

IMPROVING LOCAL AIR QUALITY PREDICTIONS USING TRANSFER LEARNING ON SATELLITE DATA AND GRAPH NEURAL NETWORKS

Finn Gueterbock

University of Bristol, UK

Raul Santos-Rodriguez

University of Bristol, UK

Jeffrey N. Clark

University of Bristol, UK

`jeff.clark@bristol.ac.uk`

ABSTRACT

Air pollution is a significant global health risk, contributing to millions of premature deaths annually. Nitrogen dioxide (NO_2), a harmful pollutant, disproportionately affects urban areas where monitoring networks are often sparse. We propose a novel method for predicting NO_2 concentrations at unmonitored locations using transfer learning with satellite and meteorological data. Leveraging the GraphSAGE framework, our approach integrates autoregression and transfer learning to enhance predictive accuracy in data-scarce regions like Bristol. Pre-trained on data from London, UK, our model achieves a 8.6% reduction in Normalised Root Mean Squared Error (NRMSE) and a 32.6% reduction in Gradient RMSE compared to a baseline model. This work demonstrates the potential of virtual sensors for cost-effective air quality monitoring, contributing to actionable insights for climate and health interventions.

1 INTRODUCTION

Air pollution is one of the leading causes of global mortality, responsible for over 8 million deaths annually (Lelieveld et al., 2023). Nitrogen dioxide (NO_2), primarily emitted from vehicles and industrial activities, has severe health impacts, particularly in urban environments. Despite its significance, the limited number of ground-based monitoring stations hinders the ability to measure air quality effectively, mitigate localised air quality challenges effectively, and evaluate the impact of policies (World Health Organization and others, 2018).

Satellite-based data provides global coverage of air quality metrics; however, its low spatial resolution restricts its applicability for localised decision-making. To address this gap, we propose a novel model that combines satellite and meteorological data with sparse ground-based measurements to simulate high-resolution NO_2 readings at unmonitored locations, creating ‘virtual sensors’.

We introduce an inductive learning framework based on GraphSAGE (Hamilton et al., 2017) that incorporates temporal dependencies through autoregression and enhances performance using transfer learning. By leveraging transfer learning from cities with better monitoring networks, this work seeks to improve the accuracy and generalizability of the model, enabling it to predict air quality in regions with fewer monitoring stations. We achieve sizeable accuracy improvements in unseen locations. This approach has the potential to enable low-cost, scalable air quality monitoring, especially in resource-constrained regions, thereby supporting global efforts toward better health outcomes and climate resilience.

2 RELATED WORK

Air quality prediction has been extensively studied using a variety of machine learning techniques. Traditional methods often rely on geostatistical interpolation, such as kriging (Cressie, 1990). While effective for some spatial analyses, these methods struggle with the complex spatiotemporal relationships in air quality data. Graph Neural Networks (GNNs) offer a more flexible approach by leveraging the relationships between data points in a network (Xu et al., 2018). For example, Muthukumar

et al. (2022) used a Graph Convolutional Network (GCN) combined with time series models to predict $\text{PM}_{2.5}$ levels, demonstrating the potential of GNNs for air quality forecasting.

Satellite data has become a key resource in air quality research, offering global coverage of atmospheric metrics such as Aerosol Optical Depth (AOD) and NO_2 column density. Masih (2019) demonstrated that machine learning models, such as Random Forests, can predict NO_2 concentrations from satellite and meteorological data. Similarly, Ghahremanloo et al. (2021) applied deep learning to estimate daily NO_2 levels, achieving promising results. However, these approaches often lack the spatial resolution necessary for accurate local predictions, particularly in urban environments.

Transfer learning has shown potential for improving model generalisability, especially in data-scarce regions. Ma et al. (2019) found that transfer learning improved air quality prediction accuracy when applied to larger temporal resolutions. Yadav et al. (2022) utilised deep transfer learning on satellite imagery to enhance air quality predictions in developing countries, demonstrating the feasibility of pre-training on well-resourced cities and fine-tuning on data-poor areas. Despite these advancements, the high spatial resolution necessary for accurate NO_2 predictions in urban environments remains a challenge.

This study aims to bridge this gap by using transfer learning and GNN-based models to provide more accurate, location-specific NO_2 predictions, addressing limitations in both spatial and temporal resolution present in previous work.

