% at 0.5% and 26.4% at 1%). Above 1% coverage, the curve enters a plateau, indicating limited marginal benefit from further sampling densification.

grid setting, the guided configuration has a mean Q error of 29.7%, whereas the unguided configuration has a mean error of 25.4%. The difference is approximately 4.4 percentage points, with the unguided configuration giving the lower mean error.

500 In the current implementation, the main constraint comes from explicit enumeration and loss comparison on the candidate parameter grid. The observation-likelihood gradient does not provide additional error reduction under this discrete search configuration. Subsequent improvements could prioritize continuous parameter optimization, allowing candidate parameters to update along the observation-likelihood gradient rather than only selecting among finite grid nodes.

5 Discussion

505 5.1 Wind-speed representativeness and observation-geometry error

The real-data experiments indicate that the main errors in NeuPlume arise from the joint constraint that wind-speed representativeness and observation geometry impose on the source parameters. As shown in Section 4.2, Flight #14 overlaps the



hourly inventory in known- U mode but yields an unstable blind-inversion posterior because of its short downwind distance and incomplete plume coverage. Flight #15 illustrates the complementary problem: blind mode adjusts the posterior wind speed to
 510 $\hat{U} \approx 13 \text{ m s}^{-1}$, more than twice the ground-station value, shifting the emission rate posterior along the Q - U direction. When the station wind speed does not represent the effective transport speed at the UAV sampling height, the posterior moves along this coupling direction and the resulting emission estimate can deviate substantially from the inventory range.

The effect of wind-speed representativeness error can be explained by the analytical scaling of steady-state passive-tracer transport. Because the conditional neural field (CNF) is trained at a unit emission rate ($Q = 1$), the inversion first estimates
 515 $k = Q/U_{\text{eff}}$ and then recovers the emission rate through $Q = k \cdot U_{\text{eff}}$. The emission rate is therefore linearly coupled to effective wind speed. This coupling follows from the linear response of the steady concentration field to source strength and its inverse scaling with advective speed; it does not depend on a particular Gaussian plume approximation. Previous studies show the same sensitivity. Meier et al. (2026) showed that wind speed contributed up to 99.4% of total uncertainty in cross-sectional-flux methane inversion. Eastwood et al. (2025) demonstrated that, in controlled-release experiments, using ERA5 reanalysis
 520 winds rather than directly measured plume velocity led to an emission rate underestimation of about 28%. Airborne methane detection and coal-mine point-source remote-sensing studies likewise identify wind-field information as a source of uncertainty in emission rate quantification Conrad et al. (2023); He et al. (2024). Blind mode can absorb a misspecified station wind speed, but it introduces the H - U degeneracy shown in Section 4.1.

Transport-model error further changes the interpretive range of this coupling. Model error can not only widen the posterior
 525 distribution but also move the posterior location nonuniformly in parameter space De Meutter et al. (2021). When the wind-profile assumptions of the Lagrangian stochastic model (LSM) are violated, the blind-mode posterior can still have systematic displacement.

Other deep-learning methods face the same wind-speed sensitivity. Zhao et al. (2026) combined a U-Net encoder with inverse Gaussian fitting and reduced single-source mean absolute percentage error (MAPE) from 38% to 28.5% on controlled-release
 530 data. That framework differs from NeuPlume in its target: it estimates only Q and does not jointly invert Q , H , and U . It returns point estimates rather than a full posterior, and it relies on dense cross-sectional concentration data rather than limited point measurements. Accordingly, its large-eddy simulation (LES)-validated setup covers a wider range of atmospheric stability (Pasquill A–D) and has been validated in 11 controlled-release experiments, providing a reference for validation scenarios not yet covered in this study.

