Atmospheric Environment 323 (2024) 120388

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Two-step AI-aided Bayesian source identification of urban-scale pollution

Elissar Al Aawar, Mohamad Abed El Rahman Hammoud, Ibrahim Hoteit *

Physical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

HIGHLIGHTS

- AI-aided Bayesian inference can efficiently identify urban air pollution sources.
- A two-step MCMC sampling method is used to identify emission location and intensity.
- A DNN approximation of the W_2 distance is employed to model the likelihood.
- The AI-aided framework achieves a three-fold reduction in computational requirements.

ARTICLE INFO

Keywords:
Urban environment
Air pollution source identification
Multistage bayesian inference
W_2 distance
Deep learning

ABSTRACT

Poor air quality produces detrimental effects worldwide; hence, it is vital to thoroughly characterize air pollution sources and effectively address and mitigate such effects. Due to the complexity of the underlying physical processes, and uncertainties in the available observations, the air pollution source identification problem is typically cast within a Bayesian inversion framework. The latter incorporates prior knowledge and observations to characterize a release event through the posterior distribution of the source parameters. In this study, we rely on two-dimensional (2D) pollutant concentration distributions as observations, and adopt the Wasserstein (W_2) distance to model the likelihood probability distribution for given emission parameters. Since the posterior distribution is estimated via random sampling that involves many forward model runs, the Bayesian framework can be computationally prohibitive for realistic urban air pollution problems that are driven by computationally demanding micro-scale flow simulations. Furthermore, computing the W_2 distance is resource-intensive . In this context, we develop a computationally efficient Bayesian framework by following (i) a two-stage approach that reduces the cost of the Bayesian inversion, and (ii) an artificial intelligence (AI) approximation of the W_2 distance. In the two-stage approach, a low-resolution dispersion model is run in the first stage to propose representative samples of emission parameters for final selection by the original high-resolution model in the second stage. In addition, we approximate the W_2 distance using a deep neural network (DNN) to achieve an appreciable reduction in the computational cost with negligible loss in the inversion performance. We design numerical experiments to test the sensitivity of the inverse solution to the characteristics of the approximative model. The results indicate that pairing the two-stage approach with the DNN approximation of the W_2 distance preserves the quality of the inverse solution, while achieving at least threefold reductions in the computational cost.

1. Introduction

Characterizing air pollution sources is indispensable in urban settings since air quality levels adversely affect public health (Turner et al., 2020; Kelly and Fussell, 2015). A detailed description of urban pollution sources facilitates designing and implementing relevant risk assessment studies and response plans (Kaginalkar et al., 2023). Pollution sources are usually characterized by their location, emission rate, and duration. In this perspective, air quality monitoring involves collecting both discrete values and continuous distributions of air pollution levels. At measurement stations, sensors record the concentrations of different air pollutants at regular intervals, resulting in discrete values of pollutant concentrations at specific times and locations (Nychka and Saltzman, 1998; Spangl et al., 2007). Conversely, advanced technologies, such as satellites, remote sensing, and light detection and ranging (LIDAR) produce concentration maps that depict the distribution of pollutants across a region (Cooper et al., 2012; Emeis and Schäfer, 2006; Gaudio et al., 2015). Moreover, they provide a continuous representation of air quality levels and a comprehensive view of air pollution patterns over a broader area (Al Aawar et al., 2023).

* Corresponding author.
E-mail address: ibrahim.hoteit@kaust.edu.sa (I. Hoteit).

https://doi.org/10.1016/j.atmosenv.2024.120388
Received 13 November 2023; Received in revised form 21 January 2024; Accepted 1 February 2024
Available online 17 February 2024

Numerous techniques were implemented for identifying the source parameters of air pollution incidents. The most intuitive approach is through backward-in-time models, integrated from observation to emission times (Bocquet, 2007; Yassin et al., 2018; Koracin et al., 2011). Such models encounter challenges related to the non-linear nature of atmospheric processes, sensitivity to input data, and spatial–temporal resolution limitations (Seinfeld and Pandis, 2016; Kukkonen et al., 2012). Additionally, uncertainties in emission inventories, inadequate dispersion representation, and computational complexity contribute to an inaccurate source estimation (Kukkonen et al., 2012; Stockwell et al., 1990). Other deterministic optimization algorithms were adopted to provide a single prediction-data best-fit, which can be subject to substantial uncertainties (Uusitalo et al., 2015; Chen et al., 2017; Khlaifi et al., 2009). In contrast, stochastic techniques were utilized to handle complex and ill-posed problems by quantifying uncertainty through probability distributions. They typically apply Bayesian inversion to compute a distribution for the source parameters conditioned on available observations (Box and Tiao, 2011). Forward models are utilized within the Bayesian framework to quantify model-data discrepancies by incorporating prior knowledge and observational data, expressed as prior and likelihood distributions, respectively (Marzouk et al., 2007; Marzouk and Najm, 2009). The posterior distribution is then obtained through the product of the prior and likelihood distributions using Bayes rule (Iglesias and Stuart, 2014). Unlike deterministic approaches, the Bayesian framework explicitly considers both observational and model uncertainties, and estimates parameter distributions through the posterior rather than single optimal values (Tikhonov et al., 2013). To evaluate the posterior distribution, random sampling strategies, such as the Markov Chain Monte Carlo (MCMC) algorithm, are commonly employed (Gamerman and Lopes, 2006). These strategies involve accepting or rejecting random samples from the prior based on predefined criteria. The accepted MCMC samples then form the posterior distribution, providing a probabilistic description of the source parameters (Chen et al., 2012).

Bayesian inference using MCMC is widely used in a range of applications, including source identification of marine pollution from oil spill contours (El Mohtar et al., 2021), radioactive release from real continental-scale measurements (Delle Monache et al., 2008), and groundwater contamination from field experiment data (Wang and Jin, 2013), among others. Several studies have also examined Bayesian inversion for source identification in urban applications using simple air pollution models such as simple Gaussian puff models (Huang et al., 2015), complex Eulerian dispersion models (Chow et al., 2008), or precomputed source-receptor pairs (Keats et al., 2007; Rajaona et al., 2015; Guo et al., 2009). These studies often rely on observational data in the form of discrete pollutant concentration measurements from either actual sensor readings (Chow et al., 2008) or synthetic scenarios (Keats et al., 2007). Gaussian likelihood distributions are most often adopted, using the L_2 norm as a discrepancy measure between the reference observations and model outputs (Keats et al., 2007). Recently, Al Aawar et al. (2023) considered observations in the form of pollutant concentration distributions rather than discrete values in the likelihood. Such distributions could be acquired using remote sensing techniques (Tratt et al., 2017; Buckland et al., 2017) or via an optimal interpolation of sensor measurements (Xu et al., 2022; Berrocal et al., 2020). In addition, Al Aawar et al. (2023) considered a realistic high-resolution Lagrangian dispersion model to properly resolve microscale flows and pollutant dispersion patterns of the urban environment. An exponential likelihood distribution was adopted with the W_2 distance as a model-to-observation discrepancy measure (Monge, 1781). The W_2 distance stems out from the optimal transport theory, and measures discrepancies between two distributions with certain conservation constraints (Villani, 2003). It is calculated by minimizing the path required to translate one distribution to the other (Lévy and Schwindt, 2018). However, the W_2 distance is computationally demanding, which escalates the cost of the underlying Bayesian inversion.

Several strategies were proposed to reduce the computational cost of Bayesian inference. This includes using a computationally cheaper surrogate of the forward model (Sraj et al., 2014, 2016). By replacing the complex and time-consuming forward model with a computationally less-demanding approximate model that retains essential features, the computational burden can be significantly reduced while still providing reasonably accurate results. Another strategy is to simplify the source identification problem by inverting for a subset of the model parameters (DiazDelaO et al., 2017), such that the focus is placed on determining a subset of the most crucial parameters instead of estimating all model parameters simultaneously. Reducing the parameter space diminishes the computational complexity, allowing for quicker and more feasible computations. Moreover, efficient sampling strategies were developed to address the time-consuming nature of traditional MCMC methods (Zhang, 2021). These techniques require fewer iterations to approximate the posterior distribution, reducing computational demands, and making Bayesian inference more feasible and accessible for complex problems and large datasets (Brooks et al., 2011). The Adaptive Metropolis (AM) and the Hamiltonian Monte Carlo (HMC) algorithms are examples of such methods. The AM method dynamically adjusts the proposal distribution during sampling, leading to faster convergence and improved exploration of the parameter space (Haario et al., 2001). Also, the HMC process, inspired by physics, simulates Hamiltonian dynamics to explore the parameter space effectively, producing less correlated samples and faster convergence (Neal et al., 2011). Other approaches rely on improving the computational time associated with accepting or rejecting the MCMC samples via the so-called multi-stage inference method (Christen and Fox, 2005). The latter represents an iterative process that involves refining source parameter estimates in multiple steps (Higdon et al., 2002). During preliminary stages, it adopts an initial set of prior beliefs to obtain a posterior distribution by a cheap approximation of the forward model. In subsequent stages, this posterior distribution serves as the new prior, allowing for continuous updates and improvements to the parameter estimates using the original forward model. This approach allows for reducing the uncertainty over multiple stages, and diminishing exhaustive model runs, which leads to faster and more efficient inversion (Ma et al., 2008; Stuart et al., 2019).

AI has recently emerged as a powerful tool in advanced air pollution source identification. Machine learning models have demonstrated their capacity to capture complex relationships among diverse factors influencing air quality, such as meteorological conditions, traffic patterns, and industrial emissions (Balogun et al., 2021; Rowley and Karakuş, 2023). One approach, the RSRF model, integrates aerosol optical depth as a remote sensing (RS) product and Random Forest (RF) for predicting ground-level $PM_{2.5}$, offering a valuable tool for pinpointing and understanding the sources of air pollution in a given region (Li and Zhang, 2019). Another study couples deep learning and physical modeling to efficiently estimate source location and emission profiles in street canyons, using measurements from air pollution monitoring stations (Zhou et al., 2022). This involves combining a DNN model, trained with data generated from numerical simulations, with an inverse method based on Tikhonov regularization to identify the source parameters. Further advancements in deep learning enable the analysis of satellite imagery to detect and monitor pollution sources (Wei et al., 2021). For instance, a recent study developed a deep learning approach to detect regional ammonia point sources from high resolution satellite imagery and data (Lei et al., 2021). Another development is AI-powered drones equipped with environmental sensors, which proved to be efficient in conducting aerial surveys of industrial areas and remote regions (Chodorek et al., 2021). This technology provides essential data on air quality and pollution sources that are otherwise challenging to access.

In this work, we propose a computationally efficient Bayesian inference framework for inferring an air pollution source in a realistic urban setting. The framework expands the work of Al Aawar et al. (2023),

which addressed the source identification of urban pollution within a Bayesian inversion framework using concentration distributions as observational data. Furthermore, we rely on Al Aawar et al. (2023) as a benchmark to assess reductions in computational costs. A two-stage MCMC algorithm is implemented here by first running a realistic Lagrangian atmospheric dispersion model on a low-resolution spatial grid to preselect the MCMC samples. The model is then integrated on a high-resolution spatial grid for final selection of the samples that go into the posterior. Furthermore, we employ the DNN proposed by Courty et al. (2017) to alleviate the computational burden associated with the evaluation of the W_2 distance in the likelihood. This network relied on a Siamese architecture (Chicco, 2021), which takes two input fields and propagates them through two neural networks with shared parameters. The network architecture is based on an encoder–decoder, where the W_2 distance is embedded within the norm of the encoder outputs, and the decoder ensures that the identity of the inputs is preserved in the reconstruction. We conduct a comprehensive series of numerical experiments to assess the effectiveness of the proposed approach for identifying air pollution sources. Additionally, we investigate the sensitivity of the algorithm to specific model features that have a significant impact on computational costs; particularly the spatial model resolution and the inclusion/exclusion of buildings, during the first MCMC stage of the algorithm.

The study is organized as follows. Section 2 presents the urban atmospheric air pollution dispersion model. Section 3 describes the proposed Bayesian inference framework. Section 4 describes the architecture of the DNN used to evaluate the W_2 distance, and its training process. Section 5 outlines the numerical experiments, whose results are analyzed in Section 6. Finally, the main takeaways are outlined in Section 7.

2. Urban air pollution model

We implement the urban air pollution dispersion model Graz Mesoscale Model-Graz Lagrangian Model (GRAMM-GRAL) (Landesregierung, 2019, 2018) to simulate a realistic urban pollution case study, set at the King Abdullah University of Science and Technology (KAUST) campus. Physical dispersion models provide a great potential in solving the challenging inverse problem of air pollution source identification, by recreating realistic pollution dispersion scenarios that closely mirror real-world conditions in the target area (Zannetti, 2013). The GRAMM-GRAL model comprises two components, GRAMM, a mesoscale wind field model, and GRAL, a microscale wind field and Lagrangian particle transport model. These models are configured on two uniformly meshed nested domains, and are driven by the initial and boundary conditions of wind speed, direction, and atmospheric stability class. The flow patterns around buildings, in a defined topographic and urban configuration, are simulated by solving the Navier–Stokes equation with the k-ϵ turbulence closure model (Landesregierung, 2019). Then, passive tracer Lagrangian transport is employed to simulate the dispersion of pollutants and compute the resulting concentration distribution corresponding to a given release source (Landesregierung, 2018).

In our implementation, we decouple GRAMM from GRAL, such that mesoscale wind fields are precomputed by GRAMM and stored as a fixed dataset, before running the MCMC algorithm. This dataset was simulated by GRAMM based on the topography, land surface features, and the prevailing meteorological conditions. After starting the MCMC algorithm, GRAL accesses the GRAMM dataset of precomputed flow fields to simulate the dispersion of pollutants on its micro-scale grid, considering the presence of buildings and other urban features. Hence, GRAL is considered the only forward model in the present study. GRAL computational domain covers KAUST campus located along the eastern coast of the Red Sea, 80 km north of Jeddah, Saudi Arabia. Fig. 1 illustrates the geographic location of the inference domain comprising the nested domains used by GRAMM-GRAL in the UTM coordinate system of latitude and longitude (zone 37N). The topography and land use data of the region were extracted from the digital elevation model of the ground elevation and coordination of information on the environment (CORINE) values (CORINE, 2018). Also, KAUST buildings were manually crafted into the urban model using their original architectural layouts.

We designed the emission scenario according to a synthetic release inspired from a fire accident at an electric power station in the nearby-to-KAUST city of Rabigh in 2021 (Rabigh, 2023). We used the location of the burning oil tank to set a point source 150 km north of Jeddah, at x = 505,010 m and y = 2,489,705 m (Fig. 2). The $PM_{2.5}$ pollutant concentration was monitored, and the emission was assumed to last for d = 120 min with a constant emission rate q = 200 kg/h. Such emission settings cause a substantial mean pollution concentration over KAUST that exceeds the World Health Organization (WHO) Air Quality Guidelines (Organization et al., 2021). We force GRAMM with meteorological conditions that directly channel the plume towards KAUST. These conditions were determined from hourly wind data between 2016 and 2018, extracted from the regional configuration of the Weather Research and Forecasting (WRF) model (Skamarock et al., 2019). The adopted wind pattern exhibits a northwest direction of 344°, an average speed of 4.5 m/s, and corresponds to atmospheric stability class C (slightly unstable as defined by Pasquill, 1961). Readers are referred to Al Aawar et al. (2023) and Dasari et al. (2020) for a more detailed description of the forecasting model environment.