3 METHODS

3.1 DATA

Surface NO_2 Measurements Ground-based NO_2 measurements for the Bristol area were obtained from the Air Quality Data Continuous dataset via the Open Data Bristol API (Bristol City Council, 2022), which provides hourly air quality data from 19 different locations across the city since 1993. For London, we use data from the London Air Quality Network Mittal (2020), which reports hourly NO_2 readings from 112 locations across Greater London. From 2018 onwards, the period of time for which satellite data is available, the datasets includes 246,572 data points across 8 locations in Bristol, and 4,182,699 across 112 locations in London (Figure 1).

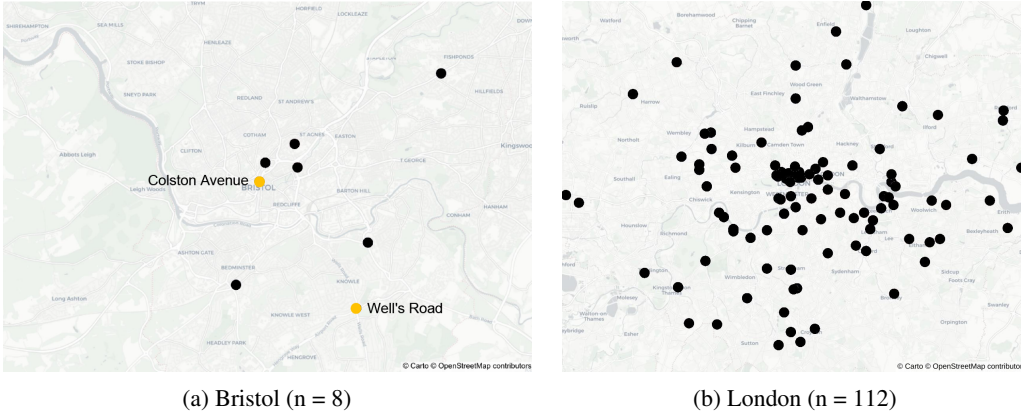


Figure 1: Maps showing the spatial distribution of NO_2 sensors locations in (a) Bristol and (b) London accessed during this study. In (a), the two Bristol sensor locations marked in orange are those for which temporal NO_2 predictions are presented in Figure 2.

Satellite Data Satellite data from the European Space Agency’s Sentinel-5P (2018–present) provides daily NO_2 concentrations and aerosol indices at a 5.5×3.5 km resolution. The satellite data is treated as static between daily measurements to match its temporal resolution.

Additional Features Hourly meteorological data from the ERA5-Land dataset Muñoz-Sabater et al. (2021) including variables such as temperature, wind, humidity, and cloud cover at a 5 km resolution were included. All features are listed in Appendix A.1. Proximity to A-roads or motorways was calculated using OS Open Roads Ordnance Survey (2022), providing a distance-to-road feature for each sensor location. All features were standardised.

3.2 MODELS

GraphSAGE aggregates data from a sensor’s local neighborhood, improving its ability to capture geographical variations. We introduce autoregression to model temporal dependencies in NO_2 concentrations, as shown by the data autocorrelation (Appendix A.1, Figure 3). This enables the model to make more accurate predictions by considering past NO_2 values. Transfer learning was conducted by pre-training on London data and fine-tuning on Bristol data. Performance was assessed against models trained on only Bristol data. For comparison purposes we evaluate three additional non-GNN models: XGBoost, MLP, and CNN, and performed transfer learning on the best performing of these baseline models. Appendix A.2 details model training.

Each model takes as input satellite, meteorological, and time-based features (Appendix A.1) and outputs hourly NO_2 predictions. We test each model on unseen locations, using RMSE (Root Mean Squared Error), NRMSE (Normalized RMSE), and Grad-RMSE (Gradient RMSE) averaged across all unseen locations to assess performance as a ‘virtual sensor’. Each model is tested on one location at a time, having trained on all the remaining locations. Results are presented for sample periods of time spanning several weeks. Code for all models and plots discussed in this report are available online ¹.

4 RESULTS AND DISCUSSION

The GraphSAGE model achieved an RMSE of $17.016 \mu\text{g}/\text{m}^3$, NRMSE of 0.526, and Grad-RMSE of $9.426 \mu\text{g}/\text{m}^3$, demonstrating improved performance compared to the MLP, XGBoost and CNN baselines (Table 1). Transfer learning from London improved predictions for both the CNN and GraphSAGE architectures, highlighting its potential for areas with fewer monitoring stations.

