COMPOUND PARAMETRIC METAMODELLING OF LARGE-EDDY SIMULATIONS FOR MICROSCALE ATMOSPHERIC DISPERSION

Bastien Nony[1]

Mélanie C. Rochoux[1], Didier Lucor[2], Thomas Jaravel[3]

[1]CECI, Université de Toulouse, CNRS, [3]LISN, CNRS Université Paris-Saclay, Orsay, France

20th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes

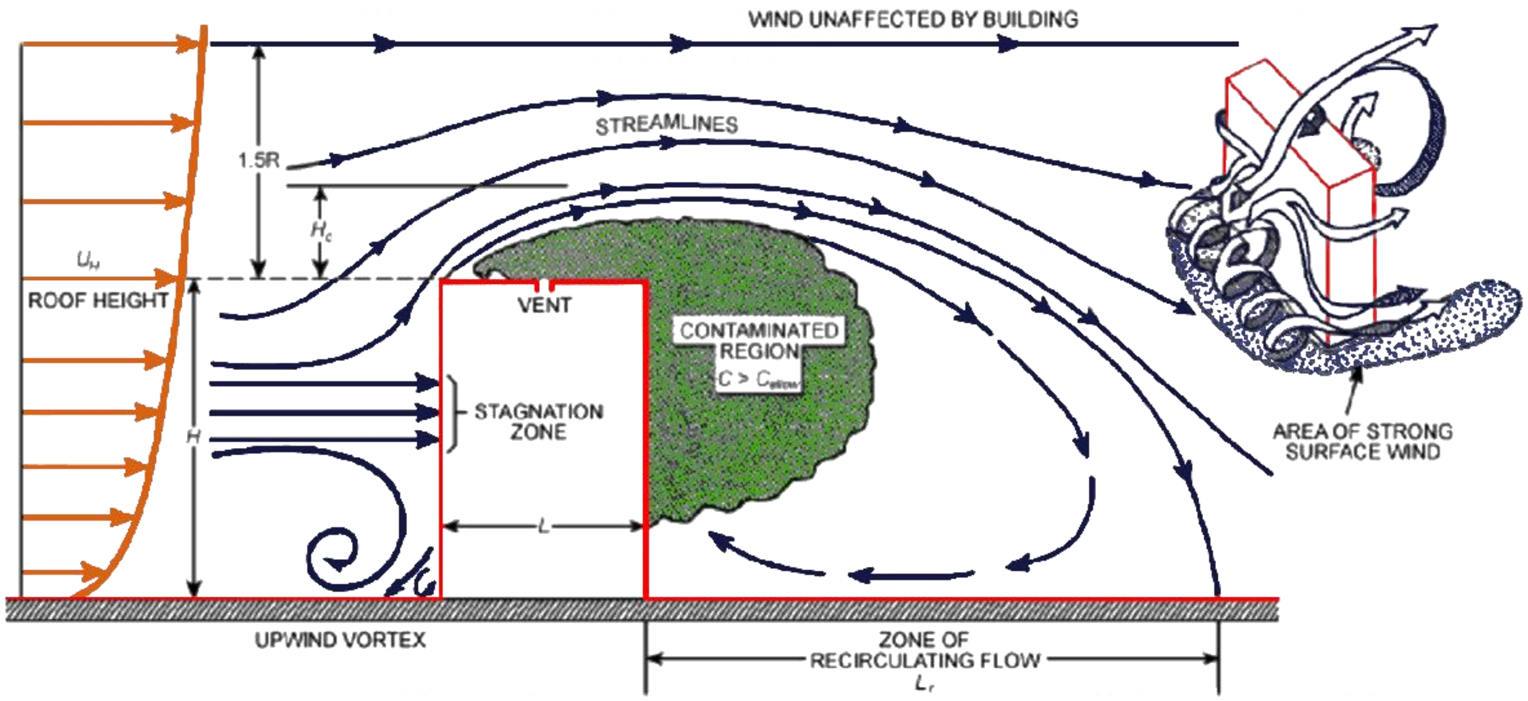
14-18 June 2021, Tartu, Estonia

Scientific challenges for microscale flow dynamics and plume dispersion

Microscale
- Evolution in a complex geometry *(urban canopy)*
- Highly dependent to near-source behaviour

Meso to microscale
- Multiscale problem *(large-scale wind forcing, turbulence, boundary layers)*

Need for a stochastic approach…
Uncertainties are not accounted for by CFD models
- Mean wind and fluctuations
- Emission source location, type of pollutant, etc.



Tominaga and Stathopoulos (2013)

CONTEXT
High fidelity modelling CFD tools

Large-eddy-simulations (LES)

+
- Can well represent complex flow features in canopies/behind obstacles
- Explicitly represent most of the eddies

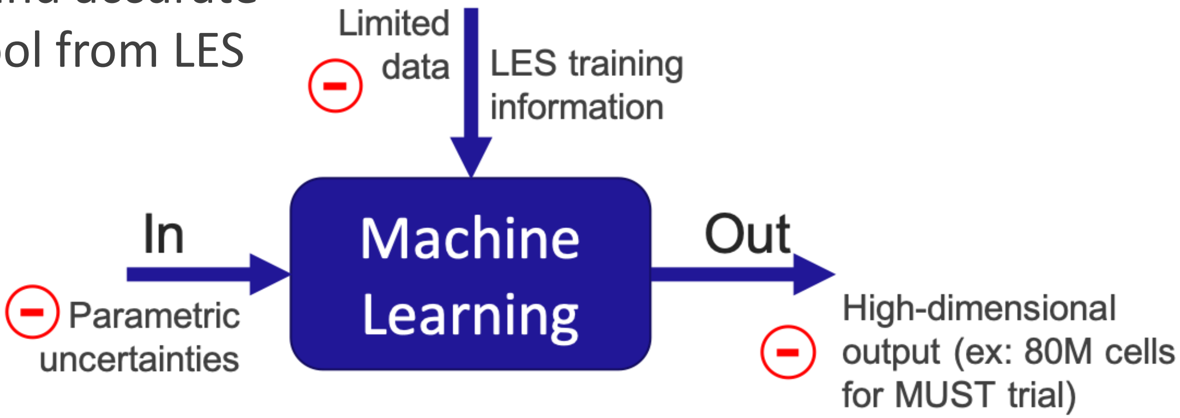
−
- Very costly (60 000 hCPU for MUST[1] trial)
- Ensemble/uncertainties

