# Deep Learning associated with Computational Fluid Dynamics to predict pollution concentration fields in urban areas

Xavier Jurado[1,2], Nicolas Reiminger[1], Marouane Benmoussa[1], José Vazquez[1,2], and Cédric Wemmert[2]

[1] AIR&D, Strasbourg, France
[2] University of Strasbourg, ICube Laboratory, France

**Abstract.** air quality is a worldwide major health issue, as an increasing number of people are living in densified cities. Several methods exist to monitor pollution levels in a city, either physical models or sensors. Computational Fluid Dynamics (CFD) is a popular and reliable approach to resolve locally pollutant dispersion in urban context for its capacity to consider complex phenomenon at local scale. Nevertheless, this method is computationally expensive and is not suitable for real time monitoring over large areas and city shape that evolves permanently. To overcome this issue, a deep learning model based on the MultiResUNetarchitecture have been trained to learn pollutant dispersion from precalculated computational fluid dynamics. This model has been used in situ on an area spanning 1km$^2$ with real values from traffic and meteorological sensors in the surroundings of Strasbourg (France) and compared against the equivalent CFD results. Classic air quality metrics shows that the Deep Learning model manages to have satisfying results against the CFD model. The similarity index used in the study shows a 62% similarity for a result obtained in minutes against the CFD result obtained in tenth of hours.

**Keywords:** Computational Fluid Dynamics ; Air pollution ; Machine Learning ; Deep Learning ; Real Time Assessment

## 1 Introduction

Air pollution is a critical worldwide health issue with about 8 million death related to it yearly, according to the World Health Organization (WHO) [1,2]. To tackle this issue, WHO provided pollution concentration values that should not be exceeded. In European Union, regulation has been enforced on the main air pollutant such as particulate matter or nitrogen dioxide [3]. To check if these values are respected, several measures have been implemented in France:

- New real estate project near pollutant sources such as heavy traffic roads, plants, or central heating system must study thoroughly air quality in the wanted area. However, these regulations are only applied at some particular timestamps and specific places.
- Sensor monitoring. But reliable sensors are expensive to acquire and maintain. For the entirety of Strasbourg city (around 80km$^2$), only 4 sensors are deployed to date.
- Simulation of the annual pollution dispersion on the entire city. However, models that allow large area to be simulated may not be adapted for urban areas because of buildings not taken into account.

Among the possible models of the third point, a popular approach in the scientific community is to create airborne pollutant dispersion maps in urban areas is to use Computational Fluid Dynamics (CFD) [4,5]. It allows to accurately consider a lot of different physical phenomena from building impact on the flow to solar radiation or chemical reaction. Indeed, pollutant dispersion concentration field error can reach less than 10% when compared to experimental data [6] and about 30% when compared to real life in situ experiments [7]. Nevertheless, the counterbalance of this method is that it is computationally expensive. For instance, to cover $1\mathrm{km}^2$, the method roughly needs around 30 million cells and can require a week of computation to converge on 96 CPUs. Furthermore, each time the building layout changes, it would require starting new simulations again. CFD is therefore not adapted for real time simulation, despite its great accuracy and detailed description of physical phenomena.
To accelerate the computation, an innovative solution based on deep learning was developed. The idea consists in training a neural network with pre-calculated CFD simulations, to create a new air quality model that can determine pollutant dispersion in a matter of minutes over a large area. Indeed, recent advances in deep learning for spatial information treatment with convolutional based architectures have proved to be able to solve issues, notably in semantic segmentation that was impossible before. A popular model, the MultiResUNet[8], heir of UNet[9], has proved to be particularly capable at handling spatial information. This model has been trained with about 5,000 examples of CFD results of pollutant dispersion from different urban areas. The input of the model is the 3D shape of the buildings, the wind force and direction, and the position of the roads, considered as the sources of pollution.
This deep learning model is then included in a wider system that uses real time meteorological, traffic and sensor data to map the concentration field in real time on an entire urban district.

