Building and Environment 287 (2026) 113863

Contents lists available at ScienceDirect

Building and Environment

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

Surrogate-based ensemble data assimilation for reducing uncertainty in large-eddy simulation of microscale pollutant dispersion

Eliott Lumet [a,b,*], Mélanie C. Rochoux [b], Thomas Jaravel [b], Simon Lacroix [c,1]

[a] Université de Toulouse, Météo-France/CNRS, CNRM, Toulouse, France
[b] Université de Toulouse, CNRS/Cerfacs/IRD, CECI, Toulouse, France
[c] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

ARTICLE INFO

Dataset link: https://github.com/eliott-lumet/esmda_ppmles

Keywords:
Data assimilation
Uncertainty quantification
Observability
Surrogate modeling
Large-eddy simulation
Microscale pollutant dispersion

ABSTRACT

By capturing the physical complexity of the interactions between atmospheric flows and the built environment, Large-Eddy Simulations (LES) could provide detailed information for risk assessment and mitigation in case of environmental emergency. However, to account for LES uncertainties and cover the range of plausible scenarios in order to support decision making, it is necessary to go beyond deterministic simulation capability. This study introduces a novel ensemble-based data assimilation algorithm to correct the LES meteorological forcing and thereby improve LES spatial predictions of pollutant concentration by making use of available measurements. This approach is demonstrated through the MUST field-scale experiment. Results show that the ensemble smoother with multiple data assimilation (ESMDA) algorithm is a good candidate to address parameter interaction effects in the relationship between uncertain meteorological forcing and LES field quantities. This iterative algorithm is computationally feasible when the LES model is replaced with a machine learning-based surrogate model, from which robust ensemble statistics can be extracted. This surrogate-based data assimilation approach can then be used to examine observability in the system. Results show that the estimation outcome is highly sensitive to the design of the observation network, and that this sensitivity may be underestimated in idealized experiments. It is therefore important to use real data assimilation to optimize sensor placement and extract informative data for modeling, thus improving our ability to monitor accidental dispersion events.

1. Introduction

By 2050, it is estimated that two-thirds of the world's population will live in cities [1]. In response, urbanized areas will expand rapidly [2], multiplying the number of interfaces with forests and industrial sites, thereby increasing the vulnerability of populations to accidental releases of pollutants into the atmosphere due to wildfires and industrial accidents. Rapid and realistic assessment of air pollution risks is therefore critical to making cities safer and more resilient.[2] To address this issue, new microscale pollutant dispersion modeling strategies must be developed, which accurately represent the pollutant concentration hotspots induced by the urban landscape, and the various uncertainties inherent to the Atmospheric Boundary Layer (ABL), in order to cover all plausible scenarios for emergency response.

Computational Fluid Dynamics (CFD) modeling approaches, such as Reynolds-averaged Navier–Stokes and Large-Eddy Simulation (LES), are increasingly recognized as essential tools for predicting atmospheric flow and pollutant dispersion in urban areas at high spatial resolution [3–5]. CFD approaches explicitly solve the flow governing equations and represent the interactions between the atmosphere and the urban landscape. However, their substantial computational cost makes them impractical for real-time applications such as emergency response. Moreover, the validation of CFD urban flow predictions against field measurements shows that their accuracy is limited, even compared to operational or empirical models [6–8]. This is because the higher the complexity of the model, the more sensitive it becomes to input data and model parameter uncertainty [9].

Microscale CFD dispersion models are subject to uncertainty in their boundary conditions: i.e. in the meteorological forcing [10–12], the urban geometry representation [13–15], and the pollutant source intensity and location [16,17]. They are also subject to structural

* Corresponding author.
E-mail address: eliott.lumet@umr-cnrm.fr (E. Lumet).

[1] Deceased author. Dr. Simon Lacroix, has sadly passed away during the preparation of the manuscript. Dr. Simon Lacroix made significant supervising on intellectual contributions to the works presented in this manuscript.

[2] UN Sustainable Development Goal #11: https://sdgs.un.org/goals/goal11 (Accessed: 27/06/2025).

https://doi.org/10.1016/j.buildenv.2025.113863
Received 30 June 2025; Received in revised form 29 September 2025; Accepted 12 October 2025
Available online 16 October 2025

Nomenclature

Abbreviations

ABL	Atmospheric Boundary Layer
ASU	Arizona State University
CFD	Computational Fluid dynamics
DA	Data Assimilation
DPID	Digi-Photoionization Detector
EnKF	Ensemble Kalman Filter
ESMDA	Ensemble Smoother with Multiple Data Assimilation
FAC2	Fraction of predictions within a factor of two of the observations
LES	Large-Eddy Simulation
MUST	Mock Urban Setting Test *(field experiment)*
NMSE	Normalized Mean Square Error
PDF	Probability Density Function
POD–GPR	Proper Orthogonal Decomposition *with* Gaussian Process Regression *(surrogate modeling technique)*
PPMLES	Perturbed-Parameter ensemble of MUST Large-Eddy Simulations *(dataset)*
UVIC	Ultraviolet Ion Collector
VG	Geometric Variance

Symbols

$<\cdot>^a$	Analysis value
$<\cdot>^b$	Background value
$<\cdot>^p$	Model prediction
$<\cdot>^t$	Truth value
$<\cdot>^{(ref)}$	Reference value
α	Inflation coefficient in the ESMDA
α_{inlet}	Inlet wind direction
$\boldsymbol{\Psi}_\ell$	ℓth POD vector mode
ϵ	Absolute error
κ	von Kármán constant
$\mathbb{E}$	Expected value
$\mathbb{P}$	Probability density function
$\mathbb{V}$	Variance
$\mathcal{H}$	Observation operator
$\mathcal{M}$	Model function
$\mathcal{N}$	Normal probability density function
$\mathbf{B}$	Background error covariance matrix
$\mathbf{c}$	Concentration field
$\mathbf{G}$	Tangent linear of the generalized observation operator $\mathcal{H}\circ\mathcal{M}$
$\mathbf{K}^*$	Kalman gain matrix
$\mathbf{R}$	Observation error covariance matrix
$\mathbf{S}$	Spatial Sobol' indice
$\mathbf{u}$	Mean inlet wind velocity vector
μ	Bootstrap replicate
σ	Standard deviation
ε_μ	Measurement relative uncertainty
$\mathbf{x}$	System state
$\mathbf{y}^o$	Concentration observations
c_t	Concentration measurement threshold
k_ℓ	ℓth POD coefficient
N_a	Number of iterations (outer loops) in the ESMDA
N_e	Ensemble size in the ESMDA
N_{obs}	Number of observations
u_*	Friction velocity
z_0	Aerodynamic roughness length
$\boldsymbol{\theta}$	Control vector of the meteorological forcing parameters

modeling uncertainty, mostly related to numerical errors and turbulence modeling assumptions and parameterization [18–22]. In addition, the turbulent nature of the ABL induces a significant aleatory and irreducible uncertainty, known as internal variability, which could explain a large part of the discrepancies between field measurements and CFD model predictions [4,6,7,23].

Recent studies have highlighted the remarkable ability of data assimilation (DA) to enhance microscale atmospheric CFD models. By optimally merging observation data with LES model predictions, DA has been successfully used to reduce uncertainties in meteorological forcing parameters [24–28], pollutant source parameters [29–31], and turbulence modeling [22]. These studies have provided tentative guidelines on how to formulate the DA problem for microscale flow applications. In particular, they have shown that *(i)* it is more relevant to estimate boundary conditions than initial conditions in the CFD context [32,33]; *(ii)* there is a need to reduce CFD model computational cost to provide a priori (or background) information to the estimation problem using either surrogate models [25,26], dimension reduction techniques [30,33,34] or DA algorithms that require few model queries [27,28].

Building on these advances, this study proposes a novel surrogate-based ensemble DA system to further improve predictions of microscale urban pollutant dispersion. The system aims to provide faster and more accurate probabilistic dispersion predictions by inferring meteorological forcing parameters through the assimilation of in-situ pollutant concentration measurements. The MUST field campaign [35] is used as a realistic case study, to demonstrate the potential and robustness of the proposed DA system. This work addresses two particular issues: (i) reducing uncertainties and identifying those that still affect state-of-the-art LES dispersion models; and (ii) analyzing the sensitivity of the DA results to the number and position of assimilated observations [25,30].

The sensitivity to the observation location is closely related to the question of observability of the system. If the available observations are not in areas influenced by the inferred global parameters, the DA problem becomes ill-posed and the DA process cannot converge properly. One novelty of this study is the implementation of the Ensemble Smoother with Multiple Data Assimilation (ESMDA) algorithm of Emerick and Reynolds [36] to better address the nonlinear relationship between the estimated parameters and the observable quantities. Another strength of the proposed DA system is the online integration of a surrogate model, referred to as POD–GPR [37,38]. Trained offline using the PPMLES[3] dataset [39], this surrogate reduces the online prediction time by five orders of magnitude compared to LES [37,40] and can therefore provide large-ensemble predictions to efficiently estimate a wide range of pollutant concentration scenarios within the ESMDA. This reduction in computational cost is leveraged to go one step further in the validation of the DA system by evaluating its robustness to initial bias in background parameters, ensemble sampling, and observation location. This provides insights into the observability and equifinality [41] properties of the DA system.

[3] Perturbed-Parameter ensemble of MUST Large-Eddy Simulations, dataset available at 10.5281/zenodo.11394347.

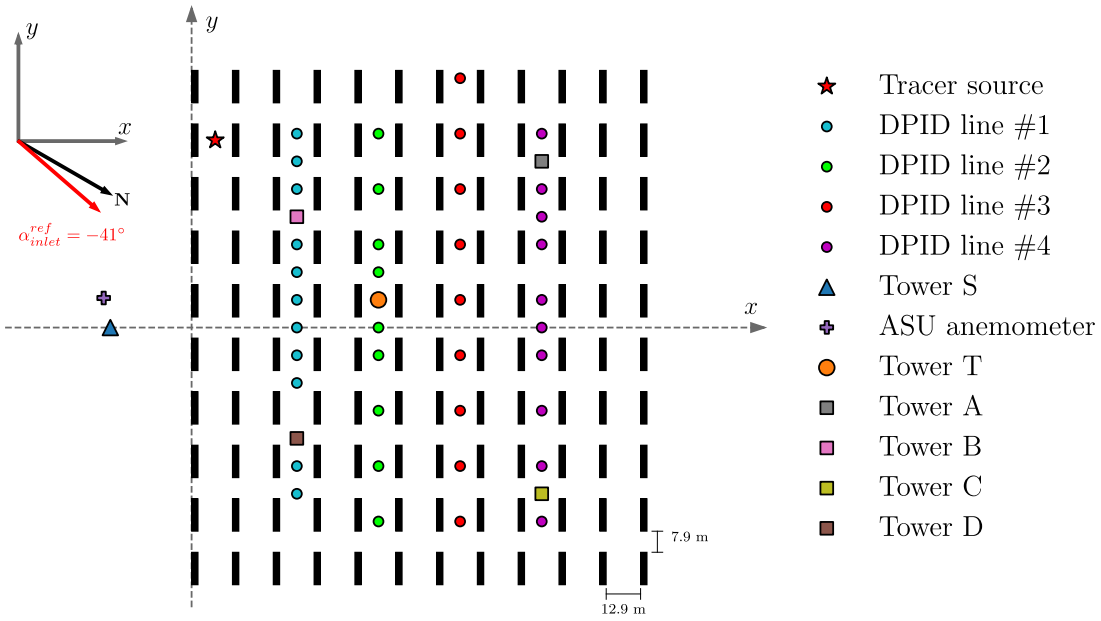

Fig. 1. Schematic view of the MUST array configuration. The coordinate system used is the same as in Yee and Biltoft [35], so that north corresponds to the angle of −30° in the x–y system. Black rectangles represent the shipping containers used to mimic the urban canopy. The red star represents the location of the propylene source for the trial #2681829. The red arrow corresponds to the reference upstream wind direction estimated from the tower S and ASU anemometer measurements. The colored circles correspond to the DPID concentration samplers, and the colored squares correspond to the UVIC concentration samplers mounted on towers A, B, C, and D (note that there is a DPID sampler at the same location as tower D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This article is organized as follows. Section 2 briefly presents the MUST use case, including the experimental trial and the existing LES and POD–GPR models. Section 3 introduces the DA system to improve LES predictions of pollutant dispersion. The validation of the DA system against the MUST trial is studied for idealized observations in Section 4 and for real observations in Section 5. Finally, the sensitivity of the DA system to observation network design is investigated in Section 6.

2. Experimental data and models for the MUST case study

This section introduces the three main components of the case study: *(i)* the MUST field data, which provide tracer concentration measurements for the DA system and its validation; *(ii)* the reference LES pollutant dispersion model [23] to be improved by DA; and *(iii)* the POD–GPR surrogate model [40] to be integrated within the DA system to reduce inline prediction computational costs.

2.1. The MUST field campaign

The Mock Urban Setting Test (MUST) field campaign was conducted in September 2001 at the US Army Dugway Proving Ground test site in the Utah desert, USA, to provide comprehensive measurements of air pollutant dispersion within an urban-like canopy [35,42]. It is a canonical case study for the development, validation and intercomparison of CFD dispersion models [43]. Although it is a highly controlled experiment (simplified geometry, isolated site and well-instrumented to define boundary conditions), local differences between CFD model predictions and observations are not yet fully explained [23,44–46]. This makes MUST a relevant case study in which to investigate the potential of DA to improve CFD predictions [28].

The MUST campaign consists of several trials in which a non-reactive gas tracer was released within an array of 10 × 12 regularly-spaced shipping containers, mimicking an idealized urban canopy (Fig. 1). This work focuses on the #2681829 trial corresponding to neutral atmospheric conditions and for which the tracer source was located between the containers at 1.8 m above ground (red star in Fig. 1). The objective here is to predict the time-averaged tracer concentration, defined over the standard 200-s analysis period extracted by Yee and Biltoft [35] to minimize mesoscale meteorological variability.

During the MUST campaign, wind velocity and direction were measured within the container array by four anemometers mounted on the central tower T at z = 4, 8, 16 and 32 m (orange circle in Fig. 1), and upstream of the container array by three anemometers mounted on the tower S at z = 4, 8 and 16 m (blue triangle in Fig. 1). Another sonic anemometer, provided by the Arizona State University (ASU), measured the wind velocity upstream of the containers at z = 1.6 m (purple plus symbol in Fig. 1). A network of meteorological stations was also installed throughout the Dugway Proving Ground, with the closest station to the container array, about 1 600 m to the southeast. Each station measures wind velocity at z = 10 m, with sampling times ranging from 5 to 15 min.

Tracer concentration was measured using 48 digi-photoionization detectors (DPID) and 24 ultraviolet ion collectors (UVIC), shown as colored circles and squares, respectively, in Fig. 1. The DPIDs have a detection threshold of 0.04 ppm versus 0.01 ppm for the UVICs. The horizontal distribution of the tracer concentration was measured by 40 DPIDs located at the same height (z = 1.6 m) within the canopy, forming four DPID lines (one color per line in Fig. 1). The vertical distribution of the tracer concentration was tracked by five towers equipped with either 8 DPIDs (tower T at z = 1, 2, 4, 6, 8, 10, 12 and 16 m) or 6 UVICs (towers A, B, C, D at z = 1, 2, 3, 4, 5, 5.9 m). For a complete description of the instruments used during the MUST campaign, the reader is referred to Biltoft [42].

2.2. Large-eddy simulation of the MUST experiment

An LES numerical model is used to reproduce the MUST trial under varying meteorological forcing. The model relies on the AVBP[4] [47,48] code to solve the LES-filtered Navier–Stokes and tracer transport

[4] AVBP documentation, see https://www.cerfacs.fr/avbp7x/

equations with the second-order Lax–Wendroff numerical scheme [47]. Subgrid momentum transport is represented with the wall-adaptative local eddy-viscosity model [49]. Tracer transport relies on a gradient-diffusion approach with a standard turbulent Schmidt number of 0.6.

The MUST field geometry is discretized with a 91-million-tetrahedron, unstructured mesh with a 0.3-m resolution at container level. This resolution is deemed sufficient based on LES guidelines to capture flow interaction with urban-type obstacles [50,51]. The shipping container height is 2.54 m high and is therefore discretized with approximately 8 grid points. Standard boundary conditions following the MUST intercomparison guidelines are used [51]. The LES model was validated in Lumet et al. [23], achieving state-of-the-art level of accuracy in reproducing the MUST trial #2681829. A complete description of the LES model, including its boundary conditions, is provided in Lumet [52], Chapter II. The following section focuses on the meteorological forcing, as it incorporates the main parameters to infer with the DA system.

