Number of words (excluding acknowledgements and references): 8930

Uncertainty-aware surrogate modeling for urban air pollutant dispersion prediction

Eliott Lumet[1 2] · Mélanie C. Rochoux[2] · Thomas Jaravel[2] · Simon Lacroix[3]

Abstract This study evaluates a surrogate modeling approach that provides rapid ensemble predictions of air pollutant dispersion in urban environments for varying meteorological forcing, while estimating irreducible and modeling uncertainties. The POD–GPR approach combining Proper Orthogonal Decomposition (POD) and Gaussian Process Regression (GPR) is applied to emulate the response surface of a Large-Eddy Simulation (LES) model of the Mock Urban Setting Test (MUST) field-scale experiment. We design and validate new methods for i) selecting the POD-latent space dimension to avoid overfitting noisy structures due to atmospheric internal variability, and ii) estimating the uncertainty in POD–GPR predictions. To train and validate the POD–GPR surrogate in an offline phase, we build a large dataset of 200 LES 3-D time-averaged concentration fields, which are subject to substantial spatial variability from near-source to background concentration and have a very large dimension of several million grid cells. The results show that POD–GPR reaches the best achievable accuracy levels, except for the highest concentration near the source, while predicting full fields at a computational cost five orders of magnitude lower than an LES. The results also show that the proposed mode selection criterion avoids perturbing the surrogate response surface, and that the uncer-

[1]Corresponding author: eliott.lumet@gmail.com

[2]CECI, Université de Toulouse, CNRS, CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 1, France

[3]LAAS-CNRS, Université de Toulouse, CNRS, 7 Avenue du Colonel Roche, BP54200, 31031 Toulouse cedex 4, France

tainty estimate explains a large part of the surrogate error and is spatially consistent with the observed internal variability. Finally, POD–GPR can be robustly trained with much smaller datasets, paving the way for application to realistic urban configurations.

Keywords Surrogate modeling · Uncertainty quantification · Microscale pollutant dispersion · Urban flow · Large-eddy simulation · Internal variability

1 Introduction

Accidental releases of pollutants into the atmosphere, such as from industrial accidents, can degrade air quality and have significant short- and long-term health impacts (EEA 2020; Manisalidis et al. 2020). In urban environments, these risks are exacerbated by high population density and reduced ventilation due to the urban canopy, leading to local pollution peaks (Fernando et al. 2001; Klein et al. 2007; Pasquier et al. 2023). For accurate mapping of these peaks and associated exposures, it is necessary to develop microscale dispersion models that take into account i) the effect of urban buildings on the local flow, and ii) the inherently multiscale and turbulent nature of the Atmospheric Boundary Layer (ABL).

To gain relevant insight into these processes, there is a growing consensus in the research community for the use of Computational Fluid Dynamics (CFD) (Blocken 2015; Tominaga et al. 2023). Advanced models based on Reynolds-Averaged Navier-Stokes (RANS) and Large-Eddy Simulation (LES) are able to represent complex flow structures, in particular due to the interactions between the atmosphere and the built environment. However, their use in operational applications remains limited because their high computational cost prevents them from being used in real time, for example in emergency response. Moreover, they still suffer from a lack of accuracy compared to field and wind tunnel measurements due to the large uncertainties involved (Schatzmann and Leitl 2011; Blocken 2014; Dauxois et al. 2021). These uncertainties can be classified into three different groups:

- *boundary condition uncertainties* due to measurement and representativeness errors in calibration data, and to boundary condition modeling assumptions, in relation to: i) the meteorological forcing (García-Sánchez et al. 2014; Lucas et al. 2016; Wise et al. 2018), ii) the urban geometry representation (Santiago et al. 2010; Montazeri and Blocken 2013; Gromke et al. 2016), and iii) the pollutant source (Winiarek et al. 2012; Spicer and Tickle 2021);

- *structural modeling uncertainties*, inherent to the model solver and its underlying modeling assumptions, mainly related to turbulence modeling (Tominaga and Stathopoulos 2007; Blocken et al. 2008; Yue Yang and Wang 2008; Tominaga and Stathopoulos 2009; Gorlé and Iaccarino 2013; Gorlé et al. 2015; Xiao et al. 2016);
- *aleatory uncertainties*, mostly due to the turbulent and therefore stochastic nature of the ABL, and referred to as internal variability, which results in an irreducible uncertainty and is largely responsible for the discrepancies between field measurements and CFD model predictions (Schatzmann and Leitl 2011; Neophytou et al. 2011; Antonioni et al. 2012; García-Sanchez et al. 2018; Dauxois et al. 2021; Lumet et al. 2024b).

In this work, we focus on atmospheric uncertainties, i.e. in how to represent the impact of large-scale atmospheric forcing uncertainties and internal variability on microscale LES field predictions. We have chosen not to consider structural modeling uncertainties, as these have been extensively studied and remain small in the LES context. Instead, we have chosen to investigate how to design a surrogate modeling approach to quantify boundary condition uncertainties in LES, while accounting for internal variability. To our knowledge, the coupling between these two sources of uncertainty has not yet been studied, while this is one challenge expressed by Dauxois et al. (2021) and Wu and Quan (2024).

Surrogate modeling, also known as reduced-order modeling, aims at accurately emulating the response surface of complex and expensive numerical models, while significantly reducing computational time. By enabling real-time and large ensemble predictions, surrogate modeling is well suited to address the dual challenges of high cost and uncertainty in LES models, making it a hot topic of research in the CFD field (Lassila et al. 2014; Vinuesa and Brunton 2022). For parametric studies, surrogate models are mostly based on fully data-driven approaches, which consist of learning the response surface of the CFD model from a dataset of reference simulations precomputed during an offline phase, to then provide fast predictions during an online phase. They have been successfully used to emulate urban wind and/or pollutant dispersion, with respect to urban geometry (Wu et al. 2021; Huang et al. 2022; Mendil et al. 2022; Kastner and Dogan 2023), or meteorological forcing and pollutant source (Margheri and Sagaut 2016; Xiang et al. 2021; Nony et al. 2023). Surrogate models are therefore valuable for ensemble prediction in more complex frameworks such as urban design optimization (Wu et al. 2021; Wu and Quan 2024), sensitivity analysis (Cheng et al. 2020; Fellmann et al. 2023), uncertainty quantification (García-Sánchez et al. 2014, 2017), and data assimilation (Mons et al. 2017; Sousa et al. 2018; Sousa and Gorlé 2019; Lumet 2024).

While surrogate models have proven to be valuable tools for dealing with uncertainties related to CFD model boundary conditions, few studies have addressed the representation of internal variability, which is at least as important (Neophytou et al. 2011; Antonioni et al. 2012; Lumet et al. 2024b). Moreover, surrogate models introduce a new form of structural uncertainty: the model reduction error, i.e. the error of the surrogate model relative to the full-order model. Our aim is to evaluate the model reduction error in a comprehensive and robust way, and to assess the ability of the surrogate model to retrieve reliable information on internal variability from the LES dataset and compare it with the model reduction error.

To this end, we adopt a surrogate modeling approach called POD–GPR (Marrel et al. 2015), which combines Proper Orthogonal Decomposition (Sirovich 1987; Berkooz et al. 1993) and Gaussian Process Regression (Rasmussen et al. 2006). It is a robust and standard method that has already been used for urban wind and pollutant dispersion prediction (Xiao et al. 2019; Xiang et al. 2021; Weerasuriya et al. 2021; Masoumi-Verki et al. 2022; Nony et al. 2023; Fellmann et al. 2023). In this study, we construct a POD–GPR model for the MUST experiment of propylene dispersion in a simplified urban canopy (Yee and Biltoft 2004). For this purpose, we generate a large dataset of 200 LES using the model validated by Lumet et al. (2024b) by varying the wind boundary forcing. We choose LES over the more common and less expensive RANS approach because: i) LES is expected to reduce structural uncertainties due to turbulence modeling compared to RANS (Gousseau et al. 2011; García-Sanchez et al. 2018), and ii) LES provides instantaneous snapshots of the most energetic atmospheric eddies and can thus be used to estimate the effect of the microscale internal variability of the ABL on tracer dispersion (Lumet et al. 2024b), which is central to the objective of this study.

The novelty of the proposed surrogate modeling approach is related to the POD latent space, i.e. the reduced space compressing the LES information, and is twofold. First, we define a method to choose a priori the POD-latent space dimension, based on the projection of the internal variability into the latent space. Secondly, knowing the internal variability in the LES data and using regression uncertainty estimates from Gaussian processes, we develop a mathematical framework for propagating these uncertainty estimates from the POD latent space to the physical space to help interpret the uncertainty results, which to our knowledge has been little studied in physical applications.

This article is structured as follows: Section 2 briefly introduces the learning dataset of LES simulations. Section 3 describes the POD–GPR surrogate modeling approach and introduces our methods to estimate prediction uncertainty and select the latent space dimension. Finally, Section 4 provides a comprehensive validation of the POD–GPR predictions, uncertainty estimates, and ability to handle reduced-size training datasets.

2 Learning dataset of large-eddy simulations

This section summarizes the key points of the LES model for the MUST experiment, which has been extensively validated in previous work (Lumet et al. 2024b) and which is used here to build the surrogate learning dataset. Details are given on the choice of the parameter space, the field quantities of interest and the associated internal variability.

2.1 The MUST field campaign

MUST is a field-scale experiment conducted in September 2001 at the US Army Dugway Proving Ground test site in Utah's desert to collect extensive measurements of urban pollutant dispersion (Biltoft 2001; Yee and Biltoft 2004). During the field campaign, a series of trials were carried out by releasing a passive tracer, propylene, at different locations within an urban-like canopy consisting of 120 regularly-spaced shipping containers. It is a canonical experiment for dispersion model validation: i) it was selected as one of the reference case studies for the COST Action 732 CFD dispersion model intercomparison (Franke et al. 2007), and ii) it has been used in a large number of CFD studies involving RANS (Hanna et al. 2004; Hsieh et al. 2007; Milliez and Carissimo 2007; Donnelly et al. 2009; Efthimiou et al. 2011; Kumar et al. 2015; Bahlali et al. 2019) or LES (Camelli et al. 2005; Antonioni et al. 2012; König 2014; Nagel et al. 2022). In this study, we focus on the trial 2681829 corresponding to neutral atmospheric conditions.

2.2 LES model of the MUST field experiment

We use the AVBP[1] (Schönfeld and Rudgyard 1999; Gicquel et al. 2011) code to build the LES model. AVBP solves the LES-filtered Navier-Stokes equations on unstructured mesh using a second-order Lax-Wendroff finite-volume centered numerical scheme (Schönfeld and Rudgyard 1999) and using pressure gradient scaling since the atmospheric flow features a low Mach number (Ramshaw et al. 1986). Tracer dispersion is modeled by the LES-filtered advection-diffusion equation using an Eulerian approach. Subgrid-scale turbulence is modeled using the Wall-Adaptative Local Eddy-Viscosity (WALE) model (Nicoud and Ducros 1999) for subgrid momentum transport, and a gradient-diffusion hypothesis for subgrid tracer transport (with the turbulent Schmidt number equal to $\mathrm{S}_c^t = 0.6$).

The computational domain is a rectangular box with dimensions of 420 m by 420 m by 50 m, discretized with a boundary-fitted mesh of 91 million tetrahedra, with a resolution ranging from 0.3 m in the canopy to 5 m at the top of the domain.

In terms of boundary conditions, a logarithmic wind profile is imposed at the inlet so

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

that the mean inlet wind velocity vector $\overline{\mathbf{u}}$ reads

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

where κ is the von Kármán constant equal to 0.4, z_0 is the aerodynamic roughness length, and u_* is the friction velocity. In addition, a synthetic turbulence injection method (Kraichnan 1970; Smirnov et al. 2001) is used to impose the upstream wind fluctuations based on a precomputed Reynolds tensor from a precursor run (corresponding to a simulation with the same surface roughness but without obstacles, and with periodic boundary conditions at the inlet and outlet inspired by Vasaturo et al. (2018)). At the lateral boundaries, symmetry boundary conditions are used. Static pressure is imposed at the outlet and top boundaries. Standard laws of the wall are imposed for the ground and obstacle boundaries. The pollutant source is modeled by a local source term in the advection-diffusion equation to match the experimental volumetric flow rate. A more detailed description of the boundary conditions is given in Lumet et al. (2024b).

To be comparable to the MUST observational time series, we need to simulate a 200-s time sequence for each snapshot of the learning dataset. Before running this time sequence, we need to initialize each simulation until first- and second-order statistics of the flow and tracer variables reach a stationary state. For this initialization, a spin-up time $t_{spin-up}$ of 1.5 times the convective time scale is used:

$$t_{spin-up} = 1.5 \times \left(\frac{L}{U_{bulk}}\right) = 1.5 \times \frac{\kappa \, H \, L}{u_* \left[(H + z_0) \ln\left(\frac{H+z_0}{z_0}\right) - H\right]}, \tag{2}$$

with $L = 420\,\mathrm{m}$ the domain length and $H = 50\,\mathrm{m}$ the domain height. This spin-up time is specific to each snapshot as the bulk velocity U_{bulk} is an uncertain quantity (Sect. 2.3). Note that the average computational cost for a given simulation of 200 s is around 15,000 core hours, which motivates the development of a surrogate model to speed up predictions.

2.3 Definition of the input parameter space

2.3.1 Choice of input parameters

In this work, we focus on atmospheric parametric uncertainties. For the surrogate model to be useful, it must capture the dependence of the tracer dispersion on the most influential and uncertain atmospheric parameters of the LES model. In preliminary work (Lumet (2024), Chapter III), we carried out one-at-a-time sensitivity analysis and showed that the inlet wind direction α_{inlet} and the friction velocity u_* are the two parameters that most significantly affect the LES mean concentration predictions. In particular, the aerodynamic roughness length z_0 is well identified in the MUST experiment (z_0 is equal to $0.045 \pm 0.005\,\mathrm{m}$ according to observations, Yee and Biltoft (2004)) and was found to have a negligible impact. For these reasons, we consider only two uncertain parameters:

$$\boldsymbol{\theta} = (\alpha_{inlet}, u_*) \,, \tag{3}$$

to define the input space of the surrogate model. Note that this choice is quite common in urban flow surrogate modeling (Margheri and Sagaut 2016; García-Sánchez et al. 2014, 2017). Note also that, under neutral conditions, the mean concentration is inversely proportional to the friction velocity and the reduction problem could thus be simplified by predicting dimensionless quantities, as done by Sousa et al. (2018) and Lamberti and Gorlé (2021). This normalization was investigated in Lumet (2024), Chapter IV, but we choose to present results with multiple input dimensions here for generalization purposes.

2.3.2 Parameter variation ranges

The surrogate model must cover a wide, but plausible and feasible, range of variation in the input parameters (Eq. 3). Based on a microclimatology constructed using all available data from the closest micrometeorological station to the MUST site (Lumet (2024), Chapter IV), all wind directions are likely to occur and more than 99% of the horizontal wind speed measurements at $z = 10\,\mathrm{m}$ are below $12\,\mathrm{m\,s^{-1}}$, which corresponds to a friction velocity u_* of $0.89\,\mathrm{m\,s^{-1}}$ and which is therefore chosen as the maximum friction velocity here. We limit the minimum friction velocity to $0.07\,\mathrm{m\,s^{-1}}$, which corresponds to a wind speed of about $1\,\mathrm{m\,s^{-1}}$ at an altitude of $10\,\mathrm{m}$, since we are interested in windy conditions. To reduce the number of LES, we also restrict the range of variation for the inlet wind direction to wind directions for which the plume crosses the array of containers. In the end, the input parameter space reads

$$\Omega_{\boldsymbol{\theta}} = [-90\,°,\ 30\,°] \times [0.07\,\mathrm{m\,s^{-1}},\ 0.89\,\mathrm{m\,s^{-1}}]. \tag{4}$$

2.3.3 Parameter space sampling

To sample the input parameter space (Eq. 4), we use Halton's sequence (1964). As a low-discrepancy sequence, it samples the space uniformly and more efficiently than a purely random sequence for a limited number of samples, avoiding redundant sampling in the same areas and it is well adapted to a small number of parameters. Figure 1 shows the location of the 200 samples thus obtained in the uncertain parameter space.

