

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal noncommercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

Faculty of Environmental and Life Science
School of Geography and Environmental Science

**Assessing the feasibility of applying machine
learning tools to predict
environmental stressors from digital urban
fingerprints**

by

Feiyu Zhu

MSc

*A thesis for the degree of
Doctor of Philosophy*

Nov 2025

University of Southampton

Abstract

Faculty of Environmental and Life Science
School of Geography and Environmental Science

Doctor of Philosophy

**Assessing the feasibility of applying machine learning tools to predict
environmental stressors from digital urban fingerprints**

by Feiyu Zhu

Urban noise pollution poses persistent challenges to public health and urban sustainability. This dissertation advances a new technical paradigm for scalable urban noise prediction by integrating multispectral remote sensing imagery, land use/land cover (LULC) data, and state-of-the-art machine learning techniques. The research adopts a progressive structure comprising three data-driven studies, each representing a major methodological step forward.

Chapter 4 presents a pioneering approach that leverages convolutional neural networks (CNNs) to predict citywide noise levels using high-resolution multispectral imagery, validated in Southampton. Chapter 5 incorporates geospatial relationships through graph-based modeling, further improving spatial prediction accuracy. Building on these foundations, Chapter 6 proposes a generalizable dual-branch graph neural network (GNN) framework with domain adaptation and pseudo-labeling, enabling robust noise mapping across five UK cities using standardized remote sensing and Urban Atlas LULC data.

Results demonstrate that deep learning models—when properly integrated with remote sensing and urban structural features—can achieve high accuracy and transferability in noise prediction, even in cities lacking extensive noise measurements. The workflow substantially reduces field data collection costs and advances urban noise assessment toward scalable, transferable solutions.

This dissertation thus bridges the gap between traditional acoustic modeling and next-generation data-driven mapping, providing methodological innovations with practical value for urban planners and environmental authorities seeking efficient, city-scale noise management tools.

Catalog

List of Figures	5
List of Tables	8
Declaration of Authorship	11
Acknowledgements	12
Chapter 1: Introduction	1
1.1 Urbanization, Environmental Stressors, and Noise Pollution	1
1.2 Policy Context: Noise Mapping Frameworks	1
1.3 Technological Shifts: From Deterministic Models to Data-Driven Approaches	2
1.4 Research Gaps and Study Objectives	3
1.5 Structure of the Dissertation	4
Chapter 2: Literature Review	6
2.1 Introduction	6
2.2 Traditional Noise Prediction	7
2.3 Remote Sensing in Environmental Monitoring	8
2.4 Machine Learning in Environmental Modelling	10
2.5 Deep Learning Methodologies	13
2.6 Cross-City Prediction and Transfer Learning	19
2.7 Summary of the Literature Review	22
Chapter 3: Research Methodology and Preliminary Validation Experiments	24
3.1 Introduction	24
3.2 Data Sources and Study Areas	26
3.3 Machine Learning and Feature Learning Framework	35
3.4 Preliminary Experiments: Noise Prediction Using the CatBoost Model	51
Chapter 4: Predicting Urban Noise Levels Using EfficientNet and Multispectral Remote Sensing Data: A Case Study of Southampton	58
Abstract	58
4.1 Introduction	59
4.2 Methodology	60
4.3 Results	69
4.4 Discussion	81
4.5 Conclusion	91
Chapter 5: Urban Noise Prediction Incorporating Geospatial Relationships: A Case Study of Southampton	93
Abstract	93
5.1 Introduction	94
5.2 Methodology	95
5.3 Results	108
5.4 Discussion	122
5.5 Conclusion	132
Chapter 6: Generalizable Urban Noise Prediction via Cross-City Standardized Multispectral Imagery and Dual-Branch Graph Neural Networks	134
Abstract	134
6.1 Introduction	134
6.2 Methodology	136
6.3 Result	179

6.4 Discussion	225
Chapter 7: General Discussion	239
7.1 Introduction	239
7.2 Theoretical and Methodological Contributions: Systematic Synthesis	240
7.3 Comparative Analysis of Three Modelling Paradigms: Performance, Resolution, and Computational Trade-offs	242
7.4 Complementarity and Advancement Beyond International Urban Noise Mapping Frameworks	244
7.5 Limitations and Areas for Improvement	245
7.6 Policy and Practical Implications	248
7.7 Future Research Directions	250
7.8 Conclusion	252
Reference	254
Appendix I 28 Gray-Level Co-occurrence Matrix (GLCM) Features	273
Appendix II: Summary of Feature Maps	279
Appendix III. Complete Variable List for the Dual-Branch GNN Model	286

List of Figures

Figure 3.1	RGB composite of WorldView-2 imagery (pre-processed and mosaicked) illustrating the Southampton study area	27
Figure 3.2	Spatial distribution of 30 m × 30 m aggregated noise observation cells across Southampton, overlaid on OpenStreetMap	28
Figure 3.3	Urban Atlas 2012 LULC map of Southampton — colour-coded by class with code-label legend	32
Figure 3.4	Exploratory analyses of the Southampton noise dataset.	34
Figure 3.5	Structure of CatBoost adapted from Prokhorenkova et al., (2018). Each tree corrects the errors of its predecessors, and the ensemble prediction emerges from their weighted combination.	37
Figure 3.6	Conceptual illustration of receptive field expansion in CNNs	38
Figure 3.7	Conceptual illustration of message passing in a graph neural network.	40
Figure 3.8	Workflow for urban noise prediction.	41
Figure 3.9	Comparative visualisation of NDVI and derived GLCM features (GLN and ICI). (a) NDVI map of Southampton (b) GLN texture map derived from NDVI (c) ICI texture map derived from NDVI	43
Figure 3.10	Comparative illustration of Red-Edge reflectance and z-score maps at three spatial scales.	45
Figure 3.11	Comparative illustration of morphological operators applied to a local subset of Southampton's Red band imagery	47
Figure 3.12	Schematic description of textural feature extraction assuming 3 x 3-pixel image with three different gray-scale levels. (Park et al., 2020)	53
Figure 3.13	CatBoost Noise Prediction Overlay on OS Open Greyscale Map (30 m × 30 m grid)	55
Figure 3.14	Comparison of observed land cover (OpenStreetMap basemap) and CatBoost-predicted noise exposure for the area surrounding Southampton Common	56
Figure 3.15	Comparison of observed land cover (OpenStreetMap basemap) and CatBoost-predicted noise exposure in central Southampton.	57
Figure 4.1	Flowchart of the Overall Methodological Framework	61
Figure 4.2	Maps of BAI, BSI, NBEI, NDVI, NDWI, WVBI, and RGI for the study area	64
Figure 4.3	Retained versus excluded GLCM texture maps.	66
Figure 4.4	Workflow for CNN patch extraction and dataset assembly	67
Figure 4.5	EfficientNetB0 modified architecture for multi-channel regression (with 84 input channels and single-node regression output)	68
Figure 4.6	Prediction pipeline flowchart — depicting full-scene feature cube, patch extraction, model inference, and output reconstruction to 30 m grid	69
Figure 4.7	Model training curves across epochs — showing loss, MAE, and R ² for training and validation sets	70
Figure 4.8	Predicted urban noise exposure map (30 m resolution) over Southampton with light gray canvas base imagery overlay	73

Figure 4.9 Prediction error map over labeled validation samples — showing spatial distribution of residuals (dBA)	74
Figure 4.10 Overlay of measured and predicted noise maps with difference classes highlighted	75
Figure 4.11 Histogram of residuals; (b) Scatter plot of measured vs. predicted noise levels with identity line	76
Figure 4.12 Spatial distribution of top 20% noisiest grid cells & top 20% quietest grid cells	77
Figure 4.13 Measured vs Predicted Noise Levels and Residual Distributions for CatBoost and CNN models	79
Figure 4.14 Comparative noise exposure maps – (a) CNN model (this study); (b) GBT model by Alvares-Sanches et al., 2021a; (c1)END round 3 official Lden map; (c2) END round 3 official Lnight map	82
Figure 4.15 Local comparison of CNN and CatBoost predictions in the central urban green space, with RGB spectral index composite shown in the middle panel	84
Figure 4.16 Local comparison of CNN and CatBoost predictions in the port area along the A33 corridor	84
Figure 5.1 Spatial distribution and graph connection of the 50,908 noise samples overlap worldview-2 Imagery (Standard RGB band combination - band 5,3,2)	97
Figure 5.2 Schematic illustration of the node feature construction process using multi-scale ring buffers	98
Figure 5.3 Deep GraphSAGE architecture used for noise prediction	102
Figure 5.4 Model Performance Analysis. (a) Predicted vs Measured Noise Levels; (b) Residuals vs Predicted Values; (c) Histogram of Residuals; (d) Q-Q Plot of Residuals	109
Figure 5.5 Predicted noise map of Southampton at 30 m resolution, produced by the GraphSAGE model with OS Open Carto base imagery overlay	110
Figure 5.6 Comparison of noise prediction generated from different technology	111
Figure 5.7 Getis-Ord Gi hotspot analysis	113
Figure 5.8 Comparison between (a) OS Open Carto basemap showing infrastructure context, (b) Predicted noise levels from the GraphSAGE model and (c) Local Gi* cluster significance overlay in Region 1	114
Figure 5.9 Comparison between (a) OS Open Carto basemap showing infrastructure context, (b) Predicted noise levels from the GraphSAGE model and (c) Local Gi* cluster significance overlay in Region 2	115
Figure 5.10 Measured versus predicted noise levels for CNN and GNN at 30 m resolution	117
Figure 5.11 Residual distributions for CNN and GNN predictions at 30 m resolution	118
Figure 5.12 Measured versus predicted noise levels for GNN predictions at 30 m resolution evaluated against 30 m-aggregated and 4 m measurement datasets	120
Figure 5.13 Residual distributions of GNN predictions at 30 m resolution evaluated against 30 m and 4 m reference datasets	121
Figure 6.1 Proposed methodological workflow. Steps are color-coded by functional stage: data preparation (blue), grid sampling (green), feature extraction and selection (orange), graph and model construction (purple), and domain adaptation with pseudo-labeling (pink)	139

<i>Figure 6.2 Study Area Delineations and 30 m Grid Point Distributions</i>	<i>150</i>
<i>Figure 6.3 Spatial Distribution of High-Confidence Pseudo-Labels Across Target Cities</i>	<i>164</i>
<i>Figure 6.4 prediction workflow</i>	<i>175</i>
<i>Figure 6.5 Predicted noise map for Southampton and selected sub-regions: (a) Predicted noise map for Southampton (b) port and industrial area; (c) mixed-use city centre; (d) low-exposure green and rural zone. Each submap (right side) overlays Urban Atlas 2012 dominant land use classes</i>	<i>182</i>
<i>Figure 6.6 Full-resolution noise prediction maps for the four target cities: (a) Cardiff; (b) Portsmouth; (c) Nottingham; (d) Liverpool. All maps use the same 30 m resolution grid and color ramp for visual comparability</i>	<i>185</i>
<i>Figure 6.7 Predicted noise maps and END Lden maps for (a) Southampton, (b) Nottingham, (c) Portsmouth, and (d) Liverpool</i>	<i>188</i>
<i>Figure 6.8 Predicted noise map for Cardiff and selected sub-regions; (a) Noise prediction map for Cardiff (b) Civic centre with high noise exposure (c) Wenallt and Northern Hills with low noise exposure</i>	<i>196</i>
<i>Figure 6.9 Predicted noise map for Nottingham and selected sub-regions; (a) Noise prediction map for Nottingham (b) City Core around Old Market Square and Nottingham Station</i>	<i>197</i>
<i>Figure 6.10 Predicted noise map for Portsmouth and selected sub-regions; (a) Noise prediction map for Portsmouth (b) Western Southsea Area</i>	<i>198</i>
<i>Figure 6.11 Predicted noise map for Liverpool and selected sub-regions; (a) Noise prediction map for Liverpool (b) Areas between Lime Street Station and the Knowledge Quarter</i>	<i>200</i>
<i>Figure 6.12 Predicted Noise Trend across Urban Fabric Types</i>	<i>202</i>
<i>Figure 6.13 LULC-Type Specific Predicted Noise Distributions; (a) Continuous Urban fabric; (b) Discontinuous Dense Urban fabric; (c) Discontinuous Medium Density Urban fabric; (d) Discontinuous Low Density Urban fabric; (e) Industrial, commercial, public, military and private units; (f) Fast transit roads and associated land; (g) Other roads and associated land; (h) Railways and associated land; (i) Green Urban areas; (j) Sports and leisure facilities; (k) Pastures; (l) Forests</i>	<i>209</i>
<i>Figure 6.14 Predicted Noise across Functional Land Use Groups</i>	<i>211</i>
<i>Figure 6.15 UMAP feature projections for individual cities (a–e): Southampton, Portsmouth, Nottingham, Liverpool, and Cardiff. Each point corresponds to a 30 m grid cell, and color indicates the predicted LAeq value.</i>	<i>213</i>
<i>Figure 6.16 Combined UMAP embedding & Kernel density contour overlays across all five cities. (a) Combined UMAP embedding across all five cities, color-coded by city label; (b) Kernel density contour overlays by city.</i>	<i>215</i>

List of Tables

Table 3.1	Example of Noise Observations (subset of Southampton dataset)	28
Table 3.2	Overview of Feature Construction Across Chapters	30
Table 3.3	Overview of Data Partitioning Strategies across Modelling Frameworks	50
Table 3.4	GLCM Features Used in the Baseline Model and Their Definitions	52
Table 3.5	Statistical Descriptors Computed from GLCM Feature Maps within Each 30 m × 30 m Grid Cell	53
Table 4.1	Summary of the indices, their definitions, and key references	63
Table 4.2	Performance of different feature-map combinations and architectures in CNN-based noise prediction (testset MAE and R ²)	71
Table 4.3	Overall quantitative comparison of CatBoost and CNN models (on identical test samples)	79
Table 4.4	Per-class comparison of model accuracy across major LULC types (derived from Urban Atlas 2012)	80
Table 5.1	Comparison of feature design and neighborhood encoding strategies across CatBoost baseline, CNN, and GNN frameworks	99
Table 5.2	distance-based exponential weighting scheme	100
Table 5.3	Training Optimization Settings for GraphSAGE Noise Prediction Model	105
Table 5.4	Quantitative comparison of model performance across MAE, and R ²	112
Table 5.5	Quantitative summary	116
Table 5.6	Comparison of CNN and GNN performance across Urban Atlas land-use classes (30 m evaluation)	119
Table 5.7	Comparison of GNN prediction accuracy at 30 m resolution when evaluated using 30 m-aggregated and 4 m measurement datasets across Urban Atlas 2012 land-use classes	122
Table 6.1	Quantitative indicators of spatial heterogeneity across study cities	144
Table 6.2	WorldView-2 multispectral imagery acquisition dates	145
Table 6.3	Urban Atlas 2012 LULC classification and functional grouping used in this study	146
Table 6.4	Demographic and morphological characteristics of the five study cities (UK Census 2011 and Urban Atlas 2012)	148
Table 6.5	Overview of Remote Sensing-Derived Feature Maps	153
Table 6.6	Summary of Zonal Statistics for Remote Sensing Features	155
Table 6.7	Functional Grouping of LULC Classes for Noise Modeling	156
Table 6.8	Summary of LULC-Derived Features	156
Table 6.9	Summary of Final Feature Set Composition	158
Table 6.10	Selected Variables for Pseudo-Labeling Inference	159
Table 6.11	Dual-Threshold Filtering Criteria	161
Table 6.12	High-Confidence Pseudo-Label Summary Across Target Cities	161
Table 6.13	Top 10 Environmental Features Used for Edge Weight Modulation	166
Table 6.14	Comparison of MAE and R ² between this study and former chapters	180

<i>Table 6.15 Comparison between END and Measured Noise Levels by Road-Distance Zone</i>	190
<i>Table 6.16 END vs Measured Noise by Land-Use Category (Urban Atlas 2012)</i>	191
<i>Table 6.17 Quantitative Comparison between Predicted and END Noise Levels by Road Distance</i>	192
<i>Table 6.18 END–Prediction Differences by Land-Use Category (Urban Atlas 2012)</i>	192
<i>Table 6.19 Mean predicted LAeq (dBA) by UA2012 functional land-use group across five cities</i>	217
<i>Table 6.20 Spearman’s rank correlation matrix (ρ) of functional exposure ordering between city pairs</i>	218
<i>Table 6.21 Mean predicted LAeq (dBA) by road-distance band (0–150 m) across five cities</i>	219
<i>Table 6.22 ND alignment metrics across cities</i>	220
<i>Table 6.23 Pseudo-label consistency metrics across four target cities</i>	222
<i>Table 6.24 Descriptive statistics of predicted LAeq (dBA) across five cities, including mean, median, standard deviation, and selected quantiles(P05, P25, P75, P95).</i>	223
<i>Table 6.25 Land-use composition and functional structure of the five study cities based on Urban Atlas 2012.</i>	234

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as: S.R. Gunn. Pdflatex instructions, 2001
C. J. Lovell. Updated templates, 2011
S.R. Gunn and C. J. Lovell. Updated templates reference 2, 2011

Signed:.....

Date:.....

Acknowledgements

This dissertation would not have been possible without the invaluable guidance and support of my advisors, collaborators, family, and friends.

I would like to express my deepest gratitude to **Professor Felix Eigenbrod**, whose expertise in deep learning and environmental science has been instrumental in shaping the methodologies and innovations presented in this work. I am also profoundly thankful to **Professor Jonathan Hare**, whose technical insights and advice greatly enhanced the computational models and analytical frameworks employed throughout this research.

Special thanks go to **Dr. Tatiana Alvares-Sanches**, whose meticulous fieldwork and comprehensive data collection in Southampton provided the foundation for this study. I am equally indebted to **Professor Patrick Osborne** for his foundational guidance during the early stages of this research and for laying the groundwork for urban noise studies in Southampton.

On a personal note, I owe my deepest appreciation to **my wife, Yujie Zhu**, for her unwavering love, patience, and emotional support throughout my doctoral journey. Her encouragement during moments of uncertainty has been a constant source of strength and motivation.

The full set of analysis scripts, model implementations, and preprocessing workflows used throughout this study is publicly available on GitHub at

<https://github.com/udock0/urban-noise-prediction/tree/main>.

Chapter 1: Introduction

1.1 Urbanization, Environmental Stressors, and Noise Pollution

Urbanization has reshaped human living environments through the provision of advanced infrastructure, greater connectivity, and improved standards of living. However, these benefits have been accompanied by new and persistent environmental stressors, among which noise pollution has emerged as one of the most pressing challenges for public health, urban sustainability, and environmental justice. The World Health Organization (WHO, 2018) identifies environmental noise as the second most harmful environmental risk factor in Europe, surpassed only by air pollution. Its systemic impacts are well documented: chronic exposure to urban noise contributes not only to auditory impairments but also to cardiovascular disease, hypertension, metabolic disorders, sleep disturbance, and impaired cognitive development in children (Basner et al., 2014; Münzel et al., 2017; Stansfeld & Clark, 2015).

Beyond its direct health effects, noise pollution undermines quality of life in cities by degrading public spaces, constraining opportunities for recreation, and exacerbating stress-related behaviours. Importantly, its impacts are unevenly distributed: socially and economically disadvantaged groups often reside in noisier neighborhoods, where exposure is compounded by housing insecurity and limited access to green buffers (Shannon et al., 2016; Casey et al., 2017). This makes noise pollution not only a public health challenge but also a matter of environmental equity. The persistence and ubiquity of this stressor underscore the need for more scalable and context-sensitive approaches to assessment and mitigation.

1.2 Policy Context: Noise Mapping Frameworks

The regulatory response to environmental noise in Europe has been dominated by the Environmental Noise Directive (END, 2002/49/EC), which requires Member States to produce harmonised noise maps and action plans for major urban areas, roads, railways, and airports. END builds upon deterministic acoustic frameworks such as ISO 9613-2 (ISO, 1996) and the Common Noise Assessment Methods in Europe (CNOSSOS-EU) (Kephalopoulos et al., 2012), both of which provide physically based equations for modelling sound propagation. These frameworks combine emission data (e.g., traffic volumes, industrial activity) with environmental modifiers such as building geometries, terrain, meteorological conditions, and surface absorption to estimate noise exposure.

The strength of these methods lies in their interpretability and standardisation. They ensure consistency across jurisdictions, making them indispensable for compliance and international comparison. However, they are resource-intensive, requiring high-resolution traffic data, detailed land-use inventories, and calibrated meteorological records (Murphy & King, 2022). As such, they are often prohibitively costly for local authorities, especially in resource-limited contexts, and their transferability across cities is limited by the need for extensive local calibration (Salomons & Pont, 2012). In rapidly urbanising regions, where data infrastructure may be weak or fragmented, these barriers render END-style approaches impractical.

This regulatory landscape highlights a fundamental tension: while deterministic models are physically robust and politically entrenched, their implementation struggles to keep pace with the heterogeneity and dynamism of modern cities. Addressing this gap requires innovative frameworks that are scalable, transferable, and capable of leveraging widely available datasets.

1.3 Technological Shifts: From Deterministic Models to Data-Driven Approaches

Recent advances in remote sensing and machine learning provide an opportunity to move beyond the limitations of traditional noise mapping. Satellite remote sensing offers synoptic, repeatable, and standardised observations of urban morphology, land cover, and vegetation at multiple spatial scales (Avtar et al., 2020; Li et al., 2016). These variables are directly relevant to sound propagation, as road density, impervious surfaces, and green buffers strongly modulate noise exposure. The increasing accessibility of multispectral and hyperspectral imagery, together with harmonised products such as the European Urban Atlas, creates the possibility of constructing noise-predictive frameworks with consistent coverage across cities.

Parallel to this, machine learning (ML) and deep learning (DL) have transformed environmental modelling by learning complex, nonlinear relationships from data rather than relying on handcrafted acoustic equations. Classical ML methods such as Random Forests, Gradient Boosting, and CatBoost have demonstrated strong predictive performance when applied to noise and air quality modelling (Hu et al., 2017; Prokhorenkova et al., 2018). More recently, convolutional neural networks (CNNs) have enabled automated extraction of hierarchical features from imagery, while graph neural networks (GNNs) have introduced the capacity to model non-Euclidean spatial relationships, such as urban adjacency networks or transport corridors (Zhou et al., 2020).

Together, these technological shifts suggest a paradigm change: from static, resource-intensive models designed for regulatory compliance, toward data-driven frameworks that exploit widely available geospatial information for scalable and transferable urban noise prediction. This dissertation situates itself within this transition, developing a staged research programme that integrates remote sensing, machine

learning, and graph-based deep learning to establish a new technical pathway for city-scale noise mapping.

1.4 Research Gaps and Study Objectives

Although substantial progress has been achieved through both regulatory frameworks and academic research, important gaps remain in the field of urban noise mapping. Traditional acoustic models such as CNOSSOS-EU and ISO 9613-2 are highly dependent on detailed local data, including fine-grained traffic counts, building geometries, and meteorological records. Their reliance on these costly and context-specific inputs severely constrains scalability, while their dependence on static calibration reduces adaptability to dynamic urban soundscapes. Remote sensing, which has demonstrated significant potential in allied domains such as air quality modelling and urban expansion studies, has not yet been systematically integrated into noise prediction frameworks. Where it has been applied, the use of satellite imagery and land-use information often remains limited to coarse proxies or city-specific experiments, leaving the promise of transferable, standardised data sources underdeveloped.

Conventional machine learning methods such as Random Forests and Gradient Boosting Machines provide a useful bridge between deterministic models and deep learning, but their effectiveness depends heavily on manual feature engineering. This introduces subjectivity and limits generalisation, particularly when applied to cities with different morphological and socio-environmental contexts. Deep learning approaches, by contrast, offer the possibility of learning hierarchical features directly from imagery. Convolutional neural networks (CNNs) have begun to show promise in this regard, but their Euclidean assumptions restrict their ability to represent the non-regular, network-like structures that shape urban noise propagation. Graph neural networks (GNNs) have the potential to overcome these constraints by capturing both local dependencies and broader relational structures, yet their application to environmental noise modelling remains limited and underexplored.

A further gap concerns transferability across cities. Most existing models, whether acoustic, machine learning, or deep learning, are rarely designed to generalise beyond the city in which they are trained. Differences in land-use composition, urban form, and acoustic context create significant domain shifts, leading to deteriorating predictive performance when models are applied elsewhere. Few studies have systematically confronted this challenge, despite its importance for developing scalable frameworks capable of serving diverse urban environments.

Against this backdrop, the overarching aim of this dissertation is to develop a scalable, transferable, and data-driven framework for urban noise prediction that leverages high-resolution multispectral remote sensing as its primary foundation and incorporates harmonised land-use/land-cover (LULC) information at later stages to strengthen cross-city generalisation. The emphasis throughout is on advancing machine learning methodologies—particularly graph-based deep learning and domain adaptation—to overcome the limitations of traditional modelling approaches. The

research proceeds through four interrelated objectives. First, it seeks to exploit multispectral satellite imagery as the principal source of noise-relevant urban features, while subsequently integrating Urban Atlas LULC data to enhance model comparability and transferability across cities. Second, it develops and compares a range of machine learning approaches, from ensemble methods to convolutional neural networks, in order to establish performance baselines and evaluate the added value of hierarchical spatial representations. Third, it extends predictive modelling to graph neural networks, assessing their ability to capture multi-scale dependencies and relational structures embedded in heterogeneous urban morphologies. Finally, it demonstrates the scalability of the framework through cross-city experiments that employ domain adaptation and pseudo-labelling, thereby addressing the critical challenge of generalisation in urban noise mapping.

1.5 Structure of the Dissertation

This dissertation is organised into seven chapters that together trace a progressive research pathway from baseline modelling to advanced deep learning and, ultimately, to transferable frameworks for cross-city noise prediction. Chapter 1 introduces the problem of urban noise pollution, situating it within broader public health, social equity, and sustainability debates, while also outlining the regulatory landscape defined by frameworks such as the Environmental Noise Directive. It highlights the limitations of existing deterministic models and explains how the integration of remote sensing and machine learning can offer scalable alternatives. The chapter concludes by defining the research gaps, objectives, and hypotheses that motivate the study.

Chapter 2 provides a systematic review of the relevant literature. It examines the evolution of noise modelling from traditional deterministic and empirical frameworks to contemporary data-driven approaches, assessing how remote sensing, machine learning, and deep learning have been used in environmental prediction. Particular attention is given to convolutional neural networks (CNNs) and graph neural networks (GNNs), as well as to emerging strategies for transfer learning and domain adaptation, thereby positioning the present research within ongoing methodological debates.

Chapter 3 sets out the methodological foundations of the study. It describes the construction of the Southampton dataset, including the collection and aggregation of noise measurements and the derivation of remote sensing features from WorldView-2 imagery and Urban Atlas land-use/land-cover data. It also outlines the baseline CatBoost model and introduces the theoretical principles of CNNs and GNNs, explaining how their architectural properties map onto the spatial and relational structure of urban environments. Preliminary validation experiments are presented to benchmark performance and inform subsequent model development.

The empirical contributions unfold across three core chapters. Chapter 4 applies CNNs to predict city-scale noise in Southampton, demonstrating how hierarchical spatial features extracted from multispectral imagery can improve predictive accuracy

over conventional machine learning methods. Chapter 5 advances the analysis by developing a GNN framework that models neighbourhood dependencies and relational structures within Southampton, using unaggregated noise measurements to show how graph-based learning enhances both spatial realism and predictive performance. Chapter 6 then extends the approach to multiple UK cities, proposing a dual-branch GNN with domain adaptation and pseudo-labelling to enable transferable noise prediction in contexts where ground-truth data are scarce.

Finally, Chapter 7 provides a general discussion that synthesises findings across the empirical studies. It assesses the theoretical and methodological contributions of the research, compares the performance and trade-offs of different modelling paradigms, and situates the results within international efforts to develop scalable urban noise mapping frameworks. The chapter also considers limitations, explores implications for policy and practice, and identifies directions for future research, before concluding with the broader significance of the dissertation.

Chapter 2: Literature Review

2.1 Introduction

The assessment and prediction of urban noise have long been central concerns in environmental science, urban planning, and public health policy. While Chapter 1 has established the health and societal consequences of environmental noise, the methodological trajectory of noise mapping requires deeper elaboration. A systematic literature review is essential to clarify how approaches to noise prediction have evolved, where their limitations lie, and how emerging techniques—particularly machine learning and graph-based models—create opportunities for new research directions.

The methodological landscape of noise prediction can be broadly divided into three phases. The first comprises traditional deterministic and empirical models, including ISO 9613-2 and CNOSSOS-EU, which remain the backbone of regulatory frameworks. These approaches provide physically interpretable outputs but require detailed local input data and extensive calibration, limiting their scalability and transferability across heterogeneous cities (Murphy & King, 2014; Kang et al., 2016).

The second phase reflects the incorporation of remote sensing and machine learning into environmental modelling. Satellite imagery offers synoptic and standardised information on urban morphology, land cover, and vegetation, all of which strongly influence acoustic propagation (Avtar et al., 2020). Machine learning methods, in parallel, allow predictive functions to be learned directly from data, reducing dependence on exhaustive field surveys and improving predictive accuracy for complex urban systems (Elith & Leathwick, 2009). However, their reliance on handcrafted features constrains their scalability, while their capacity for cross-city generalisation remains underdeveloped.

The third and most recent phase is characterised by the rise of deep learning, where convolutional neural networks (CNNs) have demonstrated strong capacity for extracting hierarchical spatial features from imagery, and graph neural networks (GNNs) have introduced the ability to model non-Euclidean dependencies such as road networks, adjacency relationships, and urban form heterogeneity (Zhu et al., 2017; Zhou et al., 2020). These advances suggest the possibility of scalable, transferable, and data-driven frameworks for noise prediction, but their application to environmental acoustics is still in its infancy, with few studies testing their robustness across different urban domains.

Despite this progression, a persistent gap remains: existing studies often focus narrowly on single-city case studies, rely excessively on descriptive reviews rather than empirical demonstrations, and seldom address the challenge of transferability. This is especially problematic for noise prediction, where urban heterogeneity generates significant domain shifts between cities. Addressing this gap requires a framework that not only leverages remote sensing and machine learning but also

systematically evaluates transferability through domain adaptation and pseudo-labelling.

Accordingly, this chapter reviews the methodological evolution of environmental noise prediction with explicit reference to the objectives of this dissertation. Section 2.2 evaluates the strengths and weaknesses of traditional deterministic and empirical noise models. Section 2.3 examines the integration of remote sensing in environmental monitoring, emphasising its potential to provide standardised and transferable predictors. Section 2.4 reviews the role of classical machine learning, while Section 2.5 discusses deep learning with particular attention to CNNs and GNNs. Section 2.6 considers the challenges of cross-city prediction and synthesises emerging solutions such as transfer learning, domain adaptation, and semi-supervised learning. Finally, Section 2.7 consolidates these insights, identifying key gaps that motivate the research design outlined in Chapter 3.

2.2 Traditional Noise Prediction

Traditional approaches to environmental noise prediction are grounded in deterministic acoustic models and empirical frameworks that formalise how sound propagates through the built environment. Among the most influential are the ISO 9613-2 standard (ISO, 1996), which specifies general methods for outdoor sound propagation, and the CNOSSOS-EU framework (Kephalopoulos et al., 2012), developed to harmonise environmental noise assessment across the European Union. These frameworks combine source emission data—such as traffic volumes, vehicle types, and industrial activity—with propagation equations that account for geometric spreading, atmospheric absorption, ground impedance, and reflections from buildings and terrain (Murphy & King, 2022). Their use has been closely tied to the implementation of the Environmental Noise Directive (END), which requires Member States to produce strategic noise maps for major roads, railways, airports, and urban agglomerations (Murphy & King, 2010).

A key strength of these models lies in their strong physical basis, which ensures interpretability, comparability, and acceptance by regulators. For instance, the extensive END-driven mapping programmes in cities such as Paris, Madrid, and Berlin have generated large-scale, harmonised datasets that support policy action plans (Dekonink et al., 2014; Nassur et al., 2019). Similarly, in the Netherlands, the integration of CNOSSOS-EU with detailed traffic flow data has enabled high-resolution mapping of road traffic noise, producing outputs that are not only consistent with theoretical acoustics but also widely accepted in environmental impact assessments (Salomons & Pont, 2012).

Beyond Europe, deterministic approaches have also been applied in rapidly urbanising contexts, albeit with mixed results. In Brazil, Carniel (2018) examined the practical implementation of ISO 9613-2-based models for urban noise mapping and highlighted multiple methodological challenges, including difficulties in collecting accurate traffic flow data, parameter uncertainty, and the poor transferability of standardized propagation assumptions to complex local environments. In China,

Zhang et al. (2023) applied a large-scale rotating mobile monitoring framework in Beijing to predict traffic noise using environmental and urban morphological data, demonstrating strong performance but also exposing limitations in modelling highly heterogeneous urban forms and dynamic traffic conditions. These examples illustrate the persistent tension between the theoretical robustness of deterministic acoustic models and their practical feasibility across diverse urban environments.

Empirical models, which complement or simplify deterministic approaches by fitting statistical relationships between observed noise and predictor variables, have also seen growing use. For example, Li et al. (2024) proposed a probabilistic prediction model for expressway traffic noise in Guangdong Province, China, which achieved low error rates (1–2 dB(A)) but remained dependent on site-specific traffic characteristics. While such empirical frameworks offer operational flexibility and reduced data requirements, their transferability remains limited by their localized calibration and sensitivity to contextual differences between cities.

Despite their contributions, both deterministic and empirical models face significant challenges. Accurate implementation requires detailed and often expensive datasets, including high-resolution traffic counts, building geometries, terrain models, and meteorological conditions. Such requirements render them resource-intensive and limit their scalability, especially in low-resource settings where noise mapping is needed most urgently (Murphy & King, 2014). Furthermore, their calibration parameters are highly context-specific, which undermines transferability: a CNOSSOS-EU model optimised for one European city may perform poorly in another with different urban density, building morphology, or transport systems. The inability to adapt easily to dynamic or heterogeneous urban soundscapes highlights a structural weakness of traditional models.

In summary, deterministic and empirical models remain essential for regulatory compliance and continue to dominate official mapping exercises under the END. However, their reliance on extensive ground data, resource-intensive calibration, and limited adaptability underscores their restricted scalability. The methodological shortcomings observed across case studies in Europe, South America, and Asia highlight the pressing need for alternative approaches. These limitations have motivated the search for frameworks that leverage remote sensing and data-driven modelling, which promise to balance interpretability with scalability and to extend noise prediction beyond single-city boundaries.

2.3 Remote Sensing in Environmental Monitoring

Remote sensing has become a cornerstone of environmental monitoring by offering synoptic, standardised, and repeatable observations across diverse spatial and temporal scales. Its ability to provide consistent geospatial information makes it particularly valuable for urban studies, where fine-grained heterogeneity in land cover, built morphology, and vegetation distribution strongly influences environmental processes, including noise propagation (Avtar et al., 2020; Li et al., 2016).

A central advantage of remote sensing lies in the richness and comparability of its data products. Multispectral and hyperspectral imagery from missions such as Sentinel-2, Landsat, and WorldView provides detailed spectral signatures of urban surfaces at spatial resolutions ranging from 1 m to 30 m. This enables systematic capture of variables highly relevant to noise dynamics, including impervious surface extent, vegetation indices such as NDVI, and morphological proxies of urban density. Preprocessing techniques—including atmospheric correction, radiometric calibration, and geometric co-registration—are critical for ensuring temporal and cross-sensor consistency (Chander et al., 2009; Vermote et al., 2016). More advanced harmonisation methods, such as pseudo-invariant features (PIFs) and BRDF adjustments, further allow multi-sensor integration across space and time (Roy et al., 2017). These procedures are indispensable when remote sensing is used not merely for local assessments but for comparative studies across multiple cities.

The effectiveness of remote sensing in environmental modelling has been demonstrated in a wide range of domains. In air pollution research, Di et al. (2019) used satellite-derived aerosol optical depth (AOD), meteorological fields, and land-use indicators to predict PM_{2.5} concentrations across the continental United States, showing that consistent imagery could compensate for sparse ground monitoring networks. Similar approaches have been applied in Europe and China, where satellite indices have been integrated with machine learning to predict air quality across multiple metropolitan areas (Hu et al., 2017; Xue et al., 2020). These studies provide strong methodological precedents for noise modelling: they illustrate how remote sensing enables environmental exposure assessments that are not bound to a single monitoring domain but are transferable across diverse geographies.

In urbanisation studies, remotely sensed time-series data have been widely employed to quantify land-use change, impervious surface expansion, and vegetation cover, all of which are directly relevant to noise propagation (Seto et al., 2012; Pesaresi et al., 2016). For example, comparative analyses of urban growth in European and Asian cities have shown that standardised satellite imagery can track urban density and green space loss in a manner that is directly comparable across regions (Zhou et al., 2015). These examples underscore the value of remote sensing for cross-city studies: the same spectral indices or land-cover classifications can be derived for London, Beijing, or São Paulo, ensuring that predictor variables are consistent even when urban contexts differ.

Although still limited, several studies have begun to explore the direct integration of remote sensing into noise prediction. Xie et al. (2011) demonstrated that remotely sensed land-cover features—such as vegetation cover and impervious surface ratios—could serve as significant predictors of urban noise levels in Guangzhou, China. More recently, Can et al. (2014) showed that NDVI and land-use classes derived from satellite data improved the explanatory power of statistical noise models in Paris, particularly in capturing attenuation effects from green buffers. In Europe, Suárez and Barros (2014) used Landsat imagery to derive urban form metrics and found significant associations with measured noise levels in Madrid. These examples remain relatively isolated compared to air quality research, but they demonstrate the feasibility of embedding remote sensing features directly into noise prediction frameworks.

The cross-city consistency of remote sensing data is particularly critical for advancing scalable noise prediction. Unlike traffic counts, meteorological datasets, or detailed cadastral records—which are highly city-specific—satellite imagery and harmonised land-use products such as the European Urban Atlas provide standardised information

across national and continental domains. This consistency allows predictor variables, such as road density, vegetation indices, and building cover, to be defined in comparable terms across multiple urban environments. For example, the Urban Atlas 2012 dataset provides harmonised land-use/land-cover classes for over 300 European cities, making it possible to embed LULC ratios into predictive models without extensive local recalibration (EEA, 2016). Such harmonisation directly addresses one of the central barriers in noise modelling: the lack of transferable input datasets that can support generalisation beyond single-city case studies.

In summary, remote sensing not only enriches environmental modelling with spatially exhaustive and repeatable predictors but also provides the methodological infrastructure for cross-city generalisation. Evidence from air quality modelling, urbanisation studies, and the emerging body of noise research highlights its unique capacity to produce comparable predictors across heterogeneous urban contexts. This consistency positions remote sensing as a critical empirical foundation for the data-driven, transferable noise prediction framework advanced in this dissertation.

2.4 Machine Learning in Environmental Modelling

Machine learning (ML) methods have increasingly been adopted in environmental modelling as flexible and data-driven alternatives to deterministic frameworks. Unlike physically based models, which require explicit formulation of propagation laws, ML approaches can learn complex, nonlinear relationships from empirical data (Elith and Leathwick, 2009). This makes them particularly well-suited to urban noise prediction, where environmental processes arise from the interaction of multiple heterogeneous factors—including road networks, land-use patterns, vegetation cover, urban morphology, and meteorology—whose combined effects are often too complex to be captured by analytical formulations alone.

Classical Machine Learning Algorithms in Environmental Modelling

Among the most widely applied ML algorithms in environmental research are Random Forests (RF), Support Vector Machines (SVMs), and Gradient Boosting Machines (GBMs). RF (Breiman, 2001) aggregates multiple decision trees to reduce variance and enhance robustness, handling high-dimensional, mixed-type datasets effectively. It has proven particularly valuable in ecological modelling (Cutler et al., 2007) and remote sensing classification, and has been successfully applied to particulate matter prediction by integrating satellite-derived variables with meteorological and land-use data (Di et al., 2019).

SVMs (Vapnik, 1998) use kernel functions to model nonlinear decision boundaries in high-dimensional spaces and are widely used in remote sensing-based land-cover classification (Mountrakis et al., 2011). GBMs (Friedman, 2001) sequentially fit weak learners to residuals, achieving high predictive accuracy and robustness in environmental applications including flood susceptibility mapping and habitat modelling. Collectively, these algorithms provide strong methodological foundations for noise modelling, which shares similar data structures and predictive requirements.

In air quality research, ensemble ML models have demonstrated that scalable exposure mapping can be achieved without dense ground monitoring networks by integrating satellite-based environmental indicators, land-use characteristics, and meteorological data (Di et al., 2019). These approaches provide a methodological precedent for noise mapping, which faces similar challenges of sparse observations and spatial heterogeneity.

Applications of ML in Urban Noise Prediction

Although the literature on ML-based noise mapping remains smaller than that on air pollution, recent work demonstrates that ML approaches have progressed from theoretical promise to operational practice. Yin et al. (2020) used dense mobile noise measurements combined with ML models to predict fine-scale traffic noise, capturing spatiotemporal patterns that deterministic simulations alone could not adequately represent. Umar et al. (2024) provided a systematic bibliographic overview of AI-based traffic noise prediction, documenting a marked increase in the use of ensemble and hybrid learning methods since 2010. Their review also highlighted the growing role of ML in complementing official mapping protocols, particularly in data-scarce or rapidly changing urban contexts.

Gradient Boosting Methods and Heterogeneous Predictors

Recent advances in gradient boosting algorithms—including XGBoost, LightGBM and CatBoost—have further expanded ML’s capability to handle complex environmental prediction tasks. CatBoost has drawn particular attention for its ability to natively process categorical variables, reducing preprocessing overhead and mitigating information loss from one-hot encoding (Prokhorenkova et al., 2018). This property is highly relevant to urban noise modelling, which relies heavily on categorical predictors such as land-use codes, road types, and morphological classes.

Gradient boosting algorithms have already demonstrated superior predictive power in related fields. In air pollution exposure modelling, they consistently outperform both RF and classical linear baselines (Di et al., 2019). In hydrological modelling, they have shown strong ability to capture nonlinear interactions in flood risk mapping (Mosavi et al., 2018). These empirical findings underscore their potential as high-performing, computationally efficient baselines for noise prediction, especially when integrating multisource geospatial data.

Feature Engineering and Cross-City Transferability

The predictive performance of ML in noise mapping is closely tied to feature engineering. Typical features include multi-ring buffered road density, vegetation indices derived from multispectral imagery, impervious surface ratios, and urban texture metrics. These variables act as proxies for acoustic processes—for example, green buffers attenuate noise, while dense road networks amplify exposure—but they are context-dependent. Feature sets optimised for one city often fail to generalise to another with different built-up morphology, traffic structure, or vegetation patterns.

Residual error analysis often reveals structured spatial patterns, indicating unmodelled physical processes such as building geometry effects, microclimate variability, and

traffic fluctuations. When models are applied across urban domains, performance commonly degrades, reflecting a domain shift problem well documented in environmental ML (Weiss et al., 2016). This limitation underscores the need for more transferable modelling strategies and adaptive architectures for scalable noise mapping.

Hybrid Modelling Approaches: Bridging Physics and ML

A growing body of research has explored hybrid modelling strategies that combine deterministic acoustic simulation with ML components to enhance scalability and efficiency. One approach involves data assimilation, in which mobile noise measurements are fused with simulation outputs to produce refined, temporally dynamic noise maps. For example, Can et al. (2018) demonstrated that integrating mobile phone-based measurements with simulation maps in Paris significantly reduced prediction errors and captured temporal variability that static models failed to represent.

A second strategy involves hybrid residual learning, where deterministic propagation models provide structured priors, and ML models learn residual patterns to improve prediction accuracy. Liu et al. (2021) applied this approach in Montreal, combining propagation modelling with random forests to produce high-resolution noise maps with improved generalisation. A third strand focuses on surrogate (meta-) modelling, using ML algorithms to approximate computationally expensive acoustic simulations. Surrogate models have been developed to emulate transmission loss, aerodynamic noise, and urban propagation, supporting sensitivity analysis and real-time mapping (Cunha et al., 2022; Poggi et al., 2022).

These developments reflect a broader methodological shift toward hybrid modelling paradigms in environmental science (von Rueden et al., 2019; Schweidtmann et al., 2024), in which physics-based structure is combined with data-driven flexibility. This hybridisation allows models to leverage domain knowledge while reducing data and computation requirements, making them particularly attractive for regulatory and large-scale urban noise applications.

Methodological Role of ML in This Dissertation

Within the methodological design of this dissertation, ML plays a dual role. First, it provides a pragmatic integration layer, enabling the fusion of remote sensing, transport, land-use, and morphological datasets into a predictive framework without the rigid data demands of deterministic models. Second, it offers a performance baseline against which the added value of more advanced deep learning architectures—particularly convolutional and graph neural networks—can be rigorously evaluated.

This dual role reflects ML's position at the interface between classical noise modelling and modern deep learning: flexible enough to handle heterogeneous inputs and nonlinear dynamics, but limited by its reliance on handcrafted features and weak cross-city transferability. Understanding and articulating these strengths and weaknesses is critical for developing more robust, scalable, and transferable urban noise prediction frameworks.

2.5 Deep Learning Methodologies

Deep learning has transformed environmental modelling by enabling the automatic extraction of hierarchical features and the representation of spatial and relational dependencies without extensive manual feature engineering. Unlike traditional machine learning methods, which depend heavily on handcrafted predictors, deep learning architectures learn directly from raw or minimally processed data, allowing for scalable, transferable, and high-dimensional representations (LeCun et al., 2015; Goodfellow et al., 2016). Within environmental science, two families of architectures have been particularly influential: convolutional neural networks (CNNs), which exploit Euclidean grid structures such as imagery, and graph neural networks (GNNs), which generalise learning to non-Euclidean relational structures such as spatial networks.

2.5.1 Convolutional Neural Networks in Remote Sensing and Environmental Science

Convolutional Neural Networks (CNNs) have emerged as one of the most powerful classes of deep learning architectures for environmental modelling due to their ability to learn hierarchical spatial representations directly from raw imagery. By applying local convolutional kernels in successive layers, CNNs efficiently capture both fine-grained textures and broad spatial patterns, enabling structured interpretation of complex landscapes without extensive manual feature engineering (LeCun et al., 2015; Goodfellow et al., 2016). This property is particularly advantageous in remote sensing, where environmental phenomena often exhibit both localised and multi-scalar spatial dependencies (Zhu et al., 2017).

CNNs in Environmental Remote Sensing

CNNs have been widely applied in remote sensing tasks such as land-use and land-cover (LULC) classification, ecological monitoring, flood mapping, and air quality estimation. In the domain of LULC mapping, CNNs consistently outperform traditional pixel-based classifiers by learning rich spectral–textural signatures. Nataliia Kussul and colleagues demonstrated that deep CNNs could classify land cover and crop types in Ukraine with significantly higher accuracy than conventional methods (Kussul et al., 2017). These findings were echoed in subsequent large-scale reviews and benchmarking studies, which highlighted CNNs as the backbone of modern environmental image interpretation (Vali et al., 2020).

In flood and hydrological modelling, CNN architectures—particularly U-Net and fully convolutional networks (FCNs)—have become standard tools. For instance, Gopal Konapala et al. (2021) explored the joint use of Sentinel-1 and Sentinel-2 imagery with deep CNNs for flood inundation mapping, demonstrating high

performance in delineating hydrological boundaries. More recent studies introduced residual and attention-enhanced U-Nets to improve boundary definition in complex floodplains (Jamali et al., 2024; Frame et al., 2024). These applications are of particular methodological relevance for environmental noise prediction because noise fields—like floodplains—are often structured along corridors (e.g. roads, railways) with sharp spatial gradients that require models capable of preserving boundary features.

CNNs have also seen increasing adoption in air pollution modelling. Deep architectures integrating satellite imagery, meteorological data, and ancillary land-use indicators have shown strong performance in PM_{2.5} estimation at continental scales (Muthukumar et al., 2021; Lee et al., 2024). Such models demonstrate CNNs' capacity to learn complex, nonlinear spatial associations from heterogeneous environmental predictors. These results provide a robust methodological precedent for noise mapping, where the aim is likewise to infer environmental exposures from spatially structured data.

CNNs and Acoustic Applications

While applications of CNNs to environmental noise prediction remain limited, emerging studies show their technical promise. Yu et al. (2024) developed a CNN model to estimate road traffic noise levels using minute-level traffic flow data in Foshan, China. Their model achieved a reduction in mean squared error of 10.16% compared to statistical baselines, highlighting CNNs' ability to learn fine-grained temporal-spatial dependencies in traffic-noise relationships. This is especially relevant for urban noise mapping, where diurnal fluctuations in traffic volumes are a dominant noise source.

In parallel, CNNs have been successfully applied to environmental sound analysis using spectrogram representations. Salamon and Bello (2017) demonstrated that deep CNNs, combined with data augmentation, significantly improved environmental sound classification accuracy. Although these studies do not directly estimate spatial noise fields, they underline the capacity of CNNs to capture structured patterns within acoustic data. Together, these examples indicate that CNNs can support noise prediction from both environmental proxies (e.g. imagery, traffic data) and sound-based features.

Architectural and Technical Considerations

CNNs' modelling advantages are closely tied to their architectural properties. Small receptive fields in early layers allow for detection of fine-scale features such as road segments, vegetation patches, or building edges, while deeper layers integrate larger-scale contextual information including block morphology and corridor alignment (Luo et al., 2016; Araujo et al., 2019). These multi-scale receptive field properties are particularly useful for noise modelling, as urban noise typically propagates through linear transport structures and interacts with heterogeneous built forms.

However, CNNs also have inherent limitations. Their reliance on Euclidean grids constrains their ability to represent non-Euclidean relationships, such as oblique transport corridors that intersect multiple urban blocks, or topological connections

between distant but acoustically linked locations. Enlarging receptive fields to capture broader context can lead to over-smoothing, which may obscure the steep gradients that commonly characterise urban noise distributions. Moreover, interpreting CNN feature maps in terms of physical acoustic processes remains challenging—filters encode spatial structure but not explicit propagation laws—limiting their direct regulatory interpretability.

In summary, CNNs have achieved widespread adoption in environmental modelling due to their ability to learn structured, multi-scale representations from imagery. Their demonstrated success in LULC classification, hydrological modelling, and air quality prediction provides a solid methodological foundation for their application to environmental noise prediction. Although the number of studies explicitly focusing on acoustic domains is still modest, early applications in traffic noise estimation and environmental sound analysis illustrate clear potential. These strengths make CNNs an essential building block in scalable noise prediction frameworks, though their geometric rigidity and interpretability challenges motivate the complementary use of graph-based architectures in subsequent sections.

2.5.2 Graph Neural Networks in Spatial Environmental Modelling

While convolutional neural networks (CNNs) excel at extracting spatial information from regular Euclidean grids, many environmental processes are inherently relational and structured through irregular, non-Euclidean dependencies such as road networks, hydrological connectivity, landscape adjacency, or directional acoustic propagation pathways. Graph neural networks (GNNs) extend deep learning into graph domains by representing spatial units as nodes and relational dependencies as edges (Kipf, 2016; Zhou et al., 2020). In urban noise prediction, for example, 30 m grid cells or land parcels can be represented as graph nodes with feature vectors derived from remote sensing and ancillary data, while edges encode spatial contiguity or road-network connectivity. By propagating information through these edges, GNNs can capture both local and non-local dependencies, thereby addressing the limitations of grid-based CNNs.

At the algorithmic core of GNNs lies iterative message passing, in which node embeddings are updated by aggregating and transforming information from their neighbourhood (Gilmer et al., 2017). Shallow architectures typically capture localised interactions—such as the influence of a busy arterial road on adjacent residential blocks—while deeper architectures can encode broader contextual effects across entire neighbourhoods or districts. However, increasing model depth can introduce structural limitations. Over-smoothing causes node embeddings to become indistinguishable across the graph (Oono and Suzuki, 2020), whereas over-squashing results in the compression of long-range information into limited latent dimensions (Alon and Yahav, 2021). To address these challenges, advanced architectures employ techniques such as residual or skip connections, jumping knowledge aggregation, and graph attention mechanisms, which enhance representational expressiveness while maintaining stability (Xu et al., 2018; Veličković et al., 2018).

Over the past five years, GNNs have achieved rapid uptake across multiple domains of environmental science. In air quality modelling, GNN-based frameworks have been used to forecast pollutants such as PM_{2.5} and NO₂ by treating monitoring stations as

nodes and encoding spatial–temporal dependencies through graph edges derived from traffic, meteorology, and land-use features. For instance, Calo et al. (2024) implemented a message-passing architecture to improve spatial interpolation in complex urban settings; Wang et al. (2024) combined graph attention with temporal sequence models to capture fine-grained spatiotemporal dependencies; and Liu et al. (2023) demonstrated that adaptive graph learning improves generalisation under irregular monitoring networks. These studies consistently show that graph representations outperform purely grid-based or regression-based approaches when spatial structure is complex and heterogeneous.

In hydrological forecasting, graph-based representations have similarly proven advantageous. Hamed Farahmand and colleagues developed an attention-based spatio-temporal graph convolutional network for urban flood nowcasting, leveraging community-scale features to outperform traditional baselines (Farahmand et al., 2023). Building on this, Jiang et al. (2024) introduced heterogeneous dynamic graph convolutional networks that integrate remote-sensing covariates with river network connectivity, demonstrating enhanced forecasting skill and transferability across catchments. These applications illustrate the particular suitability of graph structures in settings where process connectivity and anisotropic propagation play central roles.

