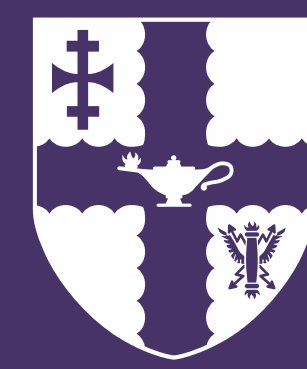


A Reproducible SUMO–FiPy Framework for Synthetic Traffic-Related Air-Quality Data

Niloufar Esfahanian, Miguel Martínez García, Sarah Dunnett



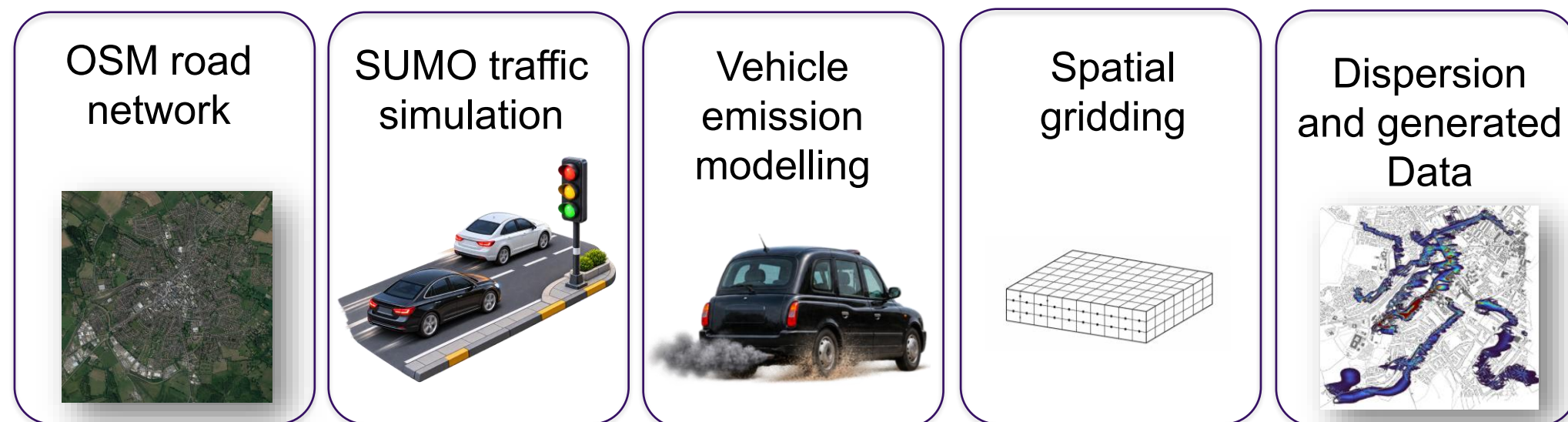
Loughborough University

1) Introduction/Motivation

Real-world traffic-emission datasets are expensive to obtain, sparse in space and time, and difficult to control. This project provides a synthetic traffic-emission simulation package that generates a coherent, high-resolution benchmark dataset for research [1,2].

- Combines road-network extraction, traffic simulation, emission modelling, and concentration analysis.
- Delivers spatio-temporal pollutant fields (NO_x , $\text{PM}_{2.5}$ and PM_{10}) and monitoring station timeseries.
- Supports custom corridors, stations, and scenario design.
- Enables reproducible experiments for model development, evaluation, and comparison.

2) Simulation workflow



- OpenStreetMap (OSM)** → Urban domain [3].
- SUMO** → Traffic simulation and Vehicle emissions [4]. **Microscopic traffic + HBEFA4** vehicle-emission model (a physics-based emission model).
- FiPy** → Pollutant dispersion [5].
- FiPy PDE Setup → Eulerian dispersion model [6].