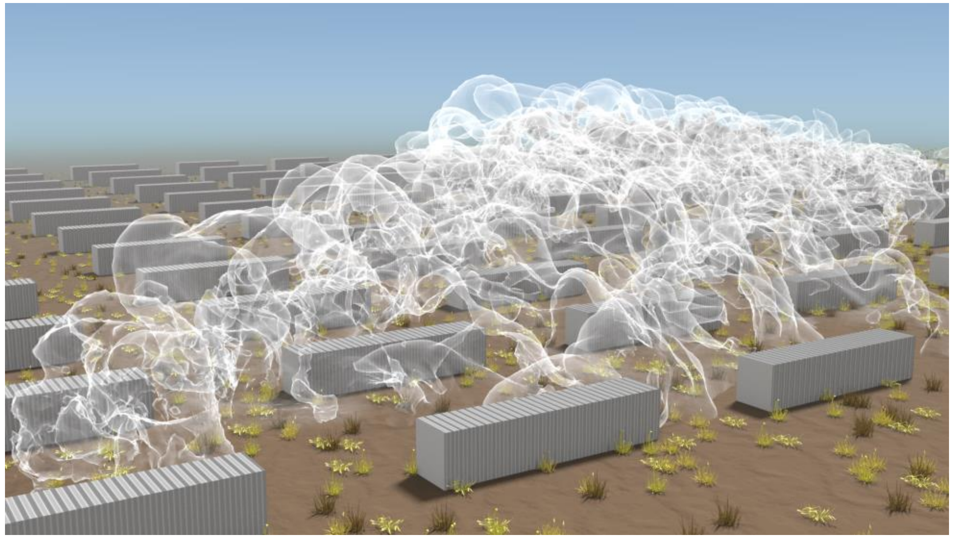
AVBP

- Compressible Navier-Stokes equations on unstructured meshes *(artificial compressibility approach for low-Mach flows)*
- Proven on many application cases (ex: aeronautical industrial applications) and evaluated for environmental flows

Objective

Build a fast and accurate predictive tool from LES information





Artist's view of the MUST case

[1]*Mock Urban Setting Case (MUST)*

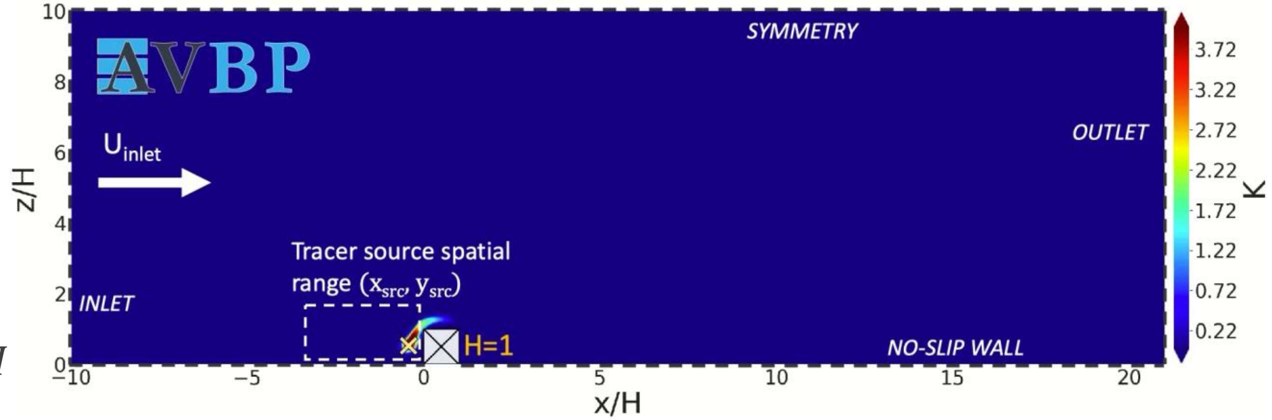
What is the most suitable machine learning metamodel for the LES microscale?

Outline

- Definition of the test case
- Metamodelling approach
- Results

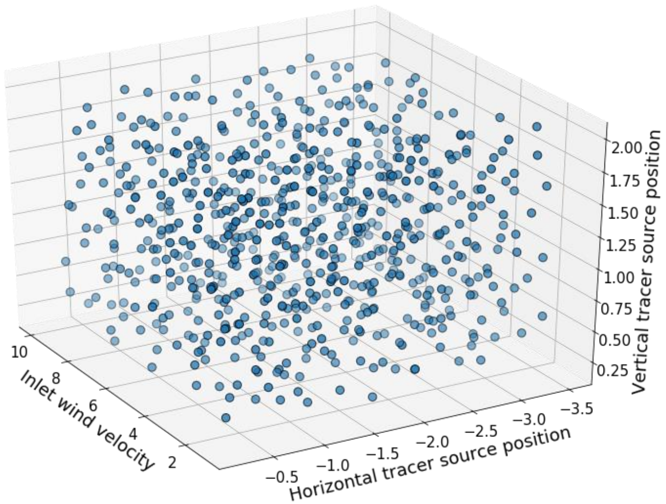
2-D flow around a surface-mounted obstacle

- Quantity of Interest K: time-averaged tracer concentration field
- Passive tracer
- Parametric uncertainties:
  - Inlet wind intensity: $U_{inlet} \in [1,10] m\ s^{-1}$
  - Emission source position: $(x_{src}, y_{src}) \in [-3.5, -0.2]H \times [0.2,2]H$



Example of a tracer concentration field K for a source centered at (-0.5,0.5) and an inlet wind of 5.5 m s^{-1}