535 Observation geometry is the second major error source and explains why inventory-compatible Flight #14 still gives an unstable blind-inversion posterior. This flight samples only the near-source region, with downwind distances of 221–238 m, and the limited downwind range makes the inversion depend strongly on steep concentration gradients. Under these conditions, small spatial offsets can translate into large intensity differences. The DPS likelihood surface therefore becomes flat, and multiple (Q, H, U) combinations can produce similar concentrations at the observation locations. This degeneracy is consistent with the sensitivity to observation geometry, platform configuration, and wind-field sampling documented in UAV and
 540 controlled-release studies Liu et al. (2024); Vollrath et al. (2026); Scheutz et al. (2025). Among them, Mohammadloo et al. (2025) found in 1,001 UAV flights that 55% had theoretical errors above 10% because of incomplete plume coverage. Our



observation-coverage ablation experiment (Section 4.4.2) supports this mechanism from a complementary direction: when coverage falls below 1%, the mean Q error increases from 26% to 42%. Reliable inversion requires observations at downwind distances beyond approximately 250 m and lateral coverage on both sides of the plume centerline, so that the likelihood surface retains enough curvature to distinguish candidate source parameters; this requirement is consistent with the multi-level transect design used in UAV AirCore quantification of coal-mine ventilation shafts (Andersen et al., 2021).

5.2 Physical scope, expected bias, and boundary effects

The current empirical results should be read as evidence for a specific LSM instance: passive, low-height, neutral releases over flat terrain with single-pass open-boundary transport. The release-height parameter is an effective height in this library, not a plume-rise model. If real initial momentum or buoyancy lifts the plume above the trained range, the posterior cannot represent the missing physics directly. It can only move toward the height upper bound or compensate through Q , U , and I . The height-boundary diagnostic shows this behavior at the edge of the synthetic training domain: Case 4 ($H = 47.9$ m) and Case 5 ($H = 44.8$ m) have posterior MAP heights 5 m below the 50 m upper grid boundary. In the field cases, Flight #11 is flagged in both blind and known- U modes, and Flight #12 is flagged in blind mode. These flags do not prove plume rise, but they show that part of the real-data posterior is controlled by the imposed height ceiling.

The wind-speed bias diagnostic gives the part of the expected bias that can be quantified without constructing a new plume-rise library. Because the unit-rate field represents concentration per unit source strength, fixed-wind inversion approximately scales Q with the effective transport speed. Flights #11, #12, and #14 have blind posterior wind speeds only 30–37% of the station value, and their blind emission rates are correspondingly lower than known- U estimates. Flight #15 shows the opposite pattern: blind mode estimates an effective wind speed 2.17 times the station value and an emission rate 3.41 times the known- U estimate. This relation quantifies the wind-speed component of expected bias; it does not quantify the full error caused by buoyancy, initial momentum, or stability mismatch.

Open-boundary effects define a second physical boundary. The boundary-loss diagnostics in Supplementary Table S4 show that most particle exits occur through the downwind boundary, consistent with open-domain outflow under unidirectional transport. Side and top losses are smaller but non-negligible in restrictive weak-wind or high-turbulence cases, indicating finite-domain truncation of plume mass. Because particles that leave the domain are deleted, the present LSM cannot represent recirculating meteorology, building wakes, or complex-terrain flows in which material exits and later re-enters the observation region.

These diagnostics support a bounded interpretation of the real-data application. The Pniówek results show that the learned physical prior can keep sparse-field estimates within the same order of magnitude as the inventory under model mismatch, but they do not validate absolute flux accuracy for plume-rise-dominated ventilation shafts. Extending NeuPlume to that regime requires regenerating the physical library with initial momentum, buoyancy, non-neutral stability, and local-flow effects represented in the forward simulator, then retraining the CNF and diffusion prior before inference.



Table 4. Applicability guidance for the current NeuPlume implementation. The entries distinguish the physical scope of the present LSM training library from the broader framework, which can be re-instantiated with a different scenario-specific forward simulator.

Condition	Current coverage	Manuscript guidance
Passive tracer without initial momentum	covered	use current instance
No buoyant plume rise	passive assumption only	treat H as effective height
Effective release height $H \leq 50$ m	covered	do not extrapolate trained weights above 50 m
Neutral stability and flat terrain	covered	apply within current LSM scope
Moderate-to-strong unidirectional wind	$U_{\text{ref}} = 1\text{--}15 \text{ m s}^{-1}$	require representative wind or infer U
Stable or unstable boundary layer	not validated	rebuild the forward library
Recirculating flow or building wake	not covered	current open boundary removes re-entry
Complex terrain	not covered	use scenario-specific LSM/LES/CFD
Time-varying emissions or multiple sources	not covered	expand the inverse parameterization