2.3. Meteorological forcing parameterization

The large-scale meteorological forcing is represented in the LES model by a standard logarithmic wind profile [53]. With this profile, corresponding to a fully-developed neutral atmospheric surface layer, the mean inlet wind velocity vector $\mathbf{u}$ reads:

$$\mathbf{u}(z) = \frac{u_*}{\kappa} \ln\left(\frac{z+z_0}{z_0}\right) \begin{pmatrix} \cos(\alpha_{inlet}) \\ \sin(\alpha_{inlet}) \\ 0 \end{pmatrix}, \tag{1}$$

where κ is the von Kármán constant equal to 0.4; z_0 is the aerodynamic roughness length estimated to be 0.045 ± 0.005 m by Yee and Biltoft [35]; α_{inlet} is the mean wind direction; and u_* is the friction velocity.

Temporal wind fluctuations are added to the mean inlet wind profile using the synthetic turbulence injection method of Smirnov et al. [54]. The imposed Reynolds stress tensor is estimated from a precursor free-field simulation inspired by the work from Vasaturo et al. [55]. Assuming that the atmosphere is at the equilibrium, the velocity fluctuations can then be rescaled accordingly to the actual wind friction velocity u_* and direction α_{inlet} as explained in Lumet et al. [23].

Control vector definition The meteorological forcing is parameterized by two parameters, which are gathered in the control vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = (\alpha_{inlet}, u_*)^{\mathrm{T}}. \tag{2}$$

$\boldsymbol{\theta}$ is the DA estimation target, implying that the DA system aims to infer $\boldsymbol{\theta}$ based on the tracer concentration measurements $\mathbf{y}^o$ in order to improve pollutant dispersion prediction.

This choice of control vector $\boldsymbol{\theta}$ is widely used in urban flows studies involving surrogate modeling [10,56,57] and/or DA [25,26,30]. With this choice, there are only two degrees of freedom for DA. This control vector could be augmented by using a more complex parameterization of the meteorological forcing [27], but that is beyond the scope of this study.

Reference meteorological parameters Reference values of the control parameters $\boldsymbol{\theta}^{(ref)} = \left(\alpha_{inlet}^{(ref)}, u_*^{(ref)}\right)^{\mathrm{T}}$ are required to carry out DA idealized experiments and validate the DA estimates of the control vector $\boldsymbol{\theta}$ (Section 4). Here the reference is defined based on prior work [23], using unassimilated wind velocity measurements from the selected MUST trial. The reference inlet wind direction, $\alpha_{inlet}^{(ref)} = -41\,°$, corresponds to the mean of the time-averaged measurements from the tower S and ASU anemometers (located 30 m upstream of the container array, see Fig. 1). The reference friction velocity, $u_*^{(ref)} = 0.73\,\mathrm{m\,s^{-1}}$, is obtained by fitting the logarithmic wind profile (Eq. (1)) to the sonic anemometer measurements.

2.4. The POD–GPR surrogate model

Predicting the 200-s averaged concentration field $\mathbf{c}$ with the LES model (Section 2.2) costs approximately 20 000 core hours (using 1344 Intel Skylake CPU cores). Such a computational load would hinder the operability of the DA system and limit the sensitivity tests that can be carried out for validation purposes. This motivates the use of a surrogate model that can quickly and accurately emulate the response of the LES tracer concentration field $\mathbf{c}$ to the selected meteorological forcing parameters $\boldsymbol{\theta}$ (Eq. (2)).

This surrogate model, designed and evaluated in Lumet et al. [40], adopts the POD–GPR machine learning approach from Marrel et al. [37]. Proper Orthogonal Decomposition (POD) reduces the dimension of the emulated field $\mathbf{c}$, from the order of million grid elements to only 10 POD modes $\{\boldsymbol{\Psi}_\ell\}_{\ell=1}^{10}$. The choice of a POD latent space dimension equal to 10 is a trade-off between minimizing reconstruction error and reducing overfitting of noise present in the training set. Then, Gaussian Process Regression (GPR) is used to learn the relationship between the input parameters $\boldsymbol{\theta}$ and the coefficients $\{k_\ell\}_{\ell=1}^{10}$ of the projection of the concentration field $\mathbf{c}$ onto the POD basis. Fig. 2 illustrates how new POD–GPR predictions are obtained (i.e. for meteorological forcing parameters that are not present in the training dataset).

The POD–GPR model was trained offline using the PPMLES[3] dataset [39] made of 200 LES simulations for varying meteorological forcing parameters

$$\boldsymbol{\theta} \in [-90\,°,\ 30\,°] \times [0.07\,\mathrm{m\,s^{-1}},\ 0.89\,\mathrm{m\,s^{-1}}], \tag{3}$$

which define the parametric bounds and thus the validity range of the surrogate model. The total computational cost of this dataset was about 5.7 million of core hours. Results show that the POD–GPR model reduces prediction time by five orders of magnitude compared to LES in the online prediction step. It achieves the best possible accuracy levels given the internal variability in the validation data, except for the highest concentrations near the source, and it is robust to training data noise [40].

An interesting feature of the POD–GPR surrogate model is that it provides a probabilistic prediction of the tracer concentration field $\mathbf{c}$, i.e. $\mathcal{M}_{\mathrm{POD-GPR}} : \boldsymbol{\theta} \mapsto \mathbb{P}(\mathbf{c}|\boldsymbol{\theta})$. These distributions accurately represent a large fraction of surrogate modeling and internal variability errors [40]. The mean and standard deviation of two different POD–GPR distribution predictions are shown in Fig. 3 for the reference (Section 2.3) and background (Section 3.3) parameter values for illustration purposes. This figure illustrates the local drop in accuracy of the POD–GPR model near the emission source, where tracer concentrations are the highest, while the surrogate model error is very low in the rest of the field. A detailed analysis in Lumet [52] showed that it is not possible with POD–GPR to achieve a remarkable level of accuracy both in the vicinity of the emission source and in the near field. The surrogate model was then calibrated for optimal performance in the near field, where the MUST sensors are located. In this study, the POD–GPR distributions are sampled to account for the surrogate modeling and internal variability errors in the DA system, as further explained in Section 3.2.2.

3. Data assimilation methodology

3.1. Key features of the data assimilation system

In this work, a novel DA system is designed to improve the estimation of the time-averaged concentration field $\mathbf{c}$ in the event of a local pollutant release. This is done by combining the LES predictions with tracer concentration measurements $\mathbf{y}^o$ to infer the meteorological forcing parameters $\boldsymbol{\theta} = (\alpha_{inlet}, u_*)^{\mathrm{T}}$ (Eq. (2)), thus corresponding to the resolution of a parameter estimation problem. Anemometer measurements were not assimilated so they can serve as independent data to validate the ability of the DA system to accurately infer the meteorological forcing parameters.

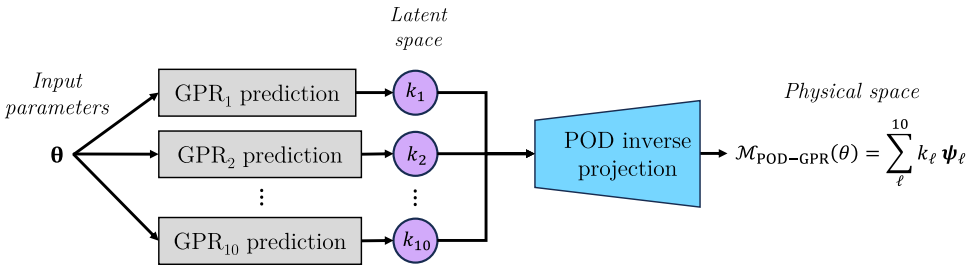

Fig. 2. Schematic of the POD–GPR surrogate model prediction principle, with one submodel for each of the ten POD modes. The pre-trained Gaussian processes (GPR_ℓ) predict the POD coefficients associated with any set of wind conditions $\boldsymbol{\theta}$ in the latent space. The inverse POD projection is applied to retrieve the associated 3-D physical field from the latent space.

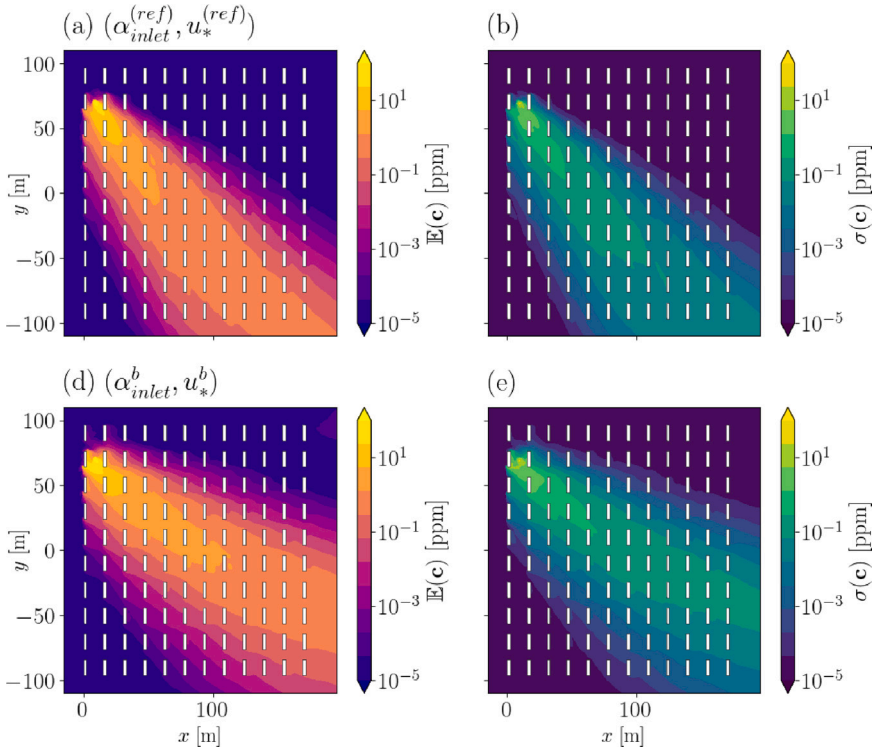

Fig. 3. Horizontal cuts at $z = 1.6\,\mathrm{m}$ of the POD–GPR surrogate model predictions for two sets of input parameters. The first row (a)–(b) corresponds to the reference parameters $(\alpha_{inlet}^{ref}, u_*^{ref}) = (-41°, 0.73\,\mathrm{m\,s^{-1}})$ estimated from tower S and ASU measurements. The second row (c)–(d) corresponds to the background parameters $(\alpha_{inlet}^{b}, u_*^{b}) = (-25°, 0.57\,\mathrm{m\,s^{-1}})$. The mean $\mathbb{E}(\mathbf{c})$ of the predicted POD–GPR concentration field distribution is given in the first column; its standard deviation $\sigma(\mathbf{c})$ is given in the second column.

The DA problem is solved using the Ensemble Smoother with Multiple Data Assimilation (ESMDA) of Emerick and Reynolds [36], which updates a prior distribution of the control parameters, called the *background*, based on the observations $\mathbf{y}^o$ and their uncertainties. The posterior distribution, called the *analysis*, can then be used to predict a new ensemble of pollutant dispersion scenarios with reduced uncertainty compared to the background. Inspired by the work of Rochoux et al. [58] and Sousa et al. [25], the LES model is replaced by a POD–GPR surrogate model (Section 2.4) within the ESMDA algorithm to substantially reduce prediction time and obtain robust ensemble statistics.

The main steps of the DA system are shown in Fig. 4. Note that in the present context, the meteorological forcing parameters $\boldsymbol{\theta}$, the measurements $\mathbf{y}^o$ and the concentration fields $\mathbf{c}$ are time-averaged and considered stationary. This implies that the underlying system is static and a single DA cycle is performed.

To go into more detail, the ESMDA algorithm is presented in Section 3.2; the modeling of the uncertainties in the background parameters and in the observations is presented in Section 3.3 and Section 3.4, respectively.

3.2. The ensemble smoother with multiple data assimilation algorithm

The ESMDA algorithm of Emerick and Reynolds [36,59] is an ensemble-based DA method that can be regarded as an iterative extension of the Ensemble Kalman Filter (EnKF) [60,61]. It has been shown to improve performance in highly nonlinear estimation problems [59, 62], which is particularly relevant for inferring wind direction, as this parameter induces strong structural changes in the field quantities of interest [25,63]. These benefits of the ESMDA algorithm, however, come at the expense of a larger number of model evaluations. This drawback is mitigated here by the use of a surrogate model.

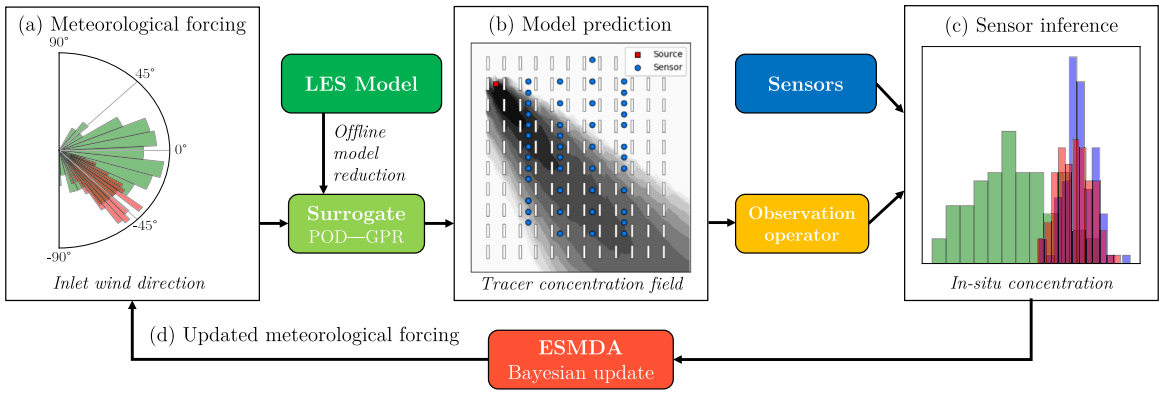

Fig. 4. Schematic diagram of the surrogate-based DA system used to estimate the meteorological forcing parameters θ, inspired by Sousa et al. [25]. (a) The background ensemble (in green) is generated to represent the range of uncertain meteorological forcing parameters. Only the inlet wind direction distribution is shown, but the friction velocity is also inferred. (b) From this background ensemble, an ensemble of tracer concentration fields is predicted using the pretrained POD–GPR surrogate model. (c) The surrogate model predictions are mapped onto the observation space and compared with the sensor measurements (in blue). (d) The ESMDA algorithm is used to solve the inverse problem and infer an updated ensemble of meteorological forcing parameters (in red), which minimizes the discrepancies with observations.

3.2.1. ESMDA theoretical principle

The ESMDA problem can be considered from a probabilistic viewpoint, as the problem of estimating the probability density function (PDF) $\mathbb{P}(\boldsymbol{\theta}|\mathbf{y}^o)$ of the forcing parameters $\boldsymbol{\theta}$ conditional on the measurements $\mathbf{y}^o$. Using Bayes' theorem, this posterior PDF reads:

$$\mathbb{P}(\boldsymbol{\theta}|\mathbf{y}^o) \propto \mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})\,\mathbb{P}(\boldsymbol{\theta}), \tag{4}$$

where $\mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})$ represents the data likelihood, i.e., the PDF of the observations $\mathbf{y}^o$ given the control parameters $\boldsymbol{\theta}$. The principle of the ESMDA is to recursively sample this posterior PDF over a predefined number of outer loops N_a for a given DA cycle, noting that it can be written as:

$$\mathbb{P}(\boldsymbol{\theta}|\mathbf{y}^o) \propto \mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})^{\left(\sum_{k=1}^{N_a} \frac{1}{\alpha_k}\right)}\,\mathbb{P}(\boldsymbol{\theta}), \tag{5}$$

so that

$$\begin{cases} \mathbb{P}(\boldsymbol{\theta}|\mathbf{y}^o) \propto \mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})^{\frac{1}{\alpha_1}}\; p_1(\boldsymbol{\theta}|\mathbf{y}^o), & \text{(6a)} \\ p_1(\boldsymbol{\theta}|\mathbf{y}^o) \propto \mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})^{\frac{1}{\alpha_2}}\; p_2(\boldsymbol{\theta}|\mathbf{y}^o), & \text{(6b)} \\ \quad\vdots \\ p_{N_a-1}(\boldsymbol{\theta}|\mathbf{y}^o) \propto \mathbb{P}(\mathbf{y}^o|\boldsymbol{\theta})^{\frac{1}{\alpha_{N_a}}}\;\mathbb{P}(\boldsymbol{\theta}), & \text{(6c)} \end{cases}$$

provided that

$$\sum_{k=1}^{N_a} \frac{1}{\alpha_k} = 1. \tag{7}$$

These N_a estimation problems are solved iteratively, using an EnKF to assimilate the same observations $\mathbf{y}^o$, but with an inflation of the observation uncertainty by α_k at each iteration. Thus, ESMDA computes a sequence of small linear updates, while EnKF computes one large linear update over a given DA cycle. This explains why ESMDA can increase the accuracy of parameter estimation in highly nonlinear problems [62]. As the EnKF, the ESMDA algorithm provides an exact solution to Bayes' theorem only when the prediction model is linear and when the parameters $\boldsymbol{\theta}$ and the observations $\mathbf{y}^o$ are unbiased and normally distributed.