3-D parameter space

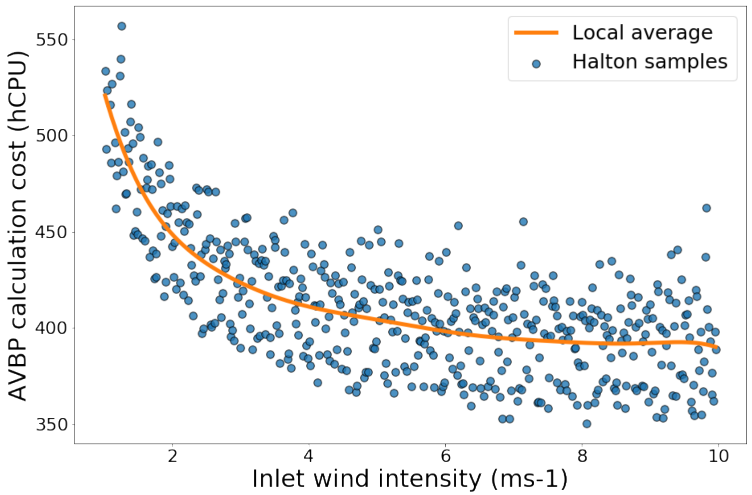


LES data set

- Densely sampled uncertain 3-D space
- 700 AVBP simulations
- Mesh resolution: $\Delta x = \Delta z = 0.04\ cm$ resolution, leading to N_{nodes} = 240,000 mesh nodes

Optimization of computational cost

- Simulation time depending on the inlet wind intensity
- Average calculation cost: 400hCPU/run



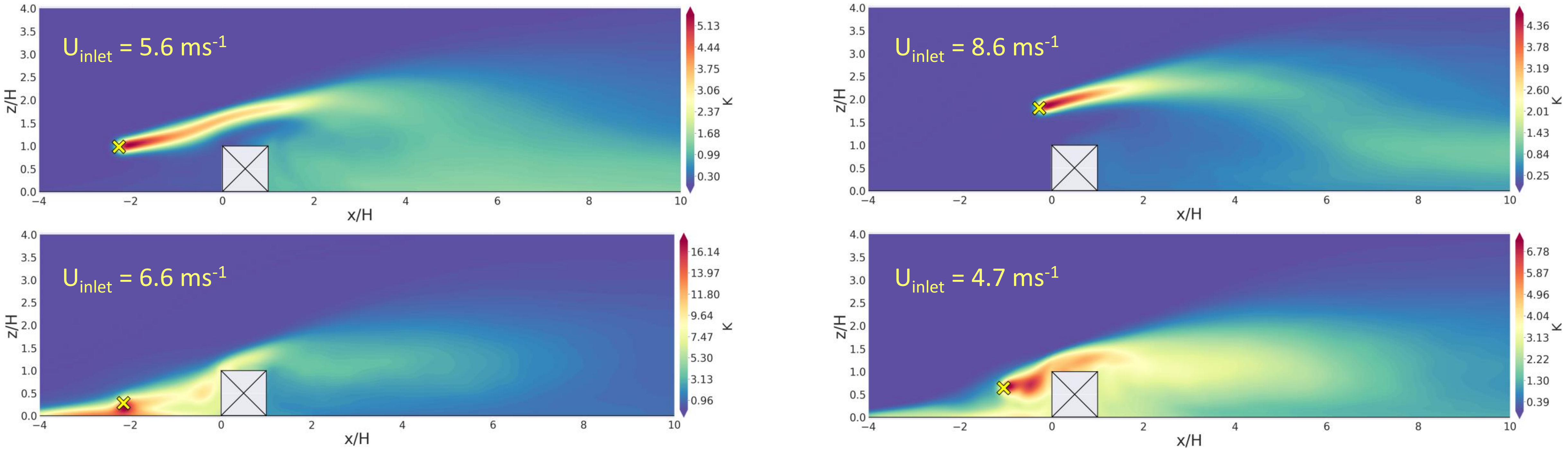
Multiple outputs for LES

- Wind flow: horizontal and vertical velocities
- Plume dispersion: averages, fluctuations
- Cross statistics between flow and tracer dispersion

Time-averaged tracer concentration statistics

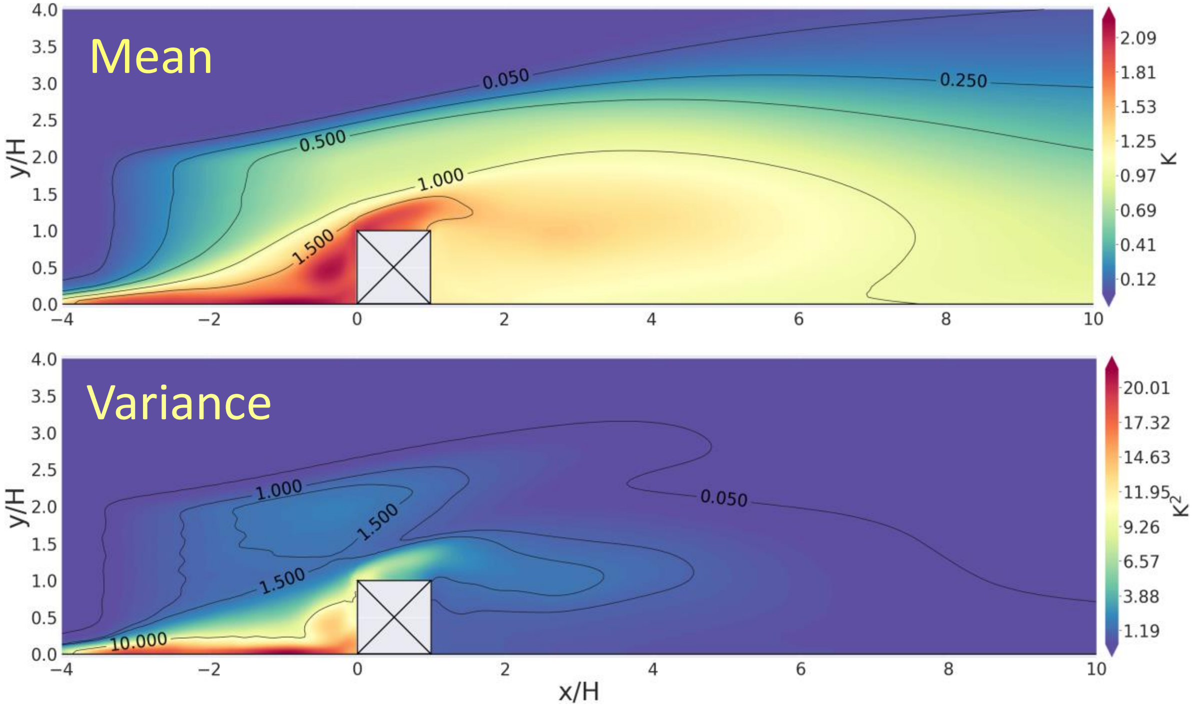
- An easy way to start
- Search for metamodels that are able to reproduce the most important flow features