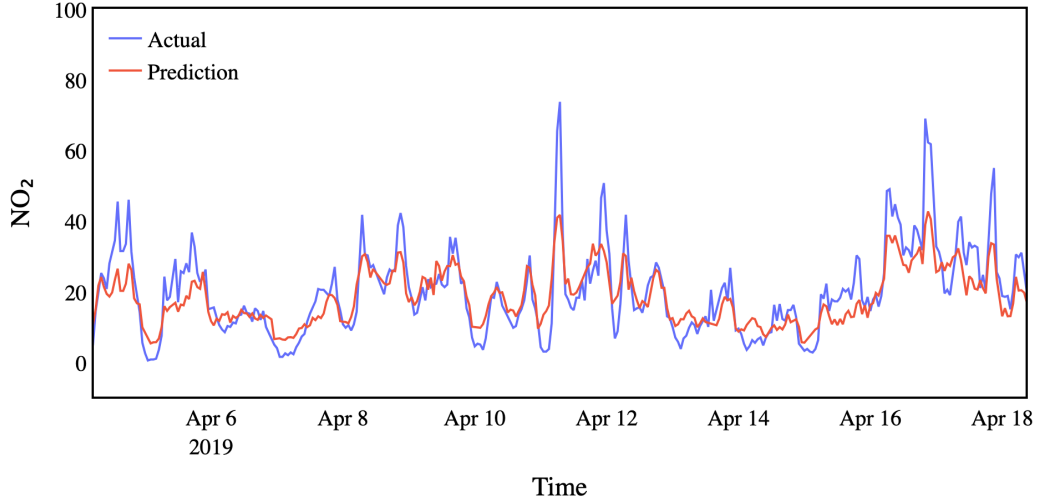
Model	RMSE ↓	NRMSE ↓	Grad-RMSE ↓
MLP	27.482	0.876	10.812
XGBoost	22.773	0.721	9.583
CNN	21.133	0.672	9.741
Transferred CNN	18.362	0.583	9.912
GraphSAGE	17.016	0.526	9.426
Transferred GraphSAGE	15.623	0.481	6.354

Table 1: Performance of all models, averaged across Bristol sensor locations (n = 8).

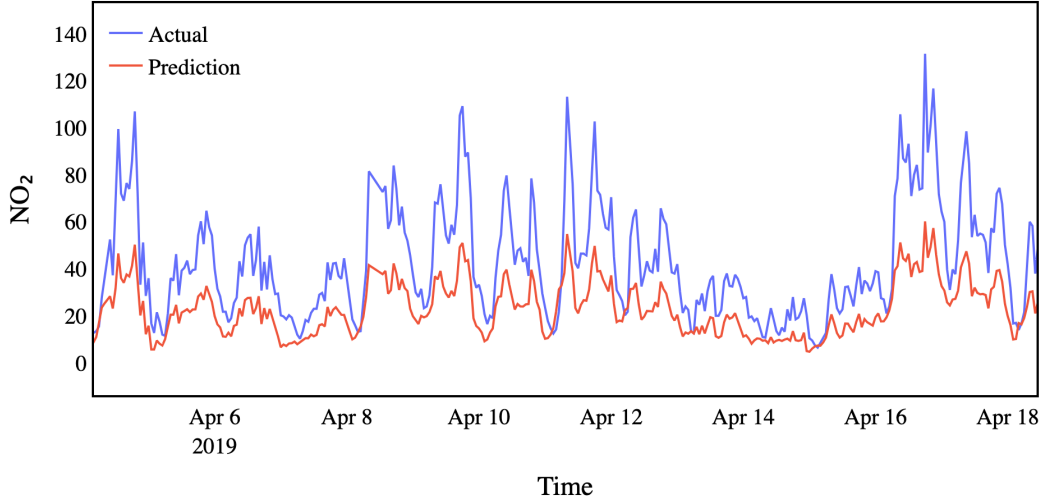
The transferred GraphSAGE model achieved the best performance across all models (Table 1), with transfer learning reducing error metrics by between 8.2% and 32.6% (Appendix A.3, Table 2). The model achieved an RMSE of $15.623 \mu\text{g}/\text{m}^3$ and an NRMSE of 0.481, both of which can be considered acceptable within the context of urban NO_2 forecasting, especially given the complexities involved with applying transfer learning to a new geographical area. Recent comparable studies utilising similar graph-based and hybrid methodologies typically report RMSE values in the range of $10\text{--}20 \mu\text{g}/\text{m}^3$, with NRMSE frequently between 30% and 50% (Wang et al., 2023; Qi et al., 2023). Although the errors reported here are somewhat higher than those achieved by specialised, locally trained models, the demonstrated accuracy remains promising for practical application in urban environments with sparse monitoring infrastructure. In particular, there is potential for creating virtual sensors in data-scarce regions, improving predictions by capturing both spatial and temporal dependencies. The addition of other data sources, such as terrain and land-use, could be incorporated to further improve performance.

