19th international EnKF workshop (2024)

Reduced-cost EnKF for parameter estimation of microscale atmospheric pollutant dispersion models

Eliott Lumet[1], Mélanie Rochoux[1], Thomas Jaravel[1] et 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 environmental and public health implications. The development of accurate microscale dispersion models is of paramount importance for predicting air pollution exposure and assessing risks, particularly in emergency situations such as accidents at industrial sites, which are often located close to densely populated urban areas. However, this is a challenging task because the structure and trajectory of pollutant plumes are strongly influenced by atmospheric flow, which is inherently multi-scale and turbulent, and interacts in complex ways with the built environment. To accurately account for these effects, there is a growing consensus in the research community for the use of high-fidelity, building-resolved models such as large-eddy simulations (LES). However, LES are very expensive and remain subject to uncertainties, particularly due to the lack of knowledge of large-scale atmospheric forcing and variability. In emergency situations, where we need to predict the location of pollutant peaks, it is essential to reduce forecast time and to estimate and quantify uncertainties in order to quickly cover different dispersion scenarios.

In this work, we design a reduced-cost data assimilation based on an ensemble Kalman filter (EnKF) that combines in situ concentration measurements with LES model predictions to reduce uncertainty in large-scale atmospheric forcing parameters. To reduce the computational cost, a surrogate model based on proper orthogonal decomposition (POD) combined with Gaussian process regression (GPR) is trained in an offline stage using a large dataset of LES simulations of mean concentration fields and replaces the LES model in the EnKF prediction step. The assimilation of measurements from the MUST field-scale experiment provides a proof-of-concept of the system's ability to reduce meteorological parametric uncertainties, correct model boundary condition biases, and thereby improve LES pollutant concentration field predictions. The use of the POD-GPR surrogate model reduces the cost of a 500-member EnKF cycle to a few tens of seconds.

In addition, we show that background sampling and the anamorphosis parameter used to normalize the concentration measurements significantly affect the EnKF estimates. We also use a bootstrap approach to quantify the uncertainty induced by the microscale internal variability of the atmospheric boundary layer. This allows us to provide a realistic non-diagonal observation error covariance matrix model and to account for the model prediction error thanks to the ability of the POD-GPR surrogate to learn this form of aleatory uncertainty from the LES ensemble. Overall, 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.