3. Bayesian inference

3.1. Modeling the likelihood

We solve the air pollution source identification inverse problem by estimating four emission parameters. We denote the solution by $\mathbf{M}$, comprising the longitude (x) and latitude (y) of the release source, its emission rate (q), and duration (d). GRAL, the forward model, takes $\mathbf{M}$ as input and predicts the resulting concentration distribution, denoted by $\mathbf{D}$. Bayes' rule estimates the posterior of the source parameters, as Box and Tiao (2011):

$$\pi(\boldsymbol{M}|D) \propto \pi(\boldsymbol{M})\pi(D|\boldsymbol{M}), \tag{1}$$

where $\pi(\boldsymbol{M}|D)$ is the posterior distribution, $\pi(\boldsymbol{M})$ is the prior, and $\pi(D|\boldsymbol{M})$ is the likelihood. The maximum a posteriori (MAP) solution is then considered the best estimate of the release parameters (Bassett and Deride, 2019).

Here, we assume no prior knowledge of $\mathbf{M}$, so that the prior distribution is a four-dimensional uniform distribution, which results in a posterior directly proportional to the likelihood. Table 1 outlines the bounds of the uniform prior distribution of $\boldsymbol{M}$. The longitude (x) and latitude (y) of the release source are given on the Universal Transverse Mercator (UTM) reference. The emission rate (q) is provided in kilograms per hour (kg/h), and the emission duration (d) is in minutes. Following Al Aawar et al. (2023), we invert for pollutant concentration distributions by considering an exponential likelihood function of the form:

$$\pi(D|M) = \frac{1}{\beta}\exp\left(\frac{-\delta}{\beta}\right), \tag{2}$$

where β is a scaling parameter, and δ is a measure of the model-data discrepancy.

We investigate different metrics for computing δ, including the Wasserstein (W_2), Hellinger (H), total variation (TV) distances, and the Kullback–Leibler (KL), and Jensen–Shannon (JS) divergences. Note that all tested measures compute distances between two distributions, which is consistent with the nature of the observations (Ciaccia et al., 1997; Sgarro, 1981). This analysis is conducted to explore the impact of different metrics on the likelihood estimation, and to assess their effectiveness in capturing the model-data agreement. We single out the

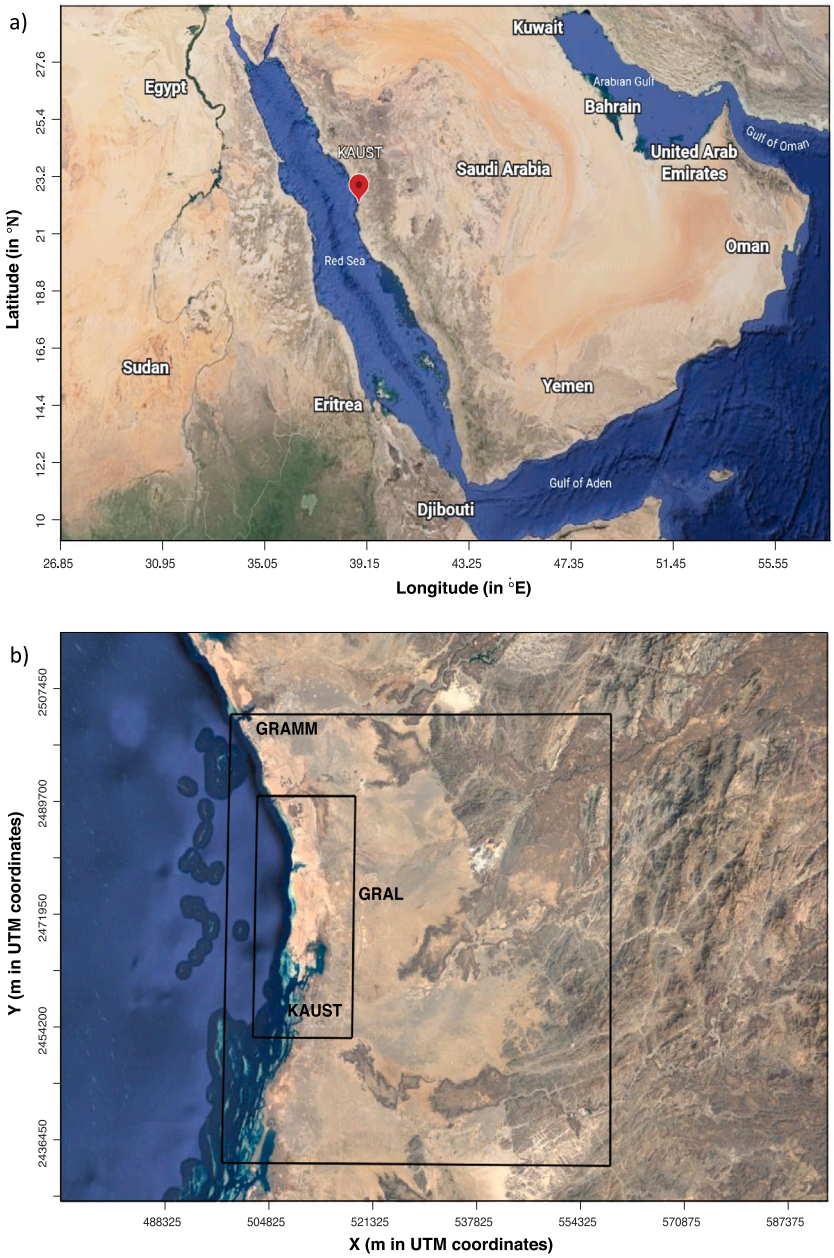

Fig. 1. Geographic location of the study area. (a) A red pin marks KAUST in a latitude–longitude coordinate system. (b) Outer and inner black rectangles denote GRAMM and GRAL domains, respectively, which are georeferenced in the UTM coordinate system. KAUST occupies the lower part of the GRAL domain.

Table 1
Lower and upper bounds of the uniform prior probability distribution.

Inverted parameter	Lower bound	Upper bound
x [UTM in m]	502,080	519,080
y [UTM in m]	2,463,680	2,501,960
q [kg/h]	10	300
d [min]	0	180

W_2 distance as the best-performing metric in modeling the likelihood, such that it results in the most representative posterior distribution for the emission parameters compared to other measures. The full details about the evaluation of alternative metrics are provided in Appendix A. Our conclusion further confirms the results of Al Aawar et al. (2023) who demonstrated that the W_2 distance achieves a better inverse solution than the L_2 distance, as typically used in the literature. The L_2 distance measures the Euclidean norm between n reference observations O_i and their corresponding model predictions P_i, and is defined as:

$$L_2 = \sqrt{\sum_{i=1}^{n} (O_i - P_i)^2}. \tag{3}$$

The selected W_2 distance could be interpreted as the minimum energy cost of moving and transforming one probability distribution to the shape of another distribution (Villani, 2003). Mathematically, the W_2 distance is obtained by solving a minimization problem over the space of transport functions while preserving certain conservation constraints (Monge, 1781). The W_2 distance between the normalized model outputs and reference observations is expressed as:

$$W_{2(p,q)} = \sqrt{\min_{\tau^* \epsilon \mathcal{T}(p,q)} J(\tau)}, \tag{4}$$

$$\text{with} \quad J(\tau) = \int L(u,v)p(u)du.$$

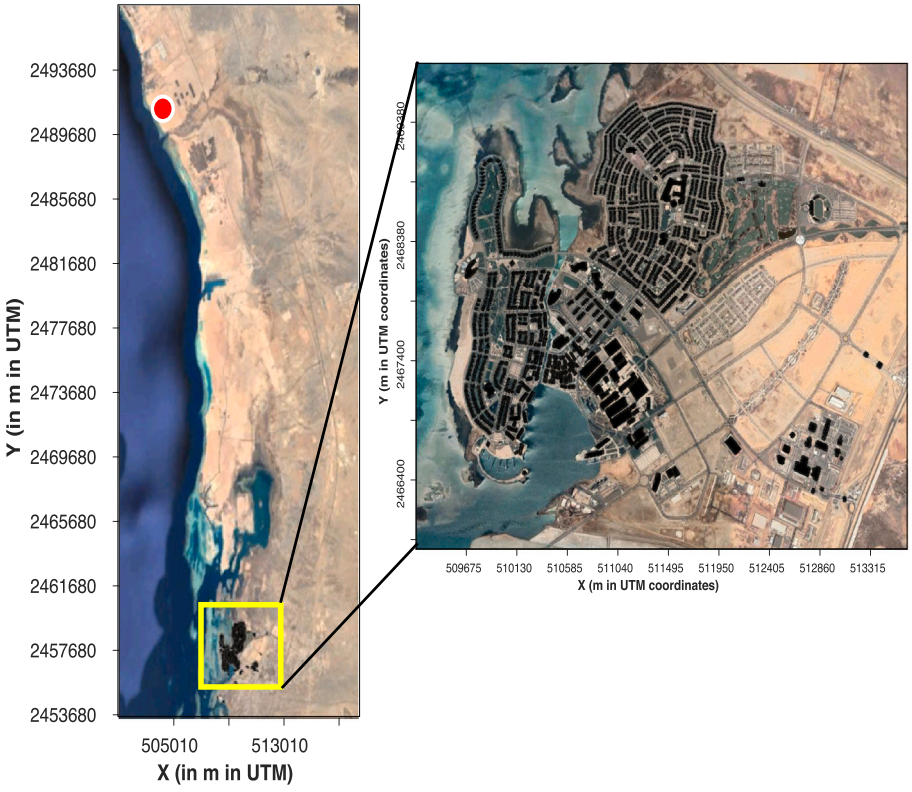

Fig. 2. Features of the experimental emission scenario. The left subplot represents the GRAL domain. The red dot marks the source location, and the yellow rectangle delimits KAUST campus, as depicted to the right.

where p and q are two distributions, $\mathcal{T}$ is the set of all sufficiently regular bijections τ mapping p to q, τ^* is the unique bijection that minimizes J, and L is a symmetric cost function (most commonly the Euclidean distance) that moves a unit from point u to point $v = \tau(u)$ (Villani, 2009; Ma et al., 2005).

3.2. The two-stage MCMC algorithm

The Metropolis–Hastings (MH) algorithm is a widely used MCMC method and a powerful tool for Bayesian inference (Chib and Greenberg, 1995). A single-stage MH-MCMC algorithm generates a chain whose asymptotic distribution is the posterior distribution (Rosenthal et al., 2011). The algorithm proceeds iteratively, starting with an initial state $\mathbf{M_{n-1}}$. It then proposes a new state $\mathbf{M_n}$ using a transitional probability distribution $f(\mathbf{M_n}|\mathbf{M}_{n-1})$. GRAL is integrated with the proposed model parameters, which are then evaluated through a suitable accept/reject criteria. The acceptance ratio is calculated between the posterior probability of the new and current states. If accepted, the new state becomes the current state for the subsequent iteration; otherwise, the current state remains unchanged. This process is repeated for a predefined number of iterations or until convergence is achieved.

Despite its advantages, this process is computationally expensive because it requires a full-forward-model run for each set of proposed emission parameters. Since 40% to 90% of the proposed samples are generally rejected (Rosenthal et al., 2011), the multi-stage MCMC methodology (Efendiev et al., 2006; Giles, 2008; Heinrich, 2001) was introduced to limit the number of model integrations by evaluating the sample first with a computationally less demanding approximative model. We adopted a two-stage MH-MCMC algorithm which operates as follows:

1. **Proposal generation:** Generate a sample of model parameters $\mathbf{M}'_n$ through a random walk sampling algorithm, i.e. using a normal distribution centered about the previous sample $\mathbf{M}_{n-1}$ with a standard deviation $\sigma_{sampler}$:

$$\boldsymbol{M}'_n \sim \mathcal{N}(\boldsymbol{M}_{n-1}, \sigma_{sampler}). \tag{5}$$

2. **Approximative and original model steps:** We denote by i the number of the current MCMC stage, such that $i = 1$ refers to the first stage (filter step), and $i = 2$ indicates the second stage (original model step). The proposed sample is

$$\boldsymbol{M}_{n,i} = \begin{cases} \boldsymbol{M}'_n & \text{with probability } \alpha_i(\boldsymbol{M}_{n-1}, \boldsymbol{M}'_n) \\ \boldsymbol{M}_{n-1} & \text{with probability } 1 - \alpha_i(\boldsymbol{M}_{n-1}, \boldsymbol{M}'_n), \end{cases} \tag{6}$$

where

$$\alpha_i(\boldsymbol{M}_{n-1}, \boldsymbol{M}'_n) = min\left(1, \frac{\pi_i(\boldsymbol{M}'_n|D)}{\pi_i(\boldsymbol{M}_{n-1}|D)}\right), \tag{7}$$

For $i = 1$, $\pi_1(\mathbf{M}|D) \propto \pi_1(D|\mathbf{M})$ is the approximate posterior distribution of the model parameters $\mathbf{M}$. Whereas, for $i = 2$, the approximate posterior acts as a prior of the original model step, such that $\pi_2(\mathbf{M}|D) \propto \pi_2(D|\mathbf{M})\pi_1(\mathbf{M}|D)$. If $\mathbf{M_{n,1}} = \mathbf{M}_{n-1}$, the fine-resolution model will not be integrated, resulting in a significant acceleration of the accept/reject process. In our implementation, we run GRAL on a coarse grid in the first stage, and on a fixed 40×40 m grid in the second stage, i.e. after the proposed sample undergoes the filter step. In addition to driving the inversion process at a lower computational cost, this approximation is proved to achieve convergence of the approximate posterior to the true one (Marzouk and Xiu, 2009). The rate of convergence mainly depends on the properties of the forward model; yet, remarkable accelerations can be achieved for relatively simple low-dimensional models (Laloy et al., 2013).

4. Deep Wasserstein embedding with a Siamese network

Siamese neural networks have been developed to compare two separate data samples (Chicco, 2021). Recently, a Siamese network was proposed to approximate the squared W_2 between two images by embedding their features in a lower-dimensional space (Courty et al., 2017). Specifically, the network was trained so that the L_2^2 distance between the features in a lower-dimensional space approximates the W_2^2 distance between the images in the original space. Originally designed for metric and similarity learning, the Siamese architecture was selected because it offers an efficient approach that enables meaningful comparisons between two distinct inputs through a replicated neural network structure (Roy et al., 2019). This design involves two branches, each receiving a different input, and the inputs are propagated through the network to learn a mapping to a new space. In this space, similarities and differences between inputs can be effectively compared and contrasted (Courty et al., 2017). Based on that, Siamese neural networks excel in metric learning tasks by naturally generating an embedding space that facilitates meaningful comparisons between data (Courty et al., 2017), aligning seamlessly with the requirements of our study, where the W_2 distance involves assessing the dissimilarity between distributions. Furthermore, a Siamese architecture directly satisfies the symmetry condition of the distance measures and was shown to offer considerable acceleration when computing the W_2 distance (Courty et al., 2017).