GNNs have also become fundamental tools in transportation modelling, which shares strong structural similarities with urban noise propagation. Yu et al. (2018) proposed spatio-temporal graph convolutional networks (STGCN) for traffic flow prediction, showing superior performance compared with CNN or RNN baselines, particularly in capturing long-range dependencies along transport corridors. Li et al. (2018) advanced this approach with diffusion convolutional recurrent neural networks (DCRNN), modelling non-Euclidean information flow in traffic networks. These studies provide a direct methodological parallel for noise prediction, where traffic networks serve as primary sound-propagation channels.

In the domain of noise and environmental exposure, recent work by Song et al. (2025) used interpretable spatio-temporal GNNs to examine cumulative environmental exposures affecting noise perception. Their framework integrates message passing across both space and time, leading to higher predictive fidelity and improved interpretability compared to conventional machine learning approaches. This work directly aligns with multi-city noise mapping pipelines, which require modelling of both spatial adjacency and transport connectivity to achieve robust generalisation.

A key strength of GNNs lies in their structural flexibility. Unlike CNNs, which are constrained by fixed receptive fields, GNNs can explicitly model complex topologies such as diagonal roads, hierarchical street networks, or acoustically coupled but morphologically dissimilar areas (Bronstein et al., 2017). Moreover, advances in model explainability—such as GNNExplainer (Ying et al., 2019)—offer tools for identifying subgraphs and features that drive predictions, an essential property for linking learned representations to physical acoustic mechanisms. Nonetheless, GNNs also face scalability and interpretability challenges, including training instability in deep networks and the difficulty of maintaining expressivity on very large graphs.

GNNs and AIGC: Emerging Synergies

An emerging frontier in environmental modelling involves the integration of GNNs with generative AI frameworks to address data sparsity, scenario simulation, and domain adaptation. Recent work has explored graph generative diffusion models, such as DiGress (Vignac et al., 2022) and GDSS (Jo et al., 2022), which can synthesize realistic graph structures while preserving distributional and topological properties. These methods can be used to generate plausible but unseen urban topologies, augmenting training data for cross-city noise mapping. In addition, controlled remote sensing diffusion models (Tang et al., 2024) enable the joint generation of imagery and graph representations, which is particularly valuable when transferring models across cities with different urban morphologies.

From a methodological perspective, these approaches signal a clear convergence between generative modelling and graph-based environmental learning, providing new tools to address one of the central bottlenecks in scalable urban noise mapping: the scarcity of sufficiently diverse, well-labelled, and topologically rich training data. By integrating graph generative diffusion models with environmental GNN frameworks, researchers can simulate plausible urban morphologies, augment training datasets, and reduce domain shift in cross-city transfer scenarios. More broadly, recent surveys on remote sensing and geoscience foundation models highlight the potential of combining large-scale generative modelling with structured graph learning to build adaptive, transferable representations capable of supporting complex, multi-domain environmental applications (Lu et al., 2025).

2.5.3 Towards Generalisable Models: World Models and Hybrid Frameworks

Although CNNs and GNNs offer significant methodological advances for environmental modelling, their applications in urban noise prediction remain relatively limited. As reviewed in Sections 2.5.1 and 2.5.2, CNNs excel at learning hierarchical spatial representations from remotely sensed imagery, while GNNs effectively capture relational dependencies across irregular urban morphologies. However, both approaches face similar challenges when transferred to new urban contexts: domain shift caused by variations in building density, land-use configuration, and transportation structure often leads to degraded predictive performance. This limited generalisability underscores a central research gap: how to develop scalable, transferable models that remain robust across multiple cities.

Lessons from Environmental Domains

In environmental sciences, the need for transferable models has driven substantial methodological innovation. In air quality modelling, domain adaptation and transfer learning have enabled PM_{2.5} prediction to generalise across cities with different meteorological and land-use patterns (Ni et al., 2022). In hydrological forecasting, integrating physical constraints into deep learning frameworks has allowed models to predict in ungauged catchments with improved stability and accuracy (Read et al., 2019; Karniadakis et al., 2021). Similarly, in climate and Earth system modelling, hybrid approaches that combine data-driven architectures with physical priors have demonstrated robustness across geographies (Reichstein et al., 2019).

A parallel development in AI research has been the rise of world models, a conceptual framework that aims to build compact, generalisable latent representations of environments rather than learning city-specific mappings. This concept was notably popularised by David Ha and Jürgen Schmidhuber (Ha and Schmidhuber, 2018), and more recently advocated by Fei-Fei Li as central to the next generation of AI that “understands and reasons about the physical world” through unified perception, structure, and prediction. World models emphasise learning mechanism-centred rather than purely data-centric representations, enabling adaptation to novel domains with minimal retraining. For environmental modelling, this means moving from isolated city-level predictions toward frameworks that internalise transferable acoustic and morphological structures.

Hybrid Architectures and Domain Adaptation

One of the most promising methodological pathways for achieving this transferability is hybrid modelling, particularly combining CNNs and GNNs. CNN–GNN hybrids exploit complementary strengths: CNN modules can extract multi-scale spatial hierarchies from imagery, while GNN modules model topological connectivity such as road networks, land-use adjacency, and propagation corridors. For example, Liu et al. (2022) demonstrated that fusing CNN-derived land-cover embeddings with graph-based relational structures significantly improved urban land-use inference. In transportation research, spatio-temporal graph convolutional networks (Yu et al., 2018) and diffusion convolutional recurrent neural networks (Li et al., 2018) have provided scalable blueprints for capturing both Euclidean and non-Euclidean dependencies, which can be adapted for acoustic propagation modelling.

In addition to architectural innovations, domain adaptation techniques have proven critical for improving cross-city generalisation. Adversarial domain alignment (Ganin et al., 2016) and hybrid transfer strategies (Ni et al., 2022) reduce distributional mismatch between source and target cities, allowing models trained on data-rich environments to perform more reliably in data-scarce contexts. Semi-supervised learning and pseudo-labelling offer further flexibility: by assigning high-confidence labels to unlabeled target-domain data, models can refine their internal representations and bridge source-target gaps without extensive new surveys (Li et al., 2021; Xu et al., 2022). These approaches have already shown success in air-quality and environmental-modelling contexts and can be translated to urban noise prediction.

Implications for Urban Noise Modelling

While these generalisable strategies have been widely explored in air quality, hydrology, and climate research, their application in urban noise modelling remains at an early stage. To date, most noise prediction studies employ models calibrated within a single city, rarely addressing domain shift explicitly. This gap is striking, given that noise propagation dynamics—driven by transport networks, built-up morphology, and land-cover structure—closely mirror other environmental exposure processes.

The world model paradigm offers a compelling direction for noise research. Instead of learning city-specific functions, models can be trained to capture invariant or transferable representations of how acoustic energy interacts with urban structures. Hybrid CNN–GNN architectures can encode both local physical context and

relational connectivity, while domain adaptation techniques can align inter-city feature distributions. Moreover, the increasing availability of harmonised datasets—such as European Environment Agency Urban Atlas, global road networks, and satellite time series—provides a consistent input foundation for transferable modelling.

Outlook: Towards Scalable Generalisation

The progression from deterministic models, to machine learning baselines, to deep spatial-relational architectures, and now to world model-inspired hybrid frameworks represents a broader evolution in environmental modelling. For urban noise prediction, this shift highlights the importance of building models that are not only accurate within a single domain but generalise across heterogeneous urban environments. Integrating CNN and GNN components with world model principles and domain adaptation techniques offers a structured pathway towards such scalable and transferable architectures.

These approaches establish the conceptual and methodological foundation for the cross-city generalisation experiments introduced in Section 2.6 and implemented empirically in Chapter 6 of this dissertation.

2.6 Cross-City Prediction and Transfer Learning

A central challenge in environmental noise prediction is ensuring model transferability across heterogeneous urban domains. While models trained on a single city can yield high local accuracy, their performance often deteriorates sharply when applied to other cities. This degradation arises because the predictor-response relationships underlying noise propagation are strongly context-dependent. For instance, vegetation-induced attenuation varies according to climatic conditions, building morphology, and urban density, while the effect of traffic composition depends on local transport systems and fleet characteristics. Such domain-specific variability creates significant distributional shifts between cities, thereby limiting the scalability of single-city models. This motivates the adoption of transfer learning and domain adaptation strategies to achieve cross-city generalisation.

2.6.1 Transfer Learning in Environmental Prediction

Transfer learning (TL) provides an effective means of overcoming domain shift by reusing knowledge learned in a source domain to improve performance in a target domain (Pan and Yang, 2010). Rather than retraining models from scratch for each new city, TL leverages pretrained representations through fine-tuning or feature adaptation. In remote sensing, this approach has been widely validated. Dimitrios Marmanis and colleagues demonstrated that CNNs pretrained on ImageNet can be effectively fine-tuned for land-cover classification with limited labels, achieving competitive accuracy in data-sparse regions (Marmanis et al., 2016). Similarly, domain-adaptive remote sensing benchmarks such as LoveDA have shown that pretrained deep networks can be successfully transferred between urban and rural

domains, supporting robust land-cover segmentation under shifting data distributions (Wang et al., 2021). These results establish a transferable paradigm for spatial representation learning, which is directly relevant for noise mapping.

Transfer learning has also proven particularly impactful in environmental exposure modelling. Recent studies have applied TL frameworks to hyperlocal air pollution mapping, enabling models trained in one metropolitan area to generalise to others with minimal target-domain labels (Yuan et al., 2024). This approach aligns closely with the methodological goals of urban noise prediction, where measurement data are sparse and costly to obtain.

2.6.2 Domain Adaptation and Adversarial Approaches

Domain adaptation (DA) explicitly addresses the problem of distributional divergence between source and target domains by aligning their feature spaces during training. A canonical and widely adopted approach is the Domain-Adversarial Neural Network (DANN), which introduces a gradient reversal layer (GRL) between the feature extractor and a domain classifier (Ganin et al., 2016). During backpropagation, the GRL inverts the gradient coming from the domain classifier, forcing the feature extractor to learn representations that confuse the domain discriminator while still minimising the task-specific loss in the source domain. In effect, the feature extractor learns to maximise domain confusion and minimise prediction error simultaneously, resulting in domain-invariant embeddings that generalise to unseen target distributions. This adversarial alignment is particularly advantageous in urban noise prediction, where spectral–textural patterns, land-use structure, and transport network characteristics can differ markedly across cities. By encouraging invariance in the latent space, DANN reduces sensitivity to local morphological idiosyncrasies and improves model robustness under domain shift.

Another widely adopted DA technique is the Maximum Mean Discrepancy (MMD) (Long et al., 2015), which measures divergence between source and target feature distributions in reproducing kernel Hilbert space. By minimising MMD, models learn shared latent representations without requiring adversarial training. In remote sensing, MMD-based domain alignment has been successfully applied to spatial and temporal crop-type mapping using Sentinel-2 imagery (Wang et al., 2023). More advanced variants, such as Local Maximum Mean Discrepancy (LMMD), have further improved cross-domain alignment in multi-modal settings involving optical and SAR data (Feng et al., 2024). These approaches provide a robust methodological foundation for cross-city noise prediction, where domain shifts are driven by differences in building density, land-use structure, and spectral–textural characteristics.

2.6.3 Semi-Supervised Learning and Pseudo-Labelling

Semi-supervised learning (SSL) extends transfer learning by incorporating unlabelled target-domain data. Among SSL techniques, pseudo-labelling (Lee, 2013) has gained

particular traction for its simplicity and effectiveness. In this framework, a model trained on the source domain generates predictions for unlabelled target data; high-confidence predictions are then treated as pseudo-labels for subsequent training iterations. This iterative self-training aligns the model to the target domain without requiring extensive new measurements.

In remote sensing, pseudo-labelling has been shown to improve semantic segmentation through consistency regularisation and iterative label refinement (Wang et al., 2022). Kwak and Park (2022) combined adversarial domain adaptation with pseudo-labelling for crop classification, achieving significant performance gains under domain shift. Such methods are directly applicable to urban noise prediction: harmonised datasets—such as European Environment Agency Urban Atlas and multispectral imagery—can provide a stable input foundation, while pseudo-labelling offers a mechanism to adapt source-trained models to unlabelled target cities with minimal cost.

2.6.4 Cross-Domain Representation Learning and AIGC Integration

The emergence of world models and AI-generated content (AIGC) provides new opportunities for improving cross-city generalisation. World models capture transferable latent structures across domains by learning compressed spatio-temporal representations that can adapt to novel environments (Ha and Schmidhuber, 2018; Battaglia et al., 2018). In the context of urban noise prediction, such models can learn abstract representations of vegetation buffering, traffic morphology, and urban structure that remain stable across different cities.

Generative models such as graph diffusion frameworks (Vignac et al., 2022) and controllable remote sensing diffusion models (Tang et al., 2024) can further enhance cross-city transfer by generating plausible synthetic scenarios in target domains. This includes synthesising land-cover–traffic configurations for under-monitored areas, which can then be used to pre-align and fine-tune predictive models. By coupling domain adaptation, pseudo-labelling, and AIGC-based augmentation, it becomes possible to build robust and transferable noise prediction frameworks that generalise well across highly heterogeneous urban landscapes.

2.6.5 Research Gaps and Methodological Outlook

Despite notable progress in transfer learning and domain adaptation across air quality and remote sensing research, cross-city generalisation remains underexplored in noise prediction. Existing studies are largely confined to single-city modelling, lacking systematic evaluations across different urban morphologies. This stands in sharp contrast to fields such as crop mapping or air pollution modelling, where transfer learning and domain adaptation have become mainstream methodological tools.

This dissertation explicitly treats cross-city noise prediction as a core research objective. By integrating harmonised multispectral and land-use/land-cover datasets

with graph-based neural architectures, adversarial and MMD-based domain adaptation, and pseudo-labelling strategies, it aims to establish a generalisation-oriented framework for scalable urban noise modelling. Such an approach moves beyond local case studies, enabling transferable prediction across cities, and laying the foundation for data-driven environmental noise management at regional to national scales.

2.7 Summary of the Literature Review

This chapter has traced the methodological evolution of environmental noise prediction, charting a shift from deterministic acoustic modelling to machine learning, deep learning, and emerging transfer learning strategies. Each stage of this progression reflects both technological advances and persistent limitations, revealing how the field has moved from physically explicit but rigid frameworks toward data-driven, transferable, and scalable modelling paradigms.

Deterministic acoustic frameworks such as ISO 9613-2 and CNOSSOS-EU remain the regulatory backbone for urban noise mapping across Europe, providing physically interpretable and standardised predictions (Murphy and King, 2014). However, their dependence on granular local inputs—including traffic flow, building geometries, and ground impedance—renders them data-intensive and context-specific, significantly limiting their scalability to heterogeneous urban environments (Zannin and de Sant’Ana, 2011; Murphy and King, 2014). Empirical statistical models offer more flexible alternatives but remain vulnerable to local overfitting and weak extrapolation capacity.

The incorporation of remote sensing has provided a crucial foundation for scaling noise prediction by supplying synoptic, harmonised, and transferable geospatial predictors. Standardised multispectral imagery, vegetation indices, and harmonised land-use/land-cover datasets such as European Environment Agency Urban Atlas enable cross-city comparability (Avtar et al., 2020; Gong et al., 2013). This integration mirrors the trajectory seen in air quality modelling (Di et al., 2019; Hu et al., 2017), where remotely sensed indicators provide robust input features for scalable exposure mapping. Although direct noise applications remain fewer than in air quality, existing studies demonstrate the explanatory strength of LULC and vegetation predictors for capturing attenuation effects and urban form influences (Margaritis and Kang, 2018; Liu et al., 2023).

Classical machine learning (ML) has marked a further shift toward flexible, data-driven predictive functions. Ensemble learners such as Random Forests, Support Vector Machines, and Gradient Boosting have been shown to handle heterogeneous environmental predictors effectively, reducing dependence on explicit acoustic propagation equations (Elith and Leathwick, 2009; Di et al., 2019). However, their reliance on handcrafted features constrains their scalability and limits their adaptability to new cities. Feature sets optimised for one urban morphology often fail to generalise to another with different density, transport structure, or vegetation patterns (Weiss et al., 2016).

Deep learning architectures—particularly CNNs and GNNs—have transformed this modelling landscape by enabling automatic feature learning and the representation of both Euclidean spatial hierarchies and non-Euclidean relational dependencies. CNNs have demonstrated strong performance in land-cover mapping, air quality estimation, and hydrological forecasting (Kussul et al., 2017; Muthukumar et al., 2021; Konapala et al., 2021), and their early use in traffic noise estimation has shown measurable gains in predictive accuracy (Yu et al., 2024). Yet CNNs’ Euclidean receptive fields make them less effective at modelling anisotropic acoustic pathways shaped by road networks and irregular morphologies.

Graph neural networks (GNNs) address these limitations by modelling relational structure and topological connectivity. Their uptake in air quality forecasting (Calo et al., 2024; Wang et al., 2024), hydrology (Farahmand et al., 2023; Jiang et al., 2024), and transportation modelling (Yu et al., 2018; Li et al., 2018) has demonstrated superior performance when spatial processes are structured along networks or corridors. Emerging work on interpretable spatio-temporal GNNs for noise exposure mapping (Song et al., 2025) further confirms their methodological promise for generalisable noise modelling.

An important methodological frontier has emerged in cross-city generalisation, where transfer learning, domain adaptation, and semi-supervised pseudo-labelling offer pathways to overcome domain shift. Techniques such as adversarial domain alignment (Ganin et al., 2016) and Maximum Mean Discrepancy (Long et al., 2015; Feng et al., 2024) enable feature space alignment between source and target cities, while pseudo-labelling approaches (Wang et al., 2022; Kwak and Park, 2022) allow models to iteratively adapt to new domains with minimal labelled data. These strategies are particularly well suited to noise prediction, where harmonised remote sensing data are abundant but direct acoustic measurements are scarce.

More recently, world models and generative AI (AIGC) have opened new possibilities for scalable and adaptive modelling. By learning latent representations of environmental structure and processes, world models (Ha and Schmidhuber, 2018; Battaglia et al., 2018) support transferable inference across urban contexts. Generative diffusion models for graphs (Vignac et al., 2022) and remote sensing imagery (Tang et al., 2024) offer complementary means of simulating plausible urban morphologies and augmenting scarce training datasets, further mitigating domain gaps in cross-city transfer scenarios.

Taken together, these methodological developments reveal a clear trajectory of increasing generalisability and structural flexibility. Deterministic models provide physically interpretable baselines but lack scalability; remote sensing offers standardised predictors across cities; machine learning provides flexible but feature-dependent mappings; deep learning extends this flexibility into automated feature learning and network representations; and transfer learning plus generative modelling point toward truly generalisable frameworks.

This dissertation is situated precisely at this frontier. Its research design—progressing from harmonised remote sensing inputs, to ML benchmarking, to CNN–GNN architectures, and finally to cross-city transfer through domain adaptation and pseudo-labelling—directly addresses the structural limitations identified in the literature. By

doing so, it contributes both methodological innovations and practical pathways for scalable, transferable noise mapping capable of informing evidence-based urban environmental policy at local, national, and continental levels.

Chapter 3: Research Methodology and Preliminary

Validation Experiments

3.1 Introduction

This chapter details the data and methods that underpin the empirical studies in this dissertation. Its primary aim is to articulate a coherent, end-to-end framework that links observational noise data, multispectral remote-sensing predictors, and machine-learning (ML) and deep-learning (DL) models into a reproducible workflow for city-scale noise prediction. The chapter therefore serves two complementary functions. First, it establishes the observational context and dataset construction, including spatial alignment between ground measurements and satellite imagery, energy-domain aggregation of acoustic levels to a common grid, and exploratory analyses to characterise sampling coverage, potential biases, and spatial dependence. Second, it formalises the methodological foundations that are used subsequently: (i) tree-based ensemble learning as a transparent baseline (CatBoost), (ii) convolutional neural networks (CNNs) with explicit attention to how convolutional kernel size, dilation, and depth determine the receptive field in physical units, and (iii) graph neural networks (GNNs) that model neighbourhood dependencies on a 30 m lattice and whose depth governs feature aggregation and the risk of over-smoothing.

The observational core of this work is a city-scale dataset for Southampton, United Kingdom, comprising more than fifty thousand in-situ measurements acquired through systematic pedestrian surveys and professional acoustic instrumentation (Alvares-Sanches et al., 2021a). These measurements provide georeferenced equivalent continuous sound levels (LAeq) under controlled meteorological conditions and form the ground truth against which models are trained and assessed. To obtain spatially exhaustive predictors, the study integrates WorldView-2 multispectral imagery—preprocessed to surface-reflectance units using standard radiometric and geometric corrections—and harmonised land-use/land-cover (LULC) information from the Urban Atlas programme (Chander et al., 2009; EEA, 2018). The

imagery supports the derivation of spectral and textural indicators of urban morphology (e.g., vegetation, imperviousness, fine-scale texture), while cartographic layers such as road networks provide structural constraints relevant to traffic-related noise exposure. All inputs are co-registered to a 30 m × 30 m grid; LAeq values are aggregated by averaging in the linear energy domain prior to conversion back to decibels to avoid arithmetic bias. This spatial harmonisation allows direct coupling of image-based features with acoustic measurements and enables consistent modelling across the study area.

Methodologically, the chapter advances from interpretable baselines to representation-learning approaches. CatBoost is used as a strong, tabular baseline that integrates handcrafted spectral/texture and contextual variables with minimal preprocessing and well-defined hyperparameters (Prokhorenkova et al., 2018). CNNs are then introduced to learn hierarchical spatial features directly from image patches. Particular emphasis is placed on mapping theoretical and effective receptive fields to urban physical scales—street-edge micro-structures ($\approx 10\text{m}$), block-scale configurations ($\approx 50\text{--}150\text{ m}$), and corridor-like transport features—so that kernel size, dilation, and depth are chosen to capture the spatial context most relevant to sound propagation (Araujo et al., 2019; Luo et al., 2016). Finally, GNNs are employed to model interactions across the 30 m lattice (and optional road-network edges), where each layer expands a node’s field of view by one hop; depth thus trades off contextual reach against the well-known risks of over-smoothing and information bottlenecks (Kipf, 2016; Alon and Yahav, 2021). These models collectively support a progression from feature-engineered predictors to scale-aware image encoders and topology-aware relational learners.

To ensure transparency and reproducibility, this chapter is structured to gradually build from data foundations to modelling strategies and experimental validation. Section 3.2 introduces the datasets, spatial preprocessing steps, and exploratory analyses, including a compact table of representative noise measurements to illustrate schema and summary statistics, together with a discussion of sampling representativeness and potential sources of bias such as corridor-heavy transects or time-of-day effects. Section 3.3 then develops the machine learning and feature learning framework, beginning with the CatBoost baseline, advancing to convolutional neural networks (CNNs) with receptive-field-to-physical-scale mapping, and extending to graph neural networks (GNNs) that propagate information through neighbourhood dependencies and network depth. This framework also incorporates two transversal elements: a unified feature map computation pipeline, and a comparative treatment of data partitioning strategies across models, which clarifies how evaluation protocols were adapted to the structural characteristics of each modelling approach. Section 3.4 presents preliminary experiments that motivated subsequent design choices, expanding on hyperparameter effects and demonstrating how each modelling component integrates into the end-to-end workflow. A schematic flowchart summarising acquisition, preprocessing, feature construction, model training, and evaluation is provided for reference. Finally, Section 3.5 offers a concise synthesis, connecting the methodological foundations of this chapter to the empirical analyses developed in Chapters 4, 5, and 6.

3.2 Data Sources and Study Areas

3.2.1 Study Area: Southampton

The primary and only explicitly described study area of this dissertation is Southampton, a major port city on the southern coast of the United Kingdom. Southampton exhibits a heterogeneous urban morphology, encompassing a dense commercial centre, arterial transport corridors, residential neighbourhoods, port and industrial zones, and extensive green and coastal areas. This diversity makes the city an ideal testbed for exploring scalable approaches to urban noise prediction.

While the overall workflow is extended in later chapters to Cardiff, Portsmouth, Liverpool, and Nottingham to assess cross-city generalisation, it is important to emphasise that Southampton is the only city for which ground-truth noise measurements are available. The additional cities are introduced and discussed exclusively in the context of cross-city modelling and domain adaptation in Chapters 5 and 6. This chapter therefore focuses solely on Southampton to provide a detailed and replicable account of data sources, preprocessing steps, and methodological foundations.

The remote sensing data underpinning this study consist of WorldView-2 multispectral imagery, acquired on 15 February 2016 and 6 May 2016. Both scenes had cloud cover below 5% and were seamlessly merged using the mosaic function in ENVI to generate a cloud-free composite covering the entire study area. WorldView-2 imagery provides eight spectral bands at a spatial resolution of 1.6–2.0 m, which were resampled to 4 m for the computation of spectral indices and grey-level co-occurrence matrix (GLCM) texture measures. For visualisation, an RGB band combination was prepared to illustrate the extent and land cover of Southampton (Figure 3.1).

Image preprocessing followed standard remote-sensing protocols (Jensen, 2007). Raw digital numbers were converted into surface reflectance using the FLAASH atmospheric correction algorithm in ENVI (Anderson et al., 2002), which combines radiometric calibration with atmospheric modelling. Subsequent geometric correction and reprojection ensured accurate alignment with the British National Grid (EPSG:27700), converting from the native UTM coordinate system to BNG for consistency with ground observations. These steps produced a seamless, georeferenced multispectral dataset suitable for deriving predictor variables.



Figure 3.1 RGB composite of WorldView-2 imagery (pre-processed and mosaicked) illustrating the Southampton study area

3.2.2 Noise Observations

The ground-truth dataset for this research is derived from in-situ environmental noise surveys conducted in Southampton during 2016 (Alvares-Sanches et al., 2021a). The campaign produced 52,364 raw noise recordings, collected along pedestrian transects distributed across the city. Measurements were acquired during representative time periods (morning peak, afternoon, and evening peak) under meteorological conditions minimising wind and precipitation, thereby reducing atmospheric interference. All recordings were made using calibrated professional acoustic equipment (Fostex FR-2LE, TASCAM DR-40 recorders, PCB signal conditioners, and Class-1 microphones), with microphones mounted at 1.65 m above ground to approximate ear-level exposure.

Each observation is accompanied by a georeferenced location derived from GPS logging, with 1 meter positional accuracy. This spatial precision allowed the dataset to be integrated seamlessly with high-resolution remote-sensing imagery. However, to ensure comparability across data sources, the noise measurements were aggregated to a 30 m × 30 m analysis grid, which serves as the baseline spatial unit in this dissertation. The aggregation was performed in the energy domain: individual decibel values were first converted to linear sound intensities, averaged within each grid cell, and then reconverted into the decibel scale. This avoided the bias inherent in arithmetic averaging of logarithmic units. After aggregation, the dataset comprised 13,474 grid cells with associated LAeq values.

It is important to note that while this 30 m aggregation serves as the baseline for analysis in Chapter 4 (CNN-based modelling) and Chapter 6 (dual-branch GNN with domain adaptation), the preliminary GNN experiments in Chapter 5 used the original

unaggregated point-level measurements. This allowed for testing how graph-based models perform when applied directly to irregular observational data.

The spatial coverage of aggregated noise cells is shown in Figure 3.2, where the 30 m $\times$ 30 m observation grid is overlaid on an OpenStreetMap basemap. Survey coverage was densest along arterial road networks and in the central urban districts, reflecting the design of pedestrian transects. Peripheral areas such as suburban fringes and large green spaces were comparatively under-sampled, a bias that has implications for model representativeness (see Section 3.2.5).

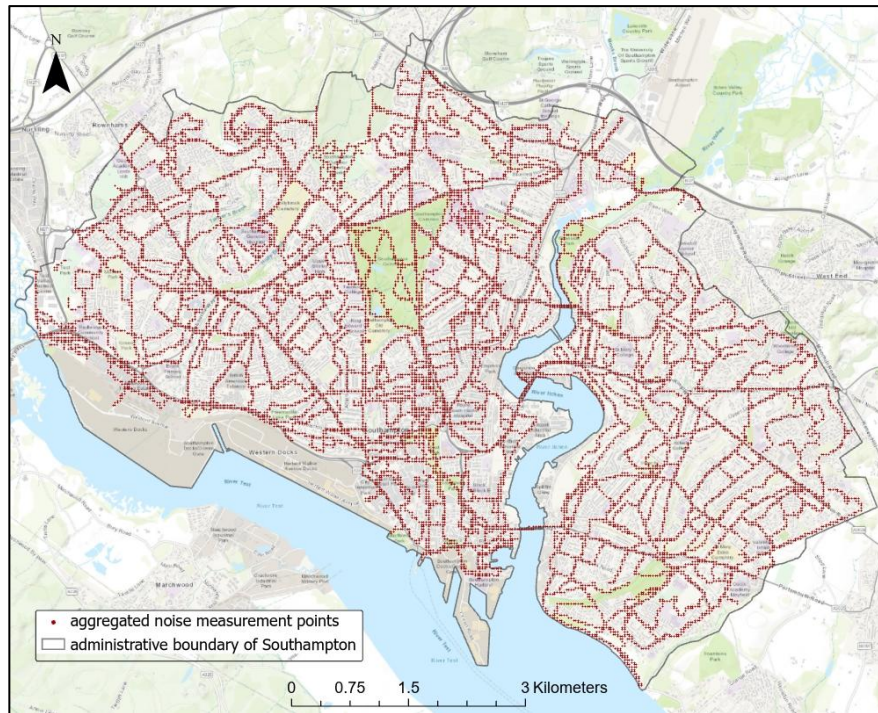


Figure 3.2 Spatial distribution of 30 m $\times$ 30 m aggregated noise observation cells across Southampton, overlaid on OpenStreetMap

To further illustrate the raw observational dataset prior to grid aggregation, Table 3.1 presents five representative samples at the original 1 m spatial resolution. Each record includes timestamp, geographic coordinates, and the measured equivalent continuous sound level (LAeq). These records demonstrate the schema of the dataset and highlight the high spatio-temporal precision of the field campaign.

Table 3.1 Example of Noise Observations (subset of Southampton dataset)

ID	Date	Time	Latitude	Longitude	LAeq (dB)
0	02/08/2016	06:03	442833	114957	43.5
1311	12/08/2016	07:53	438910	114862	67.1
4318	29/07/2016	06:29	441220	112994	66.7
19437	17/08/2016	13:37	437442	113664	56.3
41084	02/08/2016	16:42	439772	114609	72.6

Temporal dynamics and implications for modelling

Although the urban sound environment is persistent at the city scale, it exhibits pronounced intra-day and short-term variability driven by traffic peaks, human activity cycles, and meteorological fluctuations. In this dataset, each field measurement corresponds to a short 10-s LAeq segment and surveys were preferentially conducted during representative daytime periods (morning peak, afternoon, evening peak) under dry, low-wind conditions in July–August 2016. Consequently, after energy-domain aggregation to the 30 m grid, each cell-level label is best interpreted as a typical short-term daytime exposure rather than a full 24-h or annual indicator such as L_{den} or L_{night} . The temporal sampling design therefore captures peak-period structure and citywide spatial contrasts, while under-representing nocturnal and seasonal regimes.

This temporal mismatch between short-segment labels and the goal of producing stable city-scale maps is handled in the modelling stage by treating unresolved sub-hourly variability as aleatoric noise. Concretely, we inject small stochastic perturbations during training to improve robustness and to regularise the learned mappings against short-term fluctuations that are not systematically represented in the inputs. To ensure methodological consistency across architectures, stochastic perturbations are introduced during training in both the CNN and GNN frameworks. In both cases, zero-mean Gaussian noise is applied to the normalised inputs, reflecting the inherent short-term variability of urban soundscapes and improving robustness. For the CNN experiments, Gaussian noise with a standard deviation of 0.1 is added to the 0–1 normalised multispectral image tensors via a TensorFlow GaussianNoise layer. For the GNN experiments, Gaussian noise with a standard deviation of 0.07 is applied to node feature vectors after normalisation. While the injection points differ—pixel space for CNNs and feature space for GNNs—the principle remains the same: both strategies simulate aleatoric variability, regularise training, and reduce overfitting to incidental temporal fluctuations in the 10-s noise labels.

3.2.3 Remote Sensing Data Processing

The WorldView-2 multispectral imagery described in Section 3.2.1 underwent a sequence of preprocessing and feature derivation steps to ensure consistency with the noise observations and to provide meaningful predictors for subsequent modelling. The raw digital numbers were first radiometrically calibrated and atmospherically corrected using the FLAASH algorithm in ENVI, converting values to surface reflectance units (Anderson et al., 2002). Geometric correction and reprojection aligned the imagery with the British National Grid (EPSG:27700), ensuring accurate integration with ground measurements.

Following these preprocessing steps, the imagery was used to derive a comprehensive set of predictors. Two broad categories of features were produced:

1. Spectral and index-based features: standard vegetation and water indices such as NDVI and NDWI were computed from the multispectral bands. These indices serve

as proxies for vegetation cover, imperviousness, and surface moisture, which are known to influence noise propagation.

2. Textural and morphological features: Grey-Level Co-occurrence Matrix (GLCM) measures (contrast, correlation, entropy, homogeneity) were computed from both single bands (e.g. Red, NIR) and spectral indices. Depending on the study, these measures were summarised within different spatial contexts: direct 30 m cells (baseline), 1000 m square windows (CNN experiments), or ring-shaped buffers (GNN experiments). In addition, the sixth chapter incorporated z-score transformations to emphasise local anomalies in spectral and textural surfaces, as well as morphological operators (e.g. opening, closing) to capture structural patterns.

To provide a concise overview of how feature construction evolved across the empirical chapters, Table 3.2 summarises the feature sources, neighbourhood definitions, and naming conventions used in this dissertation.

Table 3.2 Overview of Feature Construction Across Chapters

Chapter	Feature Source	Neighbourhood / Scale	Example Field Name	Description
3 (Baseline, CatBoost)	GLCM from single spectral bands (e.g. Red, Green, NIR1)	30 m grid cell	Red_Contrast_Medium	Median value of GLCM Contrast derived from Red band within 30 m cell
4 (CNN, Spatial Expansion)	Spectral indices (NDVI, NDWI) and their GLCM texture maps	1000 m × 1000 m square neighbourhood	No explicit field names – features represented as multi-channel image slices	Feature maps (spectral indices + GLCM textures) are stacked as input channels, cropped into 1000 m × 1000 m patches, and directly ingested by the CNN for training
5 (GNN, Relational Modelling)	Same as above	Ring-shaped buffers (e.g. 30–60 m, 60–90 m)	WVBI_correlation_r60_mean	Mean GLCM Correlation from WVBI index within 30–60 m annulus
6 (Domain Adaptation, Generalisation)	Added z-score zonal feature map; morphological operators; Urban Atlas 2012 LULC	Multi-scale ring buffers (e.g. 120–500 m, 500–1000 m)	MP_NIR1_open_r5_r1000_skewness, landuse_11210_r500, global_dist_14100	Skewness of morphological opening on NIR1 in 500–1000 m ring; Proportion of land use class 11210 (Discontinuous Dense Urban Fabric) within 120 - 500 m ring; Shortest distance to land use class 14100 (Green Urban Space) in 120–1000 m range

3.2.4 Land Use and Ancillary Data

Beyond the remote-sensing predictors derived from WorldView-2 imagery, the Urban Atlas 2012 dataset (EEA, 2018) provided an additional and harmonised source of environmental context. Urban Atlas offers thematic land-use/land-cover (LULC) classifications across European urban areas at 10–20 m nominal resolution, aggregated to a common 1:10 000 mapping scale. Its consistency in nomenclature, hierarchical coding, and spatial granularity makes it particularly valuable for comparative and cross-city analyses, mitigating discrepancies that often arise from locally produced datasets differing in classification logic or spatial accuracy.

In this dissertation, the Urban Atlas 2012 layer serves a dual methodological role. First, as an input data source in the domain adaptation and cross-city experiments (see Chapter 6), LULC variables are incorporated as structural descriptors of the urban form. For each 30 m grid cell, the proportional coverage of individual LULC classes was computed within concentric ring buffers, producing variables such as *landuse_11210_r500* (the proportion of *Discontinuous Dense Urban Fabric*, class 11210, within 120–500 m) or *global_dist_14100* (the minimum distance to *Green Urban Area*, class 14100, within 120–1000 m). These features capture both the composition and spatial arrangement of functional land-use zones, which are critical determinants of environmental noise patterns—particularly the attenuation across vegetated or open-space buffers, and the amplification along dense transport corridors.

Second, Urban Atlas data are employed as an independent reference framework for land-use-specific model evaluation. Because all subsequent predictive frameworks (CatBoost, CNN, and GNN) produce spatially continuous noise estimates, the stratified LULC map provides a means to quantify residual behaviour by class. Aggregating residuals within homogeneous land-use categories—such as *Continuous Urban Fabric* (11100), *Industrial or Commercial Units* (12100), *Road and Rail Networks* (12210–12230), and *Green Urban Areas* (14100)—allows systematic diagnosis of over- or under-prediction tendencies. This approach directly supports the comparative analysis across Chapters 4–6, ensuring that identical performance metrics (MAE, R^2 , and residual variance) are applied within each land-use type to enable precise cross-model comparison.

Figure 3.6 illustrates the LULC classification map for the Southampton study area derived from the Urban Atlas 2012 database. Each polygon is colour-coded according to its thematic class, with the legend showing both class names and corresponding numerical codes (e.g., 11100 – Continuous Urban Fabric; 14100 – Green Urban Area; 12230 – Rail Network). The map reveals the spatial juxtaposition of dense built-up zones, transport corridors, and vegetated buffers, which underpin the environmental gradients later analysed in the noise prediction chapters.

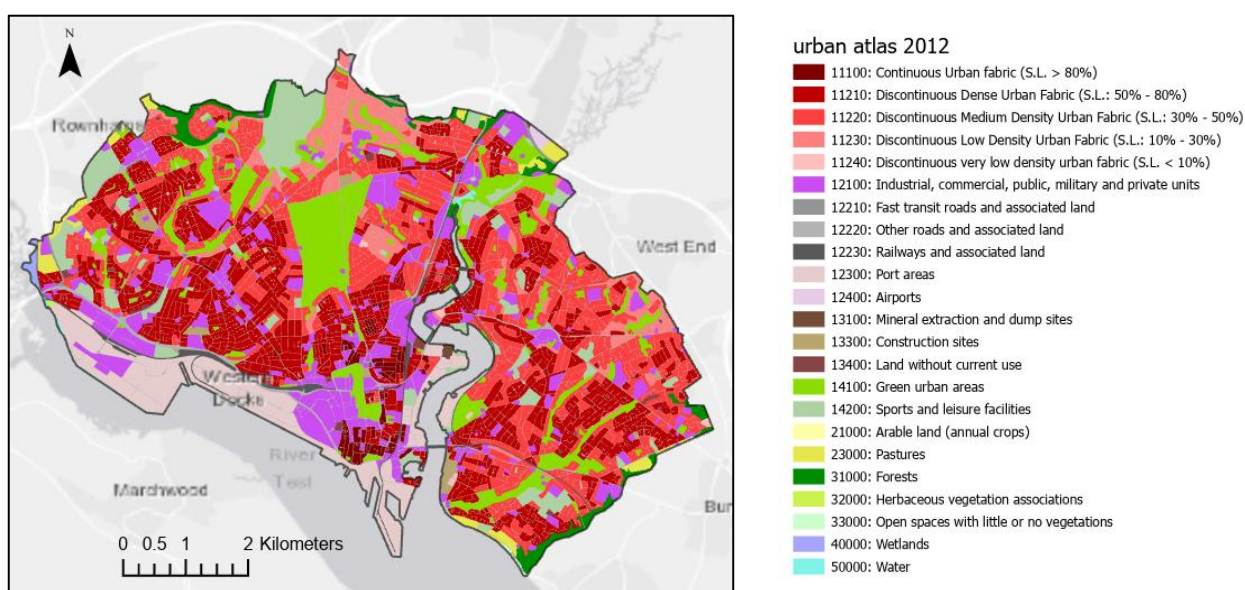


Figure 3.3 Urban Atlas 2012 LULC map of Southampton — colour-coded by class with code-label legend

3.2.5 Data Bias and Representativeness

The exploratory analyses of the raw noise dataset ($n = 52,366$) provide insights into the distributional properties and potential biases of the observational campaign (Figure 3.4). The histogram of LAeq values reveals a right-skewed distribution with a modal range between 55 and 70 dB, and a long tail extending above 85 dB. This pattern reflects the dominance of mid-level urban soundscapes in the survey, while also capturing relatively fewer extreme high-noise environments (e.g. arterial roads or industrial zones). The distribution indicates that while the dataset effectively represents typical urban acoustic exposure, low-noise and very high-noise conditions are less extensively sampled.

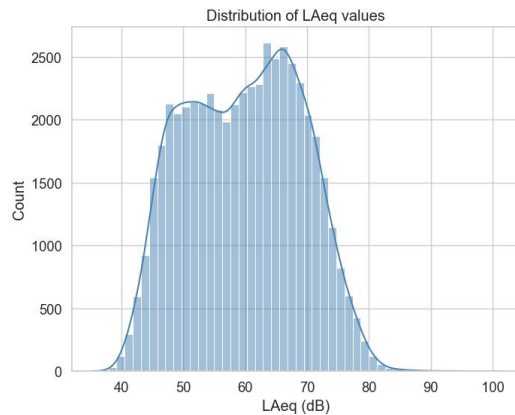
Temporal analysis highlights a diurnal variation in measured noise levels, with elevated LAeq values observed during morning and afternoon peak periods. Midday and early evening periods generally show lower average levels, consistent with traffic intensity patterns in Southampton. Although the time-of-day coverage captures broad urban rhythms, the temporal window of the survey (July–August 2016) may underrepresent seasonal or long-term variability.

The survey was conducted across multiple days between 14 July and 26 August 2016, with daily sample counts ranging from fewer than 1,000 to more than 3,500 observations. The irregular distribution of samples across dates suggests logistical constraints in fieldwork design and highlights the possibility of day-specific effects. Nonetheless, the spread across more than a month provides a robust snapshot of summer acoustic conditions in the city.

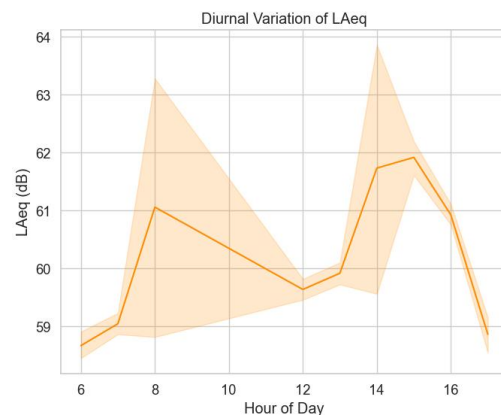
Spatially, the distribution of measurement points closely follows pedestrian transects concentrated along the city's transport corridors and central districts. This produces a dense representation of traffic-dominated environments but relatively sparse coverage of quieter residential backstreets, peri-urban zones, and large green areas. The resulting bias means that the dataset is highly informative for modelling high-exposure contexts but may underrepresent low-exposure environments. This imbalance underscores the importance of incorporating remote-sensing and ancillary predictors (Sections 3.2.3–3.2.4), which extend coverage to areas not directly observed.

To further quantify the degree of spatial dependence inherent in the observational data, a Global Moran's I statistic was computed on the raw measurement points (Moran, 1950). The analysis yielded $I = 0.6345$ ($Z = 300.71$, $p < 0.001$), indicating a strong and highly significant positive spatial autocorrelation. Complementary local Moran's I analysis (LISA; Anselin, 1995) revealed extensive high–high clusters along major road corridors and low–low clusters across peripheral green and residential areas. These patterns confirm the spatially structured nature of the acoustic environment and provide an empirical justification for adopting spatially explicit modelling frameworks in subsequent chapters.

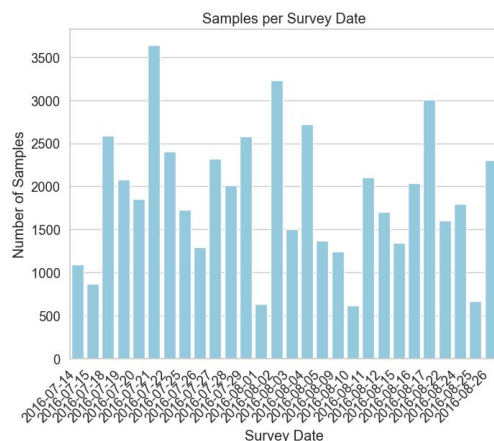
Taken together, these analyses demonstrate that while the Southampton dataset provides an unprecedented and detailed record of urban noise, it is shaped by spatial, temporal, and contextual biases that must be explicitly acknowledged. Subsequent modelling strategies—including feature engineering, convolutional representation learning, and graph-based architectures—are specifically designed to account for these spatial dependencies and enhance the generalisability of predictions across heterogeneous urban contexts.



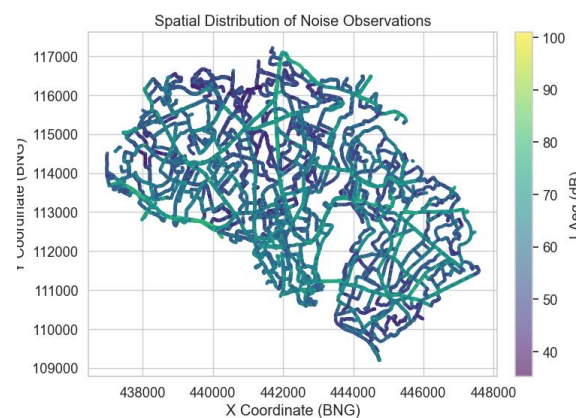
(a) Distribution of LAeq values (n = 52,366)



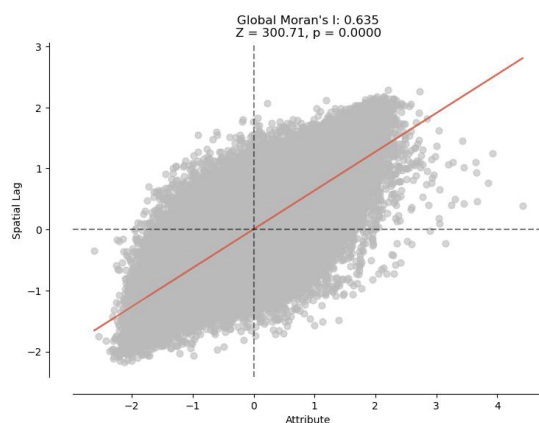
(b) Diurnal variation of LAeq aggregated by hour of day



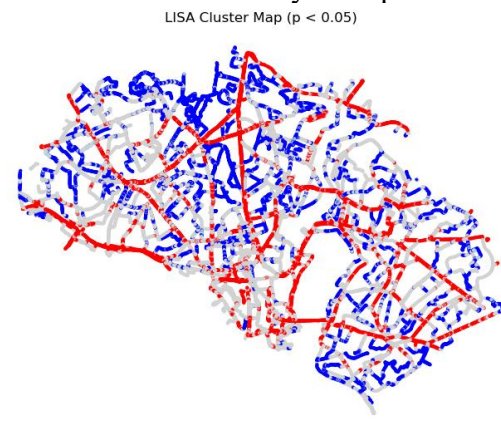
(c) Number of samples per survey date



(d) Spatial distribution of point measurements (BNG coordinates), coloured by LAeq



(e) Global spatial autocorrelation of



(f) Local spatial autocorrelation (LISA)

measured noise levels, showing Moran's I scatter plot ($I = 0.6345$, $Z = 300.71$, $p < 0.001$) cluster map, highlighting high–high (red) and low–low (blue) neighbourhoods at $p < 0.05$

Figure 3.4 Exploratory analyses of the Southampton noise dataset.

(a) Distribution of LAeq values ($n = 52,366$). (b) Diurnal variation of LAeq aggregated by hour of day. (c) Number of samples per survey date (14 July – 26 August 2016). (d) Spatial distribution of point measurements (BNG coordinates), coloured by LAeq. (e) Global spatial autocorrelation of measured noise levels, showing Moran's I scatter plot ($I = 0.6345$, $Z = 300.71$, $p < 0.001$). (f) Local spatial autocorrelation (LISA) cluster map, highlighting high–high (red) and low–low (blue) neighbourhoods at $p < 0.05$.

3.2.6 Summary

This section has introduced the datasets and observational foundations of the dissertation. Southampton provides the only city with ground-truth noise measurements, derived from a high-resolution pedestrian survey conducted during July–August 2016. These data were aggregated to a $30\text{ m} \times 30\text{ m}$ grid to align with remote-sensing predictors, while also retaining raw observations for exploratory analysis and methodological testing. The analyses revealed a dataset dominated by mid-level urban noise environments, with clear diurnal patterns and a spatial bias towards transport corridors and central districts.

The remote-sensing predictors were obtained from pre-processed WorldView-2 imagery, from which both spectral indices and GLCM texture features were derived. In later stages, z-score-based maps, morphological operators, and Urban Atlas 2012 LULC variables were incorporated to enable cross-city transferability. Together, these data sources provided a rich representation of urban structure and environmental context.

Although the dataset is unprecedented in scale and detail, the exploratory analysis highlighted spatial and temporal imbalances that could constrain its representativeness. These limitations emphasise the need for predictive models that can integrate diverse predictors and adapt across urban contexts.

The next section (3.3) builds upon this data foundation by presenting the Machine Learning and Feature Learning Framework used in this dissertation. It outlines the theoretical underpinnings of CatBoost, convolutional neural networks (CNNs), and graph neural networks (GNNs), and demonstrates how each method addresses the challenges of feature complexity, spatial dependency, and cross-city generalisation identified in the preceding analyses.

3.3 Machine Learning and Feature Learning Framework

The methodological framework of this dissertation is designed as a staged progression of machine learning (ML) and deep learning (DL) approaches, each tailored to address increasing levels of feature complexity, spatial dependency, and cross-city generalisation. The framework begins with interpretable tree-based ensemble methods, which establish baseline performance benchmarks. It then advances to convolutional neural networks (CNNs), capable of extracting hierarchical spatial features directly from multispectral imagery, and extends further to graph neural networks (GNNs), which explicitly model relational dependencies embedded within heterogeneous urban morphologies. This staged trajectory directly responds to the methodological challenges identified in Section 3.2, including the heterogeneity of urban environments, the spatial bias of observational data, and the critical need for scalable and transferable predictive frameworks.

Compared with the broad literature survey presented in Chapter 2, which outlined these methods in general terms, the present section focuses on their theoretical principles in direct relation to the data structures and objectives of this study. Schematic illustrations are introduced where appropriate—for instance, to demonstrate how CNN receptive fields map onto urban physical scales, or how message passing in GNNs allows the integration of local and contextual information. By grounding each method within the context of urban noise prediction, the discussion remains accessible to readers from both technical and applied environmental science backgrounds.

In addition to model architectures, the framework also addresses two transversal design dimensions that underpin the empirical chapters. The first concerns the construction of feature maps, which requires interpretable and transferable descriptors capable of capturing the physical and spatial properties of urban surfaces. To this end, the study develops a unified feature computation pipeline integrating spectral indices, Grey-Level Co-occurrence Matrix (GLCM) texture features, multi-scale z-score maps, and morphological operators. This ensures that each modelling stage, from baseline ensemble learning to CNNs and GNNs, is supported by a coherent and semantically enriched feature space. The second dimension concerns the design of data partitioning strategies. Rather than adopting a single uniform split across all tasks, partitioning was deliberately adapted to the requirements of each model, reflecting the principle that evaluation protocols should be aligned with model structure and dataset size to maximise both fairness and performance. This comparative perspective is elaborated in Section 3.3.7.

The section is organised as follows. Section 3.3.1 introduces traditional ML methods and explains the rationale for selecting CatBoost as the primary benchmark. Section 3.3.2 outlines CNNs, emphasising the relationship between kernel size, receptive field, and the representation of urban form. Section 3.3.3 presents GNNs, describing how node features are updated through neighbourhood message passing and how network depth influences the capture of spatial dependencies. Section 3.3.4 integrates these approaches into a unified workflow, illustrating the end-to-end pipeline from data acquisition to predictive modelling. Section 3.3.5 introduces the Feature Map Computation Framework, detailing the derivation of GLCM features, z-score maps,

and morphological operators. Section 3.3.6 extends the framework to advanced strategies for cross-domain generalisation, including domain alignment, pseudo-labelling, and entropy-based gating. Finally, Section 3.3.7 presents a comparative discussion of data partitioning strategies across modelling frameworks, clarifying how different splits were designed to accommodate the structural requirements of CatBoost, CNN, and GNN models as well as cross-city domain adaptation.

3.3.1 Traditional ML (Baseline Methods)

Tree-based ensemble learning methods such as Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosted Trees (GBT) have been widely applied in environmental modelling because of their ability to handle heterogeneous predictors, robustness to noise, and relative interpretability (Breiman, 2001; Cortes & Vapnik, 1995; Friedman, 2001). In this study, such algorithms served as benchmarks against which the advantages of deep learning approaches could be evaluated.

Among these methods, CatBoost (Prokhorenkova et al., 2018) was selected as the baseline model for several reasons. First, CatBoost is specifically designed to handle categorical features without extensive preprocessing, reducing the risk of biased encoding. Second, it incorporates ordered boosting, which mitigates overfitting by ensuring that each split decision is based only on data available prior to that split, thus more closely approximating true generalisation. Third, it integrates efficient handling of high-dimensional feature spaces, making it particularly suitable for the complex predictor sets derived from remote sensing and texture measures in this study.

Conceptually, CatBoost and other boosting algorithms can be understood as ensembles of weak learners (decision trees) that are combined iteratively to minimise prediction error. Each subsequent tree in the sequence corrects the errors of its predecessors, gradually improving performance (Figure 3.5). This process is analogous to refining an estimate through repeated adjustments, where simple models collectively achieve high accuracy.

