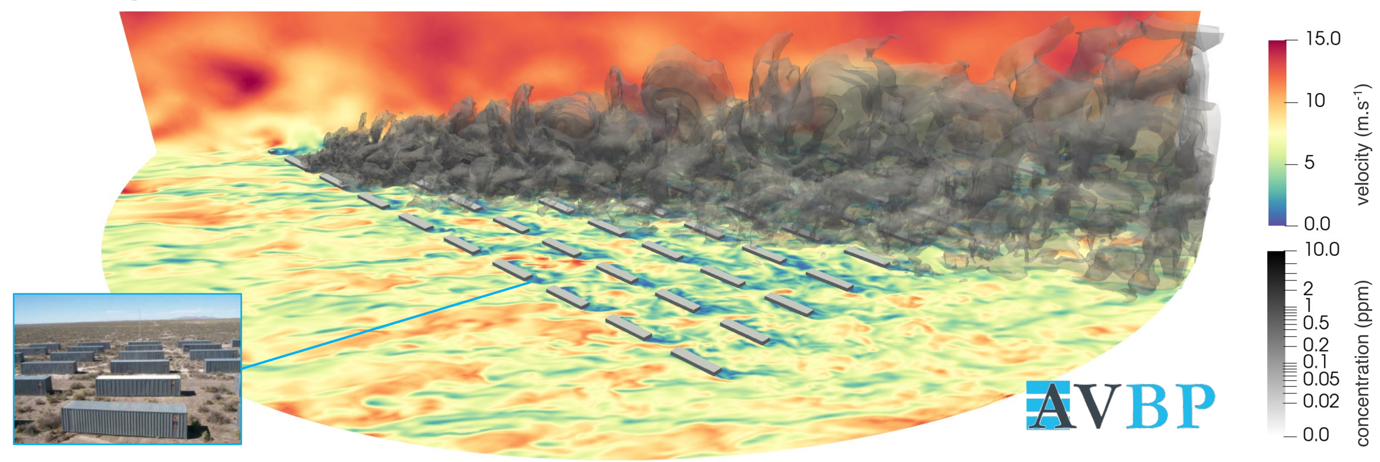
Surrogate-based data assimilation for microscale atmospheric pollutant dispersion

Eliott Lumet[1,2], Mélanie Rochoux[1], Thomas Jaravel[1] et Simon Lacroix[2]

[1] CECI, CNRS, CERFACS, Toulouse, France [2] LAAS, CNRS, Toulouse, France Contact: eliott.lumet@cerfacs.fr

1) LES for atmospheric dispersion

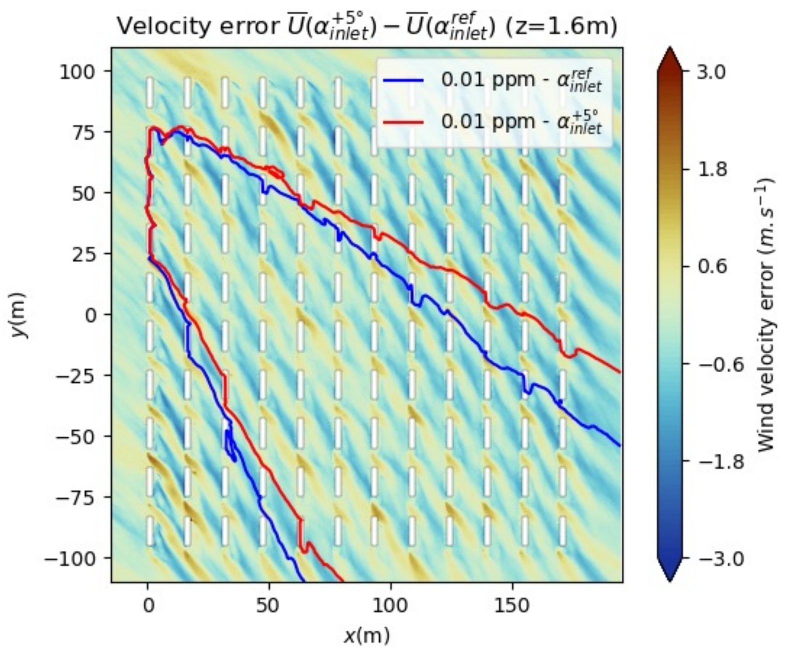
Design, evaluate and improve high fidelity models for pollutant dispersion in urban areas at microscale (≈100m)