## 2 Material and method

### 2.1 CFD air quality modeling

To train the Deep Learning architecture examples of pollutant dispersion were obtained using Computational Fluid Dynamics (CFD). The software to compute the simulation is OpenFoam 5.0 which is an open source software for numerical simulations of different kind such as fluid mechanics or radiation. The approach elected here to solve the air flow is a Reynold Averaged Navier Stokes (RANS) with a k-epsilon renormalization group (RNG) [10] performing unsteady simulation. For the pollutant dispersion a transport equation coupled with the air flow is used.

The boundary conditions for the upper and lateral boundaries are symmetry conditions, the ground as a wall with a rugosity of $z_0 = 0.1m$, the building as a wall condition, the outlet as a freestream, the inlet as a logarithmic wind profile law as proposed by [11].

For the meshing, the guidelines from [12] are respected with the top and lateral boundaries situated at 5H from the closest building including with H the height the highest building. The mesh is insensitive with cells of 0.5m nearest to the buildings. The model, equations and validation have been detailed in previous published paper [13] where the same approach has been described and properly validated.

### 2.2 Deep learning network

The Deep Learning network used to learn the CFD is the MultiResUNet from [8]. This network is first designed to be applied for segmentation. In this work, it has been converted to solve pollutant dispersion from fluid mechanics. The input are the distance from the pollutant source and the height of the buildings in the area and the output is the pollutant dispersion field. The final results covers an area of $100 \times 100m^2$ by AI predictions as showed in Figure 1. The details of the MultiResUNet architecture are presented in Figure 2.

Fig. 1: Input/output images for the Deep Learning model

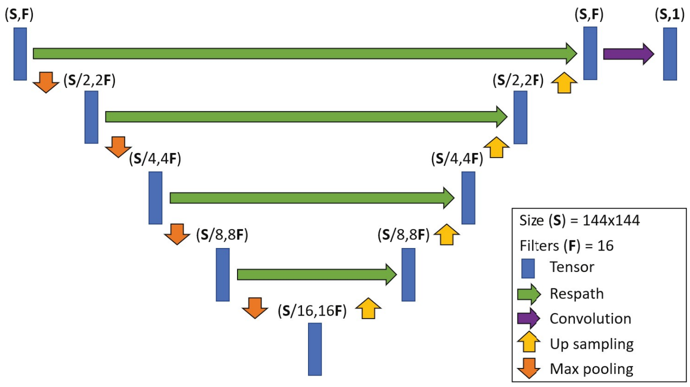



Fig. 2: Architecture details of the MultiResUNet

The loss function used is a custom loss called $J_{3D}$ and defined as followed:

$$J_{3D} = 1 - \frac{V_{pred} \bigcap V_{true}}{V_{pred} \bigcup V_{true}} \simeq 1 - \frac{min(y_i, \hat{y_i})}{max(y_i, \hat{y_i})} \tag{1}$$

where $V_{pred}$ and $V_{true}$ is the volume represented by the grayscale value of respectively the ground truth and the predicted result, $y_i$ and $\hat{y_i}$ are respectively the ground truth image and the predict deep learning result.

The dataset for the training and validation are made of around 5,000 examples of different CFD simulations with varying building layouts and pollution sources. 20% are used for the validation and 80% for the training. For the test to check on the AI capability of predicting pollutant dispersion field on unseen neighborhood, it will be compared with a real neighborhood presented in Section 2.3 that will be modelled in CFD. The training was made on 25 epochs with a patience of 5 epochs on the validation data.

## 2.3 Case study

The site is located in the surrounding of Strasbourg (GPS coordinates: 48.603468, 7.743355). The building layouts of the case study is obtained thanks to the open data of the city of Strasbourg which provide digital model of the whole city (https://data.strasbourg.eu). For the test case, a real life situation is used, the first of April of 2021 at the traffic peak which happens around 08:30 AM (to have the highest concentration related to road traffic in the area). The wind speed and directions were obtained using the API openWeatherMap with a wind speed of 2m/s and a wind direction 200°N.