575 **5.3 Steady-state and atmospheric-stability assumptions**

The neutral atmospheric-stability assumption defines the current scope of the framework. All LSM training simulations use a neutral wind profile ($\alpha = 1/7$), so the current results mainly apply to near-field point-source dispersion under moderate-to-strong wind conditions. Stability changes near-surface wind shear and vertical mixing: stable conditions generally correspond to a larger wind-shear exponent and weaker vertical mixing, whereas unstable conditions generally correspond to a smaller wind-shear exponent and stronger vertical exchange Newman and Klein (2014). In a Lagrangian stochastic model, non-neutral boundary layers also change the turbulent-velocity distribution, drift term, and implementation of the well-mixed condition Cassiani et al. (2014). Because the CNF has not been trained under these states, its reconstruction fidelity under non-neutral conditions has not been verified. Incorporating stability as a conditional parameter in the training library is a necessary step for extending the framework to a wider range of atmospheric states.

585 The steady-state emission assumption also limits the application boundary of NeuPlume. The current framework applies to continuous point sources such as ventilation shafts, pipeline leaks, and industrial stacks, but it does not directly apply to intermittent or time-varying releases. Bayesian time-varying source-term estimation methods, such as the Bayesian source-term estimation of Xu et al. (2025), can handle temporal variability by assimilating sequential observations. Such methods require dense temporal coverage and explicit turbulence parameterization. A later extension of NeuPlume could use a time-
590 conditioned diffusion model so that the latent space prior represents time-varying concentration fields, but this direction still requires separate implementation and validation.



5.4 Computational cost and scalability

NeuPlume's scalability comes from separating the physical forward model from the probabilistic inversion engine. The current implementation uses the LSM to build a parameter-to-concentration-field mapping for a single point source, neutral stability, and flat terrain. The key requirement is not a particular analytical formula or simulator, but a forward model that can generate parameterized concentration-field samples for the target scenario. Those samples can then define the field prior and, when limited measurements are available, support posterior inference of the control parameters. Multi-source emissions, complex terrain, building effects, or higher-fidelity LES/CFD models can all be added as physical front ends along this route; high-resolution LES-based methane plume simulations illustrate one such route for building scenario-specific forward libraries when simplified transport assumptions are insufficient (Jiang et al., 2025). The corresponding posterior object can also be extended from single-source (H, U, I, Q) to source locations and emission strengths for multiple sources, terrain or boundary-layer parameters, and the full concentration field.

This extension direction has a clear computational cost. Whenever the scene geometry changes, for example from flat to complex terrain or from a single-source to a multi-source configuration, the training data must be regenerated and the CNF must be retrained, because the learned concentration manifold is specific to the training parameter space. In the current implementation, the LSM ensemble and CNF training require about 50 CPU-h and 116 GPU-h, respectively. Diffusion-model training requires about 4 GPU-h, and one coarse-to-fine inversion requires about 3 GPU-h; the computational-cost breakdown is given in Appendix Table C1. These orders of magnitude make NeuPlume most appropriate for deployments in which an offline simulation library can be reused, rather than for cases that require rebuilding the full training pipeline for each new geometry.

Increasing parameter dimension further limits the current coarse-to-fine grid search strategy. If multi-source locations or time-varying emission profiles (>6 degrees of freedom) are jointly estimated, discrete grid enumeration will face combinatorial growth. The current ablation results show that, under the discrete coarse grid, the DPS observation-likelihood gradient provides no measurable benefit and even slightly increases the mean error (Section 4.4.3). This result indicates that the effective information in the present configuration mainly comes from grid enumeration. Gradient-driven continuous parameter optimization still needs separate implementation and validation. In atmospheric inversion, 4D-Var and adjoint models have been used for emission rate correction, chemical-state estimation, and high-resolution CO_2 flux inversion Elbern et al. (2007); Maksyutov et al. (2021). A more scalable approach would update source terms and meteorological parameters on a continuous parameter manifold with the observation-likelihood gradient, rather than continuing to enlarge the discrete grid. A complementary route is amortized inference: Liu et al. (2025b) demonstrated a two-stage conditional diffusion framework for partial-differential-equation (PDE) constrained inverse problems, with inference 60 times faster than DPS while maintaining an empirical coverage rate of 94.35% for the 95% credible interval. A similar strategy, if adapted to atmospheric dispersion inversion, would substantially reduce per-case computational cost.



