Clustering and Classification techniques applied to pollutant dispersion

Project description

In typical computational fluid dynamics (CFD) simulations, the concentration of a passive scalar is computed through a convection-diffusion equation that includes modelling assumptions and parameters. In particular, for a Reynolds-Averaged Navier Stokes (RANS) simulation, the concentration of a pollutant C is computed by solving the following partial differential equation

$$U_i \frac{\partial C}{\partial x_i} - D_m \frac{\partial^2 C}{\partial x_i^2} - \frac{\partial}{\partial x_i}\left[\frac{\nu_T}{Sc_T}\frac{\partial C}{\partial x_i}\right] = \dot{C}_0 \tag{1}$$

where U_i are the mean velocity components, D_m is the molecular diffusion, ν_T is the turbulent viscosity, Sc_T is the turbulent Schmidt number and $\dot{C}_0$ is a pollutant source. The parameter that is of particular interest in this project in the turbulent Schmidt number as the latter is the result of modelling assumptions and influences strongly the concentration field. The optimal formulation of Sc_T is still an open question in the community and would require further efforts to improve our understanding.

In this project, you will apply different clustering and classification algorithms on Sc_T (and thus C) to investigate some potential local formulations. The dataset corresponds to time-averaged quantities extracted from a high-fidelity CFD simulation (see more details in [1]). The data are available in two different planes, namely a vertical $y = 0$ m and an horizontal one $z = 0.035$ m (see Fig 1). The dataset is composed of 25 variables (columns) and 32520 observations (rows) in the vertical plane and 71298 observations (rows) in the horizontal plane. A description of the variables is available in the `ReadMe` file enclosed with the project assignment.

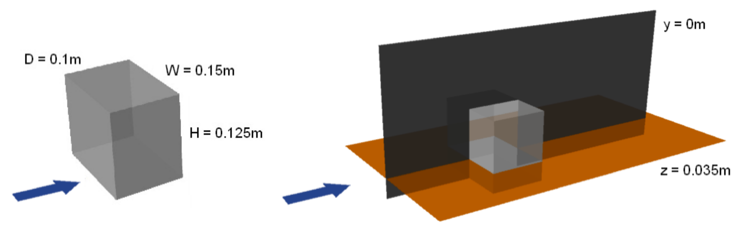

Figure 1: Schematic representation of the ground mounted building with the two planes $y = 0$ m and $z = 0.035$ m.

Tasks

In this project, you will complete the following tasks:

Task 1: The very first task is to familiarize yourself with the dataset by representing it in a readable form: use contour plots, graphs, histogram or scatter plots. You should already highlight some potential correlations and understand relationships between variables. In the present project, we are interested in the turbulent Schmidt number and the deduced concentration field, please focus your analysis on those two points. It could also be interesting to restrict the analysis to a relevant zone with respect to the geometry or the physics, if you do so, please justify and comment your choice.

Task 2: Apply the different clustering algorithms and classification that you have been introduced to on both planes. In particular, use the different techniques that you learnt during the practical sessions (e.g k-means, VQPCA, etc). Highlight any potential groups of similar regimes if there are any.

Written report

Your written report should contain the following:

Section 1: A high-level introduction and description of clustering algorithms you used and their mathematical formulations. You should also describe important hyper-parameters of the technique and what they mean.

Section 2: Results corresponding to **Tasks 1-2**. This includes any relevant figures and detailed discussion of the results.

Section 3: Final discussion and conclusions, where you make broader comments on what you found out about clustering algorithms in this project. In particular, suggestions should be made for further investigations.

As you will create many visualizations in this project, you report is expected to have many figures to support your findings, please use suitable format i.e not JPG. Make sure that you cite the relevant literature whenever needed.

If you encounter problems in the download of the data or installation of the libraries you can contact Leo Cotteleer on TEAMS or at leo.cotteleer@ulb.be. Moreover, other data planes are available (see `ReadMe`), if you want to further the analysis, feel free to ask and we will provide you additional resources!

References

[1] Riccardo Longo, Aurélie Bellemans, Marco Derudi, and Alessandro Parente. A multi-fidelity framework for the estimation of the turbulent schmidt number in the simulation of atmospheric dispersion. *Building and Environment*, 185:107066, 2020.