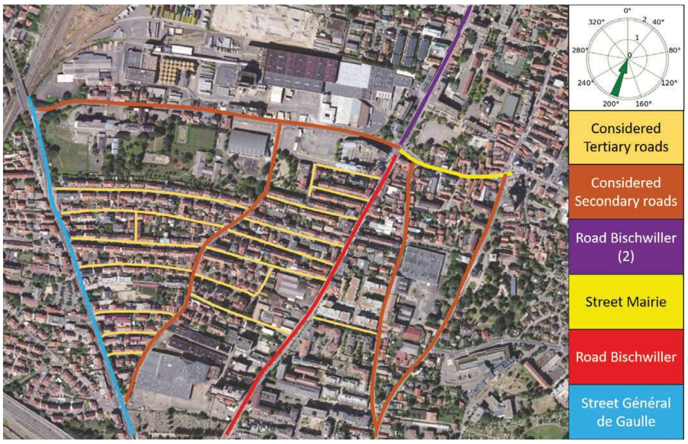



Fig. 3: Map of the Schiltigheim district with the 3 main roads used in this study

There are 27 different roads in the area. The data on traffic were obtained through the open data of the city of Strasbourg for the 4 available roads (https://data.strasbourg.eu):

- Road Bischwiller (part 1): 560 vehicles in 30 min (18.7 veh/min) with a mean velocity of 37.9km/h,
- Road Bischwiller (part 2): 784 vehicles in 30 min (26.1 veh/min) with a mean velocity of 15.5km/h,
- Street Mairie: 488 vehicles in 30 min (16.3 veh/min) with a mean velocity of 17.8km/h,
- Street General de Gaulle: 654 vehicles in 30 min (21.8 veh/min) with a mean velocity of 16.3km/h.

For other roads in the area, traffic information is lacking, thus they have been classified as secondary that will have 30% of the traffic of closest main road and tertiary that will have 5% of the closest main road. Figure 5 shows the map of the district of the study, with the three main roads and the secondary and tertiary roads. The choice of 30% and 5% is arbitrary for the sake of the example since there is no study on this traffic either with sensors or models.

Emissions are calculated based on methods proposed by the European Environment Agency (EEA) in their "EMEP/EEA Air pollutant emission inventory guidebook 2016", Tier 3 method for engine-related NOX, PM10 and PM2.5 emissions (hot and cold emissions); 2017 metropolitan fleet data found in the "OMINEA" databases provided by the Centre Interprofessionnel Technique d'Études de la Pollution Atmosphérique (share of different vehicle types, fuels and EURO standards in France).

The whole neighborhood have been modeled at once with CFD spanning an area of 1 $km^2$ made of 28 million cells. The buildings as well as the velocity magnitude field at an height of 1.5m is shown on Fig. 4.

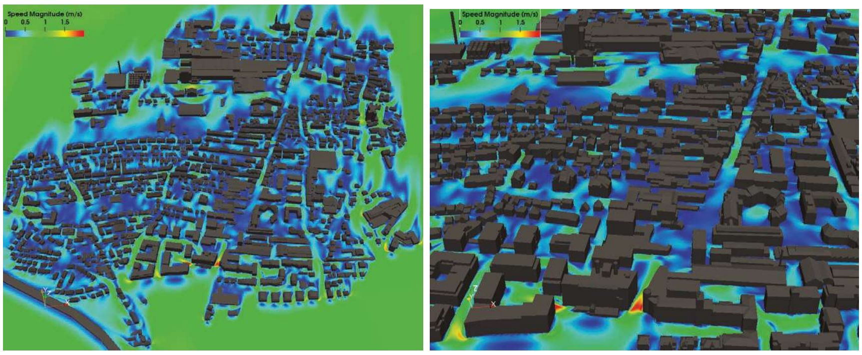



Fig. 4: Building layouts and flow field at an height of 1.5m

### 2.4 Evaluation

Seven metrics will be used, 4 from the air quality domain and three others from the computer vision. The air quality criteria have been chosen according to [14] in which the authors present several metrics with some overlapping since they evaluate the same aspect of the model. They also provides empirical threshold to consider a model as making good predictions:

