A review of Machine Learning CFD-based predictive modelling for urban air pollution.

Leonardo de Souza Sá [a].

[a] School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, leonards@eq.ufrj.br.

Abstract. This work aims to review the evolution in the study of Machine Learning CFD-based predictive modelling in general, highlighting its deficiencies and the possibility for further improvements in this area. Specifically, this paper analyses the growth of this conjoint technique applied to real-time forecasting simulations of pollutant dispersion in an urban mesh with the objective of predicting the behaviour of the flow in less time, improving performance and potentially reducing simulation costs.

Keywords. CFD, Machine Learning, Artificial Intelligence, Pollutants dispersion.

1. Introduction

Predictive models have become essential tools to avoid or mitigate tragic events, such as gas leakage from multipurpose industries. Different tools exist nowadays in an effort to construct robust models to prevent this kind of scenario. Between those is Computational Fluid Dynamics (CFD), a technique used in many scientific areas to study the behavior of fluids using numerical methods and strong algorithms that allow a deep understanding of the flow dynamics in a certain space-time frame.

This technique is of high value to the analysis of gas dispersion aiming to avoid the exposure of the population to high concentrations of pollutant molecules. Using not only CFD techniques but also Machine Learning (ML) methods together is essential to speed up the simulation process as well as to understand the scenario. In this sense, this combination might become a useful tool to improve a city management system and thus increase the welfare of its population.

Gas emissions can be very dangerous to human health and affect all countries, developed and underdeveloped. In most cases, the effect is only an olfactory discomfort due to strong odors but can rapidly scale up to acute and chronic respiratory diseases (e.g. lung cancer), heart disease, stroke and problems with the fetus. A recent study by the World Health Organization (WHO) showed that ambient air pollution was responsible for the premature death of 4.2 million people around the world in 2019 (WHO, 2022).

2. Objectives

2.1 General Objectives

The main objective of this paper is to review the most recent information available concerning Machine Learning CFD-based predictive modelling of airflow in an urban mesh observing the evolution of knowledge in the past few years.

2.2 Specific Objectives

1. Comprehend the evolution of the number of publications in ML and CFD;

2. From the results generated in the search, allow anyone to have a major understanding of the subject by listing the main findings;

3. Analyze from recent studies how the prediction of the fluid flow could help infer the pollutant dispersion in an urban mesh for the case of a street canyon.

3. Methodology

The methodology proposed by this paper consists in highlighting the most modern and relevant studies in the two major areas of this review (CFD and ML), combining these two techniques to analyze pollutant dispersion/air quality in an urban mesh. For this first part, the methodology will leverage the research functionalities provided by state-of-the-art research tools. The aim is to curate a collection of recent studies on the emulation framework, encompassing various articles within a specified time frame. These studies will be evaluated based on their relevance through citations and the temporal evolution of publications, presented through organized tables and graphical representations. This approach is intended to deeply understand each facet by harnessing the data from recent papers, thereby paving the way for the thorough development of a novel method.

3.1 Using Dimensions search platform

This first simple search consists of consulting the Dimension® database to analyze the evolution of publications in CFD and ML. Dimensions® is an accessible online platform and the world's largest linked research database with more than 138 million publications background. The search consists of obtaining general graphs both from the number of publications in the area of ML and CFD (with a focus on air pollution) separately from 2015 and onward. Results of this search using the input "CFD AND (POLLUTANT DISPERSION OR AIR QUALITY)" in the categories "Title and abstract" correspond to the term "MACHINE LEARNING" searched in "Title and Abstract" in Dimension®'s database starting from 2015.

3.2 Advanced search using Scopus

An advanced search in the Scopus® platform by Elsevier® using Boolean operators (e.g. AND, OR and NOT), as well as the specific commands of the tool is an easy way of concentrating data to the most recent studies with relevance to the field of CFD-based ML tools applied to forecasting pollutant dispersion. For example, a general search in the Scopus® platform using the command TITLE-ABS-KEY("x") gives as a result all the documents in the Scopus® database containing the "x" word in the title of the document, in the abstract or the keywords. To restrict the type of document to articles we use the command DOCTYPE(ar), which will make the search results restricted to articles only.

Tab. 1 – CFD and machine learning related articles search at Dimensions®.