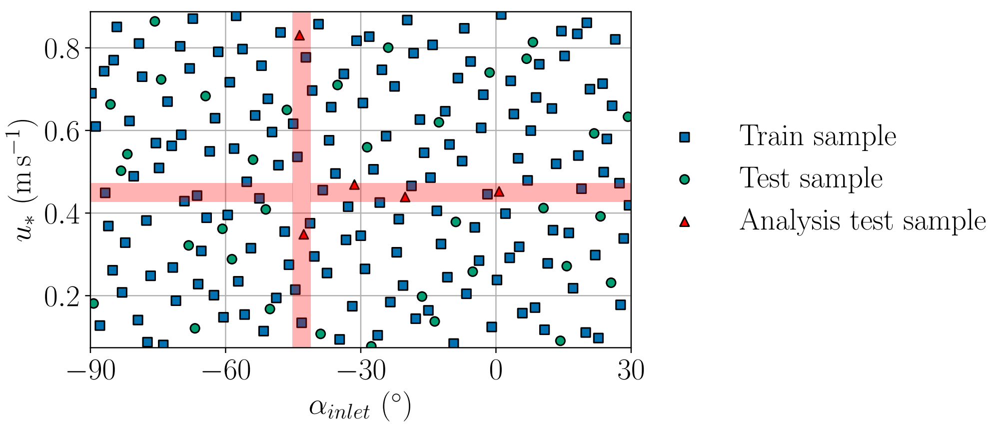

Figure 1: *Input parameter space sampling obtained with Halton's sequence. Each point is a pair of parameters for which we perform an LES prediction. The training (80%) and test (20%) sets are represented as blue squares and green circles, respectively. The horizontal red shaded area corresponds to the parameter space sub-section scanned by taking a margin of ±5 % around the constant friction speed* $u_*^{plot} = 0.45\,m\,s^{-1}$*. The vertical shaded area is similarly defined around the constant inlet wind direction* $\alpha_{inlet}^{plot} = -43\,°$ *with a margin of* $\pm 2\,°$*. The test samples within these ranges (red triangles) are used in Sect. 4.3 to evaluate the surrogate model.*

2.4 Generation of the LES dataset

We run an LES for each of the 200 input parameter samples (Fig. 1) to provide the learning dataset for the surrogate model. The main quantity of interest for the surrogate modeling approach is the 3-D mean (time-averaged) concentration field averaged over the 200-s analysis time period of the MUST experiment.

To generate this ensemble, the computational domain is rotated to align with the mean wind direction α_{inlet} to avoid inducing lateral confinement and numerical instabilities due to the shear-free boundary conditions at the domain sides. The spin-up time before collecting LES statistics is scaled by the friction velocity according to Eq. 2 to account for the slowing down of the flow establishment with decreasing u_*. Finally, the Reynolds stress tensor prescribed for the turbulent injection method is rescaled by u_*^2 following similarity theory.

The total cost of generating this LES ensemble is about 5.7 million core hours. Note that a subset of the most relevant data from these simulations, including all the data used in this study, is available in open access (Lumet et al. 2024a).

Figure 2a shows the topology of the LES ensemble with the example of the mean concentration c at one specific location within the canopy (the green square in Fig. 2b, c corresponding to the tower B in the MUST experiment). The mean concentration increases linearly with decreasing friction velocity. The dependence on the wind direction is more complex with a concentration maximum obtained for $\alpha_{inlet} \approx 30\,^\circ$ and a rapid decay in both directions down to 0 ppm as the plume no longer crosses the probe location. The two examples of horizontal cuts of the LES mean concentration fields (Fig. 2b, c) obtained for two different wind conditions highlight the high spatial variability of the fields, especially within the plumes, which is a challenge for the surrogate modeling problem.

2.5 Noise in the learning dataset

Atmospheric flows are naturally unsteady with strong variations occurring over a wide range of frequencies corresponding to the time scales of the atmospheric eddies. When considering statistics over finite temporal periods, this internal variability yields sampling

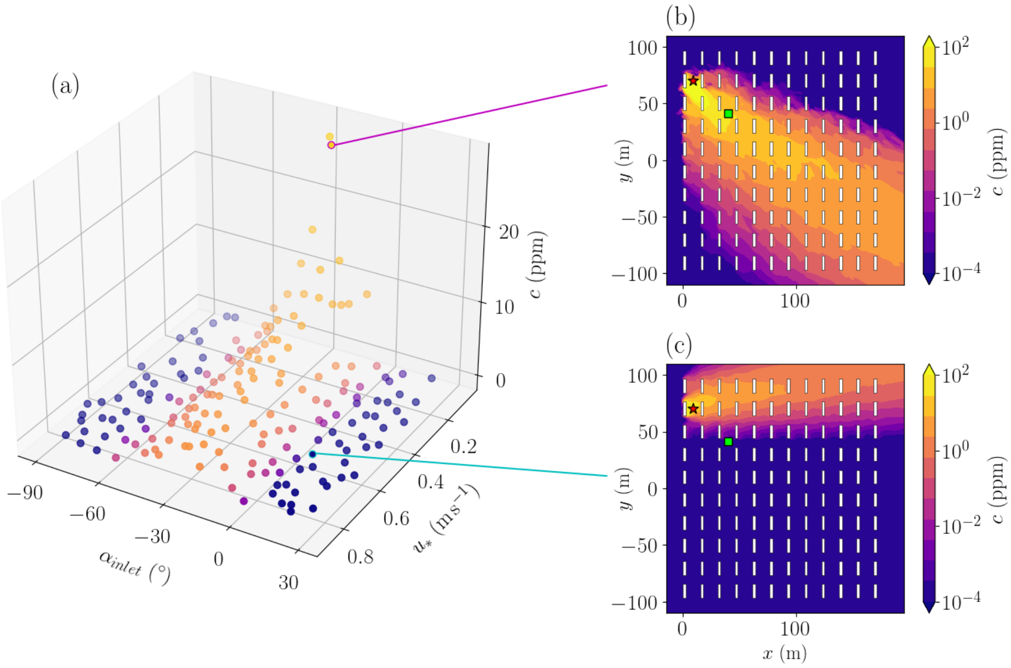

Figure 2: *(a) LES prediction of the local mean (time-averaged) concentration c at tower B at $z = 2\,m$ for each sample of parameters $\boldsymbol{\theta} = (\alpha_{inlet}, u_*)$ from Fig. 1. (b, c) Horizontal cuts of the mean concentration at $z = 1.6\,m$ for the two samples $\left(\alpha_{inlet}^{(81)}, u_*^{(81)}\right) = \left(-27.7\,°, 0.08\,m\,s^{-1}\right)$ and $\left(\alpha_{inlet}^{(133)}, u_*^{(133)}\right) = \left(7.73\,°, 0.60\,m\,s^{-1}\right)$ in (a). The green square corresponds to the tower B, and the red star corresponds to the tracer source.*

errors and is therefore a source of aleatory uncertainty, which is inherent to the physical system under study and thereby irreducible. For the MUST case, internal variability has a significant impact on the tracer concentration statistics when computed over the standard 200-s analysis period (Schatzmann et al. 2010; Lumet et al. 2024b). One of the challenges of this study is to build a surrogate model that explicitly estimates this uncertainty when emulating the mean concentration fields.

To quantify the effect of internal variability on the LES predictions, we use the stationary bootstrap approach from Lumet et al. (2024b), which relies on resampling of the sub-averages of the physical fields using the algorithm of Politis and Romano (1994) and which involves a mean bootstrap block length to account for temporal correlation between sub-averages. This approach is applied separately for each snapshot in the dataset (Fig. 1) using 1,000 bootstrap replicates to estimate the internal variability.

Figure 3 confirms that the internal microscale variability of the ABL significantly

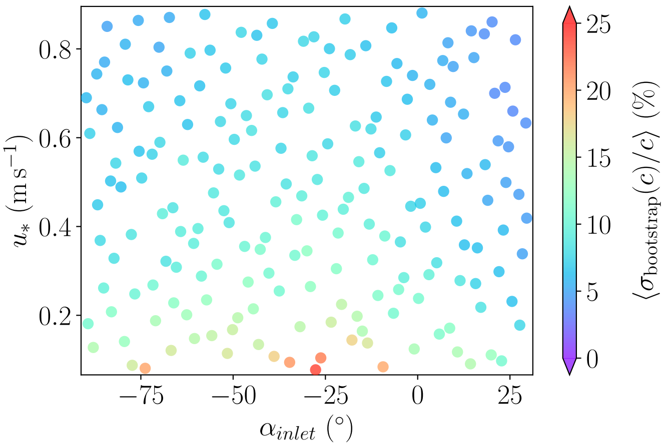

Figure 3: *Relative uncertainty of the mean concentration in the parameter space estimated using stationary bootstrap (Lumet et al. 2024b) and averaged over the whole spatial domain. Each circle corresponds to the averaged uncertainty of one LES sample of the learning dataset obtained from Halton's sequence (Fig. 1).*

affects the LES learning dataset, with spatially-averaged relative standard deviations of up to just over 20% for a few samples of the LES dataset. Looking at the mean concentration fields, these deviations can be even larger locally, especially in areas of strong gradients or close to the source. We note in Fig. 3 that the noise induced by internal variability is not homogeneous in the input parameter space, as it increases as the friction velocity decreases. This is because as advection decreases, the temporal correlation of concentration increases, which increases the uncertainty of the mean over the 200-s analysis period (less independent information to estimate the mean). We also note that the noise decreases as α_{inlet} moves away from the median value of $-30\,^\circ$, due to a zoning bias: the plume moves further outside the domain at the boundary angles (Eq. 4), and there is therefore a larger proportion of the domain where the concentration is zero at these angles.

This quantification of the noise in the learning dataset is of paramount importance for the construction and validation of surrogate models. In particular, this information can be used to select the dimension of the latent space to prevent the surrogate model from overfitting the noise associated with internal variability (Sect. 3.4). Internal variability estimates can also be used as a reference to check that the surrogate model uncertainty is not underestimated (Sect. 3.3), and as a performance target for the surrogate model

(Sect. 3.5).

3 Surrogate modeling approach

This section presents the POD–GPR surrogate modeling approach and specifies the inputs/outputs and metrics used for validation. The focus is on two points. The first point is how to estimate the uncertainty associated with POD–GPR predictions and relate it to internal variability. The second point is how to make an informed choice about the surrogate latent space dimension.

3.1 Problem statement

The goal of the surrogate model is to emulate as closely as possible the response surface of the LES model (Sect. 2.2) with respect to the input parameters $\boldsymbol{\theta} = (\alpha_{inlet}, u_*)$ defined over the space $\Omega_{\boldsymbol{\theta}}$ (Eq. 4, Sect. 2.3). This means finding a function:

$$\begin{aligned} \mathcal{M}_{\text{surrogate}} : \Omega_{\boldsymbol{\theta}} &\longrightarrow \mathbb{R}^N, \\ \boldsymbol{\theta} &\longmapsto \mathbf{y}_{\text{surrogate}}, \end{aligned} \tag{5}$$

that minimizes $\displaystyle\int_{\Omega_{\boldsymbol{\theta}}} \|\mathbf{y}_{\text{surrogate}}(\boldsymbol{\theta}) - \mathbf{y}_{\text{LES}}(\boldsymbol{\theta})\| \, \mathrm{d}\boldsymbol{\theta}$, where $\mathbf{y}_{\text{LES}} \in \mathbb{R}^N$ is the field to be emulated, discretized on a grid of N nodes, and where $\mathbf{y}_{\text{surrogate}}$ is its counterpart predicted by the surrogate. This function is obtained here by learning from the train dataset $\left\{\left(\boldsymbol{\theta}^{(i)},\ \mathbf{y}_{\text{LES}}(\boldsymbol{\theta}^{(i)})\right)\right\}_{i=1}^{N_{train}}$ with $N_{train} = 160$ (80% of the full LES dataset, see Fig. 1).

In this study, we focus on the emulation of the mean tracer concentration fields, which are noisy due to the internal variability of the ABL (Sect. 2.5). Taking into account this aleatory uncertainty in the construction and validation of the surrogate model is a key challenge we address here.

To reduce the computational cost associated with the high dimension N of the solver grid on which the fields of interest are expressed, we interpolate all the fields on an analysis mesh twice as coarse, centered around the container array, and with a height limited to 20 m as most of the tracer is located in this area. This leads to an analysis mesh of

$N = 1.88 \times 10^6$ nodes, with characteristic cell sizes ranging from 0.6 m to 4 m, which facilitates efficient model reduction. We have checked that using a coarser-resolution mesh has a negligible effect on the surrogate model accuracy (not shown here).

3.2 The POD–GPR surrogate model

3.2.1 Principle

We choose to use a POD–GPR surrogate model because it has proven to be efficient, relatively inexpensive and robust (Marrel et al. 2015; Guo and Hesthaven 2018; Nony et al. 2023). The fundamental principle of the POD–GPR approach is to combine:

i) a *reduction step* using Proper Orthogonal Decomposition (POD) (Sirovich 1987; Berkooz et al. 1993), which is very popular in fluid mechanics (Chinesta et al. 2011; Taira et al. 2017; Vinuesa and Brunton 2022) and consists in finding a low-dimensional space, called *latent space*, of dimension $L \ll N$, on which the fields to be emulated $\mathbf{y}(\boldsymbol{\theta})$ are projected;

ii) and a *regression step* using standard Gaussian Process Regression (GPR) (Rasmussen et al. 2006), which consists in learning from the train set, the relationship between the LES model input parameters $\boldsymbol{\theta}$ and the latent coefficients $\left\{k_\ell(\boldsymbol{\theta})\right\}_{\ell=1}^{L}$ resulting from the field projection onto the latent space.

This reduction-regression approach allows i) to reduce the dimension of the regression problem to L latent variables ($L \ll N$) and thereby drastically reduce the computational burden of the learning task; and ii) to separate the parametric dependence of the field from the spatial variability.