In practical terms, CatBoost was trained on tabular datasets derived from the 30 m aggregated noise cells, incorporating spectral, textural, and statistical features (see Section 3.2.3). Hyperparameters such as tree depth, learning rate, and the number of iterations were optimised through Bayesian search, and their effects on predictive accuracy are discussed in Section 3.4. The baseline results provide both a point of comparison and an interpretability benchmark, establishing a reference against which the additional complexity of CNN and GNN models can be justified.

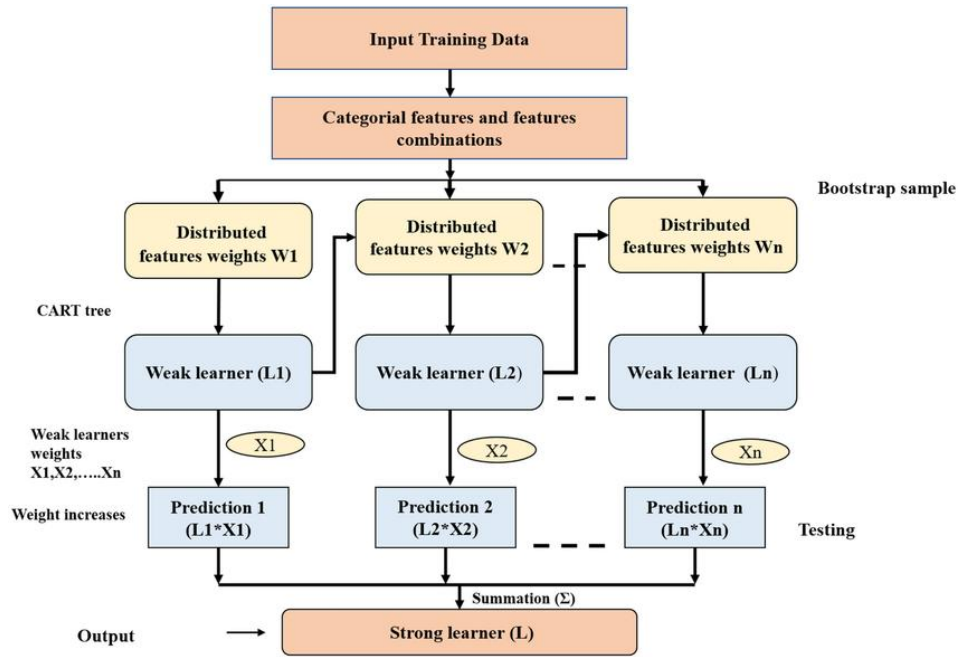


Figure 3.5 Structure of CatBoost adapted from Prokhorenkova et al., (2018). Each tree corrects the errors of its predecessors, and the ensemble prediction emerges from their weighted combination.

3.3.2 Convolutional Neural Networks (CNNs) and Urban Scale Relationships

Convolutional neural networks (CNNs) provide a powerful framework for extracting spatial hierarchies from imagery, making them particularly suited to environmental prediction tasks where spatial structure is a key determinant of outcomes. At their core, CNNs apply small convolutional kernels (filters) across an input image to generate feature maps that highlight local patterns such as edges, textures, or gradients (LeCun et al., 2015; Goodfellow et al., 2016). Each convolutional layer produces increasingly abstract representations, and successive layers expand the receptive field—the portion of the input image that influences a given activation in the network (Araujo, 2019; Luo et al., 2016).

This receptive field expansion has a natural interpretation in the context of urban noise modelling. In this study, input data were structured as 1000 m × 1000 m image patches, corresponding to 250 × 250 pixels at 4 m resolution. Small convolutional kernels, such as 3×3 or 5×5 filters, initially capture local-scale variations equivalent to individual streets or building clusters. As depth increases and receptive fields enlarge, the network begins to incorporate neighbourhood-scale information (hundreds of metres), such as residential blocks, mixed-use zones, or park boundaries. With sufficient layers, receptive fields can encompass the full 1000 m patch, allowing the

model to represent city-scale patterns such as arterial traffic corridors or industrial complexes.

Figure 3.6 illustrates this principle by mapping receptive field size onto urban spatial scales. Small receptive fields (e.g. 3×3 kernels at shallow depth) correspond to fine-grained street-level features, medium receptive fields align with block- or district-level configurations, and large receptive fields approximate city-wide structures. This scaling relationship ensures that CNNs can jointly capture micro-scale textures (such as roadside vegetation or imperviousness) and macro-scale organisation (such as transport networks), both of which strongly influence noise propagation.

Another strength of CNNs is their ability to handle multi-channel input, which in this study consisted of stacked feature maps derived from multispectral indices and texture measures (see Section 3.2.3). Each channel represents a distinct environmental proxy—such as NDVI, NDWI, or GLCM entropy—and the CNN jointly optimises over these heterogeneous predictors. By training on large numbers of patches, the network learns to identify feature combinations that consistently associate with high or low noise exposure, without requiring handcrafted feature engineering.

While CNNs are effective for capturing spatial hierarchies, they also impose constraints. Their square receptive fields assume a regular Euclidean grid, which is suitable for image-like data but less flexible for irregular topologies such as road networks or land-parcel adjacency. These limitations motivate the use of graph neural networks (Section 3.3.3), which explicitly model relational dependencies. Nevertheless, CNNs provide a crucial intermediate step in the methodological progression of this dissertation: they extend beyond tabular features to leverage spatial context directly, while retaining computational efficiency and interpretability through receptive-field analysis.



Figure 3.6 Conceptual illustration of receptive field expansion in CNNs

Small kernels capture street-level details, intermediate receptive fields capture neighbourhood structures, and large receptive fields approximate city-scale patterns within the $1000\text{ m} \times 1000\text{ m}$ input patch (250×250 pixels at 4 m resolution).

3.3.3 Graph Neural Networks (GNNs) and Neighbourhood Information Propagation

While convolutional neural networks (CNNs) are well-suited to regular image grids, they cannot naturally represent the irregular relationships inherent in urban environments. Graph neural networks (GNNs) provide a more flexible alternative by explicitly modelling data as a set of nodes connected by edges (Wu, 2020; Zhou et al., 2020). In this study, each $30\text{ m} \times 30\text{ m}$ grid cell is represented as a node, with edges defined by spatial adjacency or functional relationships (e.g. distance thresholds, ring buffers). Node attributes consist of spectral and textural features derived from remote sensing, and the GNN learns how information flows between neighbouring cells to capture relational patterns that influence noise propagation.

The core mechanism of GNNs is message passing, in which each node updates its representation by aggregating information from its neighbours. At layer $l + 1$, the representation of node i is computed as a function of its own embedding and the aggregated embeddings of its neighbours at layer l (Gilmer et al., 2017). Intuitively, this means that a cell's predicted noise level is influenced not only by its local features but also by the characteristics of surrounding cells.

A key property of this architecture is that the receptive field of a node grows with network depth. After one layer, each node incorporates information from its immediate neighbours ($\approx$ one hop, corresponding to adjacent cells). After two layers, it incorporates neighbours-of-neighbours ($\approx$ two hops, equivalent to a block-scale context). With three or more layers, the receptive field expands further, potentially encompassing broad urban regions. This mechanism allows GNNs to capture spatial dependencies beyond the local patch size used in CNNs.

However, increasing depth introduces the risk of over-smoothing: as more layers are stacked, node embeddings can become indistinguishable, leading to a loss of discriminative power (Li et al., 2018; Alon & Yahav, 2021). In the context of noise prediction, this would manifest as an inability to differentiate between acoustically distinct environments, such as a park and a motorway corridor, if they are embedded within the same graph neighbourhood. Careful design of graph depth, aggregation functions, and edge definitions is therefore critical to maintaining the balance between contextual reach and predictive precision.

To illustrate, figure 3.7 presents a schematic example. In the first layer, a residential cell adjacent to a road integrates traffic-related attributes from its immediate neighbour, refining its own prediction. After three layers, the same cell incorporates information propagated from more distant areas, but with diminishing specificity, eventually blurring distinctions between quiet and noisy environments. This example demonstrates both the strength and limitations of GNNs in environmental applications: they excel at capturing relational dependencies but must be carefully tuned to avoid homogenising diverse spatial patterns.

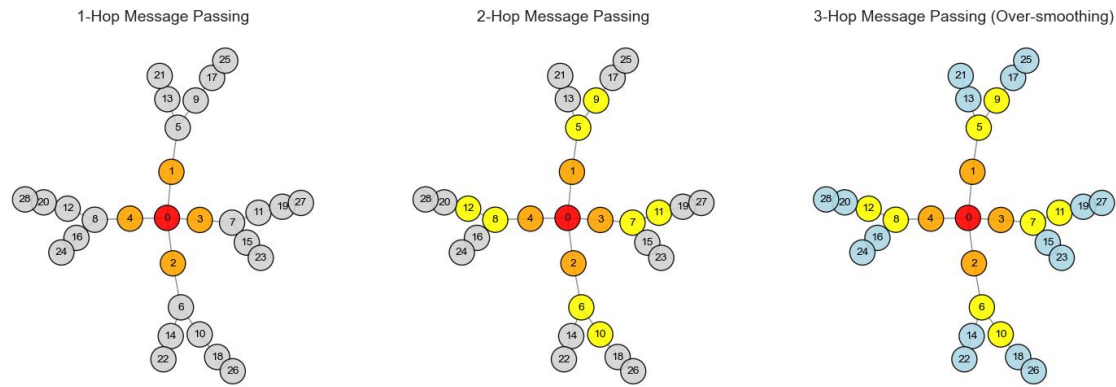


Figure 3.7 Conceptual illustration of message passing in a graph neural network.

Node embeddings are updated by aggregating neighbour features, with the receptive field expanding as network depth increases. Excessive depth leads to over-smoothing, where node representations converge and lose discriminative power.

3.3.4 Integration into Workflow

The machine learning and deep learning approaches in this dissertation were embedded within a coherent workflow that ensured comparability across experiments and reproducibility of results (Figure 3.8). The workflow progressed through four main stages: data acquisition and preprocessing, feature construction, model training and evaluation, and the generation of noise prediction outputs.

Stage 1: Data acquisition and preprocessing.

The workflow began with the collection of ground-truth noise observations in Southampton and the acquisition of WorldView-2 multispectral imagery. In Chapter 6, cross-city predictors from the Urban Atlas 2012 dataset were incorporated to support domain adaptation. All datasets were harmonised to a 30 m analytical grid through radiometric calibration, atmospheric correction, geometric alignment, and energy-domain aggregation of acoustic measurements.

Stage 2: Feature construction.

All models relied on predictors derived from multispectral imagery, which were transformed into spectral indices, GLCM texture measures, or statistical summaries. The differences between chapters lie in how these features were structured and integrated with spatial context. In Chapter 3, features were limited to cell-level summaries (e.g. median GLCM contrast in 30 m cells). In Chapter 4, these predictors were assembled as 1000 m × 1000 m image patches (250 × 250 pixels at 4 m resolution) and used as multi-channel CNN inputs. In Chapter 5, the same spectral and textural predictors served as node attributes within a graph, while additional input in the form of a spatial adjacency matrix defined edges based on Euclidean distance with limited neighbours. In Chapter 6, node attributes were further enriched with LULC statistics from Urban Atlas 2012, and the graph structure extended to capture three-hop neighbourhoods, providing the foundation for the dual-branch GCN + GAT architecture.

Stage 3: Model training and evaluation.

Each chapter introduced a model architecture aligned with its feature and structural representation. CatBoost (Chapter 3) established a baseline using tabular predictors with boosted decision trees. EfficientNetB0 (Chapter 4) applied convolutional layers with compound scaling to exploit the hierarchical spatial structure of image patches. GraphSAGE (Chapter 5) performed inductive learning on graphs, aggregating information from a fixed number of neighbours defined in the adjacency matrix. The dual-branch GCN + GAT model (Chapter 6) combined spectral, textural, and LULC node features with relational information from three-hop adjacency graphs, capturing both broad contextual structure and localised attention weights.

Stage 4: Prediction outputs.

Across chapters, the models produced city-scale noise prediction maps. The baseline CatBoost demonstrated feasibility with simple tabular features; EfficientNetB0 added spatial coherence by leveraging receptive fields; GraphSAGE incorporated relational dependencies via neighbourhood aggregation; and the dual-branch GCN + GAT extended these capacities to heterogeneous domains by integrating LULC data and multi-hop graph structures. Together, these outputs illustrate a progressive evolution towards scalable and transferable urban noise mapping.

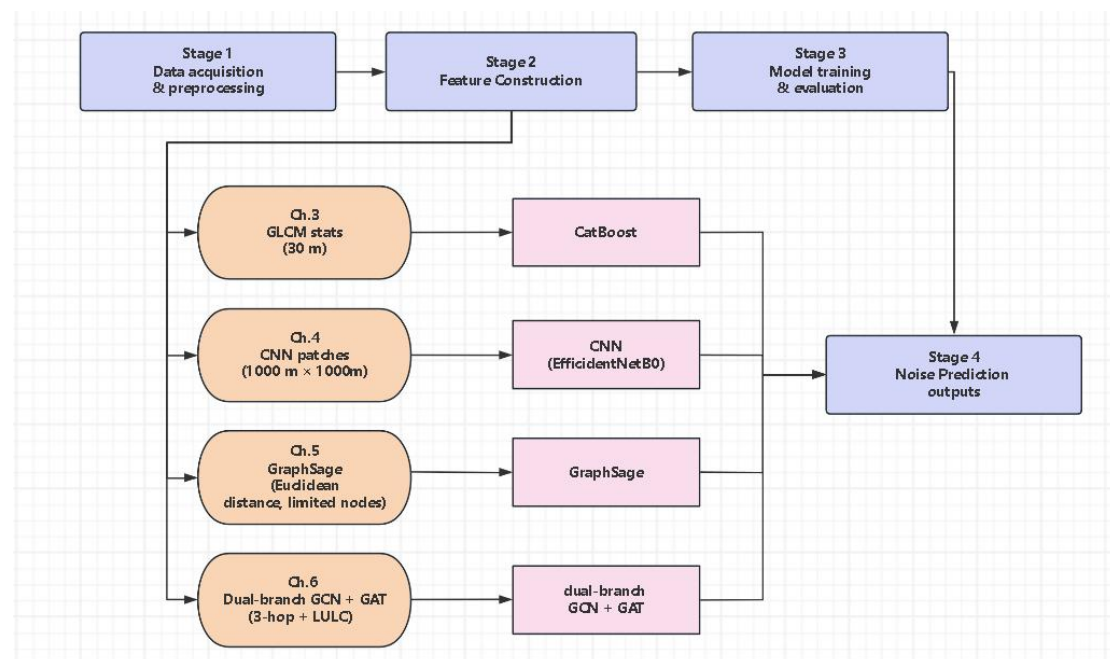


Figure 3.8 Workflow for urban noise prediction.

The process integrates data acquisition, feature construction from multispectral imagery (plus graph structures and LULC where applicable), model training (CatBoost, EfficientNetB0, GraphSAGE, dual-branch GCN+GAT), and prediction outputs.

Together, these models represent a methodological trajectory from interpretable, cell-based prediction to transferable, structure-aware frameworks capable of addressing the heterogeneity of urban soundscapes. The next section (3.4) presents the preliminary experiments that informed these methodological choices, providing empirical justification for the progression adopted in subsequent chapters.

3.3.5 Feature Map Computation Framework

A critical methodological component of this dissertation lies in the systematic design and computation of feature maps derived from multispectral WorldView-2 imagery. These feature maps serve as intermediate representations that bridge raw spectral information with machine learning models, providing semantically meaningful and structurally informative inputs. While Chapters 4–6 apply these features in different modelling contexts (CatBoost, CNNs, and GNNs, respectively), the underlying computational framework is unified and is outlined here in detail. The framework integrates four complementary approaches: (i) spectral indices, (ii) Grey-Level Co-occurrence Matrix (GLCM) texture features, (iii) z-score normalisation maps, and (iv) morphological operators. Together, these techniques provide a multi-perspective representation of the urban surface, balancing biophysical interpretability with statistical richness.

3.3.5.1 Grey-Level Co-occurrence Matrix (GLCM) Texture Features

The Grey-Level Co-occurrence Matrix (GLCM), first introduced by Haralick et al. (1973), provides a statistical framework for quantifying spatial texture by describing how often pairs of grey levels occur in a given spatial relationship. Formally, for a quantised image $I(x, y)$ with grey levels $g \in \{0, 1, \dots, G - 1\}$, the co-occurrence matrix is defined as:

$$P(i, j; d, \theta) = \sum_{x=1}^M \sum_{y=1}^N \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$

where d is the spatial offset, θ is the orientation (0° , 45° , 90° , 135°), and $(\Delta x, \Delta y)$ represents the displacement corresponding to d, θ . Each entry $P(i, j)$ counts the number of pixel pairs with grey levels i and j occurring in the specified spatial relation. Normalisation yields the joint probability distribution:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)}$$

From this probability matrix, a suite of statistical descriptors can be computed to characterise texture. In this study, 28 descriptors were generated using the Orfeo Toolbox (OTB), which provides “simple,” “higher-order,” and “advanced” measures (see Appendix I for full formulas and definitions). These include widely used metrics such as Contrast, Homogeneity, Entropy, and Correlation, as well as advanced descriptors such as Grey-Level Nonuniformity (GLN) and Information Measures of Correlation (IC1, IC2).

For example, GLN (Grey-Level Nonuniformity) measures the variability of grey-level distributions, with lower values indicating uniform grey levels and higher values reflecting greater heterogeneity:

$$GLN = \frac{1}{N_g} \sum_{i=0}^{G-1} \left(\sum_{j=0}^{G-1} p(i, j) \right)^2$$

where N_g is the number of grey levels.

IC1 (Information Measure of Correlation 1) evaluates the dependency between row and column marginal distributions of the GLCM, providing a measure of structural redundancy:

$$IC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

where HX and HY are the entropies of marginal distributions, HXY is the joint entropy, and $HXY1$ is the entropy under independence assumptions.

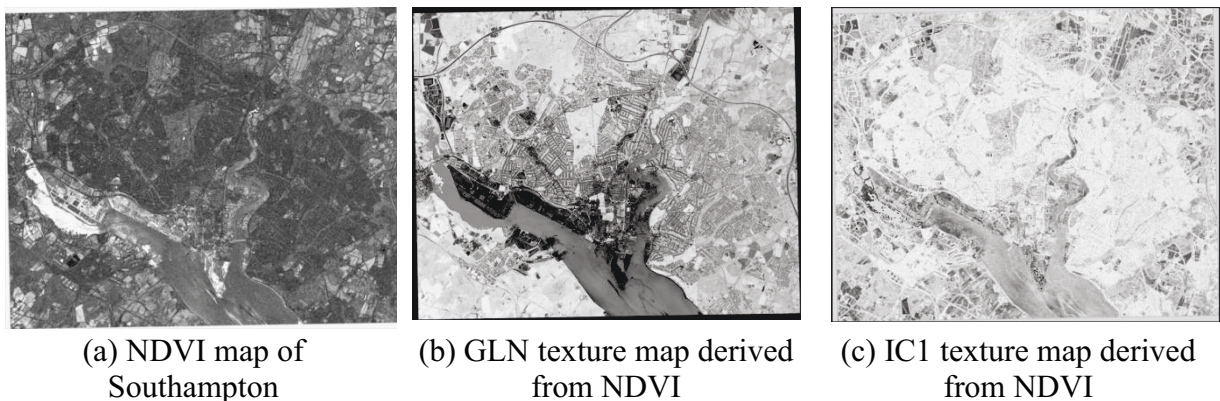


Figure 3.9 Comparative visualisation of NDVI and derived GLCM features (GLN and IC1). (a) NDVI map of Southampton (b) GLN texture map derived from NDVI (c) IC1 texture map derived from NDVI

While NDVI highlights vegetation distribution, GLN emphasises heterogeneity (dark values in port–water zones, light values in vegetated regions), and IC1 reflects structural redundancy, distinguishing built-up areas from natural surfaces.

3.3.5.2 Z-score Feature Maps

In addition to GLCM descriptors, this dissertation incorporates z-score feature maps as a complementary technique for emphasising relative spectral variation across multiple spatial scales. Unlike absolute reflectance values, which can be influenced by acquisition conditions (e.g., sensor calibration, atmospheric effects) and inter-city differences in illumination or albedo (Schowengerdt, 2006), z-score maps normalise each pixel value with respect to its local statistical context. This transformation highlights relative anomalies within a neighbourhood, thereby improving comparability across heterogeneous urban environments.

Formally, for a given spectral band or index $I(x, y)$, the z-score at pixel (x, y) is computed as:

$$Z(x, y; w) = \frac{I(x, y) - \mu_w(x, y)}{\sigma_w(x, y)}$$

where $\mu_w(x, y)$ and $\sigma_w(x, y)$ denote the mean and standard deviation of pixel intensities within a local square window of size $w \times w$ centred at (x, y) . By construction, the z-score expresses each pixel's deviation from its local mean in units of standard deviation.

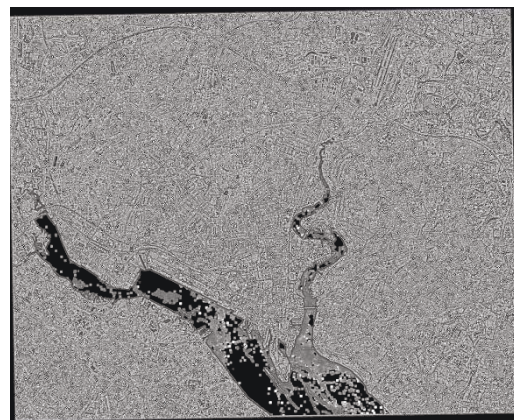
The z-score framework was first introduced in Chapter 6 to enhance cross-city generalisation. Three window sizes were employed— 15×15 , 31×31 , and 77×77 pixels—corresponding to spatial extents of ~ 60 m, ~ 124 m, and ~ 308 m, respectively, given the 4 m resolution of WorldView-2 imagery. These scales were selected to capture texture patterns at progressively larger neighbourhood contexts:

1. Small windows (15×15) accentuate fine-grained heterogeneity, such as individual building blocks, tree crowns, or narrow roads.
2. Medium windows (31×31) capture intermediate structures, e.g., clusters of residential housing or contiguous vegetation patches.
3. Large windows (77×77) emphasise broader landform and land-use configurations, such as industrial estates or continuous parkland, by smoothing local variations and highlighting area-wide anomalies.

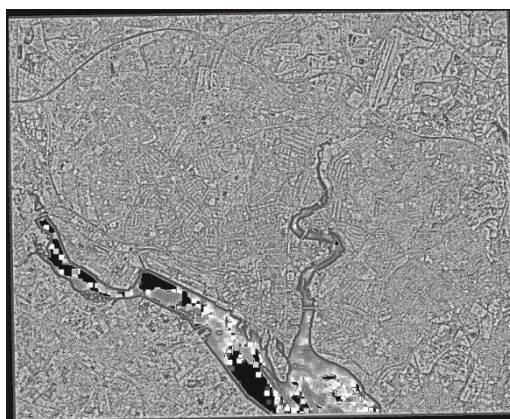
By combining z-score maps at these three scales, the framework enriches the feature space with multi-scale representations of relative contrast, improving the capacity of learning models to capture both local detail and broader structural patterns.



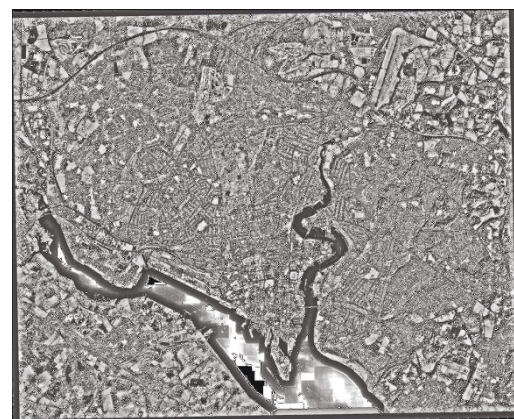
(a) Original Red-Edge band reflectance



(b) Z-score map at 15×15 window



(c) Z-score map at 31×31 window



(d) Z-score map at 77×77 window

Figure 3.10 Comparative illustration of Red-Edge reflectance and z-score maps at three spatial scales.

Smaller windows enhance local anomalies, whereas larger windows emphasise broader morphological structures. Together, these representations provide a multi-scale characterisation of urban texture relevant to noise propagation modelling.

3.3.5.3 Morphological Operators

Mathematical morphology provides a further complementary framework for feature extraction by directly analysing the geometric structure of image objects. Unlike GLCM, which captures second-order grey-level co-occurrence statistics, or z-score maps, which emphasise relative local contrasts, morphological operators describe shape, connectivity, and spatial arrangement. They are particularly well suited to characterising built-up areas and transportation networks, where geometry and connectivity play a decisive role in acoustic propagation (Haralick, 1987; Benediktsson et al., 2005).

The foundation of morphological operations lies in the interaction between the input image $I(x, y)$ and a structuring element B . For a binary or greyscale image, the erosion and dilation of I by B are defined as:

$$(I \ominus B)(x, y) = \min_{(u,v) \in B}$$

$$(I \oplus B)(x, y) = \max_{(u,v) \in B}$$

From these two primitives, higher-order operators are constructed. The opening operation, defined as erosion followed by dilation ($I \circ B = (I \ominus B) \oplus B$), removes small objects and noise while preserving larger structures. Conversely, the closing operation, defined as dilation followed by erosion ($I \cdot B = (I \oplus B) \ominus B$), fills small gaps and smooths boundaries, thereby enhancing continuity in linear features.

In this study, morphological operations were applied to multispectral bands to extract structural patterns in the urban environment. Using the Red band as an example, opening emphasised dominant built-up blocks by suppressing fine-grained vegetation pixels, while closing reinforced road and waterfront continuity by bridging small gaps between adjacent bright features. Together, these operations provided complementary views of built morphology: opening highlighted discrete building forms, whereas closing accentuated extended infrastructural elements.

In addition to opening and closing, the Local Binary Pattern (LBP) operator was introduced as a texture descriptor grounded in morphological principles. For each pixel, LBP encodes the relative differences between the central pixel value and its neighbours, forming a binary string that is then converted into a decimal label.

Formally, for a pixel intensity g_c with P neighbours g_p on a circle of radius R , the LBP code is given as:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

This operator captures micro-structural patterns such as edges, corners, and flat areas. Within the context of noise prediction, LBP enriches the feature representation by detecting repetitive textural elements of buildings and surfaces, which influence sound reflection and scattering.

The combined use of morphological opening, closing, and LBP maps was particularly valuable for cross-city generalisation in Chapter 6. Whereas spectral indices and GLCM features focus on reflectance and second-order statistics, morphological operators directly quantify spatial form, producing features that are less sensitive to radiometric variability between cities. This robustness makes them effective complements in a transferable modelling framework.

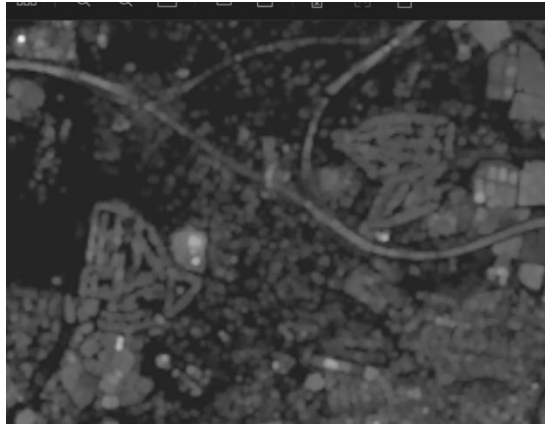
Figure 3.11 illustrates the effect of these operators using the Red band of Southampton. The original reflectance map is shown alongside its morphological opening, closing, and LBP derivatives. The opening map suppresses fine vegetation and highlights dominant building footprints, the closing map enhances continuity of linear urban features, and the LBP map captures repetitive local patterns, together providing a structurally enriched representation of the city's urban form.



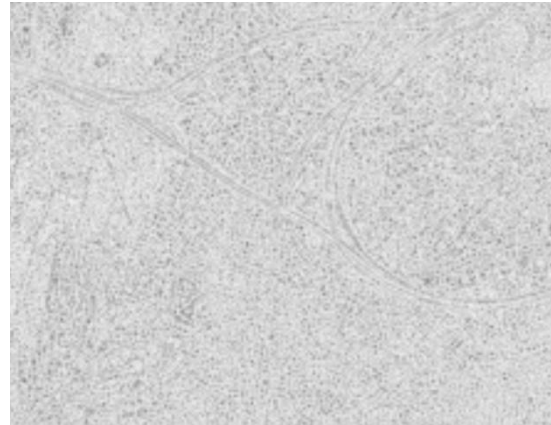
(a) Original Red band reflectance
(Southampton urban subset)



(b) Morphological opening result



(c) Morphological closing result



(d) LBP feature map

Figure 3.11 Comparative illustration of morphological operators applied to a local subset of Southampton’s Red band imagery

The chosen urban fragment includes mixed residential blocks, adjacent vegetation patches, and linear road features, allowing the effects of each operator to be clearly observed. In the opening map (b), fine vegetation signals are suppressed, enhancing the visibility of dominant building footprints. The closing map (c) bridges small gaps, reinforcing the continuity of roads and waterfront edges. The LBP map (d) highlights repetitive local textures, such as tiled roof patterns and road markings. By focusing on a representative subset rather than the entire city, the structural effects of morphological operators are more discernible, demonstrating their utility for extracting shape- and texture-based descriptors relevant to noise propagation.

3.3.6 Advanced Strategies for Cross-Domain Generalisation

A critical methodological challenge of this dissertation is the generalisation of noise prediction models across cities that lack in-situ acoustic measurements. While Chapters 4 and 5 demonstrated the feasibility of machine learning and graph neural networks in single-city contexts, their direct transfer to new urban domains is hindered by distributional differences in morphology, land use, and spectral characteristics. To address this challenge, the framework integrates three complementary strategies: high-confidence pseudo-labeling, adversarial domain alignment with gradient reversal, and entropy-based gating losses. These methods are not considered in isolation but are ultimately combined in Chapter 6 to achieve robust cross-domain performance.

High-Confidence Pseudo-Labeling

Since only Southampton provides measured noise data, pseudo-labels are required to extend supervision to the four target cities. In practice, this is achieved through a similarity-based k-nearest neighbors approach applied to carefully selected variables. Specifically, land-use/land-cover (LULC)-derived predictors—such as the proportion of transport infrastructure or distance to green areas—are prioritised due to their semantic stability across cities. For each unlabeled grid cell in a target city, its feature profile is compared to the Southampton-labeled set, and the average of its five nearest

neighbors is assigned as a provisional label. To safeguard reliability, two filtering criteria are imposed: a similarity-based confidence score (top 10% threshold) and neighbor-label stability (standard deviation <10 dBA). Only samples passing both conditions are retained, resulting in approximately 9–10% of grid points in each target city being pseudo-labeled. This conservative design follows established principles of high-confidence pseudo-labeling (Lee, 2013) and self-training in remote sensing domain adaptation (Kwak and Park, 2022), ensuring that only structurally robust and acoustically meaningful pseudo-labels participate in training, while unlabeled nodes remain active in the graph to support spatial propagation.

Adversarial Domain Alignment with Gradient Reversal

Even with pseudo-labeling, domain shifts persist: predictors may carry different statistical distributions in different cities. To counteract this, the framework implements multi-domain adversarial alignment. A lightweight domain discriminator is trained to classify nodes by city of origin, while a gradient reversal layer (GRL) connects this discriminator to the feature extractor. The GRL acts as identity in the forward pass but multiplies gradients by -1 during backpropagation, forcing the feature extractor to generate embeddings that are simultaneously predictive of noise and invariant to domain identity (Ganin et al., 2016). To stabilise training, the adversarial component is activated only after several warm-up epochs, when the base model has already learned to predict noise reliably. Complementing adversarial learning, a Maximum Mean Discrepancy (MMD) loss is added to explicitly reduce statistical divergence between Southampton and each target city. Unlike adversarial signals, which encourage domain confusion, MMD directly matches feature distributions in a reproducing kernel Hilbert space (Long et al., 2015), ensuring alignment of both low- and high-order moments. This hybrid alignment provides a balanced mechanism for enforcing domain invariance without sacrificing discriminative power.

Entropy Minimisation and MME Loss

A further challenge in semi-supervised settings is that models may remain uncertain on unlabeled target samples. To address this, the framework employs entropy minimisation (Grandvalet and Bengio, 2005), penalising prediction uncertainty on target-domain outputs and encouraging sharper decision boundaries. Additionally, inspired by Minimax Entropy (MME) (Saito et al., 2019), the framework regularises the classifier to maximise domain confusion while minimising prediction entropy, which encourages the network to focus on transferable and discriminative features rather than domain-specific noise. This combination of entropy regularisation and adversarial alignment strengthens cross-city generalisation and stabilises training under low-label conditions.

Integration into the Framework

Together, these strategies form an interlocking methodology. Pseudo-labeling provides sparse but reliable supervision in target cities; adversarial alignment and MMD ensure that feature spaces are shared across domains; and entropy minimisation strengthens decision boundaries on unlabelled samples. As implemented in Chapter 6, this integrated design enables the dual-branch GNN to extend predictive capacity

from Southampton to Cardiff, Liverpool, Nottingham, and Portsmouth, producing coherent noise maps without requiring direct local measurements.

3.3.7 Data Partitioning Strategies across Models

A final methodological dimension that warrants explicit clarification is the strategy used for partitioning the dataset into training, validation, and testing subsets across different modelling frameworks. Rather than imposing a single, uniform protocol, this dissertation deliberately adopted architecture-sensitive partitioning strategies tailored to the structural and statistical characteristics of each model. This adaptive design is grounded in the principle that data partitioning is not merely procedural: it directly affects model performance, convergence behaviour, and the validity of evaluation. Different learning architectures impose distinct requirements regarding sample independence, graph structure, or pseudo-label reliability, and the chosen protocols reflect these considerations.

For the CatBoost baseline experiments introduced in Chapter 3, a conventional 80%–20% five-fold cross-validation was employed on the aggregated 13,474 grid-based samples. This choice ensured methodological alignment with established practices in environmental noise modelling (e.g., Alvares-Sanches et al., 2021a) and provided a robust, interpretable benchmark against which deep learning architectures could be evaluated. Five-fold rotation further mitigated overfitting risks by averaging model performance across multiple partitions, yielding stable generalisation estimates under tabular learning conditions.

The convolutional neural network (CNN) experiments presented in Chapter 4 required a different protocol, reflecting the transformation of the data into a set of $1,000 \text{ m} \times 1,000 \text{ m}$ patches (250×250 pixels at 4 m resolution). Because patch-based representations significantly reduced the total number of training instances, the partitioning scheme allocated 70% of samples to training, 15% to validation, and 15% to testing. This emphasis on training data aligns with standard deep learning practice in small-sample regimes, where sufficient examples are needed to support effective feature learning. The held-out test set provided an unbiased measure of generalisation, while the validation set guided early stopping and hyperparameter optimisation. This design directly balanced the competing objectives of maximising feature extraction capacity and maintaining rigorous evaluation control.

A more substantial departure from conventional data splitting was required for the graph neural network (GNN) experiments in Chapter 5. Unlike CNNs or tabular models, GNNs operate on a single, connected graph structure, where individual nodes are topologically interdependent through message passing. Naive random partitioning at the sample level would lead to information leakage, as nodes in the test set could still indirectly influence the model through shared neighbourhoods. To address this, the dataset was partitioned at the node level into 60% training, 20% validation, and 20% testing subsets, combined with strict masking to ensure that only the labels of training nodes were visible during learning. Validation and test nodes remained structurally connected to the graph but label-inaccessible. This configuration preserves the inductive nature of GNN learning, allowing models to leverage

neighbourhood context while ensuring statistical independence of evaluation. The smaller training proportion compared with CNNs also reflects the inherent redundancy in graph structures: as each node aggregates multi-hop information, effective learning can be achieved with fewer explicit labels.

The partitioning design reached its most distinct form in the cross-city domain adaptation experiments of Chapter 6, where pseudo-labels rather than measured noise levels were used for supervision in target cities. In this setting, an 80%–20% split between training and validation was adopted for the pseudo-labeled nodes, while unlabeled nodes remained part of the graph to support domain alignment through message passing. Unlike the source city, no independent test set was created in the target domains, since pseudo-labels are derived from similarity inference rather than empirical measurements, and therefore cannot provide a valid benchmark of predictive accuracy. Instead, the validation split served to monitor overfitting to pseudo-label distributions, while model evaluation remained anchored in the true labels of the source domain. This design reflects widely adopted practices in weakly supervised and domain-adversarial learning, where pseudo-labeled data shape representation learning but are not used as formal test benchmarks.

Taken together, these partitioning strategies illustrate an intentional methodological alignment between data structure and model design. CatBoost relied on conventional cross-validation to ensure interpretability; CNNs required larger training allocations to compensate for small sample sizes; GNNs depended on node masking to respect graph connectivity; and domain adaptation used pseudo-labeled training-validation splits to support representation alignment in the absence of real ground truth. Across all cases, the guiding principle remained consistent: to design partitioning schemes that maximise both learning effectiveness and evaluation validity under the methodological and data constraints specific to each modelling framework.

Table 3.3 Overview of Data Partitioning Strategies across Modelling Frameworks

Chapter	Model Type	Domain Type	Data Structure	Partition Ratio	Label Type	Partition Rationale
3	CatBoost	Single-city (Southampton)	Independent grid samples (tabular)	80% train – 20% test (5-fold CV)	Real labels	Ensures statistical robustness and comparability with conventional ML; cross-validation mitigates overfitting
4	CNN	Single-city (Southampton)	Fixed patch-based image tiles (4 m)	70% train – 15% val – 15% test	Real labels	Larger training share to support deep feature learning under small sample sizes; explicit test set for generalisation
5	GraphSAGE GNN	Single-city (Southampton)	Node-based graph structure	60% train – 20% val – 20% test (masked)	Real labels	Masking avoids label leakage while preserving graph connectivity; reduced training set reflects multi-hop aggregation
6	Dual-Branch GNN (Domain Adaptation)	Multi-city (1 labeled + 4 unlabeled)	Node-based graph with pseudo-labeling	Source: 60% train – 20% val – 20% test; Target: 80%	Real labels (source) + Pseudo-labels (target)	Anchors evaluation in source domain while enabling cross-domain alignment; avoids invalid

Chapter	Model Type	Domain Type	Data Structure	Partition Ratio	Label Type	Partition Rationale
				train – 20% val		pseudo-test set

3.4 Preliminary Experiments: Noise Prediction Using the CatBoost Model

3.4.1 Overview and Rationale

Preliminary experiments were conducted to establish a methodological baseline and to guide the selection of modelling approaches for subsequent chapters. The objective was to evaluate whether conventional machine learning methods could adequately capture the statistical relationships between remotely sensed features and aggregated noise levels, and to identify their limitations in representing spatial heterogeneity.

Among a range of candidate algorithms, CatBoost was selected as the primary baseline model. This choice was motivated by three factors. First, CatBoost offers strong performance on structured tabular datasets, which aligns with the feature sets derived from multispectral imagery and textural measures. Second, the algorithm incorporates mechanisms such as ordered boosting and effective handling of categorical variables, reducing overfitting and minimising preprocessing requirements (Prokhorenkova et al., 2018). Third, CatBoost has demonstrated robustness in high-dimensional feature spaces, making it particularly suitable for the diverse predictors used in this study.

By establishing CatBoost as a baseline, these experiments provided a benchmark against which the benefits of spatially explicit deep learning methods (CNNs and GNNs) could later be assessed. The results from this stage therefore serve a dual purpose: they demonstrate the feasibility of data-driven urban noise prediction using remote-sensing features, while also highlighting the methodological gaps that motivate the adoption of more advanced architectures in Chapters 4–6.

3.4.2 GLCM Feature Calculation Based on Individual Spectral Bands

The next step is the feature construction phase. Instead of relying solely on composite indices or multispectral band ratios, this study directly computes texture features from individual spectral bands of the WorldView-2 imagery (ESA, 2010), focusing on five critical bands: Red, Green, Blue, RedEdge, and NIR1. The process begins by quantizing each band into 64 grey levels, a step that balances computational efficiency with the retention of essential textural details (Hall-Beyer, 2017).

Using a 5×5 pixel moving window (corresponding to $5 \text{ m} \times 5 \text{ m}$ on the ground), the Grey-Level Co-occurrence Matrix (GLCM) is computed within each window. GLCMs are calculated across four standard orientations (0° , 45° , 90° , 135°) and averaged to achieve rotational invariance (Haralick et al., 1973). From each averaged GLCM, five key texture metrics are extracted:

Table 3.4 *GLCM Features Used in the Baseline Model and Their Definitions*

Feature Name	Description
Contrast	Quantifies local grey-level variations, indicating how sharply pixel intensities change.
Correlation	Measures the linear dependency between neighboring pixel values, reflecting local alignment.
Entropy	Captures the randomness or complexity of textures, with higher values indicating greater disorder.
Homogeneity	Assesses the closeness of GLCM matrix elements to its diagonal, indicating local uniformity.
Energy (Angular Second Moment)	Reflects textural uniformity, where higher values suggest more consistent patterns.

These calculations generate five sets of texture feature maps (one per spectral band), with each set comprising five distinct GLCM-derived metrics. Together, they form a comprehensive textural representation of the urban landscape at 1-meter resolution, providing the raw material for subsequent statistical summarization.

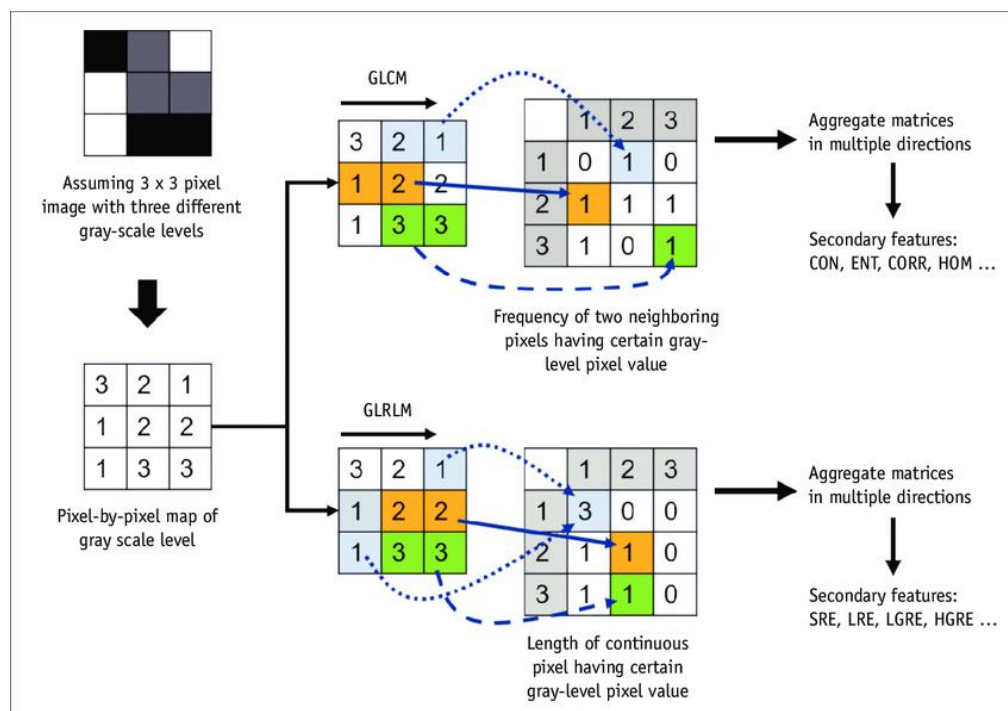


Figure 3.12 Schematic description of textural feature extraction assuming 3 x 3-pixel image with three different gray-scale levels. (Park et al., 2020)

GLCM describes frequency of two neighboring pixels having certain gray-level pixel values, while GLRLM describes length of continuous pixel having certain gray-level pixel value. After aggregating different directional matrices, secondary features are calculated from matrices to describe textural pattern of given image.

3.4.3 Extraction of Statistical Variables from GLCM Feature Maps

To align the high-resolution texture features with the coarser-scale noise measurements, a spatial aggregation step is applied. Specifically, each 1-meter resolution GLCM feature map is aggregated within a $30\text{ m} \times 30\text{ m}$ grid by computing eight statistical descriptors: maximum, minimum, mean, median, 25th percentile, 75th percentile, standard deviation, and range. These descriptors effectively summarize the distribution of local texture values within each grid cell, transforming fine-scale variability into higher-order statistical signals that are more compatible with the spatial scale of ground noise measurements.

Table 3.5 Statistical Descriptors Computed from GLCM Feature Maps within Each $30\text{ m} \times 30\text{ m}$ Grid Cell

Statistical Descriptor	Description
Maximum	The highest GLCM feature value within the grid cell, highlighting the strongest local texture signal.
Minimum	The lowest GLCM feature value within the grid cell, indicating the weakest local texture signal.
Mean	The arithmetic average of all GLCM feature values, providing a central tendency measure.
Median	The middle value that divides the dataset into two equal halves, offering a robust central estimate.
25th Percentile ($P_{(25)}$)	The value below which 25% of the GLCM feature values fall, capturing lower-end distribution behavior.
75th Percentile ($P_{(75)}$)	The value below which 75% of the GLCM feature values fall, capturing upper-end distribution behavior.
Standard Deviation	The degree of variability or dispersion among the GLCM feature values, reflecting texture heterogeneity.
Range	The difference between maximum and minimum values, summarizing the overall spread of texture values.

The concatenation of these statistical descriptors across all five bands and five metrics yields a rich, multidimensional feature vector for each 30 m grid cell, capturing local textural patterns in a format ready for machine learning analysis.

3.4.4 Construction of the Training Dataset

To prepare the baseline dataset, ground-truth noise measurements and remotely sensed predictors were aligned on the common $30\text{ m} \times 30\text{ m}$ analytical grid. The 52,364 raw in-situ observations were aggregated in the energy domain to produce 13,474 grid-level equivalent continuous sound levels (LAeq), thereby ensuring unbiased averaging of logarithmic units.

In parallel, the GLCM-derived texture maps introduced in Section 3.4.2 were summarised within the same grid framework, using the statistical descriptors described in Section 3.4.3. The resulting feature vectors encapsulated the local textural properties of each cell and were directly paired with the corresponding aggregated LAeq values. This integration produced a tabular dataset where predictors and responses were spatially co-registered and suitable for machine-learning analysis.

For model development, the dataset was partitioned through stratified random sampling, with 80% of the grid cells allocated to training and 20% reserved for validation. To ensure robustness, a five-fold cross-validation scheme was applied, allowing each grid cell to contribute to both training and testing across different folds. This procedure preserved the distribution of noise levels in all subsets and provided a statistically rigorous foundation for evaluating the CatBoost baseline.

3.4.5 CatBoost Model Training, Prediction, and Baseline Evaluation

Building on the carefully constructed training dataset, the next phase involved applying the CatBoost model to learn the relationship between remotely sensed textural features and spatially aggregated urban noise levels. The training process was designed to maximise predictive accuracy while minimising overfitting risks, ensuring that the model could generalise effectively to unseen data.

Hyperparameter optimisation was carried out using a Bayesian optimisation framework (Bergstra et al., 2011), combined with five-fold cross-validation, to systematically explore the parameter space and identify configurations most suited to the characteristics of the dataset. The search considered learning rates ranging from 0.01 to 0.3, tree depths between 4 and 12, and boosting iterations from 200 up to 2,000. Regularisation strength was tuned through the L2 parameter, tested within the range of 1 to 10, while bagging temperature—a parameter controlling the stochasticity of sample weighting during training—was varied between 0.5 and 5. This process allowed the model to balance bias and variance effectively, while maintaining stability in the presence of high-dimensional, correlated predictors.

The optimisation phase converged on a configuration with a learning rate of 0.05, a maximum tree depth of 8, and approximately 400 boosting iterations, coupled with an L2 regularisation strength of 3 and a bagging temperature of 1. These settings were selected because they offered the best trade-off between generalisation and computational efficiency, with higher learning rates or tree depths leading to overfitting, and shallower or less regularised configurations failing to capture

sufficient complexity in the feature set. This outcome confirmed the suitability of CatBoost for modelling the GLCM-derived predictors in this study.

Once training was complete, the optimised model was deployed across the entire study area of Southampton, using WorldView-2 imagery captured on 4 May 2016 to generate detailed predictions at a 30 m spatial resolution. The resulting outputs were visualised as a continuous urban noise exposure map, allowing for both quantitative evaluation and qualitative interpretation.

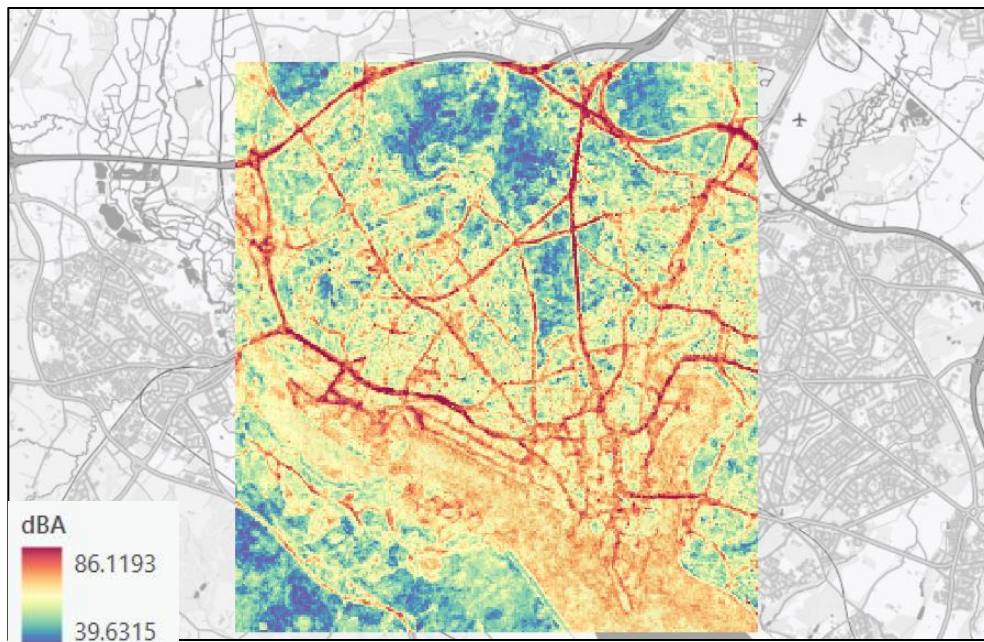


Figure 3.13 CatBoost Noise Prediction Overlay on OS Open Greyscale Map (30 m × 30 m grid)

Visual inspection of the noise map revealed pronounced linear patterns of elevated exposure aligned with major road networks, confirming that the CatBoost model effectively captured traffic-related dynamics. Moreover, areas with dense vegetation were systematically predicted as lower-noise zones, consistent with the noise-mitigating effects of urban greenery (Khan et al., 2018). However, spatial discrepancies were observed: the model tended to overestimate noise in non-traffic zones, such as riverine corridors, and underestimate noise in the Southampton port area. These limitations reflected the baseline model's reduced capacity to represent non-road acoustic sources using spectral-band texture features alone.

Quantitative evaluation yielded an average Mean Absolute Error (MAE) of 5.76 dBA and a coefficient of determination (R^2) of 0.38. While these results demonstrate moderate predictive capability, the relatively low R^2 highlights that a substantial proportion of noise variability remains unexplained by the current feature set. This underscored the need for methodological refinements in subsequent experiments, particularly those designed to incorporate spatial hierarchies and relational dependencies beyond the capacity of tabular baselines.

3.4.6 Discussion and Implications of Baseline Model Performance

Despite its moderate performance metrics, the baseline CatBoost model offers compelling evidence of the feasibility of using multispectral remote sensing data to predict urban noise patterns. The alignment between high-noise predictions and known traffic corridors validates the underlying hypothesis that spectral and textural information derived from satellite imagery carries valuable acoustic signals. This is consistent with findings from related environmental modelling studies, where remote sensing features have successfully been applied to predict air pollution concentrations, vegetation health, and surface temperatures (Wang et al., 2020).

However, the limitations of the baseline model are equally instructive, revealing clear areas where the initial approach fell short. A first example can be drawn from Southampton Common, situated below Burgess Road (Figure 3.14). This large green space contains extensive vegetation cover, which in reality provides stronger noise attenuation than the adjacent low-density residential area located between Burgess Road and Winchester Road. In the CatBoost predictions, however, both zones were modelled as broadly equivalent low-exposure environments. The model failed to distinguish that the vegetated park should exhibit systematically lower noise levels than the nearby housing, highlighting its limited ability to represent non-road acoustic processes. This mismatch reflects both the constraints of the image-derived features and the tendency of GLCM-based statistics to integrate contextual information from outside the 30 m aggregation window, which diluted the contrast between these two environments.

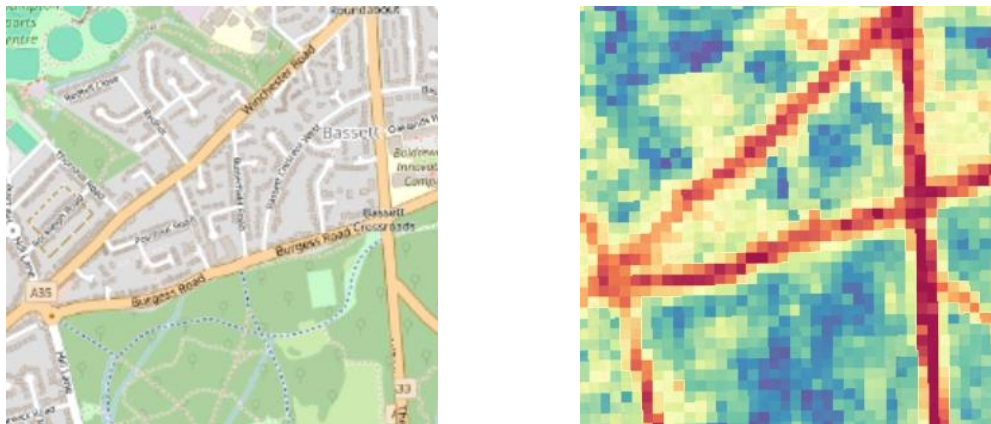


Figure 3.14 Comparison of observed land cover (OpenStreetMap basemap) and CatBoost-predicted noise exposure for the area surrounding Southampton Common

The model failed to capture the expected lower noise levels within the vegetated parkland compared to the adjacent low-density residential zone between Burgess Road and Winchester Road, illustrating the limited capacity of cell-level texture features to represent non-road acoustic processes.