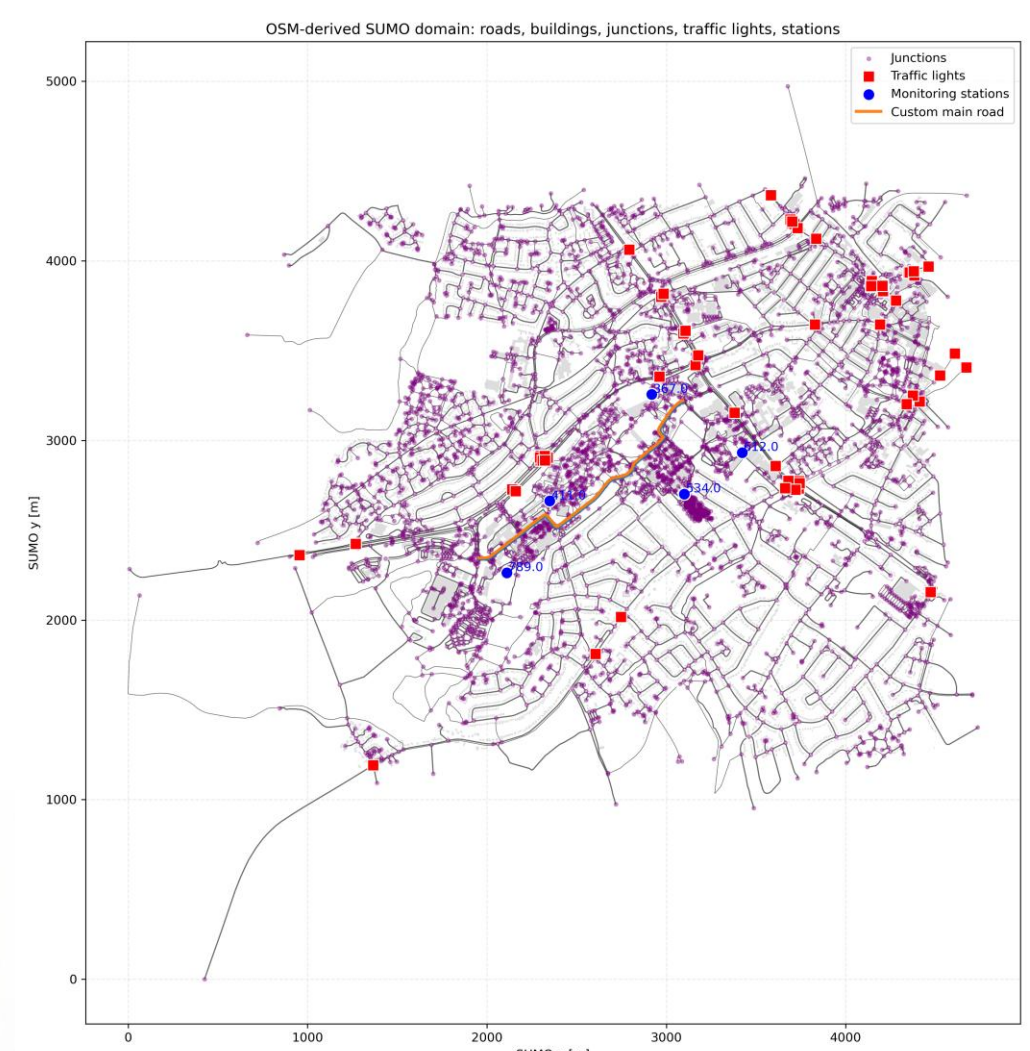
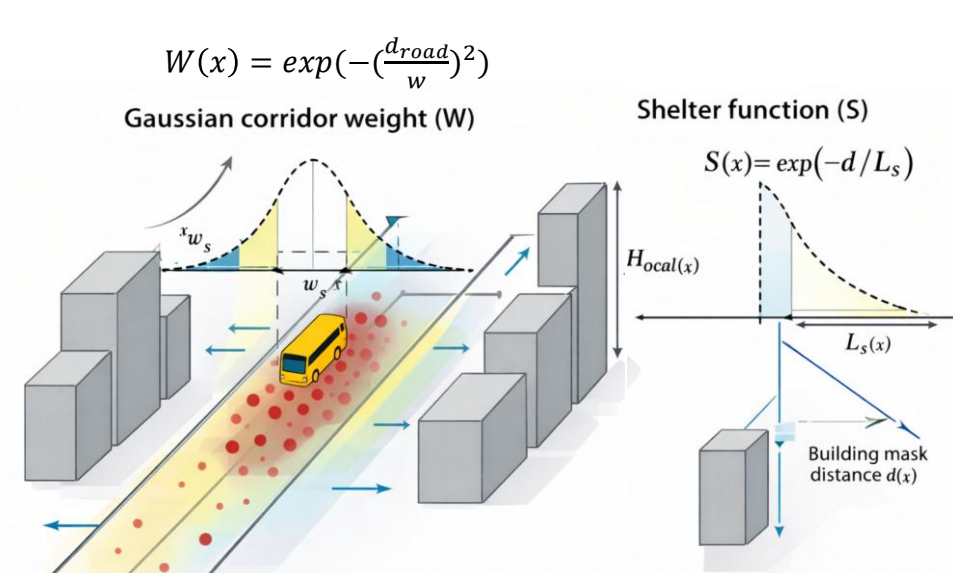
$$\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{u}C) - \nabla \cdot (D\nabla C) + S = 0$$

- Moving source: A vehicle emits mass at $q(t)$ [g/s]
 - Each source distributed into a volume: $\text{Volume} = \text{cell area} \times \text{mixing height}$
 - Emission source is Distributed over horizontal Gaussian kernel
- $$\exp\left(-\frac{(x-x_p)^2 + (y-y_p)^2}{2\sigma^2}\right)$$
- Emission rate can be converted to $\frac{q(g/s)}{A(m^2) \times H(m)} [g/m^3/s]$.

