APPLYING MACHINE LEARNING TO REPLICATE LARGE-EDDY SIMULATION RESULTS ON URBAN POLLUTANT DISPERSION

M. KURPPA[1], M. LANGE[2], H. SUOMINEN[2], E. OIKARINEN[2], R. SAVVIDES[2], L. JÄRVI[1,3] and K. PUOLAMÄKI[1,2]

[1]Institute of Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Finland.
[2]Department of Computer Science, University of Helsinki, Finland.
[3]Helsinki Institute of Sustainability Science, University of Helsinki, Finland.

Keywords: LES, machine learning, air quality, urban

INTRODUCTION

Urban pedestrian-level air quality is characterised as highly variable. This stems especially from temporally and spatially variable pollutant sources, such as cars and chimneys, and the turbulent nature of flow that transports air pollutants. Urban structures, i.e., buildings and vegetation, effectively modify the flow and generate turbulence.

In order to resolve the complex pollutant concentration fields, high-resolution modelling is needed, for which the large-eddy simulation (LES) is the most promising method. However, to apply LES for example for urban planning purposes, both high computational resources and expertise are needed.

To develop faster and computationally cheaper predictive models, application of machine learning techniques in air quality studies has increased rapidly in the past few years. Several studies have used multiple linear regression (MLR, see Krecl et al. (2019) for a review) or more advanced methods such as Random Forest (e.g., Krecl et al. (2019)) or deep neutral networks (Kim et al., 2019) to predict air pollutant concentrations based on observed values. Still, high-resolution LES that implicitly resolves the impact of urban structures on pollutant dispersion has not yet been employed in predictive model development.

This study utilises LES data output from Kurppa et al. (2018) with the aim to develop a predictive statistical model for hourly averaged street level pollutant concentrations in a complex urban environment. Furthermore, the importance of different predictor variables is assessed.

METHODS

Kurppa et al. (2018) investigated the impact of city-block orientation and variation in the building height and shape on the dispersion of traffic-related pollutants using the LES model PALM (Maronga et al., 2015). Specifically, LES simulations were conducted applying four alternative city-planning solutions and under two different meteorological conditions. Air pollutant dispersion was studied by applying a Lagrangian particle model and representing air pollutants as inert air parcels that follow the air flow and do not interact with any surface.

As a first step, different type of features or predictor variables of the simulation environment were defined for each (x, y)-point in the simulation domain, such as street width, height of the nearest building and particle emission strength. Then, different predictive models, including linear models, regression tree, and Random Forest, were trained and tested with different data combinations.

CONCLUSIONS

Fig. 1 display an example on the hourly-averaged particle density in one urban planning scenario at $z = 4$ m from LES and the corresponding prediction using regression tree algorithm, respectively. The wind is from south-west.

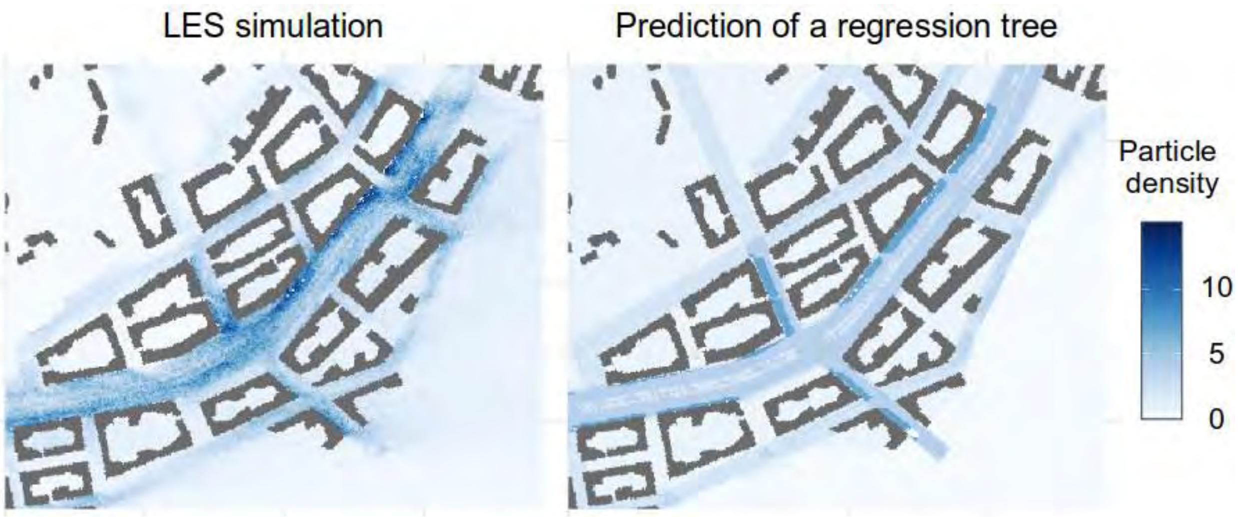

Figure 1: Hourly averaged particle densities in the city planning solutions V_{J-J} in Kurppa et al. (2018) at 4 m above ground. Left: LES simulation output. Right: prediction using a regression tree algorithm. Wind is from south-west.

In this example, the prediction captured successfully the accumulation of particles on the western side and northern part of the boulevard as well lower concentrations below street trees in the middle of the boulevard. However, other details were not predicted. The most important predictor is the particle emission, which was generally observed with all models. Nevertheless, features such as whether a point (x, y) is within a courtyard or not, or tree height were shown important.

The study is on-going and further investigations are needed to find the most accurate model and most important features, as well as to reliably estimate the accuracy of the estimates in new environments. The learned models may offer a way to replace computationally costly LES simulations in some applications. The models could be also be used to help to understand the relevant emergent interactions relevant to the spread of pollutants in urban environment.

ACKNOWLEDGEMENTS

This work was supported by the Doctoral Programme in Atmospheric Sciences (ATM-DP, University of Helsinki), the Academy of Finland (decisions 326280 and 326339) and Centre of Excellence (project no. 307331), and Helsinki Metropolitan Region Research Program.

REFERENCES

Kim et al. (2019). Development of a daily PM_{10} and $PM_{2.5}$ prediction system using a deep long short-term memory neural network model. *Atmospheric Chemistry and Physics*, **19**, 12935-12951.

Kurppa et al. (2018). Ventilation and Air Quality in City Blocks Using Large-Eddy Simulation–Urban Planning Perspective. *Atmosphere*, **9**(2), 65.

Krecl et al. (2019). Modelling urban cyclists' exposure to black carbon particles using high spatiotemporal data: A statistical approach. *Science of the Total Environment*, **679**, 115-125.

Maronga et al. (2015). The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives. *Geosci. Model Dev.*, **8**, 2515–2551.