Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Physics-Informed Surrogate Model for Rapid Prediction and Source Inference of Pollutant Dispersion in Complex Urban Environments

Zeyu Li
Tongji University
Suzhen Li
lszh@tongji.edu.cn
Tongji University

Research Article

Keywords: urban dispersion, environmental contamination transport, surrogate modeling, source inference, urban morphology, Bayesian inversion

Posted Date: May 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9434529/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Physics-Informed Surrogate Model for Rapid Prediction and Source Inference of Pollutant Dispersion in Complex Urban Environments

Zeyu Li[a], Suzhen Li[a,b]*

[a] Tongji University, College of Civil Engineering, Siping 1239, Shanghai, China, 200092;

[b] Tongji University, State Key Laboratory of Disaster Reduction in Civil Engineering, Siping 1239, Shanghai, China, 200092;

Abstract: Pollutant transport in complex urban environments is strongly shaped by building-induced flow distortion, channeling, recirculation, and geometrical blockage, which together produce highly heterogeneous concentration fields. Although high-fidelity numerical simulations can capture these effects, their computational cost limits rapid prediction and source inference. This study proposes a physics-informed surrogate modeling framework for rapid prediction and source inference of ground-level pollutant dispersion in complex urban environments under quasi-steady wind conditions. Four source-dependent flow-relative features are introduced to encode transport direction and geometry-constrained accessibility, namely along-wind distance, crosswind distance, line-of-sight indicator, and shortest-path distance. These features are used as inputs to deep-learning surrogate models, with DeepLabV3+ adopted as the primary architecture and compared against U-Net, FNO2d, and Pix2Pix. Experiments based on simulation data from eight realistic urban blocks in Paris and Grenoble show that the proposed feature encoding consistently improves cross-layout prediction, yielding R^2 (log) values of 0.72 – 0.85 across the tested architectures; eight-fold leave-

one-block-out validation gives an average R^2 (log) of 0.80. When coupled with a response-library-based Bayesian inference procedure, the framework achieves an approximately 96% finding rate, compared with about 46% for a Gaussian-plume baseline. Response-library construction for a new block requires less than 1 min, while each online inference step takes about 1 ms. Additional validation against ANSYS Fluent cases gives high R^2 (log) values while maintaining successful source finding in both cases. The results indicate that physics-informed surrogate modeling provides an efficient route for rapid prediction and source inference of urban pollutant dispersion in complex built environments.

Keywords: urban dispersion; environmental contamination transport; surrogate modeling; source inference; urban morphology; Bayesian inversion

Address all correspondence to: Suzhen Li, Tongji University, College of Civil Engineering and State Key Laboratory of Disaster Reduction in Civil Engineering, Siping 1239, Shanghai, China, 200092; E-mail: lszh@tongji.edu.cn

1. Introduction

Pollutant transport and dispersion in complex urban environments are strongly influenced by buildings, which modify local flow fields through blockage, channeling, wake formation, and recirculation. As a result, concentration fields in built environments are often highly heterogeneous and may differ substantially from the patterns predicted by simplified open-terrain formulations [1 – 3]. Understanding and modeling these building-modified transport processes is important for rapid prediction, exposure assessment, and source inference in urban areas.

In practical source-inference applications, concentration observations are commonly combined with mobile sensing, sequential inference, and search strategies [4 – 9]. However, in complex obstructed urban environments, the major bottleneck is often the forward model used to represent pollutant transport. Gaussian plume models [10] are attractive because of their simplicity and speed, but their symmetry assumptions cannot adequately represent obstruction-induced transport and street-canyon effects in realistic urban settings [11,12]. Existing source term estimation studies have shown that higher-fidelity forward models generally improve inference accuracy in complex urban environments [13 – 17]. Related advances have also been reported in sensor configuration design [18,19], data-driven source term estimation for hazardous atmospheric releases in complex chemical industrial parks [20], and federated-learning-based atmospheric source term estimation in urban environments [21]. Nevertheless, rapid source inference in previously unseen obstructed layouts remains computationally demanding because practical deployment requires evaluating a large number of candidate sources at near-real-time speed.

High-fidelity numerical approaches can, in principle, provide more accurate forward predictions and source inference. Building-resolving computational fluid dynamics (CFD), large-eddy simulation, adjoint-based inversion, and reduced-order methods have all demonstrated value for source estimation in complex environments [13 – 16,22,23]. However, such methods are generally too expensive when the source location is unknown and when concentration responses for thousands of candidate source locations must be assessed repeatedly during inference. Sampling-based

Bayesian methods such as Markov chain Monte Carlo and sequential Monte Carlo are also often too costly for rapid deployment when coupled with high-fidelity forward solvers [17,24]. As a result, there remains a practical gap between physically realistic urban dispersion modeling and computationally efficient source inference.

Deep-learning surrogates provide a promising route to bridge this gap by approximating high-dimensional transport mappings at much lower online cost. In related environmental and engineering applications, deep learning has shown strong performance in urban wind-field prediction and flow-related surrogate modeling, including U-Net, conditional generative adversarial networks, diffusion models, Fourier neural operators, graph neural networks, Transformers, and physics-informed neural networks [25 – 34]. For pollutant-dispersion prediction, data-driven models have demonstrated layout-adaptive prediction and cross-geometry generalization [35 – 38]. More directly relevant to source inference, recent studies have combined deep learning and CFD for hazardous-material release localization in outdoor chemical plants [39], investigated optimization- and inference-oriented source estimation strategies [40 – 42], and applied machine-learning-based methods to air pollution source identification [43]. Despite these advances, existing approaches still face several limitations. Many require geometry-specific retraining or scenario-specific optimization, while others do not explicitly encode source-position-dependent physical structure. Consequently, the generalization of surrogate-based source inference to previously unseen obstructed environments remains insufficiently addressed.

A key challenge is therefore not only how to construct an accurate surrogate model,

but also how to design an input representation that transfers across layouts while preserving source-dependent transport information. Conventional surrogate inputs, such as building masks, signed distance functions, wind scalars, and source-location coordinates, do not explicitly encode the physical relationship between a candidate source and the surrounding flow-accessible domain. This may limit generalization when the model is applied to new geometries. For rapid source inference, however, the forward model must distinguish among many candidate sources in a computationally efficient manner while preserving the transport structure that governs concentration patterns. Embedding such source-dependent physical priors directly into the model input may therefore improve both predictive fidelity and downstream source-inference performance.

To address this problem, this study proposes a physics-informed surrogate modeling framework for rapid prediction and source inference of a single ground-level pollutant release in a known built environment under quasi-steady wind conditions. Four source-dependent features are introduced to encode wind-relative transport structure and building-constrained accessibility, namely along-wind distance (AWD), crosswind distance (CWD), line-of-sight indicator (LOS), and shortest-path distance (SPD). Based on these features, a three-stage computational framework is developed that separates offline surrogate training, response-library construction for a previously unseen block, and Bayesian source inference through millisecond-level table lookup. In this way, the main computational burden is shifted away from the inference stage while maintaining physically informed concentration prediction.

The main contributions of this work are threefold. First, we propose a source-dependent physics-informed feature encoding that explicitly incorporates transport-direction and geometry-constrained accessibility priors into deep-learning-based dispersion prediction. Second, we develop a computational framework that integrates surrogate prediction, response-library construction, and Bayesian source inference into a unified and computationally efficient workflow. Third, through systematic experiments across eight realistic blocks in two cities, we quantify the effects of feature encoding, training-data proportion, and forward-model fidelity on both concentration prediction and source-inference performance. The study is intended to contribute to computational methods for pollutant-dispersion modeling and source inference in complex urban environments.

The remainder of this paper is organized as follows. Section 2 presents the methodology, including the mathematical formulation, physics-informed feature encoding, surrogate architecture, response-library construction, and Bayesian source-inference strategy. Section 3 describes the dataset and experimental design. Section 4 reports the results for forward prediction, source inference, and external-simulator validation, and discusses the limitations of the current framework. Section 5 concludes the paper.

2. Methodology

This study considers rapid prediction and source inference of a ground-level pollutant release in a known urban block under quasi-steady wind conditions. The building layout and local wind conditions are assumed known, and concentration

observations are collected sequentially after the release has been detected. The objective is to infer the source location efficiently while preserving physically meaningful representation of pollutant transport in a complex built environment. In this study, the specific release considered is a ground-level gas release treated as a representative atmospheric pollutant release scenario.

Figure 1 illustrates the proposed three-stage framework. In Stage 1, the deep-learning surrogate is trained using simulation data from multiple urban blocks. In Stage 2, a response library is exhaustively constructed for a previously unseen block using the trained surrogate. In Stage 3, Bayesian source inference is performed through table lookup based on the precomputed response library. In this way, most computationally intensive forward evaluations are shifted away from the inference stage: response-library construction for a new block takes approximately 1 min, and each inference step is completed in milliseconds.

The remainder of this section is organized as follows. Section 2.1 presents the mathematical formulation of dispersion prediction and source inference and introduces the notation used throughout the paper. Section 2.2 describes the physics-informed feature encoding designed for the surrogate model. Section 2.3 introduces the surrogate architecture and training procedure. Section 2.4 explains how the trained surrogate is used to construct the response library. Section 2.5 presents the Bayesian source-inference procedure based on the response library. Section 2.6 defines the evaluation metrics used for both forward prediction and source-inference performance.

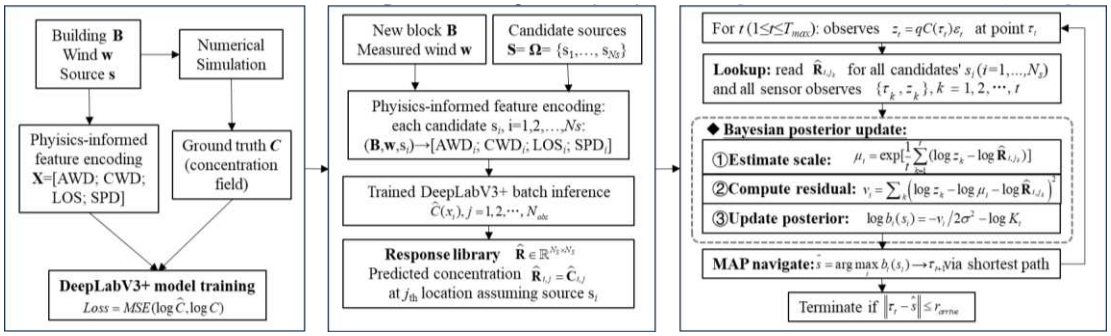


Figure 1. Methodological framework.

2.1 Mathematical formulation of dispersion prediction and source inference

An urban block is represented as a binary raster map $\mathbf{B} \in \{0,1\}^{W \times H}$ with pixel resolution Δx (m/pixel). For any pixel coordinate (x,y), $\mathbf{B}(x,y)$=1 denotes building occupancy, and $\mathbf{B}(x,y)$=0 denotes open space in which fluid can pass. The set of all fluid-domain pixels is denoted by Ω, which represents the effective space for pollutant dispersion, and its cardinality is Ns. Meteorological conditions are described by the wind direction angle θ (clockwise from north, in radians) and wind speed u (m/s), jointly denoted as $\mathbf{w} = (\theta, u)$. Accordingly, the present study focuses on single source-inference within a known urban block after detection has already occurred, rather than global exploration before first contact.

An urban block is represented as a binary raster map with pixel resolution (m/pixel). For any pixel coordinate (x, y), B(x, y) = 1 denotes building occupancy, and B(x, y) = 0 denotes open space in which fluid can pass. The set of all fluid-domain pixels is denoted by S, which represents the effective space for pollutant transport and dispersion, and its cardinality is Ns. Meteorological conditions are described by the wind direction angle θ (clockwise from north, in radians) and wind speed u (m/s), jointly denoted as w. The present study focuses on single-source inference within a

known urban block after detection has already occurred, rather than global exploration before first contact.

The true source is located at an unknown position $s^* \in \Omega$ with unknown release rate q^* (kg/s). Every fluid-domain pixel is treated as a candidate source, so the candidate source set equals the full fluid-domain pixel set, $S = \{s_1, s_2, \ldots, s_{Ns}\} = \mathbf{\Omega}$ and $s_i = (x_{s,i}, y_{s,i}) \in \Omega$, where i is the candidate source index and $(x_{s,i}, y_{s,i})$ its pixel coordinates with $|\mathbf{S}| = Ns$, preserving full spatial resolution of the posterior distribution.