4.1. Siamese network pipeline

The employed neural network consists of a convolutional neural network architecture comprising an encoder followed by a decoder. The two distributions being compared are first propagated through the encoder, resulting in one feature space embedding for each input. Courty et al. (2017) demonstrated that a Siamese neural network could be trained such that the L_2 distance between the features embedded by the encoder equals the W_2 distance. Furthermore, to ensure that the neural network preserves the identity of the inputs, a decoder is trained to map the embedded features back to the original image space.

Mathematically, we represent the encoder by the γ function that maps an input x to a lower-dimensional feature space. The W_2^2 distance could be approximated as $\left\|\gamma(x^1)-\gamma(x^2)\right\|_2^2$ using the so-called contrastive loss, such that x^1 and x^2 are the two inputs. A decoder network is represented by the ϕ function that maps the features back to the image space, ensuring that $KL(\phi(\gamma(x^1)), x^1)$ and $KL(\phi(\gamma(x^2)), x^2)$ remain small, and that the image is preserved by the network. The Siamese neural network was trained end-to-end using the following total loss function:

$$L_c = L_{embed} + \lambda L_{recons}, \tag{8}$$

$$\text{with} \quad L_{embed} = \sum_i \left\| \left\|\gamma(x_i^1)-\gamma(x_i^2)\right\|_2^2 - y_i^2 \right\|_2^2, \tag{9}$$

$$\text{and} \quad L_{recons} = \sum_i \left(KL(\phi(\gamma(x_i^1)), x_i^1) + KL(\phi(\gamma(x_i^2)), x_i^2)\right), \tag{10}$$

where L_c, L_{embed}, and L_{recons} represent the contrastive, embedding and reconstruction loss terms, respectively. The reconstruction loss is weighted by a positive constant λ, and i is the dimension of the input. In our case, x^1 and x^2 represent the observed and simulated concentration distributions. In addition, L_{embed} is expressed as the squared L_2 norm of the difference between the Euclidean distance of the embedded features and the actual W_2^2 distance, denoted by y^2. L_{recons} quantifies the dissimilarity between the reconstructed and original concentration distributions using the KL divergence and acts as a regularization term (Yu et al., 2013). Fig. 3 presents a visual illustration of the training pipeline of the implemented network.

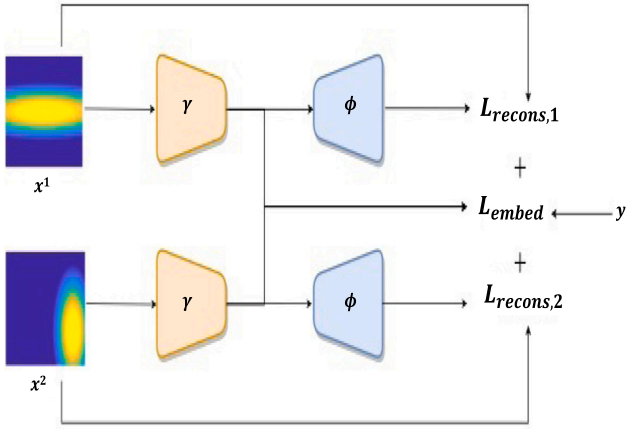

Fig. 3. Network pipeline. The used symbols are based on Eq. (10).

4.2. Network architecture and training

A supervised training approach was adopted to train the Siamese network to approximate the W_2 distance. We generate input data consisting of pairs of 2D pollutant concentration distributions, with their actual W_2 distances being used for supervision. Data generation involved sampling the source parameters from a representative distribution to ensure diversity and comprehensive coverage of the target domain. Afterwards, these parameters were fed into GRAL to create the corresponding concentration distributions. This approach allows the resulting distributions to capture a wide range of scenarios, enhancing the network's generalization and performance on unseen data. The input pollutant concentration distributions were normalized using a linear affine transformation into the range $[0, 1]$. Pairwise W_2 distances were computed *a priori* through the fluid dynamics formulation of the optimal transport problem (Farchi et al., 2016), serving as the actual W_2 distance for supervising the training process. The encoder comprises nine sequential blocks, each composed of a convolutional layer followed by a rectified linear unit (ReLU) nonlinear activation function and a batch normalization layer (Ioffe and Szegedy, 2015). The convolutional layers have a fixed kernel size of 9×9, and the number of channels gradually varies with depth, as depicted in Fig. 4. The fixed kernel size was determined after a comprehensive hyperparameter search over various kernel sizes. After experimenting with different sizes of $3\times3, 5\times5, 7\times7, 9\times9$, and 11×11, the 9×9 configuration emerged as the best choice in terms of minimizing the validation error. The selected size consistently demonstrated superior performance on the validation set, providing a trade-off between capturing complex features in the data, preventing overfitting, and improving generalization to unseen data. In this context, using a fixed kernel size proves beneficial as it captures information in small patches, but over several layers. This enables the network to obtain a more comprehensive understanding of the data, and achieve a global feature extraction by considering the cumulative effect of multiple convolution layers. Moreover, such a fixed kernel size contributes to computational efficiency, and provides stability and consistency in our model's behavior (Agrawal and Mittal, 2020). The decoder mirrors the encoder with transposed convolutions replacing the convolutional layers.

A dataset comprising 1.1 million pairs of concentration distributions and their corresponding W_2 distance values was generated. We employed an 80%–20% split ratio for partitioning the dataset into training and validation sets. Specifically, 80% of the data was allocated for training the model, while the remaining 20% was reserved for validation purposes, resulting in a training dataset of 900,000 samples and a validation dataset of the remaining 200,000 samples. The adoption of this split ratio is guided by the large size of our dataset,

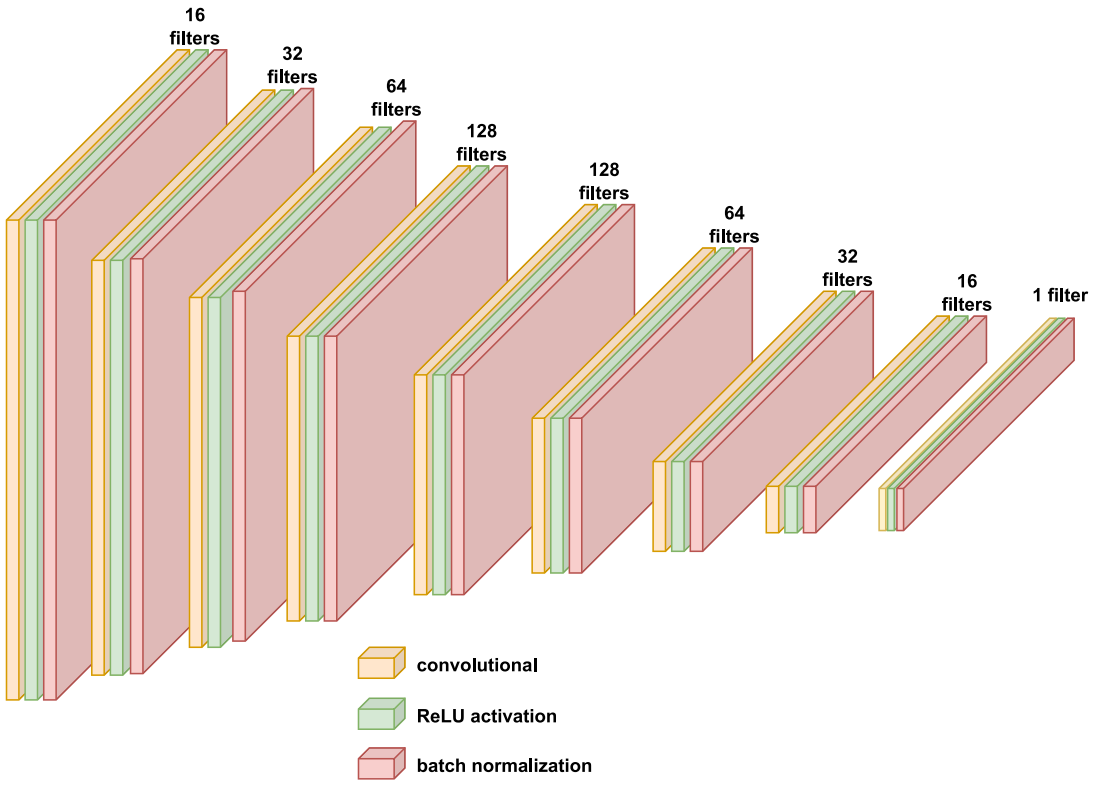

Fig. 4. Embedding network architecture. This same structure is mirrored across the decoder but with transposed convolutions instead of convolutional layers.

which allows us to allocate a significant portion for validation, while still providing the model with an adequate amount of training data. We also implemented a random and unbiased splitting approach that contributes to the representation of diverse scenarios present in our dataset, and guarantees that the model is evaluated on a statistically significant sample of the data. Such an approach enables us to draw reliable conclusions about the model's generalization performance to unseen data, thereby increasing our confidence in the reported results.

The loss function was minimized using the ADAM optimization algorithm with a weight decay of 0.0001 and an initial learning rate of 0.0001. The learning rate was adaptively modified using the *reduce on plateau* function, which drops the learning rate by a factor of 10 if the training loss does not decrease for 25 consecutive training epochs. The number of epochs refers to the number of times the learning algorithm goes through the entire training dataset. A hyperparameter search was performed for the batch size, training epochs, and the weight factor λ (see Eq. (10)). The best Siamese network was chosen as the one with the lowest L_{embed}, because a low embedding loss reliably reflects accurate W_2 estimates. Specifically, it corresponds to a batch size of 32, a pseudo-epoch size of 6400, and a λ of 0.01 with 200 training epochs.

During training, the network performance was monitored through L_c, L_{embed} and L_{recons}, with particular attention on L_{embed} because it evaluates the network performance in estimating the W_2 distance. The network is implemented through PyTorch and trained on one NVIDIA Tesla V100 32 GB GPU. The overall time required for training the Siamese network was approximately 5 h. We represent the training and validation per-epoch averaged values of the loss terms in the subplots of Fig. 5. The loss decreases with epochs, which is expected because training saturates, the longer one trains the network. The training L_{embed} plateaus around epoch 120, indicating that the network has reached its maximum learning capacity for the current configuration. In addition to its high accuracy, a trained neural network model needs only 20 ms to compute the W_2 distance between two images using a single GPU and 1 s when running the computations on a CPU.

To further examine the performance of the network, we employ the ten-fold cross-validation method (Kohavi et al., 1995). This technique involves shuffling the data, and partitioning the dataset into ten subsets or folds. The model is iteratively trained ten times, with nine folds chosen for training and the remaining one held for validation in each iteration. On each iteration of cross-validation, a new model is trained independently of the model trained during the previous iteration. The model is then validated on the corresponding validation fold, ensuring that each fold is used exactly once as a validation subset. Validation errors are recorded in each iteration, before being averaged and reported as the final performance error. We use the normalized L_2 error to evaluate the accuracy of our model's predictions against true W_2 distances, as defined by:

$$\text{Normalized } L_2 \text{ Error} = \frac{1}{N}\sum_{i=1}^{N}\frac{\|\hat{y}_i - y_i\|_2}{\|y_i\|_2}, \tag{11}$$

where $\hat{y}_i$ represents the predicted W_2 distances, y_i represents the true W_2 distances, and N is the total number of samples. After iterating over the ten folds, we observe a normalized L_2 validation error of 4.35%. The reported validation error indicates the model's effective approximation of the W_2 distance, demonstrating accuracy and consistency across multiple dataset partitions.

5. Numerical experimental design

We designed the experimental setup using synthetic observations in the form of the concentration distribution that was generated by the simulated emission scenario described in Section 2. The observation was collected 1 h after the emission stopped at an elevation of 10 m above the ground level. Fig. 6 presents the prevailing flow field and the observed concentration distribution. We considered a reference observation that covers the urban area of KAUST campus within the yellow rectangle in Fig. 6-b. The observed pollution concentrations were perturbed by a random noise to mimic realistic conditions. The observational noise is assumed unbiased and normally distributed

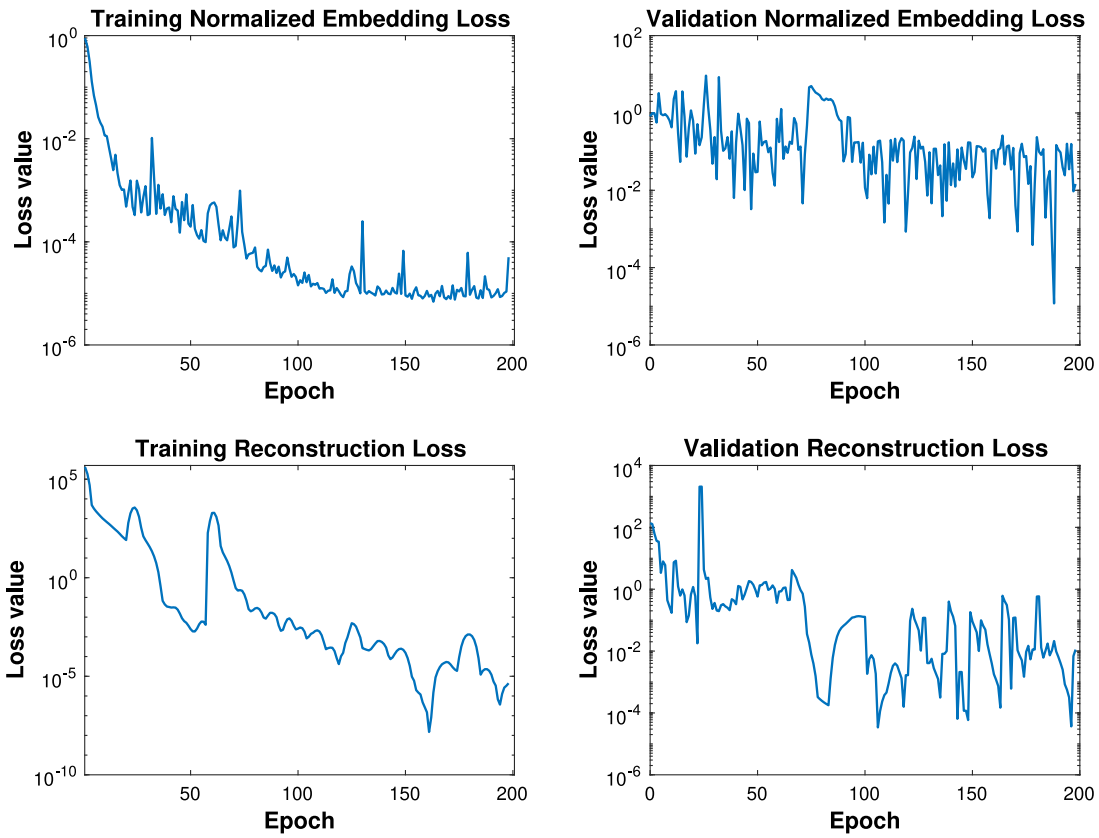

Fig. 5. Per-epoch averaged values of the loss terms in the training and validation stages. Embedding loss values are normalized by the value of the original label.

with a standard deviation of the same order as the mean pollutant concentration over the observed area.