3.2.2. ESMDA algorithm

The ESMDA algorithm follows the three following steps:

- **1. Initialization**: In the first iteration ($k = 0$), the background ensemble $\{\boldsymbol{\theta}_i^b\}_{i=1}^{N_e}$ (shown in green in Fig. 4a) is drawn from the first guess of the meteorological forcing parameters $\boldsymbol{\theta}^b$ and the background error covariance matrix $\mathbf{B}$:

$$\boldsymbol{\theta}_i^b \sim \mathcal{N}(\boldsymbol{\theta}^b, \mathbf{B}), \quad 1 \le i \le N_e. \tag{8}$$

The background sample is generated using the low-discrepancy sequence from Halton [64], rescaled to the normal distribution $\mathcal{N}(\cdot,\cdot)$. Using this quasi Monte Carlo approach instead of pure random sampling limits the variability of the ESMDA analysis and reduces sampling errors as further explained in Appendix A.

- **2. Prediction step:** Each member of the background ensemble is propagated to the state space by randomly sampling the distribution $\mathbb{P}(\mathbf{c}|\boldsymbol{\theta}_i^b)$ inferred by the POD–GPR surrogate model:

$$\mathbf{x}_i^f \sim \mathbb{P}(\mathbf{c}|\boldsymbol{\theta}_i^b), \quad 1 \le i \le N_e. \tag{9}$$

This sampling allows to account for the errors associated with the use of the surrogate model instead of the LES model and with the internal variability of the ABL [40]. Here, the predicted state $\mathbf{x}^f$ corresponds to the time-averaged concentration field $\mathbf{c}$ (Fig. 4b).

- **3. Analysis step:** Before calculating the actual analysis, the observations are perturbed so that the analysis error statistics are not underestimated [65,66]. In the ESMDA framework, the observation errors must be inflated by a factor $\sqrt{\alpha_k}$ for the same reason [59]. Consequently, an ensemble of perturbed observations is drawn from the real observations $\mathbf{y}^o$, knowing the observation error covariance matrix $\mathbf{R}$, as:

$$\mathbf{y}_i^o \sim \mathcal{N}(\mathbf{y}, \alpha_k \mathbf{R}), \quad 1 \le i \le N_e. \tag{10}$$

It is necessary to resample the perturbed observations at each iteration (or outer loop) to avoid bias from using dependent samples [62].

Once observations are perturbed, the Kalman filter update equation is applied to update each background member $\boldsymbol{\theta}_i^b$ through each analysis member $\boldsymbol{\theta}_i^a$ to better match the observations:

$$\boldsymbol{\theta}_i^a = \boldsymbol{\theta}_i^b + \mathbf{K}^* \left(\mathbf{y}_i^o - \mathcal{H}(\mathbf{x}_i^f)\right), \quad 1 \le i \le N_e, \tag{11}$$

where $\mathcal{H}(\cdot)$ is the observation operator that interpolates the predicted concentration fields $\mathbf{x}_i^f$ at the assimilated sensor locations (Fig. 4c), and where $\mathbf{K}^*$ is the Kalman gain matrix defined classically as:

$$\mathbf{K}^* = \mathbf{B}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{B}\mathbf{G}^{\mathrm{T}} + \alpha_k \mathbf{R})^{-1}, \tag{12}$$

where $\mathbf{G}$ is the tangent linear of the generalized observation operator $\mathcal{H}\circ\mathcal{M}$ (i.e. combining the observation operator $\mathcal{H}(\cdot)$ and

the model operator $\mathcal{M}(\cdot)$ to map the parameter space onto the observation space in a parameter estimation framework), and where the observation uncertainty term also involves the multiplicative factor α_k to prevent ESMDA from underestimating the analysis error statistics. In practice, the error cross-covariance matrix $\mathbf{BG}^{\mathrm{T}}$ (mapping uncertainty from the control space onto the observation space) and the error covariance matrix $\mathbf{GBG}^{\mathrm{T}}$ (in the observation space) are estimated directly from the ensemble members:

$$\begin{cases} \mathbf{BG}^{\mathrm{T}} \approx \dfrac{1}{N_e-1}\displaystyle\sum_{i=1}^{N_e}(\boldsymbol{\theta}_i^b-\overline{\boldsymbol{\theta}^b})\left(\mathcal{H}(\mathbf{x}_i^f)-\overline{\mathcal{H}(\mathbf{x}_i^f)}\right)^{\mathrm{T}}, & \text{(13a)} \\ \mathbf{GBG}^{\mathrm{T}} \approx \dfrac{1}{N_e-1}\displaystyle\sum_{i=1}^{N_e}\left(\mathcal{H}(\mathbf{x}_i^f)-\overline{\mathcal{H}(\mathbf{x}_i^f)}\right)\left(\mathcal{H}(\mathbf{x}_i^f)-\overline{\mathcal{H}(\mathbf{x}_i^f)}\right)^{\mathrm{T}}. & \text{(13b)} \end{cases}$$

where the overbar $\bar{\cdot}$ denotes the average over the ensemble. The resulting ensemble $\{\boldsymbol{\theta}_i^a\}_{i=1}^{N_e}$ obtained from Eq. (11) is called the analysis ensemble (in red in Fig. 4a). It becomes the background ensemble for the next iteration (or outer loop).

3.2.3. ESMDA settings

In this study, constant inflation coefficients are used over the iterations, i.e. $\alpha_k = N_a, 1 \le k \le N_a$. Using other sequences did not have a significant effect on the analysis, which is consistent with Emerick and Reynolds [36].

The number of outer loops, $N_a \in 1, 2, 4, 6$, is changed in Section 4 to assess the benefit of applying ESMDA to parameter estimation in microscale pollutant dispersion. This setup also enables comparison with the standard EnKF, a benchmark algorithm for parameter estimation [67], which is strictly equivalent to ESMDA when $N_a = 1$.

3.3. Background parameters and associated uncertainty

This section explains how the background parameters $\boldsymbol{\theta}^b$ and their uncertainty (represented by the background error covariance matrix $\mathbf{B}$) are defined. This is crucial because the background parameter distributions correspond to the initial first guess of the DA system (Fig. 4a).

3.3.1. Background error modeling

The background parameter uncertainty is quantified from 12 days of wind velocity measurements taken at two distinct meteorological stations, each of which was located a few kilometers away from the MUST site. The standard deviation of the difference between the two station measurements provides the coefficients of the matrix $\mathbf{B}$:

$$\mathbf{B} = \begin{pmatrix} \sigma^2(\alpha_{inlet}) & 0 \\ 0 & \sigma^2(u_*) \end{pmatrix}, \quad (14)$$

with $\sigma(\alpha_{inlet}) = 25°$ and $\sigma(u_*) = 0.09\,\mathrm{m\,s^{-1}}$. As the friction velocity and wind direction errors were found to be uncorrelated (Lumet [52], Chapter V), the matrix $\mathbf{B}$ is diagonal.

3.3.2. Choice of the background parameter values

In this study, the baseline background parameters are obtained by adding large perturbations ($\Delta\alpha_{inlet} = +16°$, $\Delta u_* = +0.16\ \mathrm{m\,s^{-1}}$) to the reference parameters $\boldsymbol{\theta}^{(ref)}$ (defined in Section 2.3). This results in the background $(\alpha_{inlet}^b, u_*^b)^{\mathrm{T}} = (-25°, 0.57\ \mathrm{m\,s^{-1}})^{\mathrm{T}}$. This choice enables assessment of the DA system's ability to correct significant bias in the meteorological forcing parameters. Such a background scenario is realistic, since substantial discrepancies often arise between the local incident flow and measurements from the nearest weather stations used to specify the meteorological forcing parameters. Furthermore, to evaluate the robustness of the DA system to background bias, this work studies the DA performance for different values of the background parameters $\boldsymbol{\theta}^b$ (Section 4.3).

3.3.3. Specific treatment of the friction velocity parameter

To ensure that the friction velocity estimated by DA is always positive, a logarithmic anamorphosis is applied to the friction velocity taking inspiration from what has been done on concentration anamorphosis by Liu et al. [68] and Defforge et al. [28] (see further discussion in Section 3.4.3). The DA algorithm estimates the logarithmic friction velocity $\widetilde{u}_* = \ln(u_*)$, instead of the friction velocity u_*. This ensures that the estimates of u_* always remain positive and that the assumption of a normal distribution of errors is better satisfied. Indeed, it was found that the assumption that $\ln(u_*)$ follows a normal distribution is plausible, unlike u_*. Using this anamorphosis, the background error covariance matrix $\mathbf{B}$ and the unbiased background parameters vector $\boldsymbol{\theta}^b$ become:

$$\mathbf{B} = \begin{pmatrix} \sigma^2(\alpha_{inlet}) & 0 \\ 0 & \sigma^2(\widetilde{u_*}) \end{pmatrix} \text{ with } \sigma^2(\widetilde{u_*}) = \ln\left(1+\frac{\sigma^2(u_*)}{(u_*^b)^2}\right), \quad (15a)$$

$$\boldsymbol{\theta}^b = (\alpha_{inlet}^b, \widetilde{u_*}^b)^{\mathrm{T}} \text{ with } \widetilde{u_*}^b = \ln(u_*^b) - \frac{1}{2}\sigma^2(\widetilde{u_*}). \quad (15b)$$

3.4. Assimilating observations of pollutant concentration

3.4.1. Baseline observation network

For this study, a first baseline observation network consists in tracer concentration measurements from 9 UVIC sensors on towers B, C and D at heights $z \in \{1, 2, 3\}$ m, which are located within the container array (Fig. 1). Unassimilated sensor data above the detection threshold are used for DA validation. The DA system is evaluated for this baseline observation network in Section 4 for idealized observations and in Section 5 for the real MUST observations.

To study the observability property of the DA system, Section 6 analyzes DA results for observation networks that differ from the baseline network in terms of the number and location of the observations.

3.4.2. Observation error modeling

In-situ tracer concentration measurements may be subject to significant uncertainty, notably due to instrument measurement errors and statistical errors resulting from the internal variability of the ABL [23]. To avoid overfitting these uncertain observations, a realistic observation error covariance matrix $\mathbf{R}$ modeling is developed for this study (see Appendix B).

3.4.3. Concentration anamorphosis

Since tracer concentration cannot be negative, the concentration anamorphosis proposed by Liu et al. [68] is used. It consists in assimilating log-transformed concentration observations $\widetilde{\mathbf{y}}^o$ instead of the actual concentration measurements $\mathbf{y}^o$:

$$\widetilde{\mathbf{y}}^o = \ln(\mathbf{y}^o + c_t), \quad (16)$$

where c_t is a threshold to avoid giving too much weight to low concentrations. In this study, this threshold is set to the maximum sensor detection threshold from the MUST experiment [35], i.e. $c_t = 0.04$ ppm. The same anamorphosis was used in other DA studies [28,30], with convincing results.

This transformation is based on the fairly standard assumption that concentration observations follow a log-normal distribution [69]. In this way, observation errors can be assumed to follow Gaussian distributions, ensuring greater consistency with the ESMDA theoretical principle and the error representation through the $\mathbf{R}$ observation error covariance matrix.

4. Assimilation of idealized observations

In this section, the DA system is validated and calibrated using *twin experiments*, also known as observing system simulation experiments (OSSE) in the literature [70]. In particular, the ESMDA and EnKF algorithms are systematically verified and compared, with attention to their sensitivity to background sampling errors.

Table 1
Setup of the baseline twin experiment for parameter estimation. The first row corresponds to the true parameters used to generate the synthetic observations given the measurement and associated error in the second row. The third row gives the background parameters and associated error. The last row gives the ESMDA parameters.

	Notation	Setup
True parameters	$\boldsymbol{\theta}^t = \left(\alpha_{inlet}^t, u_*^t\right)^{\mathrm{T}}$	$\left(-41\,^\circ, 0.73\,\mathrm{m\,s^{-1}}\right)^{\mathrm{T}}$
Observation network	$\mathbf{y}^o$	9 observations of concentration at towers B (1, 2, 3 m), C (1, 2, 3 m), and D (1, 2, 3 m)
Observation error	$\mathbf{R}$	see Appendix B
Background parameters	$\boldsymbol{\theta}^b = \left(\alpha_{inlet}^b, u_*^b\right)^{\mathrm{T}}$	$\left(-25\,^\circ, 0.57\,\mathrm{m\,s^{-1}}\right)^{\mathrm{T}}$
Background errors	$\mathbf{B} = \begin{pmatrix} \sigma^2(\alpha_{inlet}) & 0 \\ 0 & \sigma^2(u_*) \end{pmatrix}$	with $\sigma(\alpha_{inlet})25\,^\circ$, $\sigma(u_*) = 0.09\,\mathrm{m\,s^{-1}}$
Ensemble size	N_e	500 (see Appendix A)
Number of outer loops	N_a	{1, 2, 4, 6}
Anamorphosis threshold	c_t	0.04 ppm

4.1. Twin experiment principle and setup

In twin experiments, the true system state $\mathbf{x}^t$ is known and used to generate synthetic observations by adding a random noise representative of the observation error (Appendix B):

$$\mathbf{y}^o = \mathcal{H}(\mathbf{x}^t) + \epsilon^o, \quad \text{with } \epsilon^o \sim \mathcal{N}(0, \mathbf{R}). \tag{17}$$

In the present study, the true state corresponds to the POD–GPR surrogate prediction associated with the true meteorological forcing parameters, i.e. $\mathbf{x}^t = \mathcal{M}_{\mathrm{POD-GPR}}(\boldsymbol{\theta}^t)$.

Twin experiments provide an ideal testing framework as they give access to the true state of the system and control vector, which are unknown in real applications. This makes it possible to quantify the DA system accuracy in the control space (i.e. by comparing $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^t$) and in the state model space (i.e. by comparing $\mathbf{x}^a$ and $\mathbf{x}^t$). Twin experiments guarantee the existence of an exact solution to the inverse problem, thus avoiding the issue of parameter identification as a first step before moving on to the assimilation of real MUST data and offering a way to valid the DA system components.

Table 1 summarizes the setup of the baseline twin experiment (based on the baseline observation network) used to evaluate the DA system. The truth is defined using the reference estimate of the meteorological forcing parameters (Section 2.3): $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{(ref)} = (-41\,^\circ, 0.73\ \mathrm{m\,s^{-1}})^{\mathrm{T}}$. The ensemble size used in the ESMDA is set to $N_e = 500$, as shown in Appendix A to be sufficient for minimizing sampling errors.

4.2. Verification and calibration of the data assimilation system

This section presents the DA estimation results obtained with the baseline twin experiments (Table 1), both in the control space (Section 4.2.1) and the model state space (Section 4.2.3), in order to evaluate the added value of the ESMDA algorithm compared with the standard EnKF algorithm and to calibrate the number of outer loops N_a.

4.2.1. Parameter estimation step

First, the ability of the DA system to infer the meteorological forcing parameters is evaluated, as the control vector $\boldsymbol{\theta} = (\alpha_{inlet}, u_*)^{\mathrm{T}}$ is the direct DA estimation target. Table 2 shows that both ESMDA and EnKF reduce the wind direction bias compared to the background. However, the two algorithms differ significantly with regard to the friction velocity: the ESMDA is able to accurately estimate this parameter, whereas the standard EnKF fails to update it. For both parameters, the ESMDA algorithm demonstrates superior bias reduction as the number of outer loops (N_a) increases. The number of outer loops to achieve convergence remains small.

Table 2 also shows that both EnKF and ESMDA provide realistic uncertainty estimates. In all tests, the analysis errors and their error standard deviations are very close, implying that the true parameters lie within the 95% confidence interval around the estimated parameters.

Table 2
Baseline twin experiment parameter estimation error $\epsilon = \boldsymbol{\theta} - \boldsymbol{\theta}^t$ and ensemble standard deviation σ, for the background and the analysis. Results are given for the standard EnKF algorithm and the ESMDA algorithm with varying number of outer loops N_a.

	$\epsilon(\alpha_{inlet})$ (°)	$\sigma(\alpha_{inlet})$ (°)	$\epsilon(u_*)$ ($\mathrm{m\,s^{-1}}$)	$\sigma(u_*)$ ($\mathrm{m\,s^{-1}}$)
Background	16	25	−0.16	0.09
EnKF	6.4	8.8	−0.16	0.09
ESMDA – $N_a = 2$	2.4	2.4	−0.10	0.08
ESMDA – $N_a = 4$	0.9	1.8	−0.05	0.06
ESMDA – $N_a = 6$	0.6	1.6	−0.04	0.05

These estimates are consistent in that the less accurate the analysis the larger the estimated uncertainty. For instance, the EnKF does not reduce the uncertainty on the friction velocity $\sigma(u_*)$ compared to the background, as it is not able to infer this parameter.

