Imperial College London

Department of Earth Science and Engineering

MSc in Environmental Data Science and Machine Learning

Independent Research Project
Project Plan

Modelling pollution in the urban environment using neural networks

by

Hanson Shen

Email: hs1623@imperial.ac.uk
GitHub username: edsml-hs1623
Repository: https://github.com/ese-msc-2023/irp-hs1623

Supervisors:

Dr Claire E. Heaney
Dr Christopher C. Pain

June 2024

Abstract

Air pollution poses significant health risks, including respiratory infections, heart disease, lung cancer and an increased response to allergens. In urban environments, vehicle emissions contribute substantially to pollution, releasing harmful gases and particular matter. This research aims to model urban air pollution using neural networks intergrated with computational fluid dynamics (CFD). We propose utilising AI4PDEs, an in-house CFD code that leverages neural networks to solve discretized systems of equations, while assimilating real-time traffic flow and weather data to improve the accuracy of pollution concentration predictions using Variational Autoencoders (VAEs). This study addresses limitations of traditional CFD and statical models by providing a more efficient and adaptive approach to urban air quality monitoring through dynamic data assimilation. The expected outcomes include a enhanced accuracy in pollution exposure assessments and policy-making insights.

1 Introduction

Air pollution in urban environments, primarily from vehicle emissions, poses severe health risks to humans and ecosystmes. However, accurate modelling of pollution dispersion in cities is challenging due to complex air flows and diverse pollutant sources. This research is crucial as it addresses this critical public health issue by developing a robust model to predict pollution exposure in urban areas. The intergration of computational fluid dynamics (CFD) with neural networks [1] and the assimilation of real-time data [2] using Variational Autoencoders (VAEs) [7] can revolutionise urban air quality monitoring, providing actionable insights for urban palnning. By incorporating these advancements, our research anticipates to benefit society by improving living conditions and reducing health risks associated with air pollution.

Current methods for existing modelling air pollution include traditional CFD models and statistical approaches, each with inherent limitations. Traditional models, while accurate, are computational intensive and time-consuming, requiring signigicant computing resources to attain detailed simulations of urban air flows and pollution dispersion. Our project aims to bridge this gap by using AI4PDEs, which offers a novel approach by solving discretized equations using neural netwoks without the need for extensive training [1, 6]. Statistical models, on the other hand, often lack the precision for detailed urban analysis and struggle with real-time adaptability, limiting their prcaticability in dynamic environments. Previous studies have demonstrated the potential of neural networks in CFD [1, 4, 3], but the assimilation with dynamic data remains underexplored [5]. Therefore, this research also aims to incorporate traffic and weather data into the neural network-based model, improving the feasibilitiy and accuracy of predictions. By integrating AI-driven methodologies with continuous data streams, our research strives to advance the capabilities of urban air quality monitoring.

2 Methodology

This project will utilise AI4PDEs, a neural network-based CFD code that solves discretized systems of equations, to model air pollution dispersion within complex landscape across urban areas. The structured mesh image, as demonstrated in Fig. 1, represents the urban environment with detailed building and infrastructure layouts. Pollution sources, primarily vehicle emissions, will be defined within the mesh and served as the foundation for our CFD simulations. Using the AI4PDEs code, we will conduct detailed simulations of pollution movement within the urban environments, as shown in Fig. 2, providing insights into the concentration under various conditions. The method is chosen for its computational efficiency as it uses pre-configured weights based on different discretisation method, eliminating the need for training.

Figure 1: Structured mesh of urban environments.

Figure 2: Simulation of flow speed using the AI4PDEs code.

To enhance the accuracy of our predictions, we will also integrate real-time traffic flow and weather data through a process known as data assimilation, as illustrated in Fig. 3. These data will be sourced from AI-Respire projects and Open Weather, enabling dynamic updates to the model based on current conditions. Daily traffic flow data, including vehicle counts and trends, along with comtinuous weather conditions such as wind speed and direction, temperature, and humidity, will be assimilated to influence pollutant dispersion patterns within urban areas. The data assimilation process will be performed in a reduced dimension or latent space, capturing essential components of the data that are most relevant to the pollution dispersion. To achieve this, we will utilise VAEs, which compress the high-dimensional unput data into a lower-dimernsional latent space while preserving the key features necessary for accurate predictions. This compression process is demonstrated in Fig. 4, allowing our model to adapt to observational data that improves accuracy while maintaining computational efficiency and reliability.

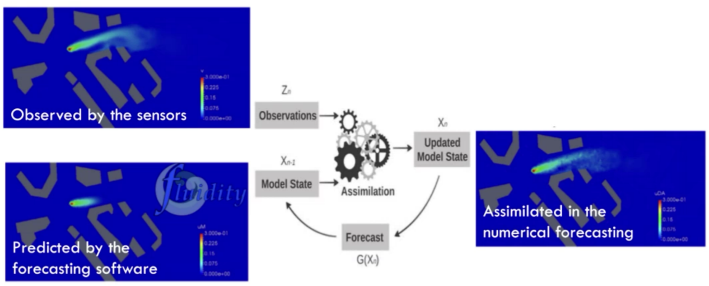

Figure 3: Process of data assimilation.