Several numerical experiments were conducted to investigate the efficiency of the proposed pollution source identification framework. Exps. 1 and 2 were conducted using the conventional W_2 distance, with Exp. 1 employing the traditional single-stage MCMC approach, and Exp. 2 adopting a two-stage approach. This facilitates a direct comparison of the performance and computational aspects between the two approaches. In Exp. 3, a single-stage MCMC algorithm was used along with the neural network W_2 approximation. Exps. 1 and 3 serve as the baseline to evaluate the efficiency of the AI-based W_2 distance approximation. Note that in all experiments using the single-stage approach, GRAL simulates the dispersion of $PM_{2.5}$ pollutants on a fine 40×40 m mesh. The two-stage MCMC approach was fully explored in Exps. 4-a through 8-a using different spatial resolutions of the approximative model in the filter stage. On top of varying the spatial resolutions, Exps. 4-b through 8-b disregard the presence of buildings in their respective domains during the first MCMC stage. The goal is to examine the performance of the multistage approach, and investigate the sensitivity of the inverted solution to the features of the filter step. Since a detailed Bayesian inversion is necessary in the second MCMC stage, the original model is run on a fixed high-resolution 40×40 m grid, resolving all buildings, for each of the two-stage experiments. The experimental setup is outlined in Table 2. A hyperparameter search was conducted for each experiment to determine the likelihood scaling parameter, β. For the simulated scenario, we generated several perturbed realizations of the model output through variations in the observational error. The selected β value is the one producing the largest likelihood value between the reference observation and those realizations. The conventional and AI-based W_2 distances result in β values of 1.5 and 1.8, respectively.

Table 2
Settings of the numerical experiments.

Experiment	MCMC mode	Buildings in filter step	Metric	Filter spatial resolution [m]
1	Single-stage	–	Conventional W_2	–
2	Two-stage	Yes	Conventional W_2	80
3	Single-stage	–	AI-based W_2	–
4-a				80
5-a				120
6-a	Two-stage	Yes	AI-based W_2	160
7-a				240
8-a				320
4-b				80
5-b				120
6-b	Two-stage	No	AI-based W_2	160
7-b				240
8-b				320

6. Results and discussion

For every experiment, ten independent randomly initialized Markov chains were generated, each comprising 100,000 Monte Carlo samples. The chains exhibit frequent oscillations over the prior bounds, indicating good mixing and efficient exploration of the model parameter space. The first 1000 samples were discarded as burn-in, and a kernel density estimation was employed to estimate the posterior distributions of the remaining samples (Peter D, 1985; Botev, 2016). We illustrate the marginal posterior probability distribution for each inverted parameter along the diagonal subplots of the upcoming corner plots. The joint posterior probability for each pair of parameters is also presented on the off-diagonal subplots of these figures. We analyze the proposed inversion framework's performance through three sets of experimental conditions. The first set involves a direct comparison between the

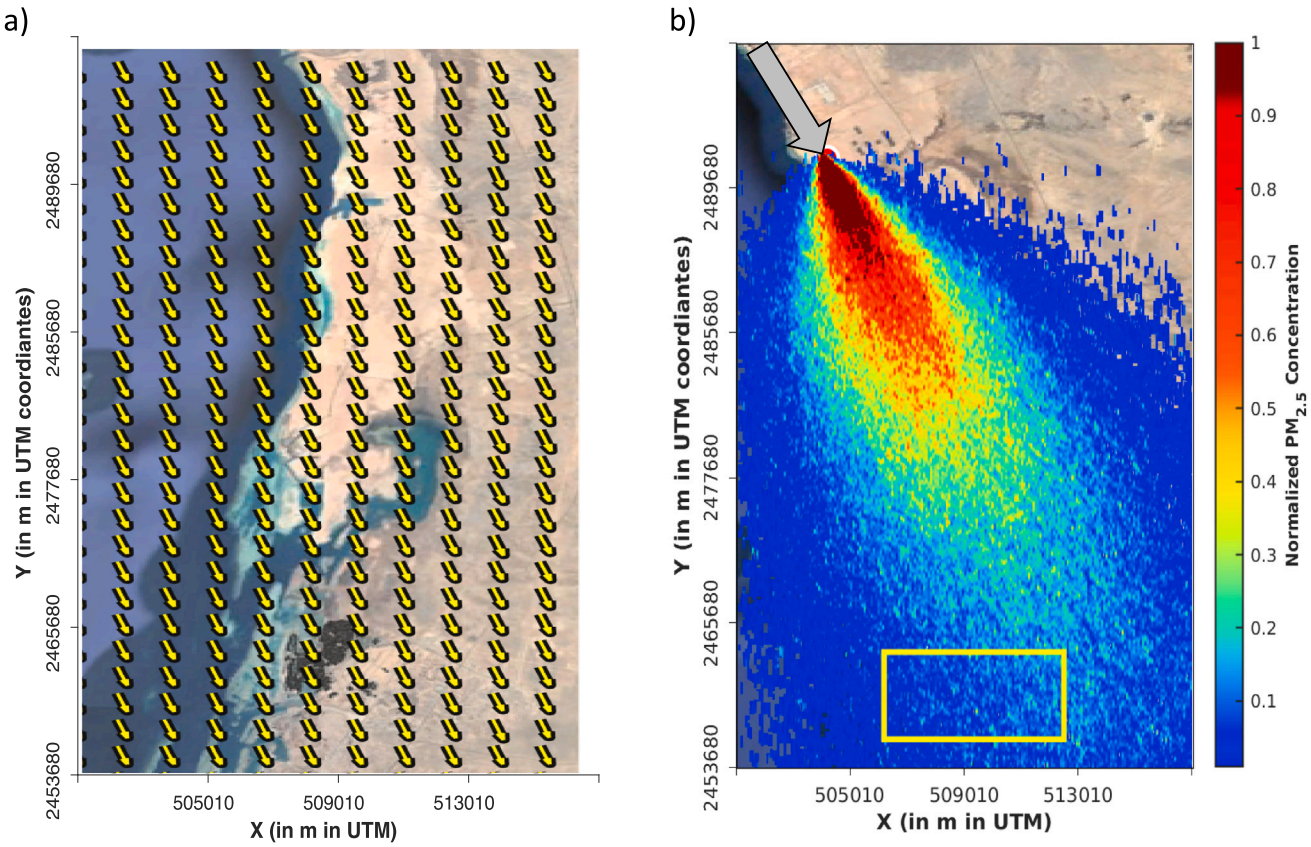

Fig. 6. (a) The microscale wind field of the simulated release case over GRAL domain. (b) Normalized values of the $PM_{2.5}$ concentrations of the base emission scenario. The gray arrow represents the prevailing wind direction.

single- and two-stage approaches, allowing for a thorough evaluation of their effectiveness and computational requirements. The second set uses the standard single-stage MCMC and contrasts the results obtained for the conventional W_2 distance with those obtained using the trained Siamese network. The third set assess the efficiency of the two-stage approach and examine the sensitivity of the inverted solution to the features of the approximative model in the filter step. Some configurations investigate the influence of different spatial resolutions, while the rest evaluate the effect of eliminating buildings from the domain during the MCMC filter stage. The inverted solutions were outlined as the MAP for each experiment and are listed in Table 3.

6.1. Single- vs. two-stage MCMC

We assess the inverse solutions computed by a single- and two-stage MCMC algorithm using the conventional W_2 distance in Exps. 1 and 2, respectively. The Markov chain was obtained with an acceptance rate 28% for Exp. 1. Fig. 7 illustrates a trace plot of the resulting Markov chains, each reaching a stationary distribution for the inverted parameters with sufficient mixing. The chain demonstrates fast convergence for x, with fluctuations stabilizing after around 2000 iterations. This signifies the high dependence of the inverted solution on the proposed x value. Such relation can be easily interpreted by analyzing the geometric orientation of the flow field with respect to the locations of the true source and the observation (Fig. 6).

The two-stage experiment reveals that the Markov chains have an acceptance rate of 20% in the first stage and 13% in the second one (see Table 4). These rates indicate the filter step effectiveness in rejecting unrepresentative samples. By examining the fraction of samples accepted in the second stage with respect to those passed into it, 65% of the samples are accepted, resulting in a more computationally efficient sampling strategy, and reducing the originally required computational time by 15%.

Figs. 8 and 9 illustrate the corner plots of the posterior distributions of the inferred parameters corresponding to Exps. 1 and 2, respectively. The posterior distributions exhibit comparable features for both experiments, at the level of the marginal and joint distributions of x and y. In particular, they illustrate relatively narrow support, with the maxima peaking close to the reference values. This suggests that the two-stage MCMC algorithm maintains accuracy similar to that with a single stage. The MAP estimate of the source location in Exp. 2 reveals that the predicted release location is only 1.38 km away from the actual source, similar to the distance obtained in Exp. 1 of 1.28 km (see Table 3). However, both experiments result in broader distributions for q and d, indicating that these parameters are more uncertain. We note that this higher uncertainty is most likely attributed to the available observation, which covers a horizontal plane at one instance in time, making it harder for the framework to predict time-related parameters (i.e. q and d) with high confidence. Similar conclusions were reached by Al Aawar et al. (2023), indicating that including additional observations in time and along the vertical dimension results in better posterior distributions for q and d. Moreover, no clear correlations between the different parameters were revealed by the joint distributions. This is also attributed to the lack of data conveyed in the single available observation, which may not be enough to establish relations between the parameters.

6.2. W_2 Distance vs. its AI-counterpart

We evaluate the inverse solutions achieved through single-stage MCMC using both the conventional and AI-based W_2 distances in Exps. 1 and 3, respectively. The Markov chains yielded acceptance rates that fall within a desirable range (Sherlock and Roberts, 2009), with

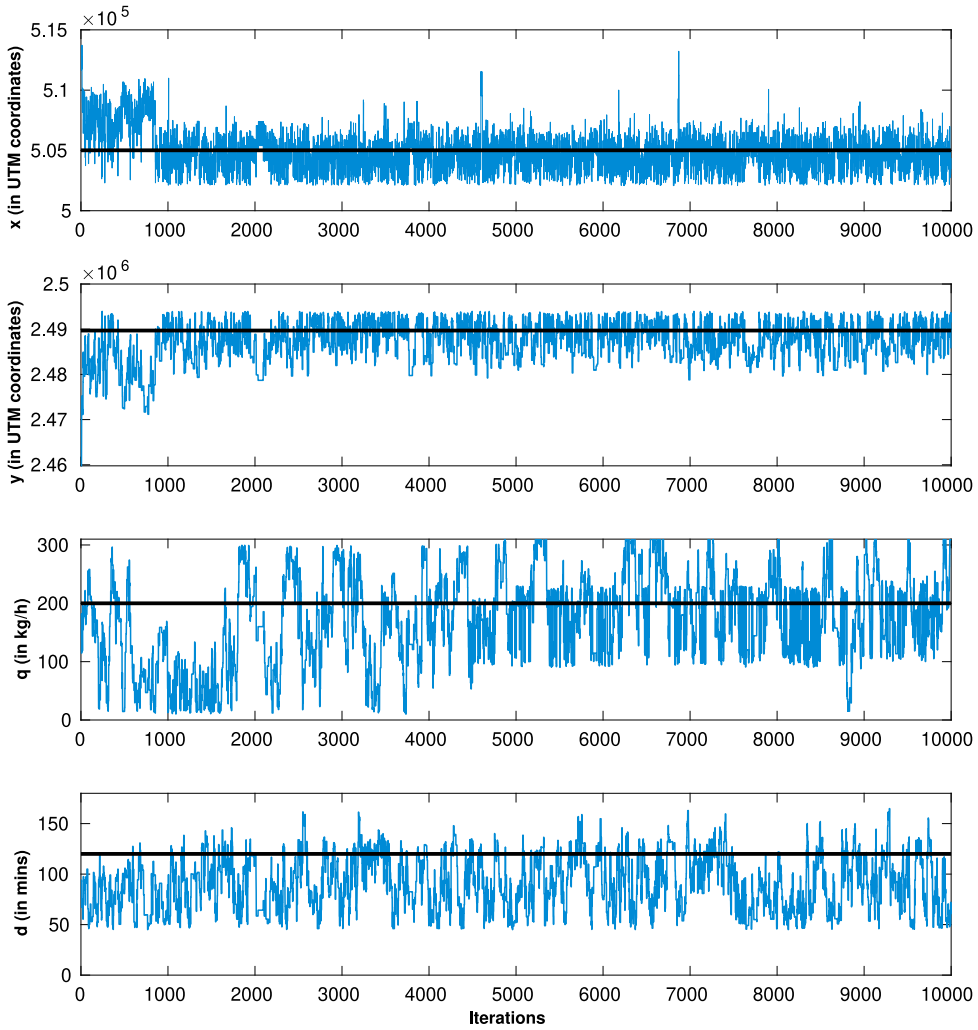

Fig. 7. Trace plot of the Markov chains obtained in Exp. 1. Horizontal black lines represent the nominal values for each inverted parameter.

values of 28% for Exp. 1, and 26% for Exp. 3. In Fig. 10, the corner plots for Exp. 3 are presented, illustrating a posterior distribution similar to that of Exp. 1 in terms of its spread and mode. These plots reveal wider posterior distributions for q and d compared to those for x and y, suggesting a more accurate source location recovery. Moreover, the MAP estimates of the source location resulting from both experiments are only 1.28 and 1.78 km from the true release source location, respectively. In both experiments, the MAP estimates corresponding to q and d lie within the high posterior probability contours, which are relatively away from the reference values. This is linked to the limitations of the available observation, which only captures a single moment in time on a horizontal plane. As also explained in Section 6.1, the joint distributions demonstrate the lack of correlations between the different parameters.

These results suggest that the AI-based W_2 distance approximation helps infer the parameters with an accuracy comparable to the conventional methodology. Furthermore, using the AI-based W_2 distance approximation achieves a 50% reduction in the overall computational time required to solve the Bayesian inverse problem compared to using the original W_2 distance (check Table 4). The reported 50% reduction in total computational time is exclusively attributed to the efficiency gained through the AI-based approximation of the W_2 distance.

6.3. Sensitivity to the features of the approximative model

The remaining numerical experiments investigate the performance of the two-stage MCMC algorithm with various filter settings, while adopting the AI-based approximation of the W_2 distance. We examine several resolutions of the approximative model, and also test the influence of excluding the buildings from the domain in the filter step. We specifically investigate these model features due to their considerable impact on computational costs. The acceptance rates in each of the stages for all two-stage experiments and the overall computational run times are presented in Table 4. The overall acceptance rate r is computed as the proportion of samples accepted in the second stage to those actually tested in this stage. High values for r (roughly over 60%) indicate that the filter step rejects a large portion of unlikely samples reducing the overall computational burdens of Bayesian inversion.

The corner plots for the inverted parameters using filters with spatial resolutions varying from 80 m to 320 m (Table 2) are represented in Figs. B.1 through B.5 in Appendix B. The resulting posterior distributions are similar to those obtained in Section 6.2. In particular, for all filter resolutions, the marginal posterior distributions relative to x and y are confined with narrow supports, as opposed to the stretched posteriors of q and d. This is inherently expressed as compact contours of joint posterior distributions for x and y, and wider contours for q and d. Additionally, we note that by employing the two-stage method, the computational costs are reduced by 65% to 87% for spatial resolutions spanning from 80 m to 320 m, respectively. These results demonstrate that the low-resolution filter efficiently rejects unlikely samples, thereby facilitating more efficient sampling and enabling a faster convergence to a stationary distribution for the inverted parameters. Note that as the observational data is represented by a single snapshot, there is no room to improve the estimates of q and d.