A few examples of the LES dataset (mean concentration fields)



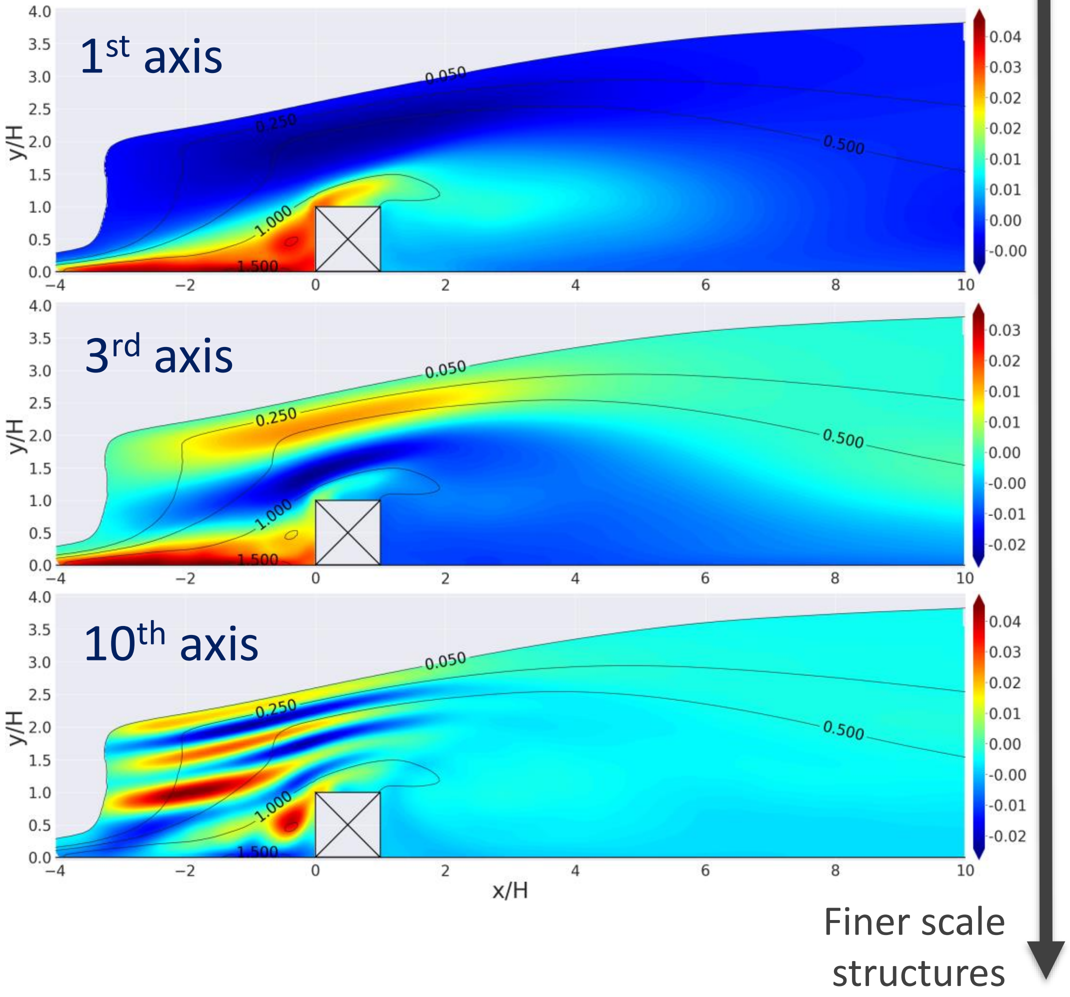
Scientific issue: High dimension output

- Account for spatial correlations
- Reduce space dimension from N_{nodes} = 240,000 to N_{POD} = 200
- Total explained variance 200 axes > 99.9 %.