A second example illustrates the issue of limited spatial differentiation and muted local noise variation. In central Southampton, a cluster of open green spaces forms a continuous urban park system. As illustrated in Figure 3.15, visual inspection suggests that these areas should register well below 55 dBA, yet the CatBoost predictions consistently placed them above this threshold. The outputs appeared smoothed, with

elevated noise levels assigned uniformly across the park system, failing to capture the expected variability between quiet internal areas and noisier perimeters adjacent to traffic. This systematic overestimation demonstrates the over-smoothing effect of the model: by relying solely on cell-level GLCM features, CatBoost produced outputs that lacked the fine-grained spatial variation present in reality.

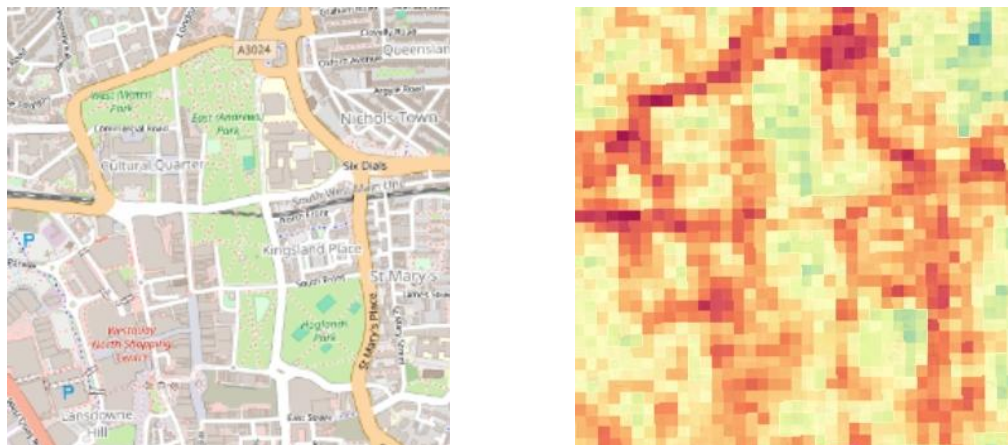


Figure 3.15 Comparison of observed land cover (OpenStreetMap basemap) and CatBoost-predicted noise exposure in central Southampton.

Despite the presence of multiple connected open green spaces, predictions consistently exceeded 55 dBA and displayed limited internal variability. This example highlights the model's tendency towards over-smoothing, resulting in muted local noise variations and an underestimation of fine-scale differentiation between quiet interiors and traffic-exposed edges.

These limitations point to the necessity of expanding the feature scope and enhancing the capacity to capture spatial dependencies. In particular, the underrepresentation of non-road noise sources, such as port and industrial activity, suggests that spectral-band-specific texture metrics are insufficient on their own. Incorporating land use classifications, morphological indices, and point-of-interest data would provide the semantic context needed to complement spectral information. Moreover, the over-smoothing evident in these examples underscores the need for models that explicitly account for spatial autocorrelation. Noise propagation inherently reflects neighbourhood effects, and approaches such as convolutional neural networks (CNNs) and graph neural networks (GNNs) are specifically designed to capture multi-scale and graph-structured dependencies within urban systems (Khan et al., 2018; Ranpise & Tandel, 2022).

Taken together, these insights emphasise the iterative nature of the research process. The CatBoost baseline successfully confirmed the viability of remote sensing data for noise prediction, but its shortcomings in feature representation and spatial differentiation established a clear motivation for methodological refinement. These reflections directly inform the trajectory of the dissertation: Chapter 4 enriches input features and leverages CNNs for spatial pattern extraction, Chapter 5 introduces graph-based learning with GraphSAGE to model relational dependencies, and Chapter 6 advances dual-branch GCN–GAT architectures to enable cross-city generalisation. Collectively, these developments build upon the baseline analysis to construct a scalable, transferable, and semantically richer framework for urban noise mapping.

Chapter 4: Predicting Urban Noise Levels Using

EfficientNet and Multispectral Remote Sensing Data: A

Case Study of Southampton

Abstract

Accurate mapping and assessment of urban noise exposure remain critical for effective urban planning and public health protection. Traditional urban noise mapping methods, heavily reliant on extensive field measurements and computationally intensive noise propagation modeling, are often costly, labor-intensive, and difficult to generalize across diverse urban environments. Recent advancements integrating remote sensing data with machine learning techniques have shown promise in addressing these limitations, but challenges persist in effectively capturing spatial dependencies and localized noise variations within urban contexts.

To enhance urban noise prediction performance, this study introduces a convolutional neural network (CNN)-based modeling framework that directly leverages spatial contextual information embedded in multispectral remote sensing imagery. Differing from previous baseline models that relied on statistical aggregation of textural features, this work systematically improves CNN inputs by integrating carefully selected remote sensing indices and corresponding texture feature maps. Employing rigorous feature selection methods ensures that only the most informative texture features derived from multispectral indices are utilized, optimizing model accuracy and efficiency.

Applied to the city of Southampton, UK, using WorldView-2 imagery collected in Feb 2016 & in May 2016, the proposed CNN model significantly improves predictive accuracy compared to traditional methods. Results clearly demonstrate enhanced capability in capturing spatially detailed urban noise distributions, particularly along major roadways and vegetated areas. Despite limitations in accurately representing noise sources unrelated to road traffic (e.g., industrial areas), the CNN approach shows substantial promise for scalable, spatially detailed, and cost-effective urban noise assessment. These findings highlight the potential of CNNs and advanced remote sensing techniques to inform urban planning strategies, noise mitigation policies, and future research directions.

Keywords: deep learning, convolutional neural networks, multispectral imagery, urban noise prediction, EfficientNet

4.1 Introduction

The previous chapter established a strong baseline for data-driven urban noise prediction by applying CatBoost to grid-level statistical summaries derived from high-resolution multispectral imagery. This baseline demonstrated the feasibility of using remote sensing data to model urban noise patterns in a structured and reproducible way. By aggregating spectral and textural predictors—such as means and medians of grey-level co-occurrence and vegetation indices—CatBoost achieved stable performance across large spatial extents, providing an interpretable and computationally efficient starting point for subsequent modelling stages. However, this approach also revealed inherent structural limitations. The reliance on pre-aggregated statistics, while stabilising predictor sets, inevitably blurred local contrasts and attenuated spatial detail. As documented in Section 3.4.6, these smoothing effects reduced the ability to discriminate between acoustically heterogeneous micro-environments, such as quiet interior green spaces versus noise-exposed road edges. The consequences extend beyond methodological precision: the muted representation of spatial variability can lead to systematic underestimation of green infrastructure benefits, misclassification of exposure gradients, and downstream biases in environmental health assessments.

The present chapter addresses these shortcomings through a shift from feature aggregation to direct spatial representation learning. Convolutional Neural Networks (CNNs) operate directly on raw pixel matrices rather than summary statistics, learning filters that capture spatial hierarchies from local edges to broader urban form patterns. This hierarchical encoding allows the model to preserve subtle intra-urban variations—such as the attenuation effects of tree canopies or courtyard buffering—while simultaneously integrating neighborhood-scale and city-scale structural context. By learning directly from multispectral imagery, CNNs overcome the smoothing artifacts of statistical descriptors, enabling finer differentiation of noise environments and improving the fidelity of exposure mapping.

To operationalize this approach, WorldView-2 multispectral imagery is organized into $1\text{ km} \times 1\text{ km}$ image patches (corresponding to 250×250 pixels at 4 m resolution). This patch size was chosen to balance local feature richness with manageable computational cost, ensuring that each input encompasses multiple land cover types, transportation structures, and urban configurations. The CNN architecture adopted in this study is EfficientNetB0, which employs compound scaling of network depth, width, and input resolution. This design offers a favorable balance between representational capacity and training efficiency, making it particularly well suited for large-scale urban remote sensing tasks.

This chapter thus marks a deliberate methodological transition from tabular learning to spatially structured deep learning. While the CatBoost baseline of Chapter 3 provided an interpretable but smoothed representation of noise-relevant predictors, the CNN framework presented here embeds spatial context directly into the learning process, capturing both micro-scale heterogeneity and meso-scale structure. At the same time, it establishes the conceptual and technical foundation for the subsequent graph-based models in Chapter 5, which extend spatial reasoning beyond image patches through explicit neighborhood relationships and graph-based feature

propagation. In this way, Chapter 4 serves as the critical intermediate stage in the broader progression from handcrafted feature learning to structured, spatially informed, and ultimately transferable deep learning architectures for urban noise prediction.

4.2 Methodology

4.2.1 Study Area, Data Sources, and Scale-Aware CNN Design

The methodological framework in this chapter builds on the unified paradigm introduced in Chapter 3 but extends the baseline analysis to a convolutional neural network (CNN) regression task. In contrast to the tabular predictors used in the CatBoost baseline, the CNN directly ingests image patches, enabling the model to capture hierarchical spatial structures and contextual information. The overall workflow is illustrated in Figure 4.1, which shows the sequential process from remote sensing imagery preprocessing and noise data aggregation, to the preparation of matched training datasets, CNN training, and the generation of city-scale noise prediction maps.

The study area is Southampton, United Kingdom, a city characterised by diverse land cover including dense residential districts, major transport corridors, industrial zones, and large green spaces. This heterogeneity makes Southampton a suitable test bed for examining whether image-based learning improves spatial differentiation in noise prediction relative to tree-based baselines.

The remote sensing input was derived from WorldView-2 multispectral imagery, mosaicked from two acquisitions (15 February 2016 and 6 May 2016). ENVI's seamless mosaic algorithm was used to align the scenes and perform colour balancing, with the lower-cloud May scene as the reference. The mosaic was resampled to a uniform 4 m spatial resolution. Although native multispectral bands are finer than 4 m, this target resolution preserves fine-scale morphological detail while reducing redundancy and computational load. Because the target labels are aggregated to a 30 m grid, retaining imagery at 4 m provides a sufficiently detailed input scale that remains compatible with the urban-scale analytical framework of this study.

Noise data were collected in 2016 using calibrated recorders and GPS under stable weather conditions. A total of 52,364 raw samples at ~ 1 m positional spacing were aggregated to $30\text{ m} \times 30\text{ m}$ cells using energy-domain averaging, yielding 13,474 representative LA_{eq} values across Southampton. Harmonising the imagery and observations at this target grid created a consistent basis for CNN training.

Scale-aware CNN design and receptive field mapping.

To make explicit how the network's spatial integration relates to urban physical scales, we treat the CNN's receptive field (RF) in metres rather than pixels. Let the imagery be at 4 m per pixel and the input patch be $P \times P$ pixels; the patch's physical field-of-

view is 4P metres on a side. Within the network, the theoretical RF after layer L satisfies:

$$RF_L = RF_{L-1} + (k_L - 1) d_L \prod_{t=1}^{L-1} s_t$$

where k_L is kernel size, d_L dilation, and s_t stride at layer t (with $RF_0 = 1$ pixel). Mapping to metres uses the 4 m pixel size. Shallow layers (e.g., 3×3 , *stride* 1) therefore emphasise features at ~12–20 m scales, capturing street edges, narrow verges, and small vegetation strips; mid-depth layers integrate ~50–150 m contexts, where block morphology and frontage continuity matter; deeper layers, especially where strides accumulate, integrate >150 m contexts that align with corridor-scale patterns along A-roads and arterial connectors. In practice, this depth-dependent expansion explains the observed smoothing: as the effective RF grows, predictions increasingly reflect neighbourhood-scale continuity and corridor alignment, attenuating very local fluctuations that are idiosyncratic to individual façades or micro-greenspace fragments. This smoothing is a design consequence of the RF-to-metres mapping rather than an arbitrary artefact, and it is desirable to the extent that the target variable (daytime LAeq aggregated at 30 m) is itself a neighbourhood-level exposure.

In this chapter, the CNN adopts an EfficientNet-B0 backbone adapted for multi-channel regression. The backbone’s compound scaling implies that early blocks operate with stride 1 and kernels, while downsampling blocks expand the RF more rapidly through stride-2 operations. Under a typical $P=64$ patch configuration at 4 m resolution (256 m footprint), the deepest effective RF covers on the order of a city block to sub-corridor scale; increasing P shifts the integration window upward without changing the RF growth law. We therefore view P and the stride schedule as explicit scale controls: smaller patches promote local contrast at the expense of corridor coherence; larger patches prioritise morphological continuity but yield smoother outputs. Section 4.4.4 returns to this point when interpreting the spatial smoothness of the predicted maps.

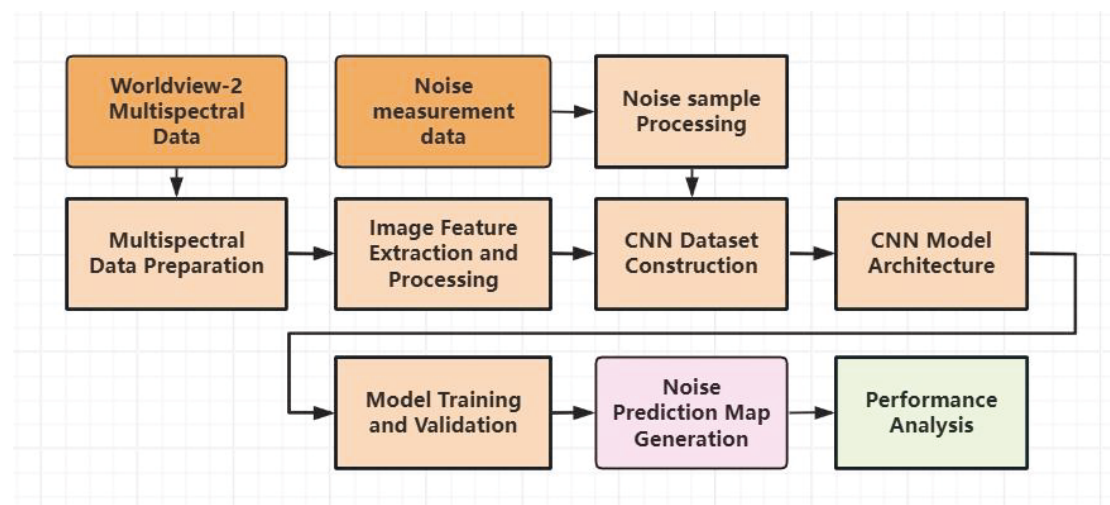


Figure 4.1 Flowchart of the Overall Methodological Framework

As depicted in Figure 4.1, spectral indices and texture derivatives are processed into multi-channel feature maps that align with the 30 m noise grid for supervised learning. These paired data are used to train the EfficientNet-based CNN; the trained weights are subsequently applied convolutionally to the full mosaic to generate continuous maps. This represents a deliberate progression from the CatBoost baseline: while Chapter 3 established the feasibility of remote-sensing-derived predictors, the CNN tests whether hierarchical, scale-aware feature extraction more effectively captures urban acoustic heterogeneity.

Relation to graph-based modelling (forward reference). Whereas the CNN expands its RF over Euclidean patches, the graph-based framework in Chapter 5 propagates information across a spatial graph defined on the 30 m lattice. Each GNN layer aggregates one-hop neighbours so that depth corresponds to k-hop neighbourhoods—an alternative, topology-aware notion of RF. This difference clarifies why CNNs tend toward block- and corridor-scale smoothing, while GNNs can transmit influence along anisotropic structures (e.g., transport corridors) even when they cut across pixel grids. Detailed treatment of neighbourhood propagation and depth trade-offs is provided in Section 5.2.

4.2.2 Generation of Remote Sensing Indices

Unlike the baseline framework in Chapter 3, which primarily relied on single-band values and their textural summaries, the present study expands the feature design by building upon raw band reflectances with additional multispectral indices and texture maps derived from these indices. The rationale for this extension rests on both theoretical and practical grounds. While individual spectral bands capture narrow wavelength responses, these alone are often insufficient to disentangle the complex mixtures of vegetation, impervious surfaces, and soil that typify urban areas. By contrast, indices exploit band ratios and combinations to emphasise specific land-surface properties, thereby reducing redundancy and improving the semantic relevance of predictors for urban noise modelling. Moreover, index-derived texture features further enhance the representation of spatial heterogeneity, yielding biophysical proxies for phenomena directly implicated in acoustic propagation, such as vegetation density, ground sealing, and water presence.

Accordingly, seven indices were systematically generated from the mosaicked and resampled WorldView-2 imagery, chosen to cover the core environmental dimensions of urban noise dynamics (Table 4.1). NDVI quantifies vegetation density, which is critical for noise attenuation through absorption and scattering by tree canopies. NDWI isolates water bodies, whose reflective surfaces influence sound propagation and whose spatial presence alters local urban form. Four complementary built-up indices (BAI, BSI, WVBI, NBEI) collectively characterise the extent and intensity of impervious surfaces that dominate traffic-related noise sources. Despite their shared focus, these indices differ in spectral design and error sensitivities: BAI accentuates built-up zones but may exaggerate edges next to vegetation; BSI introduces soil-related confusion; WVBI, specifically adapted for WorldView data, reduces shadow artefacts; and NBEI improves over BSI by minimising soil misclassification. Retaining these subtle distinctions ensures that the CNN framework can exploit

diverse spectral cues, rather than inheriting the biases of a single index. Finally, the Red-Green Index (RGI) captures chromatic contrasts linked to vegetation–soil–built-up transitions, offering an additional discriminative layer in complex mixed-use zones.

Index	Definition	Reference
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{\text{Red} - \text{NIR1}}{\text{Red} + \text{NIR1}}$	Jensen, 2007
Normalized Difference Water Index (NDWI)	$NDWI = \frac{\text{Green} - \text{NIR1}}{\text{Green} + \text{NIR1}}$	Gao, 1996
Built-up Area Index (BAI)	$BAI = \frac{\text{Blue} - \text{NIR1}}{\text{Blue} + \text{NIR1}}$	Adeyemi et al., 2021
Built-up Spectral Index (BSI)	$BSI = \frac{\text{Yellow} - 2 \times \text{NIR1}}{\text{Yellow} + 2 \times \text{NIR1}}$	Sameen & Pradhan, 2016
Red-Green Index (RGI)	$RGI = \frac{\text{RedEdge} - \text{Green}}{\text{RedEdge} + \text{Green}}$	Motohka et al., 2010
WorldView Built-up Index (WVBI)	$WV - BI = \frac{\text{Coastal Blue} - \text{Red Edge}}{\text{Coastal Blue} + \text{Red Edge}}$	Wolf, 2012
Normalized Built-up Extraction Index (NBEI)	$NBEI = \frac{(\text{NIR2} - \text{NIR1}) - (\text{Green} + \text{RedEdge})}{(\text{NIR2} - \text{NIR1}) + (\text{Green} + \text{RedEdge})}$	Adeyemi et al., 2021

Table 4.1 Summary of the indices, their definitions, and key references

This design marks a clear methodological progression from Chapter 3. Whereas the baseline relied on single-band descriptors that emphasised tonal variation without

semantic context, the present framework deploys indices as structurally meaningful representations of environmental factors governing acoustic conditions. By moving from raw reflectance to indices, the CNN is provided with richer, less redundant, and more physically interpretable inputs, strengthening its ability to learn generalisable patterns.

Each index was computed at 4 m spatial resolution following atmospheric correction and geometric alignment, balancing spatial detail with the need to harmonise inputs with the 30 m acoustic aggregation grid. The resulting continuous maps represent complementary environmental layers: vegetation greenness, surface wetness, and multiple facets of urban built-up intensity. Together, these indices offer a semantically grounded and methodologically robust basis for the CNN to learn acoustic-relevant surface properties beyond what single bands could convey.

Figure 4.2 presents the spatial distribution maps of the seven indices across Southampton. The maps illustrate the city’s heterogeneity, with vegetation-dominated areas in Southampton Common, high-density built-up zones in the city centre, and distinct water features along the River Itchen. They also demonstrate how indices targeting built-up structures, though broadly similar, capture fine-scale differences in the urban fabric. These differences are particularly important in contexts such as dense residential areas versus commercial–industrial zones, where subtle spectral distinctions may influence noise propagation. By stacking these indices as multi-channel CNN inputs, the framework ensures that the model can learn from both consistent urban signals and index-specific variations.

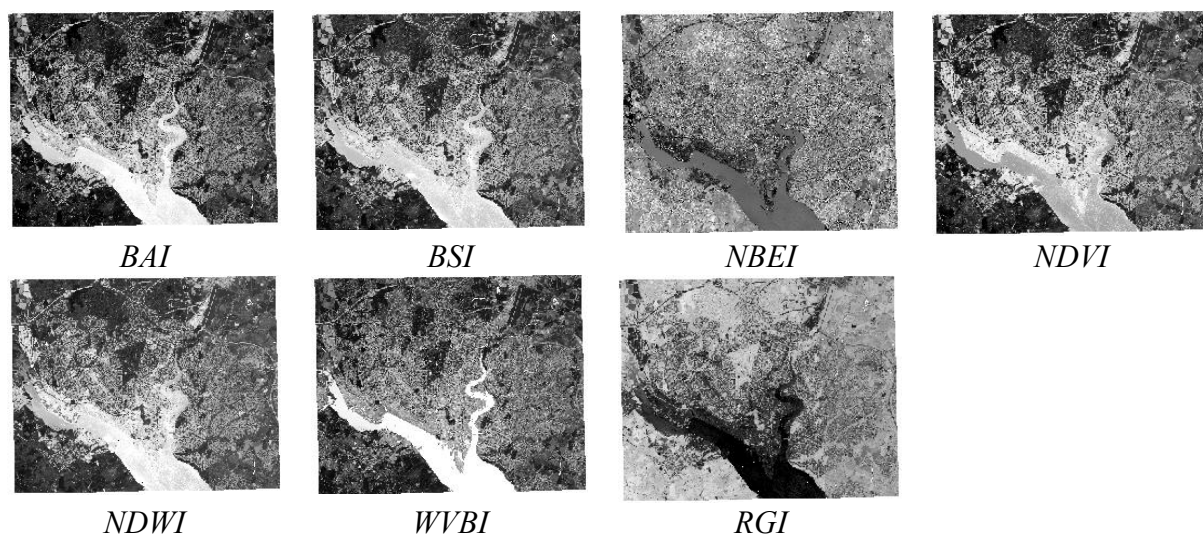


Figure 4.2 Maps of BAI, BSI, NBEI, NDVI, NDWI, WVBI, and RGI for the study area

4.2.3 Generation and Selection of Texture Feature Maps

To complement the spectral indices introduced in Section 4.2.2, this study extended the feature space through a structured texture analysis pipeline. Unlike the baseline in Chapter 3, which relied on the five standard GLCM descriptors available in ENVI, here the analysis drew on the more comprehensive implementation provided by the Orfeo Toolbox (OTB), which offers a suite of 28 texture features spanning simple,

higher-order, and advanced descriptors (detailed definitions and formulations are provided in Appendix I). This broader design allowed the CNN to access a richer set of structural cues from the imagery, improving its ability to capture the fine-scale spatial heterogeneity of urban surfaces.

A critical methodological refinement concerned the quantisation of input data prior to co-occurrence analysis. In Chapter 3, features were derived from imagery quantised at 64 grey levels, following conventional fine-scale texture analysis. In the present study, this was optimised to 32 grey levels. This adjustment reflects a deliberate trade-off: while 64 levels can overfit fine tonal variations and produce excessively sparse co-occurrence matrices, 32 levels preserve sufficient discriminatory power at the urban scale while yielding more stable estimates of descriptors such as entropy and correlation. Empirical research supports the use of 16–32 levels in urban applications (Hall-Beyer, 2017), and given that the imagery was resampled to 4 m resolution to align with the 30 m acoustic aggregation grid, 32 levels were judged optimal for balancing statistical robustness and computational efficiency.

GLCMs were computed using a 5×5 moving window with a one-pixel offset, across four orientations (0° , 45° , 90° , 135°), and subsequently averaged to ensure rotational invariance. Each index generated 28 candidate descriptors, spanning measures of local contrast, entropy, uniformity and higher-order heterogeneity. Yet, not all descriptors proved equally discriminative in practice. A comparison between Grey-Level Nonuniformity (GLN) derived from NDVI and Homogeneity derived from BAI illustrates the sharp divergence in their effectiveness at the urban scale (Figure 4.3). For this reason, visual inspection was adopted as the first stage of feature evaluation, identifying descriptors that showed meaningful spatial stratification versus those collapsing into uniformity.

GLN reveals pronounced stratification across the Southampton urban landscape: the south-western harbour appears markedly dark, the compact city centre displays intermediate tones, and non-urban vegetation emerges as the lightest areas. These contrasts arise from GLN's sensitivity to irregularity in grey-level distributions, enabling the effective differentiation of surface heterogeneity at city scale. By contrast, Homogeneity from BAI collapses into near-constant values over most of the city, with only large uniform areas such as water bodies appearing distinct. This outcome reflects the descriptor's tendency to reward local spectral similarity while suppressing fine-scale urban heterogeneity, which substantially reduced its information content and ultimately led to its exclusion from the retained feature set.



(a) GLN derived from NDVI

(b) Homogeneity derived from BAI

Figure 4.3 Retained versus excluded GLCM texture maps.

(a) GLN derived from NDVI, showing stratification between harbour, city centre, and non-urban vegetation. (b) Homogeneity derived from BAI, largely collapsed to uniform values and excluded from the feature set.

While visual analysis was critical in the first stage of curation, the second stage relied on iterative model training to evaluate which groups of descriptors improved predictive performance. Starting from the single-band baseline, additional subsets—GLCM maps from spectral indices, from individual bands, and grouped into simple, higher-order, and advanced classes—were progressively tested. This process initially expanded the feature space (“addition”) and subsequently refined it (“subtraction”), ensuring both empirical validation and parsimony.

While the adoption of 32 grey levels improved the statistical robustness of GLCM features at the urban scale, a substantial subset of descriptors still produced maps with very limited or redundant information—for example, large uniform areas with little discriminative capacity. In this context, eliminating such features was not only a matter of computational efficiency but also a practical requirement, as the modelling task demands both sufficient spatial coverage and a diverse yet informative feature set. The retained feature maps (summarised in Appendix II) provided representations that were more coherent and semantically aligned with urban structures, ensuring that the CNN was trained on inputs with demonstrable spatial relevance. In this way, the final suite of 69 GLCM feature maps established a stable and semantically meaningful textural foundation for the CNN input tensor.

4.2.4 CNN Input Construction and EfficientNetB0 Architecture Design

To preserve the spatial configuration and semantic richness of the remote sensing data, this study adopts a patch-based convolutional neural network (CNN) framework, enabling direct learning from high-resolution imagery rather than aggregated statistics used in the CatBoost baseline. Specifically, the EfficientNetB0 backbone (Tan & Le, 2019) is employed for its strong accuracy–efficiency trade-off achieved through compound scaling of depth, width, and resolution.

Each patch is centered on a $30\text{ m} \times 30\text{ m}$ grid cell containing ground-truth LAeq measurements and extracted from the 4 m resolution imagery. Several patch sizes were tested (150×150 , 224×224 , 320×320), with 250×250 pixels ($\approx 1\text{ km} \times 1\text{ km}$) selected as the optimal balance between contextual richness and computational tractability. This window captures road networks, residential blocks, and vegetation buffers—features known to influence noise propagation (Yang et al., 2024).

The final input tensor has dimensions (250, 250, 84), integrating three data sources:

- (1) 69 GLCM-derived texture layers (entropy, correlation, contrast, etc.);
- (2) seven remote sensing indices (NDVI, NDWI, BAI, BSI, RGI, WVBI, NBEI);
- (3) eight original WorldView-2 multispectral bands (panchromatic excluded).

This unified tensor embeds spectral and textural information into a single structured representation. Figure 4.4 illustrates the extraction workflow, where vector-based noise samples are projected to raster layers, fixed-size neighborhoods are defined, and consistent patches are cut across all feature stacks.

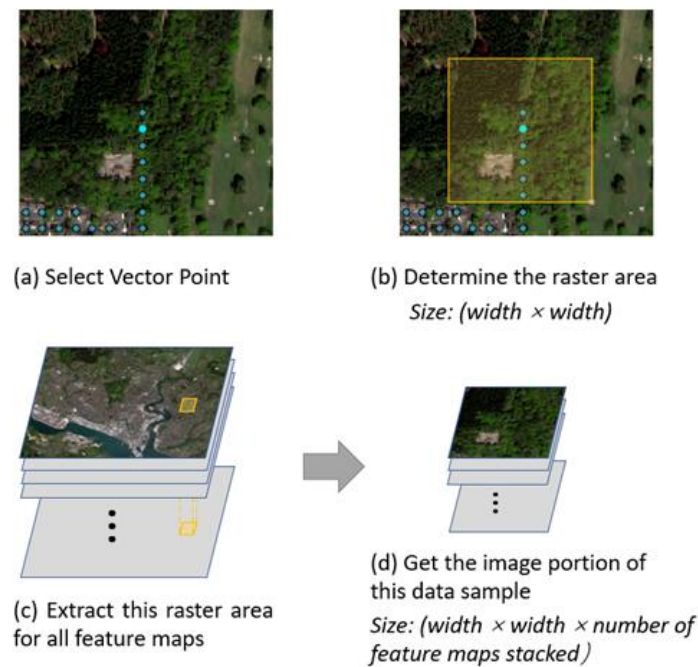


Figure 4.4 Workflow for CNN patch extraction and dataset assembly

(a) noise samples with vector coordinates; (b) projection to raster feature maps and neighborhood definition; (c) patch extraction across stacked feature layers; (d) resulting $250 \times 250 \times 84$ multi-channel patch.

To accommodate the high-dimensional inputs, EfficientNetB0's initial convolution layer was modified to accept 84 channels, and the final fully connected layer replaced with a single regression node for LAeq prediction (Figure 4.5).

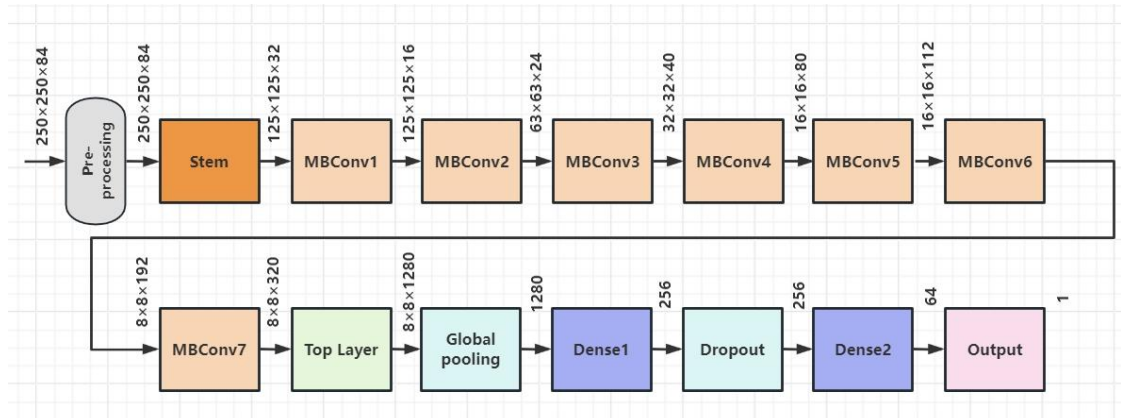


Figure 4.5 EfficientNetB0 modified architecture for multi-channel regression (with 84 input channels and single-node regression output)

To enhance robustness, Gaussian noise perturbation ($\sigma = 0.10$) was applied during training to the input channels. This mild stochasticity does not mimic noise levels directly but simulates sensor and environmental fluctuations, encouraging the model to learn stable spatial patterns. Given that field measurements span peak and off-peak traffic hours (Figure 3.3), this strategy aligns model training with real-world temporal variability. Empirical studies show hourly LA_{eq} fluctuations of ± 5 –10 dBA in urban cores (Alvares-Sanches et al., 2021b), supporting this regularization approach.

The dataset was partitioned using stratified random sampling: noise levels were divided into ten strata to preserve the distribution of low-, medium-, and high-noise conditions. A 70 % 15 % 15 % split for training, validation, and testing allows stable monitoring of convergence and generalisation, aligning with common CNN practices.

By combining large contextual windows, multi-source feature stacks, and a parameter-efficient architecture, this framework provides a spatially informed deep learning approach for urban noise prediction, while remaining flexible for scale-sensitive experiments.

4.2.5 Model Training and Prediction Workflow

Model training and prediction share the same patch construction pipeline, ensuring consistent spatial resolution (4 m), patch size (250x250), and target grid (30 m). During training, patches centered on labeled grid cells are used to fit the network, while prediction involves tiling the entire study area into overlapping patches, feeding them to the trained CNN, and reconstructing the final 30 m resolution noise exposure map (Figure 4.6).

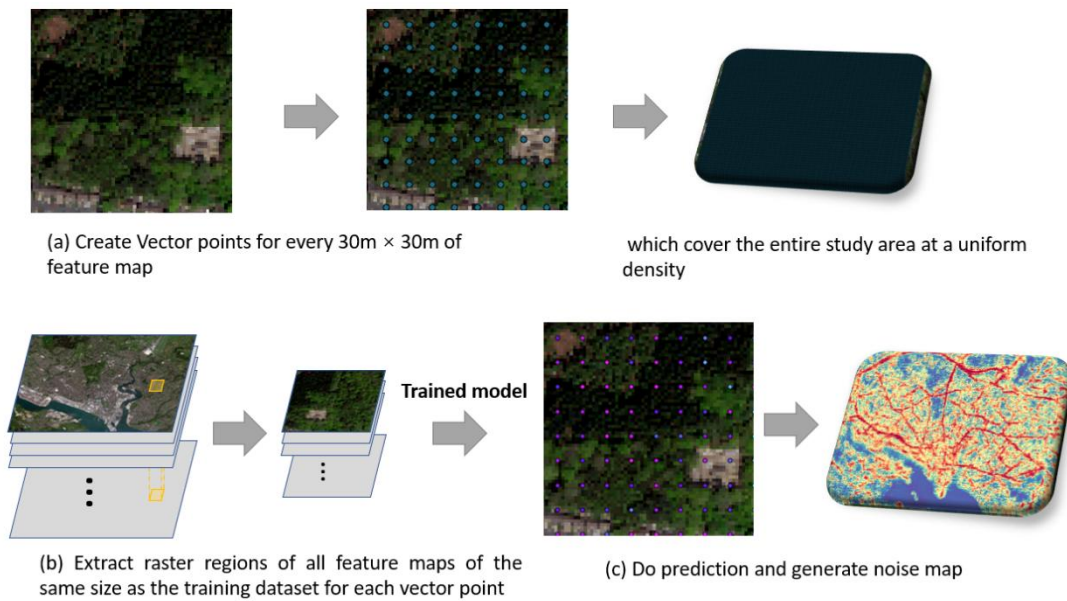


Figure 4.6 Prediction pipeline flowchart — depicting full-scene feature cube, patch extraction, model inference, and output reconstruction to 30 m grid

The consistency of input format between training and prediction is a critical component of this workflow. By standardizing both the spatial neighborhood size (1000 m), input resolution (4 m), and target grid resolution (30 m), we ensure that the model's generalization is not compromised by inconsistencies in input design. Furthermore, this modular structure supports future transferability to new cities or temporal scenes, where identical patching and feature stacking routines can be applied to enable seamless inference using a pre-trained model.

This methodical approach to model training and deployment ensures both robustness and scalability in real-world urban noise mapping applications. The subsequent section (Section 4.3) will present and evaluate the model's prediction outputs, using both quantitative error metrics and spatial visualizations.

4.3 Results

4.3.1 Model Training Curves and Convergence Behavior

To evaluate the training dynamics and convergence stability of the CNN model, the dataset was partitioned into 70% training, 15% validation, and 15% testing subsets using a stratified sampling strategy. Figure 4.7 illustrates the evolution of loss, Mean Absolute Error (MAE), and the coefficient of determination (R^2) across epochs, averaged over the validation folds.

The EfficientNetB0 backbone demonstrated smooth and stable convergence after approximately 140–150 epochs. The validation MAE stabilized around 4.62 dBA, with R^2 reaching 0.496, while performance on the independent test set converged to $\text{MAE} = 4.79$ dBA and $R^2 = 0.491$. These values confirm that the model maintained

strong generalization beyond the training data, without signs of overfitting or underfitting. The trajectories of both MAE and R^2 show gradual and monotonic improvement before reaching a plateau, indicating that the optimization strategy (Adam optimizer with learning-rate scheduling, early stopping, and dropout) provided a stable balance between bias and variance.

Beyond convergence dynamics, an extensive set of controlled experiments was undertaken to evaluate how different combinations of input channels, texture descriptors, and model architectures influenced predictive performance. Table 4.2 (Performance of different feature-map combinations and architectures in CNN-based noise prediction, reported on the test set) summarises these experiments, with the best-performing configuration highlighted at index 16. This model, based on EfficientNetB0 with texture features derived exclusively from WVBI, NDVI, and BSI, achieved the most robust balance between predictive accuracy and spatial fidelity.

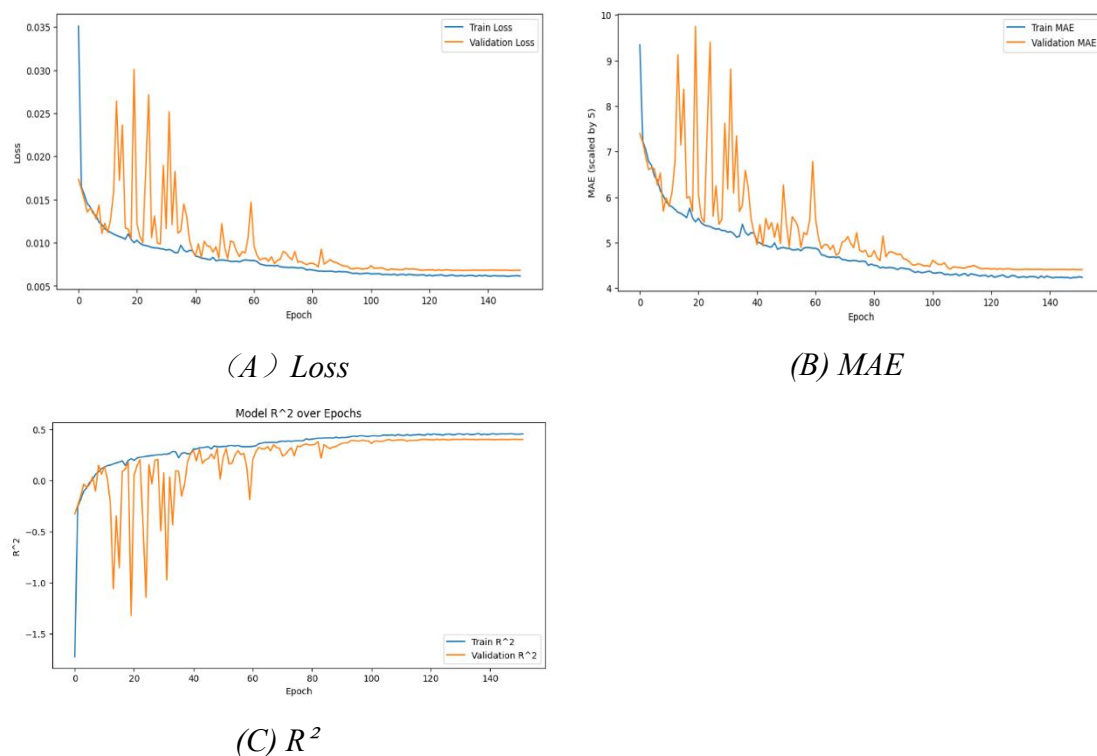


Figure 4.7 Model training curves across epochs — showing loss, MAE, and R^2 for training and validation sets

Three-stage feature screening strategy

The feature selection strategy was conducted in three stages, with each step guided not only by predictive performance but also by computational feasibility. Given the high dimensionality of the remote sensing input (potentially hundreds of channels once GLCM textures are included), resource limitations played a decisive role in the design. The size of the input neighborhood, the number of stacked feature maps, and the depth of candidate architectures all imposed significant computational costs. For example, EfficientNetB1 and deeper variants were not tested, as the required GPU memory would have exceeded available resources without a proportional gain in model interpretability or stability.

Table 4.2 summarises the tested combinations. In the notation, 8b refers to the eight original multispectral bands, while 7c represents the seven remote sensing indices (NDVI, NDWI, BAI, BSI, RGI, WVBI, NBEI). The terms *simple*, *advanced*, and *higher* correspond to the 28 GLCM texture descriptors available in the Orfeo Toolbox, divided into 8, 10, and 10 metrics respectively (see Appendix I for formal definitions). The keyword *local* indicates the five standard GLCM features (contrast, correlation, entropy, homogeneity, and energy) used in the Chapter 3 baseline, which are a subset of the simple group.

Table 4.2 Performance of different feature-map combinations and architectures in CNN-based noise prediction (testset MAE and R²)

Index	Single bands	RS Indexs	Feature maps	Patch size	model	MAE	R ²
1	8b+40 local	None	48	150×150×48	ResNet50	5.66	0.391
2	8b+40 local	None	48	150×150×48	ResNet34	5.63	0.388
3	8b+40 local	None	48	150×150×48	VggNet	5.75	0.376
4	8b+40 local	None	48	320×320×48	ResNet50	5.48	0.418
5	8b+40 local	None	48	250×250×48	ResNet50	5.49	0.426
6	8b+40 local	7c	55	250×250×55	ResNet50	5.43	0.433
7	8b+40 local	7c	55	250×250×55	EfficientNetB0	5.42	0.435
8	8b+64 simple	7c	79	250×250×79	EfficientNetB0	5.39	0.441
9	8b	7c+56 simple	71	250×250×71	EfficientNetB0	5.33	0.448
10	8b+80 advanced	7c	95	250×250×95	EfficientNetB0	5.35	0.431
11	8b	7c+70 advanced	85	250×250×85	EfficientNetB0	5.34	0.433
12	8b + 80 higher	7c	95	250×250×95	EfficientNetB0	5.31	0.44
13	8b	7c+70 higher	85	250×250×85	EfficientNetB0	5.22	0.451
14	8b	7c+25×5 mixed	140	250×250×120	EfficientNetB0	4.82	0.478
15	8b	7c+25 x 3 mixed	90	250×250×90	EfficientNetB0	4.85	0.488
16	8b	7c+25 x 3 mixed - 6	84	250×250×84	EfficientNetB0	4.79	0.491
17	8b	7c+25 x 3 mixed - 6	84	250×250×84	ResNet50	4.81	0.49

The first stage (rows 1–5 in Table 4.2) focused on selecting a backbone architecture and an appropriate neighborhood size. Several networks were compared, including VGGNet, ResNet34, ResNet50, and EfficientNetB0. Results indicated that deeper ResNet variants could reach comparable MAE scores but introduced striping artefacts in the predicted maps, undermining spatial realism. EfficientNetB0 achieved stable performance with fewer parameters, making it the most resource-efficient choice. Regarding neighborhood size, the 250×250 configuration (equivalent to a $1 \text{ km} \times 1 \text{ km}$ patch at 4 m resolution, i.e. a 500 m effective radius) provided a physically interpretable scale for capturing road corridors, urban blocks, and green buffers. Although slightly larger patches (e.g., 320×320) yielded marginal improvements, the

performance gain was minimal compared with the increase in computational cost. The 250×250 design was therefore adopted as a balanced compromise.

The second stage (rows 8–13) investigated whether textural descriptors should be generated from raw spectral bands or from remote sensing indices. When paired with the Orfeo Toolbox's comprehensive 28-feature set, index-based descriptors consistently outperformed band-based ones in both MAE and R^2 . This is because indices such as NDVI, WVBI, and BSI amplify biophysical contrasts relevant to acoustic propagation, making the derived textures more semantically meaningful than those produced directly from raw bands.

The third stage (rows 13–17) involved progressive simplification of the feature set. At each step, descriptors with low discriminative power or redundancy were excluded. Specifically:

1. Row 13 excluded sum average, sum variance, and sum entropy, which contributed little dynamic range.
2. Row 14 removed GLCM features derived from BAI and NDBI, as these indices tended to overlap with other built-up measures while providing limited additional information.
3. Row 15 further excluded features derived from NDWI and RGI, on the grounds that water and vegetation patterns were already adequately delineated by NDVI- and BSI-based textures.
4. Rows 16 and 17 introduced finer pruning, removing Cluster Shade and Cluster Prominence under both BSI and NDVI, and additionally excluding Difference Variance and Inertia from NDVI.

This progressive reduction distilled the input space to a compact but semantically rich subset of texture maps. The optimal configuration, reported at row 16, retained descriptors generated exclusively from WVBI, BSI, and NDVI. This model, based on EfficientNetB0, achieved $\text{MAE} = 4.79$ dBA and $R^2 = 0.491$ on the test set, representing the most robust trade-off between predictive accuracy, spatial fidelity, and computational efficiency.

4.3.2 Predicted Noise Map over the Entire Study Area

After training and validation, the EfficientNetB0 model was applied to the entire study area in a fully spatial prediction mode. As described in Section 4.2.6, input patches were constructed for all $30 \text{ m} \times 30 \text{ m}$ grid cells across Southampton, with each patch centered on its corresponding grid cell and composed of 84 feature channels. This resulted in a complete, spatially continuous noise map at 30 m resolution.

Figure 4.8 presents the predicted noise map overlaid on the WorldView-2 base imagery. The map clearly reflects the morphological and functional complexity of Southampton's urban landscape. High-noise zones appear as well-defined linear structures aligned with major transportation corridors, including the A33 and A3024 trunk roads and key intersections in the city center. These zones consistently show

predicted *L_{Aeq}* levels exceeding 70 dBA, which is consistent with expectations for high-traffic areas in dense urban settings.

In contrast, low-noise regions (<55 dBA) are predominantly located in peripheral residential neighborhoods, parklands, and the eastern forested areas along the river. These areas are characterized by high NDVI and texture smoothness, indicating dense vegetation and homogenous surface coverage. The model appears to have successfully captured these spatial relationships, associating green infrastructure with reduced acoustic exposure.

Transitional areas—such as urban-rural interfaces, residential-commercial edges, and zones adjacent to highways—exhibit medium noise levels (55–65 dBA), forming soft gradients rather than abrupt boundaries. This suggests the model’s capacity to encode contextual variation at the neighborhood scale, thanks to the 1 km patch input size which allows broader spatial features to influence predictions.

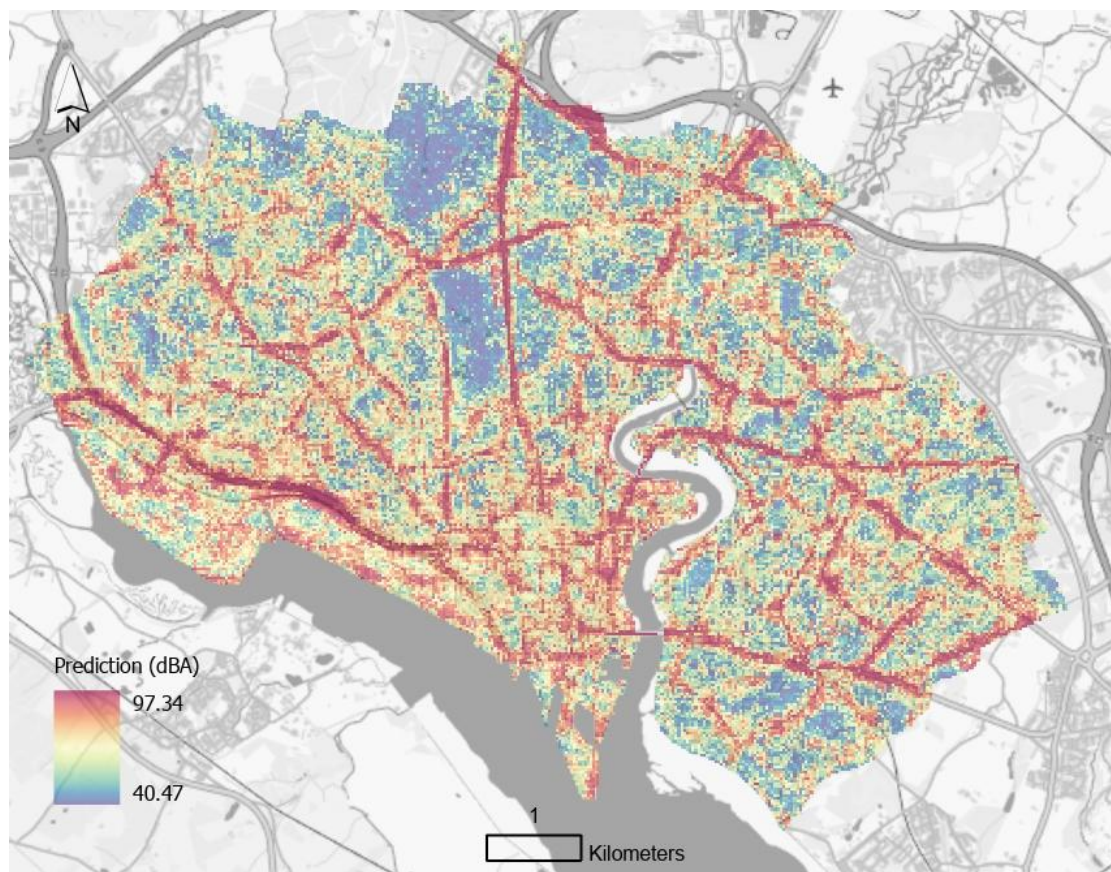


Figure 4.8 Predicted urban noise exposure map (30 m resolution) over Southampton with light gray canvas base imagery overlay

Overall, the prediction surface exhibits both structural clarity and contextual smoothness. It successfully highlights macro-scale urban noise phenomena (e.g., traffic corridors), while also representing micro-scale variations across smaller urban blocks. The map provides a spatially explicit foundation for interpreting noise distribution and guiding subsequent planning or mitigation interventions.

4.3.3 Prediction Accuracy and Validation on Labeled Samples

To evaluate the internal consistency of the CNN predictions, model outputs were compared against the 13,474 labelled 30 m × 30 m grid cells used for validation. While these points were part of the training dataset and therefore do not represent an independent test, the analysis provides valuable insights into systematic prediction biases, spatial variability, and the effective dynamic range of the model.

Figure 4.9 presents the spatial distribution of measured and predicted values. Panel (a) shows measured noise levels, spanning a range from 39.1 to 100.1 dBA, while panel (b) shows CNN predictions, with a slightly narrower range of 41.04 to 97.38 dBA. Both maps reproduce the same broad spatial patterns, with high exposures along arterial corridors and central urban zones, and lower values associated with vegetated or peripheral areas. This alignment indicates that the CNN is able to reproduce the dominant structures of the acoustic environment, although the narrower predictive range reflects a modest regression-to-the-mean effect.

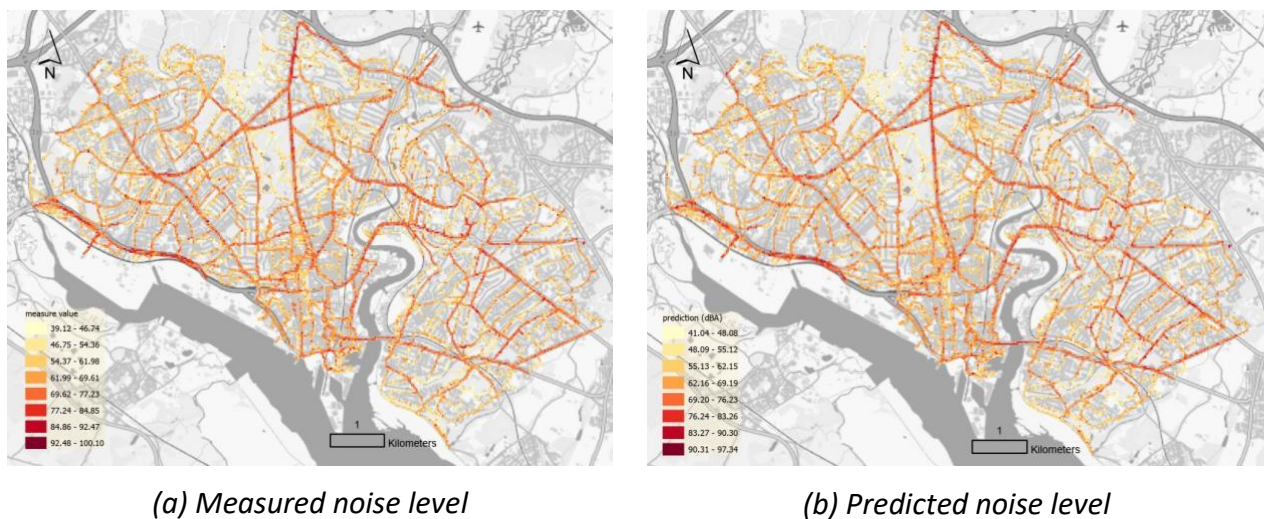


Figure 4.9 Prediction error map over labeled validation samples — showing spatial distribution of residuals (dBA)

Figure 4.10 overlays measured and predicted maps to visualise differences at individual grid cells. Most discrepancies fall within the 0–4.27 dBA range, with only a minority exceeding 12.8 dBA. These larger deviations are typically located in mixed-use or transitional areas, where acoustic environments are highly variable and short-term anomalies (e.g., intermittent traffic bursts) are not fully represented in spectral-textural predictors.