5.5 Method positioning and future validation

NeuPlume is positioned as a complement to existing source-term estimation methods, not as a replacement for all traditional methods. In far-field settings with a homogeneous wind field where the Gaussian plume assumption holds, integrated Gaussian and MB approaches remain simpler and faster and have been widely used Shah et al. (2019); Krings et al. (2013); Conley et al. (2017). NeuPlume's comparative advantage appears in three scenarios: near-field sampling dominated by non-Gaussian structure, cases that require joint estimation of multiple source and environmental parameters, and settings that need posterior uncertainty quantification for regulatory or decision support. In these settings, the posterior distribution can state which explanations are supported by the observations and which cannot be excluded, which is closer to practical decision needs than a single optimum. This methodological complementarity is consistent with Aghdasi et al. (2025), who found that variational assimilation can strongly constrain point-source emissions under in situ observations but has diminishing benefits for diffuse area sources. A recent controlled-release evaluation with fence-line sensors likewise shows that Gaussian plume inversion and backward Lagrangian stochastic methods are sensitive to wind-sector width, averaging window, and multi-source interference Mbua et al. (2025). Controlled-release assessments further show that source type, observation geometry, and platform configuration change the error structure of quantification methods Liu et al. (2024). Different source types and observation geometries call for different inversion strategies. Because the current real dataset has no certified true emission rate for individual flights, cross-method comparison cannot replace absolute-accuracy validation. Validating NeuPlume against controlled-release experiments, the established benchmark for method comparison Morales et al. (2022); Vollrath et al. (2026), is the next step. Combining UAV, ground-based, and satellite observations will further establish the framework's operational accuracy under field conditions.

6 Conclusions

This paper presents NeuPlume, a probabilistic inversion framework for point-source emissions under limited observations. It uses Lagrangian stochastic model (LSM) forward simulations to construct parameterized concentration fields for the target scenario, learns a field prior with a conditional neural field (CNF) and a latent space diffusion model, and uses diffusion posterior sampling (DPS) to infer source parameters, environmental parameters, and the complete concentration field under observational constraints. Synthetic experiments, real UAV observations, and uncertainty assessment support three conclusions.

First, across six synthetic cases under neutral atmospheric stability that cover different source heights, wind speeds, and turbulence intensities, NeuPlume achieves mean errors of 10.0%, 1.4%, and 6.1% for emission rate Q , effective release height H , and turbulence intensity I , respectively. Under the same volumetric point-observation setting, the mean Q errors of blind GP inversion and MB are 2482% and 912%. This gap indicates that the non-Gaussian concentration representation learned by the CNF provides stronger physical constraints for near-field point-source inversion than simplified dispersion assumptions.

Second, in the real UAV data from the Pniówek coal-mine ventilation shaft, NeuPlume gives posterior mean emission rates of 4.1–11.7 kt yr⁻¹ in known- U mode. Three of the four posterior intervals overlap the hourly inventory, and the remaining low estimate is associated with wind-speed representativeness error. These results support illustrative transfer from synthetic



training data to real UAV observations under model mismatch, but they do not validate absolute flux accuracy for the ventilation shaft.

660 Third, the posterior intervals achieve empirical coverage close to nominal levels on the holdout set. Among 30 holdout samples, the empirical coverage of the 68% credible interval is 73.3%, and the coverages of the 90% and 95% intervals are 97.8% and 100%, respectively. These results show that, beyond returning point estimates, NeuPlume provides posterior uncertainty estimates that reflect parameter degeneracy and observational constraint strength.

665 The ablation experiments further define the key design conditions of the current implementation. The coarse-to-fine parameter search reduces the mean Q error from 25.0% to 10.0%. When observation coverage falls below approximately 1%, inversion quality clearly decreases. These results indicate that current performance depends on local parameter refinement and sufficient spatial observational constraints.