¹<https://github.com/finngueterbock/FYP>

Samples of predicted versus actual NO_2 values for two locations (Figure 2), demonstrate temporal prediction quality over the period of several weeks, as well as the tendency to under-predict higher NO_2 values. This effect is made especially clear in Figure 2b, where the predicted values seem to follow the true values accurately, but are scaled down by some factor, leading to an overall bias. Future work should aim to address this bias, in addition to evaluating on additional metrics to measure the systematic error, both of which will be key to ensuring that the model predictions can be used to derive reliable quantitative measures of air pollutants at various scales.



(a) NO_2 predictions for Well's Road, Bristol.



(b) NO_2 predictions for Colston Avenue, Bristol.

Figure 2: Actual and predicted NO_2 values from the transferred GraphSAGE model for two locations in Bristol: (a) Well's Road - a location with typically low NO_2 values, and (b) Colston Avenue - a location with typically high NO_2 values.

While the results are promising, challenges remain, particularly around computational efficiency. The need to reconstruct the graph for each timestep imposes significant costs, which future work should address through optimised architectures. Given the strong temporal dependence observed in NO_2 (Appendix A.1, Figure 3), benchmarking against a time-series model in future work may prove valuable. We acknowledge that nearby locations may exhibit correlated air quality, which could lead to optimistic performance estimates under the employed leave-one-out evaluation. Future

work should explore alternative evaluation schemes. Additionally, validation in diverse geographical regions should be carried out to assess the model’s generalisability.

5 IMPACT AND IMPLICATIONS

The proposed approach addresses critical gaps in urban air quality monitoring by enabling high-resolution NO₂ predictions in areas with sparse monitoring networks. This work has direct implications for public health and climate policy, providing low-cost tools for assessing air quality in resource-limited settings. By integrating satellite data with machine learning, this method supports global efforts to achieve SDG 3 (Good Health and Well-being) and SDG 13 (Climate Action) (Lee et al., 2016).

Future deployments could extend to developing nations, where the lack of monitoring infrastructure exacerbates air quality challenges. There is potential to scale the approach beyond the city level to larger geographical regions. Furthermore, the approach could be applied for other pollutants for which satellite datasets are available such as methane or sulfur dioxide, enabling emissions detection and broadening the work’s impact on climate resilience and urban planning. Collaborations with local governments and environmental agencies will be essential to ensure practical application and policy integration.

6 CONCLUSION

This study demonstrates the potential of leveraging GNNs and transfer learning to address challenges in localised air quality prediction. By integrating satellite data, meteorological features, and ground-based measurements, we developed a GraphSAGE-based model capable of accurately predicting NO₂ concentrations at unmonitored locations in Bristol. Pre-training on data from London and fine-tuning on Bristol significantly improved model performance, achieving an 8.6% reduction in NRMSE and a 32.6% reduction in Gradient RMSE compared to the baseline.

Our findings highlight the feasibility of deploying virtual sensors in resource-limited settings, contributing to scalable, low-cost air quality monitoring solutions. This approach provides actionable insights for public health and urban planning, especially in cities with sparse monitoring networks.

Future work will focus on optimising the model’s computational efficiency and expanding validation across diverse geographical regions. Additionally, extending this framework to other pollutants and integrating real-time monitoring data could further enhance its utility. By addressing these challenges, this methodology has the potential to support global efforts in mitigating air pollution and combating climate change.

ACKNOWLEDGMENTS

JNC and RSR are funded by the Self-Learning Digital Twins for Sustainable Land Management grant [grant number EP/Y00597X/1] and the UKRI Turing AI Fellowship [grant number EP/V024817/1].