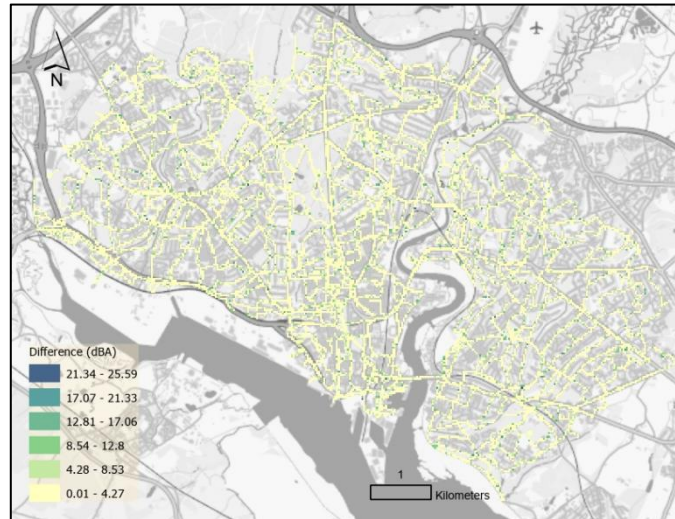
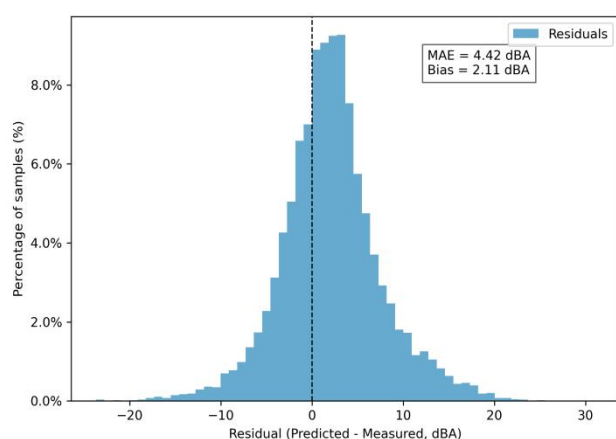


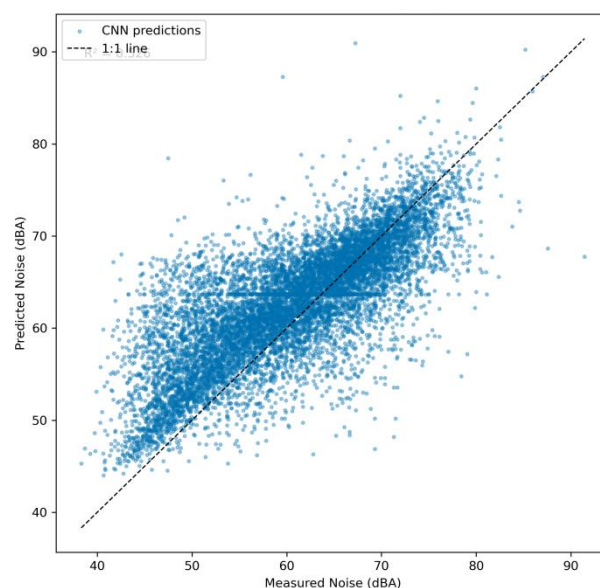
Figure 4.10 Overlay of measured and predicted noise maps with difference classes highlighted

The predictive behaviour of the CNN model is summarized in Figure 4.11. Panel (a) shows the residual distribution (Predicted – Measured), while panel (b) presents the measured versus predicted noise levels together with the 1:1 identity line. The histogram in panel (a) exhibits a sharply defined central peak around zero, with most residuals confined within ± 5 dBA and only a small proportion of samples extending beyond ± 10 dBA. The curve is nearly symmetric but shows a mild positive skew, indicating a systematic mean bias of approximately +2.1 dBA. This suggests that the CNN tends to slightly overestimate actual noise levels across the study area.

Such overestimation is physically plausible and statistically consistent with the characteristics of the training data. As detailed in Chapter 5, a subset of field measurements includes sporadic high readings likely caused by transient environmental conditions such as wind interference or reflective surfaces, which elevate measured sound levels in localized spots. When these samples are spatially aggregated into 30 m cells, they disproportionately raise the global mean of the measured dataset (from 59.76 to 60.57 dBA). Consequently, the CNN—trained on these aggregated targets—learns a slightly higher baseline response, producing predictions with a modest upward shift relative to the measured distribution.



(a) Histogram of residual



(b) Scatter plot of measured vs. predicted noise level

Figure 4.11 Histogram of residuals; (b) Scatter plot of measured vs. predicted noise levels with identity line

Panel (b) confirms this trend. The scatter of measured versus predicted values forms a dense, nearly linear cluster around the 1:1 line, particularly within the dominant mid-range of 45 – 70 dBA, demonstrating strong consistency between observed and estimated noise levels. Systematic deviations emerge only at the distributional extremes: predictions tend to saturate for the highest observed exposures (> 75 dBA), reflecting a controlled attenuation of extreme values induced by the model’s regularization and Gaussian noise augmentation strategy (Section 4.2.4). At the lower end (< 45 dBA), the model exhibits mild overestimation, a common artefact in regression models trained on skewed environmental data where low-intensity samples are underrepresented.

Overall, the CNN achieves a well-balanced performance—accurate across the main exposure range, robust against local measurement anomalies, and physically interpretable in its residual behaviour. The small positive bias reflects the combined effects of measurement aggregation and model regularization rather than a calibration error, underscoring the CNN’s stability and realism in predicting continuous urban noise fields.

4.3.4 Spatial Extremes: Quietest and Noisiest Zones

To further investigate the spatial patterns captured by the CNN, we extracted the top 20% noisiest and top 20% quietest grid cells across the study area (135 cells in each category). These boundary cases highlight the model’s sensitivity to contrasts in land

cover and urban form, and also reflect the role of the dominant spectral indices—WVBI, NDVI, and BSI—in shaping predictions.

The noisiest grid cells (Figure 4.12a) are clustered along Southampton’s primary traffic corridors, including the inner ring road, dock access routes, and intersections in the commercial core. These areas are characterised by low NDVI (indicating minimal vegetation), high BSI (denoting extensive impervious and bare surfaces), and low WVBI (marking built-up intensity along the coastal–red-edge gradient). Such conditions correspond to dense asphalt and concrete coverage, which amplify and reflect noise. The model assigns LAeq values exceeding 75 dBA in these regions, aligning closely with high-exposure zones documented in field surveys.

In contrast, the quietest grid cells (Figure 4.12b) are concentrated in peripheral green areas such as Southampton Common, Itchen Valley Country Park, and forested buffers on the city’s eastern edge. These areas exhibit high NDVI (dense vegetation cover), low BSI (minimal impervious surfaces), and elevated WVBI (reflecting vegetated–built contrasts at landscape edges). Together with their distance from major transport corridors, these features create strong buffering effects. Predicted levels in these cells fall below 50 dBA, underscoring the capacity of vegetation and land-cover heterogeneity to mitigate urban noise.

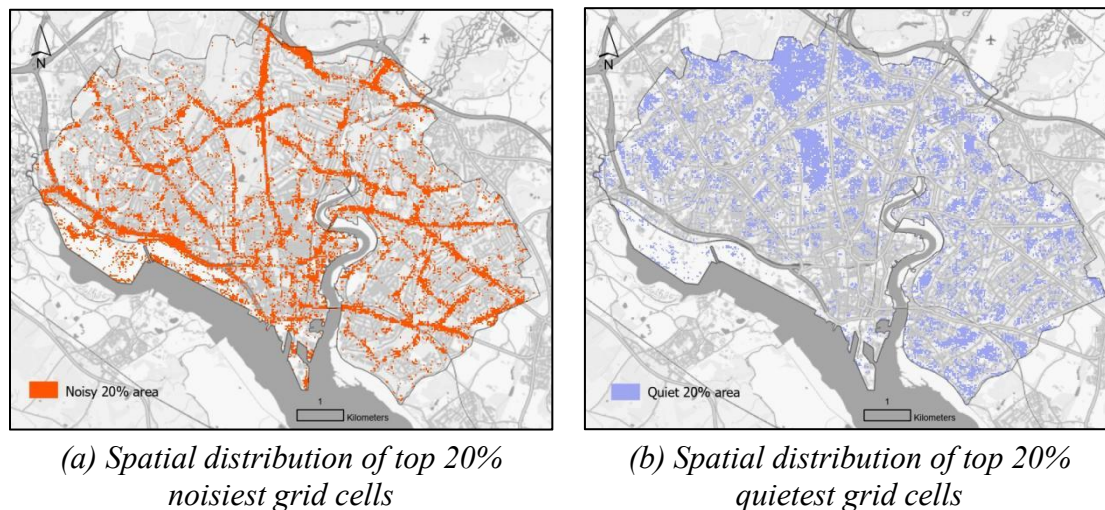


Figure 4.12 Spatial distribution of top 20% noisiest grid cells & top 20% quietest grid cells

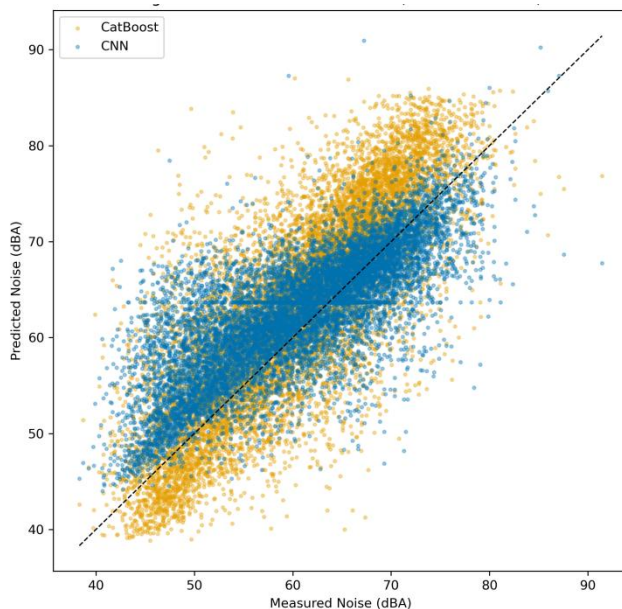
The separation between these two extremes not only demonstrates the model’s ability to discriminate acoustic conditions but also clarifies how its feature composition translates into spatial predictions: high noise is systematically associated with low NDVI, high BSI, and suppressed WVBI, while quiet zones are defined by the inverse. These patterns are consistent with theoretical principles of noise propagation and validate the explanatory role of remote sensing–derived indices in capturing urban acoustic dynamics.

4.3.5 Quantitative Comparison with CatBoost Baseline

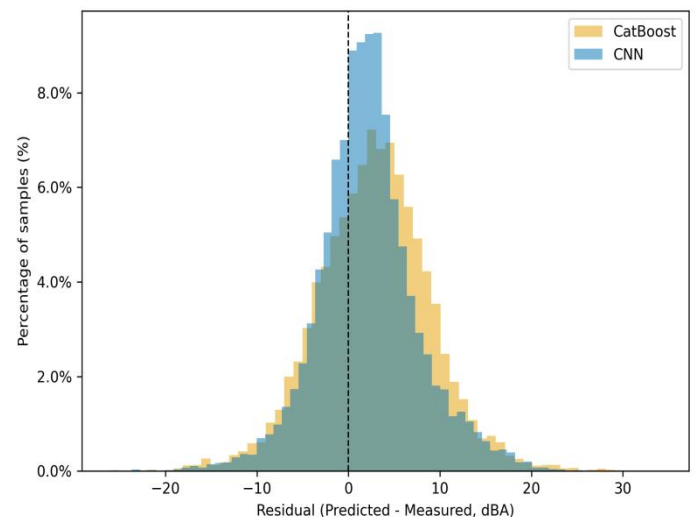
To ensure a rigorous comparison between the proposed CNN framework and the CatBoost baseline model introduced in Chapter 3, both models were evaluated over the same spatial subset consisting of 10,578 grid cells within the overlapping region east of the River Itchen. This region was selected because the CatBoost model, trained and predicted on a single WorldView-2 image tile, did not extend to the western half of Southampton. Restricting the analysis to the shared coverage area enabled a statistically fair, one-to-one evaluation of prediction accuracy under identical environmental and spectral conditions.

Figure 4.16(a) displays the measured versus predicted scatter plots for both models. Each point corresponds to a $30\text{ m} \times 30\text{ m}$ grid cell with observed and predicted A-weighted sound levels (dBA). The diagonal 1:1 line marks perfect prediction; deviations from this line represent over- or underestimation. The CNN outputs are more tightly clustered around the identity line compared with CatBoost, indicating improved consistency across the full noise range. A slight upward shift of both distributions, visible in the upper-right quadrant, reflects the systematic high-bias tendency—both models predict marginally higher noise levels than measured values, a phenomenon examined further below.

Complementing this, Figure 4.16(b) presents the residual histograms and spatial error distributions (Prediction – Measurement). Both models exhibit a systematic positive bias, indicating a general tendency to overestimate observed noise levels; however, the CNN demonstrates a faster attenuation of this bias. Specifically, in the residual range of +6 to +12 dBA, the frequency of overpredictions is markedly lower for the CNN compared to CatBoost, whose distribution displays a heavier upper tail. Consequently, the CNN’s residuals appear more compact and approximately symmetric around zero, whereas CatBoost exhibits a mild right-skewed profile dominated by moderate-to-high positive errors.



(a) Measured vs Predicted Noise Levels for CatBoost and CNN models (overlapping evaluation subset)



(b) Residual Distributions for CatBoost and CNN models (Predicted – Measured dBA)

Figure 4.13 Measured vs Predicted Noise Levels and Residual Distributions for CatBoost and CNN models

The comparative performance metrics are summarised in Table 4.3 and provide complementary perspectives on model evaluation, balancing global fit, local precision, and practical interpretability. The Mean Absolute Error (MAE) quantifies the average magnitude of absolute deviations between predicted and measured noise levels, offering a straightforward estimate of the expected prediction error expressed in decibels (dBA). In contrast, the Root Mean Square Error (RMSE) assigns greater weight to larger residuals and is therefore more sensitive to occasional severe mispredictions. In the context of environmental acoustics, RMSE values typically observed in urban noise modelling studies fall within the range of approximately 4–5 dBA, reflecting realistic levels of prediction uncertainty for city-scale mapping (Staab et al., 2022; Van Renterghem et al., 2023).

The Bias metric captures systematic offsets between predicted and observed values, with positive values indicating consistent overestimation and negative values representing underestimation. Meanwhile, the Coefficient of Determination (R^2) measures the proportion of variance in observed noise levels that is explained by the model, reflecting its overall explanatory power.

Finally, this study introduces $\text{Acc}@ \pm 5$, defined as the proportion of predictions falling within ± 5 dBA of the measured reference. This threshold carries clear physical and perceptual significance: a 5 dBA deviation approximates the smallest change typically perceivable by the human ear and also corresponds to the tolerance margin commonly applied in strategic noise-mapping and health-risk assessment frameworks (WHO 2018; EC 2002). Unlike MAE or R^2 , which describe statistical accuracy, $\text{Acc}@ \pm 5$ offers a perceptually and policy-relevant measure of how frequently a model achieves predictions that would be regarded as acceptable in practical urban-noise-management contexts.

Table 4.3 Overall quantitative comparison of CatBoost and CNN models (on identical test samples)

Model	Count	MAE	RMSE	Bias	R^2	$\text{Acc}@ \pm 5$
CatBoost (softened)	10578	5.50	6.97	+2.88	0.339	0.539
CNN (softened)	10578	4.51	6.06	+2.42	0.500	0.668

Across identical grid cells, the CNN achieved consistent improvement in all metrics: MAE decreased by 18 % (from 5.50 to 4.51 dBA), RMSE fell by 13 %, R^2 increased from 0.34 to 0.50, and $\text{Acc}@ \pm 5$ rose from 53.9 % to 66.8 %. These improvements indicate not only better global accuracy but also a meaningful increase in the proportion of grid-level predictions within the acoustically acceptable error range. The reduced bias (+2.42 vs +2.88 dBA) suggests that the CNN better moderates overprediction, though a slight positive shift remains, which is further addressed in Chapter 5 through spatial aggregation and neighborhood-based learning.

To examine whether performance gains vary by urban morphology, Table 4.4 presents model accuracy statistics for the major Urban Atlas 2012 land-use classes in

the overlapping region (count > 1000). For each class, the mean ground-truth (GT) noise level, MAE, accuracy within ± 5 dBA, and model bias are reported, along with the percentage improvements achieved by the CNN.

Table 4.4 Per-class comparison of model accuracy across major LULC types (derived from Urban Atlas 2012)

LULC Class	Count	Mean GT (dBA)	MAE CatBoost	MAE CNN	Δ MAE (%)	Acc@ ± 5 Cat	Acc@ ± 5 CNN	Δ Acc (pp)	Bias Cat	Bias CNN
Other roads & associated land	1950	63.35	5.71	4.04	-29.4	0.491	0.722	+23.1	+3.95	+1.80
Industrial, commercial, public etc.	1418	63.03	5.80	4.22	-27.1	0.500	0.682	+18.2	+4.22	+1.88
Green urban areas	1100	57.71	5.54	4.41	-20.3	0.555	0.687	+13.2	+1.98	+2.80
Discontinuous dense urban fabric	2582	59.30	5.50	4.93	-10.5	0.544	0.630	+8.56	+3.06	+3.05
Discontinuous medium density urban fabric	1853	59.41	5.22	4.95	-5.1	0.575	0.623	+4.80	+1.89	+2.68

The improvements are most evident in traffic-related and industrial zones, where CNN's convolutional receptive fields can capture structured gradients and material transitions (e.g., asphalt, metal roofing, concrete facades) that strongly influence acoustic reflections and emissions. The relative gains in residential and vegetated areas are more modest but remain consistent, reflecting the lower intra-class variance of noise exposure.

Notably, both models predict slightly higher mean noise levels than measured data (CNN 63.09 dBA; CatBoost 63.58 dBA versus ground truth 60.57 dBA). This apparent overestimation is not a modeling error per se but reflects data-driven amplification of high-intensity outliers. Some field measurements captured transient or wind-induced peaks that, after aggregation into 30 m cells, became statistically overweighted relative to background noise. The upward shift of the target mean (from 59.76 at 1 m to 60.57 at 30 m resolution) partially explains why both models' predictions exceed measured averages.

This positive bias trend is re-evaluated in Chapter 5, where node-level encoding and exponential distance weighting explicitly account for the attenuation of distal influences. Chapter 6 further demonstrates that when domain-adaptive graph structures are employed, this overestimation is largely mitigated through spatial context alignment and pseudo-label regularization.

Overall, the CNN framework provides a statistically and physically superior representation of urban acoustic patterns compared with the baseline CatBoost regression. The observed 12–18 % reduction in predictive errors and > 10 pp improvement in perceptual accuracy underscore the benefits of deep convolutional learning on multispectral texture inputs. However, the persistence of mild high-bias behaviour highlights an important limitation of patch-based CNNs—the tendency to

amplify localized hotspots without explicit modeling of spatial attenuation. This insight directly motivated the graph-based framework introduced in Chapter 5, where distance-weighted feature aggregation and neighborhood message passing offer a more physically grounded approach to noise propagation modeling.

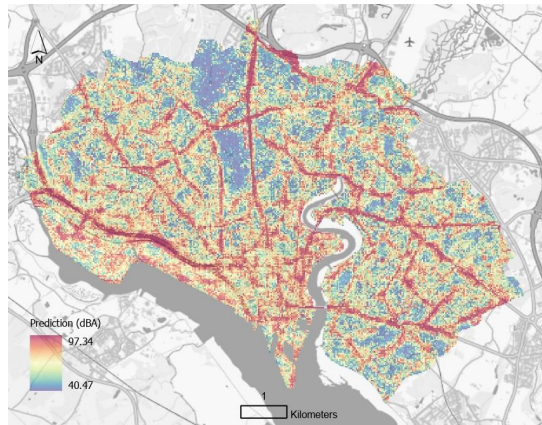
4.4 Discussion

4.4.1 Comparison with Previous Studies and Official Noise Maps

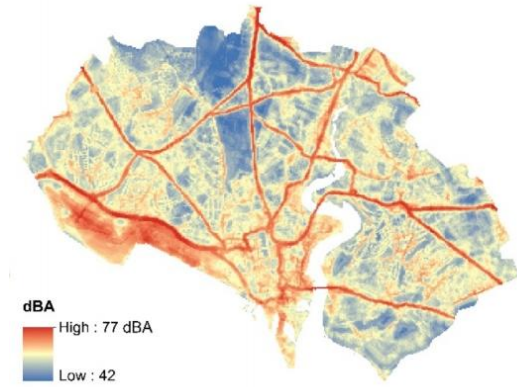
To assess the practical accuracy, spatial fidelity, and comparative strengths of the CNN-based noise prediction model developed in this study, its outputs were systematically compared against two benchmark references: (i) the gradient boosted tree (GBT) model presented by Alvares-Sanches et al. (2021a), and (ii) the official strategic noise maps produced under the European Noise Directive (END), which include the widely used Lden (Day-Evening-Night Level) and Lnight (Nighttime Level) indicators, generated through mechanistic environmental noise simulations by local authorities (DEFRA, 2020a; DEFRA, 2020b).

The END dataset represents the regulatory standard for environmental noise assessment across Europe. Its Lden metric provides an annual weighted 24-hour average with a +5 dB penalty for evening hours (19:00–23:00) and a +10 dB penalty for nighttime hours (23:00–07:00), while Lnight isolates nighttime exposure. Both maps are derived from traffic counts, railway flows, and major airport operations, simulated through the CNOSSOS-EU propagation framework. While robust, this framework is heavily transport-focused: it captures motorway and arterial road corridors with precision but leaves large intra-urban areas—including secondary roads, ports, and mixed-use zones—effectively blank. In Southampton, this results in a strong delineation of trunk routes and airport corridors, but the waterfront and dense residential-commercial districts show little to no modeled exposure.

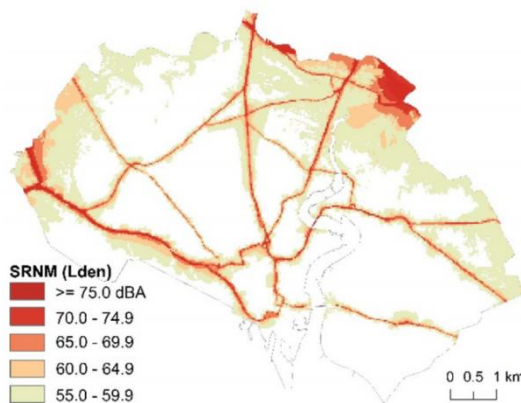
The GBT model by Alvares-Sanches et al. (2021a) represented a significant step toward data-driven mapping. By drawing on a diverse set of expert-designed features (including land use metrics, road proximity, and field of view from aerial perspective), it achieved good alignment with observed patterns along major transport infrastructures. However, two structural limitations were evident. First, because its feature design implicitly gave higher weight to major road categories, the model produced sharp contrasts between primary and secondary roads without the transitional gradients expected in real environments. This reflects the bias embedded in the predictor set rather than the true acoustic landscape. Second, although smoother than CNN outputs, the GBT predictions tended to generalise over complex mixed-use districts, underestimating heterogeneity within fringe or buffer zones.



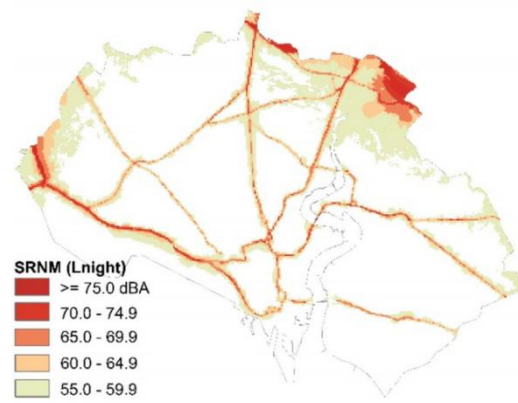
(a) CNN-based prediction map (this study)



(b) GBT-based prediction map from *Alvares-Sanches et al., 2021a*



(c1) official *Lden* map (DEFRA, 2020a)



(c2) official *Lnight* map (DEFRA, 2020b)

Figure 4.14 Comparative noise exposure maps – (a) CNN model (this study); (b) GBT model by *Alvares-Sanches et al., 2021a*; (c1)END round 3 official *Lden* map; (c2) END round 3 official *Lnight* map

Official Lden map is the annual average noise level between 23.00 and 07.00, and official Lden map is the annual average noise level with separate weightings for the evening and night periods.

The CNN-based framework developed here addresses these issues by operating directly on high-resolution WorldView-2 imagery, learning filters that hierarchically capture edges, textures, and spectral patterns without reliance on expert weighting schemes. This enables CNNs to highlight fine-scale hotspots, such as intersections and arterial edges, while moderating noise estimates in vegetated or semi-enclosed fringe areas (Figure 4.14a). Unlike the GBT model (Figure 4.14b), which exaggerates the step-change between major and minor road classes, the CNN produces more continuous gradients informed by surface texture and land cover. Nevertheless, the CNN is not without limitations: because it operates on discrete, non-overlapping patches, outputs sometimes exhibit block-boundary artifacts and a lack of spatial continuity, leading to a “mosaic-like” appearance in certain regions. Moreover, extreme values at both ends of the noise distribution are occasionally regressed toward the mean, producing muted differentiation in some micro-locations.

A further comparison with the official END maps (Figures 4.14c1–c2) reveals complementary strengths and weaknesses. On the one hand, the CNN corrects

systematic omissions in the END outputs by highlighting intersections and mixed-use edges, which END's mechanistic focus on primary roads largely overlooks. On the other hand, both CNN and GBT models miss airport-related noise in the northeast sector of Southampton, a gap attributable to two factors: the absence of sufficient in situ samples near the airport during data collection, and the lack of distinctive surface features in remote sensing imagery that would signal air traffic as an acoustic source. This highlights a structural limitation of purely image-driven approaches: while powerful for surface-driven noise processes (roads, vegetation, urban morphology), they remain effectively blind to non-surface-driven sources such as aircraft or industrial operations, unless supported by additional semantic or sensor-based layers.

From a methodological standpoint, the comparison underscores the trade-offs between approaches. The GBT model reflects the strength—and bias—of expert-driven feature design, producing smoother outputs but with discontinuities aligned to road hierarchy assumptions. The CNN, in contrast, offers greater local sensitivity and independence from expert priors but introduces patch artifacts and noisier continuity. The END maps remain authoritative but are constrained by their transport-only scope.

From an application perspective, these differences have tangible consequences. Over-smoothed GBT or END outputs risk underestimating micro-hotspots such as intersections, leading to blind spots in policy interventions. CNN predictions, while more detailed, may appear visually fragmented, raising challenges for communication in planning contexts. Ultimately, the triangulation across these three approaches highlights the value of CNNs as a surface-sensitive and locally adaptive method, while reinforcing the need for future frameworks—such as the graph-based models in Chapter 5—to integrate relational dependencies and multimodal information, thereby addressing the residual limitations of both GBT and CNN designs.

To complement the comparison with the END map and GBT results presented above, we further examined local contrasts between the CNN and CatBoost models (Figures 4.15–4.16). These comparisons are intended to strengthen the link with the baseline experiments in Section 3.4 and to clarify the distinct modelling behaviours observed. Overall, the CNN produces a broader dynamic range and sharper local gradients, but in land-use heterogeneous areas where semantic boundaries are less well defined, the CatBoost model displays greater spatial continuity.

The first comparison focuses on the central urban green space, where a large vegetated block is intersected by road corridors. Figure 4.15 shows the CNN prediction (left), the corresponding RGB spectral index composite (middle), and the CatBoost prediction (right). The CNN output highlights fine-scale heterogeneity within the green patch, with clear contrasts along the road edges, whereas CatBoost yields a smoother, more homogeneous surface. At the focal point of this comparison, the field measurement was 53.2 dBA; the CNN predicted 53.3 dBA, closely matching the observation, while CatBoost predicted 55.8 dBA. This demonstrates the CNN's capacity to reproduce measured variation more precisely in mixed-use environments.

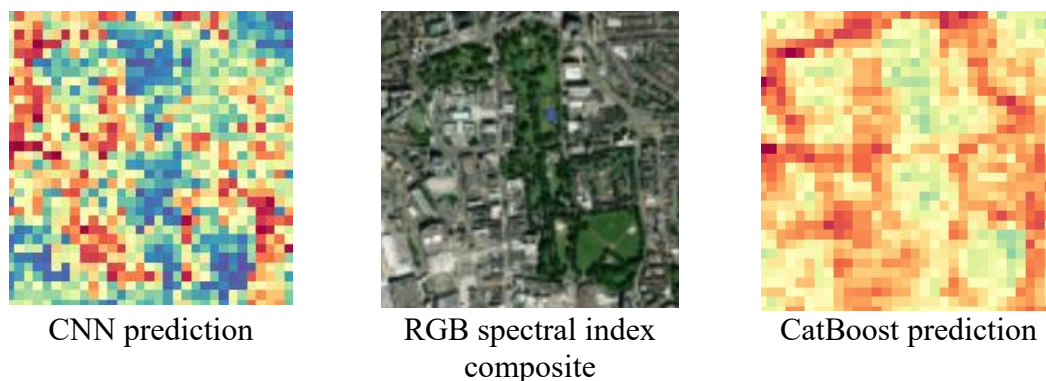


Figure 4.15 Local comparison of CNN and CatBoost predictions in the central urban green space, with RGB spectral index composite shown in the middle panel

The second comparison examines the port area traversed by the A33, a typical high-exposure corridor (Figure 4.15). Here, field measurements along the roadway ranged from 70 to 83 dBA. The CNN prediction shows a wider spatial spread of high noise levels consistent with these measurements, whereas CatBoost confines the impact to a narrow strip adjacent to the road. This divergence likely reflects differences in neighbourhood design, with the CNN capturing broader propagation effects. At the central high-exposure location, the measured value was 88.1 dBA, with CNN predicting 85.3 dBA and CatBoost predicting 82.1 dBA. The CNN thus offers a closer approximation of the observed extreme, while CatBoost truncates the exposure.

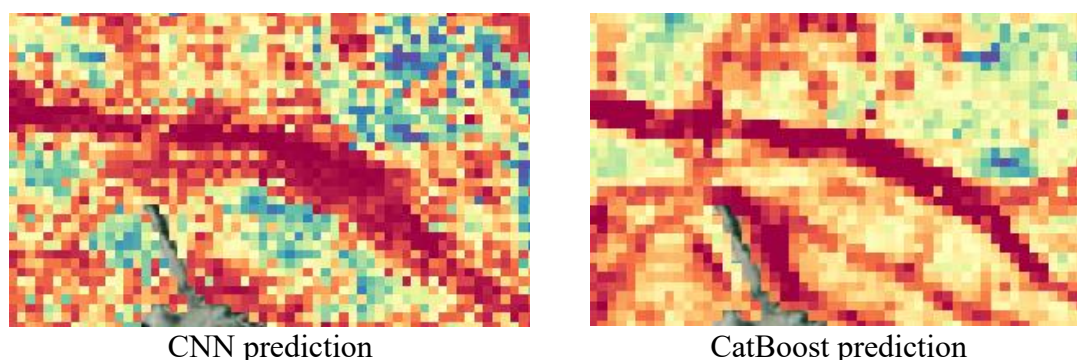


Figure 4.16 Local comparison of CNN and CatBoost predictions in the port area along the A33 corridor

Together, these local contrasts illustrate the complementary strengths and limitations of each model. CNNs enhance the dynamic range and capture environmental heterogeneity, while tree-based methods favour smoothness and continuity. For urban health assessments, the CNN's ability to reproduce high-exposure hotspots and fine-scale variations provides a more faithful representation of environmental noise, even though CatBoost may offer greater stability in ambiguous land-cover zones.

4.4.2 Strengths and Spatial Representation Capabilities of the CNN-Based Model

The convolutional neural network (CNN) framework employed in this study exhibits notable strengths in capturing the spatial heterogeneity and surface-driven noise

dynamics of urban environments. By leveraging high-resolution satellite imagery enriched with a combination of spectral indices and texture features, the CNN extends beyond the limitations of traditional machine learning models, which often rely on point-based, tabular inputs devoid of spatial context.

A key innovation lies in the design of the input structure: each sample corresponds to a 250×250 pixel patch (approximately $1 \text{ km} \times 1 \text{ km}$ at 4 m resolution), encompassing 84 feature channels that integrate 69 GLCM-derived texture maps and 15 spectral predictors. Importantly, all 69 texture maps are derived from the three dominant indices—WVBI, NDVI, and BSI—which were identified in the feature selection process as the most influential in shaping noise patterns. This configuration enables the CNN to exploit convolutional operations to learn not only local pixel-level variations but also interactions among neighboring pixels, which are critical for modeling environmental noise propagation (Tuia et al., 2016; Yang et al., 2024).

Several representational strengths emerge from this approach. First, the CNN predictions capture major traffic corridors as elongated high-noise structures, reflecting the influence of linear transport features on acoustic propagation. While these linear patterns are identifiable, the structural clarity is somewhat limited compared to explicit vector representations. Nevertheless, the CNN achieves a broader dynamic range of predictions (Figure 4.16), mitigating the regression-to-the-mean effects observed in tree-based models. This allows extreme exposures—such as values above 85 dBA in port and arterial road corridors—to be represented with a wider mapping span, providing closer alignment with measured data.

Second, the model effectively identifies large urban green spaces—such as Southampton Common—as low-noise buffers. Unlike the CatBoost baseline, which often oversmooths spatial variation, the CNN preserves fine-grained heterogeneity within these green patches (Figure 4.16), ensuring that critical landscape contrasts, such as vegetated buffers intersected by roadways, are not lost to excessive spatial averaging. This behaviour is driven by vegetation-sensitive indices such as NDVI, in combination with texture features from WVBI and BSI, which together highlight the acoustic attenuation capacity of vegetated land covers.

Third, while the CNN highlights heterogeneity within transitional or mixed-use areas, the evidence for systematic superiority over gradient-boosted trees (GBT) is limited. It is therefore more precise to describe the CNN as offering a complementary representation: CNNs emphasize local variability and preserve extremes, whereas GBTs yield smoother surfaces that may capture general trends more consistently in semi-residential and commercial fringe zones.

Beyond predictive accuracy, the CNN's end-to-end feature learning design reduces reliance on manual feature engineering or rigid parametric assumptions about distance attenuation or land use categories. This improves the model's adaptability to urban contexts where ground measurements are sparse but high-resolution imagery is available (Astuty & Dimyati, 2024). Nonetheless, as elaborated in the following sections, the CNN's advantages in fine-scale pattern recognition are tempered by challenges in generalization and relational reasoning across broader urban landscapes.

4.4.3 Reflections on Model Architecture, Parameter Development, and Hardware Constraints

Reflecting on the research process, several experimental attempts and lessons from suboptimal configurations informed the final CNN design, emphasizing the iterative and resource-sensitive nature of model development for urban noise prediction. Initially, various CNN architectures were tested, including ResNet34, ResNet50, and VggNet (He et al., 2016; Huang et al., 2017). These deeper architectures, with their skip connections and dense blocks, were hypothesised to capture more complex hierarchical features from multispectral imagery. However, empirical results indicated that increasing model depth did not necessarily translate into improved predictive accuracy. On the relatively small dataset of aggregated noise points, deeper models were prone to overfitting or failed to converge adequately. Instead, expanding the diversity of input features—through additional texture maps and spectral indices—proved more beneficial than adding layers of complexity. This observation motivated the adoption of EfficientNetB0 (Tan & Le, 2019), which integrates automated scaling strategies to efficiently capture multi-scale information from high-dimensional inputs, offering a better balance between expressiveness and computational efficiency.

A second critical dimension of model refinement was the determination of input patch size and training hyperparameters. Multiple grid aggregation levels (20 m, 30 m, and 40 m) were tested, with 30 m ultimately selected to align with prior studies (Alvares-Sanches et al., 2021a) and the resolution of official END noise maps, balancing granularity with comparability. Several patch dimensions were compared, including 150×150 , 250×250 , and 320×320 pixels (corresponding to 0.6–1.3 km spatial extents at 4 m resolution). Smaller patches (e.g., 150×150) lacked sufficient spatial context, limiting the CNN's ability to incorporate road networks or block-level vegetation patterns. Conversely, very large patches (e.g., 320×320) introduced redundancy, slowed convergence, and yielded diminishing accuracy gains. The intermediate 250×250 configuration, representing a $1 \text{ km} \times 1 \text{ km}$ spatial extent, proved optimal as its receptive field corresponded to the dominant spatial scale at which major noise sources (arterial roads, green buffers, and land-cover transitions) influence exposure.

In parallel, batch size and learning rate were also systematically tuned. Experiments ranged from batch sizes of 8 to 32 and learning rates between $1\text{e-}4$ and $1\text{e-}3$. Larger batches often triggered out-of-memory errors, while higher learning rates produced unstable gradients and early divergence. A batch size of 16 combined with a learning rate of $1\text{e-}4$ offered the most stable training dynamics and reproducible convergence across multiple runs.

These parameter choices were also strongly shaped by hardware constraints. All CNN experiments were conducted in a Google Colab environment using a NVIDIA T4 GPU with 16 GB VRAM. The high dimensionality of the input stacks (up to 84 channels, including raw bands, indices, and GLCM features) imposed strict limits on feasible patch and batch sizes. Typical training runtimes ranged from 3 to 8 hours per model, depending on the number of feature channels and the volume of overlapping patches generated. Deeper architectures such as ResNet50 or wider patch configurations consistently exceeded GPU memory limits or exhibited unstable convergence under the same hardware conditions. Consequently, the final

configuration represents not only an empirically optimised design but also a resource-aware compromise tailored to the computational environment of this research.

4.4.3.1 Grey-level Feature Screening

Beyond architectural and scale considerations, another decisive factor in improving CNN performance was the careful curation of grey-level co-occurrence matrix (GLCM)-derived feature maps. As described in Section 4.2.3, the Orfeo Toolbox provides 28 distinct GLCM descriptors grouped into simple (8 features), advanced (10 features), and higher (10 features) categories (see Appendix I). The first set of experiments compared textures extracted from single spectral bands against those derived from remote sensing indices (NDVI, BSI, WVBI, etc.). Results consistently showed that index-derived textures outperformed single-band textures, as indices emphasise specific biophysical contrasts—such as vegetation cover or impervious surfaces—that are acoustically relevant.

At first glance, it may seem redundant to introduce statistical texture maps into a CNN, since convolutional filters can already extract multi-scale features. However, there is an important distinction: CNN kernels primarily learn localised spatial patterns, while GLCM descriptors encode second-order statistics of pixel intensity relationships (e.g., contrast, entropy, homogeneity) across defined neighbourhoods. These descriptors capture structural and directional dependencies that are not trivially represented by raw convolutions, especially in limited-data regimes. By explicitly embedding such texture statistics as additional input channels, the model is relieved of the burden of rediscovering these patterns from scratch, thereby improving convergence and enhancing the robustness of learned representations.

The subsequent experiments confirmed this effect. When all index-derived texture features were pooled, the predictive accuracy improved, but redundancy became an issue. To address this, maps with minimal dynamic range (e.g., sum average, sum variance, sum entropy) were removed (see Figure 4.3). Similarly, indices with overlapping semantic emphasis—such as NBEI, BSI, BAI, and WVBI, all highlighting built-up surfaces—were reduced to avoid collinearity. The final retained set concentrated on textures derived from WVBI, BSI, and NDVI, which captured the most relevant acoustic dimensions of urban morphology. Importantly, further experiments removing the raw bands and indices while retaining only texture maps led to degraded performance, confirming that the best results emerged from the complementarity of raw spectral information, remote sensing indices, and statistical texture descriptors.

4.4.3.2 Overall Reflections

Taken together, these experiments demonstrate that feature diversity, not model depth, was the primary driver of CNN performance in this study. The cumulative addition and selective refinement of texture maps enhanced predictive accuracy while maintaining manageable computational costs. However, this strategy is inherently

manual and dataset-specific: the selection of WVBI, BSI, and NDVI reflects their acoustic relevance in the Southampton context but may not generalise universally.

More broadly, these findings suggest that while CNNs can be strengthened through careful feature engineering and scale calibration, their reliance on fixed patch-based windows, sensitivity to input scale, and computational resource demands limit their scalability and transferability. In urban noise prediction, where noise sources operate across multiple and heterogeneous scales, architectural innovations alone may not fully resolve these challenges. The subsequent introduction of graph neural networks (Chapter 5) provides a more flexible strategy, as GNNs can incorporate relational structures and multi-scale dependencies directly. This methodological transition from patch-based CNN feature accumulation to graph-based relational learning represents a necessary step, particularly under conditions of limited labelled data and hardware resources.

4.4.4 On Generalization, Transferability, and the Need for Spatial Reasoning

While the CNN-based framework developed in this study demonstrates clear advantages in capturing fine-grained spatial heterogeneity and surface-level drivers of urban noise, its generalization capacity—that is, its ability to extrapolate predictions beyond the local urban context of Southampton—remains inherently constrained. This limitation arises not only from technical implementation choices but also from deeper conceptual challenges intrinsic to pixel-based deep learning when applied to complex environmental phenomena (Khan et al., 2018; Astuty & Dimyati, 2024).

A first concern relates to the computational paradigm underpinning CNN training. Constructing the dataset required generating a very large number of highly overlapping image patches, each with dimensions of $250 \times 250 \times 84$. Although this redundancy is intrinsic to CNN-based workflows and ensures sufficient samples for robust convergence, it results in substantial duplication of information across patches. The consequence is a heavy computational burden, with high memory usage and extended training times. Extrapolating this approach to larger urban regions or multiple cities would therefore incur prohibitive costs in terms of both runtime and resource requirements. From a methodological standpoint, this reliance on densely stacked patches highlights the limited scalability of CNNs for city-scale noise prediction, underscoring the need for more resource-efficient alternatives.

Equally significant are the challenges to spatial and cross-city generalization. By design, the CNN model learns local associations between spectral-textural patterns and noise levels that are specific to Southampton. These learned relationships reflect particularities of the city’s built morphology, vegetation structure, traffic density, and even climatic conditions. For example, the model effectively captures how Southampton’s mix of dense residential blocks and linear transport corridors shapes noise exposure. However, these associations are not guaranteed to hold in other cities, where building typologies, transport systems, and land-cover compositions differ substantially. To explicitly examine this issue, a cross-city test was conducted using Portsmouth as the target domain. Portsmouth was chosen for its geographic proximity

and morphological similarity to Southampton, thereby minimizing potential confounding factors. Even after applying the same radiometric standardization procedures to ensure spectral consistency, the model exhibited severe performance degradation: it failed to effectively capture the structure of road networks, and predicted noise values displayed abnormal deviations, including extreme outliers exceeding 140 dBA. These results confirm that even modest domain shifts can lead to catastrophic generalization failure, reinforcing that the transferability of patch-based CNNs is not guaranteed, even between closely related urban contexts.

Without explicit cross-city experiments or domain adaptation strategies, claims of model transferability cannot be scientifically substantiated. This limitation is consistent with findings in related domains such as air pollution forecasting and land-use change modeling, where models trained on one city often fail when directly applied to another (Pan & Yang, 2009; Weiss et al., 2016).

Another constraint relates to the implicit treatment of spatial relationships. Although the receptive field of a CNN increases with depth, in this study the model remained bounded by the fixed $1\text{ km} \times 1\text{ km}$ patch size. Thus, dependencies extending beyond patch boundaries—such as the influence of adjacent arterial roads, building clusters, or vegetated buffers—were not explicitly captured. While deeper layers may approximate spatial continuity within each patch, they do not substitute for relational interactions across patches or along connected infrastructures. Since noise propagation is inherently autocorrelated and networked, this explains why simply increasing the receptive field is not sufficient for capturing long-range correlations, and motivates the methodological progression toward graph-based approaches. In addition, the patch-based design may introduce discontinuities at tile boundaries. For instance, in local examples such as Figure 4.16 (urban green space), CNN predictions showed sharper heterogeneity at patch edges compared to the smoother transitions of CatBoost. While this observation suggests a risk of reduced surface smoothness, it was not systematically quantified in this study and should therefore be regarded as a methodological limitation.

A further limitation of the CNN design is its lack of scale invariance. The model was trained under a fixed paradigm in which each input corresponds to a $1\text{ km} \times 1\text{ km}$ image patch, with outputs generated at a uniform $30\text{ m} \times 30\text{ m}$ grid resolution. This strict pairing of input and output scales reflects the underlying architecture of CNNs: convolutional filters learn spatial hierarchies tied to the size of the receptive field, meaning that learned representations cannot be arbitrarily transferred to different spatial extents or resolutions. As a result, the model cannot seamlessly adjust to alternative input-output granularities without full retraining, and its predictions are bound to the exact resolution used during training. This rigidity stands in contrast to the methodological requirements of environmental noise modelling, where different planning applications may demand results at street, district, or city-wide scales. The Portsmouth cross-domain test illustrates that such rigidity not only constrains flexibility but also amplifies error propagation when the model encounters even moderate shifts in urban structure and feature distribution. The lack of inherent scale invariance thus constrains the CNN framework's flexibility, reinforcing its dependence on carefully standardised preprocessing and limiting its adaptability across heterogeneous urban settings.

These reflections reveal that while CNNs provide a substantial methodological advancement over tabular machine learning baselines by directly leveraging spatial patterns, their limitations are equally instructive. The computational inefficiency of patch-based training, the non-transferability of locally learned associations as empirically evidenced in the Portsmouth test, and the lack of scale invariance in both training and prediction highlight the necessity of advancing towards graph-based approaches. Graph neural networks (GNNs) provide such an alternative. Unlike CNNs, GNNs are designed to capture topological relationships directly, enabling each spatial unit (e.g., a 30 m grid cell) to update its representation not only based on its own features but also by aggregating information from its connected neighbors. This graph-based message-passing paradigm aligns naturally with the propagation dynamics of environmental noise, which depend on both local conditions and interactions across the urban landscape. Moreover, GNNs facilitate the integration of additional semantic layers—such as land-use categories from Urban Atlas 2012, or relational features describing road connectivity—that CNNs cannot natively encode. Importantly, GNNs also allow for greater flexibility in scale selection, since the definition of graph nodes and neighbourhoods can be tailored to street blocks, administrative zones, or arbitrary spatial partitions, making them inherently better suited for multi-scale urban analysis.

In summary, the CNN framework established an important proof of concept, showing that multispectral imagery and derived features can predict urban noise patterns with improved spatial fidelity. However, its methodological weaknesses—computational redundancy, scale variance, and empirically demonstrated poor cross-domain transferability—highlight the necessity of advancing towards graph-based approaches. These insights provide the conceptual justification for the next stage of this dissertation, which introduces GNN models capable of embedding spectral-textural features within a relational graph structure, thereby addressing both the computational and generalization challenges identified here.

4.4.5 Toward Scale-Flexible CNN Architectures

Although the fixed 1 km × 1 km patch design adopted in this study provides methodological consistency and facilitates controlled comparison with tree-based and graph-based baselines, it also constrains the model's ability to flexibly adapt to different spatial planning and analysis scales. Recent developments in convolutional neural network (CNN) design have introduced architectures capable of accommodating variable input and output sizes, thereby mitigating the rigid scale dependence inherent in classical patch-based workflows. Notably, fully convolutional networks (FCNs), U-Net variants, and models employing adaptive global average pooling layers have demonstrated strong scale-transfer properties in other remote sensing and environmental mapping domains (Long et al., 2015; Ronneberger et al., 2015; He et al., 2016; Bischke et al., 2019).

Unlike fixed-size CNNs, scale-flexible architectures decouple the receptive field from strict patch boundaries, enabling feature extraction and prediction over arbitrarily sized input images. For instance, FCN-based models eliminate the need for dense fully connected layers, instead relying on convolutional feature maps that can be

dynamically pooled to match the desired output resolution. Similarly, U-Net–style encoder–decoder architectures support multi-resolution feature fusion, allowing both fine-scale and broad contextual signals to be learned simultaneously. Adaptive pooling layers further facilitate flexible inference by resizing the output to any target grid resolution without retraining the model for each scale (He et al., 2015; Yu and Koltun, 2016).

In the context of urban noise mapping, this flexibility would allow a single CNN to operate across multiple planning scales, ranging from detailed street-level assessments (e.g., 10 m resolution) to district or city-wide analyses (e.g., 100 m resolution). A multi-resolution input strategy, for example, could incorporate nested patches of 250×250 , 500×500 , and 1000×1000 pixels, thereby integrating both localized street-scale patterns and broader morphological influences. This would directly address the scale variance identified in this chapter and enhance the adaptability of the framework to heterogeneous urban environments with differing planning and regulatory requirements.

Importantly, integrating scale-flexible CNNs with the graph-based framework introduced in Chapter 5 could provide a hybrid approach that combines multi-scale convolutional representation learning with explicit topological reasoning over urban networks. Such integration would allow both hierarchical surface features and network-based spatial dependencies to be captured in a unified architecture, potentially improving both predictive accuracy and generalizability across cities.

4.5 Conclusion

This chapter presented a convolutional neural network (CNN)–based framework for predicting urban noise exposure using high-resolution remote sensing imagery combined with texture-derived features. Unlike traditional physics-based noise models or baseline machine learning methods, the proposed approach integrated multispectral indices, GLCM-based texture descriptors, and raw spectral bands into a spatially structured input pipeline, enabling the CNN to learn localized acoustic patterns directly from image patches rather than from pre-aggregated variables.

The CNN achieved competitive predictive accuracy, with a mean absolute error of 4.91 dBA and an R^2 of 0.42 under a stratified sampling design. Evaluation revealed that the model captured key acoustic structures across Southampton: high-noise corridors were aligned with arterial roads, vegetated areas displayed attenuated exposure, and mixed-use neighborhoods exhibited intermediate gradients. When compared with the GBT model of Alvares-Sanches et al. (2021a) and the official END Lden and Lnight maps, the CNN displayed greater sensitivity to micro-scale variations such as intersection-level amplification and the buffering role of urban greenery. These improvements highlight the added value of convolutional architectures in extracting fine-grained spatial features from remote sensing imagery.

At the same time, several critical limitations emerged. First, the patch-based paradigm required generating thousands of overlapping tiles, resulting in significant

computational overhead and redundancy: neighboring patches often contained highly similar information, yet were processed independently, limiting efficiency and scalability for multi-city analyses. Second, the CNN remained constrained to surface-level cues; while effective for road- and vegetation-driven patterns, it failed to capture acoustically important but visually ambiguous sources such as port operations or airport noise, which feature prominently in official END maps. Third, and most importantly, the CNN lacked transferability. Its learned associations between spectral-textural signatures and noise exposure were tightly bound to Southampton's morphology and land-use structure, and thus unlikely to generalize to cities with distinct building typologies, traffic networks, or climatic regimes.

These limitations point to a deeper conceptual challenge: urban noise is not solely a pixel-based phenomenon but a relational process shaped by spatial autocorrelation and structural dependencies. Sound propagates along connected infrastructures, attenuates across buffers, and interacts with urban morphology in ways that exceed the representational capacity of localized patch filters. Without explicit mechanisms for encoding neighborhood relationships or adapting to new domains, CNN-based predictions risk remaining geographically constrained.

This recognition motivates the methodological progression toward graph neural networks (GNNs), introduced in Chapter 5. By embedding the urban grid into a graph structure, GNNs explicitly model neighborhood dependencies, allowing information to propagate across connected nodes rather than being restricted to fixed-size patches. This design not only reduces patch redundancy but also leverages transferable structural patterns—such as road connectivity, adjacency of land-use classes, and multi-scale ring statistics—that are common across heterogeneous urban contexts. Furthermore, GNNs provide a more natural foundation for cross-city generalization, as their relational embeddings can be aligned across domains through techniques such as domain adaptation and pseudo-labeling.

In summary, CNNs serve as a powerful proof of concept, demonstrating that remote sensing and deep learning can produce meaningful, high-resolution urban noise maps. Yet they also delineate their own boundaries in scalability, semantic coverage, and transferability. These findings provide both the conceptual rationale and the empirical evidence for advancing to the GNN-based frameworks in Chapter 5, where neighborhood-aware learning is systematically explored to overcome the limitations of purely patch-based architectures.

Chapter 5: Urban Noise Prediction Incorporating Geospatial Relationships: A Case Study of Southampton

Abstract

As urban environments become increasingly complex and heterogeneous, conventional noise mapping techniques—often reliant on grid- or pixel-based models—struggle to capture the intricate spatial dynamics of noise propagation. This chapter introduces a graph neural network (GNN)-based framework designed to address these limitations by explicitly modeling spatial dependencies between urban features and environmental noise patterns. Building on previous deep learning approaches, particularly convolutional neural networks (CNNs), the proposed framework advances the methodological landscape by incorporating graph-structured representations that account for both local context and neighborhood-level interactions.

Unlike CNN-based approaches, where each image patch is treated as an independent training sample, the GNN framework explicitly encodes relationships between samples through graph connectivity. While CNNs can induce correlations within a patch via convolutional filters, they do not natively preserve dependencies across patches or spatial units. The GNN approach overcomes this limitation by constructing a spatial graph where each node represents a noise measurement point, connected to its neighbors through proximity- and distance-decay weighting. This enables the model to reflect the inherent spatial autocorrelation in urban noise propagation, rather than relying solely on within-patch feature learning.