Ensemble statistics over the 700 LES dataset

First POD axes

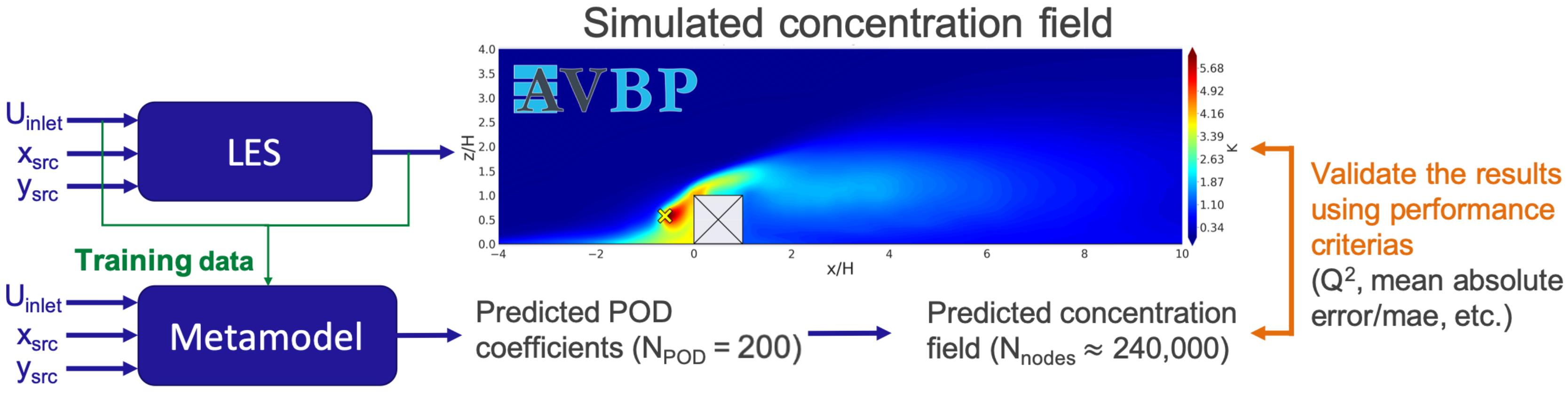


METAMODELLING METHOD

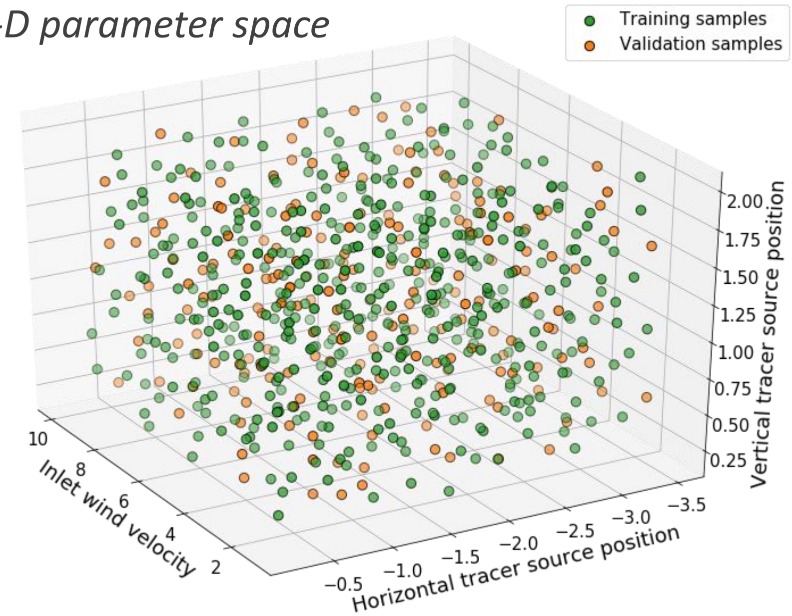
Metamodelling fomulation

700 LES dataset split

- 70% training data (490 LES) used for fitting the metamodel hyperparameters
- 30% validation data (210 LES) used for performance evaluation



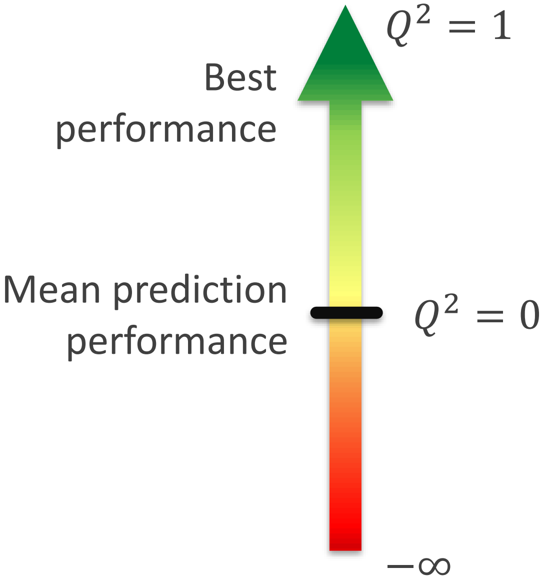
3-D parameter space



Several performance criteria

- $$Q^2 = 1 - \frac{\sum_{i=1}^{210}(y_i - \bar{y})^2}{\sum_{i=1}^{210}(y_i - \hat{y}_i)^2}$$
- $$\text{MAE} = \sum_{I=1}^{210} |y_i - \hat{y}_i|$$

- y_i, LES outputs
- $\bar{y}$, LES ensemble mean
- $\hat{y}_i$, Metamodel predictions



Looking for the most appropriate metamodels for the LES microscale

Multiple polynomial regression
Variable selection
Regularization type

Gaussian processes
Variable selection
Kernel type
Kernel's hyperparameters

Random Forest
Variable selection
Trees parameters
Loss function optimization

Gradient boosting
Variable selection
Tree parameters
Loss function optimization
Boosting parameters

Compound model
Model selection
Models hyperparameters

Model selection
- Every POD axis leads to a selective process
- One metamodel is kept per axis
- Selection relies on best Q^2 performance