Starting from an initial position, concentration measurements are collected sequentially until either reaching the source. The searcher position at step t is denoted by $\tau_t \in \Omega$, and the corresponding concentration observation is z_t (kg/m³). Each observation is modeled as the true concentration corrupted by multiplicative noise, which becomes additive Gaussian noise in log-space, as detailed in Section 2.5.

Given building layout **B**, meteorological conditions **w**, source position s^*, and release rate q^*, the steady-state concentration field satisfies the linear relationship:

$$C^*(x, y) = q^* \cdot C(x, y; \mathbf{B}, \mathbf{w}, s^*) \tag{1}$$

where $C^*(x, y)$ is the true concentration field, and $C(x, y; \mathbf{B}, \mathbf{w}, s^*)$ is the concentration response function under unit release rate. This linear homogeneity allows source-inference based on spatial pattern matching even when q^* is unknown. Urban concentration fields span multiple orders of magnitude (10^{-2} near the source to below 10^{-8} in the far field). To handle this range, log-space operations are adopted throughout: the surrogate predicts $\log_{10}\mathbf{C}$, sensor noise is modeled as additive Gaussian in log-space, and Bayesian evidence is accumulated as log-probabilities.

A deep learning surrogate model F is used to approximate the forward mapping from the encoded inputs to the unit-source concentration response **C**. For any candidate source s_i, the model predicts the normalized logarithm of the corresponding unit-release-rate concentration field:

$$\log_{10}\mathbf{C}_i(x,y;\mathbf{B},\mathbf{w},s_i) = F(\mathbf{X}_i)\,|\,(x,y;\mathbf{B},\mathbf{w},s_i)\in\Omega \tag{2}$$

where $\mathbf{C}_i(x,y;\mathbf{B},\mathbf{w},s_i)$ is the predicted unit-release-rate concentration field given building layout **B**, meteorology **w**, and candidate source s_i, and $\mathbf{X}_i$ is the physics-informed feature encoding constructed based on s_i detailed in Section 2.2. This encoding explicitly incorporates source-dependent physical priors related to transport direction, building geometry, and meteorological conditions, thereby improving the model's potential for generalization. In this study, both the surrogate output and the response library are represented in normalized $\log_{10}$-concentration form. Accordingly, the response matrix **R** stores normalized $\log_{10}$-concentration responses rather than linear concentration values.

The mathematical objective of inverse source-inference can be formulated as follows: given the sequential observations $\{z_1, z_2, \ldots, z_t\}$ collected by the searcher along the trajectory $\{\tau_1, \tau_2, \ldots, \tau_t\}$, infer the posterior probability distribution of the source position at each step:

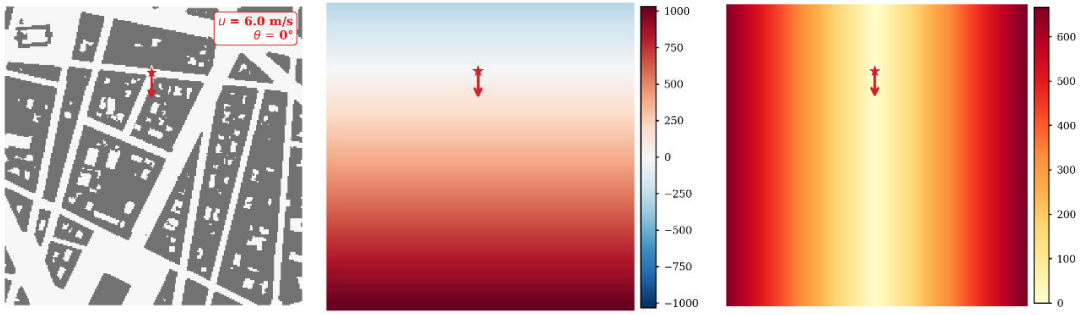
$$\mathbf{P}(s_{1:Ns}\,|\,z_{1:t}) \propto \mathrm{P}(z_{1:t}\,|\,s_{1:Ns})\cdot\mathbf{P}(s_{1:Ns}) \tag{3}$$

A uniform prior is adopted over the candidate source set $\mathbf{P}(s_{1:Ns}) = 1/Ns$. The likelihood function $\mathrm{P}(z_{1:t}\,|\,s_{1:Ns})$ is constructed from the surrogate predictions under a Gaussian observation-noise model in log-space. For each candidate source s_i, Eq. (2)

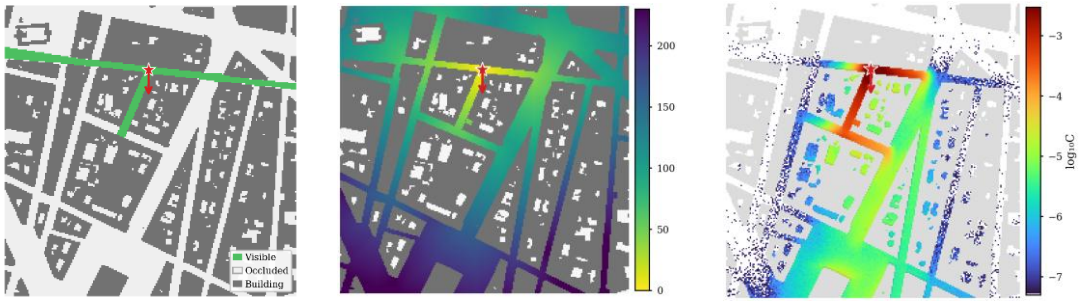
provides the corresponding unit-release-rate concentration field. Because the true release rate q^* is unknown, it is treated as a nuisance parameter via a profile-likelihood-type procedure: for each candidate source s_i, the log-scale offset is estimated by maximum likelihood, and the residual sum of squares evaluates consistency with observations $\{z_1, z_2, \ldots, z_T\}$. This yields a sequentially updated profile-likelihood-based posterior over source location.

2.2 Physics-Informed Feature Encoding

Four source-position-dependent physics-informed features are proposed. The surrogate model requires three types of physical input: candidate source location, wind conditions, and building geometry. These are encoded into a four-channel tensor $\mathbf{X}_i \in \mathbb{R}^{4\times H\times W}$, organized into two complementary groups. The first group (along-wind distance, AWD; crosswind distance, CWD) encodes the spatial structure prior of idealized open-terrain dispersion. The second group (line-of-sight indicator, LOS; shortest-path distance, SPD) encodes building-induced modifications to this idealized structure. AWD and CWD depend on wind-relative coordinates, whereas LOS and SPD are recomputed each new building layout through ray tracing and graph search. All four channels depend explicitly on the candidate source position and therefore allow different concentration fields to be predicted for different candidate sources.



(a) Input information (b) AWD (c) CWD



(d) LOS (e) SPD (f) Concentration field

Figure 2. Example of physics-informed feature encoding and concentration field.

(a) Along-Wind Distance (AWD)

For candidate source $s_i = (x_{s,i}, y_{s,i})$ and wind $\mathbf{w} = (\theta, u)$, at any grid position (x,y), $\mathbf{AWD}_i$ is a continuous field over the domain:

$$\mathbf{AWD}_i(x,y) = [(x - x_{s,i})\cos\theta + (y - y_{s,i})\sin\theta] \cdot u \tag{4}$$

where $\mathbf{AWD}_i(x,y) \in \mathbb{R}^{H \times W}$, positive values indicate downwind positions, whereas negative values indicate upwind positions. Multiplication by u accounts for wind speed and therefore reflects advective transport intensity.

(b) Crosswind Distance (CWD)

$\mathbf{CWD}_i$ is a continuous non-negative field defined as

$$\mathbf{CWD}_i(x,y) = |-(x - x_{s,i})\sin\theta + (y - y_{s,i})\cos\theta| \cdot u \tag{5}$$

where $\mathbf{CWD}_i(x,y) \in \mathbb{R}^{H \times W}$, absolute value enforces symmetry about the plume axis.

Under neutral stability, a Gaussian plume provides an analytical prior:

$$\mathbf{C}_{GP}(x,y) = \frac{q}{2\pi u \sigma_y(x)\sigma_z(x)} \exp\left(-\frac{y^2}{2\sigma_y^2(x)}\right) \tag{6}$$

where x is along-wind distance, y is crosswind distance, and $\sigma_y(x)$, $\sigma_z(x)$ are Pasquill–Gifford dispersion parameters (class D, Section 3.2). The idealized field depends solely on (x,y): monotonic decay along x, Gaussian decay along y. AWD and

CWD encode these coordinates as spatial fields, providing this structural prior without empirical parameter selection.

(c) Line of Sight (LOS)

$$\mathbf{LOS}_i = \begin{cases} 1 & \text{if straight line from } \mathrm{s}_i \text{ to } (x,y) \text{ without building} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Where $\mathbf{LOS}_i \in \{0,1\}^{H\times W}$, undefined on building pixels and masked out during training and evaluation over the fluid domain. Source–receptor lines are discretized using Bresenham's algorithm, requiring $O(\max(H,W))$ operations per ray. The sky view factor (SVF) has been shown to correlate with street-canyon pollutant concentration; LOS can be interpreted as a two-dimensional source–receptor analogue that characterizes direct building obstruction.

(d) Shortest Path Distance (SPD)

$$\mathbf{SPD}_i(x,y) = \min_{Q\in P(s_i\to(x,y))} |Q| \tag{8}$$

where $\mathbf{SPD}_i \in \mathbb{R}^{H\times W}$, undefined on building pixels and masked out during training and evaluation over the fluid domain. $P(s_i \to (x,y))$ is the set of fluid-domain paths connecting s_i to pixel (x,y), with $|Q|$ the path length. SPD is defined as the minimum path length over Q, and is computed using Dijkstra's algorithm on the 8-connected fluid-domain graph. When $\mathbf{SPD}_i(x,y)$ substantially exceeds the Euclidean distance $\|s_i-(x,y)\|_2$, it signifies strong geometric obstruction that may force transport along circuitous routes and thus lead to stronger concentration attenuation.

These four channels are assembled into the feature tensor:

$$\mathbf{X}_i = [\mathrm{AWD}_i;\mathrm{CWD}_i;\mathrm{LOS}_i;\mathrm{SPD}_i] \in \mathbb{R}^{4\times H\times W} \tag{9}$$

where the first two channels are functions of the candidate source position and

meteorological conditions, while the latter two are functions of the candidate source position and building layout.

2.3 Deep Learning Surrogate Model

The surrogate model F takes feature tensor $\mathbf{X}_i$ as input and outputs the logarithm unit-release-rate concentration field $\log_{10}\mathbf{C}$. The primary architecture is DeepLabV3+, whose atrous spatial pyramid pooling (ASPP) module [44] provides multi-scale receptive fields (dilation rates 6, 12, 18) matching the multi-scale structure of dispersion fields. The final layer is modified for single-channel regression. U-Net [45], FNO2d [46], and Pix2Pix [47] serve as baselines (Section 4.1.1).

The mean squared error in $\log_{10}$-space is computed over fluid-domain pixels:

$$\boldsymbol{L} = \frac{1}{|\Omega|}\sum_{(x,y)\in\Omega}\left[\mathbf{F}(\mathbf{X}_i)(x,y) - \log_{10}\mathbf{C}(x,y;\mathbf{B},\mathbf{w},s_i)\right]^2 \tag{10}$$

Where $\mathbf{C}$ is the true unit-release-rate concentration response generated by numerical simulations. The $\log_{10}$ transformation ensures balanced attention to high-concentration (near-source) and low-concentration (far-field) regions, as the spatial patterns of low concentrations also carry source-location information. For brevity, $\log_{10}$ is denoted as log throughout the remainder of this paper.

The optimizer is AdamW (learning rate 10^{-3}, weight decay 10^{-4}) with cosine annealing learning rate scheduling. Early stopping terminates training when validation loss does not improve for 20 consecutive epochs. Online data augmentation applies rotations (90°, 180°, and 270°) together with horizontal and vertical reflections to both the input tensor and the target field, resulting in an eightfold geometric augmentation.