- Fraction of predictions within a factor of two of observation, noted $FAC2$, a good model should respect $\simeq> 0.5$,

$$FAC2 = \text{fraction of data that satisfy } 0.5 < \frac{C_{pred}}{C_{ref}} < 2 \tag{2}$$

- Normalised Mean Squared Error, noted NMSE, a good model should respect NMSE $\simeq< 1.5$,

$$NMSE = \frac{\overline{(C_{ref} - C_{pred})^2}}{\overline{C_{pred} C_{ref}}}, \tag{3}$$

- Fraction Bias noted FB, $|\text{FB}| < 0.3$,

$$FB = \frac{(\overline{C_{ref}} - \overline{C_{pred}})}{0.5(\overline{C_{pred}} + \overline{C_{ref}})}, \tag{4}$$

- Correlation coefficient, noted $R$ (no threshold is given for this parameter),

$$R = \frac{\overline{(C_{ref} - \overline{C_{ref}})(C_{pred} - \overline{C_{pred}})}}{\sigma_{C_{pred}} \sigma_{C_{ref}}}, \tag{5}$$

The three other metrics are:

– $J_{3D}$

$$J_{3D} \simeq \frac{min(C_{ref}, C_{pred})}{max(C_{ref}, C_{pred})} \tag{6}$$

– Relative mean absolute error $MAE_{rel}$

$$MAE_{rel} = \frac{\overline{|C_{ref} - C_{pred}|}}{\overline{C_{pred}}} \tag{7}$$

– Structural similarity $SSIM$

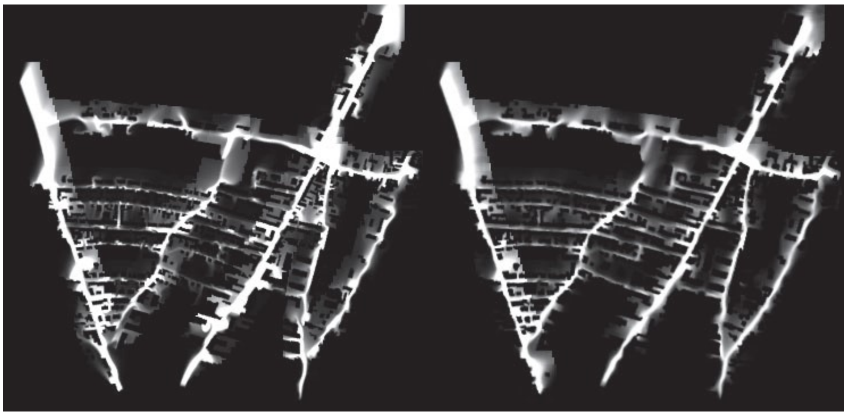
$$SSIM(A, B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c2)}{(\mu_A^2 + \mu_B^2 + c1)(\sigma_A^2 + \sigma_B^2 + c2)} \tag{8}$$

$$c_1 = (k_1 L)^2 \quad c_2 = (k_2 L)^2 \tag{9}$$

with $C_{pred}$ the model prediction concentration, $C_{ref}$ the reference concentration (ground truth), $\mu_A$ and $\mu_B$ are the respective average of A and B, $\sigma_A^2$ and $\sigma_B^2$ are the respective variances of A and B, $\sigma_{AB}$ is the covariance of $A$ and $B$, $L$ is the dynamic range of the pixel values and $k_1$ and $k_2$ are two constants respectively 0.01 and 0.03 (by default).