670 These results also define the scope of applicability and directions for extending NeuPlume. The empirical results remain limited to the LSM training library used here: passive, low-height, neutral releases over flat terrain with open boundaries. Substantially different source configurations, plume-rise physics, terrain, or atmospheric stability conditions require reconstructing the training data and retraining before inference. At the framework level, however, the physical front end and the probabilistic inverter are modular. Multiple point sources, complex terrain, non-neutral boundary layers, or higher-fidelity simulators such as LES/CFD can enter the same posterior inference workflow by redefining the forward mapping from parameters to concentration fields and learning a new scenario-specific field prior. The current discrete grid search is suitable for the low-dimensional parameter space considered here, but not for direct extension to higher-dimensional settings such as multiple sources or time-varying emissions.

Future work should prioritize continuous parameter optimization or amortized inference strategies to replace high-dimensional grid enumeration. The framework should also be validated in controlled-release experiments with known emission rates to establish its operational accuracy and calibration under field conditions.

Appendix A: Comparison between Hash-INR and SIREN architectures

680 During development of the conditional neural field component, we evaluated two candidate architectures: SIREN, which uses periodic sinusoidal activation functions, and the Hash-INR described in Section 2.3. Both architectures were trained on the same data with the same optimizer settings and loss function. For SIREN, we further compared two spatial sampling strategies: uniform random sampling and adaptive sampling that oversamples high-concentration regions, to determine whether training instability was driven by the sampling scheme.

685 Figure A1 shows the training loss curves for all three configurations.

Both SIREN variants showed initial loss reductions during the first 7 training epochs, reaching minimum values of 0.30 (uniform sampling) and 0.34 (adaptive sampling), respectively. After epoch 7, the losses of both configurations began to increase and continued to deteriorate during subsequent training. The loss of the uniform-sampling SIREN rose from 0.30 at epoch 7 to 1.94 at epoch 67. The adaptive-sampling variant showed the same qualitative failure pattern, reaching a loss of

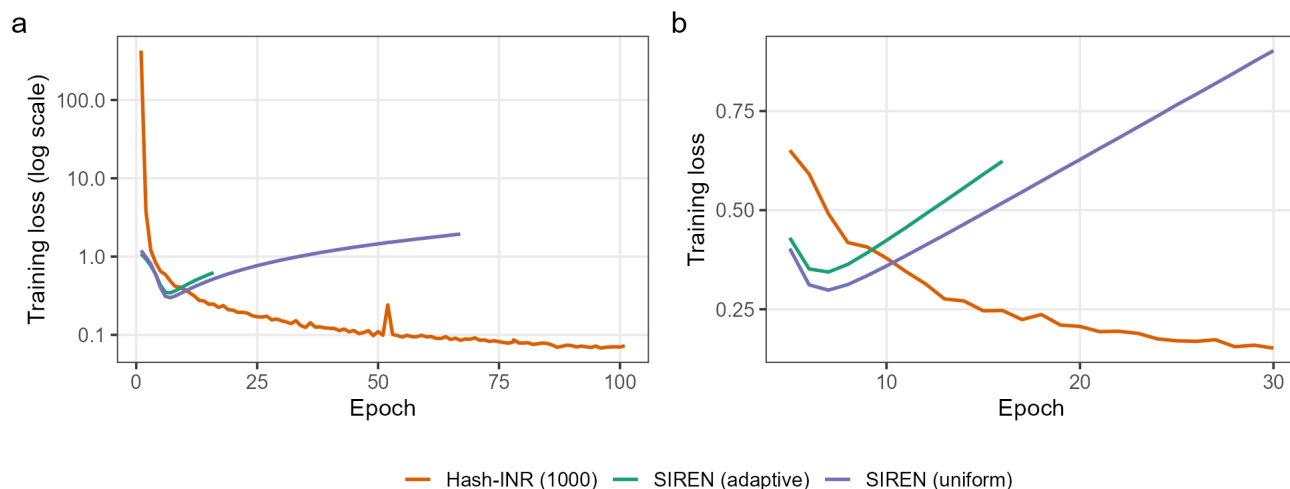


Figure A1. Training loss curves for the three CNF architectures. Both SIREN variants (uniform sampling and adaptive sampling) became unstable after training epoch 7, whereas Hash-INR converged stably to a final loss of 0.07, below the best loss achieved by the SIREN variants.