The POD–GPR model is implemented as a standard statistical learning approach, i.e. with an initial training phase consisting of i) preprocessing the LES fields, ii) building the POD reduced basis based on the train set, and iii) optimizing the GPR models in the latent space (Fig. 4a). This training phase is done offline and only once. The trained POD–GPR can then provide online field predictions for new inputs $\boldsymbol{\theta}$ as follows: i) the associated POD reduced coefficients are predicted by the fitted GPR models, and ii) the

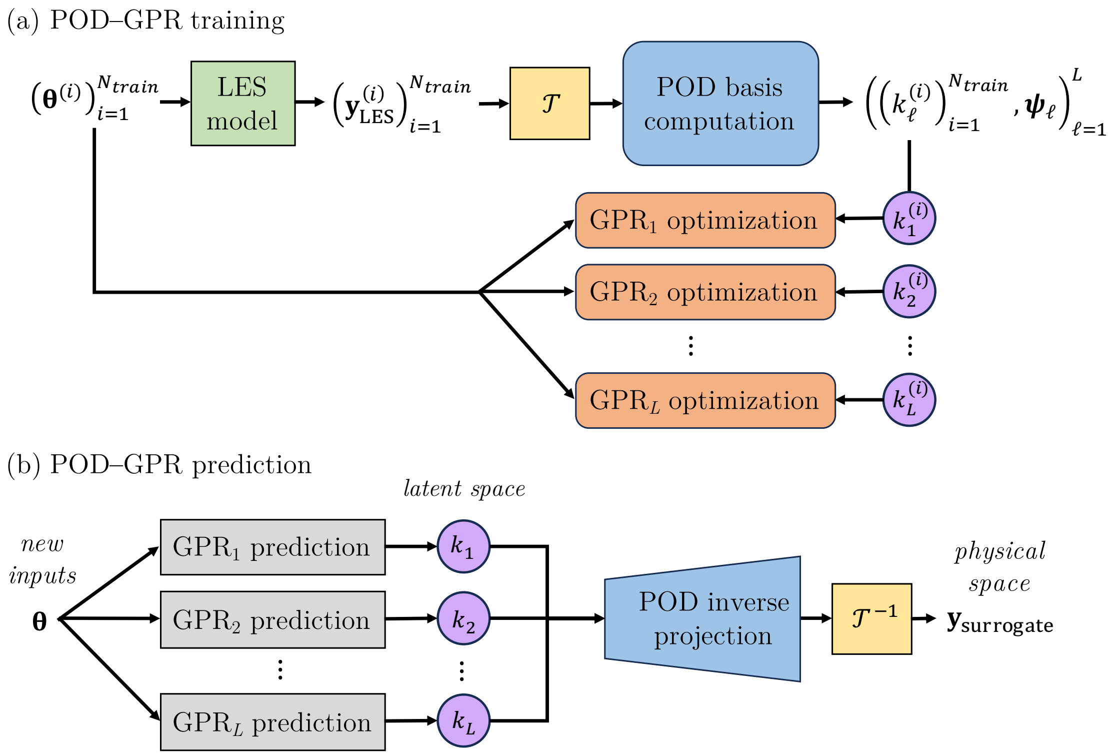

Figure 4: *Schematic of the POD–GPR surrogate model. Its operation is divided into two stages: the training phase (a), and the prediction phase (b). For the training phase, first, a preprocessing $\mathcal{T}$ is applied to the LES predicted fields, and the POD reduced basis $(\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_L)$ is built on the scaled train set; then L independent GPR models are optimized to emulate the L POD reduced coefficients $(k_1, ..., k_L)$ for the input parameters $\boldsymbol{\theta}$. For the prediction phase, the fitted GPR models predict the POD reduced coefficients associated with a given set of wind conditions $\boldsymbol{\theta}$, then the inverse POD projection and inverse scaling $\mathcal{T}^{-1}$ are applied to recover the associated physical field.*

inverse POD projection and inverse fields scaling are applied to these coefficients to recover the physical field $\mathbf{y}_{\text{surrogate}}$ (Fig. 4b). The following sections present the theoretical elements of the POD and GPR techniques required for this study.

3.2.2 Field preprocessing and dimension reduction using POD

With POD, the fields are projected linearly into the latent space generated by the L eigenvectors $\{\boldsymbol{\psi}_\ell\}_{\ell=1}^{L}$ of the train set covariance matrix associated with the L largest eigenvalues $\{\Lambda_\ell\}_{\ell=1}^{L}$. These eigenvalues are the most informative about the coherent spatial structures emerging from variations in the wind conditions $\boldsymbol{\theta} = (\alpha_{inlet},\ u_*)$. The question of how to choose L is discussed further in Sect. 3.4.

The projection of one field $\mathcal{T}(\mathbf{y})$ onto the POD latent space can be formulated as

$$\mathcal{T}(\mathbf{y}) \approx \sum_{\ell=1}^{L} \sqrt{\Lambda_\ell} k_\ell \boldsymbol{\psi}_\ell, \tag{6}$$

where $\mathcal{T}$ is a field preprocessing treatment including centering, and $\{k_\ell\}_{\ell=1}^{L}$ are the POD reduced coefficients defined as the coefficients in the projection of the given field $\mathbf{y}(\boldsymbol{\theta})$ normalized by $\sqrt{\Lambda_\ell}$. This scaling, called POD whitening (Kessy et al. 2018), ensures that the set of reduced coefficients $\{k_\ell\}_{\ell=1}^{L}$ is centered and has unit component-wise variances on average, so that the regression problem is well posed for GPR.

The orthogonality of POD modes leads to some very useful properties (Berkooz et al. 1993; Cordier and Bergmann 2006): i) the POD decomposition (Eq. 6) is the linear combination that reproduces the most variance of the original set, and ii) POD reduced coefficients are uncorrelated, i.e. $\mathrm{Cov}(k_i, k_j) = 0$, if $i \neq j$, which justifies why we build one GPR model per mode (Fig. 4).

For pollutant dispersion applications, a particular difficulty arises from the wide disparity of the concentration scale, which significantly limits POD approximation accuracy. This can be addressed by preprocessing the fields before building the POD, as this changes the meaning of the optimality and orthogonality properties of the POD modes (Schmidt and Colonius 2020), and thus conditions the POD ability to efficiently represent fields in a smaller dimension. Using a logarithmic preprocessing, which is a natural choice for concentrations that decrease exponentially with distance from the source, results in better overall projection performance for the MUST case study (not shown here – see Lumet (2024), Chapter IV, for further discussion on preprocessing strategies). This logarithmic preprocessing reads:

$$\begin{aligned} \mathcal{T} : \mathbb{R}^N &\longrightarrow \mathbb{R}^N, \\ \mathbf{y}(\mathbf{x}_k) &\longmapsto \sqrt{\frac{\omega(\mathbf{x}_k)}{\Omega}} \left[\ln(\mathbf{y}(\mathbf{x}_k) + y_t) - \langle \ln(\mathbf{y}_{\mathrm{LES}}(\mathbf{x}_k) + y_t) \rangle \right], \ 1 \leq k \leq N, \end{aligned} \tag{7}$$

where $\dfrac{\omega(\mathbf{x}_k)}{\Omega}$ is the relative volume of the node $\mathbf{x}_k$, and y_t is a threshold set to 10^{-4} ppm to avoid issues with concentration values close to zero. This choice provides an effective compromise that does not over-cut low concentrations and does not over-emphasize very low variances, which are mainly numerical noise. Note that this preprocessing also includes the centering required for POD (Berkooz et al. 1993), and volume node weighting to avoid over-weighting refined locations (Schmidt and Colonius 2020).

3.2.3 Latent coefficients estimation by Gaussian processes

Once the POD latent space is constructed, the next step is to predict the POD reduced coefficients $\{k_\ell(\boldsymbol{\theta})\}_{\ell=1}^{L}$ for any new wind conditions $\boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}$ (Fig. 4b). Since POD coefficients are uncorrelated, we simplify this vector regression problem into L scalar regression problems solved by GPR (Rasmussen et al. 2006). There are three main reasons for this choice: i) simple interpolation may fail to predict latent space components (Brunton and Kutz 2019); ii) GPR was found to be one of the best machine learning regression methods for predicting POD-reduced coefficients of LES concentration fields (Nony 2023); and iii) GPR models predict probability distributions and not just pointwise estimates, which is in line with our objective to quantify surrogate model uncertainties.

The principle of Gaussian processes (GP) is that the data distribution can be described by a Gaussian stochastic process, implying

$$k_\ell = f_\ell(\boldsymbol{\theta}) + \epsilon_\ell \text{ with } \begin{cases} f_\ell(\boldsymbol{\theta}) \sim \mathcal{GP}(\mathbf{0},\ r_\ell(\boldsymbol{\theta}, \boldsymbol{\theta}^*)),\ \forall(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \in \Omega_{\boldsymbol{\theta}}^2 \\ \epsilon_\ell \sim \mathcal{N}(0, s_\ell^2) \end{cases}, \tag{8}$$

where r_ℓ is the GP covariance function, or *kernel*, and where ϵ_ℓ is an additive Gaussian noise with variance s_ℓ^2 accounting for the fact that the k_ℓ are subject to an irreducible noise due to the internal variability of the mean concentration (Fig. 3). Note that we assume that the prior distribution of the GP is zero on average since POD reduced coefficients are centered on average.

Given the property that any finite subset of realizations of a GP follows a multivariate Gaussian distribution, the posterior probability distribution of the quantity of interest

$k_\ell^*(\boldsymbol{\theta}^*)$ knowing the training set $\{\boldsymbol{\theta}^{train}, \mathbf{K}_\ell^{train}\}$ is

$$k_\ell^*(\boldsymbol{\theta}^*)\Big|_{\{\boldsymbol{\theta}^{train}, \mathbf{K}_\ell^{train}\}} \sim \mathcal{N}\left(\mu_\ell,\ \sigma^2_{\mathrm{GP}}(k_\ell^*)\right), \tag{9}$$

with:

$$\begin{cases} \mu_\ell = r_\ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{train}) \left[r_\ell(\boldsymbol{\theta}^{train}, \boldsymbol{\theta}^{train}) + s_\ell^2 \mathbf{I}\right]^{-1} \mathbf{K}_\ell^{train}, & \text{(10a)} \\ \sigma^2_{\mathrm{GP}}(k_\ell^*) = r_\ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) + s_\ell^2 - r_\ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{train}) \left[r_\ell(\boldsymbol{\theta}^{train}, \boldsymbol{\theta}^{train}) + s_\ell^2 \mathbf{I}\right]^{-1} r_\ell(\boldsymbol{\theta}^{train}, \boldsymbol{\theta}^*). & \text{(10b)} \end{cases}$$

In the regression context, these equations give the mean GPR prediction (Eq. 10a) and the associated variance (Eq. 10b), which quantifies two forms of uncertainty: i) the uncertainty linked to the noise in the training data and related to the term $s_\ell^2\mathbf{I}$, and ii) the regression uncertainty that depends on the distance between the new input parameters $\boldsymbol{\theta}^*$ and the training parameters $\boldsymbol{\theta}^{train}$. Both equations involve the kernel r_ℓ, which is here of standard Matérn type with hyperparameter $\nu = 5/2$ (Stein 1999).

In the end, each GP has four hyperparameters: the noise variance s_ℓ^2, and three parameters involved in the Matérn kernel (Stein 1999): the maximum allowable covariance, and the length scale associated with each of the two uncertain parameters. These parameters are determined by maximum log-likelihood estimation (Hastie et al. 2009) during GP optimization (Fig. 4a).

3.3 Uncertainty estimation of POD–GPR predictions

Below we explain how the GPR estimated uncertainty (Eq. 10b) is propagated from latent space to physical space through the POD inverse projection. This is useful to quantify the uncertainty of POD–GPR field predictions.

POD–GPR predictions are defined as linear combinations of the POD reduced coefficients $k_\ell(\boldsymbol{\theta})$ (Eq. 6), which are uncorrelated (by POD modeling assumption) and normally distributed (Eq. 9). Consequently, at each grid node $\mathbf{x}_k$, the variance of the POD–GPR prediction $\mathcal{T}(\mathbf{y}(\boldsymbol{\theta}, \mathbf{x}_k))$ also follows a normal distribution:

$$\sigma^2_{\mathrm{POD-GPR}}\left(\mathcal{T}(\mathbf{y}(\boldsymbol{\theta}, \mathbf{x}_k))\right) = \sum_{\ell=1}^{L} \varLambda_\ell \, \sigma^2_{\mathrm{GP}}(k_\ell(\boldsymbol{\theta})) \, \boldsymbol{\psi}_\ell(\mathbf{x}_k)^2, \tag{11}$$

with $\sigma^2_{\mathrm{GP}}(k_\ell(\boldsymbol{\theta}))$ the ℓth GP variance (Eq. 10b).

Using log-preprocessing (Eq. 7), we deduce that the variance of the re-scaled mean concentration prediction $\mathbf{y}(\boldsymbol{\theta}, \mathbf{x}_k)$ follows a log-normal distribution:

$$\sigma^2_{\mathrm{POD-GPR}}(\mathbf{y}(\boldsymbol{\theta}, \mathbf{x}_k)) = \Big[\exp\left(s(\boldsymbol{\theta}, \mathbf{x}_k)^2\right) - 1\Big] \times \exp\left(2\, m(\boldsymbol{\theta}, \mathbf{x}_k) + s(\boldsymbol{\theta}, \mathbf{x}_k)^2\right), \tag{12}$$

where:

$$\begin{cases} m(\boldsymbol{\theta}, \mathbf{x}_k) = \sqrt{\dfrac{\varOmega}{\omega(\mathbf{x}_k)}} \displaystyle\sum_{\ell=1}^{L} \sqrt{\varLambda_\ell}\, k_\ell \, \boldsymbol{\psi}_\ell(\mathbf{x}_k) + \langle \ln(\mathbf{y}_{\mathrm{LES}} + y_t) \rangle, & \text{(13a)} \\ s(\boldsymbol{\theta}, \mathbf{x}_k)^2 = \left(\dfrac{\varOmega}{\omega(\mathbf{x}_k)}\right) \displaystyle\sum_{\ell=1}^{L} \varLambda_\ell \, \sigma^2_{\mathrm{GP}}(k_\ell(\boldsymbol{\theta})) \, \boldsymbol{\psi}_\ell(\mathbf{x}_k)^2. & \text{(13b)} \end{cases}$$

Equation 12 provides an estimate of the uncertainty associated with POD–GPR predictions. This uncertainty is the sum of the GP variances $\sigma^2_{\mathrm{GP}}(k_\ell(\boldsymbol{\theta}))$, which quantify the noise error in the training data and the regression error for each mode. In this context, these two forms of error therefore correspond to the uncertainty associated with the LES internal variability (Sect. 2.5) and to part of the structural error associated with model reduction. It is worth noting that this estimate does not include the error associated with the projection into the POD latent space.

3.4 A priori choice of latent space dimension

The choice of the POD latent space dimension is case-dependent and has a critical effect on the accuracy of the surrogate model. On the one hand, the higher the number of POD modes, the more variance of the original ensemble is captured in the POD reduced basis. On the other hand, high-order modes are likely to encode noise in the train set (Forkman et al. 2019), and are therefore best set aside to prevent GP from overfitting noise during learning. In this section, we present an innovative method to select L as a trade-off between the total variance embedded in the POD reduced basis and the amount of noise carried by the POD modes.

POD projection error First, we evaluate the POD projection error, i.e. the error obtained after reconstructing the fields projected onto the POD latent space through inverse POD transformation, for varying number of modes L following the approach adopted by Nony et al. (2023). Figure 5a shows that the POD projection normalized mean square error (NMSE) quickly decreases with the number of modes, and that a small number of modes ($L \approx 5-10$) allows to obtain very fine NMSE scores. We verify that the eigenvalues Λ_ℓ are a good proxy for quantifying the amount of information retrieved by each POD mode (Berkooz et al. 1993) and can therefore be used to select L as done by Xiao et al. (2019).