Search query
TITLE-ABS-KEY("CFD" AND "machine learning") AND DOC-TYPE(ar)

3.3 Litmap® platform

This approach consists of using the online platform Litmap® to construct a graph summarizing, for a certain group of papers, the relation between them in time, the number of citations and the citations in between them. Figure 1 illustrates how Litmap® works if we create a seed based on the reference no.[2] in the Bibliography. A circle represents an article and is accompanied by the name of the principal author and the year of publication. The circles that are connected by lines represent that there exists citations in-between them.

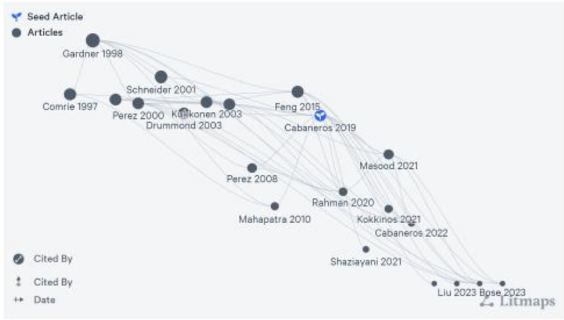

Fig. 1 - Example of using Litmap® platform for a search based on the reference no.[2]. Produced using litmap.com.

This is a good approach to better visualize the entirety of papers one is using for their research or even to find new articles connected to the one(s) you have chosen, as the artificial intelligence of Litmap® suggests you if needed. In our case, we will use the 12 articles obtained using the search input in Table 2 to create our own customized map.

Tab. 2 - Input for CFD and machine learning-related articles since 2018 relating to pollutant dispersion/air quality.

Search query
TITLE-ABS-KEY("CFD" AND "machine learning") AND TITLE-ABS-KEY("pollutant dispersion" OR "air quality" OR "air pollution") AND DOCTYPE(ar) AND PUBYEAR AFT 2017

4. Results and Discussion

4.1 Dimensions® platform

A more general search at Dimensions.ai using the input "CFD AND (POLLUTANT DISPERSION OR AIR QUALITY)" in the categories "Title and abstract" (see Table 1) gives us an overview of the effort that has been put into this area of study, represented by an augmentation in the number of articles since 2015 even though the quantity of papers is yet low showing that there is still space for growth.

Indeed, the total number found was, since 2015, 455 according to the search tool Dimensions®. The choice of using the platform Dimensions® is due to its accessibility since it has a free version. This would give an alternative to those who are not willing to pay in a very low stage of research for well-known platforms such as Scopus® but still want to have an overview about any scientific topic.

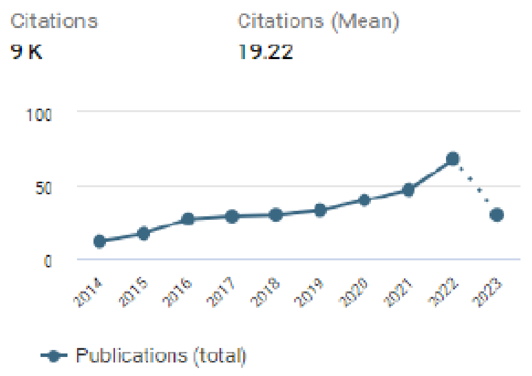

Fig 2 - Citations since 2015 using the given input. Extracted from Dimensions.ai.

The use of machine learning in many scientific areas has exploded in the past few years as a very promising tool. In Figure 2, one can observe how the number of publications has been constantly increasing, passing over 1.4 million publications since 2015. Once again the term "MACHINE LEARNING" was searched in "Title and Abstract" in Dimension®'s database.

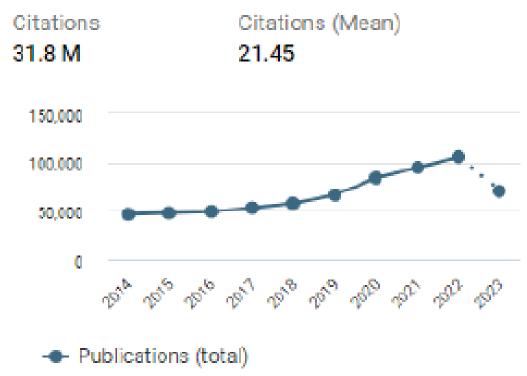

Fig. 2 - Number of publications in the area of Machine Learning since 2015. Extracted from Dimensions.ai.