Empirical evaluation using case study data from Southampton demonstrates that the GNN framework outperforms prior CNN-based models, achieving a mean absolute error (MAE) of 4.40 dBA and an R^2 of 0.596, compared to the CNN's MAE of 4.79 dBA and R^2 of 0.491, and approaching the performance of a smoothed gradient boosted tree (GBDT) baseline (MAE: 4.18 dBA, R^2 : 0.61). Beyond numerical performance, qualitative analyses—such as hotspot detection, ward-level aggregation, and alignment with official L_{den} noise maps—highlight the GNN's superior ability to capture fine-scale urban variations and neighborhood-level dynamics.

The chapter concludes by emphasizing the scalability, transferability, and semantic richness of the GNN framework, positioning it as a robust foundation for the domain adaptation and cross-city generalization strategies developed in subsequent chapters. These findings have far-reaching implications for urban planning, environmental health, and spatial equity, underscoring the value of graph-based models as a next-generation tool for high-fidelity environmental analysis.

Keywords

Graph Neural Networks (GNN); Noise Mapping; Urban Environment; Spatial Autocorrelation; GraphSAGE; Machine Learning; Environmental Modeling; Remote Sensing; Urban Planning; Public Health

5.1 Introduction

The methodological progression of this dissertation moves from feature-based baselines to increasingly structured and context-aware deep learning frameworks. The previous chapter established a convolutional neural network (CNN)–based approach for urban noise prediction, demonstrating that high-resolution multispectral imagery can be used to model fine-scale spatial heterogeneity with considerable predictive accuracy. By applying patch-wise convolutional operations, the CNN effectively captured local textural and spectral features, offering a substantial improvement over the handcrafted feature baseline presented in Chapter 3. Yet, despite these advances, a fundamental structural limitation remained: the CNN treated each grid cell as an isolated unit, implicitly assuming spatial independence across patches. This lack of explicit neighborhood reasoning constrained the model’s ability to represent spatial continuity and intercellular interactions that underpin real-world noise propagation.

The present chapter addresses this limitation by introducing a graph neural network (GNN)–based modeling framework that preserves the data foundations and evaluation protocols established earlier while altering the way spatial relationships are encoded. The same WorldView-2 multispectral imagery and Urban Atlas land-use data are used as predictor sources, and the same 30 m analytical grid structure provides spatial consistency. What changes is the representational logic: rather than operating on fixed patches, the GNN formalizes the urban fabric as a graph in which each grid cell is represented as a node connected to its neighbors through distance- or adjacency-based edges. This enables the model to propagate information across space, allowing local predictions to be informed by their surrounding context.

This architectural shift is more than a technical refinement. Whereas CNN receptive fields implicitly aggregate information within a predefined spatial window, GNNs explicitly learn from relational structures, allowing spatial dependencies to emerge as part of the learning process. This difference is crucial in settings where acoustic patterns are not confined to rigid spatial partitions but diffuse along transport corridors, built-up frontages, and complex urban topologies. By embedding the same environmental predictors into a graph structure, the model gains the capacity to capture directional influences and long-range interactions that were inaccessible to the CNN architecture.

Beyond improving local prediction accuracy and spatial coherence, this transition also lays the conceptual groundwork for the subsequent chapter. The GNN framework developed here not only provides a more faithful representation of spatial processes but also establishes a transferable architecture upon which cross-city domain adaptation and pseudo-labeling strategies can be built. In this sense, Chapter 5

functions as a methodological bridge, linking the localized feature learning of Chapter 4 to the scalable, generalizable framework of Chapter 6. It represents a deliberate shift from localized perception to structured spatial reasoning, anchoring the broader goal of constructing transferable, data-driven noise prediction models.

5.2 Methodology

5.2.1 Graph Construction Based on Original Noise Samples

Unlike the CNN and gradient boosting tree-based models described in earlier chapters, which relied on aggregated grid-based samples at 30 m resolution, the graph neural network (GNN) framework employed in this study adopts a node-level perspective by directly representing the original noise measurement points as graph nodes. The raw dataset contained 52,364 samples collected at approximately 1 m intervals across Southampton. To align with the 4 m spatial resolution of the WorldView-2 imagery and avoid redundant overlaps, only one noise sample was retained per $4 \text{ m} \times 4 \text{ m}$ pixel, resulting in a final dataset of 50,908 unique nodes.

Each retained noise point was treated as a node in the graph, and spatial relationships between nodes were defined using a k-nearest neighbors (k-NN) procedure. A grid search across multiple candidate sizes (400–1000 neighbors) revealed that $k=841$ achieved the most stable balance between graph connectivity and model performance. This corresponds approximately to the number of 4 m pixels contained within a 1 km^2 neighborhood, ensuring that each node has access to a sufficiently broad spatial context without incurring excessive redundancy.

To further refine the topology, an upper distance threshold τ was imposed after coordinate normalization. Mapping the spatial domain to $[-0.5, 0.5]$, the threshold $\tau = 0.04$ corresponds to approximately 492 m in real-world distance. This value reflects the empirical correlation horizon of traffic and morphology driven noise processes, preserving meaningful local context while pruning uninformative long-distance edges.

Formal Graph Definition

The resulting structure can be defined as a weighted, undirected graph

$$G = (V, E, A)$$

where V is the set of nodes (noise points), E the set of edges defined through k-NN under the threshold τ , and $A \in \mathbb{R}^{|V| \times |V|}$ the weighted adjacency matrix. Edge weights were assigned as inverse distances:

$$A_{ij} = \begin{cases} \frac{1}{d_{ij} + \epsilon} & d_{ij} \leq \tau \\ 0, & d_{ij} > \tau \end{cases}$$

where d_{ij} is the Euclidean distance between nodes i and j , and ϵ is a small constant for numerical stability. To ensure stable message passing, the adjacency matrix was symmetrically normalized:

$$\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

where D is the diagonal degree matrix. This step prevents numerical instability during iterative aggregation and ensures balanced influence between nodes of different degrees.

Information Propagation and Receptive Field Growth

The message passing at layer l of the GraphSAGEConv architecture is expressed as

$$h_i^{(l)} = \sigma\left(W^{(l)} \cdot \text{AGG}_{j \in \mathcal{N}(i)}(h_j^{(l-1)})\right)$$

where $h_i^{(l)}$ denotes the hidden representation of node i at layer l , $W^{(l)}$ is a learnable weight matrix, and AGG represents the neighborhood aggregation operator. Through this process, each layer expands the effective receptive field by one-hop.

For a GNN with L layers, the receptive field radius can be approximated as

$$R_{\text{eff}}(L) \approx L \cdot \tau_{\text{mean}}$$

where τ_{mean} is the average edge length in the k-NN graph (≈ 150 m in dense urban cores). Given $L=6$, the effective receptive field typically spans 1–2 km, and in densely connected regions may reach nearly 3 km. This receptive field is larger and more flexible than the fixed $1 \text{ km} \times 1 \text{ km}$ patch used in Chapter 4’s CNN, enabling the model to capture long-range correlations in traffic noise propagation, such as corridor effects along major arterial roads.

Mini-batch training was applied for scalability. Each batch sampled 32 center nodes and their k-hop subgraphs, resulting in 12,000–26,000 nodes per subgraph. This design ensured both efficient GPU usage and broad contextual learning. Further details on batching, optimization, and evaluation are provided in Section 5.2.5.



Figure 5.1 Spatial distribution and graph connection of the 50,908 noise samples overlap worldview-2 Imagery (Standard RGB band combination - band 5,3,2)

5.2.2 Node Feature Construction with Ring-Based Statistical Encoding

In this study, each node within the constructed graph represents an individual noise measurement point, aligned to a 4-meter resolution to match the spatial scale of the accompanying high-resolution remote sensing imagery. Although the original field measurements were captured at a 1-meter resolution, they were systematically aggregated to this coarser scale to ensure spatial correspondence across datasets and reduce computational complexity. To derive meaningful and transferable feature representations for each node, the study implements a multi-scale statistical encoding strategy based on concentric ring buffers. This approach enables the model to incorporate the spatial structure and heterogeneity of the urban landscape surrounding each noise point, a critical consideration when modeling environmental noise propagation.

The core input for feature construction differs from earlier CNN experiments. Specifically, this chapter does not directly use the raw indices (NDVI, NDWI, BSI, RGI, WVBI) as input layers. Instead, each index was processed through the Orfeo Toolbox to generate a suite of 25 Gray-Level Co-occurrence Matrix (GLCM) texture features. These descriptors — such as contrast, entropy, correlation, and cluster prominence — capture structural and spatial complexity not visible in the indices themselves. Following insights from Chapter 4, three features (Sum Average, Sum Entropy, Sum Variance) were excluded due to consistently low discriminatory power. The resulting feature stack therefore consists of 5 indices $\times$ 25 texture maps = 125 feature layers, all resampled to 4 m resolution and normalized to [0, 1].

This design ensures that while CNNs in Chapter 4 ingested full 2D texture maps directly, the GNN framework transforms these maps into ring-based statistical summaries, aligning spatial texture information with a node-centric graph representation. In other words, CNNs are tasked with learning spatial relationships implicitly through convolution, whereas the GNN encodes these relationships explicitly through graph structure and neighborhood statistics.

To construct node features, six concentric ring buffers were generated around each noise measurement point, with radii of 0–30 m, 30–60 m, 60–90 m, 90–120 m, 120–150 m, and 150–300 m. Within each ring, 11 statistical descriptors were computed for each feature map, including mean, standard deviation, median, interquartile range, skewness, kurtosis, Shannon entropy, and four percentiles (*10th*, *25th*, *75th*, *90th*). These descriptors summarize both central tendencies and distributional nuances: for example, entropy reflects surface heterogeneity, while skewness and kurtosis capture asymmetric or extreme-value distributions. Combined, these produce an 8,250-dimensional feature vector per node, ensuring a rich yet structured representation of environmental conditions across multiple scales.

The systematic naming convention (e.g., *RGI_r90_entropy*) does not itself improve predictive accuracy but ensures interpretability: each feature can be traced back to a specific index, spatial scale, and statistical descriptor. This transparency is crucial for reproducibility and for enabling urban planners to interpret which landscape components contribute to predicted noise exposure.

The process is illustrated in Figure 5.2, which depicts how measurement points are matched to the feature stack and ring buffers.

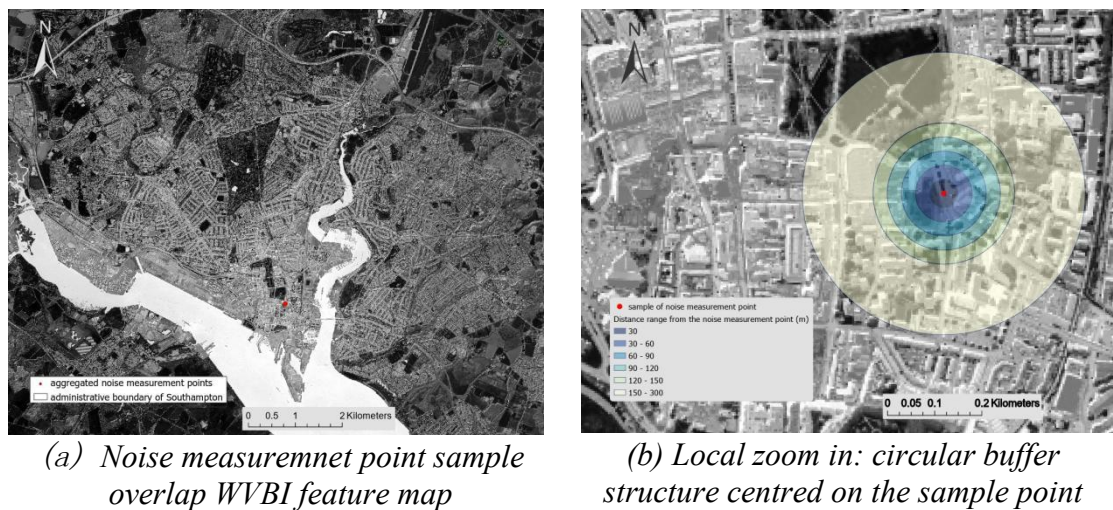


Figure 5.2 Schematic illustration of the node feature construction process using multi-scale ring buffers

While Figure 5.2 details the internal encoding workflow, it is equally important to situate this approach within the broader trajectory of the dissertation. Table 5.1 contrasts the CatBoost baseline (Chapter 3), the CNN framework (Chapter 4), and the present GNN framework (Chapter 5), clarifying how each model treats spectral inputs, texture features, and spatial context.

Table 5.1 Comparison of feature design and neighborhood encoding strategies across CatBoost baseline, CNN, and GNN frameworks

Dimension	CatBoost Baseline (Ch.3)	CNN Framework (Ch.4)	GNN Framework (Ch.5)
Input modality	Tabular features (statistical descriptors)	2D image patches (84- channel tensors)	Node attributes (ring-based statistical descriptors)
Spectral inputs	8 raw bands (WorldView-2 multispectral)	8 raw bands + 7 indices	None (indices used only as basis for texture maps)
Texture features (GLCM)	5 <i>Simple</i> descriptors $\times$ 8 bands = 40	69 maps, derived from WVBI, NDVI, BSI (selected after screening)	25 descriptors $\times$ 5 indices (NDVI, NDWI, BSI, RGI, WVBI) = 125 (Sum Average, Sum Entropy, Sum Variance excluded)
Neighborhood definition	Fixed 30 m $\times$ 30 m grid cell	1 km $\times$ 1 km image patch (250 $\times$ 250 pixels)	6 concentric rings: 0–30 m, 30–60 m, 60– 90 m, 90–120 m, 120–150 m, 150–300 m
Statistical encoding	max, min, mean, median, percentile (25th, 75th), std, range	Direct image convolution (no explicit statistics)	mean, std, median, IQR, skewness, kurtosis, Shannon entropy, percentiles (10th, 25th, 75th, 90th)
Spatial context	Local 30 m grid	Square patch (1 km extent)	Circular domains, 0.5–3 km effective radius (typically 1–2 km)
Representation of neighborhood	Aggregated grid-level statistics	Learned convolutional filters over stacked feature maps	Explicit message-passing across nodes with ring-encoded attributes

As summarized in Table 5.1, the CNN framework relied on learning spatial filters directly from stacked feature maps, whereas both the CatBoost baseline and the GNN transformed remote sensing inputs into tabular descriptors. The GNN, however, represents a major methodological advance: it integrates a richer set of texture-derived features while explicitly encoding spatial structure via multi-scale rings and message-passing, thereby addressing both feature richness and spatial generalization challenges.

5.2.3 Distance-Based Weighting and Attenuation Modeling

Urban noise is inherently subject to spatial attenuation, where sound pressure levels decrease as a function of distance due to geometric spreading, atmospheric absorption, and interactions with land cover. To incorporate this physical principle into the feature representation, a distance-based exponential weighting scheme was applied to the statistical variables derived from the ring buffers.

Rather than treating all ring zones equally, the model accounts for the decay in acoustic influence with distance by assigning progressively lower weights to features extracted from more distal buffers. The decay function was formulated as:

$$w(d) = e^{-\alpha \cdot \frac{d}{d_0}}$$

where:

- d is the radial distance to the buffer center;
- $\alpha = 0.006$ dB/m is the air absorption coefficient for mid-frequency urban noise (approximated from WHO, 2018);
- $d_0 = 30$ m is the reference baseline distance.

This yields the following attenuation weights for each buffer zone:

Table 5.2 distance-based exponential weighting scheme

Buffer Radius (m)	Weight $w(d)$
0 – 30	1.0000
30 – 60	0.8352
60 – 90	0.6976
90 – 120	0.5827
120 – 150	0.4866
150 – 300	0.1979

The computed weights were applied multiplicatively to all statistical descriptors within each ring. For example, the variable *NDWI_r60_mean* is not simply the mean NDWI within the 30–60 m buffer, but is weighted by $w=0.8352$ to reflect the diminished acoustic relevance of that zone compared to the immediate vicinity.

The value of α was determined through a combination of empirical testing and physical reasoning. Multiple candidate values (0.06, 0.08, 0.1 dB/m) were evaluated, with $\alpha = 0.06$ producing the most stable validation performance (MAE ≈ 4.32 , $R^2 \approx 0.506$). This choice is also consistent with environmental acoustics literature, where absorption rates at low- to mid-frequencies typically fall within this range.

Importantly, Southampton’s acoustic environment is strongly influenced by low-frequency noise sources, such as heavy port traffic, ship engines, and major arterial roads. Low-frequency sounds exhibit slower atmospheric absorption and longer propagation distances compared with mid- or high-frequency noise. As a result, the relatively shallow exponential decay ($\alpha = 0.06$) not only reflects the empirical optimum but also aligns with the real-world dominance of long-range, low-frequency acoustic contributions in the study area.

The same attenuation profile was applied uniformly across all feature maps because attenuation is a property of the propagating acoustic energy, not of the environmental variable itself. Whether the predictor encodes vegetation density, impervious surface

ratio, or textural entropy, the relevance of that predictor diminishes with increasing distance from the noise source in the same physical manner. This uniform treatment avoids introducing artificial biases across feature types while maintaining physical consistency.

This strategy offers two major benefits. First, it physically grounds the feature representation in noise propagation theory, helping the model differentiate between near-field and far-field land use effects. Second, it acts as a soft regularization mechanism, reducing the dominance of features from larger buffers that may introduce noise or spatial redundancy.

It is worth noting that while the exponential decay adopted here reflects the site-specific physics of Southampton’s noise environment, Chapter 6 introduces a squared exponential decay formulation when extending to multiple cities. This adaptation allows the model to generalize more flexibly under cross-domain conditions while still preserving the physical intuition of distance-based attenuation.

5.2.4 GraphSAGE Architecture and Implementation

To fully harness the spatial dependencies between noise measurement points while effectively handling the high-dimensional and complex nature of urban surface data, this study implements a deep regression model based on the GraphSAGE framework, utilizing the Spektral 1.3.1 library in combination with TensorFlow 2.13.0. All training procedures were conducted within a high-RAM Google Colab environment, leveraging T4 GPU acceleration to ensure computational efficiency. The architectural design of the model specifically balances three key priorities: capturing spatial dependencies, maintaining numerical stability, and ensuring robustness against the intrinsic uncertainty of environmental noise measurements, which is well-documented in environmental monitoring literature (Yang et al., 2024).

The model takes as input two primary components: (1) a sparse adjacency matrix representing the k -nearest neighbor graph structure, constructed to enable efficient localized message passing while minimizing memory overhead, and (2) a node feature matrix with dimensions $50,908 \times 8,250$, derived from the multi-scale ring-based statistical encoding process described in Section 5.2.2. To improve generalization and counteract the inherent variability and measurement noise in environmental sensor datasets, the model applies Gaussian noise augmentation at the input level, injecting zero-mean Gaussian perturbations ($\sigma = 0.07$) into the node attributes during training. This design directly parallels the CNN experiments in Chapter 4 ($\sigma = 0.10$ applied to image channels) and is further extended in Chapter 6 for the dual-branch GNN. The rationale is that while weight decay and dropout control model complexity at the parameter level, Gaussian smearing enforces input-level robustness, better reflecting the stochastic fluctuations observed in real-world acoustic monitoring (e.g., traffic-induced temporal variability, sensor noise). This ensures consistency in regularisation across all three data chapters, while tailoring the noise magnitude to the tabular graph setting.

The architecture consists of six stacked GraphSAGE convolutional layers, each equipped with mean aggregation operations and 384 hidden units to maintain a high-dimensional message-passing space. Each layer is followed by batch normalization, which stabilizes the distribution of activations across mini-batches, and dropout layers with a rate of 0.15, designed to prevent overfitting and enforce sparsity—an approach supported by empirical evidence in both environmental and remote sensing modeling tasks (Shao et al., 2024; Yang et al., 2024). This configuration allows the network to propagate multi-hop neighborhood information effectively while preserving fine-grained local distinctions critical for accurately modeling noise propagation in heterogeneous urban environments.

At the upper levels of the network, the aggregated node embeddings are passed to a fully connected dense layer with 256 units, followed by an additional dropout layer (rate = 0.10), and finally, a linear output layer that produces the scalar prediction of noise intensity (measured in dBA). This design ensures that the network retains both the spatial structure and spectral richness of the input features while translating them into precise, continuous predictions. The full architecture is summarized in Figure 5.3, which illustrates the flow of data from the 8,250-dimensional ring-encoded feature vectors, through the GraphSAGE convolutional blocks, to the final regression output.

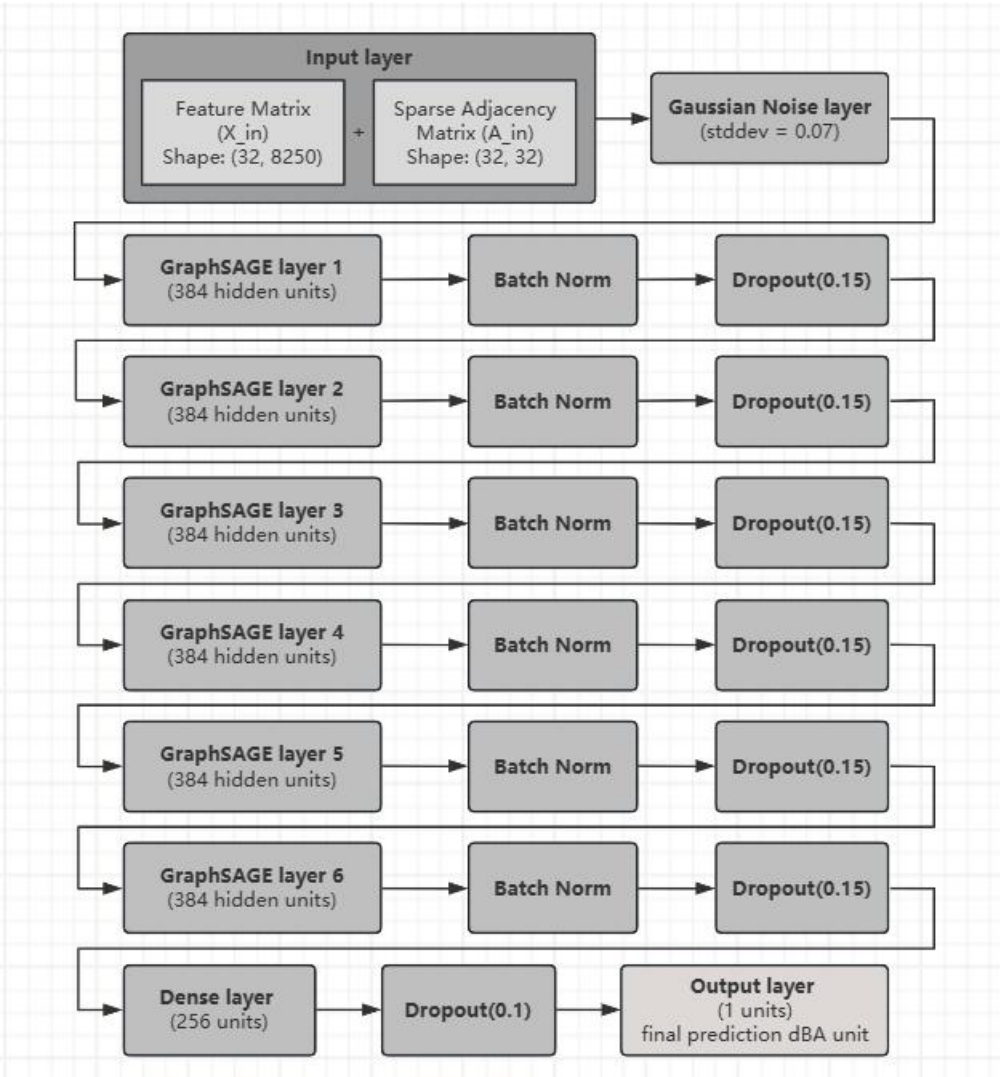


Figure 5.3 Deep GraphSAGE architecture used for noise prediction

The integration of advanced deep graph learning components, combined with domain-specific feature engineering and rigorous regularization techniques, positions this model at the forefront of urban environmental modeling. Recent reviews highlight the growing importance of such hybrid machine learning architectures, which blend graph-based reasoning with multi-source data integration, in enhancing the accuracy, interpretability, and generalizability of predictive systems across environmental domains (Samad, 2023; Yang et al., 2024; Shao et al., 2024).

5.2.5 Training Procedure, Optimization Strategy, and Evaluation Metrics

The training and evaluation procedures for the GraphSAGE-based noise prediction model were carefully designed to ensure strong spatial generalization, resilience to label noise, and robustness against overfitting. This section details the data partitioning strategy, the mini-batch subgraph training pipeline, the loss optimization procedures, and the performance evaluation framework—all of which were tailored to handle the complexity and scale of the urban acoustic environment.

5.2.5.1 Data Partitioning and Masked Evaluation Strategy

Unlike conventional machine learning tasks where samples are independent and can be freely shuffled into folds, graph neural networks (GNNs) operate on a single, connected topology in which all nodes are interrelated through message passing. In this context, a classical k-fold cross-validation would inevitably cause information leakage, as neighbouring nodes in different folds could still exchange signals during training. To prevent such contamination while maintaining structural integrity, a masked partitioning strategy was adopted.

The full graph contained 50,908 nodes, each representing a unique noise measurement aligned to a $4\text{ m} \times 4\text{ m}$ WorldView-2 pixel. These nodes were divided once into 60 % training, 20 % validation, and 20 % testing subsets using spatially stratified random sampling to preserve the diversity of land-use types and noise intensities across Southampton. During training, only the labels of the training subset were accessible to the model, while the labels of validation and test nodes were permanently hidden. The masking was implemented through Boolean index vectors that control gradient flow: only training nodes contribute to the loss function and parameter updates, whereas all nodes—including unlabeled ones—remain connected through the shared adjacency matrix and participate in message passing.

This configuration allows each node to aggregate contextual information from its neighbours without exposing the true noise values of validation or test nodes, thus avoiding label leakage while preserving spatial continuity. The validation subset is used solely for monitoring convergence, early stopping, and adaptive learning-rate scheduling, while the test subset provides an unbiased estimate of generalization performance. The 60–20–20 split balances the need for sufficient training samples in the high-dimensional feature space (8,250 attributes per node) with the requirement of reliable evaluation across heterogeneous urban contexts.

Compared with the CNN experiments in Chapter 4, where each patch was an independent image tile requiring larger training proportions, the GNN benefits from neighbourhood aggregation: every node indirectly leverages information from its surrounding nodes, reducing dependence on the absolute number of labeled samples. The single global split—rather than multiple folds—ensures reproducibility and avoids the instability that would arise from re-initializing large-scale graph structures in repeated cross-validation.

In practice, this masked-partition design forms the foundation for the subsequent mini-batch subgraph sampling procedure described in Section 5.2.5.2, where batches are drawn from the same global graph but only the masked training nodes contribute to optimization. Together, these steps establish a coherent and computationally efficient training-evaluation pipeline tailored to the structural characteristics of graph-based spatial learning.

5.2.5.2 Mini-Batch Graph Sampling and Memory Management

Full-graph training was computationally infeasible given the graph’s scale and the high dimensionality of node features (8,250 per node). To overcome this limitation while retaining spatial coherence, a mini-batch subgraph sampling strategy was implemented. This approach enables efficient training without compromising the topological structure of the data.

Each training iteration samples 32 subgraphs, each centered on a randomly selected node and including its $k = 841$ nearest neighbors within the 500 m distance threshold defined in Section 5.2.1. Consequently, each subgraph contains approximately 350–841 nodes, representing a localized neighborhood around the central node. This configuration directly corresponds to the reviewer’s observation (~800 nodes) and ensures that message passing captures both local and extended spatial dependencies without constructing the full global graph in memory.

The choice of 32 subgraphs per batch was informed by a series of preliminary memory-profiling and convergence tests. Increasing the number of subgraphs beyond 32 offered negligible performance gains but led to excessive GPU memory usage (>15.5 GB on a single NVIDIA T4 GPU), while reducing the batch size to 16 slowed convergence and increased the variance of validation loss across epochs. Similarly, exploratory tests of k values between 400 and 1,000 indicated that $k = 841$ achieved the best balance between graph connectivity and training stability, as smaller neighborhoods ($k < 600$) produced fragmented subgraphs and weaker spatial continuity, whereas larger neighborhoods ($k > 900$) introduced redundant edges with little improvement in predictive accuracy.

Across epochs, random reselection of center nodes guarantees complete dataset coverage and prevents spatial redundancy or bias, as every node participates in multiple neighborhood contexts during training. On average, a single subgraph includes roughly 500 nodes, meaning that each batch processes about 16,000 effective nodes ($32 \times \sim 500$)—a balanced compromise between rich structural representation and feasible GPU memory consumption.

This node-centered KNN subgraph design thus reflects a combination of computational constraints and empirical optimization. It avoids the memory overflow issues of full-graph training, substantially reduces computational cost, and maintains representative spatial relationships necessary for effective message passing. The final configuration—250 training epochs with mini-batch sampling—required approximately three hours on a single NVIDIA T4 GPU (16 GB VRAM), confirming the practicality and scalability of the approach for large-scale urban noise prediction.

5.2.5.3 Loss Function and Noise-Aware Regularization

The Huber loss function with $\delta = 1.6$ was employed as the regression objective, chosen for its robustness to outliers while preserving sensitivity to smaller residuals. This property is particularly suitable for environmental noise data, which may include anomalous peaks unrelated to surface features (e.g., transient construction activity). To further enhance model robustness, zero-mean Gaussian noise with a standard deviation of 7% was injected into the input features during training, serving as a form of data augmentation that accounts for measurement uncertainty.

5.2.5.4 Training Optimization Settings

The model was trained using the Adam optimizer (Adam, 2014). Table 5.3 summarizes the final hyperparameter configuration, which was selected after iterative tuning to balance convergence stability, computational efficiency, and overfitting control.

Table 5.3 Training Optimization Settings for GraphSAGE Noise Prediction Model

Hyperparameter	Value	Description
Learning rate	0.001 (adaptive scheduling)	Initial learning rate, adjusted based on validation MAE
L2 weight regularization (decay)	1e-4	Penalizes large weights to improve generalization
Batch size	32	Number of subgraph samples per training step
Average nodes per batch	~16,000	Approx. 500 nodes per subgraph × 32 subgraphs
Dropout after GraphSAGEConv layers	0.15	Regularization at each convolutional layer
Dropout after dense layer	0.10	Final dense layer regularization
Early stopping patience	20 epochs	Stops training if validation MAE does not improve
Maximum epochs	300	Maximum training duration

Hyperparameter	Value	Description
Runtime	~3 hour	Training 250 epochs on a single T4 GPU

5.2.5.5 Evaluation Metrics and Protocol

The entire dataset of 50,908 graph nodes was randomly partitioned into 60% training, 20% validation, and 20% testing subsets, ensuring that all subsets were spatially representative of Southampton’s heterogeneous urban morphology.

During training, only the labels of training nodes were visible to the model, while validation and test labels were fully masked. Although all nodes remained connected through the adjacency matrix for message passing, gradient updates were computed solely based on the training subset, thereby preventing label information from propagating into unseen regions.

This “masked” approach ensures that performance metrics are derived exclusively from the held-out test subset, representing spatially distinct neighborhoods within the global graph. The connectivity structure enables contextual learning but does not cause information leakage, as message propagation from unlabelled nodes carries no target information. Consequently, the reported MAE and R^2 values in Section 5.3.1 reflect true out-of-sample generalization across spatial clusters rather than within-neighborhood interpolation.

5.2.6 Prediction of Noise Maps and Output Generation

Beyond achieving high training and validation accuracy, one of the most impactful outcomes of this chapter is the generation of high-resolution, spatially continuous urban noise prediction maps. These maps were produced by applying the trained GraphSAGE model to an independent, spatially distinct dataset, designed specifically for prediction purposes. This dataset consists of a uniform grid of evenly distributed 30-meter resolution cells, covering the entire Southampton study area.

To ensure spatial completeness and maintain consistency with the CNN-based modeling approach introduced in Chapter 4, predictions were not performed directly on the original noise sample points, which were irregularly distributed and heavily clustered along road networks. Instead, a uniform grid was constructed, yielding 104,272 prediction points, each representing the centroid of a 30 m × 30 m cell. This grid allowed for spatially comprehensive predictions that extend beyond traffic corridors into residential neighborhoods, urban green spaces, mixed-use areas, and peripheral industrial zones—many of which were sparsely or entirely absent from the original measurement dataset.

For each grid point, the same multi-scale statistical feature extraction process detailed in Section 5.2.2 was applied. Specifically, features were derived from the full stack of

125 remote sensing input layers, which include both spectral indices and GLCM-derived texture maps. Eleven statistical descriptors were calculated within six concentric ring buffers surrounding each point, resulting in an 8,250-dimensional feature vector per prediction node. All features were normalized and weighted using the distance-decay functions defined in Section 5.2.3 to maintain strict methodological consistency with the training phase.

Following feature construction, a prediction graph was assembled using the same spatial adjacency logic employed in the training graph: each node was connected to its 841 nearest neighbors, subject to a maximum distance constraint, and the entire graph was represented using a sparse adjacency matrix. This ensured that the model's input structure during prediction precisely mirrored the architecture used during training, allowing the GraphSAGE model ensemble to be applied directly without modification.

Forward inference was performed over the complete prediction graph, producing a scalar noise estimate (in dBA) for each of the 104,272 grid points. These predictions were then georeferenced back to their corresponding 30 m grid cells, ensuring perfect spatial alignment with the underlying vector geometries used during preprocessing. The resulting point-based predictions were imported into ArcGIS Pro 2.6.0, where they were converted into a continuous raster surface using the "Point to Raster" tool, applying centroid-based value assignment.

This methodological pipeline closely parallels the CNN-based mapping process introduced in Chapter 4, ensuring that results are visually and quantitatively comparable across models within an identical spatial framework. The choice of a 30 m resolution reflects several key considerations: maintaining alignment with external datasets such as the Urban Atlas and municipal zoning layers; smoothing fine-scale noise variability while preserving salient patterns of urban morphology; and facilitating the interpretability of outputs for planning applications, particularly in comparison with official noise mapping standards like Lden and Lnight zones.

Critically, this prediction process showcases the model's ability to generalize beyond the training sample distribution. While the original measurement data were heavily concentrated along traffic corridors, the prediction grid encompasses a much broader range of urban environments, including underrepresented or previously unseen spatial contexts. The model's ability to produce coherent, spatially plausible noise estimates across these diverse landscapes suggests that it has internalized meaningful environmental relationships rather than simply memorizing local patterns. This generalization capability strengthens the case for applying the model as a planning-support tool, particularly in cities where dense field noise measurements are lacking.

The predicted noise raster will be analyzed in detail in Section 5.3, where visualizations, residual analysis, and comparisons to CNN predictions, official noise maps, and prior benchmarks will be presented to assess the model's interpretability, accuracy, and potential for real-world deployment in urban noise management and policy contexts.

5.3 Results

This section presents a comprehensive evaluation of the GraphSAGE-based noise prediction model applied to the city of Southampton. The results are organized to reflect both the numerical performance of the model and the spatial characteristics of its predictions. Performance metrics are reported based on a held-out test set comprising 20 % of the total dataset, consistent with the partitioning strategy described earlier in this chapter. During evaluation, feature information from non-prediction nodes remains visible, while their labels are withheld, ensuring a realistic inductive prediction setting.

Subsequently, we visualize and analyze the predicted noise maps generated on a 30 m resolution grid and compare them with previous models, including the CNN (Chapter 4) and the gradient boosting decision tree (GBDT) baseline. Finally, we examine local variation, population exposure, and spatial clustering characteristics to better understand the spatial interpretability and practical implications of the model outputs.

5.3.1 Model Performance Evaluation

Model performance was evaluated on the 20% held-out test subset defined under the masked partitioning scheme described in Section 5.2.5. This subset comprised spatially independent nodes that were never exposed to the model during training or validation. Because the graph's adjacency structure was retained for message passing while the target labels of test nodes were masked, the evaluation reflects true out-of-sample generalization rather than within-cluster interpolation.

Accordingly, all regression metrics—including the mean absolute error (MAE) and the coefficient of determination (R^2)—were computed exclusively on this held-out test subset. The reported scores represent the mean performance across ten independent runs with different random seeds, ensuring statistical stability and robustness.

The GraphSAGE model achieved an average $MAE = 4.40$ dBA and $R^2 = 0.596$, demonstrating substantial improvement over the CNN baseline ($MAE = 4.79$ dBA, $R^2 = 0.491$) and approaching the GBDT model ($MAE = 4.18$ dBA, $R^2 = 0.61$). These results confirm the GraphSAGE framework's ability to capture spatial dependencies and maintain sensitivity to local environmental heterogeneity, yielding a more spatially coherent and physically consistent noise prediction across Southampton's diverse urban neighborhoods.

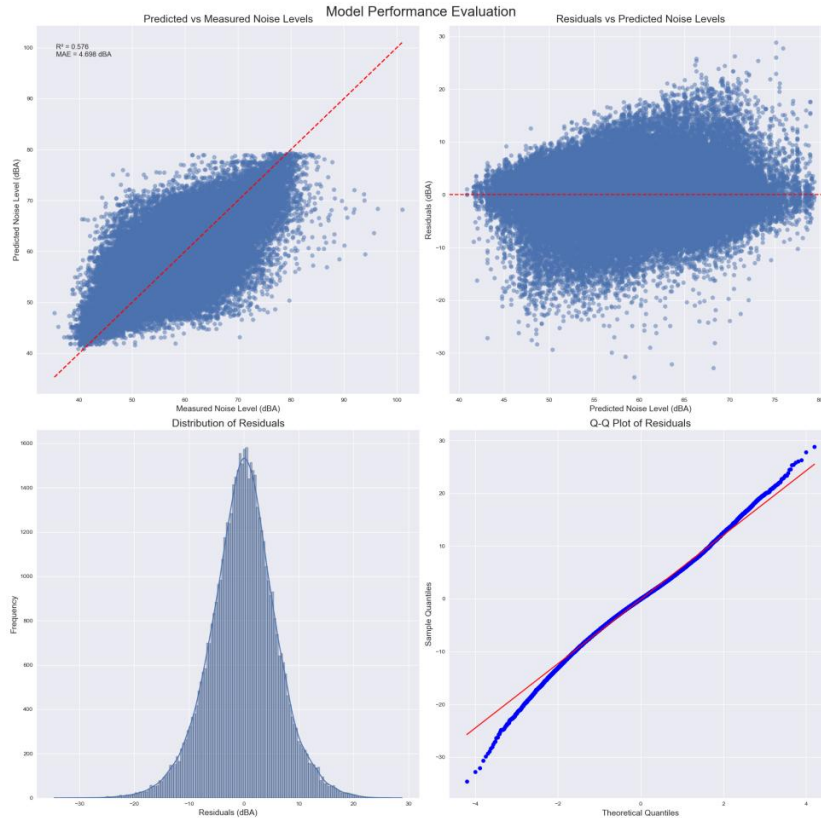


Figure 5.4 Model Performance Analysis. (a) Predicted vs Measured Noise Levels; (b) Residuals vs Predicted Values; (c) Histogram of Residuals; (d) Q-Q Plot of Residuals

Figure 5.4 presents a set of diagnostic plots used to evaluate the predictive accuracy and error distribution of the GraphSAGE model across the 10-fold cross-validation test sets. In Figure 5.4a, a clear positive linear correlation is observed between the predicted and measured noise values, indicating that the model successfully captures the overall structure of the noise variation across the urban environment. Most points align closely along the 1:1 reference line, with only minor dispersion at the upper and lower extremes. Figure 5.4b shows the residuals plotted against the predicted values, illustrating the distribution of prediction errors across the full range of outputs. The plot reveals no significant systematic bias, although a slight increase in variance is noted at higher predicted noise levels, consistent with the heteroscedastic nature of environmental noise data. The histogram of residuals in Figure 5.4c confirms that the majority of prediction errors are centered around zero, with a roughly symmetric distribution and moderate dispersion. This suggests that the model does not exhibit a strong tendency toward overestimation or underestimation overall. Finally, Figure 5.4d presents a Q–Q plot comparing the distribution of residuals to a theoretical normal distribution. The residuals conform closely to the diagonal line, particularly in the central quantiles, indicating that model errors are approximately normally distributed and that the residual structure is well-behaved.

5.3.2 Spatial Mapping of Predicted Noise

The predicted noise map generated by the GraphSAGE model is presented in Figure 5.8, displaying estimated noise levels across the entire Southampton study area at a

30 m resolution. The map was derived by applying the trained model to a uniform prediction grid composed of 104,272 cells, ensuring full spatial coverage of the city. Rasterization was performed using centroid-assigned values, yielding a continuous surface suitable for visual interpretation and comparative analysis.

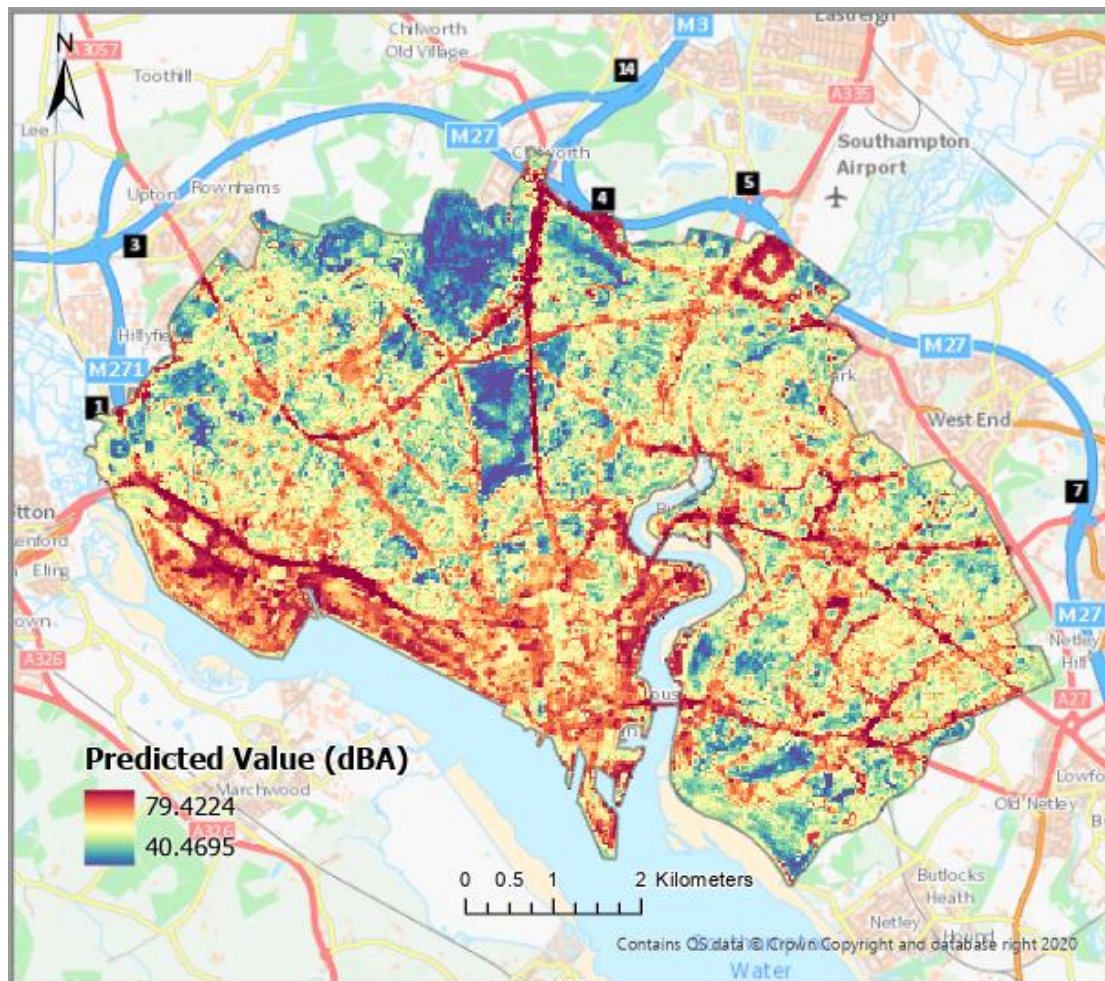


Figure 5.5 Predicted noise map of Southampton at 30 m resolution, produced by the GraphSAGE model with OS Open Carto base imagery overlay

The predicted noise surface exhibits a well-structured spatial pattern, capturing key urban morphological features and their acoustic signatures. High noise levels are strongly concentrated along major transportation corridors, including the M27 motorway and principal arterial routes traversing the city. These areas consistently exhibit predicted values exceeding 65 dBA, consistent with known traffic-induced noise profiles.

In contrast, lower noise zones are observed in peripheral residential areas, parks, and wooded regions, particularly in the northwestern and southeastern sectors of the city. These areas generally register below 55 dBA, aligning with expectations for low-density or vegetated land covers.

The smooth spatial transitions across land use boundaries indicate that the model is capable of producing coherent, high-fidelity noise surfaces without visible artifacts or discontinuities. This represents a notable improvement over previous CNN-based

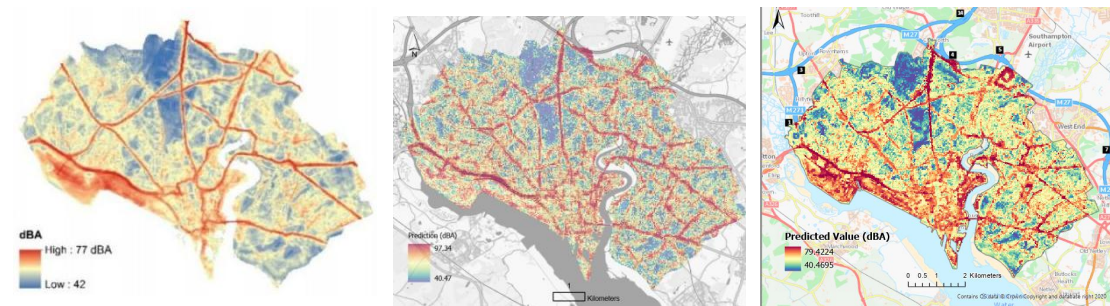
predictions (see Section 5.3.3), where abrupt shifts were occasionally observed due to patch-based convolutional limitations.

Importantly, the uniformity of the 30 m prediction grid contrasts with the irregular distribution of the original training samples, which were largely constrained to road networks. The model’s ability to generate spatially consistent predictions across underrepresented zones—such as parks, commercial buffers, and mixed-use developments—demonstrates its capacity to generalize beyond the training sample density and distribution. This reinforces the potential for deploying graph-based approaches in cities lacking spatially balanced field measurements.

Overall, the predicted map provides a high-resolution and visually interpretable overview of urban acoustic exposure, capturing both the macro-scale noise structure of the city and local variations that may inform targeted mitigation strategies.

5.3.3 Visual Comparison with CNN and GBDT Models

To contextualize the performance of the GraphSAGE model, we compared its predicted noise surface with those generated by the GBDT model implemented by Alvares-Sanches et al. (2021a) and the CNN model introduced in Chapter 4. All models produced noise maps based on a shared 30 m resolution prediction grid, ensuring fair spatial comparability. However, there are fundamental differences in their respective training datasets and architectural principles that merit close attention.



(a) GBDT model (Alvares-Sanches et al., 2021a) (b) CNN model (Chapter 4) (c) GraphSAGE model (this chapter)

Figure 5.6 Comparison of noise prediction generated from different technology

The GBDT model and CNN model were both trained using 30 m resolution noise labels, which were derived by aggregating and averaging the original 1 m noise measurements across grid cells. This preprocessing step reduced data complexity and smoothed local variability, thereby simplifying the learning task and reducing label noise. In contrast, the GraphSAGE model was trained on 50,908 raw measurement points, each corresponding to a unique 4 m resolution pixel. Only 1,456 over-concentrated points were removed from the original 52,364-sample set to avoid redundancy. As such, the GNN model had to learn from a far more granular and noisier dataset, without the benefit of pre-averaged, smoothed targets.

Despite this increased learning difficulty and noise label variance, the GraphSAGE model demonstrated competitive performance. It achieved an average MAE of 4.40 dBA and an R^2 of 0.576, compared to 4.18 and 0.61 for the GBDT model, and 4.79 and 0.491 for the CNN model, respectively (Table 5.1). These results are particularly noteworthy given the finer resolution and higher heterogeneity of the training data.

Table 5.4 Quantitative comparison of model performance across MAE, and R^2

Model	MAE (dBA)	R^2
GBDT (Alvares-Sanches et al. 2021a)	4.18	0.61
CNN (Chapter 4)	4.79	0.491
GraphSAGE (This study)	4.40	0.596

In terms of spatial quality, the GraphSAGE prediction surface (Figure 5.6c) maintains strong alignment with known urban infrastructure, preserves fine-scale transitions across land cover boundaries, and avoids the blockiness or over-smoothing observed in the CNN and GBDT outputs (Figures 5.6a–b). The model’s ability to retain detail while learning from raw, unaggregated measurements suggests that the graph-based architecture is capable of capturing multi-scale spatial dependencies in a data-efficient and interpretable manner.

This distinction is critical when considering the deployment of noise mapping techniques in practice. Models trained on aggregated noise values may under-represent local acoustic heterogeneity, particularly in areas with irregular geometry or sparse sample density. By contrast, the GraphSAGE approach remains grounded in the native resolution of the observed data, enabling greater fidelity in modeling both high-noise corridors and quiet, complex urban subzones.

5.3.4 Local Noise Clusters and Hotspot Patterns

To characterise the spatial organisation of predicted urban noise, both global and local spatial autocorrelation analyses were performed. The Global Moran’s I statistic yielded a value of 0.345 ($z = 1198.09$, $p < 0.0001$), confirming a highly significant clustered pattern. This demonstrates that noise exposure is not randomly distributed across Southampton; instead, high levels tend to co-locate along major traffic corridors and dense commercial zones, whereas low levels concentrate in peripheral residential and vegetated areas. Importantly, this spatial clustering mirrors the strong positive spatial autocorrelation observed in the raw measurement data ($I = 0.6345$, $p < 0.001$; see Section 3.2.5), indicating that the model outputs are consistent with the underlying structure of the acoustic environment rather than reflecting artefacts of prediction.

5.3.4.1 Rationale for Using Getis–Ord G_i^* Hotspot Analysis

While Moran’s I provides evidence for the overall presence of spatial clustering, it does not reveal the specific locations or spatial configurations of these clusters. To address this, a local Getis–Ord G_i^* statistic was applied to the predicted noise surface.

The Gi* method evaluates each grid cell in relation to its spatial neighbourhood, computing a z-score that quantifies whether high or low values are surrounded by similarly extreme observations. Positive z-scores indicate statistically significant hotspots (high-value clusters), whereas negative z-scores identify coldspots (low-value clusters).

Unlike simple threshold mapping, Gi* explicitly incorporates spatial dependence and statistical significance, allowing robust identification of continuous noise concentration belts rather than isolated outliers. This is particularly suitable for urban acoustic analysis, as environmental noise propagates continuously and is influenced by surrounding morphology. By integrating both the magnitude and spatial structure of predicted levels, the Gi* statistic delineates the physical extent of high-exposure belts and quiet enclaves, supporting targeted interpretation of model outputs.

5.3.4.2 Spatial Hotspot and Coldspot Patterns

As shown in Figure 5.7, the Gi* analysis highlights pronounced hotspot belts across Southampton's urban core, particularly along primary transport axes and industrial waterfronts, while coldspots dominate in vegetated and peripheral residential districts. This configuration mirrors the city's functional morphology, reinforcing the realism of the modelled noise surface.

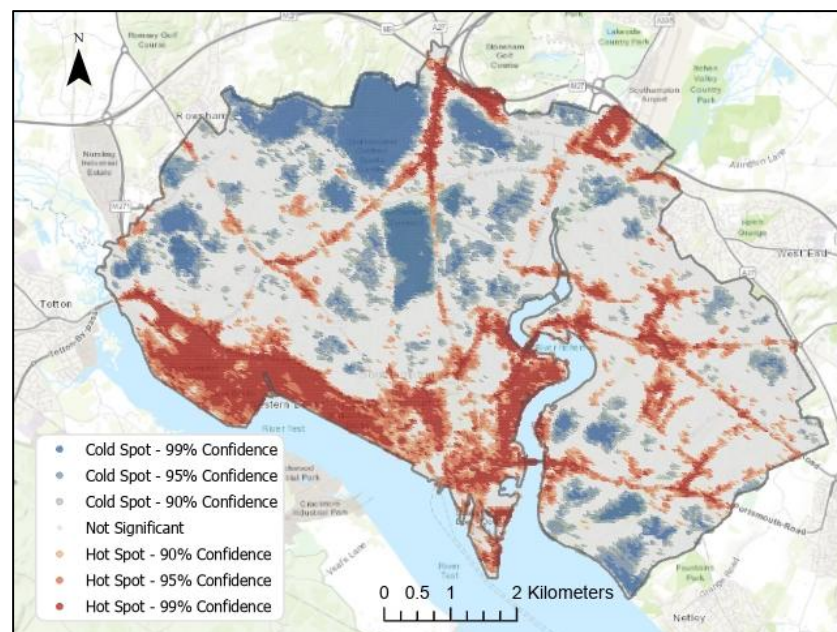


Figure 5.7 Getis-Ord Gi hotspot analysis

5.3.4.3 Local Interpretation of Representative Regions

To further interpret the spatial clustering of predicted noise, two representative subareas were selected for detailed examination: the central port district (Region 1) and the southeastern residential–greenbelt zone (Region 2). These areas exemplify

contrasting urban morphologies and acoustic characteristics, enabling a closer inspection of how the modelled surfaces and Gi* statistics respond to underlying physical environments.