RESULTS

Multiple polynomial regression

Multiple polynomial regression

|

Variable selection

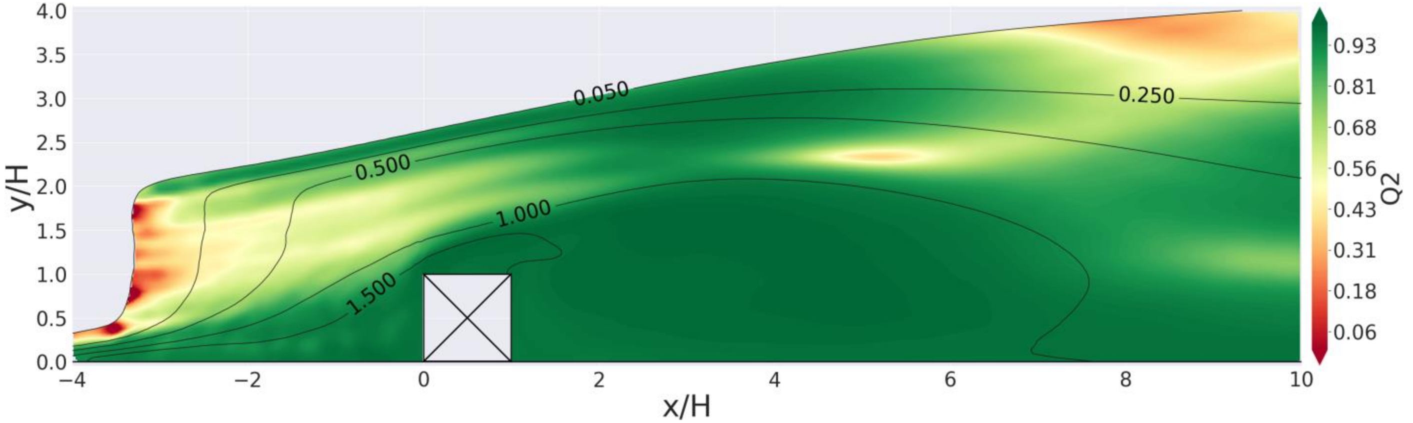
|

Regularization type

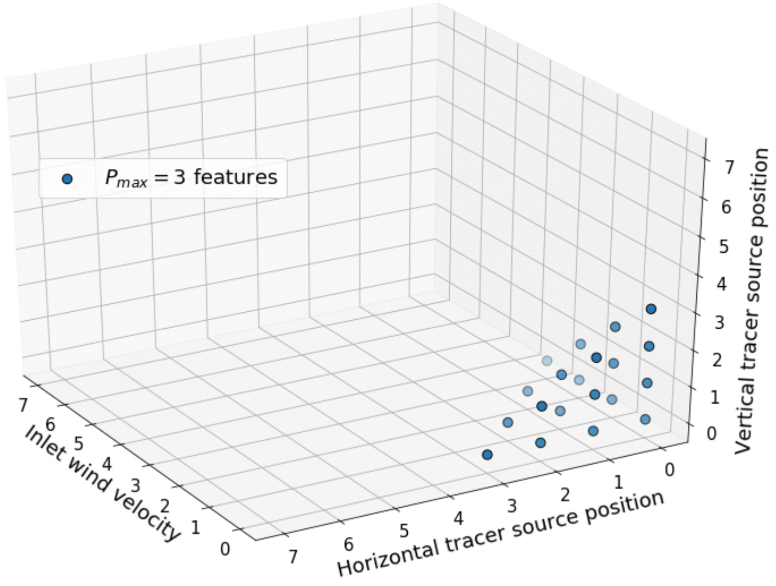
P_{max}	3	5	7
Nb. Variables	20	56	120
Q^2 of MPR	76.6 %	82.7 %	80.8 %
Q^2 of Ridge $\|\| \cdot \|\|_2$	$\leq$ 76.6%	$\leq$ 82.7 %	82.1 %
Q^2 of LASSO $\|\| \cdot \|\|_1$	$\leq$ 76.6 %	$\leq$ 82.7 %	79.5 %
Q^2 of Matching Pursuit $\|\| \cdot \|\|_0$	$\leq$ 76.6 %	$\leq$ 82.7 %	83.0 %

MPR expression

$$\hat{y} = \sum_{i+j+k \leq P_{max}} \lambda_{i,j,k}\, U_{inlet}^{i}\, x_{src}^{j}\, y_{src}^{k}$$



Q^2 response surface for the MPR without penalty and P_{max} = 5



Variable selection using the 3-D uncertain parameter polynomial combinations

RESULTS

MPR prediction fields

Prediction procedure

1. The metamodel predicts the 200 POD coefficients
2. Predicted coefficients are projected in the spatial domain using inverse POD operation

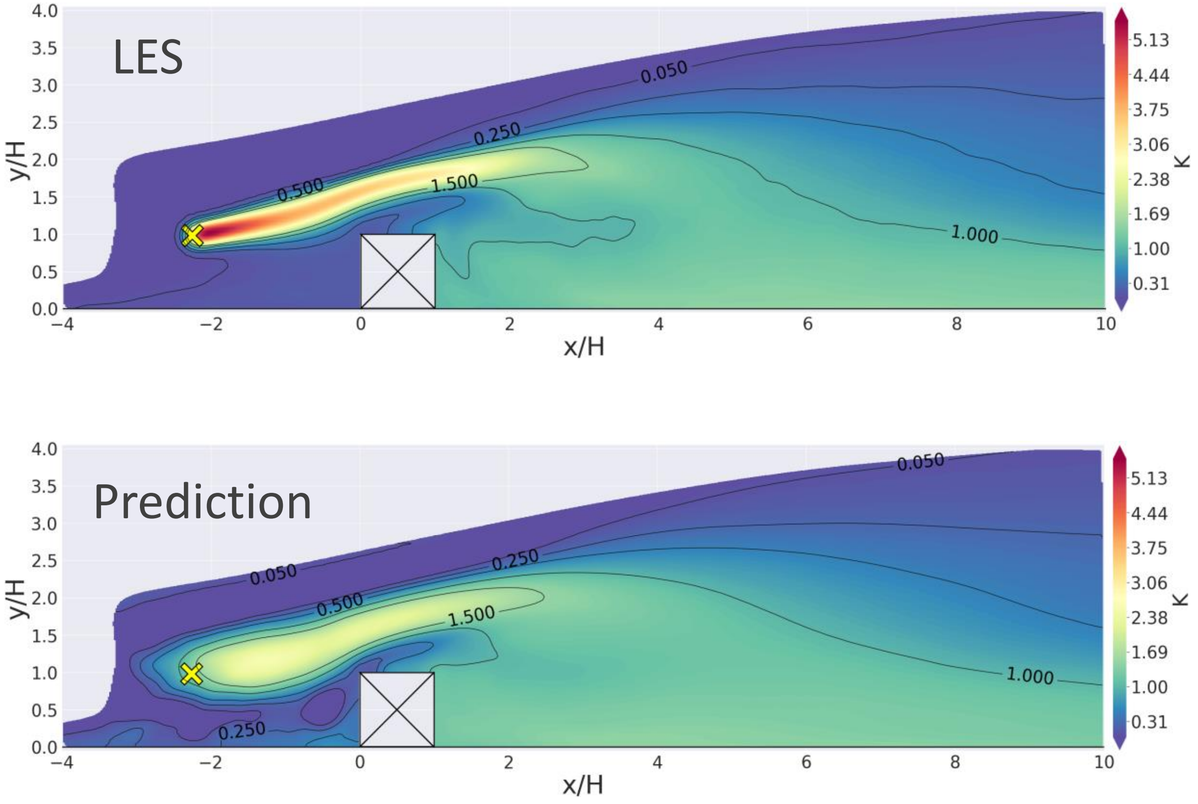
Observations

Upstream

- Coarse structure in the wake of the emission source (wider range, under-predicted peak intensity)
- Distorted over-predicted areas close to the ground

Downstream

- More steady concentration lines than LES
- Small prediction errors slightly offset the isolines