3) Demonstration Use Case: Loughborough University Campus

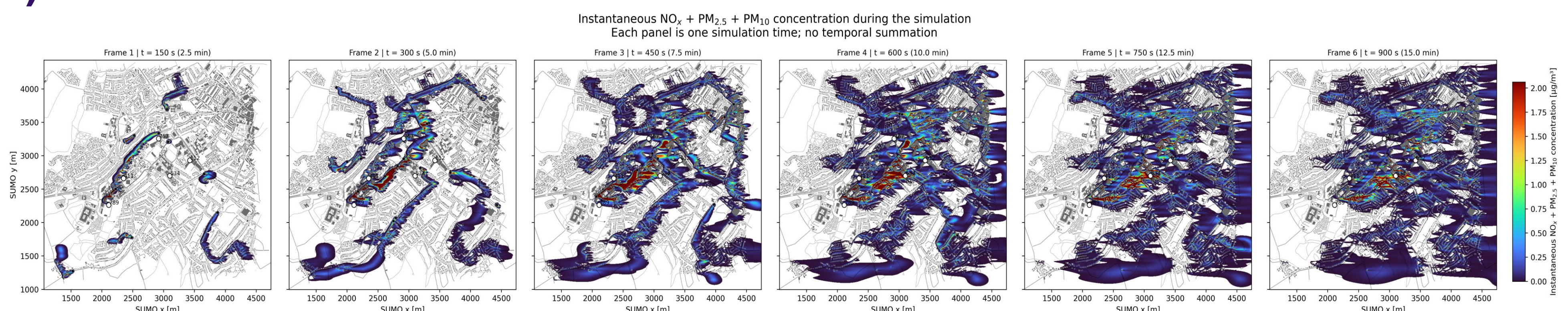
Baseline configuration:

- Simulation duration: 900s
- Timestep: 1 s
- Grid: 4×4 (m)
- Mixing height: 3 m
- Wind Speed: 2m/s
- Monitoring stations: 5
- Vehicles: 18
- Pollutants: NO_2 , $\text{PM}_{2.5}$, PM_{10}



- Real campus geometry extracted from **OpenStreetMap**.
- 5 monitoring stations** define the study area and act as receptor locations.
- 1 km buffer** around the station network defines the simulation domain.
- OSM roads, junctions, traffic lights and buildings are converted to the SUMO coordinate system.
- 18-vehicle synthetic mixed fleet** simulated over **900 s**.
- FiPy predicts pollutant concentrations across the **full spatial domain** and at **each monitoring station**.
- Demonstrates a **reusable workflow transferable to other locations and scenarios**.

4) Results & Extractable Data



- For every simulated vehicle: trajectory, x-y position, speed, route and lane, departure/arrival. Trip duration. Route length.
- Vehicle Emission Data

- Spatial dataset output, including building positions, network edges and lanes, junctions, traffic lights, etc.
- Full 2-D pollutant concentration maps over the urban domain.
- Pollutant concentration recorded (timeseries) at every station at every timestep.

5) Conclusion

- ToolBox2 provides a reproducible, reusable, and configuration-driven SUMO–FiPy framework for generating synthetic traffic-related air-quality datasets.
- Provides an end-to-end workflow from OSM-based urban-domain generation, microscopic traffic and emission simulation, pollutant dispersion, and spatio-temporal output generation.
- Scenario-based parameterisation enables controlled experiments across different traffic demand, fleet composition, meteorological conditions, urban geometries, and monitoring configurations. The simulation is specifically designed to generate physically interpretable outputs under controlled scenarios.
- The modular design makes the package extensible and reusable for different urban areas and research scenarios.

6) Future Work: Towards a City-Scale Air-Quality Digital Twin

- Scale from local domains to city-scale road networks with resultant increased traffic intensity.
- Calibrate and validate dispersion predictions against measured concentrations.
- Couple forecasting, optimisation and machine learning with the simulation engine.
- Inverse traffic–emission modelling: use the synthetic dataset to estimate the number, type, location and arrangement of vehicles on the road network needed to produce pollutant concentrations within a prescribed target range.

CONTACT

INFORMATION:

School of Aeronautical, Automotive, Chemical and Materials Engineering, Loughborough University Loughborough university Leicestershire LE11 3TU UK
n.esfahanian@lboro.ac.uk

Acknowledgement:

I would like to express my sincere gratitude to my supervisors for their invaluable guidance, continuous support, and constructive feedback throughout the course of this research.

References:

- [1] Gryech, I. et al. (2020). *Machine learning for air quality prediction using meteorological and traffic related features*. Journal of Ambient Intelligence and Smart Environments, 12(5), 379–391.
- [2] Tang, D., Zhan, Y. & Yang, F. (2024). *A review of machine learning for modeling air quality: Overlooked but important issues*. Atmospheric Research, 300.
- [3] Haklay, M. and Weber, P. (2008). *OpenStreetMap: User-Generated Street Maps*. IEEE Pervasive Computing, 7(4), 12–18. <https://doi.org/10.1109/MPRV.2008.80>
- [4] Eclipse SUMO. *Simulation of Urban MObility*. <https://eclipse.dev/sumo/>.
- [5] Guyer, J.E., Wheeler, D.W. & Warren, J.A. (2009). *FiPy: Partial Differential Equations with Python*. Computing in Science & Engineering, 11. <https://doi.org/10.1109/MCSE.2009.52>
- [6] De Visscher, A. (2014). *Air Dispersion Modeling: Foundations and Applications*. Wiley.