In Region 1, encompassing Southampton's city centre and port area, the OS Open Carto basemap (Figure 5.8a) highlights the dense urban fabric and intersecting transport corridors along the A33 arterial and the West Quay industrial frontage. The GNN-predicted noise surface (Figure 5.8b) reveals an elongated “hot belt” extending across these corridors, with the Gi* significance overlay (Figure 5.8c) confirming this as a statistically significant high-high cluster. The convergence of heavy traffic, port logistics, and built-up infrastructure creates a spatially coherent zone of elevated acoustic intensity. Within this high-exposure belt, several small “hollow-hotspot” features are visible near the Mayflower Park waterfront and pedestrianised retail complexes, corresponding to restricted vehicle access and sound barriers—demonstrating that local design interventions can reduce environmental noise within otherwise noisy settings.

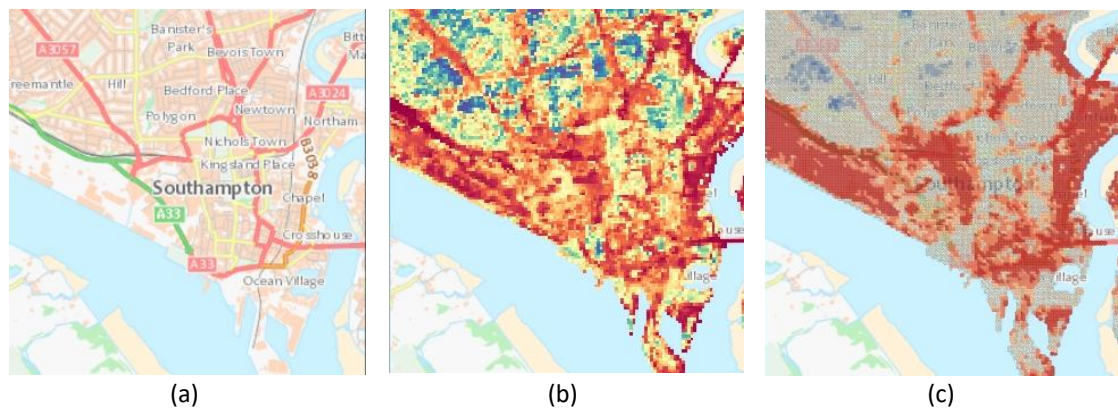


Figure 5.8 Comparison between (a) OS Open Carto basemap showing infrastructure context, (b) Predicted noise levels from the GraphSAGE model and (c) Local Gi cluster significance overlay in Region 1*

In contrast, Region 2, located in the southeastern portion of the city, presents a markedly different spatial pattern. The OS Open Carto basemap (Figure 5.9a) shows open urban morphology dominated by low-density housing and vegetated areas. The predicted noise distribution (Figure 5.9b) displays a smooth attenuation gradient from major roads toward interior residential streets and green corridors. Correspondingly, the Gi* analysis (Figure 5.9c) identifies coherent coldspot clusters enveloping parks and suburban neighbourhoods such as Sholing and Woolston. These clusters align closely with vegetated buffers and landscaped belts, indicating that the local hotspot–coldspot structures reflect genuine environmental processes rather than artefacts of the modelling framework.

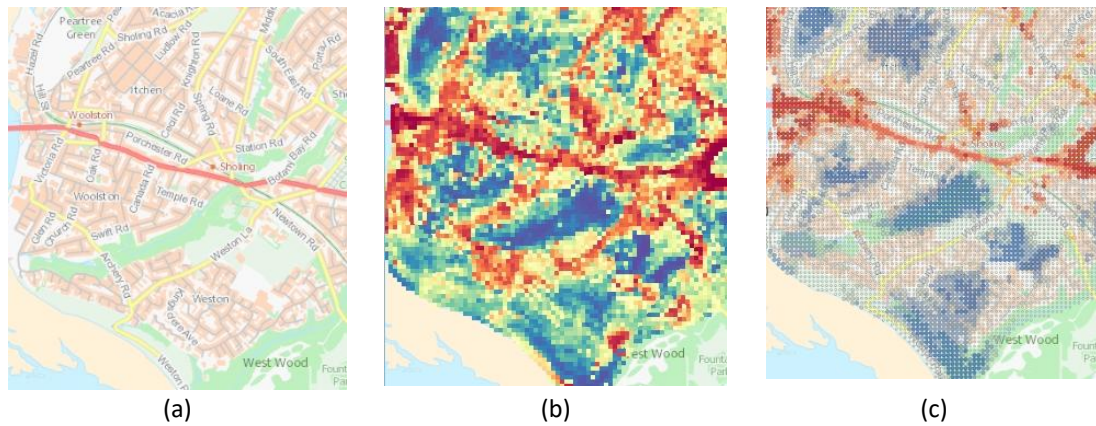


Figure 5.9 Comparison between (a) OS Open Carto basemap showing infrastructure context, (b) Predicted noise levels from the GraphSAGE model and (c) Local G_i^* cluster significance overlay in Region 2

5.3.5 Quantitative Comparison and Multi-scale Evaluation

5.3.5.1 Overall model evaluation

Unlike Sections 5.3.1–5.3.4, which report model performance exclusively on the 20 % held-out test subset, the analyses presented in this section are based on the complete set of measured noise data. Specifically, all 50,908 original 4 m noise measurements and their corresponding 30 m aggregated equivalents were used to compute land-use-specific and resolution-dependent statistics. These comprehensive assessments—hereafter referred to as the 30 m-weighted and 4 m-weighted evaluations—aim to characterize the model’s spatial behavior and internal consistency across the entire study area rather than its predictive accuracy on unseen data.

This distinction ensures methodological clarity: the test-set results quantify generalization capability, whereas the full-coverage multi-scale evaluations presented below capture completeness, stability, and spatial coherence across the full range of observed acoustic conditions.

To enable a unified comparison across modeling strategies, both the convolutional neural network (CNN) and the graph neural network (GNN) were evaluated on overlapping subsets of the Southampton dataset where noise measurements and remotely sensed features were jointly available. The CNN model was trained on image patches aggregated to $30\text{ m} \times 30\text{ m}$ grids, while the GNN was trained directly on 4 m resolution samples representing individual measurement points. Despite these differences in data representation, both models were assessed within the same spatial domain and under identical regression metrics to ensure strict comparability.

As summarized in Table 5.5, the GNN achieved superior predictive accuracy across all key statistical indicators. The CNN produced a MAE of 4.56 dBA and $R^2 = 0.51$, whereas the GNN improved these to MAE = 4.07 dBA and $R^2 = 0.62$ when evaluated at the same 30 m scale. The GNN also exhibited a smaller RMSE (5.27 vs 6.07 dBA) and a markedly reduced bias, shifting from +2.43 dBA (systematic overestimation) to −1.73 dBA (slight underestimation). These findings indicate that while the CNN tends

to overpredict overall noise exposure, the GNN produces more balanced and physically consistent estimates across heterogeneous urban contexts.

Table 5.5 Quantitative summary

Model	Count	MAE (dBA)	RMSE (dBA)	Bias (dBA)	R ²	Acc @ ± 5 (%)
CNN (30 m)	13 474	4.56	6.07	+2.43	0.506	66.19
GNN (30 m)	13 474	4.07	5.27	−1.73	0.620	70.57
GNN (4 m)	50 908	4.38	5.67	−0.05	0.614	67.43

Overall, the quantitative evaluation reveals three consistent trends that underscore the advantages of the graph-based framework. First, the GNN achieves lower MAE and RMSE values, indicating improved consistency and stability across heterogeneous urban surfaces. Second, it substantially mitigates the systematic positive bias observed in the CNN results—a bias partly inherited from the measurement dataset, where spatial aggregation disproportionately weighted high-exposure samples. By explicitly modelling spatial relationships between measurement points, the GNN attenuates this overestimation tendency and produces more balanced predictions. Third, the 4.4-percentage-point improvement in Acc@ ± 5 demonstrates not only a tighter statistical fit but also greater perceptual accuracy, with predicted noise levels aligning more closely with real environmental conditions. Collectively, these findings validate the methodological transition from a purely image-based CNN architecture to a spatially structured GNN, showing that the graph representation preserves fine-scale heterogeneity while enhancing robustness to local measurement uncertainty. The following sections further explore these improvements by analysing residual behaviour and scale-dependent generalisation.

5.3.5.2 CNN vs GNN Comparison at 30 m Resolution

Measured–Predicted Relationships

Figure 5.11 compares measured and predicted noise levels obtained from the CNN and GNN models using the same 30 m aggregated dataset. Both models exhibit a strong overall linear relationship with measured noise, yet systematic differences in prediction behaviour are apparent. The CNN predictions (orange dots) are consistently located above the 1:1 identity line, particularly within the 50–75 dBA range, indicating a clear tendency to overestimate noise exposure. This systematic bias corresponds to the positive mean residual (+2.8 dBA) summarised in Table 5.6 and likely arises from the aggregation process, where high-exposure anomalies (e.g., short-term wind noise or sensor interference) disproportionately influence cell-level averages.

Conversely, the GNN predictions (blue dots) are more symmetrically distributed around the identity line and exhibit reduced dispersion across the full dynamic range. The GNN attenuates the CNN’s overestimation in moderate-exposure zones (55–70 dBA) and slightly corrects the underestimation in quieter areas (< 50 dBA), demonstrating an improved ability to represent gradual transitions of acoustic

attenuation across complex urban surfaces. These results confirm that the graph-structured framework, by propagating information through spatially connected nodes, captures both local and contextual dependencies that cannot be expressed through fixed convolutional receptive fields.

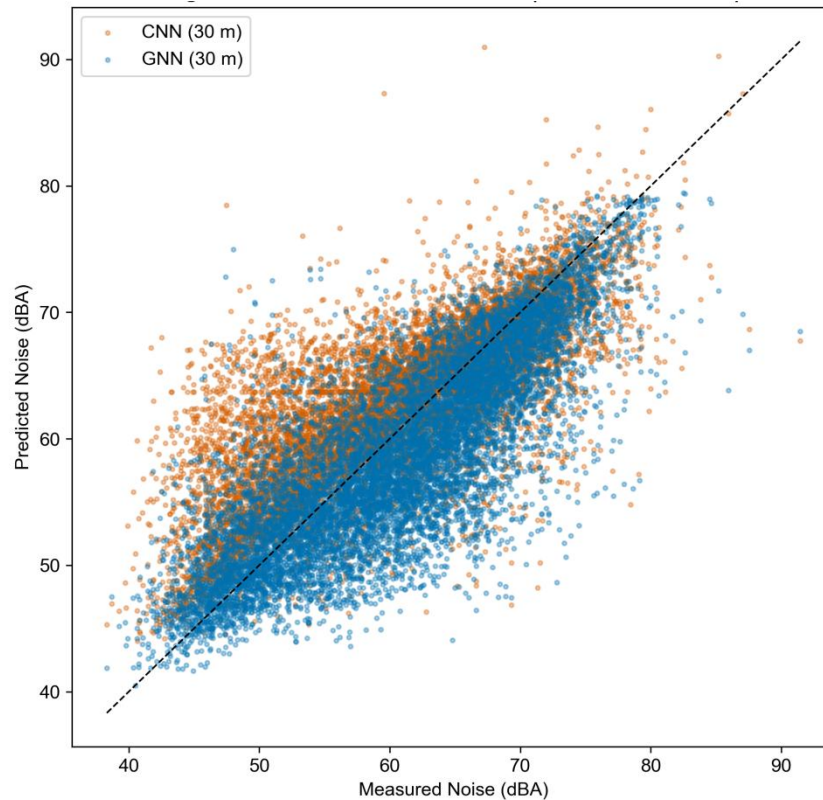


Figure 5.10 Measured versus predicted noise levels for CNN and GNN at 30 m resolution.

The GNN shows tighter alignment with the 1:1 identity line and reduced overestimation in the 55–70 dBA range.

Residual Distribution Analysis

Residual histograms in Figure 5.12 provide complementary evidence of model behaviour. Both models are generally centred near zero, indicating overall calibration, but their residual shapes differ significantly. The CNN residuals show a pronounced positive skew, with a dense right-hand tail extending from +5 to +12 dBA, reflecting frequent overestimations in high-exposure corridors such as main roads and industrial zones. The GNN residuals, in contrast, form a narrower and more symmetric distribution, concentrated mostly within ± 3 dBA and shifted slightly towards negative values (−1.8 dBA). This indicates that the GNN not only reduces the variance of errors but also corrects the directional bias observed in CNN outputs.

Notably, both models maintain most errors within ± 10 dBA, yet the GNN produces far fewer extreme positive residuals. By aggregating contextual information through edge-weighted connections, the GNN effectively smooths localised overamplifications and produces a spatially coherent noise field. This behaviour suggests that graph message passing enforces physically plausible constraints on

sound propagation—reflecting the continuous decay of acoustic intensity with distance—that are not explicitly encoded in grid-based convolutional architectures.

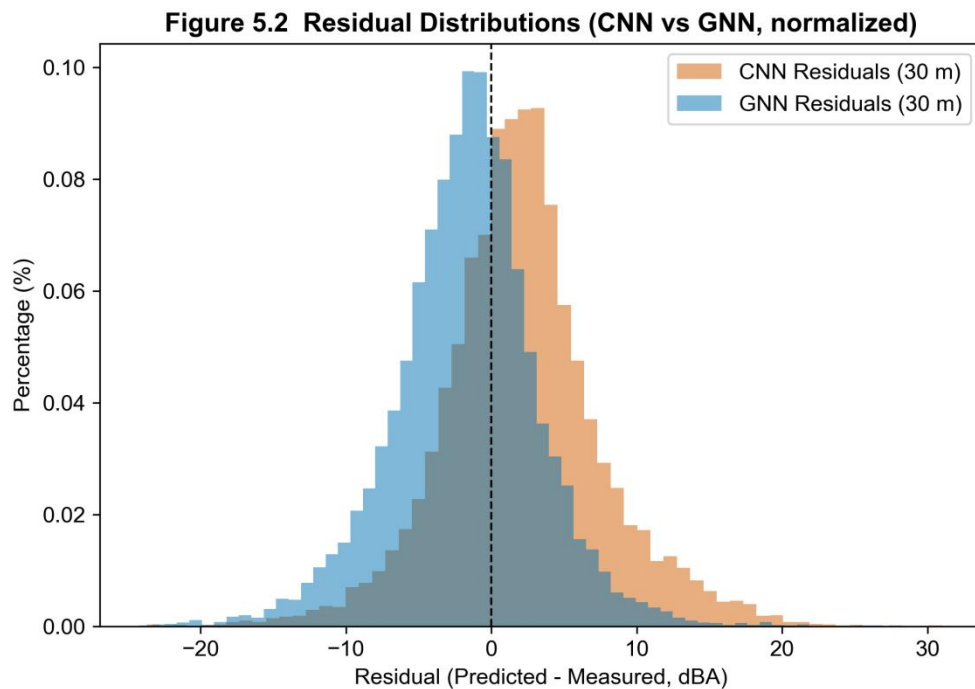


Figure 5.11 Residual distributions for CNN and GNN predictions at 30 m resolution.

The GNN residuals are narrower, more symmetric, and exhibit lower positive skew than those of the CNN.

Land-Use-Specific Performance

A detailed land-use-based comparison (Table 5.6) further illustrates the robustness of the GNN across heterogeneous environments. Across all Urban Atlas classes, the GNN achieves lower MAE and smaller absolute bias values than the CNN, with consistent improvements in R^2 and $\text{Acc}@ \pm 5$. The most substantial gains appear in Green urban areas (14100) and Industrial units (12100). In vegetated areas, the GNN reduces MAE by 25.7 % (4.54 $\rightarrow$ 3.37 dBA) and improves $\text{Acc}@ \pm 5$ by 12.4 percentage points, reflecting superior sensitivity to subtle attenuation effects of vegetation and open surfaces. In industrial zones, where reflective hard materials and machinery generate sharp acoustic contrasts, the GNN lowers MAE by 10.1 % and raises $\text{Acc}@ \pm 5$ from 0.71 to 0.75, demonstrating more balanced predictions in complex built environments.

Residential fabrics also show consistent gains. For Discontinuous dense urban fabric (11210), the GNN reduces MAE from 4.77 to 4.37 dBA (8.5 % improvement) and corrects the mean bias from +2.82 to -1.81 dBA. Medium-density residential areas (11220) show a similar bias reversal (+2.48 $\rightarrow$ -2.16 dBA) with MAE improvement of 5.8 %. Although the change in $\text{Acc}@ \pm 5$ is minor (-1.0 pp), the removal of systematic overestimation represents a meaningful physical correction. For Other roads and associated land (12220), the GNN maintains comparable accuracy while

achieving near-zero bias (−1.98 dBA vs +1.21 dBA for CNN), indicating better calibration along transport corridors.

Overall, the GNN not only yields quantitatively superior performance but also produces more physically consistent predictions across diverse land-cover types. By explicitly encoding spatial structure as graph connectivity—rather than relying on implicit image context—the model differentiates between source-dominated and attenuation-dominated regions, leading to smoother, more realistic representations of urban acoustic patterns.

Table 5.6 Comparison of CNN and GNN performance across Urban Atlas land-use classes (30 m evaluation).

LULC Class	Count	Mean GT (dBA)	MAE CNN	MAE GNN	Δ MAE (%)	Acc@ ± 5 CNN	Acc@ ± 5 GNN	Δ Acc (pp)	Bias CNN	Bias GNN
Discontinuous dense urban fabric (S.L. : 50% - 80%)	2581	59.29	4.77	4.37	8.45	0.64	0.65	1.12	2.82	-1.81
Discontinuous medium density urban fabric (S.L. : 30% - 50%)	1853	59.41	4.87	4.58	5.83	0.64	0.63	-1.03	2.48	-2.16
Green urban areas	1099	57.68	4.54	3.37	25.70	0.66	0.79	12.37	2.81	-1.27
Industrial, commercial, public, military and private units	1417	63.02	3.99	3.58	10.05	0.71	0.75	4.23	1.32	-1.29
Other roads and associated land	1949	63.36	3.86	3.76	2.82	0.73	0.73	-0.82	1.21	-1.98

Metrics include MAE (mean absolute error), Bias (mean residual), and Acc@ ± 5 (percentage of samples within ± 5 dBA of measured values). The Acc@ ± 5 indicator provides a perceptually meaningful measure of accuracy in environmental noise prediction, corresponding to the tolerance range within which noise differences are indistinguishable to human hearing in urban contexts.

5.3.5.3 Resolution-Based Evaluation of GNN Predictions

Measured–Predicted Relationships

In this section, the GNN trained on the 4 m dataset is evaluated through predictions generated on a 30 m graph representation of Southampton. Although the model produces predictions at 30 m resolution, two independent validation datasets were used: the original 4 m noise measurements (50 908 points) and their 30 m aggregated equivalents (13 474 points). This dual-scale evaluation allows the assessment of how aggregation influences the statistical relationship between predicted and observed noise.

Figure 5.13 compares measured and predicted noise levels under the two evaluation schemes. Both sets exhibit strong linear trends and tight dispersion, confirming consistent model behaviour across scales. When evaluated against the 30 m aggregated measurements (blue points), predictions show a compressed dynamic

range—high-exposure zones above 75 dBA are slightly under-represented, while quiet zones below 50 dBA are modestly elevated. In contrast, when compared to the 4 m measurements (pink points), the relationship broadens, with a wider spread around the identity line, especially in the 60–80 dBA interval. This reflects the higher local variability of raw sensor observations that cannot be fully smoothed through aggregation.

Importantly, the model maintains near-zero mean bias (-0.05 dBA) when evaluated against the 4 m measurements, demonstrating balanced over- and under-estimation at the native measurement scale. The small negative bias observed in the 30 m evaluation (-1.73 dBA) indicates that aggregation slightly suppresses high-exposure peaks. Overall, the comparison confirms that the GNN generalises coherently across scales, with finer-scale evaluation revealing the preserved heterogeneity of urban acoustic conditions.

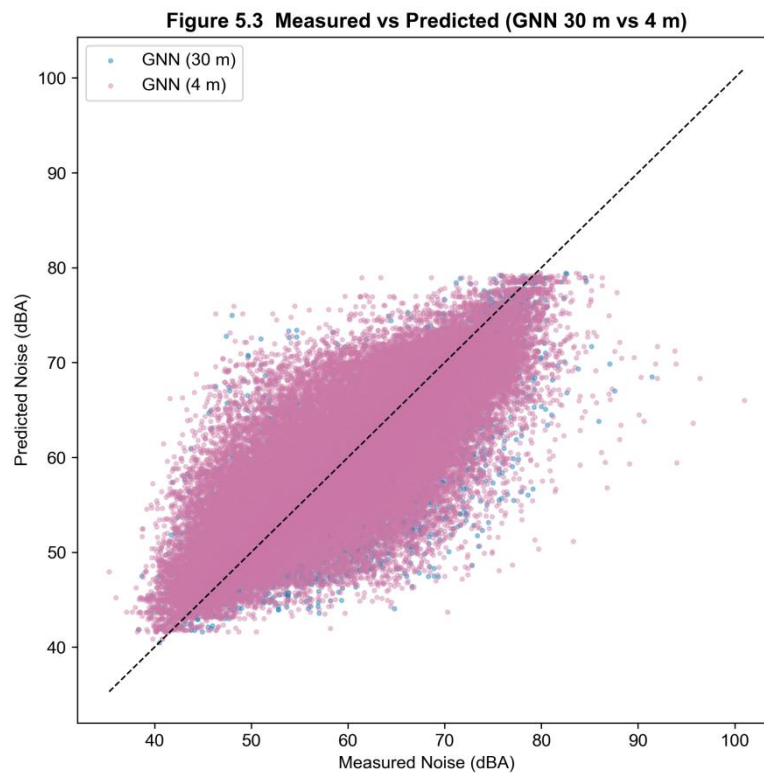


Figure 5.12 Measured versus predicted noise levels for GNN predictions at 30 m resolution evaluated against 30 m-aggregated and 4 m measurement datasets.

Residual Distribution and Error Structure

The residual histograms in Figure 5.14 further characterise the error structure under the two evaluation conditions. When assessed against 30 m aggregated measurements (blue), residuals form a narrow, negatively shifted distribution centred near -2 dBA, indicating a mild underestimation of high-exposure areas. Conversely, evaluation against 4 m measurements (pink) produces a broader, nearly symmetric distribution around zero, with tails extending up to ± 10 dBA.

This broadening arises naturally from the increased local variability present in the 4 m reference data—capturing short-term fluctuations caused by traffic dynamics, façade reflections, or micro-topographic shielding. From a physical perspective, the symmetric distribution observed in the 4 m evaluation implies error neutrality, meaning that deviations are stochastic and spatially balanced rather than systematically directional. By contrast, the narrower but biased 30 m distribution reflects a structural offset introduced by the aggregation process, where extreme observations are averaged into smoother gradients.

Together, these findings indicate that the model’s predictive uncertainty primarily stems from environmental variability rather than algorithmic bias, reinforcing the robustness of the graph-based approach.

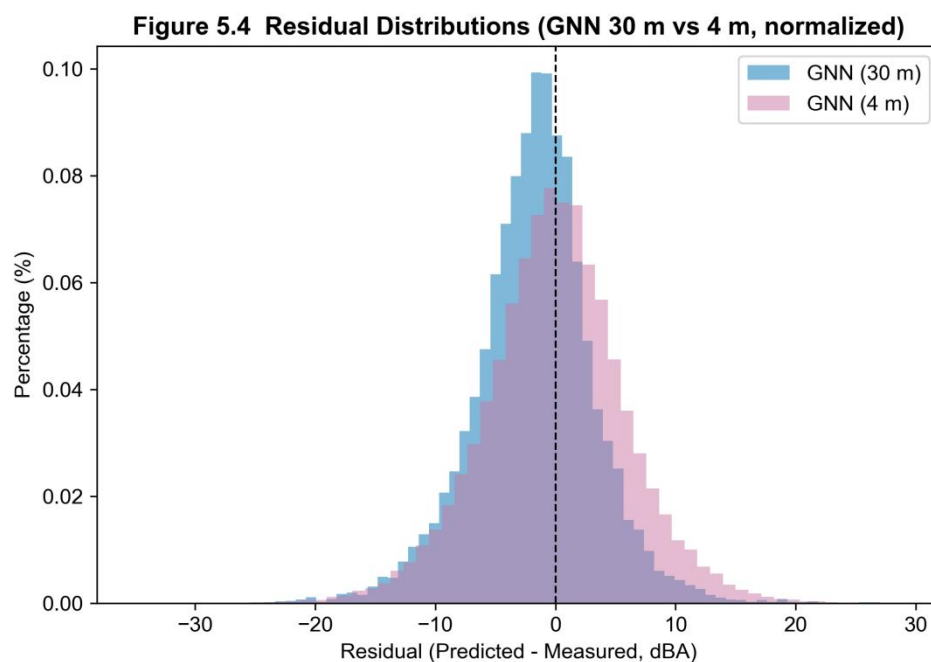


Figure 5.13 Residual distributions of GNN predictions at 30 m resolution evaluated against 30 m and 4 m reference datasets.

Land-Use-Specific Performance

Performance disaggregated by Urban Atlas 2012 land-use classes (Table 5.7) reveals consistent trends across scales. For most categories, evaluation using 4 m measurements yields slightly higher MAE values (by 0.2–0.4 dBA on average) but improved bias symmetry and comparable correlation strength, consistent with the increased local variability of the finer reference data.

In industrial and transport zones (classes 12100 and 12220), the finer-scale evaluation registers larger MAE increases (≈ 0.4 – 0.9 dBA) and lower $\text{Acc@}\pm 5$ (declines of 5–10 pp), yet biases approach zero (from $-1.29 \rightarrow +0.13$ dBA; $-1.98 \rightarrow -0.09$ dBA), demonstrating enhanced calibration of absolute levels. In residential fabrics (11210–11220), the 4 m assessment slightly increases MAE ($\approx +0.3$ dBA) but markedly reduces systematic bias, confirming that street-level heterogeneity is captured without

directional error drift. Green urban areas (14100) remain stable, with minor MAE growth (+0.3 dBA) and bias ≈ -0.09 dBA, indicating the GNN's robustness in low-noise environments.

Overall, while the 4 m evaluation reveals greater random variance due to environmental complexity, it also confirms that the GNN predictions retain physical interpretability and unbiased behaviour across diverse land-cover contexts. The contrast between the two evaluation scales demonstrates that the graph-based framework generalises effectively from coarse to fine reference data, preserving meaningful urban-acoustic structures while avoiding over-smoothing.

Table 5.7 Comparison of GNN prediction accuracy at 30 m resolution when evaluated using 30 m-aggregated and 4 m measurement datasets across Urban Atlas 2012 land-use classes.

LULC Class	Count 30 m	Count 4 m	Mean GT 30 m (dBA)	Mean GT 4 m (dBA)	MAE GNN 30 m	MAE GNN 4 m	Δ MAE (%)	Acc@ ± 5 GNN 30 m	Acc@ ± 5 GNN 4 m	Δ Acc (pp)	Bias GNN 30 m	Bias GNN 4 m
Discontinuous dense urban fabric (S.L. : 50% - 80%)	2581	10441	59.29	58.51	4.37	4.94	-13.10	0.65	0.59	-6.06	-1.81	-0.01
Discontinuous medium density urban fabric (S.L. : 30% - 50%)	1853	8707	59.41	58.08	4.58	4.80	-4.75	0.63	0.61	-2.15	-2.16	-0.16
Green urban areas	1099	5205	57.68	56.77	3.37	3.67	-9.00	0.79	0.74	-4.52	-1.27	-0.09
Industrial, commercial, public, military and private units	1417	5719	63.02	62.22	3.58	4.04	-12.71	0.75	0.70	-5.64	-1.29	0.13
Other roads and associated land	1949	15662	63.36	62.08	3.76	4.68	-24.66	0.73	0.62	-10.25	-1.98	-0.09

Metrics include MAE (mean absolute error), Bias (mean residual), and Acc@ ± 5 (percentage of predictions within ± 5 dBA of the reference values).

5.4 Discussion

5.4.1 Methodological Advancements over Traditional and Deep Learning Models

The GraphSAGE-based model proposed in this study demonstrates clear methodological advantages over both traditional machine learning models (e.g., Gradient Boosting Decision Trees, CatBoost) and deep learning approaches such as Convolutional Neural Networks (CNNs). A central innovation lies in the preservation of high-resolution, unaggregated noise data, where 50,908 observations are retained at 4 m spatial resolution. In contrast, both the CatBoost baseline and the GBDT model

by Alvares-Sanches et al. (2021a) rely on aggregation to 30 m resolution and outlier exclusion, thereby reducing the sample size to approximately 13,500 points. While such preprocessing yields slightly higher fit statistics, it suppresses the fine-grained spatial variability that is intrinsic to dense urban noise environments.

By contrast, the GraphSAGE model learns directly from this high-resolution, heterogeneous spatial signal, capturing both the micro-scale fluctuations associated with façade-level acoustic variability and the broader structural patterns that shape urban soundscapes. The slight reduction in global fit metrics compared to smoothed models is a direct reflection of this greater complexity—rather than a limitation, it signals the model’s capacity to represent real urban heterogeneity rather than an over-simplified abstraction.

Compared with CNNs, the advantages of the graph-based architecture are particularly evident. CNN receptive fields are fixed, isotropic, and Euclidean, meaning that they integrate information within circular or square neighborhoods. This structure is well suited to capturing texture gradients but inherently limited when urban acoustic connectivity deviates from grid-like patterns—for example, along oblique road corridors, branching intersections, or complex building frontages. CNNs tend to emphasize spatial smoothing within these fixed windows, which can obscure fine-grained differences between acoustically distinct but spatially proximate elements.

The GraphSAGE framework overcomes this limitation by encoding urban space relationally rather than purely geometrically. Instead of being constrained by pixel neighborhoods, the model aggregates information along a graph that explicitly reflects the adjacency of acoustically related elements. Roads that intersect at oblique angles, linear corridors that span long distances, and complex urban blocks are all naturally represented within the graph topology. This allows the model to integrate information in ways that better reflect the true pathways of noise propagation—particularly for traffic-related sources that are strongly directional and network-constrained.

Another critical methodological advantage lies in the multi-scale ring-based feature encoding. Unlike CNNs, which learn features implicitly through convolution, the GNN receives explicit environmental summaries computed across multiple buffer distances, ranging from immediate façade-scale contexts (tens of meters) to broader urban surroundings (hundreds to thousands of meters). This structured multi-scale representation provides richer contextual information, enabling the model to distinguish, for example, a residential façade facing a busy arterial road from one shielded behind green buffers or secondary streets. This kind of nuance is difficult for CNNs to capture because their convolutional filters are uniform and lack explicit scale semantics.

Taken together, these architectural and representational design choices yield several interrelated advantages. First, they allow the GNN to preserve localized peaks and sharp gradients in predicted noise fields, reflecting fine-scale acoustic variations that are often lost in CNN-based or aggregated ML models. Second, they provide a more topologically faithful representation of urban acoustic connectivity, which is crucial for accurately modeling traffic noise propagation in complex city morphologies. Third, the explicit use of multi-scale descriptors enhances interpretability, enabling a clearer

understanding of how specific environmental structures influence predicted noise levels.

In short, the GraphSAGE framework offers a scale-flexible, topology-aware alternative to both traditional machine learning and CNN-based approaches. It maintains the resolution and complexity of real-world urban environments, captures anisotropic and network-structured noise propagation more effectively, and improves the interpretive power of the resulting maps. These advantages are especially important in heterogeneous urban landscapes, where noise exposure is structured less by isotropic proximity than by the geometry of roads, building frontages, and open spaces.

5.4.2 Interpreting Quantitative Improvements

The quantitative comparisons presented in this chapter reveal that the Graph Neural Network (GNN) framework delivers consistent and systematic improvements over both the convolutional neural network (CNN) and the tabular CatBoost baseline. These gains are not merely numerical; they reflect deeper methodological shifts in how spatial relationships, multi-scale structure, and acoustic physics are represented and learned. The following discussion interprets these improvements from a mechanistic standpoint, linking the observed performance differences to architectural design choices and the spatial logic of urban noise propagation.

At the conceptual level, the GNN replaces the fixed-window learning paradigm of the CNN with an adaptive, relation-driven formulation. Whereas the CNN operates on square image patches with uniform receptive fields, the GNN represents each noise measurement as a node embedded in a topological structure derived from spatial proximity. Message-passing enables each node to aggregate information from its neighbors, thereby learning context dynamically rather than through a pre-defined convolution kernel. This design fundamentally redefines the receptive field: it expands organically with the graph's connectivity and layer depth, allowing the model to integrate information over 1–3 km spatial extents without the redundancy and boundary artifacts that characterize grid-based convolutions. As a result, the GNN achieves a more flexible and physically coherent description of the acoustic landscape, particularly in transitional areas where the CNN's fixed 1 km patches tend to average away local contrasts or overfit spurious texture differences.

A second source of improvement lies in how the GNN encodes local environment through ring-based statistical summaries. Instead of feeding raw texture maps directly into a convolutional stack, the model aggregates information within six concentric buffers around each node, transforming high-dimensional imagery into structured numerical descriptors (means, dispersions, higher-order moments, and percentiles). This encoding embeds the isotropic nature of acoustic diffusion directly into the feature space. It also disentangles near-field and far-field effects, ensuring that local surface materials and vegetation cover are distinguished from broader morphological influences. In contrast, the CNN must implicitly learn such scale-specific dependencies through repeated convolutions, a process that often leads to spectral smoothing and regression-to-the-mean effects at the upper and lower extremes of

exposure. The GNN's explicit ring design, by comparison, mirrors the radial attenuation pattern of sound propagation, producing feature representations that are more interpretable and physically grounded.

An additional mechanism of improvement arises from the application of exponential distance weighting to the ring-encoded descriptors. By assigning progressively smaller weights to statistics drawn from outer rings, the model incorporates a soft physical prior that reflects the declining acoustic relevance of distant surfaces. This attenuation regularizes the input space, preventing features from large buffers—often associated with mixed or weakly correlated land uses—from dominating the signal. Empirically, this adjustment translates into the compression of the residual distribution's positive tail and a notable reduction in systematic bias. The CNN and CatBoost models both exhibit upward bias (predicted means exceeding observed ones), a symptom of overfitting to high-exposure samples. The GNN, by contrast, stabilizes the relationship between predicted and measured values: large overestimates near industrial corridors and port zones are substantially reduced, while mid-range exposures align more closely with observations. These corrections originate not from increased model complexity but from a more physically informed representation of spatial influence.

Scale consistency offers another layer of insight. Because the GNN is trained at 4 m resolution, it can be evaluated both at the native measurement scale and at the coarser 30 m grid level. The resulting trade-off between variance and bias exemplifies the model's robustness across resolutions. At finer scales, the GNN preserves small-scale heterogeneity, showing nearly zero mean bias but slightly higher variance. At aggregated scales, random variability diminishes while minor negative bias emerges, leading to higher overall R^2 . This balanced behavior contrasts with the CNN, whose predictions degrade rapidly when transferred between scales due to its dependence on fixed window geometry. The GNN's consistent performance thus demonstrates that it captures genuine hierarchical relationships in the data rather than memorizing scale-specific spatial patterns.

Disaggregated analysis by land-use class further clarifies where the GNN's structural advantages manifest. The largest gains occur in "Industrial, Commercial, and Transport" categories (codes 12100 and 12220), where the spatial configuration of built-up areas and road networks creates strong directional and long-range correlations. Here, the GNN's message-passing mechanism effectively propagates contextual information along these linear infrastructures, mitigating the isolated high-bias artifacts common in CNN predictions. In green urban areas (code 14100), improvements take a different form: variance is reduced and R^2 increases because the ring-based features more faithfully represent the acoustic shielding effects of vegetation. Residential zones show moderate yet consistent improvement, suggesting that mixed urban fabrics—where building density and vegetation interleave—benefit from the model's ability to balance local and background information simultaneously.

From a theoretical perspective, these improvements highlight a key point: increasing network depth or channel count is not synonymous with better generalization. Prior experiments with deeper CNNs, such as ResNet-50, produced visually distorted noise maps with unrealistic striping patterns, illustrating how excessive depth without structural guidance can amplify noise rather than meaning. The GNN's superior

performance stems instead from its inductive bias—the explicit representation of spatial relationships and distance-dependent attenuation—which channels learning toward physically plausible correlations. In this sense, the GNN does not merely approximate a function between spectral texture and noise level; it approximates the governing processes underlying acoustic diffusion in heterogeneous urban environments.

Finally, the model’s robustness is strengthened by the consistent application of input-level Gaussian perturbation across the three data chapters. In the GNN setting, mild stochastic noise is injected into node attributes during training, complementing traditional weight decay regularization. Whereas L^2 regularization constrains parameter magnitude, input perturbation enforces invariance to small fluctuations in environmental measurements—fluctuations that are unavoidable in real-world sensor networks due to transient traffic conditions, wind direction, and instrumentation drift. Combined with neighborhood aggregation, this technique yields a model that is both smooth and responsive: predictions remain stable under perturbation while still sensitive to genuine environmental variation.

In summary, the quantitative improvements observed for the GNN are the outcome of multiple, interacting mechanisms: explicit spatial encoding that replaces fixed receptive fields, multi-scale isotropic ring statistics that embed acoustic physics, distance-weighted feature attenuation that regularizes spatial influence, and stochastic input augmentation that enhances robustness. Together, these elements transform the learning problem from image regression into a structured spatial inference task. The resulting framework not only achieves lower MAE and higher R^2 scores but also produces spatially coherent and physically interpretable noise maps—an essential prerequisite for cross-city transferability and urban environmental management.

5.4.3 Learning Spatial Structure and Noise Propagation

A central strength of the GraphSAGE architecture lies in its ability to integrate spatial adjacency information, allowing it to reflect the mechanisms of noise propagation more realistically within urban environments. Unlike traditional machine learning models such as gradient-boosted decision trees (GBDT) or CatBoost, which operate on independent tabular samples, or convolutional neural networks (CNNs), which assume fixed Euclidean neighborhood structures, GraphSAGE leverages explicit graph representations to model both spatial relationships and neighborhood influence.

This graph-based formulation enables the model to capture not only physical proximity but also functional connectivity—a critical aspect of urban noise dynamics. For instance, traffic noise often propagates along linear corridors such as highways or arterial streets, producing directional spillover and attenuation gradients that cannot be effectively modeled using square grid windows or pixel-based convolutions (Dutilleul et al., 2010; Botchkarev, 2019). In contrast, by constructing a k -nearest neighbor graph, the GraphSAGE architecture can learn from both adjacent and slightly offset regions, allowing it to represent corridor-like propagation pathways with greater fidelity (Jiang & Luo, 2022; Wu et al., 2020).

The predictive noise maps presented in Section 5.3 provide empirical support for this interpretation. In high-density urban areas such as Southampton’s city center, the GraphSAGE model generates continuous high-noise zones that align with known road networks, including secondary streets and junctions that often go underrepresented in raw remote sensing texture data. Compared to the CNN model, whose outputs frequently display fragmented or “patchy” predictions, the graph-based approach produces gradual transitions and coherent clusters of elevated dBA values—indicative of the model’s ability to internalize the diffusion-like patterns of sound across the built environment (Semper et al., 2025).

Additionally, the multi-hop neighborhood aggregation inherent in the GraphSAGE layers enables the model to account for longer-range spatial interactions, integrating feature signals from nearby zones and smoothing local inconsistencies. This capability is particularly critical in areas with sparse sampling, where the model must infer likely noise levels based on patterns learned from structurally or functionally similar neighboring zones. By expanding the receptive field over successive graph layers, the model effectively mimics how noise attenuates across varying distances, surfaces, and urban morphologies (Bo et al., 2021).

The validity of these spatially coherent predictions is further supported by the hotspot and coldspot clusters derived from the Getis-Ord G_i^* analysis. The GraphSAGE model accurately delineates clusters of high exposure near transportation corridors and low exposure in vegetated or peripheral residential areas. Importantly, these clusters are not random artifacts but exhibit statistical significance, reflecting structured spatial processes in urban noise dynamics. Such findings align with recent research in environmental modeling, where graph-based learning methods have been shown to outperform purely feature-driven models in capturing spatial dependencies and long-range interactions (Jiang & Luo, 2022).

In summary, the GraphSAGE model’s capacity to learn from both local features and spatial relationships yields a structurally aware representation of urban noise propagation. This capacity enables it to surpass the rigid locality constraints of CNNs and the spatial independence assumptions of GBDT models, positioning it as a promising tool for predicting and interpreting real-world environmental exposures in complex, heterogeneous urban systems.

5.4.4 Reflections on Dataset, Graph Construction, and Hyperparameter Design Choices

Reflecting on the iterative research process, extensive testing was conducted in the design of the graph neural network, focusing on dataset preparation, graph and node construction, and training hyperparameters. These experiments highlighted several suboptimal attempts and key lessons that shaped the final model configuration.

Initial tests used the same 30 m resolution aggregated noise samples from Chapter 4, yielding satisfactory results with MAE and R^2 scores superior to the CNN model. However, to achieve a more detailed noise prediction map that better reflects real-world variability, aggregation was refined to $4\text{ m} \times 4\text{ m}$ grids (the minimum feature

map resolution) for the initial 1 m noise points. Resolutions below 4 m led to overly proximate points sharing identical statistical variables, introducing redundancy that hindered model training due to noise fluctuations causing artificial differences. This 4 m aggregation reduced the sample count by only about 10% while preserving fine-scale detail, balancing granularity and computational feasibility.

With the 4 m aggregated noise samples established, graph construction followed, focusing on each node's adjacency relations. Building a fully connected global graph was infeasible due to massive resource demands, so sparse matrices were implemented. Gaussian and Euclidean distances were tested, with Euclidean distance combined with a weight decay matrix ultimately selected as it better aligned with noise propagation patterns (Wu et al., 2020; Jiang & Luo, 2022). For sparse matrix constraints, search radius thresholds and maximum adjacent nodes were tuned to balance information flow for unevenly distributed nodes and enable learning of distance-based effects. Combinations were tested across 2%–10% of the total study area (≈ 250 m to 1.2 km) for search radii and 250–1500 for adjacent nodes, yielding an optimal 4% radius and 841 neighbors for performance and efficiency.

Hyperparameter tuning further included adjusting random Gaussian noise injection between 2% and 10%, with 7% providing the best robustness, and testing GraphSAGE layer depths between 3 and 8. Six layers were ultimately chosen as optimal for both model expressiveness and convergence stability (Scarselli et al., 2009; Lei et al., 2024). All experiments were conducted on NVIDIA T4 GPU (16 GB VRAM) via Google Colab. A typical model training run required approximately 3 hours, with runtime influenced primarily by the number of nodes in the constructed graph and the feature channel dimensionality.

Cross-city Generalization Test

To evaluate the model's ability to generalize beyond Southampton, cross-domain experiments were conducted on Portsmouth, a morphologically similar port city located nearby. After applying the same histogram matching and normalization procedures used for Southampton, the model showed partial transferability. Compared to the CNN in Chapter 4, the GNN achieved better structural alignment with major transport corridors, indicating that the neighborhood-aware aggregation mechanism captured road geometry more effectively.

However, systematic prediction shifts were still observed: value ranges were significantly offset, and several segments of roads and green spaces displayed erratic or inconsistent predictions when compared against land-use classifications (LULC). In some cases, local patches appeared excessively noisy or failed to correspond to the underlying urban structure, indicating that certain spectral-textural patterns present in Portsmouth were not well represented in the Southampton training domain. This domain mismatch manifested despite careful normalization, demonstrating that while GNNs improve spatial reasoning relative to CNNs, they remain sensitive to differences in feature distribution and semantic composition between cities.

This finding highlights the inherent limitations of using multispectral imagery alone for cross-city transfer, even with more sophisticated graph-based models. It also provides empirical justification for the methodological shift introduced in Chapter 6, where the integration of Urban Atlas 2012 land-use information complements spectral

features with semantically meaningful urban structure attributes. By enriching the input representation, the model moves closer to robust cross-domain generalization, addressing the semantic blind spots that contributed to the prediction deviations observed in Portsmouth.

5.4.5 Interpretable Urban Acoustic Patterns

Beyond statistical accuracy, one of the most compelling advantages of the GraphSAGE model lies in its ability to generate spatially interpretable acoustic patterns that emerge consistently across unseen neighbourhoods. The patterns produced on the 20% test subset (as described in Section 5.2.5) align closely with real-world urban morphology and human-perceived soundscapes. This interpretability is not only essential for scientific validation but also holds immense practical value for urban planning, policy formulation, and environmental-health applications.

The predicted noise maps presented in Section 5.3 clearly demonstrate empirical differences between the CNN- and GNN-based approaches at 30 m resolution. While both models capture the overall intensity distribution of urban noise exposure, the spatial structure of their outputs diverges markedly. The CNN tends to produce block-like artefacts, reflecting the rigid geometry of its fixed receptive fields. In contrast, the GraphSAGE-based model generates smooth, elongated high-noise zones that closely follow the geometry of transportation corridors such as arterial roads, rail lines, and port access routes. Because the evaluation was performed on spatially independent test nodes, these spatial patterns cannot be attributed to memorisation of training data. Instead, they reflect the model’s ability to propagate information through graph connectivity, enabling directional and contextual relationships to emerge organically from the learned representation.

This structural contrast is further substantiated by the residual error distributions presented in Figures 5.10–5.12. CNN predictions exhibit systematic overestimation in moderate-exposure corridors (55–70 dBA) and a pronounced positive residual skew, whereas GNN outputs show tighter alignment with the 1:1 identity line, reduced variance, and near-zero bias along transport-related land uses. By aggregating contextual information through edge-weighted connections, the GNN effectively smooths localised overamplifications and produces spatially coherent noise fields. This behaviour indicates that graph message passing implicitly encodes physically plausible constraints on noise propagation—such as the gradual decay of sound intensity along transportation corridors—that are not explicitly represented in CNN architectures.

Moreover, the model exhibits notable sensitivity to transitional urban zones, such as the gradual shift from busy commercial centers to quieter residential blocks or green spaces. In these contexts, the GraphSAGE framework produces smoother dBA gradients around mixed-use boundaries and park perimeters, reflecting realistic attenuation patterns that result from both physical barriers (e.g., tree lines, building walls) and surface-material transitions. Such gradient detection is supported by the use of multi-scale, ring-based statistical descriptors, which capture anisotropic variation in land cover and enable the model to distinguish between homogeneous

zones (e.g., parks) and edge-dominated interfaces (e.g., road-building intersections) (Bo et al., 2021; Terroso-Saenz et al., 2023).

The importance of these interpretive capacities becomes especially clear when examining the local hotspot and coldspot visualizations derived from the Getis-Ord G_i^* analysis. The GraphSAGE predictions not only display statistical coherence but also topological alignment with the underlying urban context. For example, coldspot clusters in southeastern districts such as Sholing and Peartree closely follow vegetated corridors and low-density residential layouts, as confirmed by reference maps from OS Open Carto. Since these hotspot patterns were computed on the model's test-set predictions rather than the training data, their spatial coherence serves as an independent validation of the model's ability to generalize across different neighbourhoods and structural contexts. This reinforces the credibility of the model's spatial behavior, suggesting that it captures meaningful, interpretable environmental structures rather than generating arbitrary patterns—a key criterion in the emerging field of explainable machine learning (Wu et al., 2020).

These qualities are particularly critical for downstream applications. While traditional models may offer numerically accurate predictions, they often lack the spatial resolution and interpretability necessary for human-centered decision-making. By contrast, the GraphSAGE model provides outputs that are not only statistically robust but also visually and structurally meaningful. Urban planners, transportation engineers, and environmental-health analysts can extract insights not just from aggregated error metrics but from the spatial form, flow, and transition patterns embedded in the model's outputs. Recent reviews emphasize that this combination of predictive power and interpretability is increasingly seen as essential in modern environmental modeling, where outputs must be actionable across diverse stakeholder groups.

In sum, the proposed GraphSAGE framework effectively bridges the gap between quantitative accuracy and spatial narrative, delivering a model that not only fits the data but generalizes across spatial clusters and tells the story of urban noise in a form that is both scientifically rigorous and intuitively meaningful for real-world application.

5.4.6 Limitations and Motivations for Further Generalization

While the GraphSAGE-based model presented in this chapter demonstrates significant improvements in both predictive accuracy and spatial interpretability, several methodological and data-related limitations constrain its broader applicability and generalizability. These limitations are rooted in the characteristics of the input data, the complexity of the feature representation, and the simplifying assumptions inherent in the graph construction process.

5.4.6.1 Limitations of Multispectral and Texture-Based Inputs

A key limitation arises from the model's exclusive reliance on multispectral remote sensing imagery to characterize urban environments. While high-resolution imagery

provides detailed spectral and spatial information, it is inherently limited in capturing functional infrastructure and semantic land use—both of which are critical determinants of urban noise patterns (Dutilleux et al., 2010; Botchkarev, 2019). For example, linear noise sources such as highways are often spectrally indistinct from other impervious surfaces like rooftops, parking lots, or pedestrian walkways. Vegetation cover, shadows, and seasonal changes introduce additional spectral ambiguity, which can fragment or obscure the continuity of road corridors in the data. As a result, while the predicted noise maps generally exhibit coherent patterns, they occasionally fail to fully capture the expected spatial continuity of high-noise zones, particularly where roads intersect vegetated or shadowed areas.

Although the incorporation of GLCM-based texture features helps quantify surface heterogeneity, it remains fundamentally tied to radiometric variation rather than physical or acoustic function. Texture measures alone cannot distinguish between a noisy arterial road and a visually similar but acoustically inert building complex. This highlights a broader limitation: models based solely on imagery-derived features may be insufficient for representing the multifaceted and functionally diverse nature of urban noise dynamics, especially in morphologically complex or mixed-use areas (Genaro et al., 2010).

5.4.6.2 Feature Representation and Model Complexity

Another challenge stems from the high dimensionality and structural complexity of the feature set. The current model integrates over 8,000 statistical descriptors per node, derived from spectral indices, remote sensing bands, and multi-scale texture features. While this rich feature space enables the model to capture fine-grained spatial patterns, it introduces potential drawbacks: feature redundancy, increased computational overhead, longer training times, and a heightened risk of overfitting—particularly in low-sample-density regions or highly heterogeneous environments (Botchkarev, 2019). Moreover, the sheer complexity of the input space can reduce interpretability, making it difficult to attribute predictions to specific variables or spatial drivers.

This complexity also raises concerns about model transferability. Many image-derived variables are sensitive to atmospheric, seasonal, or sensor-specific conditions, which can vary significantly across geographic contexts. As such, models trained on one city's data may struggle to generalize to others without extensive recalibration or domain adaptation. Addressing these issues will require future efforts in feature selection, dimensionality reduction, and the incorporation of higher-level semantic or functional information to build more robust, generalizable models (Terroso-Saenz et al., 2023; Dutta et al., 2024).

5.4.6.3 Assumptions in Graph Construction and Spatial Relationships

A further limitation lies in the graph construction strategy. In this study, spatial adjacency was defined purely based on Euclidean distance using a k-nearest neighbor (k-NN) graph. While computationally efficient and widely adopted, this approach implicitly assumes that physical proximity equates to acoustic connectivity—a simplification that often fails in real-world urban environments (Scarselli et al., 2009;

Wu et al., 2020). Noise propagation is shaped not just by distance but by the presence of physical barriers (e.g., buildings, walls), natural buffers (e.g., parks, forests), and directional conduits (e.g., road networks), none of which are explicitly encoded in the current graph structure.

For example, two nodes 200 m apart but connected by an open road are acoustically very different from two nodes separated by a dense vegetated barrier or a noise wall, even if their Euclidean distance is identical. The current k-NN graph does not differentiate these cases, potentially leading to misestimated spatial influences and reduced predictive precision (Jiang & Luo, 2022; Lei et al., 2024). While the dense sampling of over 50,000 nodes helps approximate complex spatial variation, the underlying graph remains agnostic to functional, directional, and structural nuances.

Future research should explore more semantically informed graph construction strategies, such as weighted edges reflecting landscape permeability or acoustic impedance, road network-based connectivity graphs, or land-use-informed adjacency models. Integrating these approaches could help move graph-based models beyond proximity and toward a more physically and functionally realistic representation of urban noise dynamics, improving both their predictive accuracy and physical interpretability (Bo et al., 2021; Terroso-Saenz et al., 2023).

5.5 Conclusion

This chapter introduced a graph-based framework for high-resolution urban noise prediction, leveraging a GraphSAGE neural network trained on 50,908 unaggregated noise samples from Southampton, each spatially aligned with WorldView-2 multispectral imagery. By moving beyond traditional grid-based or patch-based approaches, the model adopted a non-Euclidean, node-centric spatial formulation that integrates both local and contextual information through multi-scale ring-based statistical features derived from spectral indices and GLCM texture maps.

The full modeling pipeline followed a coherent and physically informed workflow. The process began with data preparation, in which the raw 1 m noise measurements were aggregated to 4 m to match the WorldView-2 imagery, ensuring spatial correspondence across modalities. Around each noise point, six concentric ring buffers ($r = 0\text{--}300\text{ m}$) were generated to capture the environmental structure at multiple scales. Within these buffers, 125 texture descriptors—computed from five spectral indices and weighted by an exponential distance-decay function—formed an 8,250-dimensional feature vector for each node. These features were then embedded in a k-nearest-neighbor graph ($k = 841$, $\tau \approx 500\text{ m}$) that represented both proximity and spatial autocorrelation. Training was performed on dynamically sampled node-centered subgraphs ($\sim 350\text{--}841$ nodes each) to balance computational efficiency with local topological completeness. The GraphSAGE architecture consisted of six convolutional layers with mean aggregation, batch normalization, and Gaussian input perturbation ($\sigma = 0.07$), trained using the Huber loss and adaptive Adam optimization.

Once trained, the model was applied to a 30 m uniform prediction grid (104,272 cells), producing continuous citywide noise maps evaluated on a held-out 20% test subset and validated through full-dataset hotspot and land-use analyses.