2.4 Response-library construction from surrogate-predicted fields

Given the wind information $\mathbf{w}=(\theta,u)$ and building map **B**, the trained surrogate model $\mathbf{F}$ is exhaustively evaluated over all candidate source $s_i \in \mathbf{S}$ to construct a response matrix $\mathbf{R}$ for subsequent inference. Because the candidate source set is defined as the full fluid-domain pixel set, exhaustive precomputation remains tractable while preserving the full spatial structure of the posterior distribution and avoiding approximation errors associated with particle-based representations.

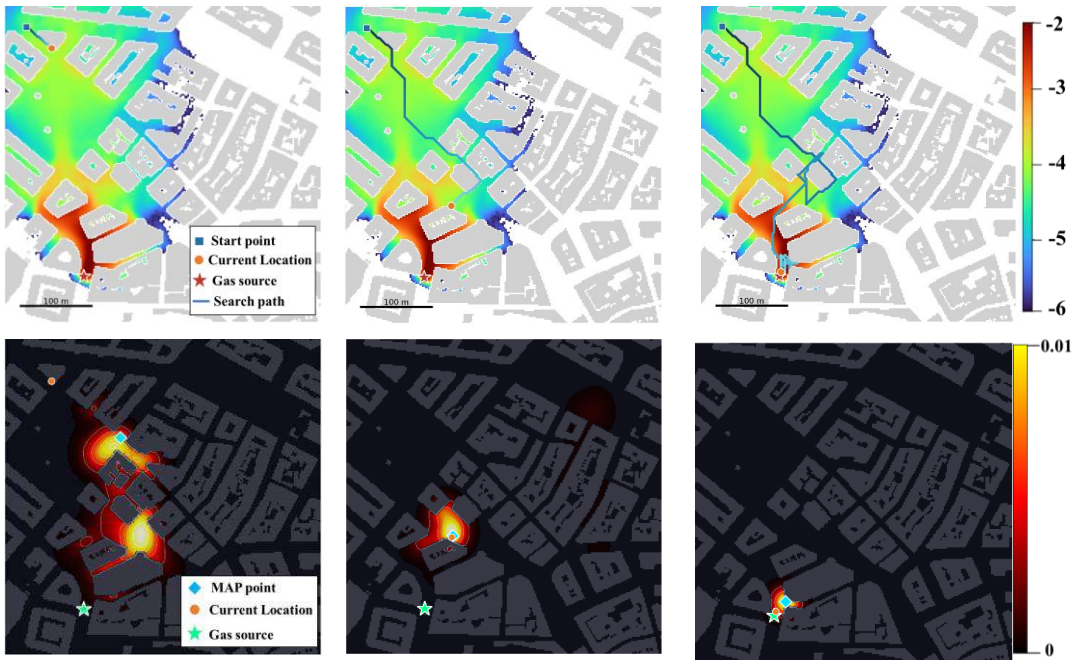
For each candidate source S_i, $\mathrm{AWD}_i, \mathrm{CWD}_i$ are computed using Eq.(4) and (5) whereas LOS_i are computed using Eq.(7) and (8). These four fields are then assembled into the feature tensor $\mathbf{X}_i$ according to Eq.(9). The tensors $\{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_{Ns}\}$ are fed into the surrogate model in batch to obtain the predicted unit-source concentration fields. The predicted normalized log-concentration values at all fluid-domain pixels are then extracted and arranged into the response matrix $\mathbf{R}\in\square^{Ns\times Ns}$, where $\mathbf{R}_{i,j}=\mathbf{C}_{i,j}$ denotes the normalized $\log_{10}$ concentration at the j_{th} fluid-domain pixel for a unit-source release from candidate source S_i.

2.5 Sequential Bayesian source inference based on the response library

With response matrix $\mathbf{R}$ constructed, sequential source inference is performed. At each step t, three actions are executed: concentration measurement (Section 2.5.1), probability update (Section 2.5.2), and selection of the next sampling position (Section 2.5.3). This cycle continues until the source is reached or step budget $T_{\max}$ is exhausted.

For each candidate source, the unknown release rate is treated as a nuisance parameter, appearing as an additive offset in log-space and estimated by maximum likelihood. The resulting residual sum of squares yields a profile-likelihood-type

estimate for source-location inference. This shifts the problem from matching absolute magnitude to matching the spatial decay pattern, thereby enabling robust source inference even when the true release rate differs by orders of magnitude. Figure 3 illustrates a representative inference process.



(a) Step 1 (b) Step 8 (c) Step 18

Figure 3. Example search path and evolving posterior probability distribution.

2.5.1 Sequential concentration observations

The source-inference procedure is activated only after the measured concentration exceeds a predefined alarm threshold, so that subsequent log-space updates are performed on positive observations within the detectable range. At step *t*, the concentration observation. Portable-sensor error is modeled as multiplicative in concentration space, which corresponds to additive Gaussian noise in log-space:

$$\log z_t = \log q^* + \log C(\tau_t; s^*) + \varepsilon_t \tag{11}$$

where $C(\tau_t; s^*)$ is true unit-release-rate response at τ_t, unknown release rate q^* and source position s^*, noise standard deviation $\varepsilon_t \square N(0, \sigma_{obs}^2)$. Assuming independent

Gaussian observation noise with constant variance in log-space, the unknown release-rate effect for each candidate source appears as an additive constant offset, whose maximum-likelihood estimate is the mean observation–prediction difference.

2.5.2 Bayesian posterior update

Using observations $\{z_1, z_t, \ldots, z_t\}$, update posterior probability $b_t(s_i)$ for each candidate s_i under a uniform prior and the Gaussian log-noise model described above. For each s_i, retrieve its unit-release-rate predictions at all past positions from $\mathbf{R}$: R_{i,j_k} for k=1,2,⋯,t, where j_k indexes the pixel of past position τ_k. R_{i,j_k} is the concentration expected at τ_k if source were s_i with unit release rate. By Eq. (1), $\log q^*$ is a uniform constant offset in log-space. For candidate s_i, the maximum likelihood estimate of this offset is the mean of observation-prediction differences:

$$\mu_i = \frac{1}{t}\sum_{k=1}^{t}(\log z_k - \log \mathbf{R}_{i,j_k}) \tag{12}$$

If s_i is the true source, $\mu_i \approx \log q^*$.

After removing estimated release rate, compute residual sum of squares—how well the spatial decay pattern matches observations:

$$v_i = \sum_{k=1}^{t}(\log z_k - \mu_i - \log \mathbf{R}_{i,j_k})^2 \tag{13}$$

Smaller v_i means better pattern match, independent of unknown absolute scale. Posterior probability follows directly:

$$\log b_t(s_i) = -\frac{v_i}{2\sigma_{obs}^2} - \log K_t \tag{14}$$

Where K_t normalization constant ensuring $\sum_i b_t(s_i) = 1$.

Therefore, the posterior used in this study is a profile-likelihood-based posterior

over source location, obtained under a uniform prior and a Gaussian log-space observation model.

2.5.3 MAP-guided sampling rule

The maximum a posteriori (MAP) candidate is selected as the current target:

$$\hat{s}_t = \arg\max_{s_i \in \mathbf{S}} b_t(s_i) \tag{15}$$

At each decision round, the next sampling position is chosen along the shortest fluid-domain path toward the current MAP target $\hat{s}_t$ subject to a maximum movement radius of 25 pixels (approximately 50 m) per step. Thus, one step in this study refers to one decision-and-update cycle rather than one-pixel motion. If the posterior mode shifts at a later step, the path is replanned accordingly. The inference process is declared successful when the sampling position first enters the arrival region of the true source, $\left\|\tau_t - s^*\right\| \le r_{\text{arrive}}$, where $r_{\text{arrive}} = 4\text{m}$ in the study; otherwise the procedure continues until $T_{\max}$. A MAP-guided sampling rule is adopted to enable controlled comparison of likelihood models: in DL+MAP versus GP+MAP, the sampling rule is identical, so performance differences primarily reflect likelihood fidelity.

2.6 Evaluation metrics

(a) Forward prediction accuracy

Log-space $R^2(\log)$ over fluid domain pixels only:

$$R^2(\log) = 1 - \frac{\sum_{(x,y)\in\Omega}[\log C_i(x,y) - \log C(x,y;s_i)]^2}{\sum_{(x,y)\in\Omega}[\log C(x,y;s_i) - \overline{\log C}]^2} \tag{16}$$

where $\overline{\log C}$ is mean true log-concentration over Ω. $R^2(\log) = 1$ indicates perfect prediction.

(b) Source-inference performance

Finding rate is defined as the proportion of trials in which the searcher reaches within r_{arrive} of the true source s^* within $T_{\max}$ steps.

Path efficiency is defined as the ratio of the shortest feasible path length from the start location to the true source to the actual path length traveled by the searcher:

$$\text{PE} = L_{\text{shortest}} / L_{\text{actual}} \tag{17}$$

where L_{shortest} is shortest Dijkstra path length from start to s^*, L_{actual} is actual path traveled. Larger path efficiency indicates a more direct search path.

Median search steps denote the median number of steps required to reach the source among successful trials only. Here, the median refers to the middle value after sorting the step counts in ascending order.

3. Experimental setup

3.1 Dataset

The simulation data used in this study were generated using the PMSS (Parallel Micro-Swift-Spray) atmospheric dispersion modeling system [48]. The dataset is publicly available on Zenodo [38]. PMSS couples a diagnostic wind-field model (Micro-Swift) with a Lagrangian particle-dispersion model (Spray). Compared with full CFD solvers, this framework provides substantially higher computational efficiency while still resolving building-scale flow features, making it suitable for rapid-response simulations at the urban-neighborhood scale. Each simulation represents a two-hour time-integrated near-surface concentration field generated by a point source located at a height of 2 m under steady meteorological conditions.

The dataset covers Paris and Grenoble. Paris has 18 wind directions × 3 speeds

(1.5, 3.5, 6.0 m/s) = 54 combinations; Grenoble has 36 directions × 3 speeds = 108 combinations. Each city is divided into four 224×224-pixel blocks (448 m × 448 m). After excluding sources within 30 pixels of boundaries, 15,804 valid samples remain.

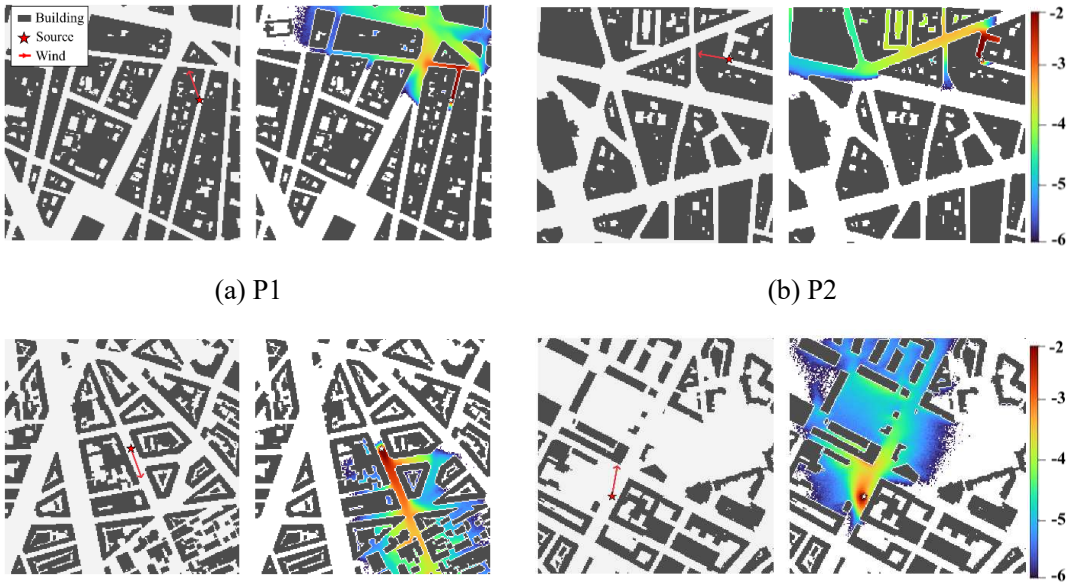
Paris blocks have λ of 56.5%–63.1% with compact, similar layouts; Grenoble blocks span λ 32.0%–49.3% with greater morphological variation (Figure 4). This contrast provides a useful test bed for cross-domain generalization.