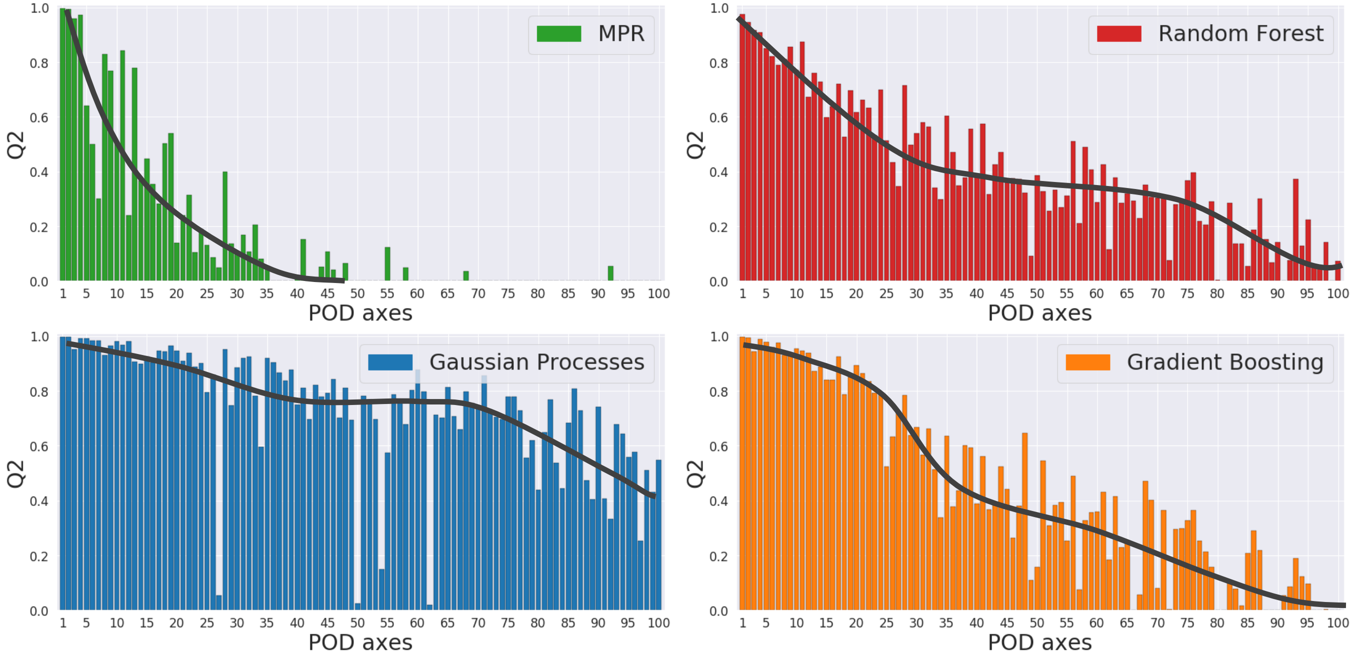
Mean concentration fields for a validation simulation with U_{inlet} = 5.6 ms^{-1}

Q^2 performance evaluation of metamodels

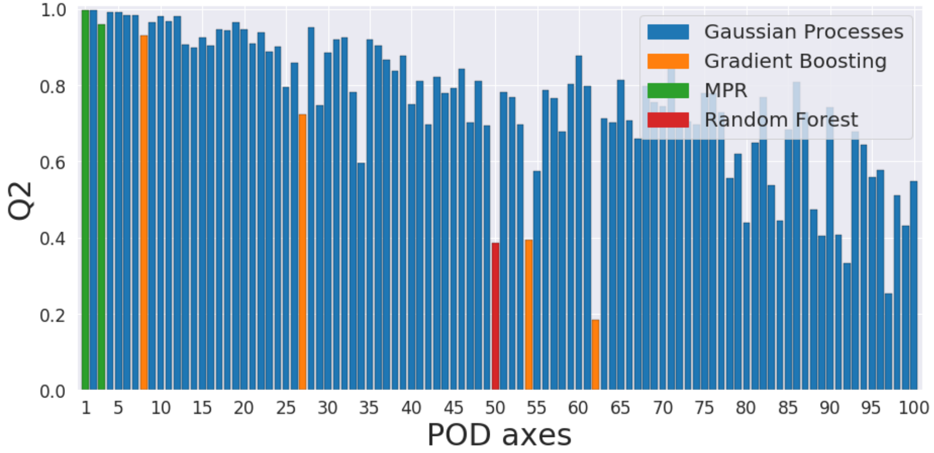
- MPR performs well on the 5 first POD axes
- Gradient Boosting performances decrease linearly from axis 20
- Random Forest performs poorly on first axes but the decay is slower than gradient boosting
- Gaussian processes maintain a good level of performance on the first 100 axes

Compound model composition

- 4 MPR
- 6 Gradient Boosting
- 15 Random Forest
- 172 Gaussian Processes



Q^2 performances on the first 100 POD axes of 4 families of metamodels



In this case the compound model is essentially a combination of Gaussian process metamodels

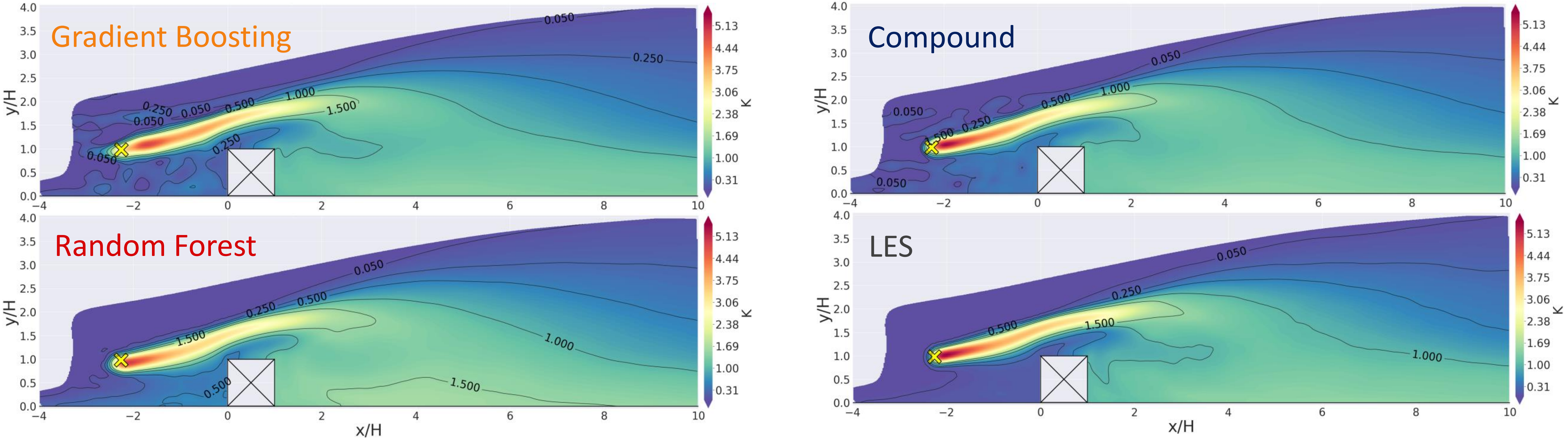
RESULTS

Compound prediction fields

Gradient Boosting
- Noisy prediction near the source
- Underestimation of peak concentrations