## 3 Results

To evaluate the deep learning capabilities to be applied in real life situation, a comparison has been made with real world data at the traffic at 08:30AM in the south of Schiltigheim, France the first of April 2021 between results from a CFD simulation and our deep learning approach on the $NO_x$ dispersion from traffic emissions. The results proposed respectively by the CFD and MultiResUNet for the whole neighborhood are shown on Fig.5



(a) CFD result (b) MultiResUNet result

Fig. 5: Maps of the studied district and comparison of the two results)

It can be tedious to compare the results between the CFD and the deep learning network since the CFD determines the dispersion in 3D while the deep learning approach works in 2D only at a given height. Nonetheless, the CFD needed one week of computation on 96 CPU while the deep learning network needed around 3 minutes on a GTX 1080Ti GPU, representing a speed up by x3000. To evaluate the accuracy of the predictions, the metrics presented above were computed between the prediction and the CFD considered as the ground truth and are presented below on Table 1.

| **Metrics** | $FAC2$ | $NMSE$ | $FB$ | $R$ | $MAE_{rel}$ | $J_{3D}$ | $SSIM$ |
|---|---|---|---|---|---|---|---|
| **Score** | **0.818** | 1.565 | **0.176** | 0.851 | 0.431 | 0.620 | 0.768 |
| **Expected values** | $> 0.5$ | $< 1.5$ | $< 0.3$ | 1 | 0 | 1 | 1 |

Table 1: Evaluation of the quality of the dispersion model given by the deep learning approach.

## 4 Conclusion

As demonstrated by our work, deep learning has proved to be able to predict results close to CFD for air pollutant dispersion. Moreover, the MultiResUNet architecture was able to compute the dispersion in a matter of minutes over a wide area against several days for the CFD. This makes the Deep Learning approach a potential model to predict in real time over large scale the pollutant dispersion from traffic related pollution.

## References


1. WHO: Mortality and burden of disease from ambient air pollution, Global Health Observatory data (2016)
2. WHO: Mortality from household air pollution, Global Health Observatory data (2016)
3. EU: Directive 2008/50/EC of the european parliament and of the council of 21 May 2008 on ambient air quality and cleaner air for Europe. European Union. (2008)
4. Reiminger, N., Jurado, X., Vazquez, J., Wemmert, C., Blond, N., Dufresne, M., Wertel, J.: Effects of wind speed and atmospheric stability on the air pollution reduction rate induced by noise barriers. Journal of Wind Engineering and Industrial Aerodynamics **200** (May 2020) 104160
5. Santiago, J.L., Martilli, A., Martin, F.: On dry deposition modelling of atmospheric pollutants on vegetation at the microscale : application to the impact of street vegetation on air quality. Boundary-Layer Meteorology **162** (2017) 451–474
6. Reiminger, N., Vazquez, J., Blond, N., Dufresne, M., Wertel, J.: How pollutant concentrations evolve in step-down street canyons as a function of buildings geometric properties. (2019)
7. Rivas, E., Santiago, J.L., Lechón, Y., Martín, F., Ariño, A., Pons, J.J., Santamaría, J.M.: CFD modelling of air quality in Pamplona City (Spain): Assessment, stations spatial representativeness and health impacts valuation. Science of the Total Environment (2019) 19
8. Ibtehaz, N., Rahman, M.S.: MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. Neural Networks **121** (January 2020) 74–87 arXiv: 1902.04049.
9. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs] (May 2015) arXiv: 1505.04597.

10. Yakhot, V., Orszag, S.A., Thangam, S., Gatski, T.B., Speziale, C.G.: Development of turbulence models for shear flows by a double expansion technique. Physics of Fluids A: Fluid Dynamics **4**(7) (July 1992) 1510–1520
11. Richards, P.J., Hoxey, R.P.: Appropriate boundary conditions for computational wind engineering models using the k-E turbulence model. (1993) 9
12. Franke, J., Hellsten, A., Schlünzen, H., Carissimo, B.: Best practice guideline for the CFD simulation of flows in the urban environment. COST Action 732 (2007)
13. Reiminger, N., Vazquez, J., Blond, N., Dufresne, M., Wertel, J.: CFD evaluation of mean pollutant concentration variations in step-down street canyons. Journal of Wind Engineering and Industrial Aerodynamics **196** (January 2020) 104032
14. Chang, J.C., Hanna, S.R.: Air quality model performance evaluation. Meteorology and Atmospheric Physics **87**(1-3) (September 2004)