4.2.2. Parameter sensitivity analysis

To better understand the difficulty of estimating friction velocity and to gain further insight into the DA results, the sensitivity of the predicted concentration to the inferred parameters is assessed using the first-order and second-order Sobol' indices ([71]):

$$\begin{cases} \mathbf{S}_{\alpha_{inlet}} = \dfrac{\mathbb{V}(\mathbb{E}(\mathbf{c}|\alpha_{inlet}))}{\mathbb{V}(\mathbf{c})}, & \text{(18a)} \\ \mathbf{S}_{u_*} = \dfrac{\mathbb{V}(\mathbb{E}(\mathbf{c}|u_*))}{\mathbb{V}(\mathbf{c})}, & \text{(18b)} \\ \mathbf{S}_{\alpha_{inlet},u_*} = \dfrac{\mathbb{V}(\mathbb{E}(\mathbf{c}|\alpha_{inlet}, u_*))}{\mathbb{V}(\mathbf{c})} - \mathbf{S}_{\alpha_{inlet}} - \mathbf{S}_{u_*}, & \text{(18c)} \end{cases}$$

where $\mathbb{V}(\cdot)$ denotes the variance and $\mathbb{E}(\cdot|\cdot)$ the conditional expected value. These Sobol' indices quantify the fraction of concentration variance explained by each input parameter as well as by the combined effect of the two parameters (the closer these indices are to one, the greater the contribution.). In this study, these spatial indices are estimated using the algorithm from Saltelli et al. [72] with 8 192 POD–GPR model evaluations. Fig. 5 shows that the variance of the tracer concentration prediction is mainly explained by the inlet wind direction α_{inlet}, while the friction velocity u_* alone has a negligible contribution except near the source. This result is obtained for each vertical level of the model, Fig. 5 showing a representative example at $z = 1.6$ m. This is because variations in α_{inlet} cause large changes in the plume position, while u_* only has an inversely proportional effect on the concentration. The higher sensitivity of concentration to wind direction explains why the DA algorithms, especially EnKF, better infer this parameter than friction velocity (Table 2).

Another noteworthy and rather unusual result is that the second-order Sobol' index reaches significant levels ($\mathbf{S}_{\alpha_{inlet},u_*} > 0.35$ on average), far exceeding the first-order index $\mathbf{S}_{u_*}$ linked to friction velocity

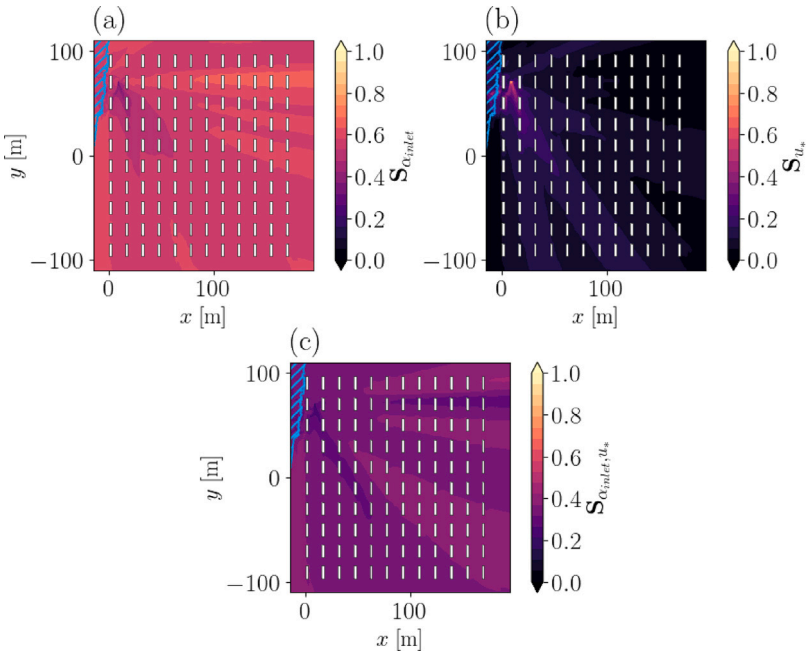

Fig. 5. Horizontal cuts at $z = 1.6$ m of the Sobol' index maps. (a) First-order index quantifying the sensitivity of the concentration POD–GPR prediction to the wind direction α_{inlet}. (b) First-order index for the friction velocity u_*. (c) Second-order Sobol' index showing the coupled effect of α_{inlet} and u^*. The blue hatched area corresponds to mesh nodes for which 99% of the concentration samples used to estimate the Sobol' indices are under 10^{-4} ppm.

(Fig. 5). This important coupled effect of the two parameters α_{inlet} and u_* is explained by the fact that the concentration relation with u_* is strongly dependent on the wind angle: if the plume passes at a given location, then the local concentration is approximately inversely proportional to u_*; if the location is outside the plume, the local concentration becomes independent of u_*. Hence, the sensitivity of the concentration to the friction velocity is conditioned by the wind direction, which explains the significant gain in accuracy of the ESMDA over the standard EnKF due to its iterative nature: during the first iteration, the ESMDA mainly corrects the predominant parameter α_{inlet} finding the most likely plume location; then the effect of u_* becomes more observable in the next iterations, allowing the ESMDA to infer it.

4.2.3. Propagation step to model state correction

In the ESMDA framework, the best estimate of the control vector is defined as the mean of the analysis ensemble $\{\theta_i^a\}_{i=1}^{N_e}$ (Eq. (11)). Using this mean estimate as input for the surrogate model generates a new estimate of the concentration field corresponding to the corrected meteorological forcing (Fig. 4). Fig. 6 shows that the gain in accuracy into the control space propagates well to the state space, with an analysis concentration field very close to the true field when using the ESMDA with $N_a = 4$ outer loops. This demonstrates the ability of the DA system to propagate corrections from the parameter space to the model state space (which is not directly estimated by the DA algorithm).

4.2.4. Computational efficiency

The DA system appears to be highly cost-effective, with runtimes of 50 s for the EnKF and 190 s for the ESMDA with $N_a = 4$ outer loops (using a single Intel Ice Lake processor core). For comparison, a single assimilation with the LES model and 10 members would require 200 000 core hours. This illustrates the relevance of using a surrogate model and proves that the proposed DA framework could be suitable for real-time applications. Note that most of the execution time corresponds to the $(N_a \times N_e + 1)$ surrogate model calls, and that this process could be significantly accelerated by parallelizing model calls in each DA outer loop.

4.3. Robustness to the choice of the background

To evaluate the performance of the DA system in a systematic way, a Monte Carlo experiment is carried out by evaluating both the EnKF and ESMDA (with $N_a = 4$ outer loops) for a range of background conditions, i.e. for 50 different sets of background parameters. Fig. 7 summarizes the DA results in terms of analysis bias versus background bias for the two control parameters, the inlet wind direction (a, b) and the friction velocity (c, d).

The ESMDA systematically improves the wind direction estimation regardless of the initial background bias. The gain in accuracy is less important for the EnKF. For some samples, the analysis is even less accurate than the background (Figs. 7a, b). Furthermore, the analysis correction of the wind direction tends to deteriorate when the initial background bias on the friction velocity $\epsilon^b(u_*)$ is large, for both EnKF and ESMDA. This is explained by the fact that, during its first iteration, the DA algorithm tends to overcorrect α_{inlet} to explain deviations in the observations that are actually due to bias in u_*.

Regarding the friction velocity estimation, Fig. 7c confirms that the EnKF is not able to correct the friction velocity regardless of the initial background bias. In contrast, Fig. 7d shows that the ESMDA with $N_a = 4$ outer loops significantly improves the friction velocity estimation.

In conclusion, the twin experiments demonstrate that the ESMDA algorithm is more accurate than the standard stochastic EnKF algorithm. Specifically, the iterative analysis of the ESMDA allows the friction velocity u_* to be inferred. This is not possible for the EnKF due to insufficient observability. This is because the effect of u_* on the observations is coupled with the effect of the inlet wind direction, while having a lesser impact on observations. Furthermore, using the POD–GPR surrogate model substantially reduces the computational cost of the DA process and makes it possible to carry out a wide range of

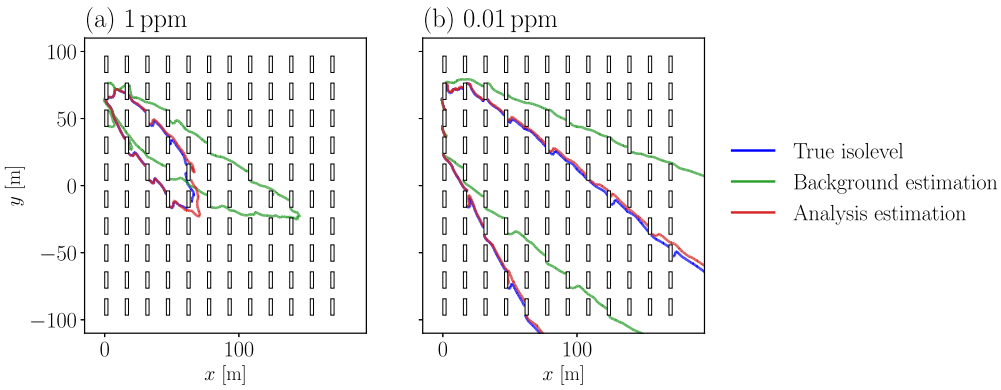

Fig. 6. Contour levels of tracer concentration for 1 ppm (a) and 0.01 ppm (b) at $z = 1.6$ m obtained in the baseline twin experiment (Table 1). Comparison between background estimates (green lines), the analysis estimates obtained by the ESMDA with $N_a = 4$ (red lines) and ground truth (blue lines).

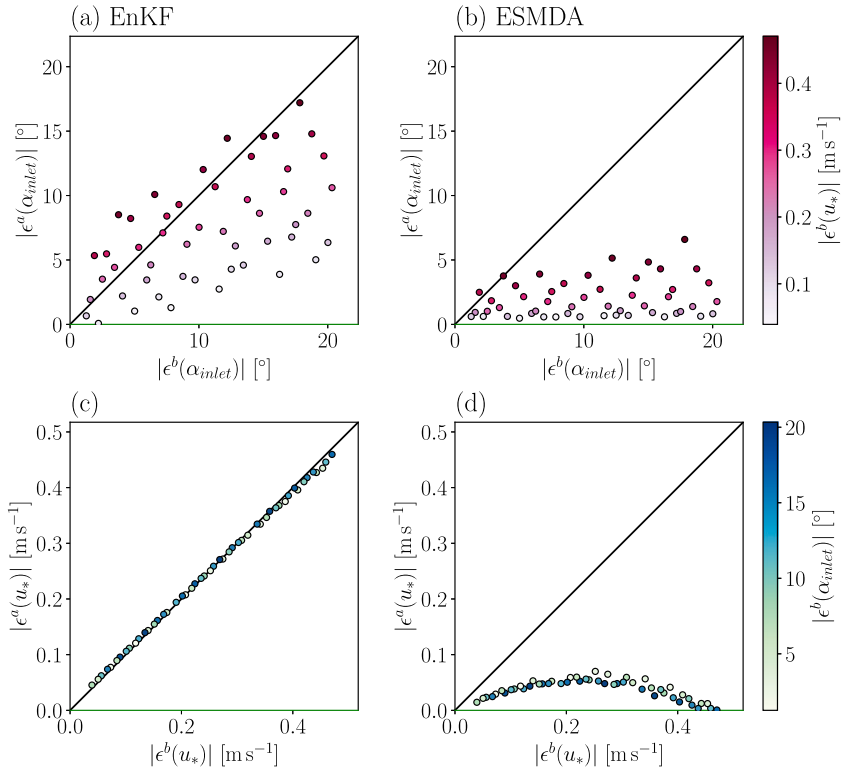

Fig. 7. Absolute error of the EnKF (a, c) and ESMDA (b, d) parameter estimates $|\epsilon^a| = |\theta^a - \theta^t|$ for varying background bias $|\epsilon^b| = |\theta^b - \theta^t|$. The bias of the inlet wind direction α_{inlet} estimates is shown in (a, b) and is colored according to the bias of the friction velocity u_*. The bias of the friction velocity u^* estimates, colored by the bias of the inlet wind direction, is shown in (c, d).

twin experiments to: *(i)* establish guidelines for the ESMDA parameters before assimilating real measurements ($N_e = 500$ ensemble members and $N_a = 4$ outer loops in the following) ; and *(ii)* demonstrate the robustness of the DA system to background sampling errors.

5. Assimilation of the real MUST field observations

This section moves to the assimilation of the real tracer concentration measurements from the MUST trial #2681829. Assimilating real observations introduces uncontrolled model and observation biases, making the assimilation problem more challenging. The same background parameters and observation network as in the baseline twin experiment are retained (Table 1). The ESMDA algorithm settings ($N_a = 4$, $N_e = 500$) follow the twin experiments guidelines from Section 4.

The evaluation is carried out in two stages. First, the ability of the ESMDA to correct the meteorological forcing parameters is assessed (Section 5.1). Then, it is determined whether correcting these parameters improves the prediction of the tracer concentration field (Section 5.2).

5.1. Estimation of the meteorological forcing parameters

The accuracy of the ESMDA algorithm is evaluated by comparing its estimation of the meteorological forcing parameters θ^a to the reference parameters for this MUST trial, $\theta^{(ref)} = (-41\,°, 0.73\,\mathrm{m\,s^{-1}})$, which is independent of the DA experiment (Section 2.3).

Fig. 8 shows that the ESMDA successfully corrects both the inlet wind direction and the friction velocity, with the analysis ensemble averages being closer to the reference than those of the background. However, there is a slight overcorrection of the inlet wind direction

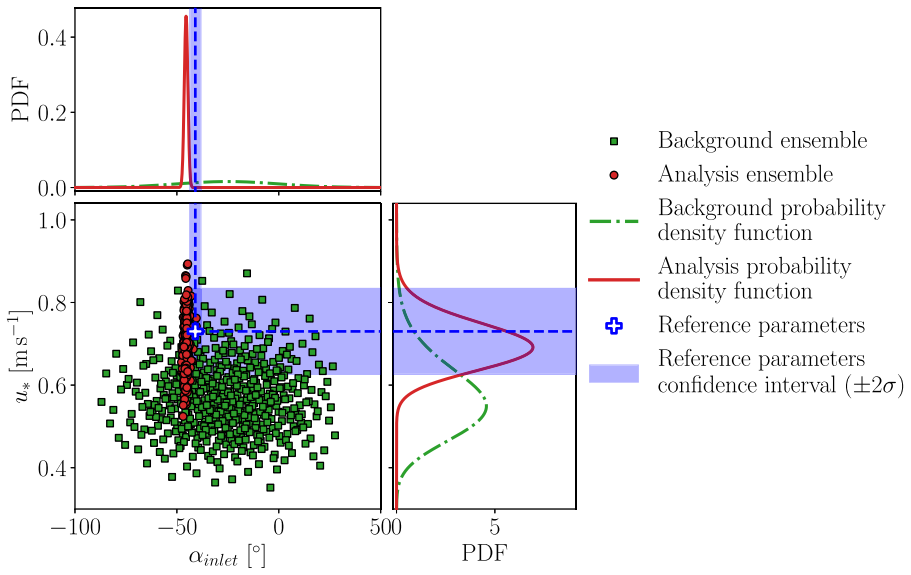

Fig. 8. Background and analysis ESMDA ensembles obtained when assimilating the real MUST data. The associated normal and log-normal PDF of the inlet wind direction and friction velocity are represented in the upper and right panels, respectively. The blue plus symbol indicates the reference parameters, $(\alpha_{inlet}^{(ref)}, u_*^{(ref)})^{\mathrm{T}} = (-41\,°, 0.73\,\mathrm{m\,s^{-1}})^{\mathrm{T}}$, derived from unassimilated wind velocity measurements. The blue shaded areas represent the associated measurement uncertainty.

compared to the reference ($\alpha_{inlet}^{a} - \alpha_{inlet}^{(ref)} = -4.5°$). This is due to bias in either the prediction of tracer concentration or its measurement, as evidenced in Section 5.2.

Assimilating tracer concentration measurements also reduces the uncertainty in the meteorological forcing compared to the background, particularly with regard to the wind direction (Fig. 8). Furthermore, the ESMDA uncertainty estimates are realistic as they are of the same order of magnitude as the uncertainty of the reference meteorological measurements $\boldsymbol{\theta}^{(ref)}$. This good performance is partly due to the realistic observation error model $\mathbf{R}$ adopted (see Appendix B).