Random Forest
- Smooth prediction
- shifted predictions near the source
- Underestimation of peak concentrations

Compound
- Similar results to Gaussian processes
- Noisy prediction near the source
- Good prediction of the plume structure and peak concentrations



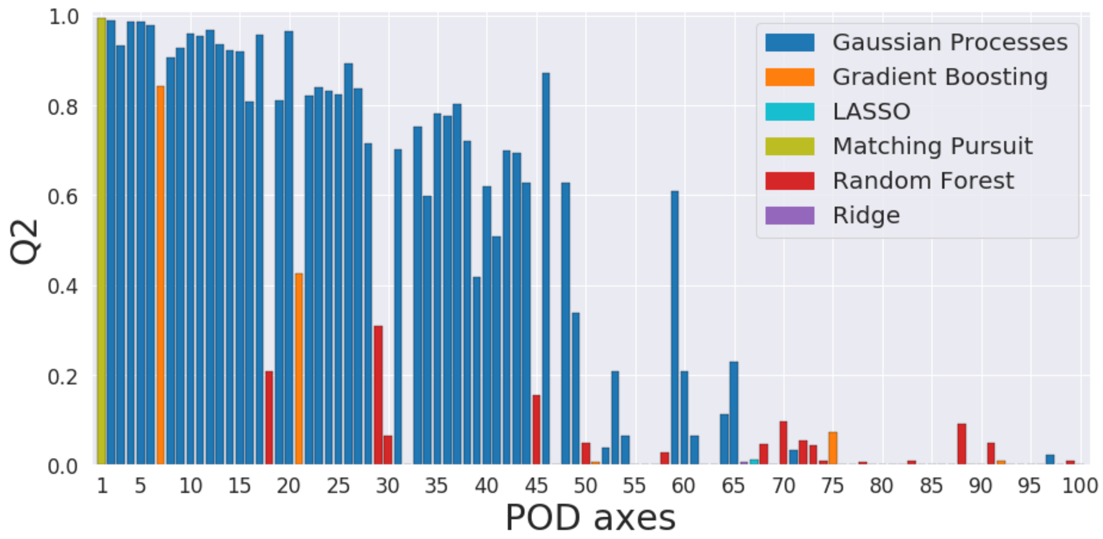
Mean concentration fields for a validation simulation with $U_{inlet} = 5.6\ m\ s^{-1}$

Robustness to the lack of training data

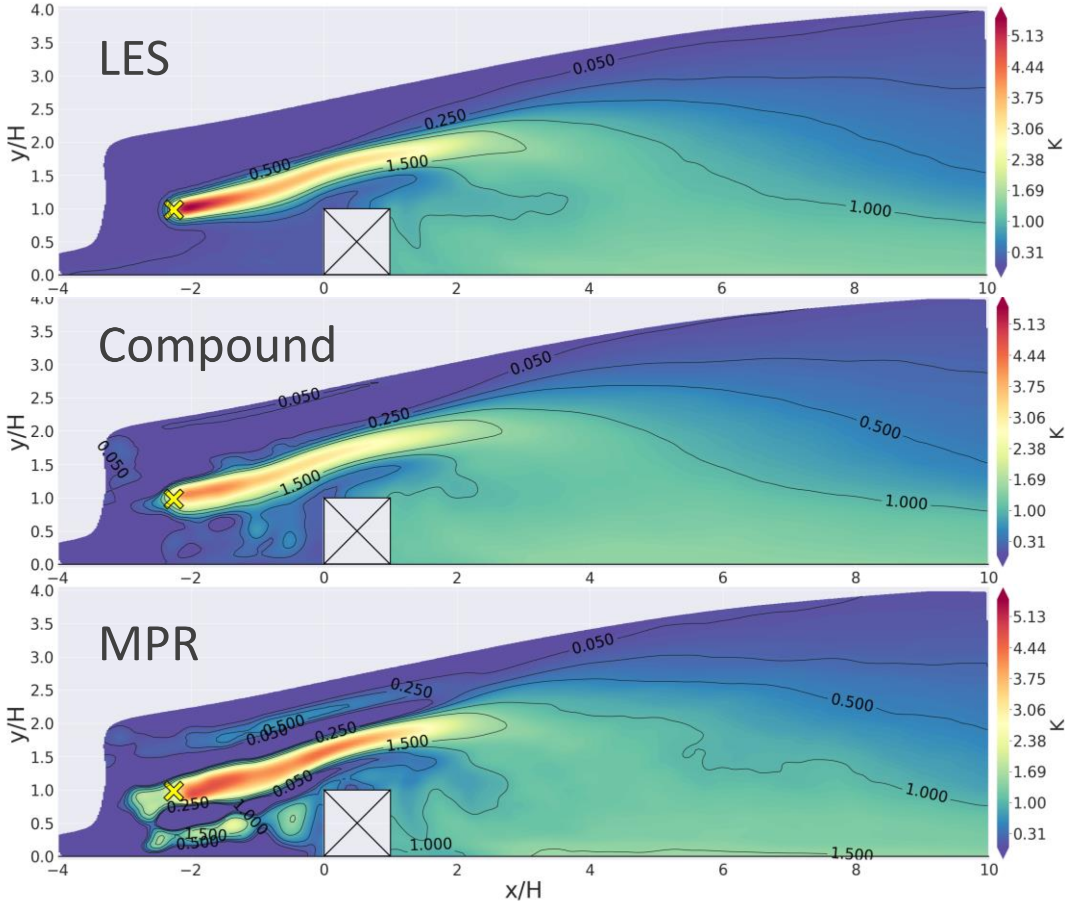
Training data is reduced to 100 LES

POD basis is reduced to 100 axes

- High POD modes can’t be considered due to lack of data
- Very noisy predictions near the emission source
- Strong errors in predicting peak concentrations



Q^2 performances on the first 100 POD axes of 4 families of metamodels



Mean concentration fields for a validation simulation with U_{inlet} = 5.6 ms^{-1}

Perspectives

Towards a real-test case: Mock Urban Test-Case/MUST

High cost of simulation: 60,000 hCPU

Issue

- 2-D test case study showed a minimum ensemble of 100 simulations was necessary for good convergence of performance statistics
- Need for reducing simulation cost

Idea: new problem decomposition

- use LES to metamodel atmospheric flows without tracer
- Simulate plume flow using cheaper CFD models (e.g. RANS)