4.2 Advanced search using Scopus®

The first search in articles containing the words CFD and machine learning gives a total number of 597 documents found. This being stated, it can be created a new and better search entry (always in Scopus®) taking into account only articles from the last 5 years (since 2018). This will help narrow the search to the most recent scientific knowledge. The result is obtained and shown in Table 3.

Tab. 3 - CFD and machine learning-related articles since 2018.

Search query	Result
TITLE-ABS-KEY("CFD" AND "machine learning") AND DOC-TYPE(ar) AND PUBYEAR AFT 2017	580

As expected by the past figure, the number is still high since the trend of this study is recent. Now to focus on the subject of our interest we will add a new restriction that is to the scenario of CFD-based machine learning techniques applied to pollutant dispersion/air quality according to the entry query in Table 4.

Tab. 4 - CFD and machine learning-related articles since 2018 relating to pollutant dispersion/air quality.

Search query	Result
TITLE-ABS-KEY("CFD" AND "machine learning") AND TITLE-ABS-KEY("pollutant dispersion" OR "air quality" OR "air pollution") AND DOC-TYPE(ar) AND PUBYEAR AFT 2017	12

Now the result is extremely reduced to only 12 articles been published since 2018, which highlights the importance of the present thesis of fomenting new studies in this area. Research by hand was done in Google Scholar using the Keywords for the search: CFD, pollutant dispersion, and urban, which gives us an extensive list of recent studies that suits well this subject area and that would be of great base to develop more studies in the area of CFD-based ML techniques applied to pollutant dispersion in urban areas. The articles obtained using this approach highlight the significance of this research area. These served as reference throughout the paper, aiding in a deeper comprehension of the subject. All utilized information is duly listed in the Bibliography. Again, an interesting result to be observed is how the interest in this research area has grown exponentially in the last five years compared to the beginning of the decade, when almost nothing was published correlating CFD and machine learning techniques. In fact, a more general analysis of the graphs in Figure 1 and Figure 2 shows that both machine learning and CFD techniques in air quality have been constantly growing in the past almost 10 years. One can note that, since 2015, when machine learning started growing at a fast pace, a tendency in the publication of articles published using CFD-based machine learning techniques has appeared, suggesting that with the maturity of new ML techniques arriving, new methods were integrated to optimize current CFD models and also develop new ones.

4.3 Customized Litmap®

The 12 articles found by using the search query in Table 4 cover multiple areas, like Environmental Science, Engineering, Physics and Astronomy, etc. These papers are related in Table 3 and are represented by the map in Figure 3.

Tab. 3 - CFD and machine learning-related articles since 2018.

Ref.	Year	Title
[3]	2023	Optimizing block morphology for reducing traffic pollutant concentration in adjacent external spaces of street canyons: A machine learning approach
[4]	2023	A data-driven adversarial machine learning for 3D surrogates of unstructured computational fluid dynamic simulations
[5]	2023	Data-driven assessment of arch vortices in simplified urban flows
[6]	2023	Application of a Machine Learning Method for Prediction of Urban Neighborhood-Scale Air Pollution
[7]	2022	Unobtrusive Cardio-Respiratory Assessment for Different Indoor Environmental Conditions
[8]	2022	Prediction and optimization of thermal comfort, IAQ and energy consumption of typical air-conditioned rooms based on a hybrid prediction model
[9]	2022	BIM-supported sensor placement optimization based on genetic algorithm for multi-zone thermal comfort and IAQ monitoring
[10]	2021	Bim and data-driven predictive analysis of optimum thermal comfort for indoor environment
[11]	2021	Designing roadside green infrastructure to mitigate traffic-related air pollution using machine learning
[12]	2020	A Reduced Order Deep Data Assimilation model
[13]	2019	Implicit definition of flow patterns in street canyons recirculation zone-using exploratory quantitative and qualitative methods
[14]	2019	Wind field reconstruction for the dispersion modelling of accidental chemical spills on complex geometry

The circle between "Rodrigues 2022" and "Quilodran-Casas 2023" corresponds to Wai and Yu (2023). Note how the correlation that corresponds to the line connecting the dots is scarce and shows how the recent studies haven't been focusing in one single area, and mostly that it hasn't been many developments yet, which contributes to the importance of this review and propitiating new papers to be written.