It should be noted that the reference wind direction is outside of the confidence interval provided by the analysis ensemble $|\alpha_{inlet}^{a} - \alpha_{inlet}^{(ref)}| > 3\sigma(\alpha_{inlet}^{a})$ (this is not the case for the friction velocity). This lack of reliability, which was not observed in the twin experiments, suggests that other forms of uncertainty (e.g. model bias) might involved and not captured by the ESMDA.

5.2. Evaluation of the updated tracer concentration prediction

This section assesses whether the ESMDA parameter estimation improves the prediction of the quantity of interest, i.e. the time-averaged tracer concentration field.

5.2.1. Dispersion prediction accuracy metrics

To quantify the gain in accuracy achieved with DA, the set of N_{obs} observed time-averaged concentrations $\mathbf{c}^o$ is compared with the collocated DA estimates $\mathbf{c}^p$, defined as the model prediction associated with either the background parameters $\mathbf{c}^p = \mathcal{M}_{\mathrm{POD-GPR}}\left(\overline{\boldsymbol{\theta}^b}\right)$ or the ESMDA ensemble-averaged analysis $\mathbf{c}^p = \mathcal{M}_{\mathrm{POD-GPR}}\left(\overline{\boldsymbol{\theta}^a}\right)$. This is done using the standard statistical metrics for air quality model evaluation [73], including the normalized mean square error (NMSE), the fraction of predictions within a factor of two of the observations (FAC2), and the geometric variance (VG):

$$\mathrm{NMSE} = \frac{\langle (\mathbf{c}^o - \mathbf{c}^p)^2 \rangle}{\langle \mathbf{c}^o \rangle \langle \mathbf{c}^p \rangle}, \tag{19}$$

Table 3
Air quality metrics quantifying the global agreement between the background and analysis concentration predictions and experimental measurements. The scores obtained by the reference LES model prediction $\mathcal{M}_{\mathrm{LES}}(\boldsymbol{\theta}^{(ref)})$ are also given for comparison. The metrics are calculated separately over the set of the assimilated observations and over the validation set (made of unassimilated observations).

		NMSE	FAC2	VG
Perfect score		0	1	1
"Acceptable" [73]		< 4	> 0.5	< 1.6
Assimilated obs.	Background	4.12	$^3/_9 \approx 33\%$	7.47
	Analysis	0.04	$^9/_9 = 100\%$	1.02
	Reference LES	0.62	$^5/_9 \approx 56\%$	2.16
Validation obs.	Background	1.37	$^{11}/_{38} \approx 29\%$	23.6
	Analysis	0.43	$^{18}/_{38} \approx 47\%$	2.73
	Reference LES	0.31	$^{22}/_{38} \approx 58\%$	2.12

$$\mathrm{FAC2} = \frac{1}{N_{obs}} \sum_{k=1}^{N_{obs}} \mathbb{1}_{\mathrm{FAC2}}(k) \text{ with } \mathbb{1}_{\mathrm{FAC2}}(k) = \begin{cases} 1 \text{ if } 0.5 \leq \mathbf{c}_k^p / \mathbf{c}_k^o \leq 2, \\ 0 \text{ else,} \end{cases} \tag{20}$$

$$\mathrm{VG} = \exp\left(\left\langle \left(\ln \widetilde{c}_o - \ln \widetilde{c}_p \right)^2 \right\rangle \right), \tag{21}$$

where $\langle \cdot \rangle$ denotes the average over the N_{obs} samples. The measurements below the instrument detection threshold c_t (0.04 ppm for DPIDs and 0.01 ppm for UVICs) are excluded from the metrics calculations. For each metric, the perfect and acceptable scores according to Chang and Hanna [73] are indicated in Table 3.

For verification purposes, the metrics are first calculated over the set of the assimilated observations (DPID sensors from towers B, C, and D at $z = 1, 2, 3$ m). Table 3 shows that the updated concentration estimate is very close to the assimilated observations, as evidenced by the near perfect NMSE, FAC2, and VG scores.

Then, the metrics are calculated over all the unassimilated observations referred to as validation observations. Table 3 shows that the updated concentration estimate is significantly improved compared to the background. The ESMDA highly reduces the NMSE and VG

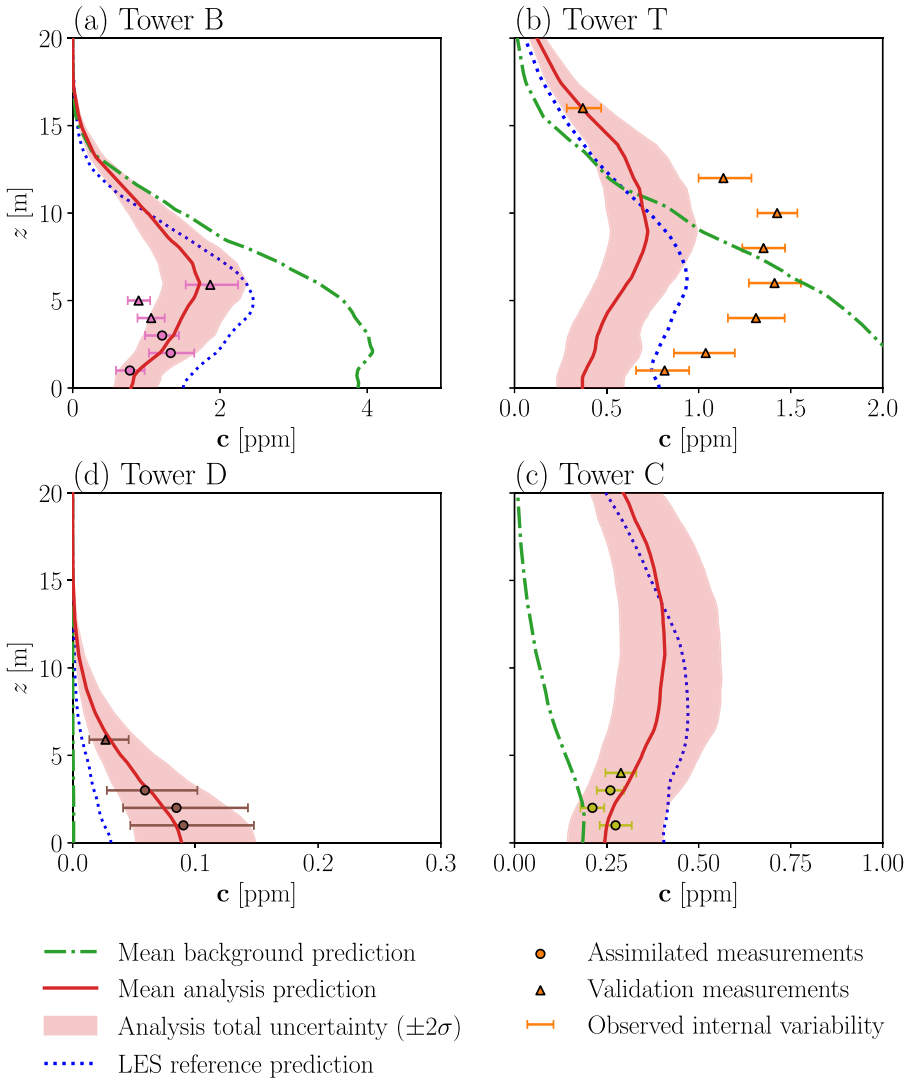

Fig. 9. Vertical profiles of time-averaged concentration predicted by the POD–GPR model using the background (green dash-dotted) and analysis (red solid) ensemble-mean parameters. Predictions from the reference parameters $\theta^{(ref)}$ are shown as blue dotted lines. Results are shown for towers B, T, C, and D (Fig. 1). Crosses and triangles indicate assimilated and unassimilated measurements, respectively. Error bars reflect ABL internal variability. Red shaded areas represent the 95% confidence intervals from the analysis ensemble and POD–GPR uncertainty.

quadratic errors, reaching the acceptable levels of accuracy defined by Chang and Hanna [73]. However, this is not the case for the FAC2 metric, for which only 47% of the updated predictions fall within a factor of two away of the observations (the acceptable limit is 50%). The analysis cannot reach the 58% FAC2 score obtained by the reference LES prediction $\mathcal{M}_{\text{LES}}(\theta^{(ref)})$, suggesting that the DA system may be overfitting the assimilated observations. It is therefore necessary to further investigate the sensitivity of the analysis estimate to the observation locations to find possible explanations for this DA behavior.

5.2.2. Tracer concentration horizontal and vertical profiles

To better identify where the DA update lacks accuracy, the horizontal and vertical tracer concentration profiles associated with the background and analysis ensembles are examined.

Fig. 9 shows that the DA system successfully recalibrates the POD–GPR prediction with the observations at towers B, C and D compared to the background prediction and to the reference prediction obtained with the reference meteorological forcing parameters $\theta^{(ref)}$. This is consistent with the very convincing metrics scores obtained over the assimilated observations (Table 3). Note that the spatial consistency of the model predictions improves the agreement with the unassimilated observations located higher up on towers B, C and D. At tower T, where no observation is assimilated, the DA results are less good with an overall underestimation of the concentration. Yet, the shape of the vertical profile predicted after the analysis is more consistent than that obtained with the background. Note that the reference LES prediction $\mathcal{M}_{\text{LES}}(\theta^{(ref)})$ also underestimates tracer concentration at tower T.

Using the $N_e = 500$ ensemble members of the ESMDA, the uncertainty in the meteorological forcing parameters is propagated through the surrogate model to quantify the uncertainty on the tracer concentration prediction. In addition, for each ensemble member, the POD–GPR model provides a probabilistic distribution of the concentration field, which accounts for the atmospheric internal variability [40]. Fig. 9 shows that the resulting total uncertainty in the tracer concentration prediction explains well the differences between the model prediction and the observations from towers B, C, and D, but not from tower T.

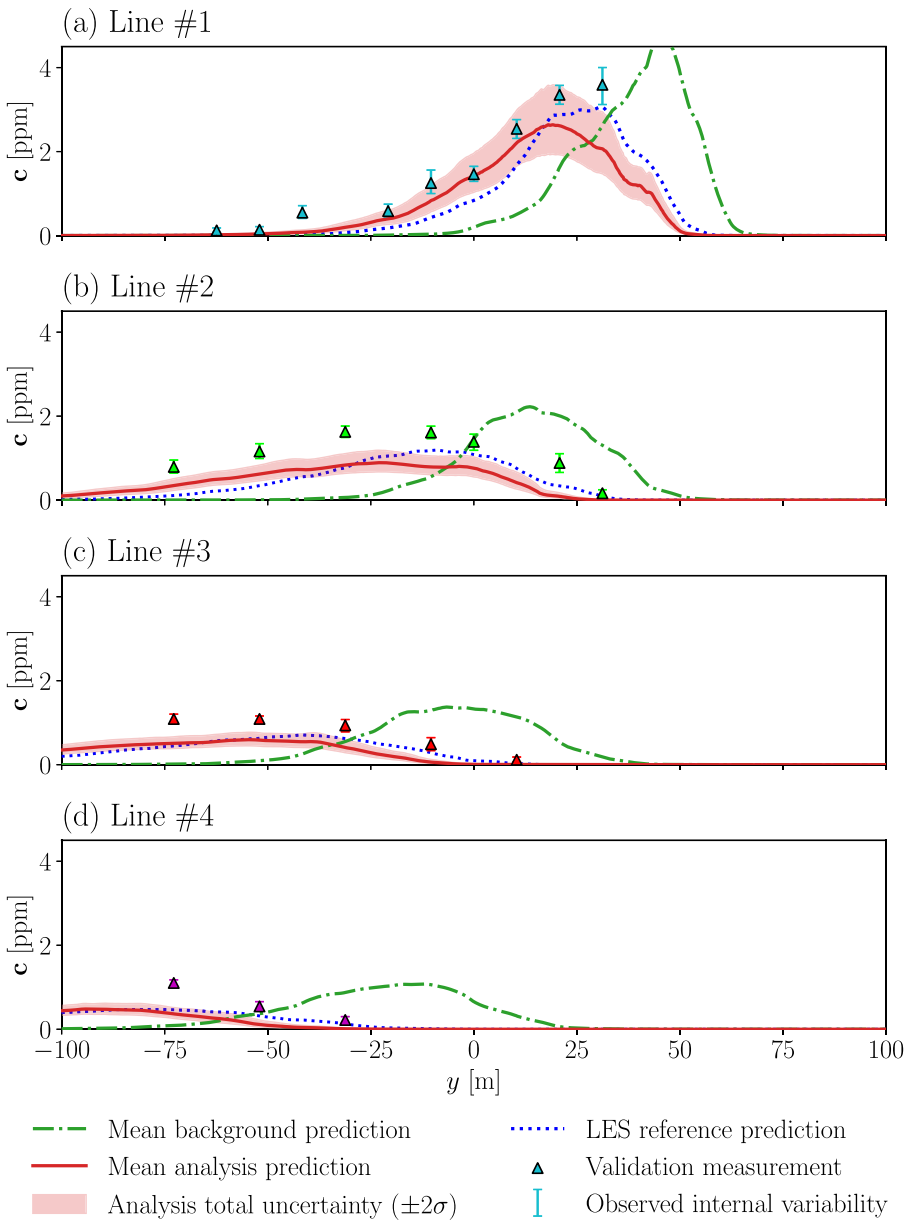

Fig. 10. Horizontal profiles of time-averaged concentration predicted by the POD–GPR model using the background (green dash-dotted) and analysis (red solid) ensemble-mean parameters. Predictions from the reference parameters $\theta^{(ref)}$ are shown as blue dotted lines. Each row corresponds to a DPID sensor line (Fig. 1). Triangles indicate unassimilated measurements with error bars representing the ABL variability. Red shaded areas represent the 95% confidence intervals from the analysis ensemble and POD–GPR uncertainty.

This suggests that there are other forms of uncertainty involved that limit model prediction accuracy. Note that there may be an inconsistency between the measurements from the DPID sensors at tower T and the UVIC sensors at towers B, C, and D.

Regarding the horizontal profiles of the tracer concentration, Fig. 10 confirms that the DA process improves the prediction compared to the background, while no observations are assimilated on the $z = 1.6$-m plane. This is mainly due to the wind direction correction, which corrects the originally misaligned plume centerline direction. However, despite the good plume shape prediction, the analysis tends to underestimate concentration. As before, these differences from the DPID measurements are not explained by the meteorological forcing uncertainty and internal variability. The same underestimation is obtained by the reference LES prediction. This shows that the lack of DA accuracy is due to an internal model bias, rather than an incorrect parameter correction. In particular, the fact that the model underestimation increases with the distance from the emission source may indicate that the LES model predicts a plume spreading too much at this height, as already observed in Lumet et al. [23].

In conclusion, assimilating the real MUST observations improves tracer concentration prediction by inferring the meteorological forcing parameters. This results in acceptable air quality metrics (except for the FAC2 that is slightly below the acceptable limit of 50%) and a predicted plume shape that is consistent with unassimilated observations. However, since the DA system does not directly correct the model state but only input parameters, the accuracy of the analysis prediction remains limited by internal model biases. Although interesting and with great potential to overcome this issue, it is beyond the scope of this work to develop an estimation approach able to jointly estimate the model

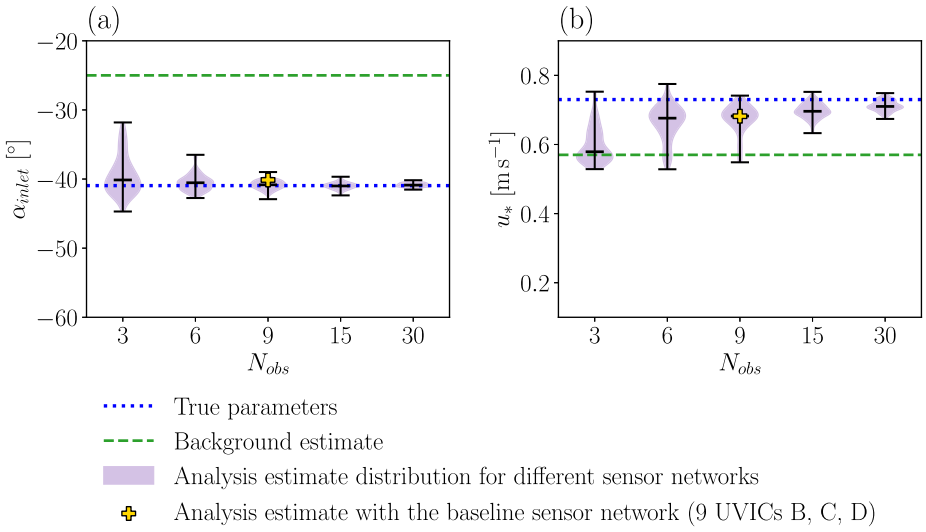

Fig. 11. Distributions of ESMDA ensemble-averaged estimates of wind direction α_{inlet} (a) and friction velocity u_* (b) for varying numbers of assimilated observations N_{obs}. Each violin plot represents 100 DA experiments with different synthetic observation combinations. True values (blue dotted lines), background estimates (green dashed lines), and baseline sensor network results (yellow crosses) are also shown.

state and the model parameters [74]. To gain a better understanding of the DA system behavior, this study opted to conduct a more in-depth analysis of observability. This involves examining the extent to which the quality of the DA results depends on the choice of assimilated observations, as well as investigating whether better DA results could be achieved with certain observation network configurations. This aims to provide insight into LES model structural errors and areas for model improvement.

