EGU24-17987, updated on 02 Nov 2024
https://doi.org/10.5194/egusphere-egu24-17987
EGU General Assembly 2024

Surrogate-based data assimilation for microscale atmospheric pollutant dispersion

Eliott Lumet[1], Mélanie Rochoux[1], Thomas Jaravel[1], and Simon Lacroix[2]
[1]CECI, CNRS CERFACS, Toulouse, France
[2]LAAS, CNRS, Toulouse, France

Microscale pollutant dispersion is a critical aspect of air quality assessment with significant implications for the environment and public health. Designing accurate microscale dispersion models is of paramount importance for predicting air pollution exposure and assessing risks, in particular in emergency situations such as accidents at industrial sites, which are often close to densely-populated urban areas. However, this is a challenging task, as the structure and trajectory of pollutant plumes are strongly influenced by atmospheric flow, which is inherently multi-scale, turbulent, and interacts in complex ways with the built environment. To accurately capture the airflow and dispersion patterns induced by the built environment, large-eddy simulations (LES) are recognized as a high-fidelity numerical approach. However, LES are very costly and remain subject to uncertainties, partly due to the lack of knowledge and variability of the large-scale atmospheric forcing. In emergency situations, it is essential to quantify and reduce these uncertainties in order to better predict where pollutant concentration peaks occur.

In this work, to cope with the computational cost of LES, while providing the best possible information on the processes involved, we design and validate a data assimilation algorithm based on an ensemble Kalman filter (EnKF) that combines in situ concentration measurements with LES information. This LES information is obtained through a surrogate model, based on proper orthogonal decomposition (POD) combined with Gaussian process regression (GPR), which was trained in an offline stage using a large dataset of LES simulations, and which predicts the time-averaged concentration spatial fields for varying large-scale atmospheric conditions. The application of our data assimilation approach to the MUST field-scale experiment provides a proof-of-concept of the system's ability to reduce meteorological parametric uncertainties, correct bias in the model boundary conditions and thereby improve LES pollutant concentration field predictions. The use of the POD-GPR reduced-order model enables generating ensemble predictions that accurately account for the strong nonlinearities of the LES model, in just a few tens of seconds.

In addition, particular attention is paid to the representation of the errors, in particular to the internal variability of the atmospheric boundary layer that induces variability in the LES statistical fields and in the in-situ measurements. We design a bootstrap approach to quantify the significant effect of atmospheric internal variability on microscale dispersion, and we show that GPRs are able

to learn this source of noise. Finally, we take internal variability into account in the data assimilation system considering that it is a source of model error and of observation error. This provides a more robust data assimilation framework with a more realistic description of the errors, which will be of interest for dispersion applications in real urban areas.