REFERENCES

- Bristol City Council. Air quality data continuous, 2022. URL <https://opendata.bristol.gov.uk/explore/dataset/air-quality-data-continuous/information/?disjunctive.location>. [Last accessed 28/10/22].
- Lia Chatzidiakou, Anika Krause, Olalekan AM Popoola, Andrea Di Antonio, Mike Kellaway, Yiqun Han, Freya A Squires, Teng Wang, Hanbin Zhang, Qi Wang, et al. Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments. *Atmospheric measurement techniques*, 12(8):4643–4657, 2019.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

- Noel Cressie. The origins of kriging. *Mathematical geology*, 22(3):239–252, 1990.
- CSIRO’s Data61. Stellargraph machine learning library. <https://github.com/stellargraph/stellargraph>, 2018.
- Masoud Ghahremanloo, Yannic Lops, Yunsoo Choi, and Bijan Yeganeh. Deep learning estimation of daily ground-level no₂ concentrations from remote sensing data. *Journal of Geophysical Research: Atmospheres*, 126(21):e2021JD034925, 2021.
- William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, pp. 1025–1035, 2017. ISBN 9781510860964.
- Bandy X Lee, Finn Kjaerulf, Shannon Turner, Larry Cohen, Peter D Donnelly, Robert Muggah, Rachel Davis, Anna Realini, Berit Kieselbach, Lori Snyder MacGregor, et al. Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy*, 37:13–31, 2016.
- Jos Lelieveld, Andy Haines, Richard Burnett, Cathryn Tonne, Klaus Klingmüller, Thomas Münzel, and Andrea Pozzer. Air pollution deaths attributable to fossil fuels: observational and modelling study. *bmj*, 383, 2023.
- Jun Ma, Jack CP Cheng, Changqing Lin, Yi Tan, and Jingcheng Zhang. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214:116885, 2019.
- Adven Masih. Application of random forest algorithm to predict the atmospheric concentration of no₂. In *2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, pp. 252–255. IEEE, 2019.
- Louise Mittal. London air quality network summary report, 2020. URL https://londonair.org.uk/london/reports/2020_LAQN_Report.pdf. [Last accessed 28/11/22].
- Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 2021.
- Pratyush Muthukumar, Emmanuel Cocom, Kabir Nagrecha, Dawn Comer, Irene Burga, Jeremy Taub, Chisato Fukuda Calvert, Jeanne Holm, and Mohammad Pourhomayoun. Predicting pm_{2.5} atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Quality, Atmosphere & Health*, 15(7):1221–1234, 2022.
- Ordnance Survey. Os open roads. <https://www.ordnancesurvey.co.uk/business-government/products/open-roads.html>, 2022.
- Junyu Qi, Bin Liu, Lei Zhang, Jie Huang, and Honglei Zhao. Graph neural network-based spatiotemporal air quality prediction model with satellite remote sensing and meteorological data. *Environmental Pollution*, 316:120596, 2023.
- James Thomas and Sam Gunner. Local air – mapping the local environment using e-scooters, 2023. URL <https://www.localair.uk/>.
- Yujie Wang, Shuang Li, Ke Zhang, Hongyu Liu, and Wei Huang. A hybrid deep learning approach for no₂ concentration forecasting combining transformer and lstm. *Atmospheric Pollution Research*, 14(2):101615, 2023.
- World Health Organization and others. Air pollution and child health: prescribing clean air: summary. Technical report, World Health Organization, 2018.
- Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Nishant Yadav, Meytar Sorek-Hamer, Michael Von Pohle, Ata Akbari Asanjan, Adwait Sahasrabhojane, Esra Suel, Raphael Arku, Violet Lingenfelter, Michael Brauer, Majid Ezzati, et al. Deep transfer learning on satellite imagery improves air quality estimates in developing nations. *arXiv preprint arXiv:2202.08890*, 2022.

A APPENDIX

A.1 DATA

The following features were used to predict local NO₂ measurements:

Satellite Features: Tropospheric NO₂ column number density, absorbing aerosol index;

Meteorological Features: Wind speed, wind gust speed, wind direction, vapour pressure deficit, temperature, surface pressure, relative humidity, dew point, cloud cover percentage;

Time-based Features: Day of the week, week of the year and time of day.