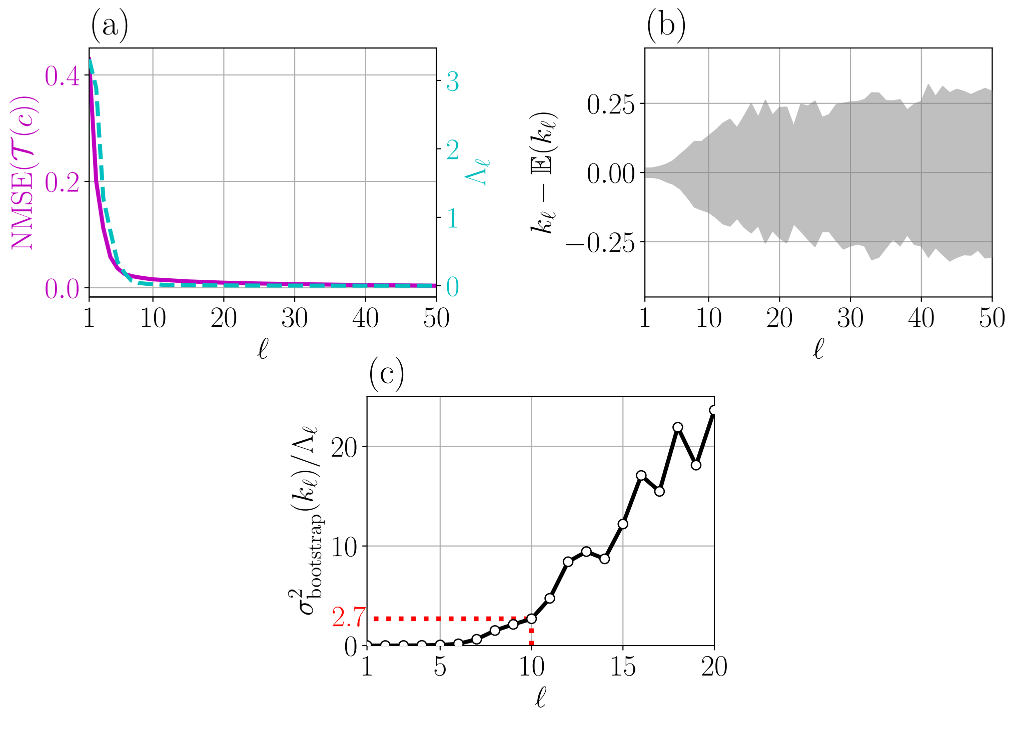

Figure 5: *(a) POD projection error evaluated over the train set with NMSE (Eq. 15) as function of the number of modes retained, and POD eigenvalues Λ_ℓ associated with each mode ℓ. (b) Spread of the difference between POD reduced coefficients k_ℓ replicates and their mean $\mathbb{E}(k_\ell)$. The spread is defined by the 2.5th and 97.5th percentiles and is averaged over the train set. (c) Ratio between the averaged variance of the reduced coefficient bootstrap replicates $\sigma^2_{\text{bootstrap}}(k_\ell)$ and the POD eigenvalue Λ_ℓ associated with each mode ℓ. The red dotted line indicates the number of modes selected for this study.*

Internal variability in the POD latent space To quantify how the noise caused by internal variability is captured by each POD mode, we project the bootstrap replicates of the LES fields onto the POD latent space constructed with the original fields (Sect. 2.5). We thereby obtain $1,000$ realizations of the POD reduced coefficients k_ℓ associated with each field in the dataset. Figure 5b shows that the spread of the reduced coefficient replicates increases significantly when going to higher order modes (i.e. for increasing ℓ). In particular, for $\ell \leq 5$, the spread of the bootstrap replicates of the reduced coefficients remain small ($< 10\,\%$), implying that these modes correspond to systematic patterns associated with the plume structure and its dependence on the wind conditions. The variability of the POD reduced coefficients then increases rapidly before reaching a plateau from $\ell \geq 15$. At this plateau, the spread of the k_ℓ replicates varies by about $\pm 25\%$. This implies that field features linked to internal variability are mainly captured by higher order modes, which is consistent with the literature (Forkman et al. 2019). This in turn implies that we need to limit the number of modes L to avoid introducing noise into the POD-GPR surrogate model.

A priori criterion to choose the POD latent space dimension Based on these findings, we propose to measure the ratio between the internal variability noise and the fraction of the total ensemble variance represented by each mode defined as

$$\frac{\sigma^2_{\text{bootstrap}}(k_\ell)}{\Lambda_\ell}, \tag{14}$$

where $\sigma^2_{\text{bootstrap}}(k_\ell)$ is the variance of the POD reduced coefficients replicates averaged over the train set, and where Λ_ℓ is the ℓth eigenvalue in the POD decomposition.

The ratio in Eq. 14 is shown in Fig. 5c and provides a way to choose the latent space dimension L that minimizes both the noise and the POD projection error, and it has the advantage of being completely a priori as it does not require either the test set or the evaluation of the full POD–GPR model. Results show that this ratio is close to zero for the first six modes and then increases sharply with mode order. We therefore choose to truncate the POD decomposition before the inflection point using $L = 10$ modes to project the mean concentration fields. This approach for selecting the latent space dimension is evaluated a posteriori in Sect 4.3.

3.5 Surrogate validation methodology

We present now the metrics used to quantify the surrogate model reduction error, before estimating the best values achievable for each metric given the internal variability.

3.5.1 Quantification of the surrogate error

The POD–GPR model accuracy is estimated on a set of independent test samples ($N_{test} = 40$, corresponding to 20% of the full LES dataset, see Fig. 1). This is essential to assess the ability of the model to generalize information from the train set to new meteorological forcing parameters ($\alpha_{inlet},\ u_*$).

To assess the surrogate error, we use standard air quality metrics from Chang and Hanna (2004). These metrics compare the mean concentration field predicted by the surrogate model $\mathbf{c}_{\text{surrogate}}$ with the LES counterpart $\mathbf{c}_{\text{LES}}$ in terms of normalized mean square error (NMSE), fraction of predictions within a factor of two of observations (FAC2), geometric variance (VG), and figure of merit in space (FMS):

$$\text{NMSE} = \frac{\langle\, (\mathbf{c}_{\text{LES}} - \mathbf{c}_{\text{surrogate}})^2\,\rangle}{\langle\, \mathbf{c}_{\text{LES}}\,\rangle\,\langle\, \mathbf{c}_{\text{surrogate}}\,\rangle}, \tag{15}$$

$$\text{FAC2} = \langle\, \xi\,\rangle \text{ with } \xi(\mathbf{x}_k) = \begin{cases} 1 \text{ if } 0.5 \leq \mathbf{c}_{\text{surrogate}}(\mathbf{x}_k)\,/\,\mathbf{c}_{\text{LES}}(\mathbf{x}_k) \leq 2, \\ 1 \text{ if } \mathbf{c}_{\text{surrogate}}(\mathbf{x}_k) \leq c_t \text{ and } \mathbf{c}_{\text{LES}}(\mathbf{x}_k) \leq c_t, \\ 0 \text{ else,} \end{cases} \tag{16}$$

$$\text{VG} = \exp\left(\langle\, (\ln \widetilde{\mathbf{c}}_{\text{LES}} - \ln \widetilde{\mathbf{c}}_{\text{surrogate}})^2\,\rangle\right), \tag{17}$$

$$\text{FMS}(c_\ell) = \frac{\Omega_\cap(c_\ell)}{\Omega_\cup(c_\ell)}, \tag{18}$$

where $\langle\cdot\rangle$ denotes spatial averaging weighted by the dual volume of the node $\mathbf{x}_k$, c_t is a concentration threshold defining $\widetilde{\mathbf{c}} = \max(\mathbf{c},\, c_t)$, as suggested by Chang and Hanna

(2004) and Schatzmann et al. (2010) to avoid issues with values close to zero in FAC2 and VG metrics. In this study, we use $c_t = 10^{-4}$ ppm, considering that errors on lower concentrations are mainly due to numerical noise. Finally, $\Omega_{\cap}(c_\ell)$ denotes the volume, in m^3, of the domain in which both $\mathbf{c}_{\text{surrogate}}$ and $\mathbf{c}_{\text{LES}}$ are over a user-specified tracer value c_ℓ. Conversely, $\Omega_{\cup}(c_\ell)$ denotes the volume where $\mathbf{c}_{\text{surrogate}} \geq c_\ell$ or $\mathbf{c}_{\text{LES}} \geq c_\ell$.

The use of different metrics than the loss used during training is important because of the multi-order nature of the concentration field. NMSE is more sensitive to errors at high concentrations, while VG assesses prediction accuracy at low concentrations. FMS quantifies how close the two plume shapes are relative to a given concentration level. The scores that a perfect model would obtain are reported in Table 1.

3.5.2 Estimation of the internal variability

LES data are noisy due to internal variability (Sect 2.5). It would therefore be pointless to try to build a surrogate model whose accuracy exceeds this uncertainty. To quantify the error due to internal variability alone, we use the bootstrap approach proposed in Lumet et al. (2024b) to generate two independent sets of bootstrap replicates of the same LES field. We then compute the average difference between each pair of replicates using the metrics introduced in Sect. 3.5.1. For each metric, we obtain the amount of error due to internal variability only, which is the expected error when comparing two independent realizations of the mean concentration fields for the same input parameters.

This is done for every LES sample in the dataset, and the ensemble-averaged internal variability errors give an upper bound estimate of the best overall accuracy achievable for each metric when validating the POD–GPR surrogate model.

4 Surrogate model validation

In this section, we present a thorough evaluation of the POD–GPR surrogate model. We first assess its accuracy over the test set and its efficiency (Sect. 4.1). We then validate the innovative aspects of our approach: the POD–GPR uncertainty estimation (Sect. 4.2), and the selection of the number of POD modes (Sect. 4.3). Finally, we study how the

POD–GPR model behaves when reducing the train set (Sect. 4.4). All results are given for the mean concentration field, but the POD–GPR approach can be applied to any LES field (Lumet (2024), Appendix B).

4.1 Evaluation of the surrogate model field predictions

We evaluate here the POD–GPR predictions of mean concentration following the methodology introduced in Sect. 3.5, using the mean internal variability error as the reference for validation. We use a latent space dimension of $L = 10$ in accordance with the informed choice made in Sect. 3.4.

Prediction accuracy The overall performance of the surrogate model is quantified using standard air quality metrics (Sect. 3.5.1). Table 1 shows the obtained scores averaged over the test set. Overall, the POD–GPR model yields very satisfactory results, with most scores close to the error due to internal variability only, which is the best achievable accuracy. However, the results for NMSE and FMS(1 ppm) remain relatively far from the internal variability error, indicating that POD–GPR is less good at predicting high concentration values.

Table 1: *Prediction accuracy of the POD–GPR surrogate model evaluated using the metrics defined in Sect. 3.5.1 and averaged over the test set. The standard deviations of the scores over the test set are also given, as well as the individual scores for test samples #81 and #187, which represent the lowest and highest FAC2 scores achieved by the POD–GPR, respectively. For comparison, the perfect scores for the metrics, the mean error due to internal variability only (Sect. 3.5.2) and the mean error due to standalone reduction dimension are given.*

	FAC2	NMSE	VG	FMS (1 ppm)	FMS (0.01 ppm)
Perfect score	1	0	0	1	1
Internal variability	0.95	1.80	1.39	0.83	0.93
POD projection error	0.91	20.4	1.33	0.75	0.93
POD–GPR prediction error	**0.91**	**20.6**	**1.39**	**0.75**	**0.92**
Standard deviation	0.04	43.2	0.68	0.11	0.03
Test sample #81	0.74	23.4	5.25	0.79	0.85
Test sample #187	0.96	8.08	1.07	0.86	0.94

Table 1 also shows that the POD–GPR prediction errors are almost identical to the standalone POD projection errors (i.e. errors obtained by simply reconstructing the test fields after projection onto the POD basis by inverse POD transformation). This implies that the accuracy of the POD–GPR model is mostly limited by the accuracy of the POD and not by the GPR. In addition, the poor prediction performance for high concentrations can be explained by the fact that the POD is not well adapted to the multiscale and nonlinear nature of the concentration fields. In particular, the use of a logarithmic preprocessing before the POD degrades the reconstruction of high concentrations in the vicinity of the emission source, but has the advantage of preserving the other metrics and in particular the shape of the plume compared to linear processing (Lumet (2024), Chapter IV).

There is quite a large spread of POD–GPR errors across the test samples, especially for the quadratic metrics NMSE and VG, indicating the presence of test sample outliers. This variability over the input parameter space is mainly explained by the fact that as the friction velocity decreases, the internal variability increases (Fig. 3), which makes the mean concentration noisier and therefore more difficult to predict. In addition, FMS(1 ppm), and to a lesser extent FMS(0.01 ppm) and FAC2, are subject to a zoning effect as they depend on the size of the plume within the domain of interest (Eqs. 16, 18). For example, these scores are improved when the wind direction carries the plume outside the container array (i.e. for $\alpha_{inlet} \approx 30\,^{\circ}$or $\alpha_{inlet} \approx -90\,^{\circ}$).

Field prediction examples For a more detailed assessment of the POD–GPR model accuracy, we also examine its predictions in the physical space. Figures 6a, b, c, and d compare 2-D cuts of the mean concentration at $z = 1.6\,\mathrm{m}$ predicted by LES and POD–GPR. Results are given for the test sample #187 $\left(\alpha_{inlet}^{(187)}, u_*^{(187)}\right) = \left(21.8\,^{\circ}, 0.59\,\mathrm{m\,s^{-1}}\right)$ for which POD–GPR obtains the best FAC2 score over the test set, and for the test sample #81 $\left(\alpha_{inlet}^{(81)}, u_*^{(81)}\right) = \left(-27.7\,^{\circ}, 0.08\,\mathrm{m\,s^{-1}}\right)$ associated with the worst FAC2 score. The global scores obtained for these two particular snapshots are summarized in Table 1.

In both cases, the POD–GPR model reproduces well the main features of the LES concentration field, in particular the shape and orientation of the plume. The spatial

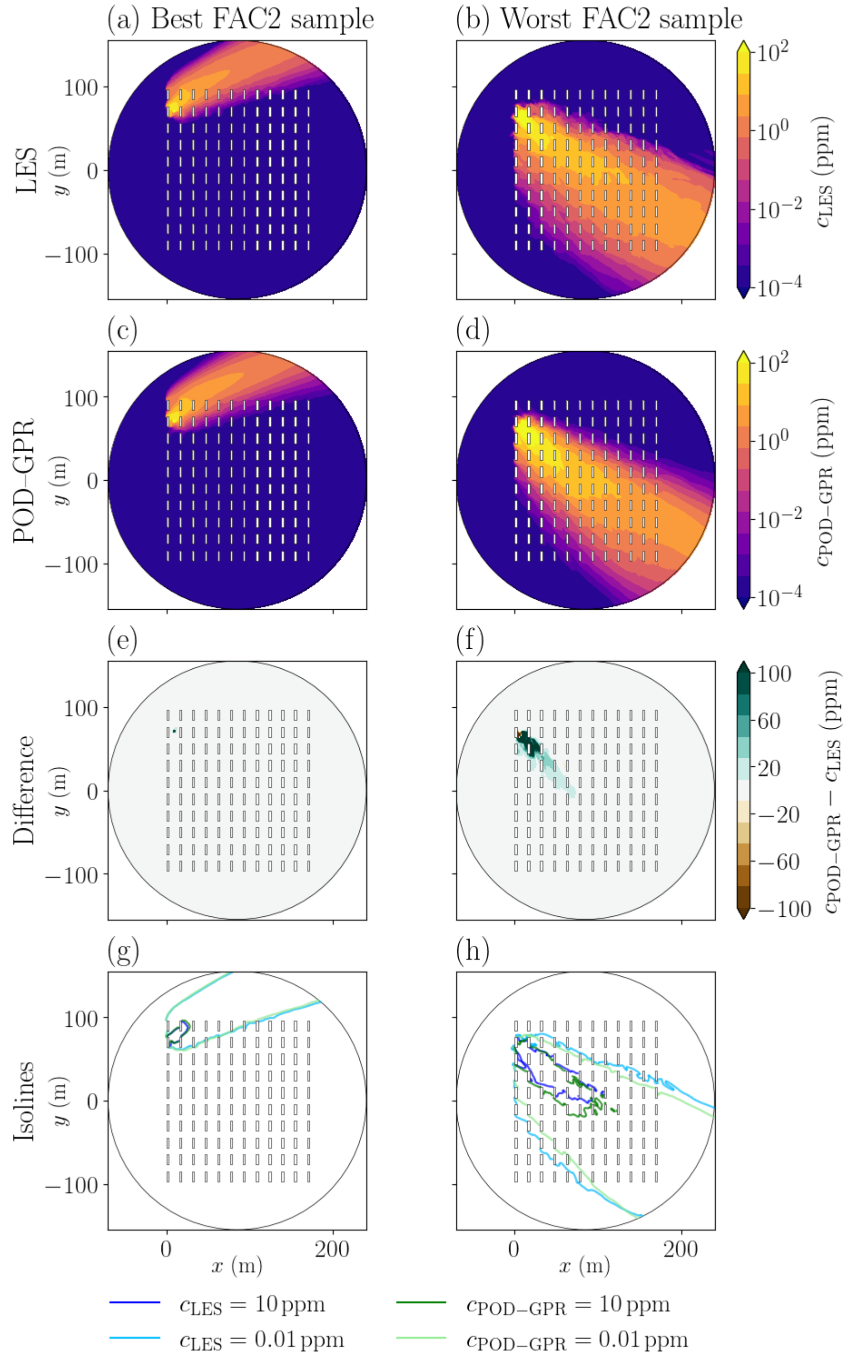

Figure 6: *Horizontal cuts at $z = 1.6\,m$ of two test mean concentration fields estimated by LES (a, b) and POD–GPR (c, d), and the absolute difference between the two (e, f). The left column corresponds to the test sample #187 for which POD–GPR achieves the best FAC2 (Eq. 16) score over the test set, and the right column corresponds to the test sample #81 which is associated with the worst FAC2 score. The LES and POD–GPR predictions of 0.01 ppm and 10 ppm iso concentration levels are shown in (g, h).*

distribution of the different concentration levels is also well reproduced, which is confirmed by the near superposition of the 0.01 ppm and 10 ppm concentration contour lines between LES and POD–GPR (Fig. 6d, h).