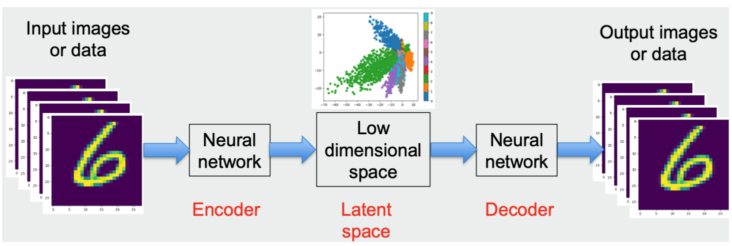

Figure 4: Concept of Variational Autoencoder.

All the simulations will utilise GPUs from Imperial College's High-Performance Computing (HPC) facility to leverage the computational efficiency. Besides, the solver tolerances will be adjusted to ensure the solutions are consistent with traditional CFD models. The model will be further validated against existing air quality measurements from monitoring stations, and senitivity analysis will be performed to understand the influence of various parameters.

The justification for this approach lies in its potential for high efficiency and predictive capability, making it more practical for dynamic urban environments. However, potential limitations include the dependency on the data quality, availability of real-time data, and the computational complexity for large-scale simulations.

3 Expected Outcomes

The deliverables will include a validated model, comprehensive pollution exposure maps, and an innovative framework for continuous data integration and assimilation using VAEs. By leveraging AI4PDEs intergrated with real-time data, we anticipate developing a robust neural network-based CFD model capable of reflecting dynamic conditions and simulating pollution dispersion in urban environments with high precision. This model will generate detailed pollution maps that can inform public health strategies and urban planning, offering practical applications to mitigate pollution-related health risks and improve urban living conditions. Preliminary findings have indicated the feasibility of using AI4PDEs for pollution modelling, ensuring the project is on a solid path to success. The maps will provide visual and quantitative insights, enabling policymakes to make informed decision about air quality management. The framework for dynamic data integration and assimilation ensures the accuracy and practical utility under varying conditions.

However, challenges such as variability in data quality and challenging computational demands despite neural network efficiency will need to be addressed. To overcome this, extensive validation precedures will be implemented using the powerful computing resources provided by Imperial College's High-Performance Computing (HPC) cluster to handle data-intensive tasks and ensure accuracy on a large scale. By tackling these issues, the project aims to deliver a reliable and practical tool for urban quality management,

4 Future Plan

Over the next three month, the porject will proceed according to a sturctured timeline aimed at achieving key milestones. In the first month, activities will commence with a thorough literature review and the establishment of a modelling framework based on the AI4PDEs code. This includes building a sample model thats uses a Variational Autoencoder (VAE) to compress data into a reduced space. Additionally, setup of the High-Performance Computing (HPC) facility provided by Imperial College and acquirement of necessary traffic and weather data from AI-Respire project and Open Weather will occur. In the second month, the emphasis will shift towards the intergration of the neural network-based CFD model with real-time data, followed by testing to ensure funtionality and accuracy under varying environmental conditions. The final month will be dedicated to refining the model based on preliminary results and completing the dissertation. This phase will involve further validation and adjustments to optimize the predictive capabilities.

Progress to date include a comprehensive literature review, setup of HPC facility, and the construction of a sample model that effectively compresses the data using a Variational Autoencoder(VAE).

References

[1] Boyang Chen, Claire E. Heaney, and Christopher C. Pain. Using ai libraries for incompressible computational fluid dynamics. *arXiv preprint arXiv:2402.17913*, 2024.

[2] Geer A. J. Learning earth system models from observations: machine learning or data assimilation? *Philosophical Transactions of the Royal Society A*, 379, 2021.

[3] Toby R. F. Phillips, Claire E. Heaney, Boyang Chen, Andrew G. Buchan, and Christopher C. Pain. Solving the discretised neutron diffusion equations using neural networks. *International Journal for Numerical Methods in Engineering*, 2023.

[4] Toby R. F. Phillips, Claire E. Heaney, Boyang Chen, and Christopher C. Pain. Solving the discretised boltzmann transport equations using neural networks: Applications in neutron transport. *arXiv:2301.09991*, 2023.

[5] Maike Sonnewald, Redouane Lguensat, Daniel C. Jones, Peter D. Dueben, Julien Brajard, and V. Balaji. Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environmental Research Letters*, 16(7), 2021.

[6] Nils Thuerey, Konstantin Weißenow, Lukas Prantl, and Xiangyu Hu. Deep learning methods for reynolds-averaged navier–stokes simulations of airfoil flows. *AIAA Journal*, 68(1), 2019.

[7] Yi Xiao, Qilong Jia, Wei Xue, and Lei Bai. Vae-var: Variational-autoencoder-enhanced variational assimilation. *arXiv:2405.13711*, 1, 2024.