Large-eddy simulation of propylene dispersion in an idealized urban canopy. Reproduction of the #2681829 trial of the MUST field campaign (Yee et al. 2004)

A. Why do we use Large-Eddy Simulation (LES)?

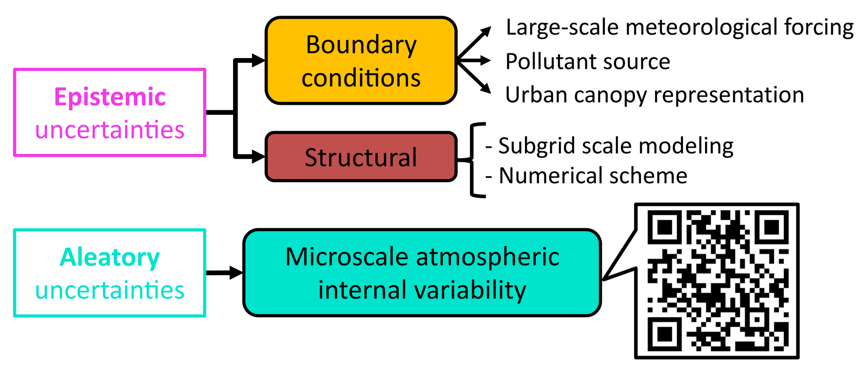
i. Explicitly takes into account the effect of the urban canopy on atmospheric flow
ii. Allows to track temporal evolution of the quantities of interest
iii. Resolves the largest turbulence scales
- ⊕ Reduced modeling uncertainty
- ⊕ Better representation the effect of atmospheric variability on the plume



Inlet wind direction perturbation effect on mean velocity and mean concentration within the array of containers of the MUST field campaign

B. Which model limitations?

- ⊖ **Cost:** One 200-s simulation ⇔ 20 000h$_{CPU}$
- ⊖ **High uncertainties** (Dauxois et al. 2021):



LES remains very uncertain despite its substantial computational cost

2) Surrogate modeling

Emulate the response surface of the LES model at reduced computing cost

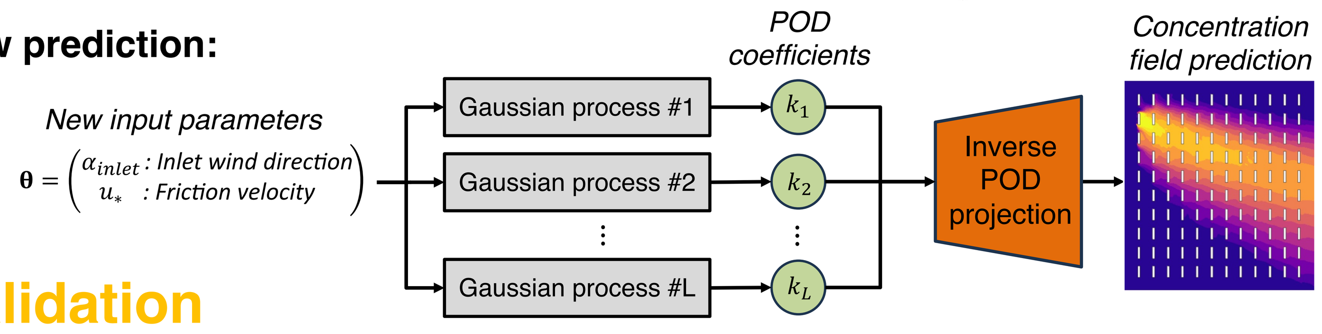
A. Learning database of LES simulations

- **Sensitivity analysis:** Wind velocity and direction ≫ rugosity, turbulence intensity, SGS model, …
- **Microclimatology** to define realistic parameter ranges
- **Input parameter space sampling** using the low-discrepancy Halton sequence

Generation of an ensemble of 200 LES for varying wind boundary conditions (≈ 6 Mh$_{CPU}$)