However, for the sample with the worst FAC2 (#81), the POD–GPR underestimates the spanwise spread of the plume and significantly overestimates the mean concentration near the emission source (Fig. 6g). This is consistent with the poor NMSE obtained (Table 1) and this is due to the poor reproduction of high concentrations by the POD with logarithmic preprocessing. For this sample (corresponding to a low friction velocity and therefore subject to substantial internal variability) the POD–GPR tends to smooth the irregularities observed at the edges of the plume, thus poorly predicting the local abrupt decrease in concentration.

Efficiency In terms of computational cost, it takes approximately 30 s to train the POD–GPR model using a single core of an Intel Ice Lake CPU. This includes field preprocessing, POD basis decomposition and GPR optimization. This training cost is insignificant compared to the cost of building the training dataset (Sect. 2.4). Once trained, the model provides a prediction of the full 3-D concentration field in about 0.03 s. This approach is therefore compatible with applications requiring a large ensemble of predictions and/or real-time predictions.

4.2 Assessment of the surrogate model uncertainty estimation

We evaluate here the ability of the POD–GPR model to provide realistic uncertainty estimates by comparing them to the actual surrogate error over the test set and to the internal variability present in the LES dataset.

Uncertainty reliability Figure 7a shows the uncertainty reliability diagram comparing actual surrogate error (y-axis) and surrogate model uncertainty estimates (x-axis). The POD–GPR uncertainty is underestimated compared to the actual POD–GPR error for most domain nodes, especially for the lowest concentration values. Nevertheless, the estimated trend is consistent, i.e. the larger the actual error, the larger the prior estimate.

Furthermore, the overall level of precision is satisfactory as the estimated uncertainty is in the right order of magnitude (within the green dashed lines) for 98% of the domain nodes. This is confirmed by the response surface of the POD–GPR (Fig. 12a, b), as the predicted envelopes appear to cover the test samples well. We can therefore be confident in the uncertainty predicted by the POD–GPR surrogate model despite a tendency to underestimate.

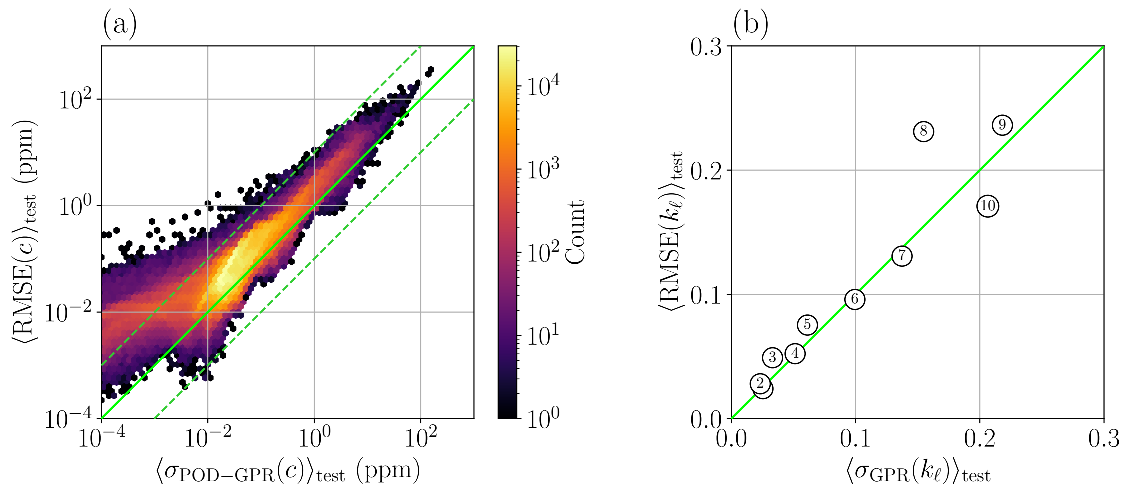

Figure 7: *Reliability diagrams in the physical space and in the latent space: (a) Root mean square error (RMSE) of the POD–GPR concentration prediction over the test set versus the POD–GPR estimated uncertainty at each node where the concentration is larger than the tolerance $c_t = 10^{-4}$ ppm. Each hexagon is colored according to the number of node points in the hexagon. (b) RMSE of the GP prediction of the POD reduced coefficients k_ℓ over the test set versus the GP estimated uncertainty, each mode ℓ is represented by a numbered circle (the POD latent space dimension is $L = 10$). The green solid lines correspond to the identity function, and the dashed lines in (a) show the range of plus or minus one order of magnitude.*

To further investigate the cause of this underestimation, the uncertainty reliability is examined directly in the latent space in Fig. 7b. We find that for the estimation of the reduced POD coefficients by the GPs, the uncertainty estimate is very close to the error made on average, except for the high-order modes 8 and 10. This increase in error for higher-order modes is consistent with the fact that they are more affected by internal variability (Fig. 5b). The following conclusions can be drawn: i) the variance of the GP posterior distribution (Eq. 10b) is realistic, and ii) the underestimation observed in the physical space in Fig. 7a comes from the inverse POD projection. This is consistent with the fact that the POD projection error is not taken into account when estimating the total POD–GPR uncertainty (Sect. 3.3), yet the total POD–GPR error is essentially due to the POD projection error as indicated in Table 1.

Ability to estimate internal variability a posteriori We now examine the nature of the estimated uncertainty in more detail, and assess the proportion due to internal variability. The first point is to study how the noise of the LES fields projected onto the POD latent space is captured by GPR. Figure 8 shows that the values of the GP variance hyperparameters s_ℓ^2 obtained by maximum likelihood estimation are very close to the maximum level of internal variability of the POD reduced coefficients over the train set estimated by bootstrap. This is a strong result because the bootstrap estimates of the internal variability are not used to train the GPs.

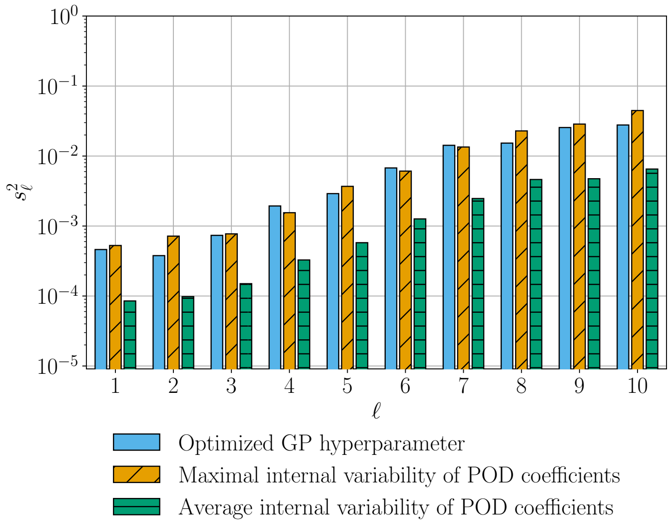

Figure 8: *GP noise variance s_ℓ^2 hyperparameter obtained by log-likelihood maximization for each mode ℓ as blue bars, and maximal (resp. average) noise on the POD reduced coefficients over the train set as orange (resp. green) bars.*

The fact that the GP noise variance parameter matches the maximum level of internal variability (Fig. 8) implies that GPs overestimate the variance of the POD reduced coefficients for most samples where the internal variability is low. This is a structural limitation due to the fact that the GP additive noise does not depend on the input parameter space (Eq. 8), while the variance due to internal variability does (Fig. 3). As a result, in the physical space, the POD–GPR uncertainty predictions tend to be underestimated compared to the actual internal variability for samples where the internal variability is high, while they are overestimated for samples with low internal variability. This could be partially overcome in the future by implementing input-dependent noise variance hyperparameters, as suggested by Miyagusuku et al. (2015).

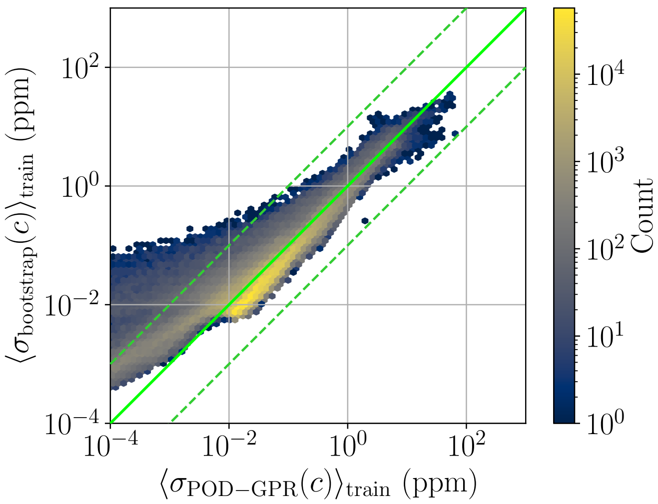

Figure 9: *Internal variability of the mean concentration estimated by bootstrap and averaged over the train set versus the POD–GPR estimated uncertainty at each node where the concentration is larger than the tolerance $c_t = 10^{-4}$ ppm. Each hexagon is colored according to the number of node points in the hexagon. The green solid lines correspond to the identity function and the dashed lines show the range of plus or minus one order of magnitude.*

Figure 9 shows that the uncertainty estimated by the POD–GPR is overall consistent with the LES internal variability over the train set, as the level of variability is in the right order of magnitude for 99% of the domain nodes. For most of the domain, the POD–GPR tends to overestimate the internal variability (hexagonal cells of high density beyond the green line), which is consistent with the GP noise matching the maximum level of internal variability in the latent space (Fig. 8). Note that this analysis is performed over the train set since for theses samples the GPR regression covariance is zero, and thus the POD–GPR uncertainty estimate only corresponds to the estimated internal variability. Finally, we note that the estimated uncertainty envelopes are consistent with the LES internal variability when looking at the POD–GPR response surfaces (Fig. 12a, b).

In this internal variability analysis, the second point is to evaluate the spatial consistency of the POD–GPR uncertainty estimates with respect to the spatial distribu-

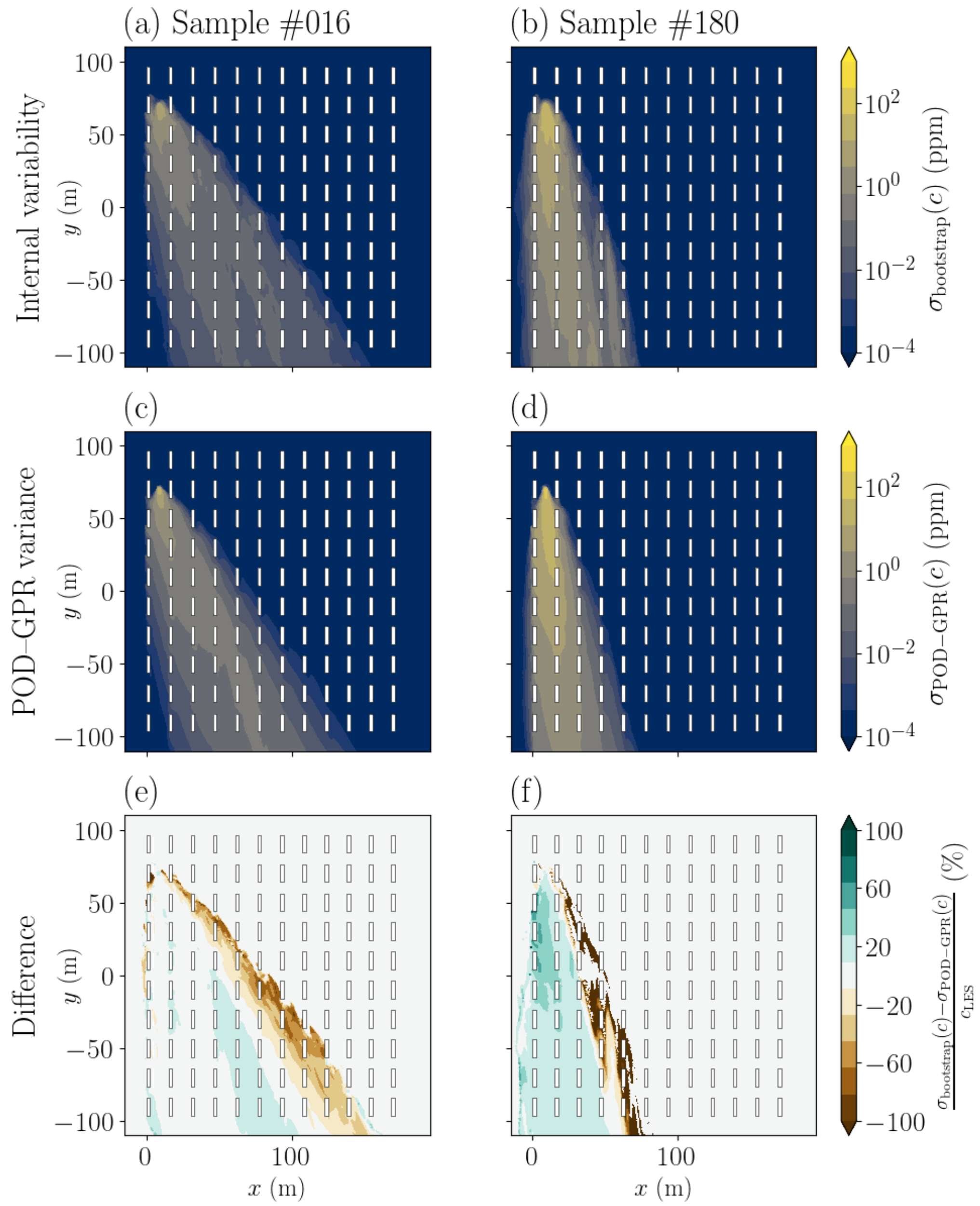

Figure 10: *Horizontal cuts at $z = 1.6\,m$ of the standard deviation of the mean concentration induced by internal variability estimated using bootstrap (a, b), predicted by POD–GPR (c, d), and the relative difference between the two (e, f). The left column corresponds to the train sample #016 and the right column corresponds to the train sample #180*

tion of internal variability to verify that the uncertainty is properly propagated from the POD latent space to the physical space (Sect. 3.3). We find that the variance predicted by the POD–GPR is consistent with the internal variability of the LES in terms of magnitude and structure, as shown in Fig. 10 for the train sample #016

$\left(\alpha_{inlet}^{(016)}, u_*^{(016)}\right) = \left(-79.5\,°, 0.14\,\mathrm{m\,s^{-1}}\right)$ for which the POD–GPR uncertainty estimate is the closest to the internal variability estimated by bootstrap and for the train sample #180 $\left(\alpha_{inlet}^{(180)}, u_*^{(180)}\right) = \left(-58.1\,°, 0.56\,\mathrm{m\,s^{-1}}\right)$, where POD–GPR overestimates the internal variability the most. Despite the overall agreement, the POD–GPR variability estimates appear to be overestimated within the plume and significantly underestimated near the plume edges (Fig. 10e, f), which is consistent with the overall tendency to underestimate low internal variability levels (Fig. 9). This is explained by the fact that there are high concentration gradients near the plume edges and thus high internal variability levels, a feature not well represented by the POD projection, which is based solely on mean concentration and not on its variability.