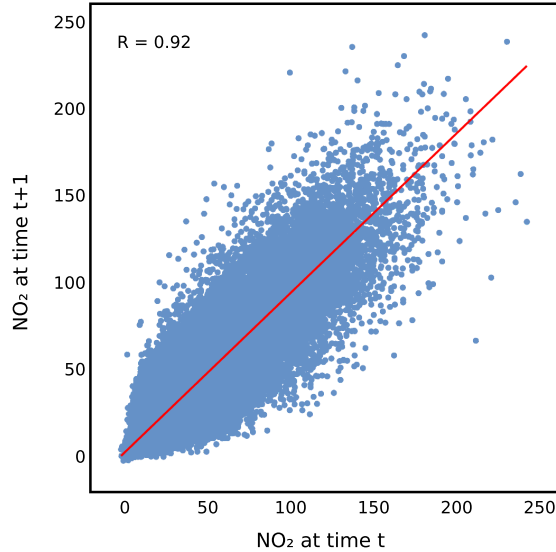


Figure 3: NO₂ values at time t vs NO₂ values at time $t + 1$.

A.2 MODEL TRAINING

GraphSAGE Since the GraphSAGE algorithm (Hamilton et al., 2017) works by sampling from a node’s local neighbourhood of connected nodes, feeding in the data all at once in this way does not allow the model to see other timesteps in the future or the past, as the graphs at each timestep are not connected to each other.

In order to address this issue of not having a continuous representation of time in the graph, we include the predicted NO₂ value from a node’s previous timestep as an input for the current prediction in a process called autoregression. Autoregression is a technique used to model time series data, where each data point is predicted based on the previous predicted values. During training, we include the previous timestep’s actual NO₂ value at each node as a feature for that node. When nodes at a certain timestep are missing, the last recorded value for that time of day is used in its place. This should not greatly affect our results due to the high correlation the NO₂ values have with time. This enables the model to learn how to aggregate the satellite data, meteorological data and previous NO₂ values for each node and its neighbours. The use of autoregression is particularly effective for modeling time series data that exhibit a high degree of autocorrelation, as is the case with the NO₂ concentration data. Autocorrelation is a measure of the degree to which a data point is correlated

with its preceding data points, and can be visualised using an autocorrelation plot (Figure 3). As can be observed from the plot, there is a significant degree of correlation between the NO_2 concentration at each timestep and the NO_2 concentration at the next timestep.

To predict on an unseen node, we must first initialise the node with a value for the NO_2 at the previous timestep. During development of the model, this was achieved by including the actual NO_2 value for the first timestep, however in reality this initial sample could be provided by an air quality sampling scheme such as the Breathe London scheme, which involves children wearing portable air quality monitors on their backpacks Chatzidiakou et al. (2019), or LocalAir, an e-scooter based air quality monitoring scheme in Bristol Thomas & Gunner (2023). These methods of sampling are designed to be cheap and versatile in their applications so could be used in countries lacking infrastructure to provide a baseline NO_2 reading for a new location. If no such sample is available, it would also be possible to simply provide an estimate for the NO_2 concentration at a particular location, as the model only has access to a single previous value, forcing it to focus on the change in NO_2 at each timestep.

We compare the performance of four different node aggregator functions, mean, max pooling, mean pooling and attentional aggregator, as defined in the StellarGraph python library Data61 (2018), and chose mean pooling. To improve the model’s ability to generalise to unseen data, we include a dropout layer with a drop out rate of 0.5 before the activation layer.

Other parameters for the model include the number of hops away from each node to sample from, the maximum number of nodes to sample at each hop and the number of neurons to use when aggregating each node and its neighbours. Since the maximum number of hops possible from any node in the Bristol graph is 2, it was decided that we would perform 2 aggregations, sampling nodes at one hop, then two hops from each node. At each of these steps, a maximum of 3 and 5 nodes would be sampled respectively. Other parameters for the model such as the number of neurons for aggregation and the learning rate were selected by trial and error.

Non-GNN baseline models Baseline model parameters were as follows:

- an XGBoost model Chen & Guestrin (2016) with 100 decision trees;
- a multilayer perceptron model (MLP) with 2 fully connected layers, a dropout layer with a rate of 0.5, and 2 more fully connected layers;
- a convolutional neural network (CNN) model with 2 convolutional layers, a dropout layer with a rate of 0.5, and 2 fully connected layers.

A.3 MODEL PERFORMANCE

Table 2 illustrates the performance improvements observed using transfer learning.

Model	RMSE ↓	NRMSE ↓	Grad-RMSE ↓
GraphSAGE	17.016	0.526	9.426
Transferred GraphSAGE	15.623	0.481	6.354
Percentage Improvement	8.185	8.576	32.593

Table 2: Comparison of NO_2 prediction performance between GraphSAGE and Transferred GraphSAGE models.