Fig. 3 - Litmap of the 12 articles found.

4.4 The case of a street canyon

A recent study by Kwak et al. (2018) analyzed the concentration of common urban atmospheric pollutants (NOx, BC, pPAH and PN) in urban areas using mobile monitoring and CFD modeling. This article showed that in street canyons (streets of great length generally composed of several lanes, tall buildings around them, with poor ventilation and a large concentration of emitting vehicles), there is an important correlation between poor air quality and the presence of heavy-duty diesel vehicles (HDDVs). This correlation shows the relevance of this study since those HDDV vehicles are diesel exhausted, classified as Group 1 carcinogens by the International Agency for Research on Cancer (IARC). In fact, heavy vehicles emitting diesel are largely responsible for the emission of pollutants that are later dispersed into the urban mesh (50% of particle number (PN), and 60% of particulate matter emissions). The study uses the RANS model, one of the most common, with the re-normalization group (RNG) k-.turbulence closure scheme to simulate the gas dispersion in a street canyon in the Seoul (South Korea) metropolitan area. The high concentration of tall buildings in the entourage of the main roadway is responsible for enclosing the flow in the longitudinal path, with few concurrent streets. This study can then be interesting to help us understand how the dispersion of pollutants works in an urban mesh, as this is usually composed of a set of street canyons, with some isolated vegetation regions (e.g. squares), and a large volume of vehicle traffic. Therefore, in addition to airflow simulations around buildings, the study of street canyons is complementary and helps understand what types of measures could be taken by authorities to better manage urban traffic to avoid exposure of their population to high volumes of polluting gases. For example, it is known that along the street canyon, the airspeed has low magnitude because there are no exit routes for it since we are surrounded by tall commercial and residential buildings. On the contrary, at signs of road intersections, there are side winds that accelerate the local wind and thus contribute to the dispersion of pollutants. So, a simple solution one could think of is to avoid the construction of buildings too close to each other in a way to allow ventilation between the buildings that break this longitudinal flow in the street canyon and consequently disperse the polluting gases that concentrate and affect the population's health. Proposing an urban geometry that would allow strong winds to push away the pollutants from its source and thus from pedestrians and residences along the street canyon would be a good approach to mitigate the pollutant dispersion in an urban mesh that affects its inhabitant's health. However, studies have shown that this would only spread the pollutants to nearby living blocks and thus enlarge the exposure area. In fact, the pollutants can have a wide range of impact of more than 300m from the source. For example, the furthest diffusion distance carbon monoxide (CO) can reach from a road source is between 100m and 200m. Thus, spreading the gas would not be a real solution but would displace the issue to other areas. A more effective alternative is proposed by Wu and Chen (2023): guiding the pollutants into the UCL, lowering the concentration in the pedestrian zones. The study proposes a ML approach to find an optimized block configuration based on a data set from CFD simulations. The use of ML is known to have sped up the CFD simulations by 800 times. The modification of the block morphology changes the pathway of pollutants not only for the adjacent buildings but also the inner buildings in the block (considering only perpendicular winds from the street canyon). Adding a building of a certain height affects the pollutant concentration by the formation of accumulation zones in the street canyon.

In consequence, the study of how block morphology and block layouts contribute to pollutant dispersion remains important. An often ignored morphological indicator, the standard deviation of building height (BHstd), for example, is an important parameter to analyse the changes in mean flows in horizontal and vertical directions and consequently the pathways and rates of pollutant transport. Indeed, urban morphology is an important parameter and can not be ignored since it affects the airflow patterns, turbulence intensity and ventilation performance in urban blocks and street canyons (Wu and Chen, 2023). Many empirical studies do not show a singular linear relationship between BHstd and pollutant concentration. In fact, for a block of buildings facing a main road of a street canyon, Wu and Chen (2023) suggest that a height of 33m for buildings facing the street canyon would be the minimum for mitigating the spread of gases into the building block. It also suggests as a general method that tall buildings facing the main road and small buildings inside the block would be an optimal configuration for preventing a threatening concentration of traffic-related air pollutants inside the block. One can conclude then that a lot of discussion and research can still be produced in this area to give further insights and better options for urban planning and improvement of urban human health.