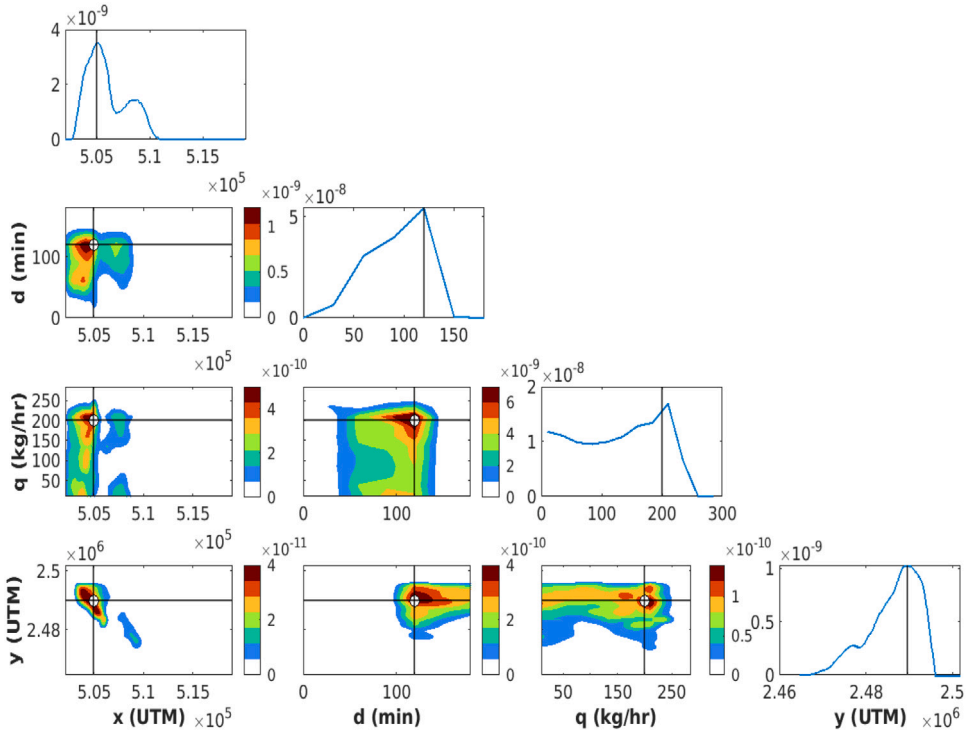

Fig. 8. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 1. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

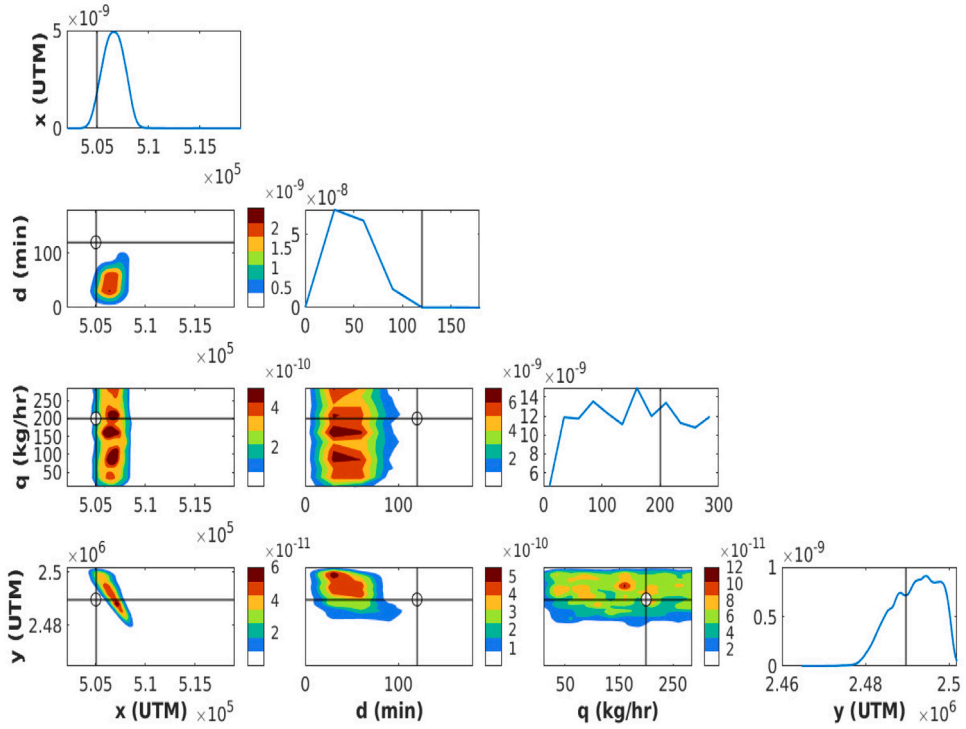

Fig. 9. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 2. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

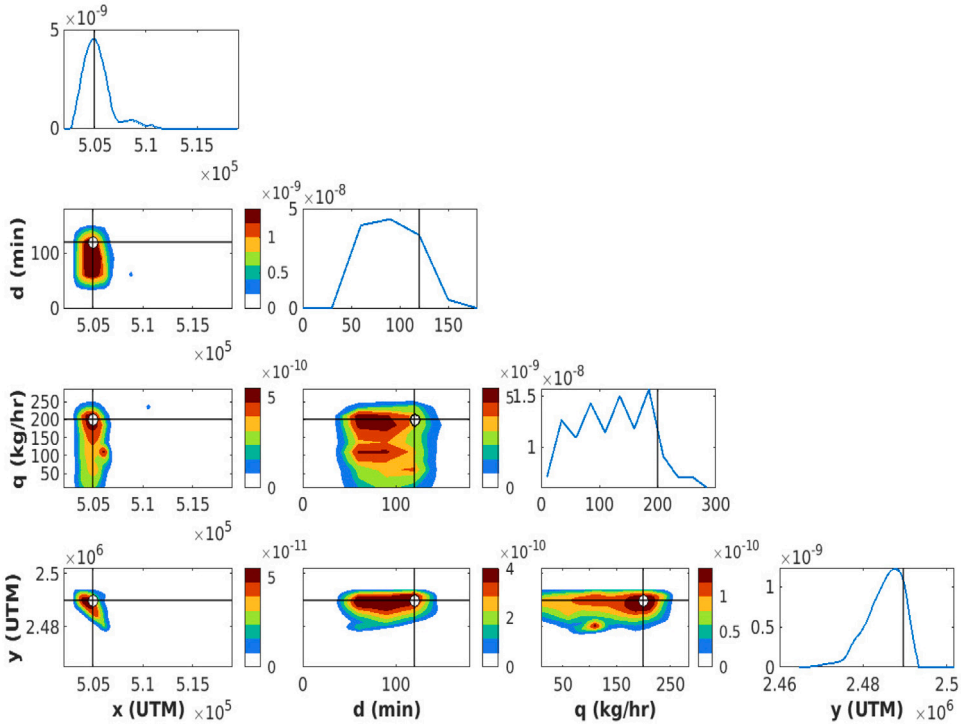

Fig. 10. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 3. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

Examining Exps. 4-a through 8-a in further detail, as the filter model's spatial resolution is coarsened, the high posterior probability zones diverge from their reference values. This is also confirmed by the MAP predictions, such that the coarser resolution filters result in a less accurate MAP estimate. Note that the errors increase as the filter model's resolution coarsens, with a notable inflection point observed at a resolution of 240 m. Beyond this point, the errors sharply escalate. Specifically, with a 320 m resolution filter, the inverted source location deviates by 5.16 km from the reference location, whereas the finer spatial filter yields deviations of at most 2.5 km. Results obtained from Exps. 4-b through 8-b align closely with those of Exps. 4-a through 8-a, respectively, and a concise analysis is provided for completeness. When buildings are excluded from the filter step, the overall run time significantly decreases compared to previous experiments, ranging from a reduction of 71% to 89.5% for the considered filter resolutions. However, this efficiency gain comes at the expense of compromised solution accuracy. The resulting inverted distribution exhibits a greater uncertainty and broader spread, with the MAP estimates deviating further from the reference solution as compared to previous experiments (Table 3). These findings highlight the need to find a trade-off between the desired accuracy and the available computational resources, where more substantial savings can be achieved with coarser filters, but at the cost of a less accurate pollution source identification.

Finally, to further illustrate the results in terms of the relevant atmospheric environment, Fig. 11 shows the spatial distribution of the L_2 error values between the true and predicted concentration fields corresponding to Exp.4-a MAP source estimates. The heatmap reveals that the majority of locations exhibit lower error values, suggesting that the framework performs remarkably for a substantial portion of the study area. However, areas located far from the observation sites exhibit larger L_2 errors, highlighting the importance of data for Bayesian inference. It is not uncommon for areas situated at a considerable distance from the observation locations to demonstrate higher discrepancies. Hence, it is natural for errors to grow with the distance from the observation locations.

Table 3
Inverted solution for each experiment as represented by the MAP. The distance between the inverted source and true location is denoted by δ. Reference values are bold.

Experiment	x (UTM)	y (UTM)	q [kg/h]	d [min]	δ [km]
1	504,990	2,488,430	210	120	1.28
2	506,390	2,489,730	160	60	1.38
3	505,090	2,487,930	185	90	1.78
4-a	506,390	2,490,630	135	150	1.66
5-a	506,790	2,489,630	85	30	1.78
6-a	506,290	2,491,430	160	60	2.15
7-a	506,990	2,488,130	210	30	2.53
8-a	507,190	2,485,030	235	30	5.16
4-b	505,390	2,487,230	160	30	2.51
5-b	507,190	2,487,930	210	60	2.81
6-b	506,890	2,487,430	210	90	2.95
7-b	507,490	2,486,830	160	60	3.8
8-b	507,590	2,483,930	235	60	6.33
Reference values	**505,010**	**2,489,707**	**200**	**120**	–

7. Summary and conclusions

This study addresses the Bayesian inverse problem of characterizing air pollution emission sources in urban environments. We explore numerical techniques to alleviate the computational challenges of Bayesian inversion, utilizing a two-stage MCMC approach alongside a DNN for approximating the dissimilarity between observed and simulated concentration distributions, quantified by the W_2 distance. Generally, a Bayesian inverse problem is solved using an MCMC algorithm, requiring full-forward runs for each posterior sample, making it inefficient for large-scale models. A two-stage approach would improve

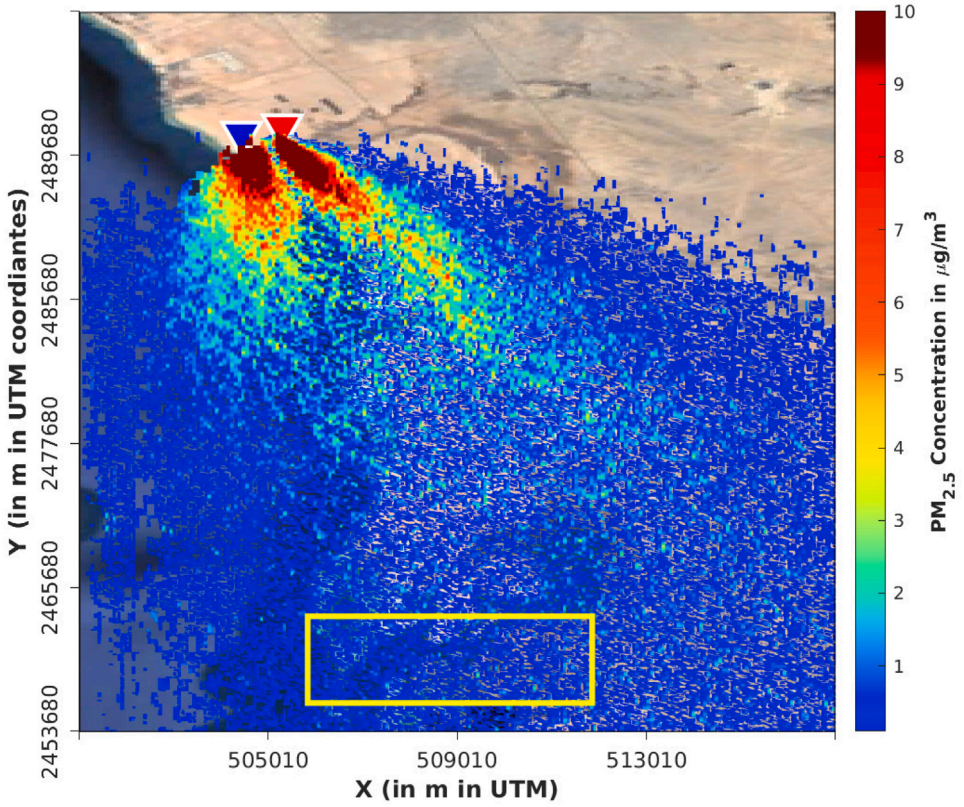

Fig. 11. Spatial distribution of the L_2 error between true and predicted concentrations in Exp. 4-a. The blue triangle marks the true location of the emission source, whereas the red one shows the predicted location based upon the MAP source estimates. The yellow rectangle delimits the reference observation, covering the urban area of KAUST campus.

Table 4
Acceptance rates and total run times in each experiment. The acceptance rate in the first stage is denoted by r_1 and in the second stage by r_2. These ratios are with respect to the total number of samples in a chain. The overall acceptance rate, denoted by r, is represented by the fraction of samples accepted in the second stage with respect to those entering this stage.

Experiment	r_1 [%]	r_2 [%]	$r = \frac{r_2}{r_1}$ [%]	Total run time [h]
1	28	–	–	143
2	20	13	65	121
3	26	–	–	72
4-a	24	15	62.5	50
5-a	19.7	13	66	40
6-a	13	9.2	70	31
7-a	21	13.7	65	24
8-a	23	14	61	19
4-b	24	14	58.3	42
5-b	20.6	14	68	33
6-b	19	12.6	66	25
7-b	19.7	12.8	65	19
8-b	23	15	65	15

MCMC sampling by rejecting unrepresentative samples simulated on a coarser-resolution mesh, and avoiding the need for high-resolution model runs. Furthermore, a Siamese neural network was trained to approximate the W_2 distance between two concentration distributions, providing a swift estimation for a computationally intensive optimization problem. This study also investigates the sensitivity of the inverse solution to the characteristics of the approximative model of the filter step, and offers insights into the impact of different model configurations.

The numerical experiments indicate that employing an AI-based approximation of the W_2 distance reduces the computational time required for the complete Bayesian inversion problem by half. Furthermore, the study examines solving the Bayesian inverse problem using the two-stage MCMC algorithm and the AI-based W_2 approximation. The results indicate that further savings could be achieved by combining the two methods, where the overall computational time is at least 65% less than that of the single-stage MCMC, resolving the W_2 distance through optimization. Computational savings also depend on the approximative model used in the filter step. In particular, these savings depends on whether the approximative model resolves buildings or not, and on its spatial resolution. Note however that the inversion accuracy may be compromised due to underlying assumptions and excessively low-resolution filter grids. Hence, a trade-off between accuracy and computational costs is deemed necessary.

Knowing that most computational expenses are now related to the physical urban flow model, one may resort to machine learning emulations and polynomial chaos expansions to further reduce the overall computational costs. We aim to investigate such ideas in future work.

Funding

This work was supported by the Office of the Vice President of Research at the King Abdullah University of Science and Technology (KAUST) under the Virtual Red Sea Initiative (Grant #REP/1/3268-01-01).