(a) Paris city map(~1km^2) (b) Paris blocks(448m×448m)

(c) Grenoble city map(~1km^2) (d) Grenoble blocks(448m×448m)

Figure 4. 8 blocks of two cities in France



(a) P1 (b) P2

(c) G1 (d) G2

Figure 5. 4 examples of concentration

3.2 Experimental design

The experiments include both forward prediction and source-inference evaluation. All experimental cases are summarized in Table 1, with feature-encoding schemes, ablation settings, surrogate architectures, and sequential-inference parameters detailed in Tables 2 – 5, respectively. Cases 1 – 7 focus on forward-prediction performance and generalization, Cases 8 – 10 evaluate source-inference performance, and Case 11 provides validation against independent simulations.

Table 1. Overview of the Experimental Plan

Case	Scenario	Section	Training set	Training Proportion	Testing set	Generalization type	Feature type	Model
1	Forward prediction	4.1.1	P1\P2\ G1\G2	100%	P3\P4\ G3\G4	Cross-block	3 encoding schemes in Table 2	4 types of surrogate models in Table 4
2			P1-4		G1-4	Cross-city		
3			G1-4		P1-4	Cross-city		
4		4.1.2	P1\P2\ G1\G2	1%\2.5%\ 5%\10%\ 25%\50%	P3\P4\ G3\G4	Cross-block	Physics-informed feature	DeepLabV3+
4b			P1\P2\ G1\G2		P3\P4\ G3\G4	Cross-block (Feature ablation in		

						Table 3 without data augmentation)		
5			P1-4		G1-4	Cross-city		
6			G1-4		P1-4	Cross-city		
7		4.1.3	One block	100%	Other blocks	Cross-city		DeepLabV3+
8	Source location	4.2	P1\P2\G1\G2	1%\2.5%\5%\10%\25%\50%\100%	P3\P4\G3\G4	Cross-block		4 types of pollutant source location method in Table 5
9			P1-4		G1-4	Cross-city		
10			G1-4		P1-4	Cross-city		
11	External-simulator validation	4.3	P1\P2\G1\G2	100%	Fluent simulations	Cross-simulations		DeepLabV3+

Table 2. Three types of feature encoding schemes

No.	Type	Notation
1	Physics-informed Feature	**AWD+CWD+LOS+SPD**
2	Conventional inputs + Gaussian-plume prior	Building mask + wind speed + wind direction + source location + Gaussian-plume prior
3	Conventional inputs only	Building mask + wind speed + wind direction + source location

Table 3. Feature-ablation settings for Case 4b without data augmentation

No.	Type	Notation
1	Physics-informed feature encoding	**AWD+CWD+LOS+SPD**
2	Replace LOS and SPD with building mask	AWD + CWD + Building
3	Conventional meteorological encoding with LOS and SPD	Wind speed + Wind direction + LOS + SPD
4	Without alongwind distance	CWD+LOS+SPD
5	Without crosswind distance	AWD+LOS+SPD
6	Without line-of-sight indicator	AWD+CWD+SPD
7	Without shortest-path distance	AWD+CWD+LOS

Table 4. Four surrogate architectures evaluated in this study

Model	DeepLabV3+[44]	U-Net [45]	FNO2d [46]	Pix2Pix [47]
Parameters	6.68M	31.39M	4.74M	57.18M
Core mechanism	Xception backbone + ASPP	Encoder-decoder + skip connections	Fourier-space global convolution	Conditional U-Net generator + PatchGAN discriminator
Key configuratio	base_ch = 32, output stride = 16,	64-128-256-512-1024,	Width = 64; 4 Fourier layers;	ngf = 64, num_downs = 8;

n	ASPP rates = 6/12/18	no batch normalization (BN), 3×3 conv with p = 1	spectral modes = 12 × 12	ndf = 64, n_layers = 3; no output Tanh
Resolution strategy	Encoder-decoder with bilinear upsampling	Symmetric down/up-sampling with transposed convolutions	Full-resolution processing throughout	Reflect-pad to 256 × 256, then crop back to original size
Main inductive bias	Multi-scale local-context aggregation	Local hierarchical feature fusion	Global spectral operator learning	Adversarial conditional image-to-image translation

Table 5. Parameters for the four pollution source-inference methods

Parameter	Settings
Likelihood function models	DeepLabV3+(DL)+MAP\Gaussian Plume (GP)+MAP\ Gradient chemotaxis\Upwind walk
Arrival radius r_{arrive}	$\leqslant$ 2px (4 m)
Step budget $T_{\max}$	$\leqslant$ 300steps
Step movement radius r_{move}	$\leqslant$ 25px/step (~50 m)
Observation noise σ_{obs}	{0.3,0.5,1.0}

In Cases 8–10, for each test source, initial observation positions are sampled from fluid-domain locations that satisfy the post-detection trigger condition described in Section 2.5.1, i.e., positions where the simulated concentration exceeds the alarm threshold. Positions within the arrival radius of the true source are excluded. For each test source, identical initial positions are used for all compared methods. Three baseline methods are considered. In Upwind walk, no concentration information is used and the next sampling position is selected directly in the upwind direction. In Gradient chemotaxis, the next sampling position is selected toward the neighboring pixel with the highest measured concentration. In GP+MAP, the same MAP-guided sampling rule as in the proposed method is retained, but the likelihood is constructed from a neutral-stability Gaussian-plume response library based on the Pasquill–Gifford class-D

formulation [10], using a = 0.068, b = 0.908, and c = 0 in the empirical dispersion-parameter relation adopted in this study. The four methods compared in the source-inference experiments are therefore DL+MAP (proposed), GP+MAP, Gradient chemotaxis, and Upwind walk. The DL+MAP versus GP+MAP comparison is particularly important because the sampling rule is held fixed, so any performance difference can be attributed mainly to the fidelity of the likelihood model. By contrast, Gradient chemotaxis and Upwind walk are included as lower-complexity reference behaviors rather than as state-of-the-art decision-theoretic baselines.

4. Results and Discussion

This section presents the results in four parts. Section 4.1 evaluates forward prediction performance, including the comparative effects of feature encoding and network architecture (Section 4.1.1), sample efficiency and feature-ablation analysis under DeepLabV3+ (Section 4.1.2), and single-block-to-others cross-validation (Section 4.1.3). Section 4.2 evaluates source-inference performance, including comparisons across methods and observation-noise levels (Section 4.2.1) and data-efficiency experiments that examine the relationship between forward prediction accuracy and source-inference performance (Section 4.2.2). Section 4.3 provides independent validation using external numerical simulations. Section 4.4 discusses the limitations of the current framework.

4.1 Forward prediction performance

4.1.1 Cross-comparison of feature encoding and architectures

Table 6 reports test-set $R^2(\log)$ for four surrogate architectures and three input-

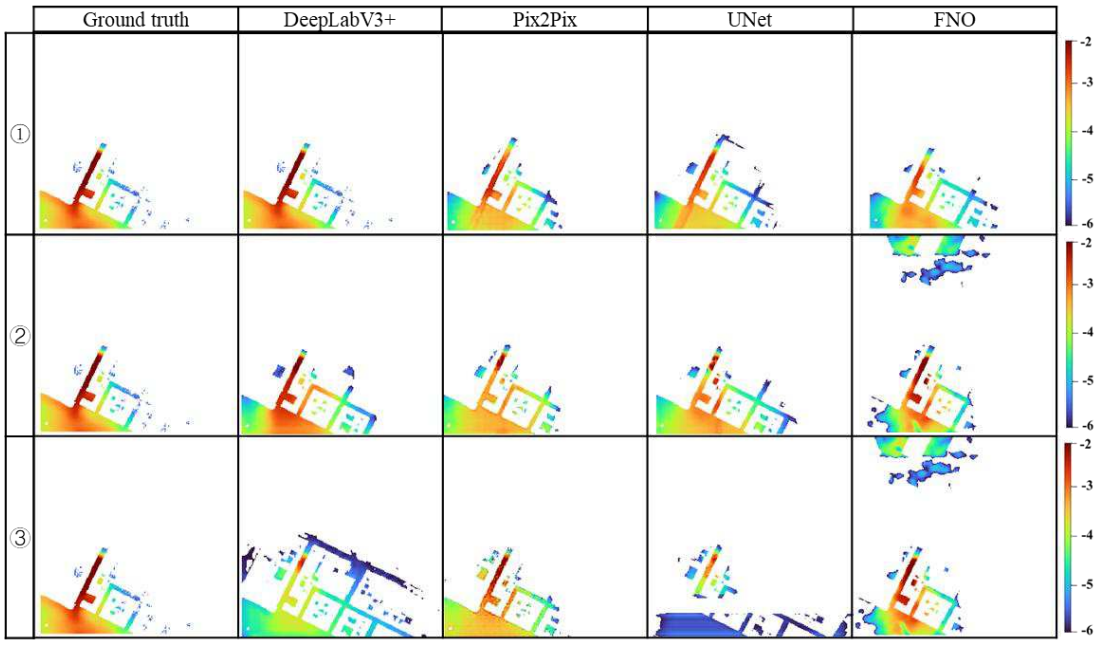
encoding schemes across the three generalization scenarios defined in Cases 1–3. Two main observations can be made.

First, prediction accuracy is more strongly influenced by feature encoding than by architecture. The physics-informed features consistently outperform conventional inputs with Gaussian-plume prior, which in turn outperform conventional inputs alone (R^2(log) 0.720–0.850 vs. negative values for conventional-only in cross-city cases). This confirms that source-dependent priors are critical for robust generalization (Figure 6).

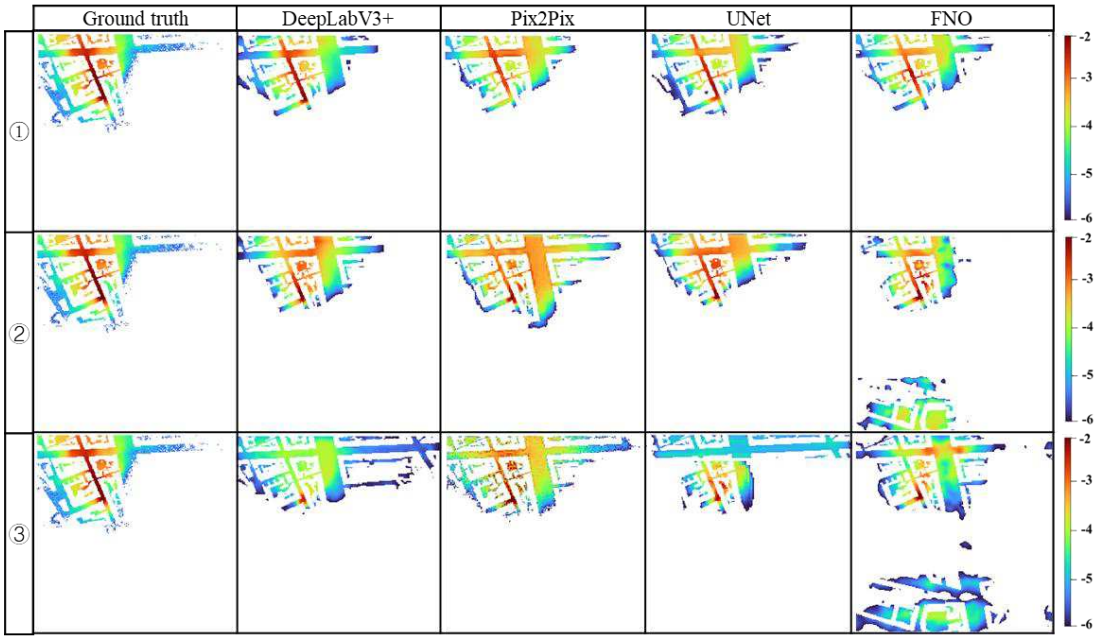
Second, DeepLabV3+ achieves the best overall performance with a favorable accuracy-to-parameter ratio (6.68M parameters), likely due to its ASPP module providing multi-scale receptive fields matching the physical structure of dispersion fields. FNO2d and U-Net perform less reliably under weaker encoding schemes. Based on these results, all subsequent experiments adopt DeepLabV3+.