6. Sensitivity to the observation network

While previous DA experiments were carried out using the same baseline observation network (Section 3.4.1), the following analysis investigates how the design of the observation network, i.e. the number and location of the assimilated sensors, affects DA accuracy in relation to system observability. The theoretical sensitivity to the observation network is first assessed using twin experiments (Section 6.1), before turning to real data (Section 6.2).

6.1. Theoretical impact of observations using twin experiments

To assess the DA sensitivity to the observation network, the baseline twin experiment (Table 1) is reproduced, while varying the observation network design, in terms of number of sensors (N_{obs}) and in terms of location of these sensors (among the 47 sensors that capture a significant signal during the MUST campaign). Five different sizes of sensor network are tested with $N_{obs} \in \{3, 6, 9, 15, 30\}$. For each network size, 100 sensor arrangements are randomly generated and tested, resulting in a total of 5×100 DA experiments. This was found sufficient to obtain statistically converged results (not shown). As for each ESMDA experiment $N_e = 500$ ensemble members and $N_a = 4$ outer loops are used, this observability study requires a total of one million surrogate model predictions ($N_a \times N_e = 2\,000$ model predictions for each of the 500 DA experiments), made possible through the use of a surrogate-based DA system.

Fig. 11 shows the analysis distribution of the parameters $\theta = (\alpha_{inlet}, u_*)^{\mathrm{T}}$ estimated by the ESMDA over the different sensor networks. The results obtained using the baseline sensor network are indicated by yellow crosses ($N_{obs} = 9$). When assimilating a small number of sensors ($N_{obs} \leq 9$), the analysis estimates highly depend on the location of these sensors, with a standard deviation reaching 5.4 ° for the wind direction estimates when $N_{obs} = 3$. Conversely, the variability on the mean analysis estimates decreases as the number of assimilated observations increases. This demonstrates that the information provided by the observations is sufficiently detailed when $N_{obs} > 9$ to overcome the local effects of wind/container interactions on the tracer concentration and retrieve physical parameters of the LES model that have a global meaning. Indeed, when there are too few observations, these parameters are adjusted to match the local effects measured by these observations. These effects can differ greatly within the canopy and explain the variability in the analysis estimates observed in Fig. 11 for $N_{obs} \leq 9$.

Fig. 11 also shows that on average, the ESMDA accurately infers the wind direction regardless of the number of sensors N_{obs}, but is unable to correct the friction velocity bias when $N_{obs} < 6$. These shortcomings can be overcome by increasing the number of assimilated sensors, as this both increases the robustness of the ESMDA analysis to sensor location and improves the friction velocity mean estimates.

6.2. Analysis of the impact of the real MUST field observations

In this section, the observation network sensitivity analysis from Section 6.1 is reproduced with the actual MUST experimental data. Unlike in the twin experiments, the DA process is now subject to observational and model biases, which provides a more realistic assessment of the effect of the observation network design on the DA system and provide further insights into its observability property.

For each size of sensor network $N_{obs} \in \{3, 6, 9, 15\}$, the ESMDA with $N_a = 4$ outer loops and $N_e = 500$ ensemble members is used to assimilate the real MUST measurements from 300 different sensor networks. Among these networks, 100 are obtained by randomly selecting N_{obs} observation positions from the 47 possible ones as in Section 6.1 (Fig. 1); 100 networks correspond to horizontal sensors only, i.e. DPID sensors on the plane $z = 1.6$ m; and 100 networks correspond to vertical sensors only, i.e. from towers B, C, D, and T.

As shown by the red distributions (combining horizontal and vertical observations) in Fig. 12, the DA system correctly estimates the wind direction, regardless of the size of the observation network. However, it

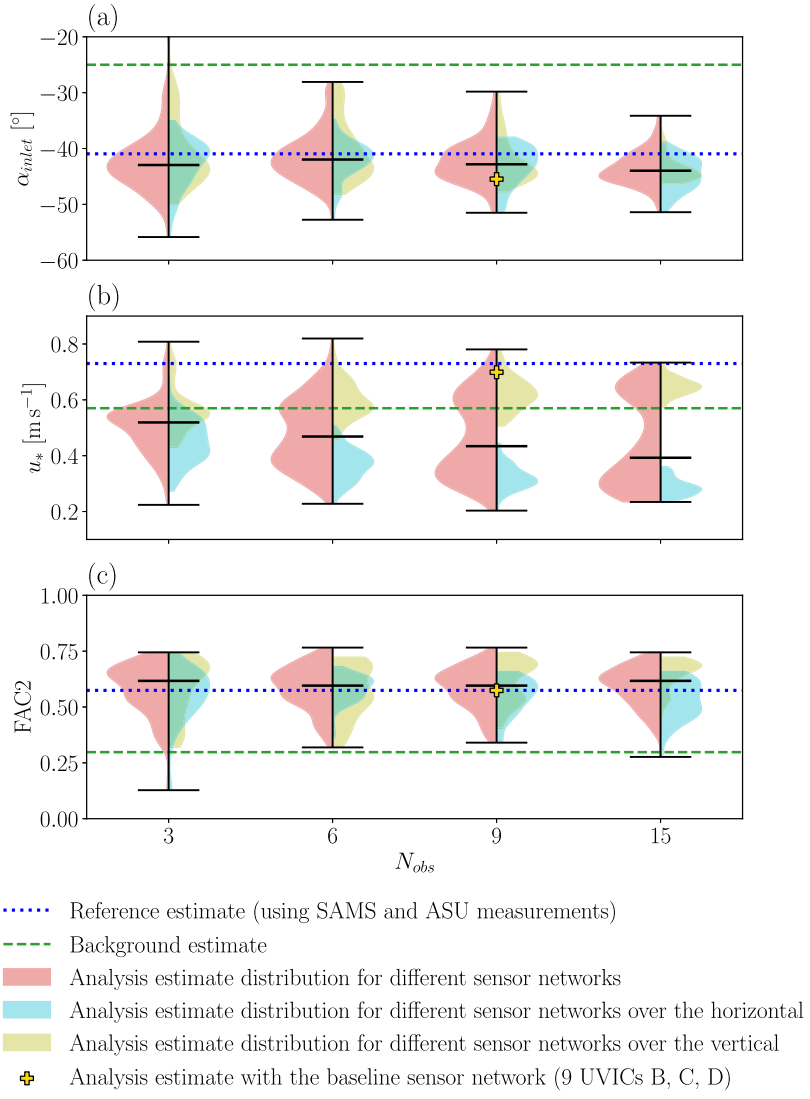

Fig. 12. Comparison of ESMDA ensemble-averaged estimates of wind direction α_{inlet} (a), friction velocity u_* (b), and FAC2 metric (c) for varying numbers of assimilated observations N_{obs}. Shaded violin plots show Is it fair to say that the uniform heights of the containers and regular arrangement of the containers yield a log-normal distribution for concentration observations? If yes, how does this translate to a heterogeneous distribution of buildings? Can this ESMDA approach be applied to such a complex array of buildings, which is usually the case in dense metropolitan cities? Please comment on this to demonstrate the practicality of this approach. distributions from 300 DA experiments: all sensor networks (red), horizontal-only sensors, i.e. DPIDs at $z = 1.6$ m (blue), and vertical mast-only sensors (yellow). Reference values (blue dotted lines), background estimates (green dashed lines), and baseline sensor network results (yellow crosses) are also shown.

fails to correct the friction velocity on average. The analysis even tends to worsen the friction velocity estimates compared to the background. This issue gets worse when the number of assimilated observations N_{obs} increases.

Fig. 12 also shows that the distribution of the friction velocity estimates becomes bimodal when $N_{obs} > 3$. This issue was not present in twin experiments (Fig. 11b). This can be explained by the occurrence of two distinct cases. On the one hand, when the observation network mainly consists of DPIDs along the plane $z = 1.6$ m (horizontal sensors), the ESMDA considerably underestimates the friction velocity. On the other hand, when the observation network mainly consists of sensors from the vertical towers B, C, D and T (vertical sensors), the ESMDA significantly improves friction velocity estimation. This behavior is evidenced by the distinct distributions over the horizontal and vertical subsets in Fig. 12b (blue and yellow distributions). This can be explained by the tendency of the model to underestimate tracer concentration at $z = 1.6$-m altitude (Fig. 10). When these DPID sensor measurements are assimilated, the ESMDA artificially reduces the friction velocity in order to compensate for model underestimation and to better match the observed concentration since the predicted concentration is inversely proportional to friction velocity. In other words, friction velocity loses its physical meaning and becomes an adjustment variable of the system due to other biases in the model that are not accounted for. These two distinct cases also explain why the sensitivity to the observational network remains high when assimilating $N_{obs} = 15$ observations: the two observational subsets provide conflicting information on the meteorological parameters and induce an equifinality issue, which makes the inverse problem partly ill-posed.

Even if the friction velocity is not properly estimated, the DA system still leads to good FAC2 scores regardless of the observation network (Fig. 12c), the FAC2 score being above 50% in most cases and

thereby satisfying the acceptability criterion of Chang and Hanna [73]. The same is true when only assimilating DPID measurements at $z = 1.6$ m, despite the artificial friction velocity correction. This is because the friction velocity has less influence on the predicted concentration field, and thus on FAC2, than the wind direction, which is accurately estimated by the ESMDA. Good scores can thus be obtained for the wrong reasons; in this case, the overestimation of wind friction velocity compensates for the fact that the LES model overestimates the plume dispersion. This highlights the limitations of using only global metrics to assess model performance.

In conclusion, the assimilation of actual MUST data highlights the limitations of idealized twin experiments in optimizing sensor placement (i.e. identifying the most informative observations for the uncertain parameters to be inferred). It is shown that increasing the number of assimilated observations does not necessarily reduce the spread of the DA estimates because of conflicting measurements and compensating effects. This implies that conclusions from twin experiments may not be applicable to field conditions when model and observation biases are present. Nevertheless, the sensitivity analysis carried out for real field measurements has demonstrated the robustness of the ESMDA system. This system can accurately estimate wind direction and improve concentration field prediction regardless of the location of the assimilated sensors.

7. Discussion

This study provides a proof-of-concept of a surrogate-based ensemble DA system that efficiently reduces uncertainties in meteorological forcing parameters and improves urban pollutant dispersion predictions. This system is validated by assimilating tracer concentration measurements from the MUST simplified urban geometry in near-neutral atmospheric conditions. As the quantities of interest are stationary, a single DA cycle is performed. Nevertheless, the same system could be applied sequentially with sliding averages to capture large-scale temporal variations in meteorological forcing and their impact on pollutant dispersion.

Application to more realistic, heterogeneous, urban geometries and broader parameter ranges would require construction of a new LES training database adapted to the specific case study. The number of degrees of freedom of the DA system, limited here to two parameters, could also be expanded to refine the representation of the meteorological forcing [28], to account for variability in atmospheric stability, or to correct pollutant source parameters [30]. Increasing the number of degrees of freedom, however, may challenge the ESMDA assumption of normally-distributed errors. This limitation could be overcome thanks to the low computational cost of the POD–GPR surrogate model, which makes more advanced ensemble-based methods such as particle filters [75] computationally feasible.

This study provides insights into the origins of uncertainty in LES pollutant dispersion predictions. Correcting meteorological forcing parameters significantly improves plume representation in the vertical direction and near the source. However, persistent underestimation of concentration close to the ground and downstream cannot be explained by forcing conditions or internal variability alone. This finding challenges the common assumption that LES structural model uncertainties are negligible compared to forcing uncertainties [4,6,10], and underscores the need for systematic sensitivity analysis and multi-model comparison.

Within the standalone parameter estimation framework, the DA system tends to overcorrect the meteorological forcing parameters in order to compensate for these structural model biases. This behavior reduces the reliability of the corrected fields and highlights the importance of explicitly representing model uncertainties. One possible solution is to sample forecasts from multiple LES models within the ESMDA framework [76,77], thereby incorporating model diversity into the assimilation process. Another promising solution is to carry out joint state–parameter estimation [74,78–80]. This would allow errors in the plume model predictions to be corrected directly and provide more realistic parameter estimates, as the parameters would no longer be adjustment variables.

Finally, this study confirms that this type of DA system for urban flow prediction is highly sensitive to the design of the observation network [25,30]. Moreover, while classical twin experiments suggest that increasing the number of sensors improves estimation robustness, results based on real MUST measurements show that biases in the model and/or observations can lead to contradictory analysis trends depending on sensor placement. This calls into question the robustness of optimal sensor placement strategies derived solely from idealized numerical experiments. This highlights the need to explicitly consider possible model biases in future methodologies.

8. Conclusion

This study demonstrates the feasibility and benefits of a surrogate-based ensemble DA system for LES of urban pollutant dispersion, through the MUST field campaign. The system infers meteorological forcing parameters (wind direction and friction velocity) through assimilation of in-situ pollutant concentration measurements using the ESMDA algorithm [36]. The computational burden of LES in a multi-query uncertainty quantification context is mitigated by the POD–GPR surrogate model [40].

The substantial speed-up provided by the surrogate model enables large ensemble size, thereby improving DA estimation accuracy, while making the DA system tractable for real-time applications. It is also a strength for carrying out a detailed validation of the DA system. The key findings are:

- The iterative nature of ESMDA systematically improves the estimation of meteorological forcing parameters compared to the standard stochastic EnKF, ensuring accurate parameter correction regardless of background bias. In particular, the ESMDA better tackles nonlinear effects, such as the coupled interaction of friction velocity and wind direction on the plume. This highlights the suitability of the easy-to-implement ESMDA algorithm for nonlinear parameter estimation problems in CFD urban flow modeling.
- LES plume predictions are improved by the DA system near the source and in the vertical, but persistent underestimation near the ground and downstream remains. These errors cannot be explained solely by meteorological forcing biases or internal variability. This indicates significant structural uncertainties in LES that should be investigated and accounted for in future developments.
- DA performance is highly sensitive to sensor placement, and this sensitivity is underestimated when assessed with synthetic measurements. The combination of model and real measurement biases leads to contradictory corrections depending on sensor locations. This shows that sensor optimization strategies based solely on idealized twin experiments may be overly optimistic.

This proof-of-concept study confirms that surrogate-based DA can reduce uncertainty in urban LES predictions, while also revealing critical challenges related to observation network design and LES structural modeling that must be addressed in future research.

CRediT authorship contribution statement

Eliott Lumet: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Mélanie C. Rochoux:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Thomas Jaravel:** Validation, Supervision, Methodology, Formal analysis, Conceptualization.

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.

Acknowledgments

Eliott Lumet's PhD thesis was funded by the Université Fédérale Toulouse Midi-Pyrénées together with the Région Occitanie (ADI20, AtmoDrones project, 2020–2023). This work was granted access to the HPC resources of GENCI-TGCC/CINES (A0062A10822 project, 2020–2022).

Appendix A. Reducing background sampling error using the halton low-discrepancy sequence

A limitation of the EnKF and ESMDA algorithms is the use of ensembles of limited size, which introduces sampling errors [77,81]. While sampling errors due to observation perturbation (Eq. (10)) can be avoided by using more sophisticated DA implementations such as the ensemble transform Kalman filter [82,83], sampling errors due to background ensemble sampling (Eq. (8)) remain.

To assess the impact of background sampling errors in the present case study, the variability of the DA estimates is evaluated when the background sampling seed is changed for a given ensemble size. This is done by replicating the baseline twin experiment (Table 1) with different seeds. It was found that between 500 seeds for $N_e = 10$ and 100 seeds for $N_e = 1,000$ was enough to reach statistical convergence. It is worth noting that sampling errors due to observation perturbations (Eq. (10)) and model error (Eq. (9)) were found to be significantly less important. Attention is therefore directed to background sampling in the following.

Fig. A.1 shows that, as expected, the spread of the DA analysis errors decreases with the ensemble size N_e for both EnKF and ESMDA. For a given ensemble size, this spread is smaller for the ESMDA than for the EnKF. Nevertheless, even at a rather large ensemble size of $N_e = 500$ for only 2 control parameters, the effect of background sampling errors remains significant as it induces deviation of the EnKF α_{inlet}-estimate of up to 1.8°.