690 0.62 at epoch 16 before training was terminated. Neither sampling strategy eliminated the instability, indicating that, under this training setup, the convergence problem of SIREN was not mainly caused by the sampling distribution.

By contrast, Hash-INR trained under all 1000 conditions converged stably over 100 training epochs, with the loss decreasing from 427.09 at epoch 1 to 0.07 at epoch 100. Its final loss was lower than the best value reached by either SIREN variant and also markedly lower than the final loss of the uniform-sampling SIREN. Based on this convergence stability and reconstruction
695 accuracy, NeuPlume uses Hash-INR as the conditional neural field component.

Appendix B: Inversion hyperparameters

Table B1 lists hyperparameters for the three-stage DPS inversion workflow in Section 2.4. The coarse-to-fine grid search defines discrete parameter spaces for Stages 1 and 2; DPS sampling and emission rate optimization control the Stage 3 reverse diffusion process and learnable scaling factor. Boltzmann temperature controls pseudo-posterior sharpness from the grid-search
700 loss surface. The synthetic experiments (Table 2) use the fixed settings below. The real-data experiments (Table 3) use the same DPS and optimization settings, but search I_{ref} on a discrete grid and set $n_{\text{fine}} = 3$.



Table B1. Hyperparameters of the three-stage DPS inversion workflow.

Parameter	Value	Description
<i>Stage 1: coarse grid search</i>		
H grid	2, 5, 10, 15, 20, 30, 40, 50 m	Stack height candidates
U_{ref} grid	0.8–15.0 m s^{-1} (11 levels)	Wind speed candidates
I_{ref}	Optimized by DPS	Turbulence intensity (continuous)
Samples per point (n_{coarse})	1	Coarse-grid DPS runs
<i>Stage 2: fine grid search</i>		
ΔH offsets	$\{-5, 0, +5\}$ m	Fine grid around coarse-grid posterior maximum
ΔU_{ref} offsets	$\{-1.0, -0.5, 0, +0.5, +1.0\}$ m s^{-1}	Fine grid around coarse-grid posterior maximum
ΔI_{ref} offsets	$\{-0.10, -0.05, 0, +0.05, +0.10\}$	Fine grid around coarse-grid posterior maximum
Samples per point (n_{fine})	1	Fine-grid DPS runs
<i>DPS sampling (Stages 1–3)</i>		
Diffusion steps T	1000	Reverse trajectory
Noise schedule	Linear, $\beta \in [10^{-4}, 2 \times 10^{-2}]$	Variance schedule
Guidance scale s	2.5	Classifier-free guidance weight
Measurement scale ζ	0.1	Observation guidance weight (Eq. 6)
Parameter learning rate	0.01	Adam rate for θ , Stages 1–2
Warm-up steps	300	Pre-optimization steps
Loss function	Smooth L_1	Measurement fidelity term
<i>Stage 3: emission rate optimization</i>		
k learning rate	5×10^{-4}	Adam rate for $\log k$
k warm-up ratio	0.0	Pre-update fraction of T
k mixing weight	0.7	$Q = 0.7k_{\text{robust}} + 0.3k_{\text{LS}}$
<i>Uncertainty quantification</i>		
Boltzmann temperature τ	1.0	Pseudo-posterior sharpness



Appendix C: Computational cost

Table C1 reports the order of computational cost for the current implementation. The offline training cost mainly comes from LSM data generation and CNF training. Diffusion model training and single-case inversion are less expensive, indicating that the computational burden is concentrated mainly in the construction of the reusable forward model library.

Table C1. Computational cost breakdown (hardware: NVIDIA RTX 4080 SUPER, 16 GB VRAM).

Stage	Task	Time	Hardware
<i>Offline training (one-time)</i>			
1	LSM data generation	~50 CPU-h	CPU
2	Hash implicit neural representation (CNF) training	~116 GPU-h	GPU
3	Diffusion model training	~4 GPU-h	GPU
<i>Online deployment (per case)</i>			
4	Inversion (coarse-to-fine search)	~3 GPU-h	GPU

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.