Table 6. Test-set R^2 (log) for four architectures and three encoding schemes in Cases 1–3

Case	Features	DeepLabV3+	U-Net	FNO2d	Pix2Pix
1	①Physics-informed feature	0.850	0.833	0.792	0.771
	②Conventional + Gaussian-plume	0.789	0.770	0.498	0.723
	③Conventional feature	0.390	0.210	0.268	0.260
2	①Physics-informed feature	0.819	0.773	0.757	0.720
	②Conventional + Gaussian-plume	0.746	0.747	0.255	0.678
	③Conventional feature	-0.412	0.027	-0.137	-0.651
3	①Physics-informed feature	0.845	0.803	0.781	0.761
	②Conventional + Gaussian-plume	0.758	0.732	0.405	0.692
	③Conventional feature	-0.269	0.031	-0.160	-0.484



(a) Example 1 under different feature encodings and surrogate architectures for two test cases



(b) Example 2 under different feature encodings and surrogate architectures for two test cases

Figure 6. Representative concentration-field predictions in Case 1

4.1.2 Sample efficiency and feature ablation under DeepLabV3+

Table 7 reports R^2 (log) for DeepLabV3+ with the full physics-informed feature encoding in Cases 4 – 6 at different training-data proportions. Figure 7(a) shows the corresponding average R^2 (log) across the three transfer scenarios. Overall, the model exhibits relatively modest data requirements. Cross-block training (Case 4), which

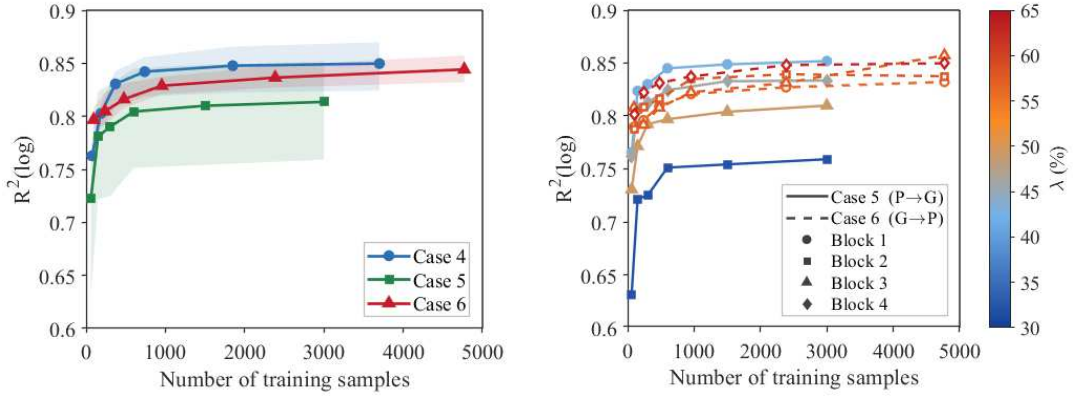
includes blocks from both cities, yields the best overall prediction performance. All three scenarios show a consistent pattern of rapid initial improvement followed by diminishing returns. The largest gains occur between 1% and 5% of the training data, and the performance curves flatten thereafter. In all three cases, the average performance at 5% data already reaches more than 97% of the corresponding value at 50% data. This indicates that, under the present architecture and setting, the surrogate can achieve strong generalization with a relatively limited number of training samples.

Table 7. R^2 (log) at different training data proportions

Case	Training set Proportion	1.0%	2.5%	5.0%	10.0%	25.0%	50.0%
4	Training set number	74	185	370	740	1850	3700
	Testing P3	0.752	0.811	0.843	0.856	0.866	0.870
	Testing P4	0.778	0.805	0.842	0.851	0.864	0.865
	Testing G3	0.757	0.787	0.809	0.819	0.822	0.824
	Testing G4	0.765	0.808	0.829	0.843	0.840	0.840
	Average	0.763	0.803	0.831	0.842	0.848	0.850
5	Training set number	60	150	300	600	1500	3000
	Testing G1	0.766	0.824	0.830	0.845	0.849	0.852
	Testing G2	0.631	0.722	0.725	0.751	0.754	0.759
	Testing G3	0.731	0.772	0.792	0.797	0.804	0.810
	Testing G4	0.762	0.807	0.813	0.825	0.833	0.834
	Average	0.722	0.781	0.790	0.805	0.810	0.814
6	Training set number	95	238	477	954	2385	4770
	Testing P1	0.790	0.796	0.810	0.821	0.827	0.832
	Testing P2	0.788	0.809	0.816	0.835	0.840	0.838
	Testing P3	0.808	0.792	0.808	0.822	0.831	0.857
	Testing P4	0.802	0.823	0.831	0.837	0.848	0.850
	Average	0.793	0.809	0.820	0.831	0.840	0.844

Figure 7(b) shows a clear asymmetry between cross-city transfer directions. Grenoble-to-Paris (Case 6) starts high even at 1% data (R^2(log) 0.788–0.808), whereas Paris-to-Grenoble (Case 5) is more dispersed (0.631–0.766), with Grenoble Block 2 consistently lagging. This asymmetry reflects differences in building-density coverage: Paris blocks have a narrow building-density range (56.5%–63.1%), whereas Grenoble

blocks span a broader range (32.0%–49.3%), with Block 2 having the lowest building density.



(a) Average R^2(log) with data proportion (b) Block-wise R^2(log) in Case5\6

Figure 7. R^2 (log) under different training-data proportions for the three generalization scenarios

Case 4b evaluates feature ablations under the same cross-block setting as Case 4 using DeepLabV3+ only. As shown in Tables 3 and 8, the full physics-informed feature encoding consistently performs best across all training-data ratios, improving from 0.746 at 1% training data to 0.808 at 50%. Among the ablated components, along-wind distance (AWD) is the most important, since removing it causes the largest performance drop by a wide margin. Removing CWD, LOS, or SPD results in comparatively smaller degradations, suggesting that these features play complementary roles. The conventional meteorological encoding variant also remains clearly inferior to the full physics-informed encoding. Overall, the results show that the benefit of the proposed encoding mainly comes from explicit source-relative physics-informed descriptors, especially AWD.

Table 8. Test R^2(log) under different training-data ratios for feature ablations in Case 4b

Training set Proportion	1.0%	2.5%	5.0%	10.0%	25.0%	50.0%
Training set number	74	185	370	740	1850	3700
AWD+CWD+LOS+SPD (Proposed)	0.746	0.773	0.784	0.793	0.806	0.808
AWD + CWD + Building	0.680	0.729	0.742	0.745	0.755	0.765
Wind speed + Wind direction + LOS + SPD	0.244	0.440	0.505	0.430	0.460	0.481

CWD+LOS+SPD	0.003	0.154	0.235	0.260	0.267	0.207
AWD+LOS+SPD	0.734	0.768	0.775	0.780	0.794	0.794
AWD+CWD+SPD	0.728	0.765	0.781	0.794	0.796	0.804
AWD+CWD+LOS	0.737	0.753	0.773	0.772	0.771	0.778

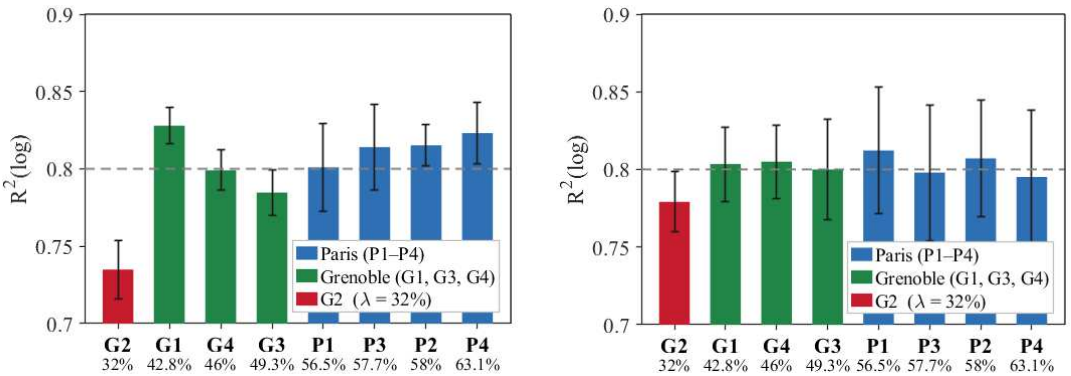
4.1.3 Single-block-to-others transfer experiment

In Case 7, each fold uses one block for training (with eightfold augmentation) and the remaining seven for testing. The overall mean R^2(log) is 0.800, with same-city mean 0.811 versus cross-city 0.792 (difference 0.019), indicating strong single-block transfer including cross-city cases (Table 9).

Table 9. R^2 (log) of single-block-to-others transfer experiment

Block	G1	G2	G3	G4	P1	P2	P3	P4	Row mean
G1(Fold1)	—	0.752	0.804	0.811	0.802	0.820	0.821	0.816	0.804
G2(Fold2)	0.806	—	0.775	0.789	0.747	0.788	0.763	0.787	0.779
G3(Fold3)	0.837	0.741	—	0.787	0.791	0.821	0.798	0.827	0.800
G4(Fold4)	0.838	0.762	0.798	—	0.795	0.811	0.817	0.814	0.805
P1(Fold5)	0.837	0.734	0.781	0.816	—	0.830	0.850	0.837	0.812
P2(Fold6)	0.829	0.732	0.789	0.806	0.818	—	0.831	0.845	0.807
P3(Fold7)	0.821	0.717	0.761	0.804	0.831	0.817	—	0.835	0.798
P4(Fold8)	0.827	0.707	0.786	0.783	0.824	0.821	0.820	—	0.795
Col mean	0.828	0.735	0.785	0.799	0.801	0.815	0.814	0.823	0.800

Prediction accuracy is more sensitive to target-block morphology than to training-block identity. G2 (lowest λ =32%) consistently shows the lowest R^2(log) as both training and test block, suggesting that broader building-density coverage in the training set may improve generalization.



(a) Performance grouped by test block

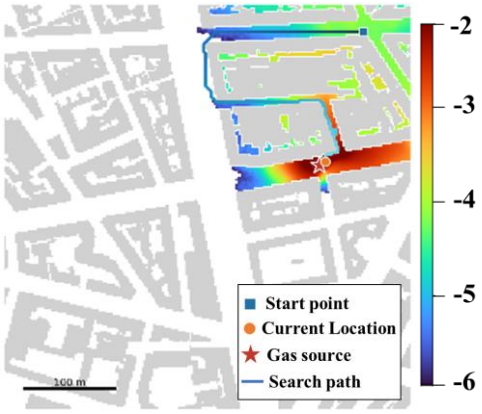
(b) Performance grouped by training block

Figure 8. Single-block-to-others 8-fold cross-validation results

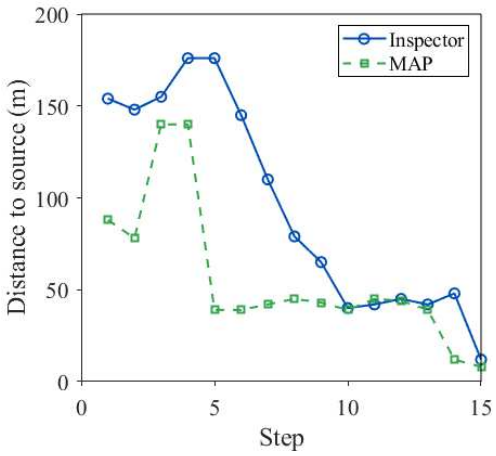
4.2 Source-inference performance based on surrogate-predicted response fields

4.2.1 Comparison of source-inference methods under different observation-noise levels

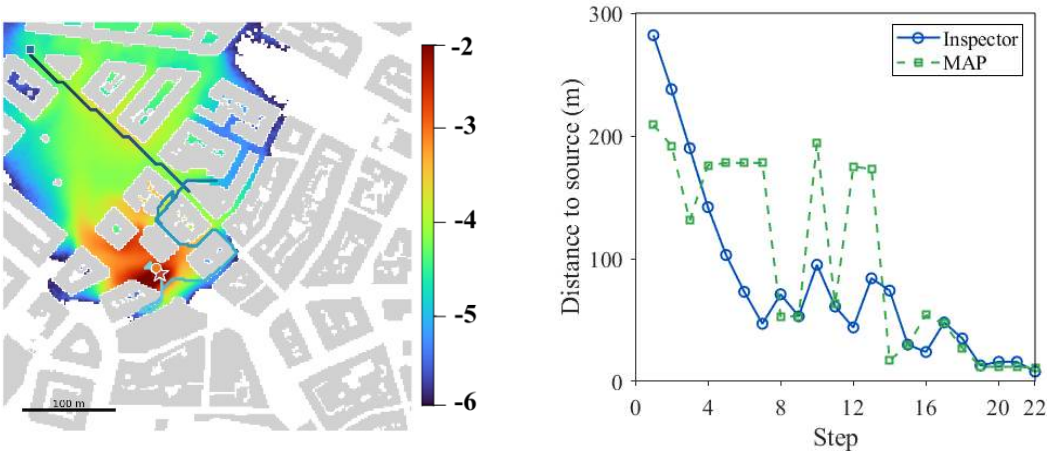
Four source-inference methods are compared under the Case 8 setting at three observation-noise levels. The DeepLabV3+ surrogate is trained on Blocks 1 – 2 from both cities. Each test sample is assigned a randomly sampled release rate from the predefined range described in Section 3, so that inference performance can be evaluated under release-rate uncertainty. The arrival radius is set to 2 pixels (4 m), the step budget is 300 steps, and the per-step movement radius is 25 pixels (50 m). For both Bayesian methods, the same profile-likelihood-type treatment is used to account for the unknown release rate.