To reduce the sensitivity of the analysis on the background sampling, it is possible to use a quasi-Monte Carlo approach or a quasi-random number sequence to generate the background ensemble [84]. In this work, as the control parameter space is low-dimensional ($p = 2$), the Halton sequence ([64]) is used. As a low-discrepancy sequence, this sequence covers more efficiently the uncertain space than a purely random sampling by avoiding redundant samples in the same areas.

In practice, the background ensemble is generated as follows: first, the Halton sequence is used to sample the unit disk $[0, 1]^p$ with $p = 2$ the dimension of the control space; then the background ensemble $\{\boldsymbol{\theta}_i^b\}_{i=1}^N$ is generated by mapping the Halton samples with the quantile function of the Gaussian distribution assumed in the EnKF–ESMDA framework ($\boldsymbol{\theta}_i^b \sim \mathcal{N}(\boldsymbol{\theta}^b, \mathbf{B}), 1 \leq i \leq N$, $\mathbf{B}$ being the background error covariance matrix). In the present case, this can be done independently for each component since the $\mathbf{B}$-matrix is diagonal.

Fig. A.1 shows that the parameters estimated when generating the background ensemble with the Halton sequence are very close to the Monte Carlo estimates averaged over varying background sampling seeds, both with the EnKF and the ESMDA algorithms. Using the Halton sequence provides a convincing way to suppress the variability of the analysis induced by background sampling. It should be recommended as good practice when employing ensemble DA methods, to prevent the analysis estimates from depending on the chosen background sampling seed

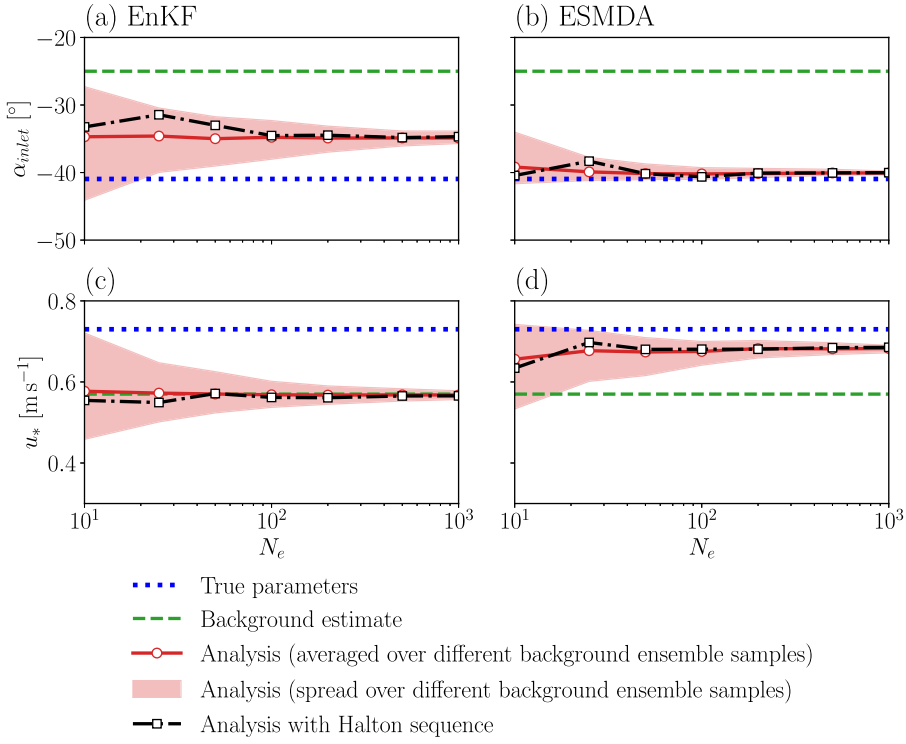

Fig. A.1. Wind direction α_{inlet} and friction velocity u_* estimated by the EnKF (a, c) and by the ESMDA (b, d) in the baseline twin experiment (Table 1) and for varying ensemble size Ne. The background estimates and the true values are shown as green dotted lines and blue dotted lines, respectively. The red solid lines correspond to the averaged analysis errors, estimated by a Monte Carlo procedure with varying random seeds to sample the background ensemble. The red shaded area represents the associated 95% confidence interval of the analysis error. The black dashed–dotted lines correspond to the analysis estimate obtained when sampling the background ensemble with the Halton sequence.

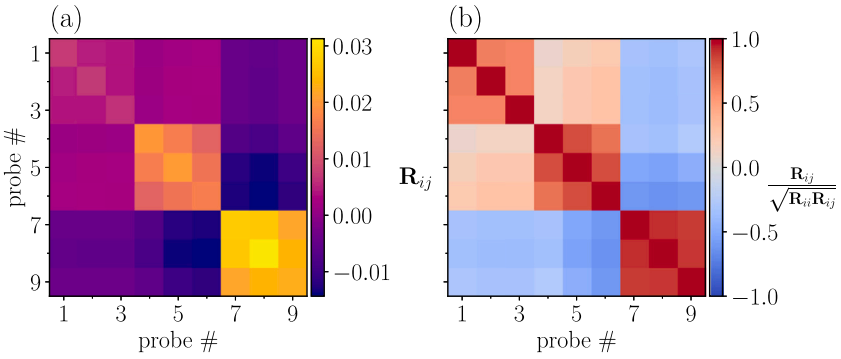

Fig. B.1. Observation error covariance matrix accounting for the sensor measurement error and the internal variability error estimated using bootstrap replicates of the log-transformed concentration measurements without (a) and with normalization (b). Rows and column indices correspond to the sensors from which measurements are assimilated for the baseline observation network, arranged in the following order: tower C (1, 2, 3 m), B (1, 2, 3 m), and D (1, 2, 3 m).

Appendix B. Observation error covariance modeling

Observation errors come from three independent sources: *(i)* the instrumental errors, *(ii)* the representativeness errors, which correspond to errors due to the discretization of the system state and the related interpolations in the observation operator, and *(iii)* the aleatory uncertainty due to the internal variability of the ABL, which has been shown to be significant in the MUST campaign [23,43].

Given the fine grid resolution used to represent the system state, the representativeness errors are negligible. Instrumental errors were not reported in the MUST technical report [42]; it is therefore assumed that the UVIC and DPID sensors used during the MUST campaign have the same uncertainty as the miniPID 200B model from the same manufacturer, i.e. $\varepsilon_\mu = 5\%$ relative uncertainty [85].

To account for ABL internal variability, which has not been considered in previous DA studies, the stationary bootstrap procedure based on resampling of sub-averages proposed in Lumet et al. [23] is employed. This method involves generating B replicates $\left\{\mu_b(\mathbf{y}^o)\right\}_{b=1}^{B}$ of the observations $\mathbf{y}^o$ based on the original experimental time series. Using these replicates, which are representative of the changes in the time-averaged value of the 200-s analysis period induced by the ABL internal variability, the observation internal variability error covariance matrix $\mathbf{R}^{\mathrm{IV}}$ can be estimated as

$$\mathbf{R}^{\mathrm{IV}} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\mu_b(\mathbf{y}^o) - \widehat{\mu(\mathbf{y}^o)}\right)\left(\mu_b(\mathbf{y}^o) - \widehat{\mu(\mathbf{y}^o)}\right)^T, \tag{22}$$

with $\widehat{\mu(\mathbf{y}^o)} = \frac{1}{B}\sum_{b=1}^{B}\mu_k(\mathbf{y}^o)$ the ensemble mean of the replicates. To limit sampling errors, a very large number of bootstrap replicates is used ($B = 50\,000$). The stationary bootstrap block length is set equal to the average correlation time over each concentration measurement (18.5 s here). One benefit of this approach is that it provides estimates of the cross-covariance terms of the observation error covariance matrix $\mathbf{R}$. These cross-covariance terms characterize the spatial correlations of the errors between sensors. These correlations are often assumed to be zero due to a lack of information [25,28,33].

In summary, the observation error covariance matrix $\mathbf{R}$, which accounts for both instrumental and internal variability errors, is defined as

$$\mathbf{R} = \mathbf{R}^{\mu} + \mathbf{R}^{\mathrm{IV}} = \sigma_\mu^2(\mathbf{y}^o)\,\mathbf{I} + \mathbf{R}^{\mathrm{IV}}, \tag{23}$$

where $\sigma_\mu^2(\mathbf{y}^o) = \ln\left(1+(\varepsilon_\mu)^2\right)$ is the variance of the log-transformed concentration measurement errors, assuming that the concentration is log-normally distributed. Fig. B.1 shows the resulting $\mathbf{R}$-matrix for the baseline observation network (Section 3.4.1). Since the instrumental error is assumed to be independent for each sensor, the $\mathbf{R}$-matrix has large values on its diagonal, while the extra-diagonal coefficients correspond to error correlations due to the ABL internal variability. The distribution of these covariances is physically-consistent since they decrease with increasing distance between the sensors. Three blocks of high covariances emerge, corresponding to the three groups of sensors on towers B, C, and D. Note that the errors of the sensors on the towers B and C are significantly correlated since these towers are aligned with the flow direction, while this is not the case for the sensors on tower D as they are isolated at the edge of the plume (Fig. 1).

Data availability

The ESMDA code developed for this study is available at https://github.com/eliott-lumet/esmda_ppmles. The dataset used to train the POD–GPR surrogate model is available in a public repository [86]. The other data and codes used in the study are available from the corresponding author upon reasonable request.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division, World Urbanization Prospects: The 2018 Revision, Technical report, in: ST/ESA/SER.A/420, United Nations, New York, 2019, URL https://population.un.org/wup/assets/WUP2018-Report.pdf.

[2] K.C. Seto, B. Güneralp, L.R. Hutyra, Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools, Proc. Natl. Acad. Sci. 109 (40) (2012) 16083–16088, http://dx.doi.org/10.1073/pnas.1211658109.

[3] B. Blocken, Computational fluid dynamics for urban physics: Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations, Build. Environ. (ISSN: 0360-1323) 91 (2015) 219–245, http://dx.doi.org/10.1016/j.buildenv.2015.02.015, Fifty Year Anniversary for Building and Environment.

[4] T. Dauxois, T. Peacock, P. Bauer, C.P. Caulfield, C. Cenedese, C. Gorlé, G. Haller, G.N. Ivey, P.F. Linden, E. Meiburg, N. Pinardi, N.M. Vriend, A.W. Woods, Confronting grand challenges in environmental fluid mechanics, Phys. Rev. Fluids 6 (2021) 020501, http://dx.doi.org/10.1103/PhysRevFluids.6.020501.

[5] Y. Tominaga, L.L. Wang, Z.J. Zhai, T. Stathopoulos, Accuracy of CFD simulations in urban aerodynamics and microclimate: Progress and challenges, Build. Environ. (ISSN: 0360-1323) 243 (2023) 110723, http://dx.doi.org/10.1016/j.buildenv.2023.110723.

[6] M. Schatzmann, B. Leitl, Issues with validation of urban flow and dispersion CFD models, J. Wind Eng. Ind. Aerodyn. (ISSN: 0167-6105) 99 (4) (2011) 169–186, http://dx.doi.org/10.1016/j.jweia.2011.01.005, The Fifth International Symposium on Computational Wind Engineering.

[7] M. Neophytou, A. Gowardhan, M. Brown, An inter-comparison of three urban wind models using Oklahoma city joint urban 2003 wind field measurements, J. Wind Eng. Ind. Aerodyn. 99 (4) (2011) 357–368, http://dx.doi.org/10.1016/j.jweia.2011.01.010.

[8] G. Antonioni, S. Burkhart, J. Burman, A. Dejoan, A. Fusco, R. Gaasbeek, T. Gjesdal, A. Jäppinen, K. Riikonen, P. Morra, O. Parmhed, J. Santiago, Comparison of CFD and operational dispersion models in an urban-like environment, Atmos. Environ. (ISSN: 1352-2310) 47 (2012) 365–372, http://dx.doi.org/10.1016/j.atmosenv.2011.10.053.

[9] S. Hanna, Plume dispersion and concentration fluctuations in the atmosphere. Encyclopedia of environmental control technology, Air Pollut. Control. 2 (1989) 547–582.

[10] C. García-Sánchez, D. Philips, C. Gorlé, Quantifying inflow uncertainties for CFD simulations of the flow in downtown Oklahoma City, Build. Environ. (ISSN: 0360-1323) 78 (2014) 118–129, http://dx.doi.org/10.1016/j.buildenv.2014.04.013.

[11] D.D. Lucas, A. Gowardhan, P. Cameron-Smith, R.L. Baskett, Impact of meteorological inflow uncertainty on tracer transport and source estimation in urban atmospheres, Atmos. Environ. (ISSN: 1352-2310) 143 (2016) 120–132, http://dx.doi.org/10.1016/j.atmosenv.2016.08.019.

[12] D. Wise, V. Boppana, K. Li, H. Poh, Effects of minor changes in the mean inlet wind direction on urban flow simulations, Sustain. Cities Soc. (ISSN: 2210-6707) 37 (2018) 492–500, http://dx.doi.org/10.1016/j.scs.2017.11.041.

[13] J.L. Santiago, A. Dejoan, A. Martilli, F. Martin, A. Pinelli, Comparison between large-eddy simulation and Reynolds-averaged Navier–Stokes computations for the MUST field experiment. Part I: Study of the flow for an incident wind directed perpendicularly to the front array of containers, Bound.-Layer Meteorol. 135 (1) (2010) 109–132, http://dx.doi.org/10.1007/s10546-010-9466-3.

[14] H. Montazeri, B. Blocken, CFD simulation of wind-induced pressure coefficients on buildings with and without balconies: Validation and sensitivity analysis, Build Env. (ISSN: 0360-1323) 60 (2013) 137–149, http://dx.doi.org/10.1016/j.buildenv.2012.11.012.

[15] C. Gromke, N. Jamarkattel, B. Ruck, Influence of roadside hedgerows on air quality in urban street canyons, Atmos. Environ. (ISSN: 1352-2310) 139 (2016) 75–86, http://dx.doi.org/10.1016/j.atmosenv.2016.05.014.

[16] V. Winiarek, M. Bocquet, O. Saunier, A. Mathieu, Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant, J. Geophys. Res.: Atmos. 117 (D5) (2012) http://dx.doi.org/10.1029/2011JD016932.

[17] T.O. Spicer, G. Tickle, Simplified source description for atmospheric dispersion model comparison of the Jack Rabbit II chlorine field experiments, Atmos. Environ. (ISSN: 1352-2310) 244 (2021) 117866, http://dx.doi.org/10.1016/j.atmosenv.2020.117866.

[18] Y. Tominaga, T. Stathopoulos, Turbulent Schmidt numbers for CFD analysis with various types of flowfield, Atmos. Environ. (ISSN: 1352-2310) 41 (37) (2007) 8091–8099, http://dx.doi.org/10.1016/j.atmosenv.2007.06.054.

[19] Y. Tominaga, T. Stathopoulos, Numerical simulation of dispersion around an isolated cubic building: Comparison of various types of k–ϵ models, Atmos. Environ. (ISSN: 1352-2310) 43 (20) (2009) 3200–3210, http://dx.doi.org/10.1016/j.atmosenv.2009.03.038.

[20] C. Gorlé, G. Iaccarino, A framework for epistemic uncertainty quantification of turbulent scalar flux models for Reynolds-averaged Navier-Stokes simulations, Phys. Fluids (ISSN: 1070-6631) 25 (5) (2013) 055105, http://dx.doi.org/10.1063/1.4807067.

[21] C. Gorlé, C. Garcia-Sanchez, G. Iaccarino, Quantifying inflow and RANS turbulence model form uncertainties for wind engineering flows, J. Wind Eng. Ind. Aerodyn. 144 (2015) 202–212, http://dx.doi.org/10.1016/j.jweia.2015.03.025.

[22] H. Xiao, J.-L. Wu, J.-X. Wang, R. Sun, C. Roy, Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach, J. Comput. Phys. (ISSN: 0021-9991) 324 (2016) 115–136, http://dx.doi.org/10.1016/j.jcp.2016.07.038.

[23] E. Lumet, T. Jaravel, M.C. Rochoux, O. Vermorel, S. Lacroix, Assessing the internal variability of large-eddy simulations for microscale pollutant dispersion prediction in an idealized urban environment, Bound.-Layer Meteorol. (ISSN: 1573-1472) 190 (2) (2024) 9, http://dx.doi.org/10.1007/s10546-023-00853-7.

[24] X. Li, F. Xue, Bayesian inversion of inflow direction and speed in urban dispersion simulations, Build. Environ. (ISSN: 0360-1323) 144 (2018) 555–564, http://dx.doi.org/10.1016/j.buildenv.2018.08.042.

[25] J. Sousa, C. García-Sánchez, C. Gorlé, Improving urban flow predictions through data assimilation, Build. Environ. (ISSN: 0360-1323) 132 (2018) 282–290, http://dx.doi.org/10.1016/j.buildenv.2018.01.032.