5. Conclusion

It was highlighted the importance of the simultaneous work of CFD and ML methods to provide fast results in the prediction of pollutant gas dispersion in an urban mesh and how this would be of high value for any authorities responding to emergency scenarios as well as to the increase of the population's welfare. This necessity comes from alerting global scenario of pollution/air quality in many countries from all continents. In this paper, it was reviewed recent papers on CFD-based machine learning algorithms for predicting pollutant dispersion. It has also been discussed the evolution of recent works in CFD, ML and both of them together applied to the field of pollutant dispersion, when we realized that the number of papers from the last 5 years is not very large, indicating that there is still a huge space for improvement in this research area. Nonetheless, we still were able to analyze with these results how the formation of accumulation zones and the urban mesh morphology could affect pollutant dispersion especially in a street canyon and its surroundings.

From the 12 papers we found in the combination of the subjects, not all of them were necessarily applied to an urban mesh, which narrows, even more, the information available in the scientific community to give insights into urban planning improvement. This suggests that further studies are necessary yet to understand deeply the dynamics of pollutant dispersion in urban meshes. These studies could also open ways to other fields that would improve even more the analysis, such as thermal comfort.

6. References

[1] World Health Organization. Ambient (outdoor) air pollution. 19 Dec. 2022. Available at: www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

[2] Cabaneros S.M., Calautit J.K., Hughes B.R. A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Software. 2019. pp. 285–304.

[3] Wu Y., Chen H. Optimizing block morphology for reducing traffic pollutant concentration in adjacent external spaces of street canyons: A machine learning approach. Building and Environment. 2023. Volume 242, 110587.

[4] Quilodran-casas C., Arcucci R. A data-driven adversarial machine learning for 3D surrogates of unstructured computational fluid dynamic simulations. Physica A: Statistical Mechanics and its Applications. 2023. Volume 615, 128564.

[5] Martinez-Sanchez A. et al. Data-driven assessment of arch vortices in simplified urban flows. International Journal of Heat and Fluid Flow. 2023. Volume 100, 109101.

[6] Wai K.-M., Yu P. K. N. Application of a Machine Learning Method for Prediction of Urban Neighborhood-Scale Air Pollution. International Journal of Environmental Research and Public Health. 2023. 20, no. 3: 2412.

[7] Rodrigues M. C. J., Postolache O., Cercas F. Unobtrusive Cardio-Respiratory Assessment for Different Indoor Environmental Conditions. IEEE Sensors Journal. 2022. vol. 22, no. 23, pp. 23243-23257.

[8] Hou F. et al. Prediction and optimization of thermal comfort, IAQ and energy consumption of typical air-conditioned rooms based on a hybrid prediction model. Building and Environment. 2022. Volume 225, 109576.

[9] Cheng J. C. P. et al. BIM-supported sensor placement optimization based on genetic algorithm for multi-zone thermal comfort and IAQ monitoring. Building and Environment. 2022. Volume 216, 108997.

[10] Gan V. J. L. et al. BIM and Data-Driven Predictive Analysis of Optimum Thermal Comfort for Indoor Environment. Sensors. 2021. 21, no. 13: 4401.

[11] Hashad K. et al. Designing roadside green infrastructure to mitigate traffic-related air pollution using machine learning. Science of The Total Environment. 2021. Volume 773, 144760.

[12] Quilodran Casas C. et al. A Reduced Order Deep Data Assimilation model. Physica D: Nonlinear Phenomena. 2020. Volume 412, 132615.

[13] Chatzilichailidis A. E. et al. Implicit Definition of Flow Patterns in Street Canyons—Recirculation Zone—Using Exploratory Quantitative and Qualitative Methods. Atmosphere. 2019. 10, no. 12: 794.

[14] Wang B., Qian F., Zhong W. Wind field reconstruction for the dispersion modeling of accidental chemical spills on complex geometry. Chinese Journal of Chemical Engineering. 2019. Volume 27, Issue 11, Pages 2712-2724.