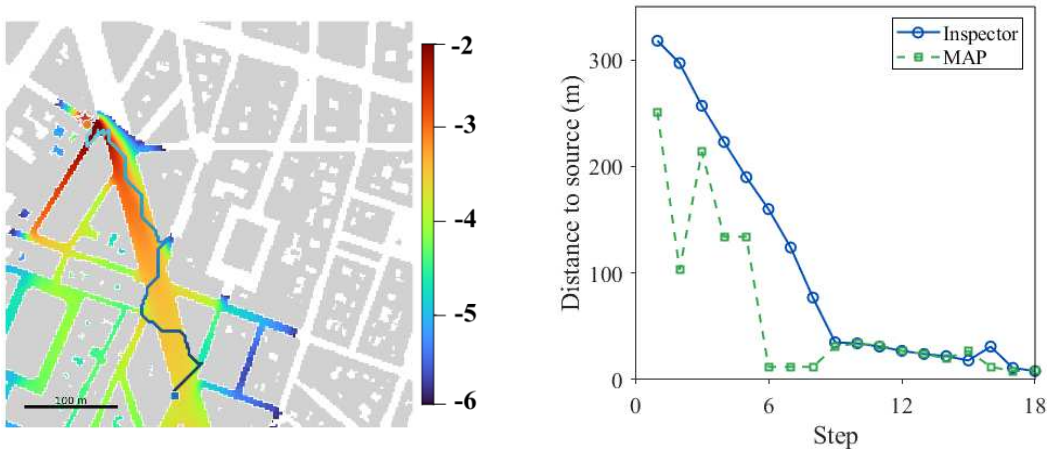
(a) Search path of example 1



(b) Distance to source of example 1



(c) Search path of example 2 (d) Distance to source of example 2



(e) Search path of example 3 (f) Distance to source of example 3

Figure 9. Examples of source-inference trajectories

Figure 9 shows representative source-inference trajectories. Table 10 summarizes the quantitative results. DL+MAP performs best at all noise levels, maintaining a finding rate above 96%. As noise increases, the median number of steps rises from 20 to 33 and path efficiency decreases from 0.535 to 0.370, but DL+MAP still substantially outperforms all other methods.

The DL+MAP versus GP+MAP comparison is particularly informative because both share the same sampling rule. Replacing the Gaussian-plume likelihood with the DL surrogate increases the finding rate from 46% to 96%, reduces the median number of steps from 99 – 104 to 20 – 33, and improves path efficiency from 0.11 – 0.12 to 0.37 – 0.54, confirming the central role of likelihood fidelity. Gradient chemotaxis

achieves a finding rate of 52% – 55% with 71 – 77 median steps, whereas Upwind walk achieves only 23%.

Table 10. Pollution source-inference performance at different observation-noise levels

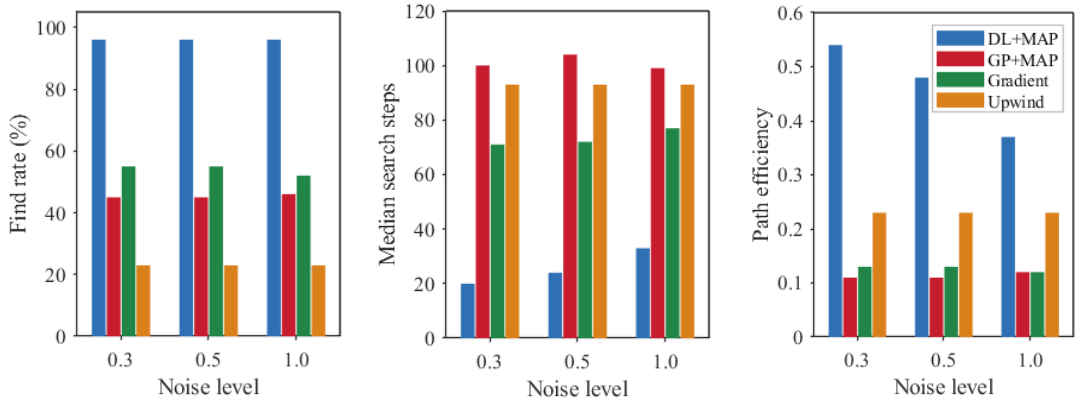
Noise level	Method	Finding rate (%)	Median Search steps	Path efficiency
0.3	DL+MAP	96.43	20	0.535
	GP+MAP	45.52	100	0.114
	Gradient	55.23	71	0.126
	Upwind	23.40	93	0.233
0.5	DL+MAP	96.48	24	0.483
	GP+MAP	45.46	104	0.112
	Gradient	54.04	72	0.125
	Upwind	23.40	93	0.233
1	DL+MAP	96.26	33	0.370
	GP+MAP	45.99	99	0.119
	Gradient	51.71	77	0.118
	Upwind	23.40	93	0.233

Table 11 summarizes the computational cost on a single NVIDIA T4 GPU. During on-site preparation, candidate source locations and sensor locations are discretized over the fluid-domain pixels, resulting in approximately 4,000 candidate-source grids and 4,000 sensor grids in the current implementation. For each candidate source, LOS and SPD feature maps are computed using accelerated parallelization, requiring approximately 20 s and about 2 GB of memory. Subsequently, DeepLabV3+ performs batched GPU inference to predict the normalized log-concentration response for each candidate source, thereby constructing the response matrix R. This inference stage takes approximately 40 s, and storing the response matrix in float32 format requires about 1 GB of memory.

Table 11. Computational cost summary

Stage	Time	Frequency
Offline: data generation, preprocessing, and surrogate training	7 working days	One-time
On-site preparation: response library	1 min/block	Per search event

Stage	Time	Frequency
construction (incl. LOS/SPD computation)		
Online: per-step decision	1 ms/step	Every step



(a) Finding rate (b) Median search steps (c) Path efficiency

Figure 10. Comparisons of 4 types of source location methods

4.2.2 Data efficiency of source inference

Tables 12 and 13 report finding rate and path efficiency versus training-data proportion (1% – 50%) for the three generalization settings in Cases 8 – 10. Overall, the finding rate reaches saturation more rapidly than path efficiency. With only 150 – 240 training samples, all three settings exceed a 95% finding rate. By contrast, path efficiency continues to improve up to approximately 10% of the training data, corresponding to 370 – 950 samples depending on the case.

Table 12. Path efficiency at different training data proportions

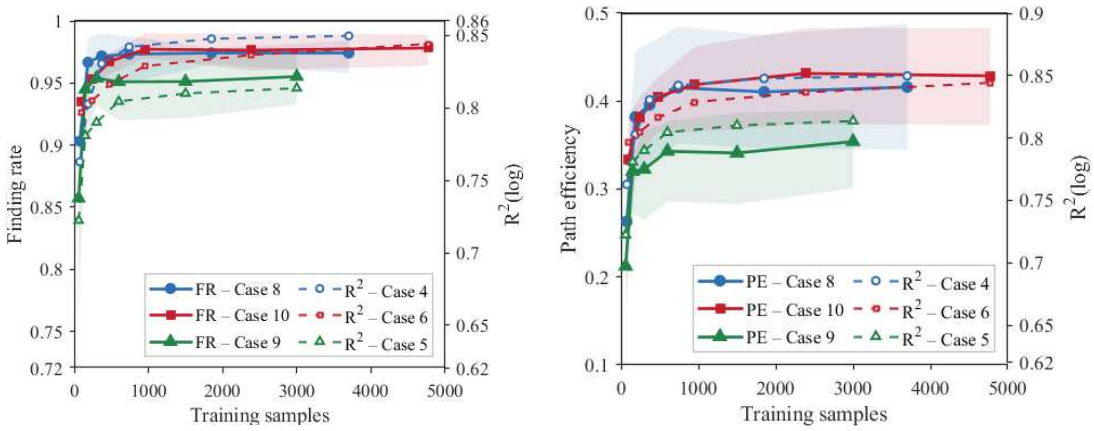
Case	Training set proportion	1.0%	2.5%	5.0%	10.0%	25.0%	50.0%
8	Training set number	74	185	370	740	1850	3700
	Testing P3	0.320	0.457	0.466	0.485	0.475	0.488
	Testing P4	0.280	0.406	0.426	0.442	0.435	0.442
	Testing G3	0.238	0.355	0.363	0.382	0.386	0.389
	Testing G4	0.215	0.310	0.328	0.352	0.346	0.345
	Average	0.248	0.362	0.376	0.396	0.392	0.396
9	Training set number	60	150	300	600	1500	3000
	Testing G1	0.229	0.334	0.355	0.363	0.359	0.371
	Testing G2	0.158	0.272	0.265	0.286	0.283	0.301

	Testing G3	0.239	0.340	0.366	0.377	0.387	0.390
	Testing G4	0.223	0.335	0.304	0.346	0.335	0.354
	Average	0.218	0.326	0.328	0.349	0.347	0.360
10	Training set number	95	238	477	954	2385	4770
	Testing P1	0.330	0.388	0.408	0.424	0.435	0.436
	Testing P2	0.276	0.341	0.364	0.365	0.374	0.373
	Testing P3	0.381	0.404	0.421	0.462	0.483	0.483
	Testing P4	0.349	0.393	0.427	0.424	0.434	0.423
	Average	0.332	0.382	0.406	0.418	0.430	0.427

Table 13. Finding rate at different training data proportions

Case	Training set proportion	1.0%	2.5%	5.0%	10.0%	25.0%	50.0%
8	Training set number	74	185	370	740	1850	3700
	Testing P3	0.913	0.986	0.990	0.988	0.976	0.984
	Testing P4	0.909	0.961	0.975	0.974	0.978	0.977
	Testing G3	0.892	0.948	0.950	0.950	0.965	0.958
	Testing G4	0.899	0.972	0.972	0.981	0.978	0.978
	Average	0.901	0.964	0.968	0.971	0.974	0.972
9	Training set number	60	150	300	600	1500	3000
	Testing G1	0.873	0.960	0.963	0.967	0.963	0.959
	Testing G2	0.764	0.908	0.932	0.920	0.922	0.933
	Testing G3	0.890	0.941	0.956	0.944	0.943	0.954
	Testing G4	0.902	0.971	0.965	0.973	0.977	0.975
	Average	0.869	0.950	0.957	0.955	0.956	0.959
10	Training set number	95	238	477	954	2385	4770
	Testing P1	0.938	0.949	0.974	0.976	0.979	0.982
	Testing P2	0.920	0.953	0.957	0.970	0.960	0.964
	Testing P3	0.941	0.955	0.966	0.990	0.990	0.989
	Testing P4	0.940	0.955	0.972	0.973	0.977	0.978
	Average	0.935	0.952	0.969	0.976	0.976	0.978

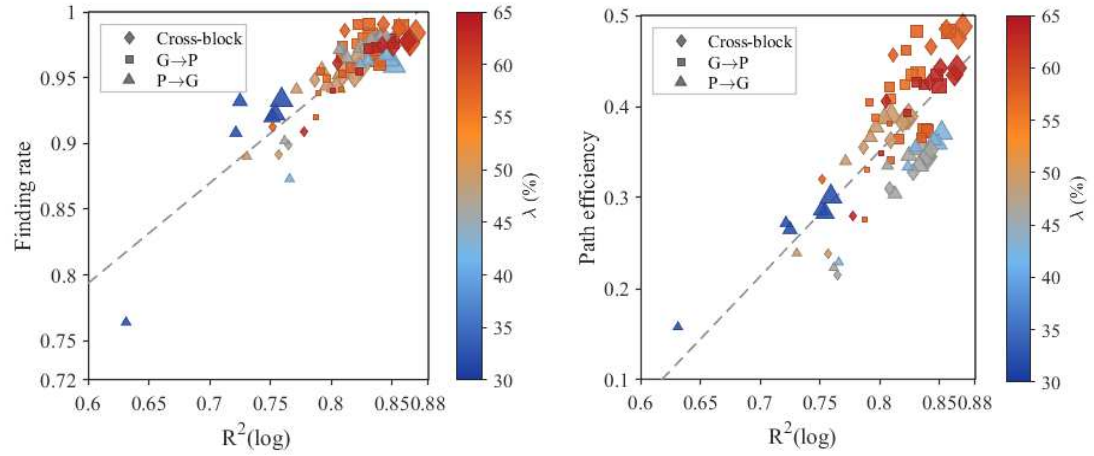
As shown in Figure 11(a), finding rate increases sharply between 1% and 2.5% of training data and stabilizes after 5%, with 150–240 samples sufficient to exceed 95% (Figure 11(a)). Path efficiency saturates later, around 10% (Figure 11(b)), indicating that forward prediction improvements continue benefiting path efficiency after finding rate has plateaued. This reflects different posterior-quality requirements: finding rate requires only roughly correct direction, whereas path efficiency benefits from a sharper posterior.