In summary, the POD–GPR uncertainty estimates derived in Sect. 3.3 i) represent, in a spatially coherent manner, the inherent internal variability of the mean concentration field thanks to the ability of the GPs to accurately infer the level of noise in the train set, and ii) properly explain the actual surrogate errors at predicting the mean concentration. This particularly reinforces the robustness of the POD–GPR and its relevance to uncertainty quantification applications.

4.3 A posteriori validation of the latent space dimension

We revisit our choice of the number of POD modes ($L = 10$) obtained by following the a priori statistical approach we propose in Sect. 3.4. For this purpose, we evaluate the effect of the number of modes L on the performance of the full POD–GPR model on the test set (i.e. by varying L from 5 to 50 in the construction of the POD–GPR model).

Validation metrics Figure 11 shows how the metrics defined in Sect. 3.5.1 change when modifying the POD latent space dimension L. The POD–GPR prediction accuracy over the test set increases with the number of modes and reaches a plateau for a larger number of modes ($L \approx 15$–25) than the NMSE on the train set used in our mode choice approach (Fig. 5). This may indicate that integrating a larger number of modes into the POD–GPR model could lead to improved surrogate model accuracy.

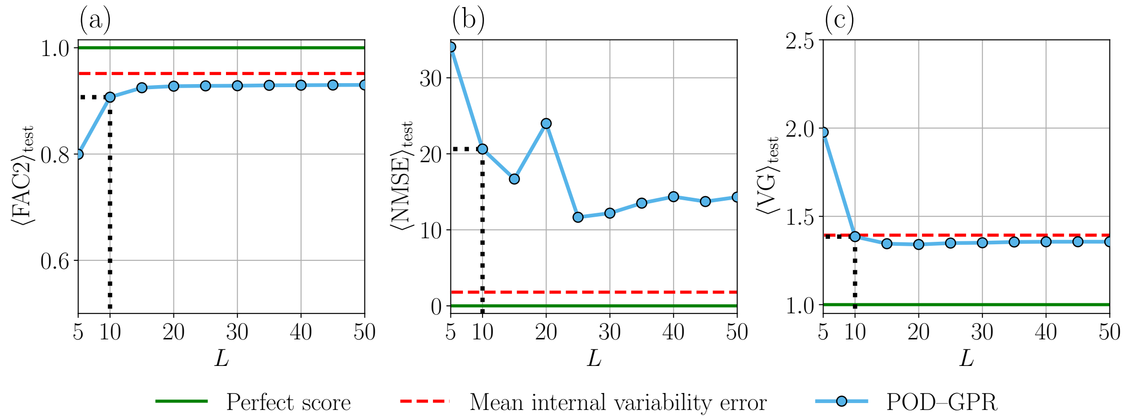

Figure 11: *POD–GPR prediction error as a function of the number of the modes L and evaluated with FAC2 (a), NMSE (b), VG (c) averaged over the test set. Green lines correspond to perfect scores; and red dashed lines correspond to the mean level of error due to internal variability only. Error levels corresponding to the selected number of modes ($L = 10$) are shown as black dotted lines.*

Response surfaces As an additional diagnostic, Fig. 12 shows that using a larger number of modes significantly deteriorates the POD–GPR response surfaces, making them very noisy and implausible as, with $L = 50$ modes (Fig. 12e, f), the POD–GPR model is no longer able to retrieve the inversely proportional dependence of concentration on friction velocity expected from theory and retrieved for the configuration with $L = 10$ modes (Fig. 12a, b). This degradation is due to the fact that high-order modes mostly account for noisy structures due to internal variability (Fig. 5b), and are therefore not informative on systematic structures related to the wind conditions. As a result, when including high-order modes, the GPs attempt to learn unphysical dependence on the input parameters, resulting in the shortwave noise observed in Fig. 12. Still, the increase in uncertainty with the response surface deterioration suggests that the POD–GPR uncertainty estimate is robust. However, the fact that the degradation of the POD–GPR response surface is not seen by the global metrics, which continue to improve as the number of modes increases (Fig. 11), shows that one should not draw conclusions based on scalar metrics alone.

In the light of these tests, our prior selection method for the latent space dimension is convincing. The resulting trade-off of $L = 10$ yields good validation scores, while avoiding the problem of response surface noise. However, we acknowledge that using a slightly larger number of modes ($L \approx 15$–20) would also be appropriate and even slightly improve

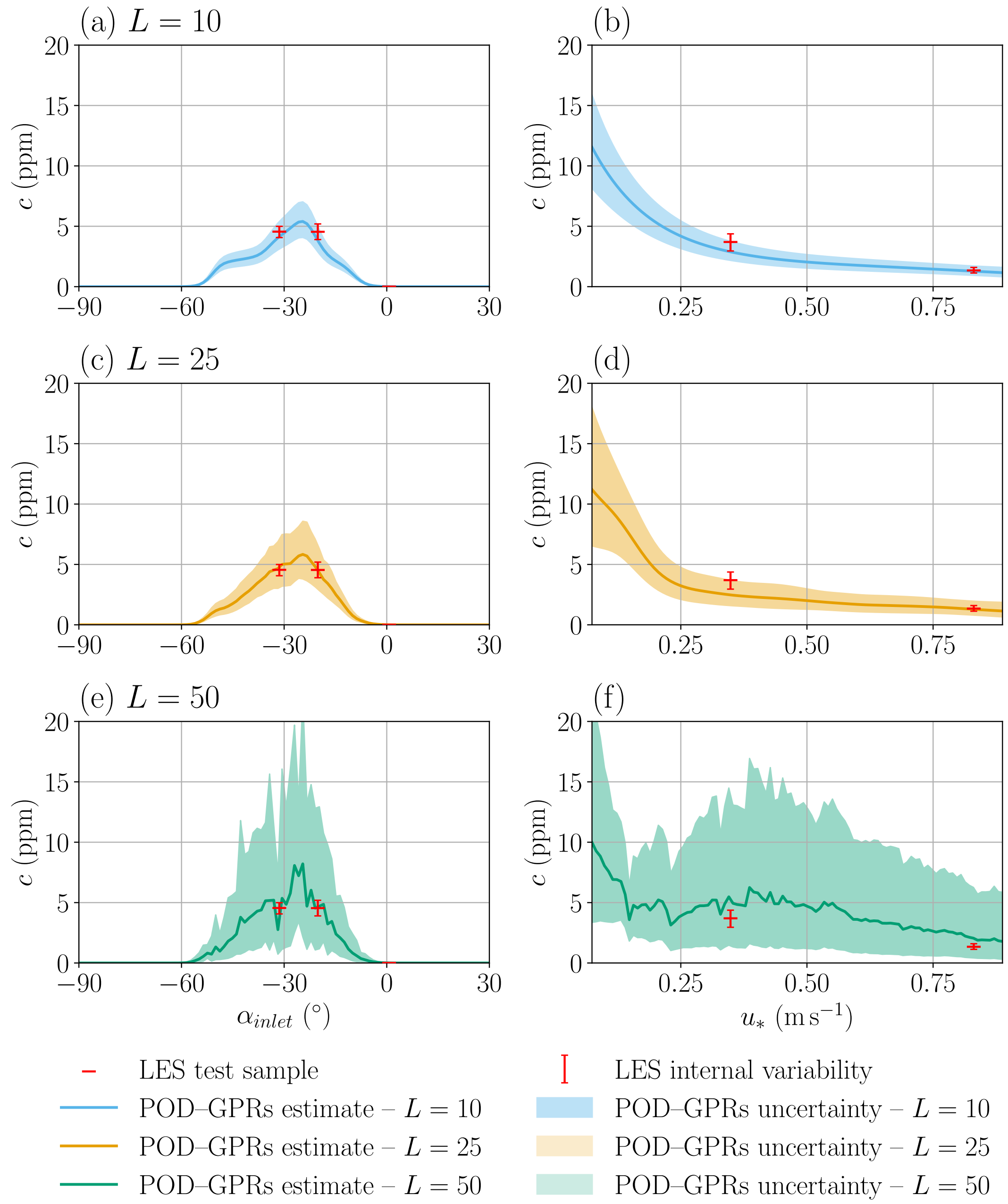

Figure 12: *POD–GPR prediction of the mean concentration at tower B at $z = 2$ m (see tower location in Fig. 2) as a function of the inlet wind direction α_{inlet} (a, c, e), and of the friction velocity u_* (b, d, f). Shaded areas correspond to the 95% confidence intervals estimated by the POD–GPR according to the procedure detailed in Sect. 3.3. Each row corresponds to the results obtained with a different latent space dimension $L \in \{10, 25, 50\}$. When varying one parameter, the other is set constant to either $u_*^{plot} = 0.45\,m\,s^{-1}$ (a, c, e), or $\alpha_{inlet}^{plot} = -43\,°$(b, d, f), and the test samples closest to the two segments of parameter space thus scanned (see Fig. 1) are represented by horizontal red bars. The uncertainty on LES test samples induced by internal variability is depicted as red vertical error bars.*

the surrogate model accuracy. Defining an optimal criterion for latent space dimension selection based on the noise/signal ratio defined in Eq. 14 is therefore an interesting prospect, but requires more validation cases.

4.4 Robustness to training set reduction

In order to assess the potential of the POD–GPR approach for future applications, we examine how the POD–GPR accuracy evolves as the size of the train set decreases (without changing the test set). This is particularly important to investigate the possible trade-offs between the ability of the model to generalize from training data and the substantial cost of building the LES training dataset.

The surrogate model is trained for decreasing train set sizes from $N_{train} = 160$ to $N_{train} = 40$ by keeping only the first samples in Halton's sequence. To make the comparison fair, we systematically evaluate the averaged prediction errors over the same test set of $N_{test} = 40$ samples. Results are shown in Fig. 13a, b, c and d in terms of FAC2, VG, FMS(1 ppm), FMS(0.01 ppm). The decrease in accuracy is fairly constrained and evolves linearly with the train set size, with a loss of 0.08 in FAC2 and 0.12 in VG for every 10 training samples removed. More importantly, the accuracy decreases less rapidly than that of the nearest neighbor model (1–NN), which trivially predicts the mean concentration field as equal to the closest train field in the parameter space (see Appendix). This is especially true for the low concentration values, as the VG score is significantly higher with the 1–NN model than with the POD–GPR model for small train set sizes (Fig. 13b).

Regarding the NMSE metric (Fig. 13e), the evolution with N_{train} is quite chaotic for the POD–GPR and worse than for the 1–NN approach. As previously mentioned, this is related to the high POD projection error near the source when using the logarithmic transformation, and we can consider that the POD–GPR approach with the present preprocessing is not designed to make predictions near the source, regardless of the train set size.

Figure 14 shows that the POD–GPR uncertainty predictions are very robust to train set size reduction. We find that, on average, the POD–GPR uncertainty predictions explain

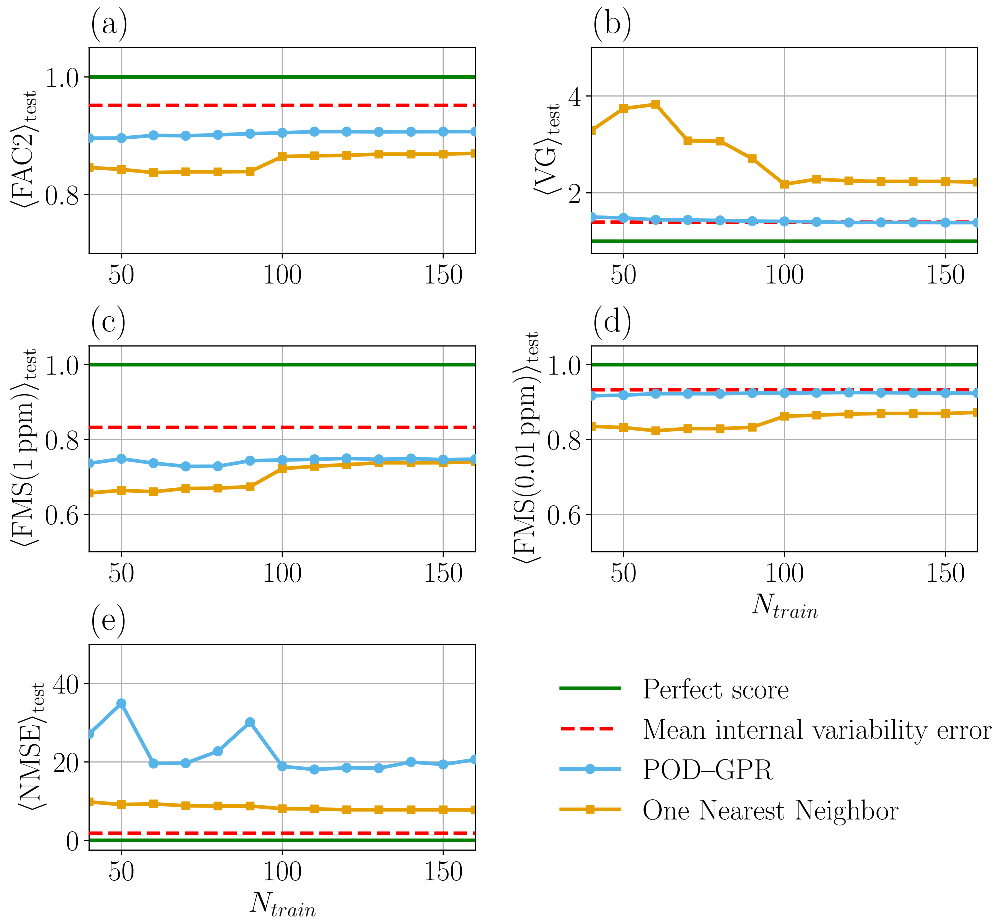

Figure 13: *Surrogate modeling errors for decreasing train set sizes. The mean concentration prediction error is assessed using the metrics defined in Sect. 3.5.1: namely FAC2 (a), VG (b), FMS (c, d), and NMSE (e). Results are given for the POD–GPR as blue circles and the 1–NN model as orange squares. Perfect scores are represented as green lines; and red dashed lines correspond to the mean level of error due to internal variability only.*

overall well its actual error over the test set even with only 40 train samples (Fig. 14a, b, c). Similarly, the ability of the POD–GPR to represent the internal variability of the mean concentration is well preserved (Fig. 14d, e, f), although we note a tendency to underestimate it when the train set size is reduced, as there are fewer close neighboring points for the GPs to estimate the noise in this case.