CRediT authorship contribution statement

Elissar Al Aawar: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mohamad Abed El Rahman Hammoud:** Conceptualization, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **Ibrahim Hoteit:** Conceptualization, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Sensitivity to the discrepancy measures

We perform experiments to assess different metrics for modeling the exponential likelihood, as detailed in Section 3.1. In this appendix, we begin by presenting the mathematical definitions of these metrics, along with their corresponding experimental configurations. Then, we analyze the sensitivity of the inverted solution to the selected discrepancy measures through the interpretation of the resulting posterior distributions.

A.1. Model-data misfit measures

Our comprehensive sensitivity analysis covered several measures, including the Wasserstein (W_2), Hellinger (H), and total variation (TV) distances, as well as the Kullback–Leibler (KL) and Jensen–Shannon (JS) divergences. Since the W_2 distance is discussed in Section 3.1, we now proceed to define the remaining metrics. The H distance is a special case of the f-divergence measuring the difference between two probability distributions. It is generally called the bounded analog of the Euclidean distance in the probability space (Bhattacharyya, 1943; Aherne et al., 1998). In particular, the H distance reduces the f-divergence to the ℓ_2 distance between the two distributions $\sqrt{p}$ and $\sqrt{q}$, which can be expressed as follows:

$$H_{(p,q)} = \left(1 - \sum_{x\epsilon X} \sqrt{p(x)q(x)}\right)^{\frac{1}{2}}. \tag{A.1}$$

Furthermore, we assess the total variation (TV) distance, which is the greatest possible difference between the probabilities of two distributions (Devroye et al., 2018; Bhattacharyya et al., 2022). The TV distance could be further reduced to the L_1 norm of the difference between the two distributions (Levin and Peres, 2017):

$$TV_{(p,q)} = \sup_{A\epsilon X} |p(A) - q(A)| = \frac{1}{2}\sum_{x\epsilon X} |p(x) - q(x)|, \tag{A.2}$$

where A is a subset of the probability space X.

In addition, we use the Kullback–Leibler (KL) divergence, a non-symmetric measure of the difference between two probability distributions (Joyce, 2011). The KL divergence is the expected log of the ratio of two distributions if the expected value is drawn from the probability distribution p, given as

$$KL(p,q) = \sum_{x\epsilon X} p(x)\log\left(\frac{p(x)}{q(x)}\right). \tag{A.3}$$

We also test the Jensen–Shannon (JS) divergence, which is the symmetric and smooth version of the KL divergence (Nielsen, 2019), expressed as

$$JS(p,q) = \frac{1}{2}KL\left(p,\ \frac{p+q}{2}\right) + \frac{1}{2}KL\left(q,\ \frac{p+q}{2}\right). \tag{A.4}$$

The two-stage MCMC approach was implemented in a set of numerical experiments, each utilizing a distinct dissimilarity measure. This approach aims to evaluate the performance of the multistage method, and explore how the inverted solution's sensitivity varies with different discrepancy measures. Similarly as explained in Section 5, a hyperparameter search was conducted for each experiment to determine the likelihood scaling parameter, β. The relevant experimental settings are listed in Table A.1.

Table A.1
Numerical experimental settings for analyzing the sensitivity to model-data misfit measures.

Experiment	Metric	β
9	H	0.2
10	TV	0.5
11	KL	0.8
12	JS	0.4

A.2. Results and discussion

Four additional numerical experiments were conducted to investigate the performance of the two-stage MCMC algorithm with various distance measures, as listed in Table A.1. The inverted solutions were outlined as the MAP for each experiment and are shown in Table A.2. The acceptance rates in each of the stages for all two-stage experiments are represented in Table A.3. Note that each of these experiments requires a computational time of 50 h.

Figs. A.1 through A.4 illustrate the corner plots corresponding to each of the conducted experiments. The corner plots corresponding to Exp. 9, which considers the H distance as a discrepancy measure, show posterior distributions with ample supports over the prior bounds. Moreover, the plot indicates multiple high posterior probability regions; thus, the relevant combinations of source parameters have equal posterior probabilities. Similar results and interpretation were obtained with Exp. 10, demonstrating the results for the TV distance, as presented in Fig. A.2. Flat marginal distributions and wide contours of joint distributions characterize the resulting posterior. Multiple modes are also present, especially for q and d. The solution represented by the MAP designates an emission source 11 km away from the reference in both experiments.

Exp. 11 uses the KL divergence in its likelihood term. The distributions point to an emission source off from the reference location by 4 km. Moreover, the contours have narrower supports (Fig. A.3), translating to higher confidence in the inverted solution. This setup produces a much better solution than the previous two experiments; however, as expected, the predicted q and d are still far from the reference compared to using the W_2 distance.

Finally, Exp. 12 uses the JS divergence and the corresponding corner plot is shown in Fig. A.4. The posterior distributions of x and y are relatively compact with one maximum, making the results easier to interpret. However, the highest posterior areas are far from the reference values. The results suggest that the solution has less uncertainty than Exps. 9 and 10; however, the MAP indicates that the predicted solution is far from the reference. The posterior distributions for q and d are broad with high uncertainty.

We further analyze the results of Exps. 9 and 10 by linking in the behavior of the H and TV distances, which are bounded measures over the posterior probability space (Oosterhoff and van Zwet, 2012). The bounded property leads to similar likelihood values for concentration distributions sufficiently far apart. The MCMC algorithm accepts or rejects different samples with equal probabilities in this case. In Daskalakis and Pan (2017), the authors reveal that two measures between distributions p and q are related via the inequality $H^2(p,q) \leq TV(p,q) \leq \sqrt{2}H(p,q)$. Thus, the TV and H distances penalize the dissimilarity between the model outputs and the reference observations analogously. Because of this penalty, the framework predicts sources with inaccurate emission intensities around 11 km from their actual location (see Table A.2). The distributions in Exp. 10 seem sharper than those in Exp. 9, especially those corresponding to x and y. This finding could be because the TV distance is less sensitive to outliers in the distribution (Markatou and Chen, 2018).

In contrast, we interpret the better performance of the KL divergence in Exp. 11 by two of its main features: it is not bounded, and it

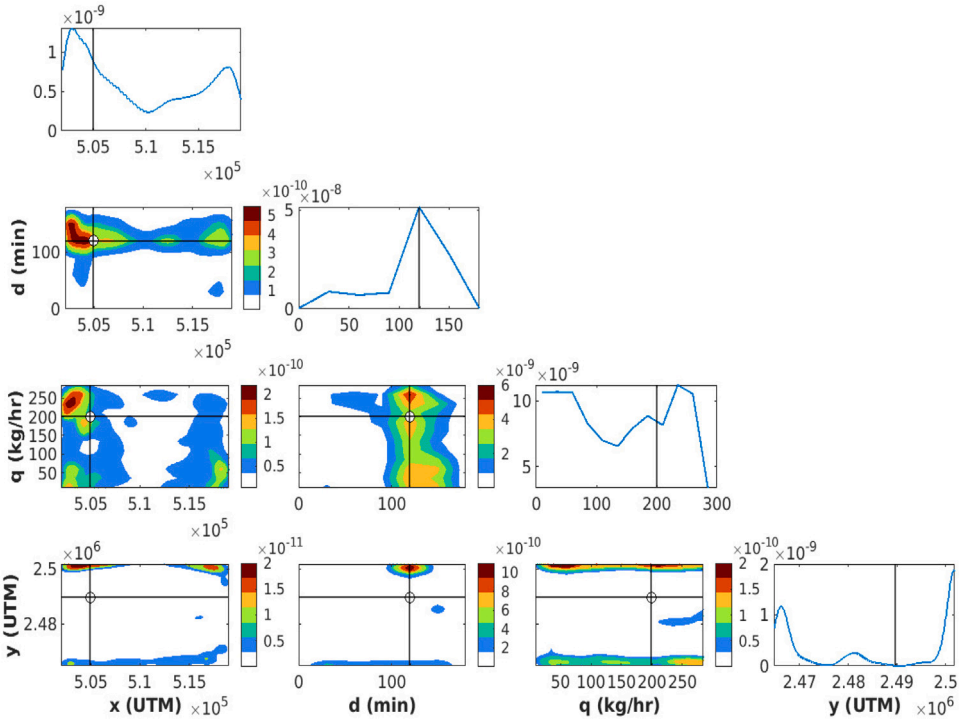

Fig. A.1. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 9. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

relates to the two previous distances through $TV(p,q) \leq \sqrt{2}H(p,q) \leq \sqrt{KL(p,q)}$ (Daskalakis and Pan, 2017). This inequality suggests that the KL divergence quadratically penalizes the mismatch between two distributions p and q, amplifying the value of the mismatch in comparison to the H and TV distances. By pairing such a desired property with its unboundedness, we deduce that plugging the KL divergence into the likelihood expression is highly beneficial. This approach facilitates distinguishing whether the model outputs are becoming closer to the reference observation, which does not apply to the bounded metrics. Eventually, these properties result in a more certain posterior distribution and better solution. The solution of Exp. 11 using the KL divergence is the second-most accurate, after those using the W_2 distance. The low acceptance rate in the second stage of this experiment (Table A.3) reveals the strict acceptance criteria in the case of KL divergence. Regarding the JS divergence, which is adapted from the KL divergence, it is significant noting that it is a smoothed version. Hence, it penalizes the mismatch less aggressively than the KL divergence, leading to less accurate inference solutions.

Adding up to the discussion in Section 6, these results suggest the W_2 distance results in the most representative posterior distribution for the emission parameters compared to other measures. Specifically, the MAP approximations corresponding to the case where the W_2 distance was employed are closer to the reference solution than those obtained using other measures.

Table A.2
Inverted solution for experiments 9 through 12. The are consistent with those presented in Table 3. Reference values are bold.

Experiment	x (UTM)	y (UTM)	q [kg/h]	d [min]	δ [km]
9	502,790	2,501,030	235	120	11.54
10	509,190	2,479,230	35	120	11.28
11	505,090	2,493,830	60	60	4.12
12	508,390	2,480,530	235	90	9.78
Reference values	**505,010**	**2,489,707**	**200**	**120**	–

Table A.3
Acceptance rates and total run times in experiment 9 through 12. The notations are consistent with those presented in Table 4.

Experiment	r_1 [%]	r_2 [%]	$r = \frac{r_2}{r_1}$ [%]
9	24	14	58
10	29	18	62
11	24	8	34
12	23	19	83

Appendix B. Corner plots for the sensitivity analysis of the approximative model features

We provide the results of the sensitivity analysis against the features of the approximative model through a collection of corner plots. These figures present illustrations of the posterior distributions obtained for the numerical experiments defined in Table 2.

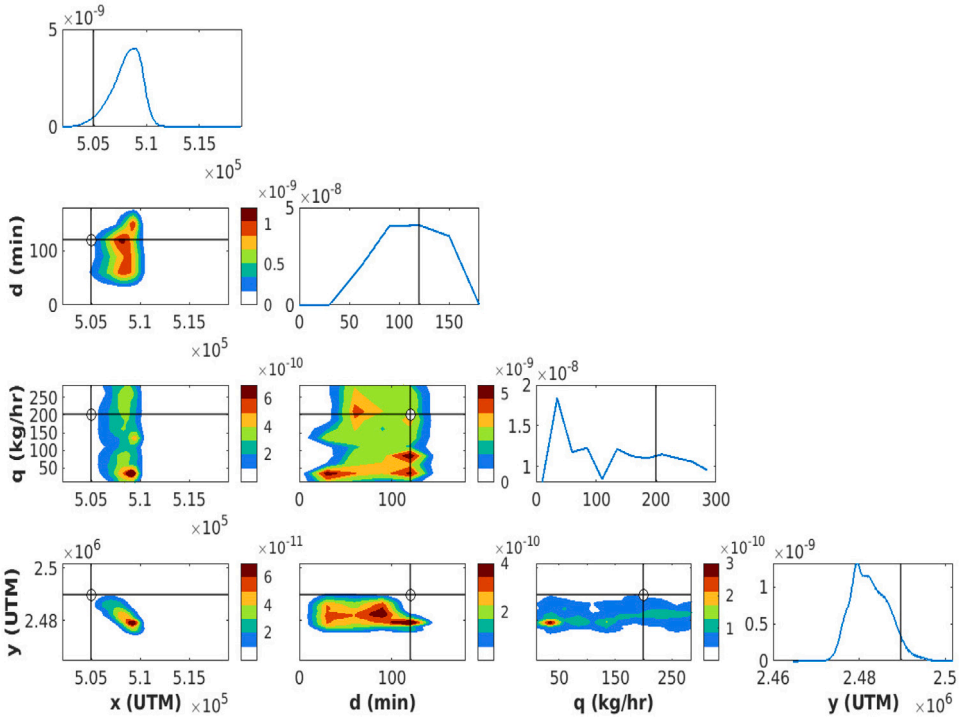

Fig. A.2. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 10. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

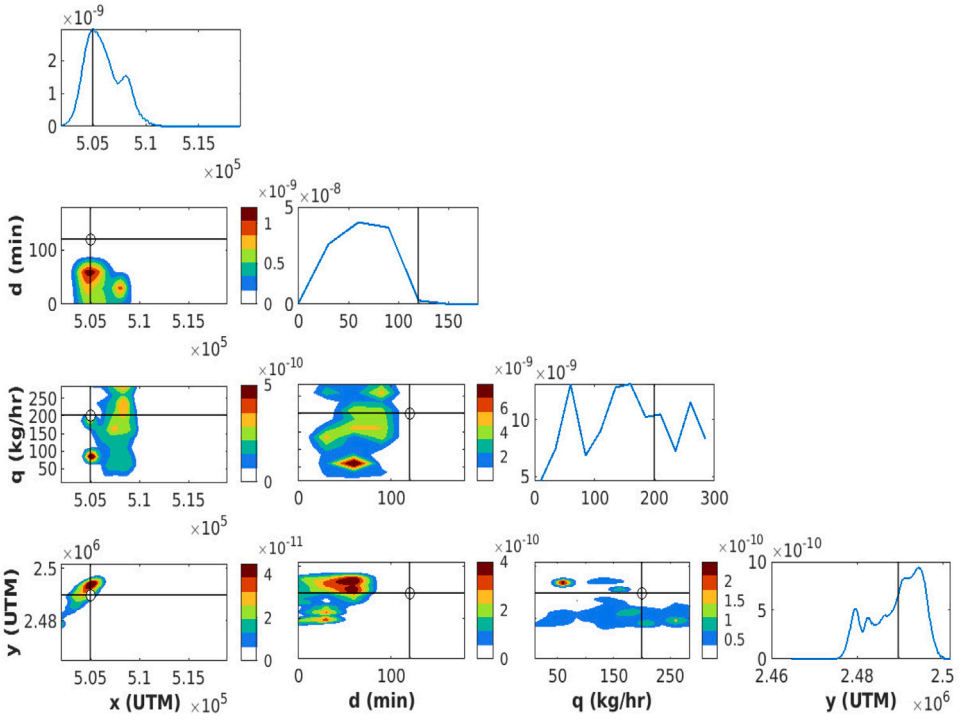

Fig. A.3. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 11. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