(a) Finding rate (b) Path efficiency

Figure 11. Source-inference performance versus training-data proportion in Cases 8–10

Figure 12 further shows that test blocks with higher forward-prediction accuracy also tend to exhibit higher finding rate and path efficiency. This pattern supports a consistent link between forward-model fidelity and pollution source-inference performance across different urban morphologies.



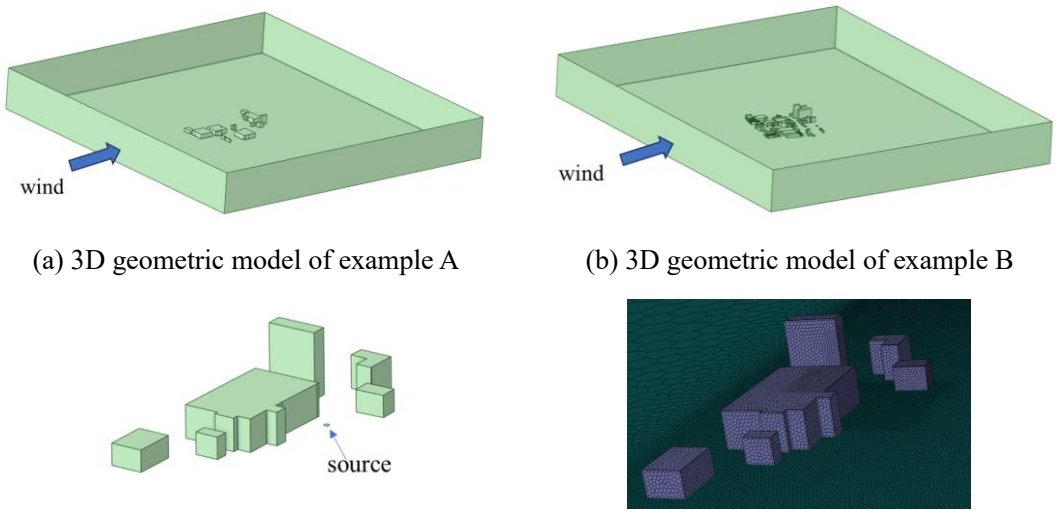
(a) Finding rate versus $R^2(\log)$ (b) Path efficiency versus $R^2(\log)$

Figure 12. Relationship between forward prediction and source-inference performance across blocks

4.3 External validation against independent Fluent simulations

To further assess the predictive capability of the proposed framework beyond the PMSS-based training data, two independent urban dispersion cases were constructed in ANSYS Fluent without fine-tuning the surrogate. A standard CFD setup was adopted,

including a boundary-layer inlet profile, pressure outlet, no-slip walls, local mesh refinement near buildings and source, and steady Reynolds-averaged Navier – Stokes (RANS) simulation with passive-scalar transport. Near-surface concentration fields at pedestrian height served as the reference fields (Figure 13).

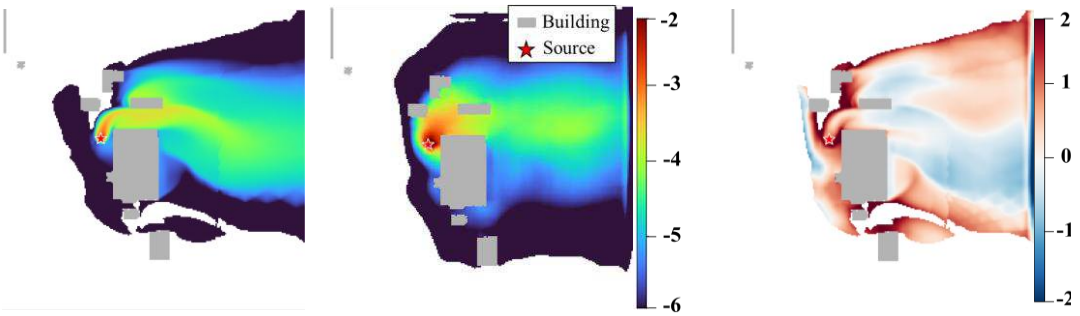


(a) 3D geometric model of example A (b) 3D geometric model of example B

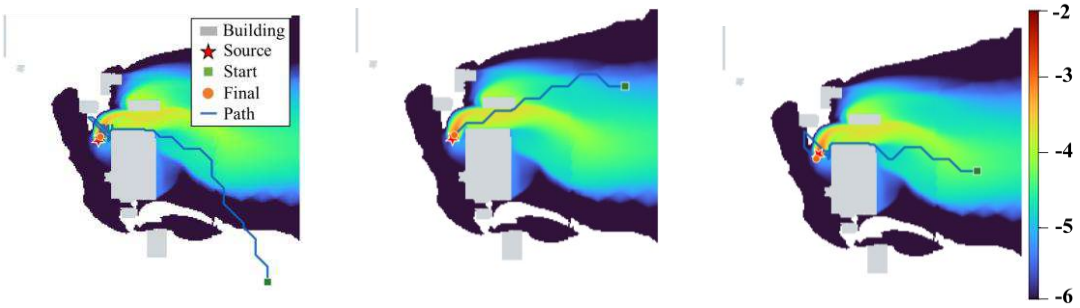
(c) local details of validation example A (d) computational mesh of example A

Figure 13. Fluent modeling details

Figures 14 and 15 compare the surrogate predictions with the Fluent reference fields for validation examples A and B, respectively. For example A, the model achieves R^2 (log) = 0.667, whereas for example B it achieves 0.572. In each example, 20 downstream initial positions are used for source-inference evaluation, with an arrival radius of 4 m. Both examples achieve a 100% finding rate. The median numbers of steps are 30 and 32, respectively, and the corresponding mean path efficiencies are 0.34 and 0.38. These results provide limited but encouraging preliminary cross-simulator evidence that the proposed framework remains functional in two external simulation cases not seen during training.

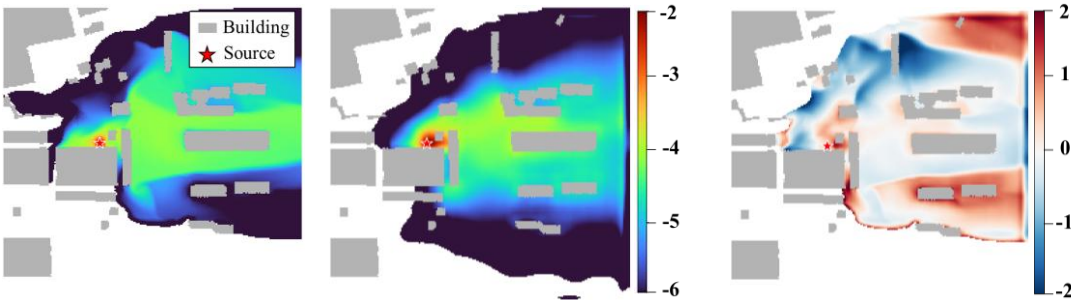


(a) Ground truth (b) Prediction (c) Path efficiency

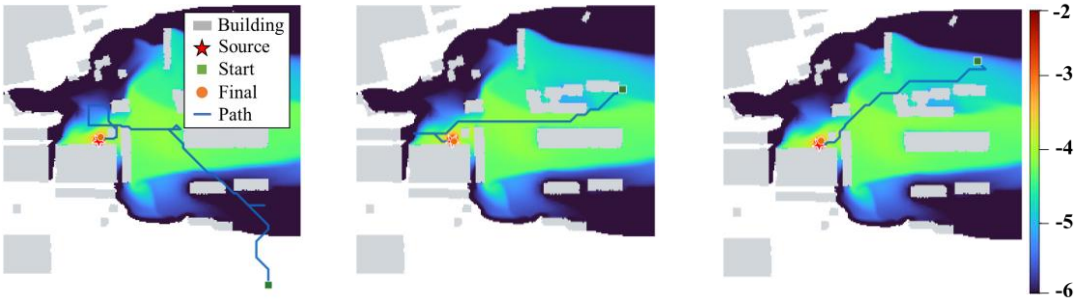


(d) Source search path 1 (e) Source search path 2 (f) Source search path 3

Figure 14. Fluent validation example A: surrogate prediction, source-search paths



(a) Ground truth (b) Prediction (c) Path efficiency



(d) Source search path 1 (e) Source search path 2 (f) Source search path 3

Figure 15. Fluent validation example B: surrogate prediction, source-search paths

4.4 Limitations

The present framework has several limitations. First, the framework is formulated for threshold-triggered source-inference under quasi-steady release and meteorological

conditions, which may not hold under rapidly varying winds, transient release events, or pre-detection exploration scenarios. Second, although the training data are generated from three-dimensional simulations, the surrogate model uses only two-dimensional building geometry and predicts near-surface concentration fields at pedestrian height, which limits its applicability to ground-level search. Third, the current formulation considers only a single release source and would require extension to handle multi-source scenarios. Fourth, most evaluations are based on simulation data, with only limited external validation using Fluent cases; field validation under real urban conditions is still needed. Finally, while the present analysis highlights building density as an important factor, other urban morphological parameters may also influence forward prediction and Pollution source-inference performance and deserve further investigation. Additionally, real sensor errors may exhibit censoring, temporal correlation, or heteroscedasticity beyond the assumed constant-variance Gaussian log-space model. The response-library construction scales with candidate-source and sensor pixel counts, so larger domains or finer resolutions would increase preparation cost.

5. Conclusions

This study proposed a physics-informed surrogate modeling framework for rapid prediction and source inference of a single ground-level pollutant release in complex urban environments under quasi-steady wind conditions. The framework integrates source-dependent feature encoding, deep-learning-based forward prediction, response-library construction, and Bayesian source inference into a unified computational workflow. Based on the present results, the main conclusions are as follows.

(1) The proposed physics-informed feature encoding substantially improves forward prediction and cross-layout generalization of pollutant dispersion. Across four surrogate architectures, the physics-informed inputs consistently outperform conventional input formulations, and DeepLabV3+ achieves the best overall performance, with R^2 (log) values ranging from 0.72 to 0.85. In the eight-fold leave-one-block-out evaluation, the average R^2 (log) reaches 0.80, indicating that source-dependent physical encoding can support transferable surrogate prediction in previously unseen urban environments.

(2) Forward-model fidelity plays a central role in source-inference performance. Under the same MAP-guided sampling rule, replacing the Gaussian-plume likelihood with the surrogate-based likelihood increases the finding rate from about 46% to about 96%, reduces the median number of steps from 99 – 104 to 20 – 33, and improves path efficiency from 0.11 – 0.12 to 0.37 – 0.54. These results indicate that, for source inference in obstructed urban environments, improving the fidelity of the computational likelihood model can be more important than modifying the sampling rule alone.