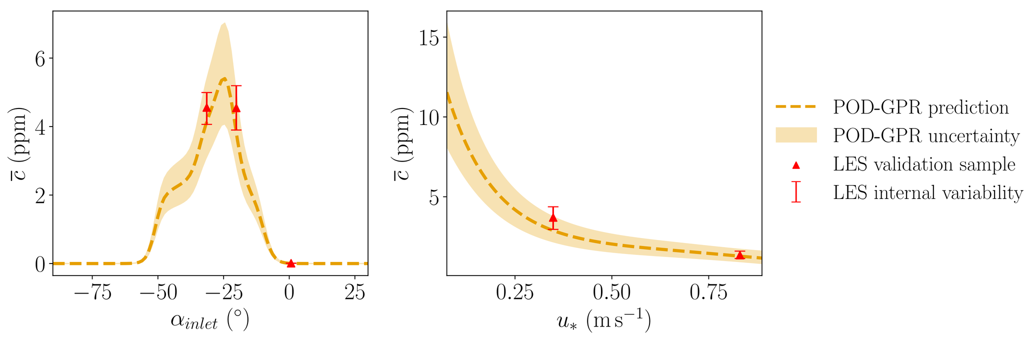
B. The POD-GPR surrogate model

- **Two-step approach:** (Nony et al. 2023) i. Dimension reduction using Proper Orthogonal Decomposition (POD) ii. Estimation of POD coefficients using Gaussian Process Regressors
- **New prediction:**



C. Validation

i. **Validation** on a test set of 40 LES shows near-maximum accuracy given the noise in the database (except for very high concentrations near the source)
ii. **Computational costs:** Prediction ≈ 0.03s versus Training ≈ 30s
iii. **Surrogate model uncertainty** = Model reduction uncertainty + Internal variability



Estimation of the concentration at a point in the domain by the POD-GPR surrogate model as a function of the two input parameters (α_{inlet}, u_*)

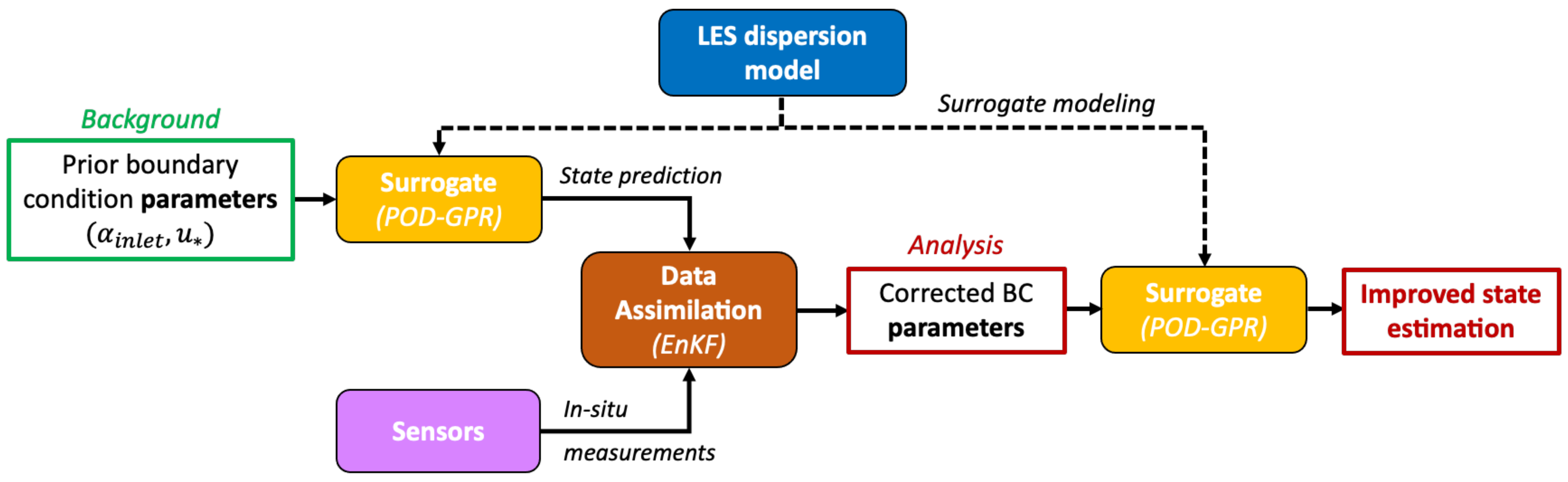
The POD-GPR surrogate model can accurately and efficiently replace the LES model while representing part of the uncertainties involved