Financial support. This work was supported by the National Natural Science Foundation of China (Grant U25D9004 and 42171464), the Fundamental Research Funds for the Central Universities (ZNJC202415 and 2042026kf0075), the National Key R&D Program of China (2024YFC3015600), and the Science and Technology Program of Hubei Provincial (2025BEB017).

Acknowledgements. The numerical calculations reported in this paper were performed on the supercomputing system at the Supercomputing Center of Wuhan University. Figure 1 was prepared as an AI-assisted schematic; an initial draft was generated using the gemini-3-pro-image-preview model, and the authors subsequently reviewed and manually revised the layout, icons, labels, and scientific content. The authors take full responsibility for the final figure.



References

- Aghdasi, S., Rayner, P., Deutscher, N., and Silver, J.: Exploring uncertainty reduction in high-resolution methane emissions in
720 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.
- Andersen, T., Vinkovic, K., de Vries, M., Kers, B., Necki, J., Swolkien, J., Roiger, A., Peters, W., and Chen, H.: Quantifying methane
emissions from coal mining ventilation shafts using an unmanned aerial vehicle (UAV)-based active AirCore system, *Atmospheric Envi-
ronment: X*, 12, 100 135, <https://doi.org/10.1016/j.aeaoa.2021.100135>, 2021.
- 725 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
730 Meteorology*, <https://doi.org/10.1007/s10546-014-9976-5>, 2014.
- 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., Faloona, 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
735 gases, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-10-3345-2017>, 2017.
- 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,
740 *GEOSCIENTIFIC MODEL DEVELOPMENT*, <https://doi.org/10.5194/gmd-14-1237-2021>, 2021.
- 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
745 Sciences of the United States of America*, <https://doi.org/10.1073/pnas.2507350122>, 2025.
- Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion,
Atmospheric Chemistry and Physics, <https://doi.org/10.5194/acp-7-3749-2007>, 2007.
- 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.
- 750 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.
- 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.



- Jiang, Y., Zhang, L., Zhang, X., Cao, X., Dou, H., Zhang, L., Yan, H., Wang, Y., Si, Y., and Chen, B.: Instrument Performance Analysis for Methane Point Source Retrieval and Estimation Using Remote Sensing Technique, *Remote Sensing*, 17, 634, <https://doi.org/10.3390/rs17040634>, 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.
- 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.
- 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., 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.
- Maksyutov, S., Oda, T., Saito, M., Janardanan, R., Belikov, D., Kaiser, J. W., Zhuravlev, R., and Ganshin, A.: Technical note: A high-resolution inverse modelling technique for estimating surface CO₂ fluxes based on the NIES-TM-FLEXPART coupled transport model and its adjoint, *Atmospheric Chemistry and Physics*, <https://doi.org/10.5194/acp-21-1245-2021>, 2021.
- Mbua, M., Riddick, S., Kiplimo, E., Shonkwiler, K., Hodshire, A., and Zimmerle, D.: Evaluating the feasibility of using downwind methods to quantify point source oil and gas emissions using continuously monitoring fence-line sensors, *ATMOSPHERIC MEASUREMENT TECHNIQUES*, <https://doi.org/10.5194/amt-18-5687-2025>, 2025.
- Meier, S., Vöggtli, 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.



- 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.
- 795 Morales, R., Ravelid, J., Vinkovic, K., Korbe, P., Tuzson, B., Emmenegger, L., Chen, H., Schmidt, M., Humbel, S., and Brunner, D.: 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.
- 800 Scheutz, C., Knudsen, J., Vechi, 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.
- 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.
- 805 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*, <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.
- 810 Thomson, D. J.: Criteria for the Selection of Stochastic Models of Particle Trajectories in Turbulent Flows, *Journal of Fluid Mechanics*, <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.
- 815 Vollrath, C., Barchyn, T., Munn, A., Wearmouth, C., and Hugenholtz, C.: A human-portable mass flux method for methane 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*, <https://doi.org/10.5194/acp-19-4345-2019>, 2019.
- 820 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 MATERIALS*, <https://doi.org/10.1016/j.jhazmat.2025.138171>, 2025.
- 825 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.



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