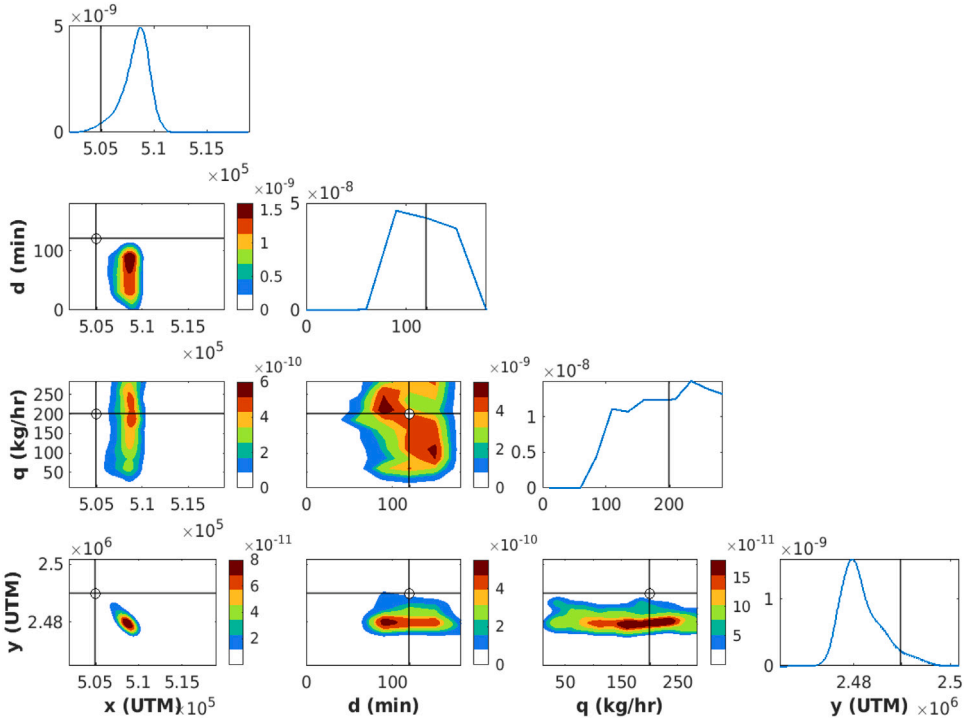

Fig. A.4. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 12. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

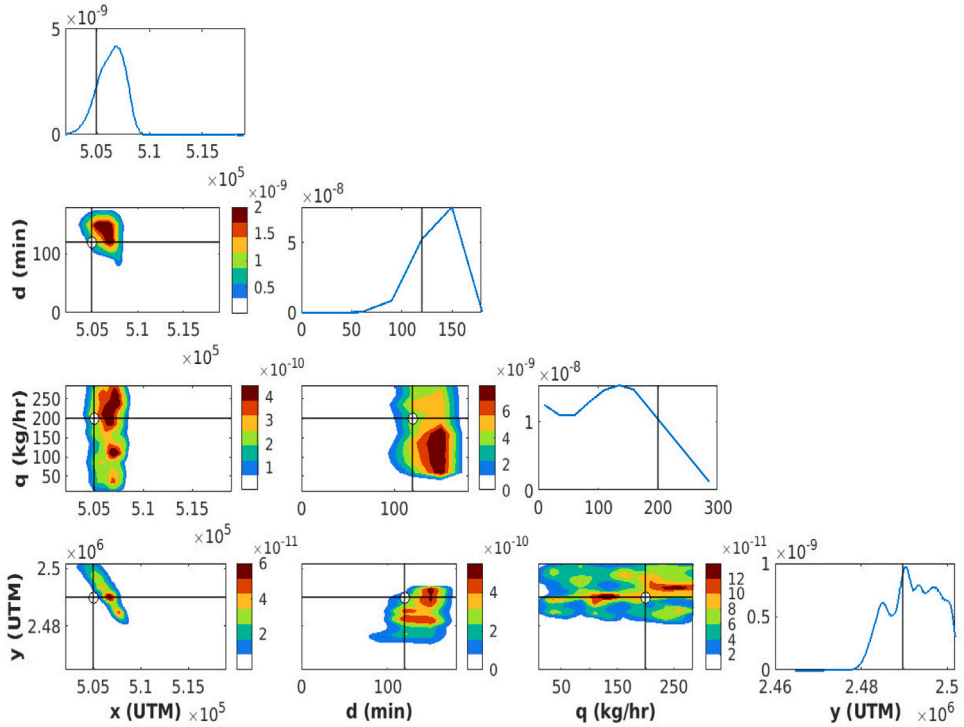

Fig. B.1. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 4-a. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

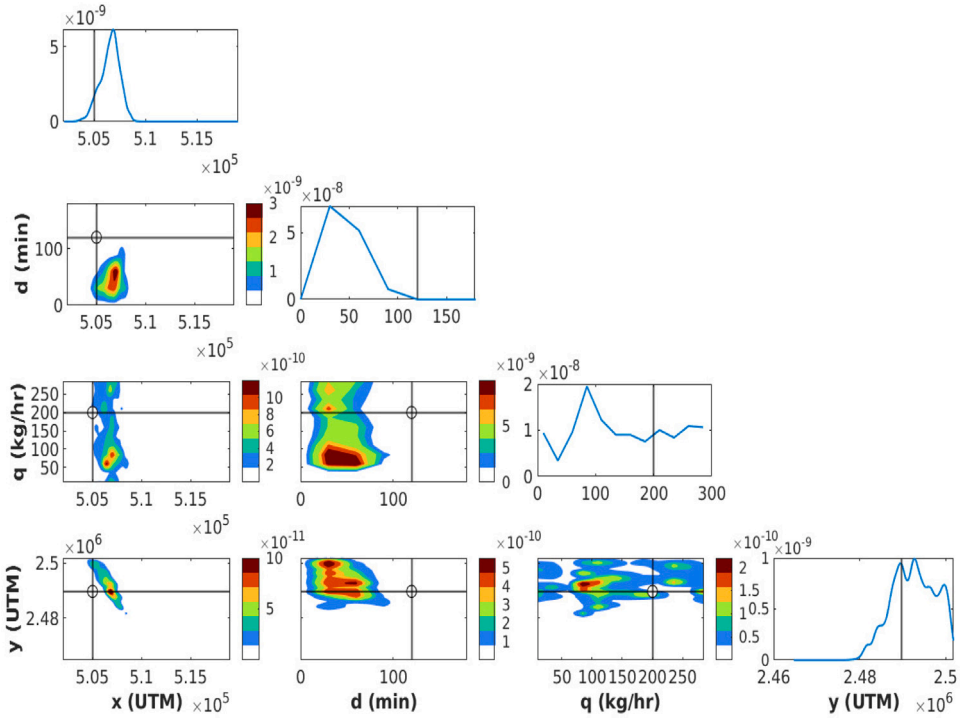

Fig. B.2. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 5-a. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

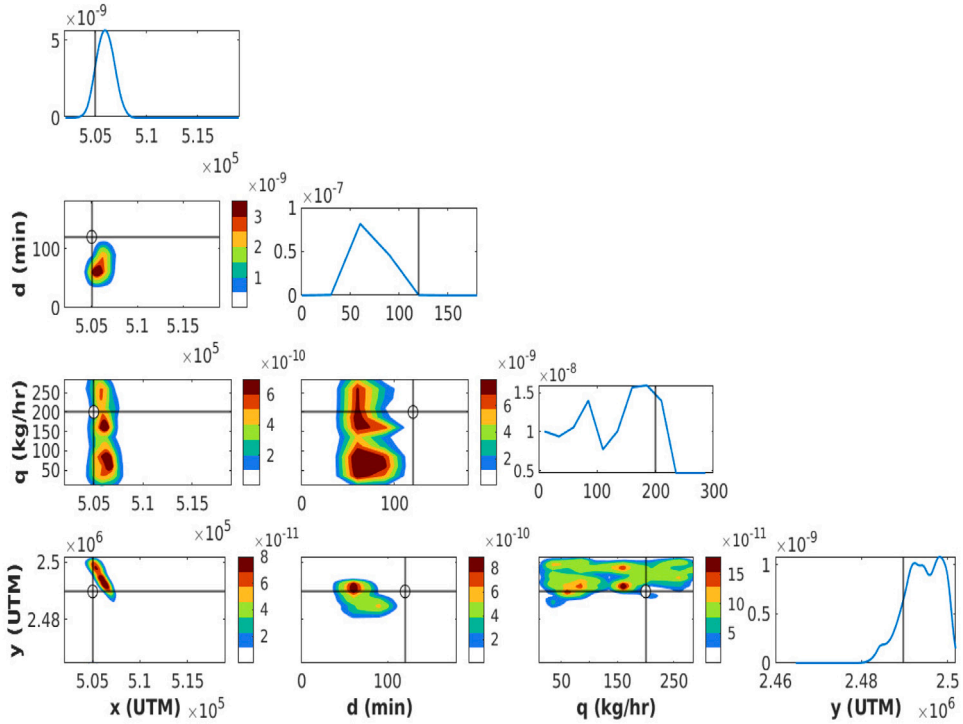

Fig. B.3. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 6-a. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

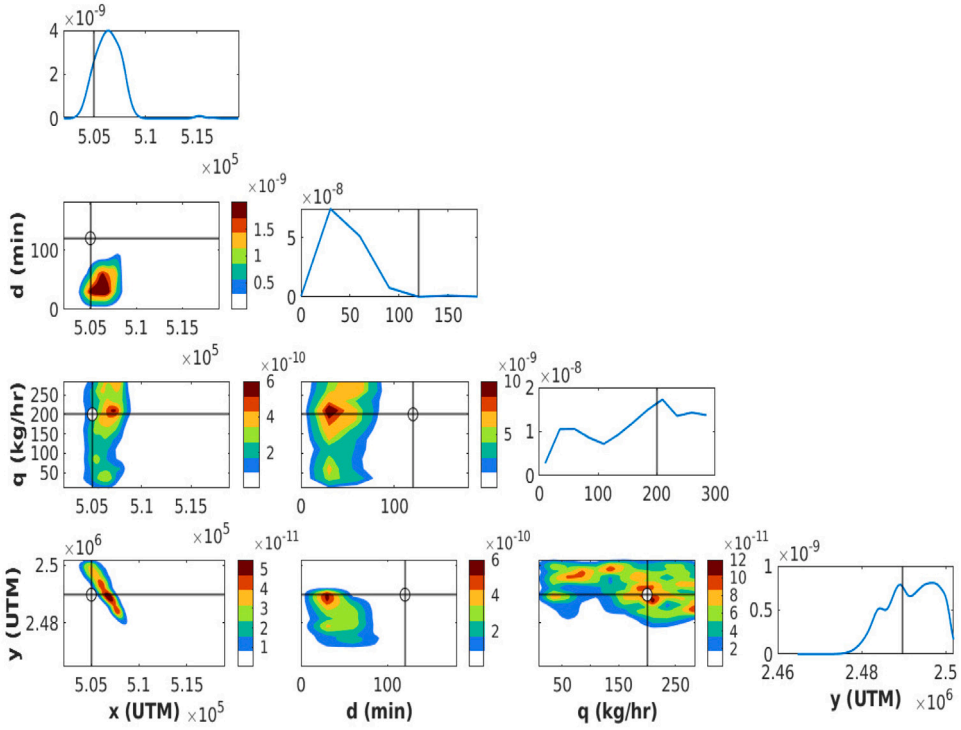

Fig. B.4. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 7-a. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

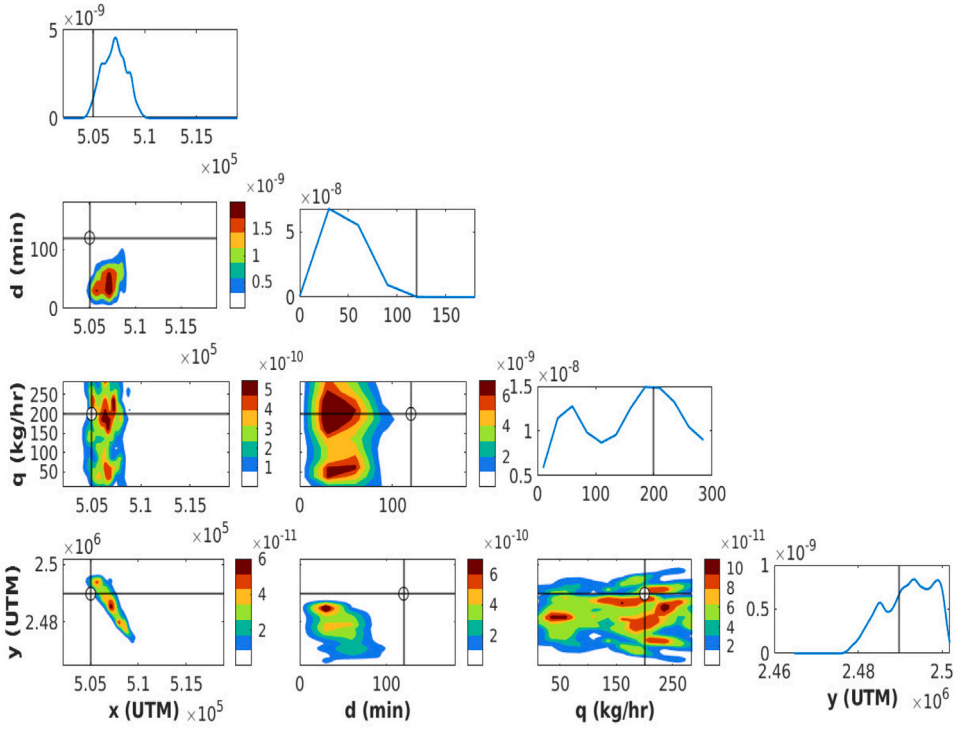

Fig. B.5. Corner plot of the marginal and joint probabilities of the inverted parameters extracted from the posterior distribution in Exp. 8-a. Black lines represent the nominal values for each inverted parameter. The boundaries represented in each subplot fall within the prior limits.

References

Agrawal, A., Mittal, N., 2020. Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. Vis. Comput. 36 (2), 405–412.

Aherne, F.J., Thacker, N.A., Rockett, P.I., 1998. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. Kybernetika 34 (4), 363–368.

Al Aawar, E., El Mohtar, S., Lakkis, I., Alduwais, A.K., Hoteit, I., 2023. Bayesian source identification of urban-scale air pollution from point and field concentration measurements. Comput. Geosci. 1–22.

Balogun, A.-L., Tella, A., Baloo, L., Adebisi, N., 2021. A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. Urban Clim. 40, 100989.

Bassett, R., Deride, J., 2019. Maximum a posteriori estimators as a limit of Bayes estimators. Math. Program. 174, 129–144.

Berrocal, V.J., Guan, Y., Muyskens, A., Wang, H., Reich, B.J., Mulholland, J.A., Chang, H.H., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM2. 5 concentration. Atmos. Environ. 222, 117130.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distribution. Bull. Calcutta Math. Soc. 35, 99–110.

Bhattacharyya, A., Gayen, S., Meel, K.S., Myrisiotis, D., Pavan, A., Vinodchandran, N., 2022. On approximating total variation distance. arXiv preprint arXiv:2206.07209.

Bocquet, M., 2007. High-resolution reconstruction of a tracer dispersion event: application to ETEX. Q. J. R. Meteorol. Soc. 133 (625), 1013–1026.

Botev, Z., 2016. Fast multivariate kernel density estimation for high dimensions.