(3) The framework shows favorable data efficiency. In the tested settings, approximately 150 – 240 training samples are sufficient for the finding rate to exceed 95%, while the largest gains in forward prediction are obtained at low training-data proportions before the performance curves flatten. This suggests that source-dependent physics-informed encoding can improve sample efficiency and make surrogate-assisted source inference more practical when training data are limited.

(4) The proposed three-stage workflow provides high computational efficiency. Response-library construction for a new block requires less than 1 min, while each Bayesian inference step requires about 1 ms. This separation of offline training, response-library construction, and low-cost inference offers a practical computational compromise between model fidelity and computational efficiency for source-inference problems involving many candidate release locations.

(5) The current framework still has several limitations. It is restricted to single-source source inference under quasi-steady conditions with known geometry and wind information, and most evaluations remain simulation-based, with only limited external validation. In addition, the present surrogate predicts near-surface concentration fields from two-dimensional layout information, which constrains its applicability to ground-level inference scenarios. Future work will focus on extending the framework to more strongly time-varying conditions, multi-source releases, broader uncertainty treatment, and more extensive validation across independent simulators and field-relevant datasets.

Overall, the proposed framework provides an efficient route for rapid prediction and source inference of pollutant dispersion in complex urban environments where simplified dispersion models may be insufficient. The framework may also support future developments in urban environmental fluid mechanics, fast-response dispersion modeling, and data-driven inverse analysis.

Declarations

Ethical Approval

Not applicable.

Funding

This research was funded by the National Key Research and Development Program of China (Grant No. 2024YFC3808800).

Availability of data and materials

The dataset generated and/or analyzed during the current study are available from the corresponding author upon reasonable request

Reference

[1] Wu, X., Zhou, Y., Zhao, X., Wang, J., Hou, L., & Fu, M. (2026). Dynamic probabilistic risk assessment of urban gas pipeline leakage considering multitype uncertainties. *Process Safety and Environmental Protection*, 108415.

[2] Lin, Y., Huang, H., Wang, J., & Zhang, X. (2026). Sensor configuration optimization for source term estimation of time-varying emissions. *Sustainable Cities and Society*, 107186.

[3] Fernando, H. J. S., Lee, S. M., Anderson, J., Princevac, M., Pardyjak, E., & Grossman-Clarke, S. (2001). Urban fluid mechanics: air circulation and contaminant dispersion in cities. *Environmental fluid mechanics*, *1*(1), 107-164.

[4] Hutchinson, M., Liu, C., & Chen, W. H. (2018). Information-based search for an atmospheric release using a mobile robot: Algorithm and experiments. *IEEE Transactions on Control Systems Technology*, *27*(6), 2388-2402.

[5] Loisy, A., & Heinonen, R. A. (2023). Deep reinforcement learning for the olfactory search POMDP: a quantitative benchmark. *The European Physical Journal E*, *46*(3), 17.

[6] Vergassola, M., Villermaux, E., & Shraiman, B. I. (2007). 'Infotaxis' as a strategy for searching without gradients. *Nature*, *445*(7126), 406-409.

[7] Hutchinson, M., Oh, H., & Chen, W. H. (2018). Entrotaxis as a strategy for autonomous search and source reconstruction in turbulent conditions. *Information Fusion*, *42*, 179-189.

[8] Wang, L., & Pang, S. (2023). Robotic Odor Source Localization via End-to-End Recurrent Deep Reinforcement Learning. In *2023 Seventh IEEE International Conference on Robotic Computing (IRC)* (pp. 43-50). IEEE.

[9] Wang, L., Pang, S., & Li, J. (2020). Olfactory-based navigation via model-based reinforcement learning and fuzzy inference methods. *IEEE Transactions on Fuzzy Systems*, *29*(10), 3014-3027.

[10] Turner, D. B. (2020). *Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling*. CRC press.

[11] Shen, Z., Cui, G., & Zhang, Z. (2017). Turbulent dispersion of pollutants in urban-type canopies under stable stratification conditions. *Atmospheric Environment*, *156*, 1-14.

[12] Moon, K., Hwang, J. M., Kim, B. G., Lee, C., & Choi, J. I. (2014). Large-eddy simulation of turbulent flow and dispersion over a complex urban street canyon. *Environmental Fluid Mechanics*, *14*(6), 1381-1403.

[13] Xue, F., Kikumoto, H., Li, X., & Ooka, R. (2018). Bayesian source term estimation of atmospheric releases in urban areas using LES approach. *Journal of hazardous materials*, *349*, 68-78.

[14] Jia, H., & Kikumoto, H. (2021). Source term estimation in complex urban environments based on Bayesian inference and unsteady adjoint equations simulated via large eddy simulation. *Building and Environment*, *193*, 107669.

[15] Zhang, Y., Zhou, X., Zhu, J., & Kikumoto, H. (2026). Source term estimation based on the reduced-order adjoint equation. *Building and Environment*, 114197.

[16] Efthimiou, G. C., Kovalets, I. V., Venetsanos, A., Andronopoulos, S., Argyropoulos, C. D., & Kakosimos, K. (2017). An optimized inverse modelling method for determining the location and strength of a point source releasing airborne material in urban environment. *Atmospheric Environment*, *170*, 118-129.

[17] Zhao, J., Li, J., Bai, Y., Zhou, W., Zhang, Y., & Wei, J. (2024). Research on leakage detection technology of natural gas pipeline based on modified Gaussian plume model and Markov chain Monte Carlo method. *Process Safety and Environmental Protection*, *182*, 314-326.

[18] Liu, Z., & Li, X. (2024). A new sensor configuration design method for source term estimation in urban neighborhood with complex conditions under different wind directions. *Building and Environment*, *252*, 111210.

[19] Liu, Z., & Li, X. (2025). Influence of number of sensors on source term estimation performance in complex urban neighborhood. *Sustainable Cities and Society*, *121*, 106174.

[20] Ni, C., Lang, Z., Wang, B., Li, A., Cao, C., Du, W., & Qian, F. (2025). Data-driven source term estimation of hazardous gas leakages in complex chemical industrial parks. *Computers & Chemical Engineering*, 109530.

[21] Xu, J., Du, W., Xu, Q., Dong, J., & Wang, B. (2021). Federated learning based atmospheric source term estimation in urban environments. *Computers & Chemical Engineering*, *155*, 107505.

[22] Septier, F., Armand, P., & Duchenne, C. (2020). A bayesian inference procedure based on inverse dispersion modelling for source term estimation in built-up environments. *Atmospheric Environment*, *242*, 117733.

[23] Septier, F., Armand, P., & Duchenne, C. (2022). Sequential Monte Carlo sampler applied to source term estimation in complex atmospheric environments. *Atmospheric environment*, *269*, 118822.

[24] Albani, R. A., Albani, V. V., Migon, H. S., & Neto, A. J. S. (2021). Uncertainty quantification and atmospheric source estimation with a discrepancy-based and a state-dependent adaptative MCMC. *Environmental Pollution*, *290*, 118039.

[25] Clarke, A., Giljarhus, K. E. T., Oggiano, L., Saddington, A., & Depuru-Mohan, K. (2025). Deep learning for urban wind prediction: An MLP-Mixer approach with 3D encoding. *Building and Environment*, 113495.

[26] Calzolari, G., & Liu, W. (2025). Accelerating Large Eddy Simulations of Urban Airflow with Generative Adversarial Networks. *Building and Environment*, 113622.

[27] Milla-Val, J., Montañés, C., & Fueyo, N. (2024). Adversarial image-to-image model to obtain highly detailed wind fields from mesoscale simulations in urban environments. *Building and Environment*, *266*, 112123.

[28] Tahmasebi, S., Tian, G., Qin, S., Marey, A., Wang, L. L., & Rayegan, S. (2025). Using diffusion models for reducing spatiotemporal errors of deep learning based urban

microclimate predictions at post-processing stage. *Physics of Fluids*, *37*(3).

[29] Chen, C., Tian, G., Qin, S., Yang, S., Geng, D., Zhan, D., ... & Wang, L. (2026, January). Urban wind field prediction using Fourier Neural Operators across different wind directions and cities. In *Building Simulation* (pp. 1-15). Beijing: Tsinghua University Press.

[30] Shao, X., Liu, Z., Zhang, S., Zhao, Z., & Hu, C. (2023). PIGNN-CFD: A physics-informed graph neural network for rapid predicting urban wind field defined on unstructured mesh. *Building and Environment*, *232*, 110056.

[31] Wang, H., Ma, W., Niu, J., & You, R. (2025). A geometry-conditioned transformer solver for predicting three-dimensional wind distribution in urban microclimate design. *Urban Climate*, *64*, 102634.

[32] Wu, Y., Cao, Z., Lei, Y., Han, Y., Yuan, M., Wang, L., & Zhou, X. (2026). Multi-Constraint Physics-Informed Generative Adversarial Networks (PIGAN) Enable Small-Data Learning for Urban Wind Field Prediction. *Building and Environment*, 114202.

[33] Tian, G., Qin, S., Geng, D., Zhan, D., Yang, S., Liu, P., ... & Wang, L. L. (2025). AI-enhanced large-eddy simulation for urban wind flows: accelerating spin-up with Fourier neural operator initialization. *Building and Environment*, 114046.

[34] Chen, T., Li, R., Hu, X., Zhang, B., Liu, Y., Wang, L. L., & Gao, N. (2025). Machine learning as CFD surrogate models for rapid prediction of building-related physical fields: A review of methods and state-of-the-art. *Building and Environment*, 113667.

[35] Song, D., Lee, K., Phark, C., & Jung, S. (2021). Spatiotemporal and layout-adaptive prediction of leak gas dispersion by encoding-prediction neural network. *process safety and Environmental Protection*, *151*, 365-372.

[36] Cui, D., Liu, S., Xu, X., Lin, P., & Hu, G. (2025). Rapid prediction of pollutant dispersion in residential blocks using generative adversarial networks. *Urban Climate*, *62*, 102533.

[37] Mendil, M., Leirens, S., Armand, P., & Duchenne, C. (2022). Hazardous atmospheric dispersion in urban areas: A Deep Learning approach for emergency pollution forecast. *Environmental Modelling & Software*, *152*, 105387.

[38] Mendil, M., Leirens, S., Novello, P., Duchenne, C., & Armand, P. (2025). A 3D discrepancy modeling framework for urban pollution prediction in accelerated time. Environmental Modelling & Software, 106662

[39] Kim, H., Park, M., Kim, C. W., & Shin, D. (2019). Source localization for hazardous material release in an outdoor chemical plant via a combination of LSTM-RNN and CFD simulation. *Computers & Chemical Engineering*, *125*, 476-489.

[40] Gao, P., Jiang, J. C., Xu, Q., Wu, J., Tang, Y., & Huang, A. C. (2026). Inversion of the source of gas pipeline leaks using a gaussian plume optimization approach. *Journal of Loss Prevention in the Process Industries*, 105909.

[41] Jing, Q., Song, X., Sun, B., Li, Y., & Zhang, L. (2025). Efficient estimation of natural gas leakage source terms using physical information and improved particle filtering. Reliability Engineering & System Safety, 260, 110989.

[42] Newman, T., Nemeth, C., Jones, M., & Jonathan, P. (2025). Probabilistic inversion

modeling of gas emissions: A gradient-based MCMC estimation of Gaussian plume parameters. *The Annals of Applied Statistics*, *19*(4), 2937-2956.

[43] Hu, J., Cui, C., Gao, J., Zhang, A., Liu, X., Zhang, S., & Fang, Y. (2026). A Method for Locating Air Pollution Sources Combining Sensor Network and Machine Learning. *Water, Air, & Soil Pollution*, *237*(7), 421.

[44] Wang, Z., Wang, J., Yang, K., Wang, L., Su, F., & Chen, X. (2022). Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. *Computers & Geosciences*, *158*, 104969.

[45] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.

[46] Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, *24*(89), 1-97.

[47] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[48] Oldrini, O., Armand, P., Duchenne, C., Olry, C., Moussafir, J., & Tinarelli, G. (2017). Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. *Environmental Fluid Mechanics*, *17*(5), 997-1014.