In summary, the ability of the POD–GPR model to generalize from a training dataset of limited size is better than for the 1–NN baseline approach, justifying the use of such a more sophisticated surrogate model. We find that for this problem, 40 LES training samples are sufficient to achieve good levels of accuracy for most metrics. Furthermore,

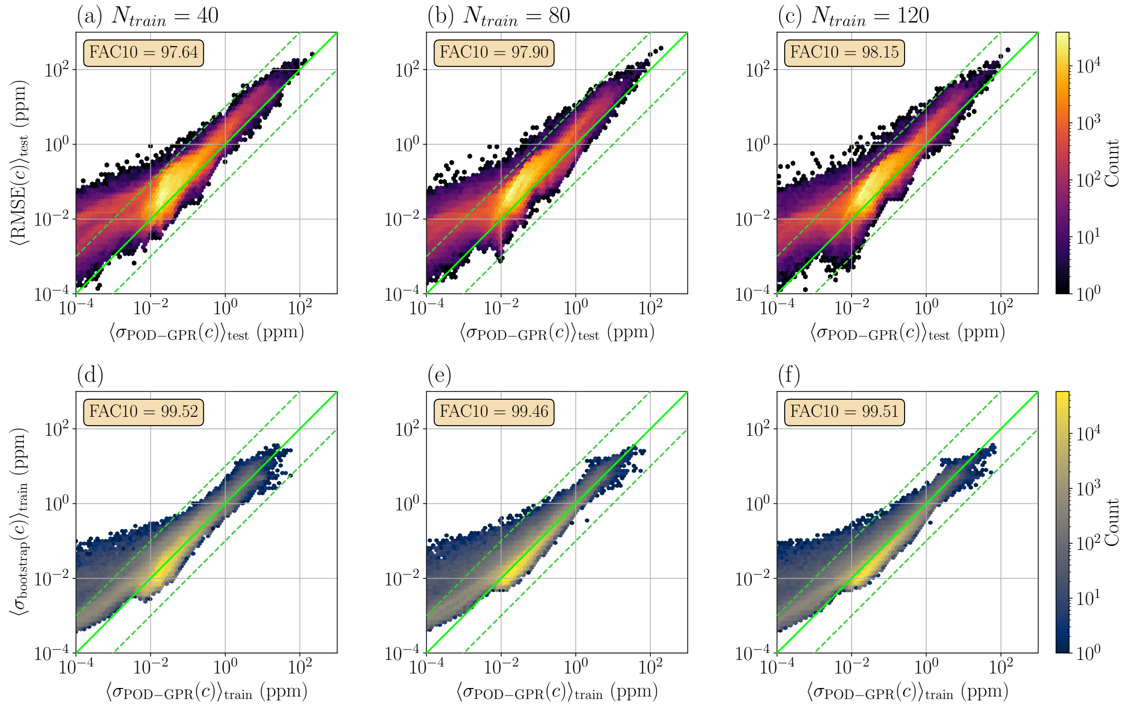

Figure 14: *Reliability diagrams of the POD–GPR uncertainty estimates for varying train set size* $N_{train} \in \{40, 80, 120\}$. *(a, b, c) Root mean square error (RMSE) of the POD–GPR concentration prediction over the test set and (d, e, f) internal variability of the mean concentration estimated by bootstrap and averaged over the train set, both versus the POD–GPR estimated uncertainty at each node where the concentration is larger than the tolerance* $c_t = 10^{-4}$ *ppm. Each hexagon is colored according to the number of node points in the hexagon. The green solid lines correspond to the identity function and the dashed lines show the range of plus or minus one order of magnitude. The FAC10 scores give the percentage of points between the two dashed lines (similarly as in Eq. 16).*

the uncertainty estimates provided by POD–GPR remain consistent as the training set size decreases, despite a tendency to overestimate.

5 Conclusion

In this study, a data-driven surrogate dispersion model based on the two-stage POD–GPR approach was trained and rigorously evaluated using a large dataset of 200 LES simulations reproducing microscale dispersion scenarios of the field-scale MUST experiment for varying meteorological forcing. The resulting surrogate model is able to capture well the general plume shape within the canopy, approaching the best achievable accuracy given the internal variability in the LES data, while being very computationally efficient.

The main novelty of this study is the in-depth analysis of the POD-GPR surrogate model uncertainty and of the weight of internal variability, thus meeting the need expressed by Dauxois et al. (2021); Tominaga et al. (2023) and Wu and Quan (2024). Future developments are required to account for the POD projection error in the POD–GPR uncertainty estimate to avoid error underestimation. But the present uncertainty estimates already explain the differences between the POD–GPR predictions and the LES references quite well, being in the right order of magnitude in 97% of cases. This work thus represents an important methodological step towards the representation of total uncertainty in microscale urban pollutant dispersion, as aleatory and modeling uncertainties have not been considered in most uncertainty quantification (García-Sánchez et al. 2014, 2017) and data assimilation (Xiao et al. 2016; Mons et al. 2017; Sousa et al. 2018; Sousa and Gorlé 2019; Defforge et al. 2019, 2021) studies to date.

A second important contribution of this study is the method for selecting a priori the POD latent space dimension, which is based on a trade-off between the accuracy of the POD reconstruction and the noise captured by the POD modes estimated by bootstrap as in Lumet et al. (2024b). The threshold used here to make this trade-off need to be consolidated and made more objective in future studies by considering a wide range of cases. For this study, the retained dimension ($L = 10$) is smaller than the dimension chosen based on the standalone reconstruction error (Xiao et al. 2019; Nony et al. 2023), but this choice is justified by the fact that using more modes ($L > 25$) significantly noises and degrades the POD–GPR response surface despite slightly better global metrics such as FAC2 and NMSE. This highlights that a surrogate model validation process learning from LES data, especially for the concentration variable, should not be based solely on global metrics but requires more local and structural analyses.

In this study, the main shortcoming of the POD–GPR approach is its lack of accuracy in areas of high concentration, i.e. close to the source. This is mainly due to POD, as a linear transformation is not well suited to the wide disparity in concentration scales and introduces projection errors. A promising way to overcome this issue is the mixture-of-experts approach (El Garroussi et al. 2020), whose key idea is to train several POD–GPR models, each corresponding to a different preprocessing, to capture the different concentration scales (Lumet (2024), Appendix B). Another promising perspective is the use of nonlinear dimension reduction techniques such as neural network autoencoders (Murata et al. 2020; Xiang et al. 2021; Masoumi-Verki et al. 2022; Nony 2023). However, a difficulty lies in the interpretation of the nonlinear latent space and in the identification of the internal variability noise.

In the future, using deep learning for dimension reduction and/or learning the dependency on new parameters such as the source location will require significantly larger training datasets, which may not be feasible due to the computational cost of LES. Defining the minimum amount of LES data required for training is therefore a key issue in LES emulator development. In this study, the POD–GPR surrogate model copes very well with a reduction of the train set down to 40 samples for two input parameters. The number of training samples may be further reduced by applying adaptive sampling methods to target learning zones (Picheny et al. 2010; Braconnier et al. 2011). Multifidelity approaches (Lamberti and Gorlé 2021; Nony 2023) are also promising to to enrich a train set by including information from a lower fidelity, lower cost model, while retaining the more accurate information provided by LES, thus paving the way for the use of the uncertainty-aware POD–GPR surrogate model for more general and more complex urban pollutant dispersion studies.

Acknowledgements 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). The authors acknowledge Bastien Nony for preliminary code development and helpful discussion.

Availability of data and codes The dataset used in this paper is openly available on a public repository (Lumet et al. 2024a). A notebook describing the construction and validation of the POD–GPR surrogate model is openly available at `https://github.com/eliott-lumet/pod_gpr_ppmles`. Other analysis codes developed for this study are available from the corresponding author upon reasonable request.

Declaration of competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Appendix: The nearest neigbor control surrogate model

We use a Nearest Neighbor model (1–NN) as a simple baseline model against which we compare the POD–GPR accuracy. It is an appropriate control model because it represents the generalization error obtained by simply querying the available simulation dataset, and thus represents the minimum level of error that the POD–GPR must exceed to be worth using. The 1–NN is a classical k-Nearest Neighbor (k–NN) model (Hastie et al. 2009) with only one neighbor ($k = 1$). The 1–NN prediction is simply defined as the nearest LES field in the training dataset:

$$\mathbf{y}_{\text{surrogate}}(\boldsymbol{\theta}) = \mathbf{y}_{\text{LES}}^{train}(\boldsymbol{\theta}^*), \text{ with } \boldsymbol{\theta}^* = \min_{1 \leq i \leq N_{train}} d(\boldsymbol{\theta}_i^{train}, \boldsymbol{\theta}), \tag{19}$$

where d is the Euclidean distance in a rescaled input space:

$$d(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) = \sqrt{\left(\frac{\alpha_{\text{inlet}}^{(2)} - \alpha_{\text{inlet}}^{(1)}}{\alpha_{\text{inlet}}^{\max} - \alpha_{\text{inlet}}^{\min}}\right)^2 + \zeta^2 \left(\frac{u_*^{(2)} - u_*^{(1)}}{u_*^{\max} - u_*^{\min}}\right)^2} \tag{20}$$

where $\alpha_{\text{inlet}}^{\min}$, $\alpha_{\text{inlet}}^{\max}$, $u_*^{\min}$, and $u_*^{\max}$ are the input space boundaries, and ζ is a rescaling factor that distorts the distances in the parameter space.

The hyperparameter ζ gives more or less weight to the friction velocity when searching for the closest LES field in the dataset (Eq. 19). It is optimized during training by cross-validation (Hastie et al. 2009) with 8-fold resampling of the train set. The best compromise between RMSE, VG and FMS(1 ppm) scores is obtained for $\zeta = 0.275$, which reduces the distances along the friction velocity axis and therefore gives more weight to the inlet wind direction parameter.

References

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

Bahlali, M. L., Dupont, E., and Carissimo, B. (2019). Atmospheric dispersion using a Lagrangian stochastic approach: Application to an idealized urban area under neutral and stable meteorological conditions. *Journal of Wind Engineering and Industrial Aerodynamics*, 193:103976. ISSN 0167-6105. DOI: 10.1016/j.jweia.2019.103976.

Berkooz, G., Holmes, P., and Lumley, J. L. (1993). The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575. DOI: 10.1146/annurev.fl.25.010193.002543.

Biltoft, C. (2001). 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.

Blocken, B., Stathopoulos, T., Saathoff, P., and Wang, X. (2008). Numerical evaluation of pollutant dispersion in the built environment: Comparisons between models and experiments. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10):1817–1831. ISSN 0167-6105. DOI: 10.1016/j.jweia.2008.02.049. 4th International Symposium on Computational Wind Engineering (CWE2006).

Blocken, B. (2014). 50 years of computational wind engineering: Past, present and future. *Journal of Wind Engineering and Industrial Aerodynamics*, 129:69–102. ISSN 0167-6105. DOI: 10.1016/j.jweia.2014.03.008.

Blocken, B. (2015). Computational Fluid Dynamics for urban physics: Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations. *Building and Environment*, 91:219–245. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2015.02.015. Fifty Year Anniversary for Building and Environment.

Braconnier, T., Ferrier, M., Jouhaud, J.-C., Montagnac, M., and Sagaut, P. (2011). Towards an adaptive POD/SVD surrogate model for aeronautic design. *Computers & Fluids*, 40(1):195–209. ISSN 0045-7930. DOI: 10.1016/j.compfluid.2010.09.002.

Brunton, S. L. and Kutz, J. N. (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press. DOI: 10.1017/9781108380690.

Camelli, F., Lohner, R., and Hanna, S. (2005). VLES study of MUST experiment. In *43rd AIAA Aerospace Sciences Meeting and Exhibit*. DOI: 10.2514/6.2005-1279.

Chang, J. and Hanna, S. (2004). Air quality model performance evaluation. *Meteorol. Atm. Phys*, 87(1):167–196. DOI: 10.1007/s00703-003-0070-7.

Cheng, K., Lu, Z., Ling, C., and Zhou, S. (2020). Surrogate-assisted global sensitivity analysis: an overview. *Structural and Multidisciplinary Optimization*, 61:1187–1213. DOI: 10.1007/s00158-019-02413-5.

Chinesta, F., Ladeveze, P., and Cueto, E. (2011). A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395–404. ISSN 1886-1784. DOI: 10.1007/s11831-011-9064-7.

Cordier, L. and Bergmann, M. (2006). Réduction de dynamique par décomposition orthogonale aux valeurs propres (POD) (in French). Lecture notes, Ecole de printemps OCET. URL `https://www.math.u-bordeaux.fr/~mbergman/PDF/OuvrageSynthese/OCET06.pdf`. Accessed: 2023-12-01.

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

Defforge, C. L., Carissimo, B., Bocquet, M., Bresson, R., and Armand, P. (2021). Improving numerical dispersion modelling in built environments with data assimilation using the iterative ensemble Kalman smoother. *Boundary-Layer Meteorology*, 179(2): 209–240. ISSN 1573-1472. DOI: 10.1007/s10546-020-00588-9.

Defforge, C. L., Carissimo, B., Bocquet, M., Bresson, R., and Armand, P. (2019). Improving CFD atmospheric simulations at local scale for wind resource assessment using the iterative ensemble Kalman smoother. *Journal of Wind Engineering and Industrial Aerodynamics*, 189:243–257. ISSN 0167-6105. DOI: 10.1016/j.jweia.2019.03.030.

Donnelly, R., Lyons, T., and Flassak, T. (2009). Evaluation of results of a numerical simulation of dispersion in an idealised urban area for emergency response modelling. *Atmos. Environ.*, 43(29):4416–4423. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2009.05.038.

EEA. (2020). Air quality in Europe. 2020 report, European Environment Agency. URL `https://www.eea.europa.eu/publications/air-quality-in-europe-2020`.

Efthimiou, G. C., Bartzis, J. G., and Koutsourakis, N. (2011). Modelling concentration fluctuations and individual exposure in complex urban environments. *Journal of Wind Engineering and Industrial Aerodynamics*, 99(4):349–356. ISSN 0167-6105. DOI: 10.1016/j.jweia.2010.12.007. The Fifth International Symposium on Computational Wind Engineering.

El Garroussi, S., Ricci, S., De Lozzo, M., Goutal, N., and Lucor, D. (2020). Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions. In *2020 TELEMAC-MASCARET User Conference*. URL `https://hal.science/hal-03444227/document`. Accessed: 2023-12-01.

Fellmann, N., Pasquier, M., Blanchet-Scalliet, C., Helbert, C., Spagnol, A., and Sinoquet, D. (2023). Sensitivity analysis for sets : application to pollutant concentration maps.

Fernando, H. J. S., Lee, S. M., Anderson, J., Princevac, M., Pardyjak, E., and Grossman-Clarke, S. (2001). Urban fluid mechanics: Air circulation and contaminant dispersion in cities. *Environ. Fluid Mech.*, 1(1):107–164. DOI: 10.1023/A:1011504001479.

Forkman, J., Josse, J., and Piepho, H.-P. (2019). Hypothesis tests for principal component analysis when variables are standardized. *Journal of Agricultural, Biological and Environmental Statistics*, 24:289–308. DOI: 10.1007/s13253-019-00355-5.

Franke, J., Hellsten, A., Schlünzen, H., and Carissimo, B. (2007). Best practice guideline for the CFD simulation of flows in the urban environmen. Technical report, COST European Cooperation in Science and Technology. URL `https://hal.science/hal-04181390`. Accessed: 2023-12-01.

García-Sanchez, C., van Beeck, J., and Gorlé, C. (2018). Predictive large eddy simulations for urban flows: Challenges and opportunities. *Building and Environment*, 139:146–156. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2018.05.007.