[26] J. Sousa, C. Gorlé, Computational urban flow predictions with Bayesian inference: Validation with field data, Build. Environ. (ISSN: 0360-1323) 154 (2019) 13–22, http://dx.doi.org/10.1016/j.buildenv.2019.02.028.

[27] C.L. Defforge, B. Carissimo, M. Bocquet, R. Bresson, P. Armand, Improving CFD atmospheric simulations at local scale for wind resource assessment using the iterative ensemble Kalman smoother, J. Wind Eng. Ind. Aerodyn. (ISSN: 0167-6105) 189 (2019) 243–257, http://dx.doi.org/10.1016/j.jweia.2019.03.030.

[28] C.L. Defforge, B. Carissimo, M. Bocquet, R. Bresson, P. Armand, Improving numerical dispersion modelling in built environments with data assimilation using the iterative ensemble Kalman smoother, Bound.-Layer Meteorol. (ISSN: 1573-1472) 179 (2) (2021) 209–240, http://dx.doi.org/10.1007/s10546-020-00588-9.

[29] A. Keats, E. Yee, F.-S. Lien, Bayesian inference for source determination with applications to a complex urban environment, Atmos. Environ. (ISSN: 1352-2310) 41 (3) (2007) 465–479, http://dx.doi.org/10.1016/j.atmosenv.2006.08.044.

[30] V. Mons, L. Margheri, J.-C. Chassaing, P. Sagaut, Data assimilation-based reconstruction of urban pollutant release characteristics, J. Wind Eng. Ind. Aerodyn. (ISSN: 0167-6105) 169 (2017) 232–250, http://dx.doi.org/10.1016/j.jweia.2017.07.007.

[31] E. Launay, V. Hergault, M. Bocquet, J. Dumont Le Brazidec, Y. Roustan, Bayesian inversion of emissions from large urban fire using in situ observations, Atmos. Environ. (ISSN: 1352-2310) 323 (2024) 120391, http://dx.doi.org/10.1016/j.atmosenv.2024.120391.

[32] C. Defforge, Data assimilation for micrometeorological applications with the fluid dynamics model Code_Saturne (Ph.D. thesis), Université Paris-Est, 2019, URL https://theses.hal.science/tel-02318713v2/document. (Accessed 01 December 2023).

[33] E. Aristodemou, R. Arcucci, L. Mottet, A. Robins, C. Pain, Y.-K. Guo, Enhancing CFD-LES air pollution prediction accuracy using data assimilation, Build. Environ. (ISSN: 0360-1323) 165 (2019) 106383, http://dx.doi.org/10.1016/j.buildenv.2019.106383.

[34] P. Bauweraerts, J. Meyers, Reconstruction of turbulent flow fields from lidar measurements using large-eddy simulation, J. Fluid Mech. 906 (2021) A17, http://dx.doi.org/10.1017/jfm.2020.805.

[35] E. Yee, C.A. Biltoft, Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles, Bound.-Layer Meteorol. 111 (3) (2004) 363–415, http://dx.doi.org/10.1023/B:BOUN.0000016496.83909.ee.

[36] A.A. Emerick, A.C. Reynolds, Ensemble smoother with multiple data assimilation, Comput. Geosci. (ISSN: 0098-3004) 55 (2013) 3–15, http://dx.doi.org/10.1016/j.cageo.2012.03.011, Ensemble Kalman filter for data assimilation.

[37] A. Marrel, N. Perot, C. Mottet, Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators, Stoch. Environ. Res. Risk Assess. (ISSN: 1436-3259) 29 (3) (2015) 959–974, http://dx.doi.org/10.1007/s00477-014-0927-y.

[38] B.X. Nony, M.C. Rochoux, T. Jaravel, D. Lucor, Reduced-order modeling for parameterized large-eddy simulations of atmospheric pollutant dispersion, Stoch. Env. Res. Risk Assess. (ISSN: 1436-3259) 37 (6) (2023) 2117–2144, http://dx.doi.org/10.1007/s00477-023-02383-7.

[39] E. Lumet, T. Jaravel, M.C. Rochoux, Dataset of microscale atmospheric flow and pollutant concentration large-eddy simulations for varying mesoscale meteorological forcing in an idealized urban environment, Data Brief (ISSN: 2352-3409) 58 (2025) 111285, http://dx.doi.org/10.1016/j.dib.2025.111285.

[40] E. Lumet, M.C. Rochoux, T. Jaravel, S. Lacroix, Uncertainty-aware surrogate modeling for urban air pollutant dispersion prediction, Build. Environ. (ISSN: 0360-1323) 267 (2025) 112287, http://dx.doi.org/10.1016/j.buildenv.2024.112287.

[41] K. Beven, J. Freer, Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol. (ISSN: 0022-1694) 249 (1) (2001) 11–29, http://dx.doi.org/10.1016/S0022-1694(01)00421-8.

[42] C. Biltoft, Customer report for Mock Urban Setting Test, DPG Document No. WDTC-FR-01-121, West Desert Test Center, U.S. Army Dugway Proving Ground, Utah, USA, 2001.

[43] M. Schatzmann, H. Olesen, J. Franke, COST 732 Model Evaluation Case Studies: Approach and Results, Technical Report, University of Hamburg, Meteorological Institute, 2010, URL https://www.researchgate.net/profile/George-Efthimiou-3/post/Has-fluent-been-compared-to-starccm/attachment/59d6585379197b80779ae4bd/AS%3A538043318628353%401505290931380/download/5th_Docu_May_10.pdf. (Accessed 01 December 2023).

[44] M. Milliez, B. Carissimo, Numerical simulations of pollutant dispersion in an idealized urban area, for different meteorological conditions, Bound.-Layer Meteorol. 122 (2) (2007) 321–342, http://dx.doi.org/10.1007/s10546-006-9110-4.

[45] M. König, Large-Eddy Simulation Modelling for Urban Scale (Ph.D. thesis), University of Leipzig, 2014, URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=baab9d7b41623099c1b6d840c11821b8e31fac9b. (Accessed 05 August 2024).

[46] T. Nagel, R. Schoetter, V. Masson, C. Lac, B. Carissimo, Numerical analysis of the atmospheric boundary-layer turbulence influence on microscale transport of pollutant in an idealized urban environment, Bound.-Layer Meteorol. 184 (1) (2022) 113–141, http://dx.doi.org/10.1007/s10546-022-00697-7.

[47] T. Schönfeld, M. Rudgyard, Steady and unsteady flow simulations using the hybrid flow solver AVBP, AIAA J. 37 (11) (1999) 1378–1385, http://dx.doi.org/10.2514/2.636.

[48] L.Y. Gicquel, N. Gourdain, J.-F. Boussuge, H. Deniau, G. Staffelbach, P. Wolf, T. Poinsot, High performance parallel computing of flows in complex geometries, Comptes Rendus Mécanique (ISSN: 1631-0721) 339 (2) (2011) 104–124, http://dx.doi.org/10.1016/j.crme.2010.11.006, High Performance Computing.

[49] F. Nicoud, F. Ducros, Subgrid-scale stress modelling based on the square of the velocity gradient tensor, Flow Turbul. Combust. 62 (3) (1999) 183–200, http://dx.doi.org/10.1023/A:1009995426001.

[50] Y. Tominaga, A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, T. Shirasawa, AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings, J. Wind Eng. Ind. Aerodyn. (ISSN: 0167-6105) 96 (10) (2008) 1749–1761, http://dx.doi.org/10.1016/j.jweia.2008.02.058, 4th International Symposium on Computational Wind Engineering (CWE2006).

[51] J. Franke, A. Hellsten, H. Schlünzen, B. Carissimo, Best Practice Guideline for the CFD Simulation of Flows in the Urban Environmen, Technical Report, COST European Cooperation in Science and Technology, 2007, URL https://hal.science/hal-04181390. (Accessed 01 December 2023).

[52] E. Lumet, Assessing and reducing uncertainty in large-eddy simulation for microscale atmospheric dispersion (Ph.D. thesis), Université Toulouse III - Paul Sabatier, 2024, URL https://theses.fr/2024TLSES003. (Accessed 30 May 2024).

[53] P. Richards, R. Hoxey, Appropriate boundary conditions for computational wind engineering models using the k–ϵ turbulence model, J. Wind Eng. Ind. Aerodyn. (ISSN: 0167-6105) 46–47 (1993) 145–153, http://dx.doi.org/10.1016/0167-6105(93)90124-7, Proceedings of the 1st International on Computational Wind Engineering.

[54] A. Smirnov, S. Shi, I. Celik, Random flow generation technique for large eddy simulations and particle-dynamics modeling, J. Fluids Eng. (ISSN: 0098-2202) 123 (2) (2001) 359–371, http://dx.doi.org/10.1115/1.1369598.

[55] R. Vasaturo, I. Kalkman, B. Blocken, P. van Wesemael, Large eddy simulation of the neutral atmospheric boundary layer: Performance evaluation of three inflow methods for terrains with different roughness, J. Wind Eng. Ind. Aerodyn. 173 (2018) 241–261, http://dx.doi.org/10.1016/j.jweia.2017.11.025.

[56] L. Margheri, P. Sagaut, A hybrid anchored-ANOVA – POD/Kriging method for uncertainty quantification in unsteady high-fidelity CFD simulations, J. Comput. Phys. (ISSN: 0021-9991) 324 (2016) 137–173, http://dx.doi.org/10.1016/j.jcp.2016.07.036.

[57] C. García-Sánchez, G. Van Tendeloo, C. Gorlé, Quantifying inflow uncertainties in RANS simulations of urban pollutant dispersion, Atmos. Environ. (ISSN: 1352-2310) 161 (2017) 263–273, http://dx.doi.org/10.1016/j.atmosenv.2017.04.019.

[58] M.C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, A. Trouvé, Towards predictive data-driven simulations of wildfire spread – part I: Reduced-cost ensemble Kalman filter based on a polynomial chaos surrogate model for parameter estimation, Nat. Hazards Earth Syst. Sci. 14 (11) (2014) 2951–2973, http://dx.doi.org/10.5194/nhess-14-2951-2014.

[59] A.A. Emerick, A.C. Reynolds, History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations, Comput. Geosci. (ISSN: 1573-1499) 16 (3) (2012) 639–659, http://dx.doi.org/10.1007/s10596-012-9275-5.

[60] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res.: Ocean. 99 (C5) (1994) 10143–10162, http://dx.doi.org/10.1029/94JC00572.

[61] G. Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean. Dyn. (ISSN: 1616-7228) 53 (4) (2003) 343–367, http://dx.doi.org/10.1007/s10236-003-0036-9.

[62] G. Evensen, F.C. Vossepoel, P.J. van Leeuwen, Iterative ensemble smoothers for data assimilation in coupled nonlinear multiscale models, Mon. Weather Rev. 152 (6) (2024) 1277–1301, http://dx.doi.org/10.1175/MWR-D-23-0239.1.

[63] C. Zhang, M. Rochoux, W. Tang, M. Gollner, J.-B. Filippi, A. Trouvé, Evaluation of a data-driven wildland fire spread forecast model with spatially-distributed parameter estimation in simulations of the FireFlux I field-scale experiment, Fire Saf. J. (ISSN: 0379-7112) 91 (2017) 758–767, http://dx.doi.org/10.1016/j.firesaf.2017.03.057, URL http://www.sciencedirect.com/science/article/pii/S0379711217301005. Fire Safety Science: Proceedings of the 12th International Symposium.

[64] J.H. Halton, Algorithm 247: Radical-inverse quasi-random point sequence, Commun. ACM 7 (12) (1964) 701–702, http://dx.doi.org/10.1145/355588.365104.

[65] G. Burgers, P.J. van Leeuwen, G. Evensen, Analysis scheme in the ensemble Kalman filter, Mon. Weather Rev. 126 (6) (1998) 1719–1724, http://dx.doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

[66] P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev. 126 (3) (1998) 796–811, http://dx.doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

[67] A. Carrassi, M. Bocquet, L. Bertino, G. Evensen, Data assimilation in the geosciences: An overview of methods, issues, and perspectives, WIREs Clim. Chang. 9 (5) (2018) e535, http://dx.doi.org/10.1002/wcc.535.

[68] Y. Liu, J.-M. Haussaire, M. Bocquet, Y. Roustan, O. Saunier, A. Mathieu, Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima Daiichi accidental releases of radionuclides, Q. J. R. Meteorol. Soc. 143 (708) (2017) 2886–2901, http://dx.doi.org/10.1002/qj.3138.

[69] M. Cassiani, M.B. Bertagni, M. Marro, P. Salizzoni, Concentration fluctuations from localized atmospheric releases, Bound.-Layer Meteorol. (ISSN: 1573-1472) 177 (2) (2020) 461–510, http://dx.doi.org/10.1007/s10546-020-00547-4.

[70] C.P. Arnold, C.H. Dey, Observing-systems simulation experiments: Past, present, and future, Bull. Am. Meteorol. Soc. 67 (6) (1986) 687–695, http://dx.doi.org/10.1175/1520-0477(1986)067<0687:OSSEPP>2.0.CO;2.

[71] I.M. Sobol', On sensitivity estimation for nonlinear mathematical models (in Russian), Mat. Model. 2 (1) (1990) 112–118, URL http://mi.mathnet.ru/mm2320. (Accessed 01 December 2023).

[72] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Comm. 181 (2) (2010) 259–270, http://dx.doi.org/10.1016/j.cpc.2009.09.018.

[73] J. Chang, S. Hanna, Air quality model performance evaluation, Meteorol. Atm. Phys 87 (1) (2004) 167–196, http://dx.doi.org/10.1007/s00703-003-0070-7.

[74] C. Zhang, A. Collin, P. Moireau, A. Trouvé, M.C. Rochoux, State-parameter estimation approach for data-driven wildland fire spread modeling: Application to the 2012 RxCADRE S5 field-scale experiment, Fire Saf. J. (ISSN: 0379-7112) 105 (2019) 286–299, http://dx.doi.org/10.1016/j.firesaf.2019.03.009.

[75] N.J. Gordon, D.J. Salmond, A.F. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc. F (ISSN: 0956-375X) 140 (1993) http://dx.doi.org/10.1049/ip-f-2.1993.0015, 107–113(6).

[76] P.L. Houtekamer, H.L. Mitchell, P. Gérard, B. Mark, C. Martin, S. Lubos, H. Bjarne, Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, Mon. Weather Rev. 133 (3) (2005) 604–620, http://dx.doi.org/10.1175/MWR-2864.1.

[77] M. Ehrendorfer, A review of issues in ensemble-based Kalman filtering, Meteorol. Z. 16 (6) (2007) 795–818, http://dx.doi.org/10.1127/0941-2948/2007/0256.

[78] G. Evensen, The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Syst. Mag. 29 (3) (2009) 83–104, http://dx.doi.org/10.1109/MCS.2009.932223.

[79] P.J. Smith, G.D. Thornhill, S.L. Dance, A.S. Lawless, D.C. Mason, N.K. Nichols, Data assimilation for state and parameter estimation: application to morphodynamic modelling, Q. J. R. Meteorol. Soc. 139 (671) (2013) 314–327, http://dx.doi.org/10.1002/qj.1944.

[80] Y. Ruckstuhl, T. Janjić, Combined state-parameter estimation with the LETKF for convective-scale weather forecasting, Mon. Weather Rev. 148 (4) (2020) 1607–1628, http://dx.doi.org/10.1175/MWR-D-19-0233.1.

[81] M. Asch, M. Bocquet, M. Nodet, Data Assimilation: Methods, Algorithms, and Applications, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016, http://dx.doi.org/10.1137/1.9781611974546.

[82] C.H. Bishop, B.J. Etherton, S.J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, Mon. Weather Rev. 129 (3) (2001) 420–436, http://dx.doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

[83] P.L. Houtekamer, F. Zhang, Review of the ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev. 144 (12) (2016) 4489–4532, http://dx.doi.org/10.1175/MWR-D-15-0440.1.

[84] A. Hirvoas, C. Prieur, E. Arnaud, F. Caleyron, M. Munoz Zuniga, Quantification and reduction of uncertainties in a wind turbine numerical model based on a global sensitivity analysis and a recursive Bayesian inference approach, Internat. J. Numer. Methods Engrg. 122 (10) (2021) 2528–2544, http://dx.doi.org/10.1002/nme.6630.

[85] Aurora, Instruction manual model 200b miniPID fast response miniature photoionization detector, 2012, URL https://bulletins.aurorascientific.com/manuals/200B.Manual.pdf. (Accessed 09 December 2024).

[86] E. Lumet, T. Jaravel, M.C. Rochoux, PPMLES – perturbed-parameter ensemble of MUST large-eddy simulations, 2024, http://dx.doi.org/10.5281/zenodo.11394347, Dataset. Zenodo.