Box, G.E., Tiao, G.C., 2011. Bayesian Inference in Statistical Analysis, vol. 40, John Wiley & Sons.

Brooks, S., Gelman, A., Jones, G., Meng, X.-L., 2011. Handbook of Markov Chain Monte Carlo. CRC Press.

Buckland, K.N., Young, S.J., Keim, E.R., Johnson, B.R., Johnson, P.D., Tratt, D.M., 2017. Tracking and quantification of gaseous chemical plumes from anthropogenic emission sources within the Los Angeles Basin. Remote Sens. Environ. 201, 275–296.

Chen, M.-H., Shao, Q.-M., Ibrahim, J.G., 2012. Monte Carlo Methods in Bayesian Computation. Springer Science & Business Media.

Chen, Y., Wang, S., Han, W., Xiong, Y., Wang, W., Tong, L., 2017. A new air pollution source identification method based on remotely sensed aerosol and improved glowworm swarm optimization. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (8), 3454–3464.

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Amer. Statist. 49 (4), 327–335.

Chicco, D., 2021. Siamese neural networks: An overview. Artif. Neural Netw. 73–94.

Chodorek, A., Chodorek, R.R., Sitek, P., 2021. UAV-based and WebRTC-based open universal framework to monitor urban and industrial areas. Sensors 21 (12), 4061.

Chow, F.K., Kosović, B., Chan, S., 2008. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. J. Appl. Meteorol. Climatol. 47 (6), 1553–1572.

Christen, J.A., Fox, C., 2005. Markov chain Monte Carlo using an approximation. J. Comput. Graph. Statist. 14 (4), 795–810.

Ciaccia, P., Patella, M., Zezula, P., 1997. M-tree: An efficient access method for similarity search in metric spaces. In: Vldb. Vol. 97, pp. 426–435.

Cooper, M.J., Martin, R.V., van Donkelaar, A., Lamsal, L., Brauer, M., Brook, J.R., 2012. A satellite-based multi-pollutant index of global air quality.

2018. CORINE land cover. URL https://land.copernicus.eu/.

Courty, N., Flamary, R., Ducoffe, M., 2017. Learning wasserstein embeddings. arXiv preprint arXiv:1710.07457.

Dasari, H.P., Desamsetti, S., Langodan, S., Karumuri, R.K., Singh, S., Hoteit, I., 2020. Atmospheric conditions and air quality assessment over NEOM, kingdom of Saudi Arabia. Atmos. Environ. 230, 117489.

Daskalakis, C., Pan, Q., 2017. Square Hellinger subadditivity for Bayesian networks and its applications to identity testing. In: Conference on Learning Theory. PMLR, pp. 697–703.

Delle Monache, L., Lundquist, J.K., Kosović, B., Johannesson, G., Dyer, K.M., Aines, R.D., Chow, F.K., Belles, R.D., Hanley, W.G., Larsen, S.C., et al., 2008. Bayesian inference and Markov chain Monte Carlo sampling to reconstruct a contaminant source on a continental scale. J. Appl. Meteorol. Climatol. 47 (10), 2600–2613.

Devroye, L., Mehrabian, A., Reddad, T., 2018. The total variation distance between high-dimensional Gaussians. arXiv preprint arXiv:1810.08693.

DiazDelaO, F., Garbuno-Inigo, A., Au, S., Yoshida, I., 2017. Bayesian updating and model class selection with subset simulation. Comput. Methods Appl. Mech. Engrg. 317, 1102–1121.

Efendiev, Y., Hou, T., Luo, W., 2006. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. SIAM J. Sci. Comput. 28 (2), 776–803.

El Mohtar, S., Ait-El-Fquih, B., Knio, O., Lakkis, I., Hoteit, I., 2021. Bayesian identification of oil spill source parameters from image contours. Mar. Pollut. Bull. 169, 112514.

Emeis, S., Schäfer, K., 2006. Remote sensing methods to investigate boundary-layer structures relevant to air pollution in cities. Bound.-Lay. Meteorol. 121, 377–385.

Farchi, A., Bocquet, M., Roustan, Y., Mathieu, A., Quérel, A., 2016. Using the Wasserstein distance to compare fields of pollutants: application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident. Tellus B 68 (1), 31682.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. CRC Press.

Gaudio, P., Gelfusa, M., Malizia, A., Parracino, S., Richetta, M., De Leo, L., Perrimezzi, C., Bellecci, C., 2015. Detection and monitoring of pollutant sources with Lidar/Dial techniques. In: Journal of Physics: Conference Series, vol. 658, IOP Publishing, 012004.

Giles, M.B., 2008. Multilevel monte carlo path simulation. Oper. Res. 56 (3), 607–617.

Guo, S., Yang, R., Zhang, H., Weng, W., Fan, W., 2009. Source identification for unsteady atmospheric dispersion of hazardous materials using Markov Chain Monte Carlo method. Int. J. Heat Mass Transfer 52 (17–18), 3955–3962.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 223–242.

Heinrich, S., 2001. Multilevel monte carlo methods. In: Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3. Springer, pp. 58–67.

Higdon, D., Lee, H., Bi, Z., 2002. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. IEEE Trans. Signal Process. 50 (2), 389–399.

Huang, Z., Wang, Y., Yu, Q., Ma, W., Zhang, Y., Chen, L., 2015. Source area identification with observation from limited monitor sites for air pollution episodes in industrial parks. Atmos. Environ. 122, 1–9.

Iglesias, M., Stuart, A.M., 2014. Inverse problems and uncertainty quantification. SIAM News 20, 2–3.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. pmlr, pp. 448–456.

Joyce, J.M., 2011. Kullback-leibler divergence. In: International Encyclopedia of Statistical Science. Springer, pp. 720–722.

Kaginalkar, A., Kumar, S., Gargava, P., Niyogi, D., 2023. Stakeholder analysis for designing an urban air quality data governance ecosystem in smart cities. Urban Clim. 48, 101403.

Keats, A., Yee, E., Lien, F.-S., 2007. Bayesian inference for source determination with applications to a complex urban environment. Atmos. Environ. 41 (3), 465–479.

Kelly, F.J., Fussell, J.C., 2015. Air pollution and public health: emerging hazards and improved understanding of risk. Environ. Geochem. Health 37, 631–649.

Khlaifi, A., Ionescu, A., Candau, Y., 2009. Pollution source identification using a coupled diffusion model with a genetic algorithm. Math. Comput. Simulation 79 (12), 3500–3510.

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. Vol. 14, Montreal, Canada, pp. 1137–1145.

Koracin, D., Vellore, R., Lowenthal, D.H., Watson, J.G., Koracin, J., McCord, T., DuBois, D.W., Chen, L.-W.A., Kumar, N., Knipping, E.M., et al., 2011. Regional source identification using Lagrangian stochastic particle dispersion and HYSPLIT backward-trajectory models. J. Air Waste Manage. Assoc. 61 (6), 660–672.

Kukkonen, J., Olsson, T., Schultz, D.M., Baklanov, A., Klein, T., Miranda, A., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., et al., 2012. A review of operational, regional-scale, chemical weather forecasting models in Europe. Atmos. Chem. Phys. 12 (1), 1–87.

Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. Water Resour. Res. 49 (5), 2664–2682.

Landesregierung, A.D.S., 2018. Documentation of the Lagrangian particle model GRAL (Graz Lagrangian model) vs. 19.1.

Landesregierung, A.D.S., 2019. Documentation of the prognostic mesoscale model GRAMM (Graz Mesoscale Model) Version 19.1.

Lei, M., Cheng, T., Li, X., Shi, S., Zuo, X., Guo, H., Wu, Y., 2021. Atmospheric ammonia point source detection technique at regional scale using high resolution satellite imagery and deep learning. Atmos. Res. 257, 105587.

Levin, D.A., Peres, Y., 2017. Markov Chains and Mixing Times, vol. 107, American Mathematical Soc..

Lévy, B., Schwindt, E.L., 2018. Notions of optimal transport theory and how to implement them on a computer. Comput. Graph. 72, 135–148.

Li, X., Zhang, X., 2019. Predicting ground-level PM2. 5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning approach. Environ. Pollut. 249, 735–749.

Ma, X., Al-Harbi, M., Datta-Gupta, A., Efendiev, Y., 2008. An efficient two-stage sampling method for uncertainty quantification in history matching geological models. SPE J. 13 (01), 77–87.

Ma, X.-N., Trudinger, N.S., Wang, X.-J., 2005. Regularity of potential functions of the optimal transportation problem. Arch. Ration. Mech. Anal. 177 (2), 151–183.

Markatou, M., Chen, Y., 2018. Non-quadratic distances in model assessment. Entropy 20 (6), 464.

Marzouk, Y.M., Najm, H.N., 2009. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. J. Comput. Phys. 228 (6), 1862–1902.

Marzouk, Y.M., Najm, H.N., Rahn, L.A., 2007. Stochastic spectral methods for efficient Bayesian solution of inverse problems. J. Comput. Phys. 224 (2), 560–586.
Marzouk, Y., Xiu, D., 2009. A stochastic collocation approach to Bayesian inference in inverse problems.
Monge, G., 1781. Mémoire sur la théorie des déblais et des remblais. Hist. Acad. R. Sci. Paris.
Neal, R.M., et al., 2011. MCMC using Hamiltonian dynamics. In: Handbook of Markov Chain Monte Carlo. Vol. 2, Chapman and Hall/CRC, p. 2, (11).
Nielsen, F., 2019. On the Jensen–Shannon symmetrization of distances relying on abstract means. Entropy 21 (5), 485.
Nychka, D., Saltzman, N., 1998. Design of air-quality monitoring networks. In: Case Studies in Environmental Statistics. Springer, pp. 51–76.
Oosterhoff, J., van Zwet, W.R., 2012. A Note on Contiguity and Hellinger Distance. Springer.
Organization, W.H., et al., 2021. WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary.
Pasquill, F., 1961. The Estimation of the Dispersion of Windborne Material. Meteor. Mag., 90, 33-49., 1976: Atmospheric Dispersion Parameters in Gaussian Plume Modeling, Part 11. Possible Requirements for Change in the Turner Workbook Values. EPA Publication No. Tech. rep., EPA-600/4-76-030b, Environmental Protection Agency, Research Triangle Park ....
Peter D, H., 1985. Kernel estimation of a distribution function. Comm. Statist. Theory Methods 14 (3), 605–620.
2023. Fire incident, Rabigh electricity power station, Saudi Arabia. https://allsourceanalysis.com/wp-content/uploads/2020/09/SR-SA-Fire-Incident-Rabigh-Electricity-Power-Station-Saudi-Arabia.pdf, accessed: 2023-02-19.
Rajaona, H., Septier, F., Armand, P., Delignon, Y., Olry, C., Albergel, A., Moussafir, J., 2015. An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release. Atmos. Environ. 122, 748–762.
Rosenthal, J.S., et al., 2011. Optimal proposal distributions and adaptive MCMC. In: Handbook of Markov Chain Monte Carlo. Vol. 4, Chapman & Hall/CRC, Boca Raton, FL, (10.1201).
Rowley, A., Karakuş, O., 2023. Predicting air quality via multimodal AI and satellite imagery. Remote Sens. Environ. 293, 113609.
Roy, S.K., Harandi, M., Nock, R., Hartley, R., 2019. Siamese networks: The tale of two manifolds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3046–3055.
Seinfeld, J.H., Pandis, S.N., 2016. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. John Wiley & Sons.
Sgarro, A., 1981. Informational divergence and the dissimilarity of probability distributions. Calcolo 18 (3), 293–302.
Sherlock, C., Roberts, G., 2009. Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. Bernoulli 15 (3).
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., et al., 2019. A Description of the Advanced Research WRF Model Version 4. Vol. 145, National Center for Atmospheric Research, Boulder, CO, USA, p. 550, (145).
Spangl, W., Schneider, J., Moosmann, L., Nagl, C., 2007. Representativeness and Classification of Air Quality Monitoring Stations. Umweltbundesamt Dessau-Roßlau, Germany.
Sraj, I., Mandli, K.T., Knio, O.M., Dawson, C.N., Hoteit, I., 2014. Uncertainty quantification and inference of Manning's friction coefficients using DART buoy data during the Tōhoku tsunami. Ocean Model. 83, 82–97.
Sraj, I., Zedler, S.E., Knio, O.M., Jackson, C.S., Hoteit, I., 2016. Polynomial chaos–based bayesian inference of k-profile parameterization in a general circulation model of the tropical pacific. Mon. Weather Rev. 144 (12), 4621–4640.
Stockwell, W.R., Middleton, P., Chang, J.S., Tang, X., 1990. The second generation regional acid deposition model chemical mechanism for regional air quality modeling. J. Geophys. Res.: Atmos. 95 (D10), 16343–16367.
Stuart, G.K., Minkoff, S.E., Pereira, F., 2019. A two-stage Markov chain Monte Carlo method for seismic inversion and uncertainty quantification. Geophysics 84 (6), R1003–R1020.
Tikhonov, A.N., Goncharsky, A., Stepanov, V., Yagola, A.G., 2013. Numerical Methods for the Solution of Ill-Posed Problems, vol. 328, Springer Science & Business Media.
Tratt, D.M., Young, S.J., Hackwell, J.A., Rudy, D.J., Warren, D.W., Vore, A.G., Johnson, P.D., 2017. MAHI: An airborne mid-infrared imaging spectrometer for industrial emissions monitoring. IEEE Trans. Geosci. Remote Sens. 55 (8), 4558–4566.
Turner, M.C., Andersen, Z.J., Baccarelli, A., Diver, W.R., Gapstur, S.M., Pope III, C.A., Prada, D., Samet, J., Thurston, G., Cohen, A., 2020. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. CA: Cancer J. Clin. 70 (6), 460–479.
Uusitalo, L., Lehikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environ. Model. Softw. 63, 24–31.
Villani, C., 2003. Topics in Optimal Transportation. no. 58, American Mathematical Soc..
Villani, C., 2009. Grundlehren Math. Wiss., Springer-Verlag.
Wang, H., Jin, X., 2013. Characterization of groundwater contaminant source using Bayesian method. Stoch. Environ. Res. Risk Assess. 27, 867–876.
Wei, X., Bai, K., Chang, N.-B., Gao, W., 2021. Multi-source hierarchical data fusion for high-resolution AOD mapping in a forest fire event. Int. J. Appl. Earth Obs. Geoinf. 102, 102366.
Xu, C., Wang, J., Hu, M., Wang, W., 2022. A new method for interpolation of missing air quality data at monitor stations. Environ. Int. 169, 107538.
Yassin, M.F., Almutairi, S.K., Al-Hemoud, A., 2018. Dust storms backward Trajectories' and source identification over Kuwait. Atmos. Res. 212, 158–171.
Yu, W., Zeng, G., Luo, P., Zhuang, F., He, Q., Shi, Z., 2013. Embedding with autoencoder regularization. In: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13. Springer, pp. 208–223.
Zannetti, P., 2013. Air Pollution Modeling: Theories, Computational Methods and Available Software. Springer Science & Business Media.
Zhang, J., 2021. Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey. Wiley Interdiscip. Rev. Comput. Stat. 13 (5), e1539.
Zhou, Y., An, Y., Huang, W., Chen, C., You, R., 2022. A combined deep learning and physical modelling method for estimating air pollutants' source location and emission profile in street canyons. Build. Environ. 219, 109246.