García-Sánchez, C., Philips, D., and Gorlé, C. (2014). Quantifying inflow uncertainties for CFD simulations of the flow in downtown Oklahoma City. *Building and Environment*, 78:118–129. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2014.04.013.

García-Sánchez, C., Van Tendeloo, G., and Gorlé, C. (2017). Quantifying inflow uncertainties in RANS simulations of urban pollutant dispersion. *Atmospheric Environment*, 161:263–273. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2017.04.019.

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

Gorlé, C. and Iaccarino, G. (2013). A framework for epistemic uncertainty quantification of turbulent scalar flux models for Reynolds-averaged Navier-Stokes simulations. *Physics of Fluids*, 25(5):055105. ISSN 1070-6631. DOI: 10.1063/1.4807067.

Gorlé, C., Garcia-Sanchez, C., and Iaccarino, G. (2015). Quantifying inflow and RANS turbulence model form uncertainties for wind engineering flows. *Journal of Wind Engineering and Industrial Aerodynamics*, 144:202–212. DOI: 10.1016/j.jweia.2015.03.025.

Gousseau, P., Blocken, B., Stathopoulos, T., and van Heijst, G. (2011). CFD simulation of near-field pollutant dispersion on a high-resolution grid: A case study by LES and RANS for a building group in downtown Montreal. *Atmos. Environ.*, 45(2):428–438. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2010.09.065.

Gromke, C., Jamarkattel, N., and Ruck, B. (2016). Influence of roadside hedgerows on air quality in urban street canyons. *Atmospheric Environment*, 139:75–86. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2016.05.014.

Guo, M. and Hesthaven, J. S. (2018). Reduced order modeling for nonlinear structural analysis using gaussian process regression. *Computer Methods in Applied Mechanics and Engineering*, 341:807–826. ISSN 0045-7825. DOI: 10.1016/j.cma.2018.07.017.

Halton, J. H. (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702. DOI: 10.1145/355588.365104.

Hanna, S. R., Hansen, O. R., and Dharmavaram, S. (2004). FLACS CFD air quality model performance evaluation with Kit Fox, MUST, Prairie Grass, and EMU observations. *Atmos. Environ.*, 38(28):4675–4687. DOI: 10.1016/j.atmosenv.2004.05.041.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer. DOI: 10.1007/978-0-387-21606-5.

Hsieh, K.-J., Lien, F.-S., and Yee, E. (2007). Numerical modeling of passive scalar dispersion in an urban canopy layer. *Journal of Wind Engineering and Industrial Aerodynamics*, 95(12):1611–1636. ISSN 0167-6105. DOI: 10.1016/j.jweia.2007.02.028.

Huang, C., Zhang, G., Yao, J., Wang, X., Calautit, J. K., Zhao, C., An, N., and Peng, X. (2022). Accelerated environmental performance-driven urban design with generative adversarial network. *Building and Environment*, 224:109575. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2022.109575.

Kastner, P. and Dogan, T. (2023). A GAN-Based Surrogate Model for Instantaneous

Urban Wind Flow Prediction. *Building and Environment*, 242:110384. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2023.110384.

Kessy, A., Lewin, A., and Strimmer, K. (2018). Optimal whitening and decorrelation. *The American Statistician*, 72(4):309–314. DOI: 10.1080/00031305.2016.1277159.

Klein, P., Leitl, B., and Schatzmann, M. (2007). Driving physical mechanisms of flow and dispersion in urban canopies. *Int. J. Climatol.*, 27(14):1887–1907. DOI: 10.1002/joc.1581.

König, M. (2014). *Large-eddy simulation modelling for urban scale.* PhD thesis, University of Leipzig. URL `https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=baab9d7b41623099c1b6d840c11821b8e31fac9b`. Accessed: 2024-08-05.

Kraichnan, R. H. (1970). Diffusion by a random velocity field. *Phys. Fluids*, 13(1):22–31. DOI: 10.1063/1.1692799.

Kumar, P., Feiz, A.-A., Ngae, P., Singh, S. K., and Issartel, J.-P. (2015). CFD simulation of short-range plume dispersion from a point release in an urban like environment. *Atmos. Environ.*, 122:645–656. DOI: 10.1016/j.atmosenv.2015.10.027.

Lamberti, G. and Gorlé, C. (2021). A multi-fidelity machine learning framework to predict wind loads on buildings. *Journal of Wind Engineering and Industrial Aerodynamics*, 214:104647. ISSN 0167-6105. DOI: 10.1016/j.jweia.2021.104647.

Lassila, T., Manzoni, A., Quarteroni, A., and Rozza, G. (2014). Model order reduction in fluid dynamics: challenges and perspectives. *Reduced Order Methods for modeling and computational reduction*, pages 235–273. DOI: 10.1007/978-3-319-02090-7_9.

Lucas, D. D., Gowardhan, A., Cameron-Smith, P., and Baskett, R. L. (2016). Impact of meteorological inflow uncertainty on tracer transport and source estimation in urban atmospheres. *Atmospheric Environment*, 143:120–132. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2016.08.019.

Lumet, E. (2024). *Assessing and reducing uncertainty in large-eddy simulation for microscale atmospheric dispersion*. PhD thesis, Université Toulouse III - Paul Sabatier. URL `https://theses.fr/2024TLSES003`. Accessed: 2024-05-30.

Lumet, E., Jaravel, T., and Rochoux, M. C. (2024)a. PPMLES – Perturbed-Parameter ensemble of MUST Large-Eddy Simulations. Dataset on Zenodo. To be published.

Lumet, E., Jaravel, T., Rochoux, M. C., Vermorel, O., and Lacroix, S. (2024)b. Assessing the Internal Variability of Large-Eddy Simulations for Microscale Pollutant Dispersion Prediction in an Idealized Urban Environment. *Boundary-Layer Meteorology*, 190(2): 9. ISSN 1573-1472. DOI: 10.1007/s10546-023-00853-7.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., and Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: A review. *Frontiers in Public Health*, 8. ISSN 2296-2565. DOI: 10.3389/fpubh.2020.00014.

Margheri, L. and Sagaut, P. (2016). A hybrid anchored-ANOVA – POD/Kriging method for uncertainty quantification in unsteady high-fidelity CFD simulations. *Journal of Computational Physics*, 324:137–173. ISSN 0021-9991. DOI: 10.1016/j.jcp.2016.07.036.

Marrel, A., Perot, N., and Mottet, C. (2015). Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, 29(3):959–974. ISSN 1436-3259. DOI: 10.1007/s00477-014-0927-y.

Masoumi-Verki, S., Haghighat, F., and Eicker, U. (2022). A review of advances towards efficient reduced-order models (ROM) for predicting urban airflow and pollutant dispersion. *Building and Environment*, 216:108966. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2022.108966.

Mendil, M., Leirens, S., Armand, P., and Duchenne, C. (2022). Hazardous atmospheric dispersion in urban areas: A Deep Learning approach for emergency pollution forecast. *Environmental Modelling & Software*, 152:105387. ISSN 1364-8152. DOI: 10.1016/j.envsoft.2022.105387.

Milliez, M. and Carissimo, B. (2007). Numerical simulations of pollutant dispersion in an idealized urban area, for different meteorological conditions. *Boundary-Layer Meteorology*, 122(2):321–342. DOI: 10.1007/s10546-006-9110-4.

Miyagusuku, R., Yamashita, A., and Asama, H. (2015). Gaussian processes with input-dependent noise variance for wireless signal strength-based localization. In *2015 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 1–6. DOI: 10.1109/SSRR.2015.7442993.

Mons, V., Margheri, L., Chassaing, J.-C., and Sagaut, P. (2017). Data assimilation-based reconstruction of urban pollutant release characteristics. *Journal of Wind Engineering and Industrial Aerodynamics*, 169:232–250. ISSN 0167-6105. DOI: 10.1016/j.jweia.2017.07.007.

Montazeri, H. and Blocken, B. (2013). CFD simulation of wind-induced pressure coefficients on buildings with and without balconies: Validation and sensitivity analysis. *Build Environ.*, 60:137–149. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2012.11.012.

Murata, T., Fukami, K., and Fukagata, K. (2020). Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *Journal of Fluid Mechanics*, 882:A13. DOI: 10.1017/jfm.2019.822.

Nagel, T., Schoetter, R., Masson, V., Lac, C., and Carissimo, B. (2022). Numerical analysis of the atmospheric boundary-layer turbulence influence on microscale transport of pollutant in an idealized urban environment. *Boundary-Layer Meteorology*, 184(1): 113–141. DOI: 10.1007/s10546-022-00697-7.

Neophytou, M., Gowardhan, A., and Brown, M. (2011). An inter-comparison of three urban wind models using Oklahoma City Joint Urban 2003 wind field measurements. *Journal of Wind Engineering and Industrial Aerodynamics*, 99(4):357–368. DOI: 10.1016/j.jweia.2011.01.010.

Nicoud, F. and Ducros, F. (1999). Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow Turbul. Combust.*, 62(3):183–200. DOI: 10.1023/A:1009995426001.

Nony, B. X., Rochoux, M. C., Jaravel, T., and Lucor, D. (2023). Reduced-order modeling for parameterized large-eddy simulations of atmospheric pollutant dispersion. *Stoch. Environ. Res. Risk Assess.*, 37(6):2117–2144. ISSN 1436-3259. DOI: 10.1007/s00477-023-02383-7.

Nony, B. X. (2023). *Reduced-order models under uncertainties for microscale atmospheric pollutant dispersion in urban areas: exploring learning algorithms for high-fidelity model emulation*. Phd thesis, Université de Toulouse, France.

Pasquier, M., Jay, S., Jacob, J., and Sagaut, P. (2023). A Lattice-Boltzmann-based modelling chain for traffic-related atmospheric pollutant dispersion at the local urban scale. *Building and Environment*, 242:110562. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2023.110562.

Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7):071008. ISSN 1050-0472. DOI: 10.1115/1.4001873.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *J. Am. Stat. Assoc.*, 89(428):1303–1313. DOI: 10.1080/01621459.1994.10476870.

Ramshaw, J., O'Rourke, P., and Amsden, A. (1986). Acoustic damping for explicit calculations of fluid flow at low Mach number. Technical report no. LA–10641-MS, Los Alamos National Laboratories, USA. URL `https://inis.iaea.org/collection/NCLCollectionStore/_Public/17/074/17074782.pdf`. Accessed: 2023-12-01.

Rasmussen, C. E., Williams, C. K., et al. (2006). *Gaussian processes for machine learning*, volume 1. Springer. DOI: 10.7551/mitpress/3206.001.0001.

Santiago, J. L., Dejoan, A., Martilli, A., Martin, F., and Pinelli, A. (2010). 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. *Boundary-Layer Meteorology*, 135(1): 109–132. DOI: 10.1007/s10546-010-9466-3.

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

Schatzmann, M., Olesen, H., and Franke, J. (2010). COST 732 model evaluation case studies: approach and results. Technical report, University of Hamburg, Meteorological Institute. 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: 2023-12-01.

Schmidt, O. T. and Colonius, T. (2020). Guide to spectral proper orthogonal decomposition. *AIAA journal*, 58(3):1023–1033. DOI: 10.2514/1.J058809.

Schönfeld, T. and Rudgyard, M. (1999). Steady and unsteady flow simulations using the hybrid flow solver AVBP. *AIAA journal*, 37(11):1378–1385. DOI: 10.2514/2.636.

Sirovich, L. (1987). Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quarterly of applied mathematics*, 45(3):561–571. DOI: 10.1090/qam/910462.

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

Sousa, J. and Gorlé, C. (2019). Computational urban flow predictions with Bayesian inference: Validation with field data. *Building and Environment*, 154:13–22. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2019.02.028.

Sousa, J., García-Sánchez, C., and Gorlé, C. (2018). Improving urban flow predictions through data assimilation. *Building and Environment*, 132:282–290. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2018.01.032.

Spicer, T. O. and Tickle, G. (2021). Simplified source description for atmospheric dispersion model comparison of the Jack Rabbit II chlorine field experiments. *Atmospheric Environment*, 244:117866. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2020.117866.

Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer Series in Statistics. Springer Science & Business Media. DOI: 10.1007/978-1-4612-1494-6.

Taira, K., Brunton, S. L., Dawson, S. T. M., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., and Ukeiley, L. S. (2017). Modal analysis of fluid flows: An overview. *AIAA Journal*, 55(12):4013–4041. DOI: 10.2514/1.J056060.

Tominaga, Y. and Stathopoulos, T. (2007). Turbulent Schmidt numbers for CFD analysis with various types of flowfield. *Atmospheric Environment*, 41(37):8091–8099. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2007.06.054.

Tominaga, Y. and Stathopoulos, T. (2009). Numerical simulation of dispersion around an isolated cubic building: Comparison of various types of k–ϵ models. *Atmospheric Environment*, 43(20):3200–3210. ISSN 1352-2310. DOI: 10.1016/j.atmosenv.2009.03.038.

Tominaga, Y., Wang, L. L., Zhai, Z. J., and Stathopoulos, T. (2023). Accuracy of CFD simulations in urban aerodynamics and microclimate: Progress and challenges. *Building and Environment*, 243:110723. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2023.110723.

Vasaturo, R., Kalkman, I., Blocken, B., and van Wesemael, P. (2018). 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: 241–261. DOI: 10.1016/j.jweia.2017.11.025.

Vinuesa, R. and Brunton, S. L. (2022). Enhancing computational fluid dynamics with machine learning. *Nature Computational Science*, 2(6):358–366. ISSN 2662-8457. DOI: 10.1038/s43588-022-00264-7.

Weerasuriya, A., Zhang, X., Lu, B., Tse, K., and Liu, C. (2021). A gaussian process-based emulator for modeling pedestrian-level wind field. *Building and Environment*, 188:107500. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2020.107500.

Winiarek, V., Bocquet, M., Saunier, O., and Mathieu, A. (2012). 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. *Journal of Geophysical Research: Atmospheres*, 117(D5). DOI: 10.1029/2011JD016932.

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

Wu, Y. and Quan, S. J. (2024). A review of surrogate-assisted design optimization for improving urban wind environment. *Building and Environment*, 253:111157. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2023.111157.

Wu, Y., Zhan, Q., Quan, S. J., Fan, Y., and Yang, Y. (2021). A surrogate-assisted optimization framework for microclimate-sensitive urban design practice. *Building and Environment*, 195:107661. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2021.107661.

Xiang, S., Fu, X., Zhou, J., Wang, Y., Zhang, Y., Hu, X., Xu, J., Liu, H., Liu, J., Ma, J., and Tao, S. (2021). Non-intrusive reduced order model of urban airflow with dynamic boundary conditions. *Building and Environment*, 187:107397. ISSN 0360-1323. DOI: 10.1016/j.buildenv.2020.107397.

Xiao, D., Heaney, C., Fang, F., Mottet, L., Hu, R., Bistrian, D., Aristodemou, E., Navon, I., and Pain, C. (2019). A domain decomposition non-intrusive reduced order model for turbulent flows. *Computers & Fluids*, 182:15–27. ISSN 0045-7930. DOI: 10.1016/j.compfluid.2019.02.012.

Xiao, H., Wu, J.-L., Wang, J.-X., Sun, R., and Roy, C. (2016). Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach. *Journal of Computational Physics*, 324:115–136. ISSN 0021-9991. DOI: 10.1016/j.jcp.2016.07.038.

Yee, E. and Biltoft, C. A. (2004). Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Boundary-Layer Meteorology*, 111(3): 363–415. DOI: 10.1023/B:BOUN.0000016496.83909.ee.

Yue Yang, G.-W. H. and Wang, L.-P. (2008). Effects of subgrid-scale modeling on lagrangian statistics in large-eddy simulation. *Journal of Turbulence*, 9:N8. DOI: 10.1080/14685240801905360.