3) Data assimilation for wind boundary condition parameters estimation

Assimilate in situ measurements to estimate large-scale wind boundary conditions and improve LES concentration field prediction

A. Data assimilation framework

- **System state:** Mean concentration field
- **Control vector:** Boundary condition parameters (α_{inlet}, u_*) as initial conditions quickly vanish at microscale (Defforge 2019)
- **Observations:** 13 concentration measurements at different locations
- **Anamorphosis** (Defforge et al. 2021)
- **Data assimilation method:** Ensemble Kalman Filter (EnKF)
- **Ensemble size:** 500 members (not a problem using the surrogate model)



B. Errors modeling

Robust and realistic errors models are necessary to ensures pertinent analysis

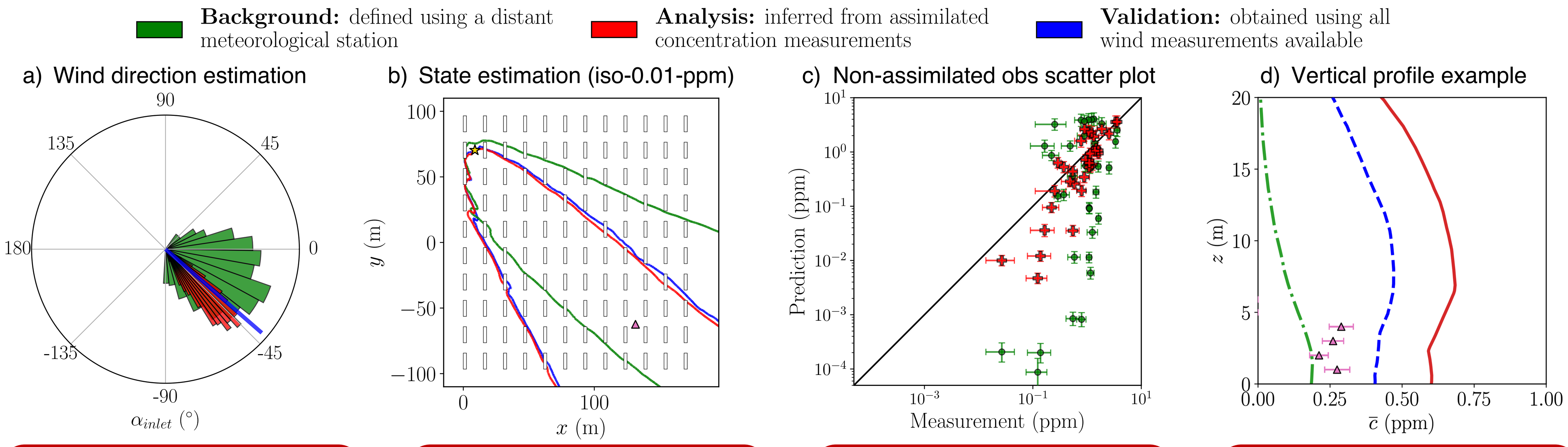
Background error B = Errors when calibrating the model boundary conditions using distant meteorological stations → Micro-climatology → $\sigma(\alpha_{inlet}^b) = 25°$, $\sigma(u_*^b) = 0.09\text{m.s}^{-1}$

Observation error R: - Instrument error (*unknown*) - Atmospheric internal variability → Bootstrap Estimation

Model error: POD-GPR uncertainty estimation: - Surrogate modeling error - Atmospheric internal variability

C. Assimilation of field measurements

Background: defined using a distant meteorological station **Analysis:** inferred from assimilated concentration measurements **Validation:** obtained using all wind measurements available



The data assimilation framework accurately estimates wind direction…

… which improves the accuracy of state prediction

Estimation is also improved for non-assimilated observation locations

Data assimilation can worsen the estimation where the LES model is biased

4) Take-home messages

- The use of a surrogate model enables real-time data assimilation and large ensemble size while also providing an estimation of the uncertainties involved
- The proposed data assimilation framework efficiently corrects wind boundary conditions and improve pollutant dispersion predictions
- **Perspectives:** i) state-parameter estimation, ii) optimal sensor placement, and iii) assimilation of plume images

Poster & materials