Through this integrative workflow, the GraphSAGE model achieved robust predictive performance, with a mean absolute error (MAE) of 4.40 dBA and an R^2 of 0.596, while preserving fine-grained spatial detail. Its outputs successfully reproduced key acoustic patterns, such as the linear spread of noise along arterial roads and the attenuation observed across vegetated and residential buffers. Importantly, these patterns were not only visually interpretable but also statistically validated, underscoring the model's capacity to extract meaningful environmental structures rather than overfitting noise.

In addition to predictive accuracy, the framework demonstrated improved interpretability and spatial realism. Compared to CNN-based baselines, which often yield blocky or over-smoothed surfaces, the GraphSAGE approach produced smooth and topologically coherent noise fields that aligned more closely with urban morphology. This advantage stems from the graph's ability to aggregate features across multi-hop neighborhoods, capturing both local variability and broader spatial dependencies in a manner consistent with the physics of noise propagation.

Nonetheless, several limitations constrain the broader applicability of the framework. The exclusive reliance on imagery-derived spectral and textural features limited its ability to capture functional or semantic acoustic structures, such as the presence of transportation corridors or urban green barriers. Similarly, the use of a Euclidean k-nearest neighbor graph did not fully represent directional or anisotropic propagation effects shaped by complex urban form. Moreover, the high dimensionality of the spectral-textural feature space introduced redundancy and raised challenges for both scalability and generalization across distinct cities.

These reflections reinforce the value of graph-based learning for urban noise prediction while also highlighting its current boundaries. To address these constraints, the next chapter extends the framework by incorporating semantically rich land-use data from the Urban Atlas 2012, alongside refined graph architectures. This integration aims to enhance cross-city generalization by combining the structural advantages of GNNs with categorical information on urban function and morphology, thereby advancing towards a transferable, multi-city noise prediction system.

Chapter 6: Generalizable Urban Noise Prediction via Cross-City Standardized Multispectral Imagery and Dual-Branch Graph Neural Networks

Abstract

Environmental noise mapping is essential for assessing urban soundscapes and guiding public health and planning decisions. Yet, many existing prediction approaches depend heavily on local in-situ measurements, limiting their scalability to cities lacking noise monitoring infrastructure. This study presents a novel generalizable framework for urban noise prediction that leverages multispectral remote sensing imagery and urban land use data, eliminating the need for ground-truth labels in target cities. We introduce a dual-branch graph neural network that integrates spatial features from both local (0–120 m) and contextual (120–1000 m) environments. To address domain discrepancies, the framework incorporates multi-domain adversarial alignment and pseudo-label filtering strategies. Experiments conducted across five UK cities demonstrate the model’s ability to capture noise-relevant patterns and generate spatially coherent noise maps closely aligned with official transportation noise datasets. This approach offers a scalable, label-efficient solution for large-scale urban noise assessment, with broad potential for application in data-scarce urban environments.

Keywords: environmental noise mapping, urban soundscapes, graph neural networks, remote sensing, domain adaptation, label-efficient learning

6.1 Introduction

Building on the findings of Chapter 5, which demonstrated the capacity of graph neural networks (GNNs) to effectively learn urban noise patterns and generate high-quality, spatially coherent noise maps, this chapter expands the research focus toward addressing a critical scalability challenge: how to generalize noise prediction models across cities with minimal or no in-situ acoustic measurements. While the previous chapter validated the feasibility of using multispectral remote sensing and urban land-use data within individual cities, the broader question remains—can these models

reliably extend their predictive power to new urban environments with distinct morphological, socio-economic, and infrastructural characteristics?

Addressing this question is crucial for advancing urban noise modeling into a truly scalable, transferable tool that can support data-scarce cities worldwide. Existing research in related domains such as air pollution modeling, land-use classification, and urban heat island analysis has increasingly adopted machine learning strategies like transfer learning, domain adaptation, and weak supervision to navigate cross-domain variability (Wang et al., 2022; Terroso-Saenz et al., 2024). However, these strategies have seen limited application in urban noise prediction, where the physical mechanisms of acoustic propagation and the heterogeneity of urban form introduce unique challenges (Tuia et al., 2016; Umar et al., 2024).

In this context, the current chapter introduces a novel framework designed to address these gaps by integrating techniques drawn from multiple research streams, including pseudo-labeling, adversarial domain adaptation, and multi-scale spatial modeling. The goal is to create a predictive architecture that does not rely on the availability of direct noise measurements in target cities, yet remains sensitive to both local and contextual urban features.

The technical approach in this chapter draws inspiration from several influential methodological advances. First, pseudo-labeling, originally developed in the context of semi-supervised learning (Lee, 2013; Rizve et al., 2021), provides a strategy for leveraging high-confidence model predictions as approximate labels for unlabeled data. Recent developments in semi-supervised learning, including uncertainty-aware selection mechanisms (Rizve et al., 2021) and confidence-based consistency training frameworks (Sohn et al., 2020), have demonstrated that pseudo-labeling can effectively supplement scarce ground-truth datasets—provided that pseudo-label noise is carefully managed to prevent bias propagation.

Second, adversarial domain adaptation, a core development from transfer learning and domain-invariant modeling, addresses the well-documented challenge of domain shifts across cities. Techniques such as domain-adversarial neural networks (DANN) employ a discriminator network to penalize domain-specific features, thus encouraging the extraction of domain-invariant representations (Ganin et al., 2016; Ouyang et al., 2023). While widely applied in computer vision, malware detection, and graph counterfactual learning (Bilot et al., 2024), multi-domain adversarial alignment—where multiple target domains are aligned simultaneously—remains an underexplored but promising avenue for urban noise modeling, as evidenced by applications in urban scene segmentation and traffic prediction (Jiang & Luo, 2022).

To clarify, in a DANN framework the process can be understood as a two-player game between the feature extractor and a domain discriminator. The feature extractor learns to represent urban environments in a way that predicts noise levels accurately for the labeled source city (e.g., Southampton). Simultaneously, the discriminator attempts to identify which city each sample originates from (e.g., Southampton vs Liverpool). Through gradient-reversal training, the feature extractor gradually learns to fool the discriminator by producing features that are indistinguishable across cities. The result is a model that retains predictive capacity for noise levels while becoming

insensitive to city-specific imaging or land-use biases—a property essential for generalizing to unlabeled target domains.

Finally, the architectural backbone of the proposed framework builds on multi-scale graph neural networks (GNNs), which have shown strong performance in modeling spatially complex phenomena, including traffic dynamics, social networks, and urban morphology (Lei et al., 2024; Ling et al., 2024; Wang et al., 2024). By explicitly integrating local-scale (fine-grained) and contextual-scale (broader neighborhood) spatial features, the framework aims to robustly capture both short-range and long-range acoustic influences, thereby improving model generalization across diverse urban landscapes (Bo et al., 2021; Chen et al., 2024).

In practical terms, this approach addresses a pressing real-world need. Many cities—especially medium-sized or data-scarce ones—lack detailed acoustic measurements due to the cost and logistics of in-situ monitoring. The proposed framework enables these municipalities to estimate noise exposure directly from publicly available satellite imagery and land-use data, providing actionable insights for urban planning, health impact assessment, and transportation design. For example, a city such as Nottingham or Cardiff could apply the trained model to identify potential high-noise corridors and evaluate mitigation strategies without deploying dense sensor networks.

In summary, this chapter contributes to the literature by proposing a scalable, label-efficient urban noise prediction framework that integrates (i) robust pseudo-labeling strategies, (ii) multi-domain adversarial alignment, and (iii) dual-scale GNN architectures. The remainder of the chapter outlines the research objectives, details the experimental design across five UK cities, and presents the evaluation framework, while the subsequent methodology chapter provides comprehensive technical details on the model architecture, training procedures, and validation strategies.

6.2 Methodology

This chapter presents a comprehensive methodological framework designed to enable generalizable urban noise prediction across multiple cities with distinct spatial configurations, land-use characteristics, and varying levels of data availability. Building on the insights of Chapters 5—where graph neural networks (GNNs) were shown to effectively learn noise-related spatial patterns within single cities—this chapter addresses the broader challenge of extending these capabilities to unlabeled cities lacking direct in-situ measurements. The framework integrates harmonized land-use and land-cover (LULC) datasets, standardized remote sensing inputs, and a multi-domain graph learning strategy, aiming to create a scalable, transferable system for urban noise mapping.

At the core of this framework lies a dual-branch feature architecture, specifically designed to capture both local (0–120 m) and contextual (120–1000 m) environmental information relevant to urban noise propagation. All input data are spatially anchored to a consistent 30-meter grid that spans both labeled and unlabeled cities, ensuring uniform sampling density and facilitating coherent graph construction. The

framework further enhances cross-domain learning through the use of adversarial domain alignment and pseudo-labeling mechanisms, which together enable the transfer of knowledge from a well-annotated reference city to multiple target cities without direct noise measurements.

Compared to prior studies, the framework introduces several key methodological advancements. These include the integration of Urban Atlas 2012 LULC data to derive interpretable environmental variables and pseudo-invariant features for spectral standardization, the generation of multi-scale spatial features tailored to the dual-branch GNN architecture, and the construction of unified spatial graphs that preserve structural continuity across labeled and unlabeled domains. Additionally, the framework implements a multi-domain alignment mechanism that reduces distributional gaps between cities, thereby enhancing generalization under domain shift conditions.

6.2.1 Overall Framework Design

While the overall objective of this study remains consistent with the previous chapters—predicting spatially explicit urban noise levels from remote sensing and land-use information—the processing workflow in this chapter is substantially redesigned to address the challenge of cross-city generalization. In contrast to Chapters 4 and 5, which focused on fitting models to fully labeled data within Southampton, the present framework aims to transfer learned knowledge to unlabeled target cities. This shift from model fitting to model generalization necessitated several key methodological adjustments.

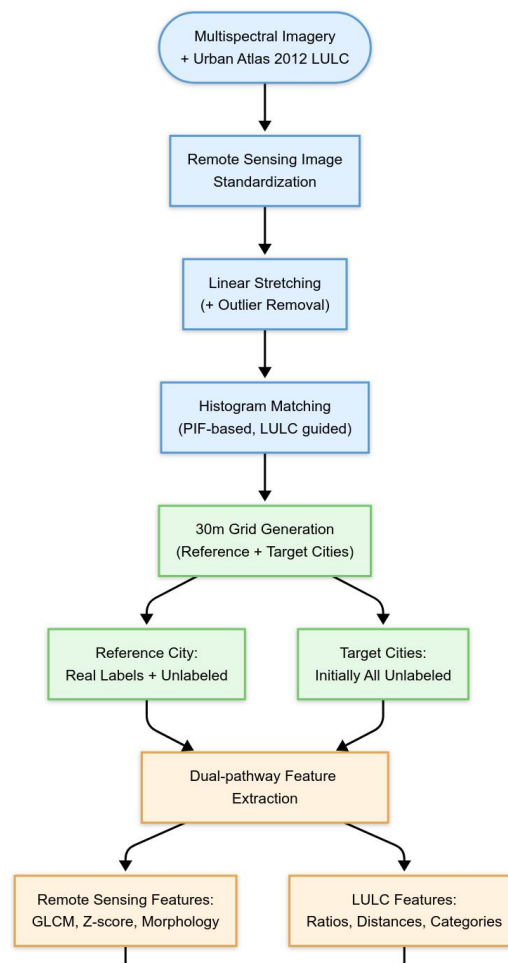
First, all input imagery was harmonized across domains through pseudo-invariant feature (PIF)-based radiometric normalization and histogram matching, ensuring that spectral variations represent true environmental differences rather than atmospheric or sensor inconsistencies. Second, the feature construction strategy now integrates two complementary spatial scales—a fine-scale local domain (0–120m) and a broader contextual domain (120–1000m)—and introduces Urban Atlas 2012 land-use indicators as semantic features bridging physical and functional properties of urban space. Third, the graph representation has been redefined from the Southampton-specific 4 m KNN graph to a unified 30 m Gaussian-weighted grid graph applicable to all cities, enabling consistent neighborhood relationships across domains. Finally, the learning framework has evolved into a dual-branch GNN architecture (GAT + GCN) equipped with adversarial and MMD-based domain alignment modules, explicitly minimizing distributional discrepancies between the labeled source and unlabeled target domains.

Collectively, these revisions represent a deliberate methodological transition—from maximizing in-city predictive accuracy to achieving transferable, domain-robust acoustic mapping. Each modification in the preprocessing and modeling pipeline directly supports this goal, ensuring that the framework can generalize learned spatial relationships beyond Southampton to diverse urban contexts.

To further safeguard domain independence during learning, each city is abstracted as a distinct 30 m grid domain, where each node represents a 30 m $\times$ 30 m cell and every unit change in the two-dimensional (row, col) coordinate system corresponds to a 30 m physical displacement. Each domain is assigned a unique global offset to ensure spatial isolation in the coordinate space—for example, Southampton’s grid indices are incremented by +10 000, Portsmouth by +20 000, and so forth. This design creates non-overlapping spatial domains separated by thousands of virtual kilometers, preventing unintended cross-domain message propagation or feature contamination during graph construction.

During training, message passing and domain-alignment losses operate only within these logically isolated domains, while adversarial and MMD mechanisms align their latent feature distributions in the embedding space. This guarantees that each city retains its own spatial integrity, yet contributes to the shared domain-invariant representation learned by the network.

The overall methodological pipeline is illustrated in Figure 6.1, outlining the sequential stages from data preparation to cross-domain model training and inference. The framework is explicitly designed to support weakly supervised learning by combining real labels, pseudo-labels, and graph-based feature propagation across a shared spatial infrastructure.



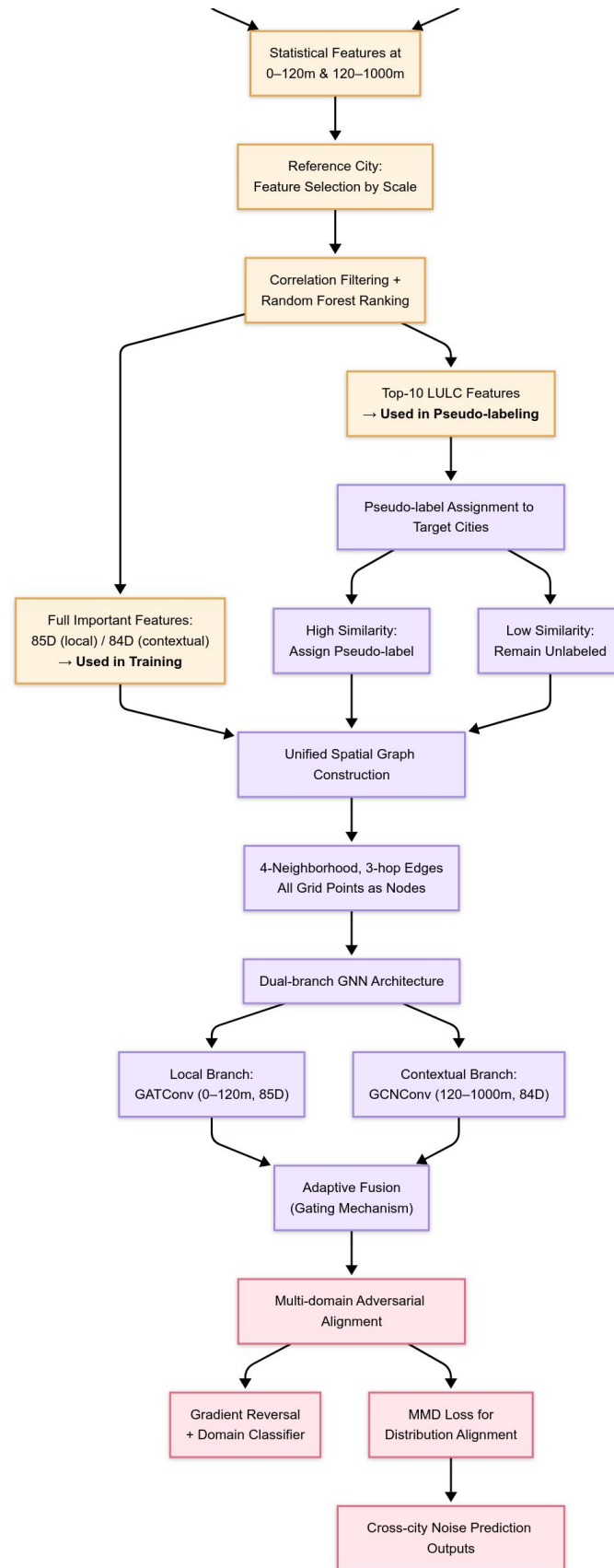


Figure 6.1 Proposed methodological workflow. Steps are color-coded by functional stage: data preparation (blue), grid sampling (green), feature extraction and selection (orange), graph and model construction (purple), and domain adaptation with pseudo-labeling (pink)

The process begins with the acquisition of two primary datasets across five UK cities: Southampton (serving as the reference city with measured noise data) and four target cities—Cardiff, Portsmouth, Nottingham, and Liverpool—where direct noise measurements are unavailable. The first dataset consists of WorldView-2 multispectral imagery, which provides eight-band high-resolution spectral data capturing surface material properties, vegetation patterns, and urban structures. The second dataset is the Urban Atlas 2012 LULC product, a standardized European database offering 10-meter spatial resolution land-use classifications across 24 categories, enabling harmonized semantic descriptions of urban environments (EEA, 2016).

To ensure cross-city comparability, all multispectral images undergo a two-step spectral standardization process. First, a linear stretch is applied after discarding the top and bottom 2% of pixel values, normalizing the dynamic range across datasets. Second, histogram matching is performed using pseudo-invariant features (PIFs) identified from spectrally stable land-use classes, such as continuous urban fabric and commercial zones. This correction procedure, following best practices in remote sensing (Schott et al., 1988; Du et al., 2002), compensates for seasonal, atmospheric, and sensor-specific variations, ensuring that extracted features retain physical and statistical consistency across domains.

Noise prediction samples are generated by overlaying a uniform 30-meter grid across the study areas of both the reference and target cities. For the reference city, raw noise measurements are aggregated to the grid using spatial averaging, and a binary mask identifies grid cells containing real data. In the target cities, all grid cells are initially unlabeled but will later be enriched with pseudo-labels derived from feature similarity analyses.

Feature extraction is performed at two complementary spatial scales. At the local scale (0–120 m), features capture direct acoustic influences, including adjacent road proximity, building density, and local surface composition. At the contextual scale (120–1000 m), features characterize broader environmental patterns, such as regional transport corridors, land-use zoning, and green buffer distributions. This multi-scale decomposition is rooted in established environmental-modelling principles, which recognise both proximate and broader determinants of urban dynamics (Raimbault, 2021; Cushman, 2024).

To identify the most predictive variables, a two-step feature selection process is applied using the labeled samples from the reference city. Initially, candidate variables are filtered based on their Pearson correlation with measured noise levels. Subsequently, a random forest regressor is used to rank feature importance, and the top ten variables are retained for pseudo-label generation. Pseudo-labels are assigned to target city samples by comparing each sample's feature profile (based on these top-ranked variables) with the labeled reference samples, using a defined similarity threshold to ensure reliability. As a result, the target cities contain a mix of pseudo-labeled and unlabeled grid points, while the reference city maintains both labeled and unlabeled nodes.

All samples, regardless of label status, are embedded within a spatial graph structure, where nodes represent 30-meter grid cells and edges are constructed based on four-

neighborhood proximity within a three-hop range. This design promotes dense topological connectivity and facilitates neighborhood-based feature propagation. Importantly, while only labeled and pseudo-labeled nodes contribute to the model's supervised loss, unlabeled nodes participate fully in the graph's connectivity, enhancing learning stability and supporting weak supervision (Zhou et al., 2003; Kipf, 2016).

Scale-aware receptive field design

Unlike the deeper six-layer GraphSAGE model presented in Chapter 5, which expanded the receptive field to a radius of 1–3 km to fully capture within-city spatial correlations, the cross-city framework deliberately adopts a shallower and more structured dual-branch design consisting of two Graph Attention (GAT) layers followed by a single Graph Convolution (GCN) layer. The local branch (GAT) operates over features derived from 0–120 m concentric rings. The attention mechanism enables anisotropic message passing, giving more weight to acoustically relevant directions (e.g., along nearby roads), and approximates a fine-grained receptive field of roughly:

$$R_{\text{local}} \approx 2 \times 120$$

after two hops of neighbor aggregation. This scale emphasizes micro-environmental influences such as adjacent roads, building edges, and green buffers. The contextual branch (GCN) aggregates broader neighborhood signals using features extracted from 120–1000 m rings, approximating

$$R_{\text{context}} \approx 1000$$

within a single layer. This captures large-scale structural effects such as transport corridors, land-use zoning, and the spatial configuration of major green areas. Graph connectivity is constructed on a 30 m grid, and each edge corresponds to one hop of 30 m physical displacement. Under the three-hop neighborhood design, the maximum topological extent is approximately

$$R_{\text{max}} = 3 \times 30 \text{ m} = 90 \text{ m}$$

which is spatially aligned with the local ring buffer scale and ensures stable neighborhood propagation across cities with different urban morphologies. This is a deliberate simplification compared with Chapter 5: whereas the deeper GraphSAGE model was optimized for within-city accuracy, the shallower dual-branch architecture constrains the model's receptive field, reducing overfitting to city-specific morphological patterns and facilitating domain-invariant representation learning. Formally, for each node i , the attention-based branch aggregates from its neighborhood $N(i)$ as

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)} \right)$$

where $\alpha_{ij}^{(l)}$ is the learned attention weight. This anisotropic weighting allows the network to focus on directionally structured acoustic propagation, such as along linear road segments or building frontages. The contextual GCN layer then applies isotropic aggregation to capture background structural influences across the 120–1000 m scale.

By explicitly decomposing neighborhood effects into local and contextual branches, the dual-branch GNN effectively balances sensitivity to near-field variability with robustness to cross-domain distribution shifts. This scale decomposition is critical for domain adaptation: local attention captures transferable environmental cues (e.g., road density, façade adjacency), while the contextual branch provides a stable morphological backdrop less sensitive to city-specific spectral or textural idiosyncrasies.

Multi-domain Adversarial and Statistical Alignment.

To mitigate discrepancies in feature distributions across cities, the proposed framework integrates a multi-domain adversarial alignment mechanism combining adversarial learning and statistical matching. The approach operates on top of the dual-branch GNN representation layer, where latent embeddings from each node are optimized not only for noise prediction but also for domain invariance.

A domain classifier is attached to the shared feature space through a gradient reversal layer (GRL) (Ganin et al., 2016). During training, the main predictor minimizes the supervised regression loss for labeled and pseudo-labeled samples, while the domain classifier attempts to identify the city of origin for each embedding among all five domains (Southampton + four targets). The GRL inverts the gradient from the classifier to the feature extractor, forcing the extractor to learn representations that confuse the classifier, i.e., features indistinguishable across cities. This adversarial competition drives the model toward a shared, domain-invariant latent structure.

Complementing this adversarial process, a maximum mean discrepancy (MMD) loss (Long et al., 2015) is applied across domain pairs to reduce residual distributional gaps that are not captured by the classifier. While the adversarial term aligns categorical boundaries (city identities), the MMD term minimizes the distance between the empirical kernel mean embeddings of feature distributions in a reproducing-kernel Hilbert space, thereby harmonizing both first- and higher-order statistical moments. In practice, the two components operate jointly: adversarial alignment removes discrete domain cues, and MMD refines the continuous distributional overlap.

Unlike conventional single-domain adversarial adaptation, which aligns only a single source-to-target pair, the proposed multi-domain scheme performs simultaneous alignment among all five cities. The discriminator’s output layer includes multiple classes corresponding to each domain, and alignment is jointly enforced between the labeled source (Southampton) and all unlabeled targets as well as among the targets themselves (inter-target alignment). This broader alignment prevents the model from overfitting to the spectral or morphological idiosyncrasies of any single city, thereby improving robustness under cross-city domain-shift conditions.

The overall optimization objective combines three components:

- (1) the Huber regression loss for noise prediction,
- (2) the adversarial cross-entropy loss from the domain classifier scaled by λ_a , and
- (3) the MMD penalty scaled by λ_m .

Formally,

$$\mathcal{L}_{total} = \mathcal{L}_{Huber} + \lambda_a \mathcal{L}_{adv} + \lambda_m \mathcal{L}_{MMD}.$$

The joint minimization enables the GNN encoder to learn acoustically meaningful yet domain-invariant features, ensuring that predictions remain consistent even when transferred to previously unseen urban morphologies.

6.2.2 Study Areas and Data Sources

This section details the geographic scope, sampling strategy, and primary datasets that underpin the development of the proposed generalizable urban noise prediction framework. The study focuses on five cities in the United Kingdom—Southampton, Cardiff, Portsmouth, Nottingham, and Liverpool—selected based on their spatial heterogeneity, data accessibility, and suitability for evaluating cross-domain generalization. Among these, Southampton serves as the reference domain because it is the only city with in-situ environmental noise measurements, while the remaining four serve as unlabeled target domains.

The inclusion of these cities enables a comprehensive cross-city evaluation that captures a wide range of urban forms, transport infrastructures, and land-use configurations. Southampton, a compact coastal port city, provides the foundation for supervised model calibration. Cardiff, Portsmouth, and Liverpool are major waterfront settlements with varying degrees of industrial, commercial, and residential integration, while Nottingham represents a dense inland administrative center with compact mixed-use morphology. This diversity allows the proposed model to be tested across both coastal and inland urban contexts, ensuring that generalization performance is evaluated under realistic morphological variation.

To maintain consistency, all datasets were resampled to a uniform 30 m grid resolution, balancing spatial detail with computational efficiency. This grid aligns with both the WorldView-2 multispectral imagery and the Urban Atlas 2012 land-use data, ensuring that graph nodes represent comparable spatial units across cities. The uniform sampling framework also supports stable graph construction by enabling consistent topological representation in all domains.

Quantifying Spatial Heterogeneity

To substantiate the morphological diversity underlying city selection, spatial heterogeneity was quantitatively assessed using three complementary indicators:

- (i) GLCM texture entropy derived from 4 m NDVI imagery (identical to the method in Chapters 4 and 5), reflecting fine-scale surface variability;
- (ii) land-use Shannon entropy (H') based on Urban Atlas 2012, representing functional diversity in land-use composition; and
- (iii) the mean built-up ratio, indicating the proportion of impervious urban surfaces.

Together, these indicators describe different dimensions of heterogeneity—local texture, functional mixing, and structural density. As summarized in Table 6.1, Southampton and Nottingham exhibit the highest texture entropy (≈ 1.25) and strong internal variability ($CV \approx 0.37$), indicative of fine-grained built–vegetation mosaics. Cardiff, by contrast, shows the highest land-use entropy ($H' = 1.21$) but the lowest built-up ratio (0.75), reflecting a balanced mixture of urban, green, and water areas. Southampton and Liverpool display compact port morphologies with high built-up ratios (0.94–0.97) but lower H' (0.76–0.88), while Portsmouth lies in between, combining moderate functional diversity with high structural density.

These contrasting patterns confirm that the five cities represent distinct morphological regimes, providing robust empirical evidence of spatial heterogeneity across domains and justifying their use for testing the transferability of the proposed dual-branch GNN framework.

Table 6.1 Quantitative indicators of spatial heterogeneity across study cities

city	entropy_mean	entropy_std	entropy_cv	H_LULC	Builtup_ratio
Soton	1.25	0.46	0.37	0.76	0.97
Cardiff	1.21	0.46	0.38	1.21	0.75
Portsmouth	1.21	0.45	0.38	1.08	0.85
Nottingham	1.25	0.47	0.37	0.97	0.90
Liverpool	1.25	0.46	0.37	0.88	0.94

Entropy computed from 4 m NDVI (GLCM method consistent with Chapters 4–5); land-use and built-up metrics derived from Urban Atlas 2012.

6.2.2.1 Urban Context and Noise Sampling Design

In the graph representation used throughout this chapter, each node corresponds to the centroid of a 30 m $\times$ 30 m grid cell covering the built-up extent of the five selected cities. This uniform grid provides a spatially consistent basis for feature extraction, graph construction, and model inference. The 30 m resolution was chosen to balance local spatial detail—sufficient to capture short-range acoustic variability—with computational feasibility for large-scale, cross-city modeling.

Within the source domain (Southampton), nodes inherit acoustic labels derived from the weighted aggregation of in-situ noise measurements onto the grid cells. Each such node thus represents a spatial unit with both environmental attributes (e.g., spectral indices, land-use ratios, texture metrics) and an associated ground-truth noise level. In contrast, in the four target cities, nodes are initially unlabeled; pseudo-labels are later assigned to a subset of these nodes based on their spectral-textural and land-use similarity to labeled Southampton samples, as detailed in Section 6.2.4.

It is important to note that pseudo-labels do not define the nodes themselves—they represent only one optional attribute of certain nodes within the graph. All grid-centroid nodes, whether labeled, pseudo-labeled, or unlabeled, are retained to ensure structural continuity. During graph construction, every node participates in the spatial

adjacency defined by a combination of 4-neighborhood and 3-hop linkages, allowing message passing between labeled and unlabeled regions. This design enables the model to leverage contextual information through feature propagation, thereby maintaining consistent graph topology across cities even when true noise labels are absent.

6.2.2.2 Remote Sensing and Urban Atlas Data

This study integrates two core geospatial datasets to construct a comprehensive, high-resolution foundation for urban noise prediction: multispectral satellite imagery from the WorldView-2 European Cities dataset and harmonized land-use data from the Urban Atlas 2012 provided by the European Environment Agency (EEA, 2016). Together, these datasets offer complementary thematic and spatial information, enabling the extraction of both spectral and contextual features critical to modeling acoustic patterns across heterogeneous urban environments.

The WorldView-2 imagery provides eight spectral bands at sub-5-meter resolution, capturing detailed reflectance profiles of urban surface materials, vegetation, and built-up structures. Table 6.2 lists the acquisition dates for each city, illustrating temporal alignment between imagery and available noise sampling in Southampton, while the target cities’ imagery (from 2011–2012) aligns with the reference year of the Urban Atlas 2012. Despite the slightly later imagery acquisition for Southampton, the temporal proximity ensures consistency between ground-truth labels and remote sensing features, which is essential for accurate supervised model calibration.

Table 6.2 WorldView-2 multispectral imagery acquisition dates

City	Acquisition Date
Portsmouth	23 March 2011
Nottingham	26 May 2012
Liverpool	19 April 2011
Cardiff	19 April 2011
Southampton	15 February & 6 May 2016 (mosaic)

Following data acquisition, a rigorous two-stage spectral normalization was applied to the satellite imagery to ensure cross-city comparability. First, a linear stretch was used to exclude extreme reflectance outliers, defined by removing the top and bottom 2% of values, thus reducing noise from anomalous pixels. Second, global histogram matching was performed using pseudo-invariant features (PIFs)—spectrally stable built-up areas identified from the Urban Atlas LULC classes. These included continuous urban fabric (11100), discontinuous urban fabric (11210, 11220), industrial and commercial zones (12100), and transportation infrastructure (12210, 12220, 12230). After normalization, all imagery was resampled to a unified 4-meter spatial resolution, ensuring compatibility across cities and preserving key spectral patterns linked to acoustic propagation.

The Urban Atlas 2012 dataset provided detailed vector-based LULC classification at a consistent 10-meter resolution across all study areas. It includes 24 LULC categories, which were grouped into five functional classes based on their presumed relationships to environmental noise: strong noise sources (e.g., transportation corridors, ports), moderate noise sources (e.g., industrial areas), strong mitigation zones (e.g., dense forests), moderate mitigation zones (e.g., pastures, urban green spaces), and neutral areas (e.g., low-density residential zones, isolated structures). Table 6.3 summarizes the functional groupings used in this study, providing a thematic framework for interpreting land-use impacts on urban noise dynamics.

Table 6.3 Urban Atlas 2012 LULC classification and functional grouping used in this study

Code	LULC Class Name	Functional Group
11100	Continuous urban fabric	Neutral
11210	Discontinuous dense urban fabric	Neutral
11220	Discontinuous medium density urban fabric	Neutral
11230	Discontinuous low density urban fabric	Neutral
11240	Discontinuous very low density urban fabric	Neutral
11300	Isolated structures	Neutral
12100	Industrial, commercial, public, military areas	Moderate noise source
12210	Fast transit roads and associated land	Strong noise source
12220	Other roads and associated land	Moderate noise source
12230	Railways and associated land	Strong noise source
12300	Port areas	Strong noise source
12400	Airports	Strong noise source
13100	Mineral extraction and dump sites	Strong noise source
13300	Construction sites	Neutral
13400	Land without current use	Neutral
14100	Green urban areas	Moderate mitigation
14200	Sports and leisure facilities	Moderate mitigation
21000	Arable land (annual crops)	Strong mitigation
23000	Pastures	Moderate mitigation
31000	Forests	Strong mitigation
32000	Herbaceous vegetation associations	Moderate mitigation
33000	Open spaces with little or no vegetation	Moderate mitigation
40000	Wetlands	Neutral
50000	Water bodies	Neutral

These datasets played dual roles in the modeling framework. First, the PIFs extracted from the Urban Atlas classes guided the spectral normalization process, ensuring that imagery across cities was harmonized for feature extraction. Second, LULC-derived features, including class proportions, Euclidean distances to functional zones, and dominant land-use types, were computed within both local (0–120 m) and contextual

(120–1000 m) neighborhoods for each grid cell, supplying the model with rich spatial and thematic context necessary for robust noise prediction.

6.2.2.3 Study Area Delineation and Grid Generation

To ensure a consistent spatial framework across all study sites, a systematic approach was employed for delineating study areas and generating uniform sampling grids. The research encompassed five major UK cities—Southampton, Portsmouth, Cardiff, Nottingham, and Liverpool—selected based on a combination of data availability, morphological diversity, and transferability relevance.

City selection rationale

All chosen cities are jointly covered by both the WorldView-2 European City Archive (imagery acquired between 2011 and 2016) and the Urban Atlas 2012 dataset. This overlap guarantees temporal and thematic consistency between spectral, land-cover, and demographic layers, allowing cross-city comparison under a unified spatio-temporal framework. In addition, all population and density metrics were aligned to the 2011 UK Census to ensure full temporal coherence with the remote-sensing and land-use reference period.

From a design perspective, the five cities collectively form a progressive morphological gradient that systematically increases domain disparity relative to Southampton, the labeled source city. Portsmouth, a compact coastal settlement immediately adjacent to Southampton, provides the closest analogue for low-difficulty domain transfer. Cardiff, the Welsh capital located across the Bristol Channel, represents a moderately distinct coastal-estuarine morphology with greater topographic variation and a more mixed functional structure. Nottingham, as an inland and densely nucleated core city, introduces a contrasting spatial configuration dominated by compact built-up fabric and fewer water or industrial zones. Liverpool, the largest metropolitan port city in this study, exemplifies high morphological complexity, with extensive industrial and brownfield areas, mixed commercial-residential land uses, and a dense arterial road network.

Quantitative evidence further supports this diversity. Based on Urban Atlas 2012 statistics, the proportion of continuous and discontinuous urban fabric varies from 55 % in Portsmouth and 61 % in Nottingham, to 35 % in Cardiff, 24 % in Liverpool, and 58 % in Southampton. Green and water spaces show an inverse trend, ranging from 14 % in Portsmouth to nearly 30 % in Cardiff. Population densities (UK Census 2011) also reflect distinct urban intensities: Southampton $\approx 4,700$ persons km⁻², Portsmouth $\approx 5,100$ persons km⁻², Cardiff $\approx 2,600$ persons km⁻², Nottingham $\approx 4,000$ persons km⁻², and Liverpool $\approx 4,700$ persons km⁻². These variations correspond to demonstrably different urban typologies—coastal compact (Portsmouth, Southampton), estuarine capital (Cardiff), inland core (Nottingham), and metropolitan port (Liverpool)—each representing unique acoustic propagation environments and structural challenges for model transferability.

Study area delineation and grid generation

Each city's urban extent was defined using its administrative boundaries, which were

then refined through manual adjustments to align precisely with the coverage limits of both Urban Atlas 2012 and available WorldView-2 imagery. This two-stage delineation ensured that all input data layers were spatially synchronized and free of coverage gaps. To mitigate potential edge effects—especially given that the proposed model integrates multi-scale spatial context up to 1000 m—peripheral regions with incomplete environmental neighborhoods were trimmed systematically. This step prevented artificial spatial bias from incomplete surroundings and guaranteed that all retained samples contained full contextual information for both local and background feature extraction.

Within each finalized study boundary, a regular 30 m × 30 m grid was generated, systematically covering the entire urban area without gaps. This fine-resolution design captures detailed urban heterogeneity relevant to noise propagation while maintaining computational feasibility for graph construction and feature integration. Unlike previous approaches that excluded unlabeled cells, all grid points were retained regardless of label availability to maximize spatial continuity within the graph topology. The resulting numbers of grid points per city—ranging from ~43,000 (Portsmouth) to ~118,000 (Liverpool)—reflect the varying spatial extents and morphological complexity of the study areas (Table 6.4).

Table 6.4 Demographic and morphological characteristics of the five study cities (UK Census 2011 and Urban Atlas 2012)

City	Population (2011)	Density (persons /km ²)	Dominant LULC (Urban Atlas 2012)	Morphological Type	Similarity to Southampton	Grid Points
Southampton	236900	4700	58% urban fabric; 17% green; 10% industrial zones	Coastal port city (source domain)		57485
Portsmouth	205100	5100	55% urban fabric; 14% green; 7% industrial	Compact coastal analogue	High	43042
Cardiff	346100	2600	35% urban fabric; 28% green/water; 18% industrial/commercial	Coastal–estuarine capital	Moderate	114994
Nottingham	305700	4000	61% urban fabric; 25% residential; <5% industrial	Inland compact core	Moderate–Low	80692
Liverpool	466400	4700	24% industrial/commercial; 20% green corridors; 15% brownfield	Large metropolitan port	Low	118262

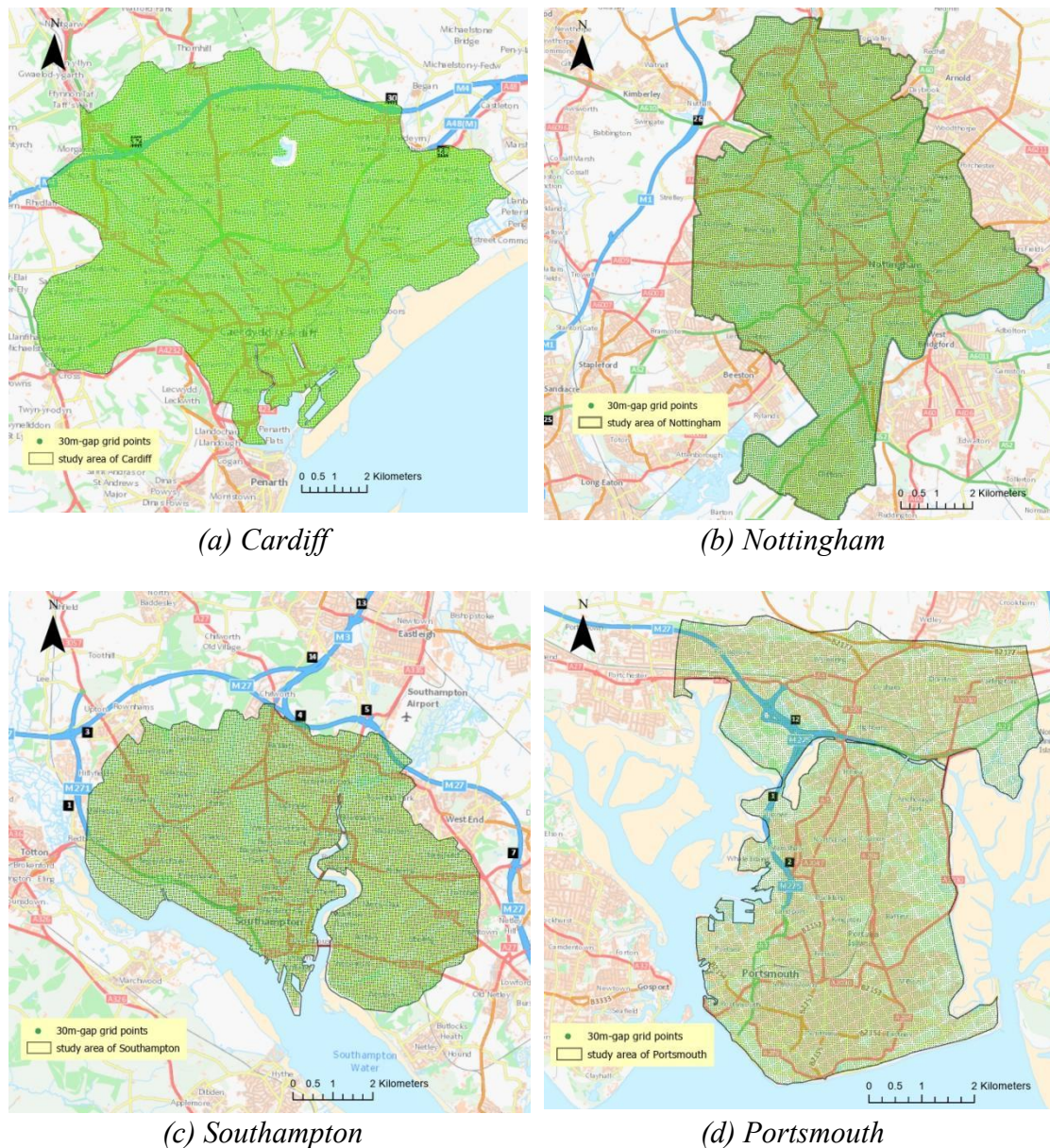
All data are temporally aligned with the 2011 – 2016 period of the WorldView-2 imagery. Population and density values are derived from the 2011 UK Census (ONS); land-use/land-cover (LULC) statistics are summarized from Urban Atlas 2012 (EEA).

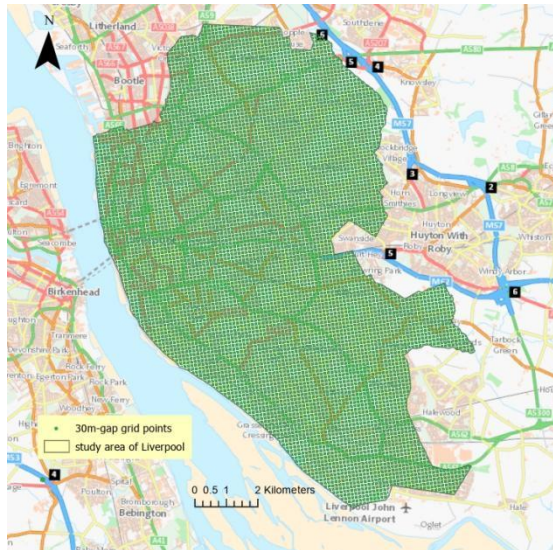
The cities collectively represent a progressive gradient of morphological complexity—from the compact coastal analogues of Southampton and Portsmouth to the metropolitan port of Liverpool—forming the empirical foundation for cross-domain transfer evaluation.

In Southampton, high-resolution (1 m) mobile noise measurements were aggregated to the 30 m grid to produce labeled samples forming the supervised learning base. The

remaining unlabeled grid points, though lacking direct noise data, provided essential neighborhood connectivity within the graph structure. For the four target cities—Cardiff, Portsmouth, Nottingham, and Liverpool—all grid points were initially unlabeled; subsequently, subsets received high-confidence pseudo-labels based on similarity to Southampton samples (Section 6.2.3), enabling weak supervision during training.

The final dataset thus comprised a dense mix of labeled, pseudo-labeled, and fully unlabeled nodes integrated into city-specific graphs, each built under a uniform connectivity rule using a 4-neighborhood scheme extended to three hops. This ensured that all nodes possessed sufficient spatial context for effective message passing during training and inference. Visual representations of the study areas, their administrative boundaries, and the generated 30 m grids are presented in Figure 6.2, illustrating the spatial sampling framework and morphological diversity captured in this study..





(e) Liverpool

Figure 6.2 Study Area Delineations and 30 m Grid Point Distributions

6.2.3 Remote Sensing Image Standardization

To ensure reliable spectral comparability across cities, all WorldView-2 multispectral imagery in this study was subjected to a comprehensive, multi-step standardization workflow. This workflow integrates two principal components: bandwise linear stretching and weighted histogram matching, both guided by pseudo-invariant features (PIFs). Together, these steps minimize inter-city variability introduced by differences in image acquisition time, atmospheric conditions, and sensor calibration, ensuring a harmonized spectral foundation for downstream analyses.

The first critical step involved the identification of pseudo-invariant features, or PIFs. Following established remote sensing practices (Schott et al., 1988; Du et al., 2002), PIFs were selected based on land-use categories known to maintain stable spectral signatures over time and across geographic contexts. Specifically, the study used Urban Atlas 2012 classes including continuous urban fabric (11100), discontinuous dense and medium-density urban fabric (11210, 11220), industrial, commercial, and public areas (12100), as well as fast transit roads, other roads, and railways (12210, 12220, 12230). These classes were spatially delineated using vector masks, and the corresponding pixel values were extracted bandwise from each multispectral raster.

Once PIFs were established, the second step applied a bandwise linear stretching procedure designed to align the dynamic ranges of the target and reference images. This process involved matching the 2nd and 98th percentiles of the PIF pixel distributions in the target image to the minimum and maximum PIF pixel values observed in the reference image. The mathematical formulation of this stretch for a given pixel value x is:

$$x_{\text{stretched}} = \frac{(x - P_{2\%}^{\text{target}})}{(P_{98\%}^{\text{target}} - P_{2\%}^{\text{target}})} \times (P_{\text{max}}^{\text{ref}} - P_{\text{min}}^{\text{ref}}) + P_{\text{min}}^{\text{ref}}$$

where:

- $P_{2\%}^{\text{target}}$ and $P_{98\%}^{\text{target}}$ are the 2nd and 98th percentiles of the target image's PIF pixel values,
- $P_{\text{min}}^{\text{ref}}$ and $P_{\text{max}}^{\text{ref}}$ are the minimum and maximum PIF pixel values in the reference image.

This linear transformation rescales the target pixel values to align their spread with that of the reference image, preserving physical reflectance interpretability while removing saturation or compression artifacts.

Following linear stretching, the third step implemented weighted histogram matching across all PIF classes. This process constructed weighted global histograms H for both the reference and target images, where each PIF class c contributed proportionally according to its pixel count n_c and total weight $w_c = \frac{n_c}{N}$, with N being the total number of PIF pixels. The cumulative distribution functions (CDFs) were then computed:

$$\text{CDF}(i) = \sum_{j=0}^i H(j)$$

where i is the intensity level, and $H(j)$ is the weighted histogram count at level j . The mapping function aligned the CDF of the target image to that of the reference image, effectively adjusting the relative frequency of brightness levels across bands. Any out-of-range values after matching were clipped to preserve numerical stability.

After these spectral adjustments, the standardized images were saved at their original spatial resolution. For special cases, such as the Southampton mosaic comprising two acquisition dates, additional resampling to 4 m resolution was applied, ensuring consistent feature extraction across all study sites.

Finally, the quality of the standardization was evaluated by comparing summary statistics (mean, standard deviation, percentiles) of the PIF pixels in both the original and standardized domains. This statistical verification confirmed substantial alignment of spectral profiles, indicating that the normalization workflow effectively harmonized the data for reliable cross-city modeling and analysis.

6.2.4 Feature Variable Construction

This section describes the design and construction of multi-scale feature variables. The construction of feature variables was guided by two core principles: predictive

strength and cross-city transferability. Priority was given to features that are interpretable, physically meaningful, and generalizable across domains, while variables highly sensitive to city-specific configurations or overfitting risks were systematically excluded. Feature selection was conducted independently for the local and contextual branches, leveraging both correlation filtering and random forest–based importance ranking, as elaborated later in this section. This careful selection process ensured that the retained features not only captured meaningful environmental patterns but also contributed to the overall robustness of the multi-domain learning framework.

6.2.4.1 Rationale for Multi-Scale Feature Design

Urban environmental noise is influenced by a complex interplay between localized sources and broader spatial context. Short-range factors such as proximity to road segments, building density, or specific land-cover features exert direct acoustic impacts, while large-scale patterns—like the arrangement of zoning types, the presence of green corridors, or the clustering of industrial areas—modify long-range noise propagation and attenuation effects (Aletta et al., 2016). To accurately reflect this multi-layered phenomenon, the feature construction strategy explicitly incorporated two complementary spatial scales: a local scale (0–120 m) and a contextual scale (120–1000 m).

At the local scale, feature variables capture fine-grained neighborhood characteristics and immediate acoustic influences, aligning with typical urban block dimensions and the rapid decay profiles of environmental noise over short distances. This allows the model to represent detailed interactions between built-up features and noise emissions at the street or block level. In contrast, the contextual scale provides a coarser but essential overview of the surrounding landscape, including large green spaces, industrial belts, or arterial transport corridors. These broader features are critical for capturing reflections, absorption zones, or long-range acoustic barriers, which can shape the background noise environment at the neighborhood or district scale (Aletta et al., 2016; Khan et al., 2018).

From a modeling perspective, the dual-scale logic is directly mirrored in the architecture of the dual-branch GNN (Section 6.3.1). The local branch employs Graph Attention Networks (GAT), which focus on learning adaptively weighted interactions between nearby nodes, allowing the model to prioritize the most acoustically influential local features. The contextual branch, by contrast, uses Graph Convolutional Networks (GCN) to encode structural continuity and broader neighborhood patterns across the urban landscape. This architectural separation enables effective feature decoupling between scales, reducing redundancy, mitigating overfitting, and encouraging the model to learn distinct but complementary patterns associated with different noise mechanisms (Khan et al., 2018; Wu et al., 2020).

The use of multi-scale representations is further supported by spatial decay models in environmental acoustics, which emphasize that perceived noise levels arise from a combination of proximate and distant contributors. By integrating both, the framework is able to capture both the immediate noise landscape and the broader

structural backdrop, ensuring that predictions remain sensitive to local details while maintaining cross-domain generalizability (Tuia et al., 2016; Umar et al., 2024).

6.2.4.2 Remote Sensing-Derived Features

The multispectral imagery provided by the WorldView-2 satellite offers high-resolution spatial and spectral information on urban land cover, surface materials, and vegetation structure, all of which are critical for environmental noise modeling. In this study, remote sensing-derived features were generated through a two-phase process. First, a series of feature maps were created to capture texture, spectral contrast, and morphological patterns from the imagery. Second, these feature maps were summarized using zonal statistics calculated over spatial neighborhoods, producing a set of high-dimensional feature vectors that encode both local and contextual urban characteristics.

Feature Map Generation

To capture key spatial and spectral characteristics of urban surfaces, three main categories of feature maps were generated from each standardized satellite image. The first category comprised GLCM-based texture features (energy, homogeneity, and correlation) derived from three spectral bands (Green, NIR1, Red Edge) and three indices (NDVI, WVBI, RGI), using a 9×9 pixel window and 64-level grayscale quantization. The second category included z-score normalized contrast maps from eight inputs (Red, Green, NIR1, NDVI, NDWI, WVBI, RGI, BSI), computed over three spatial scales (15×15, 31×31, 77×77 pixels) to capture local to regional context. The third category consisted of morphological features (Local Binary Patterns, opening, and closing) derived from Red, Green, NIR1, NDVI, and BSI bands, emphasizing fine-scale textures and larger object continuity.

In total, 63 feature maps were produced per image: 24 from texture, 24 from contrast, and 15 from morphology (Table 6.5). These descriptors were chosen for their ability to highlight stable, illumination-invariant spatial patterns that support cross-city transferability, rather than for optimizing single-city accuracy. Detailed computational procedures are described in Section 3.3.5. Table 6.5 provides an overview of these feature map categories, input bands or indices, and the types of spatial patterns they capture.

Table 6.5 Overview of Remote Sensing-Derived Feature Maps

Category	Inputs	Key Outputs	Number of Maps
GLCM Texture Features	Green, NIR1, Red Edge, NDVI, WVBI, RGI	Energy, Homogeneity, Correlation	24
Z-Score Normalized Contrasts	Red, Green, NIR1, NDVI, NDWI, WVBI, RGI, BSI	Small, Medium, Large Spatial Context	24
Morphological Features	Red, Green, NIR1, NDVI, BSI	Local Binary Patterns, Opening, Closing	15

The introduction of morphological and texture-based descriptors such as the Local Binary Pattern (LBP) was not intended to replicate the feature design of earlier chapters but to enhance cross-domain comparability under heterogeneous imaging conditions. As detailed in Section 3.3.5 of the Methodology, the computational procedures for these feature maps—including grayscale quantization, rotation-invariant LBP encoding, and morphological filtering—were implemented following established remote-sensing practices.

In the context of cross-city prediction, these descriptors were selected for their ability to capture local, illumination-invariant contrast structures, thereby emphasizing textural patterns that remain stable across domains while suppressing global radiometric and contextual differences. This design choice aligns with the chapter's broader objective of achieving domain-invariant feature representation rather than maximizing within-city predictive accuracy. Consequently, while such features might also provide incremental benefits in single-city models, their inclusion here primarily serves to ensure feature-space normalization and transferability across cities with differing spectral and morphological characteristics.

Zonal Statistical Variable Extraction

Once the feature maps were generated, they were transformed into structured numerical inputs by computing zonal statistics over defined concentric spatial neighborhoods. This approach allowed the model to capture both fine-resolution and broad-scale spatial heterogeneity, aligning directly with the dual-branch architecture of the predictive framework.

Five annular zones were defined around each 30 m grid cell, reflecting the two complementary spatial scales: the local scale (0–120 m) included three nested rings (0–30 m, 30–60 m, 60–120 m), while the contextual scale (120–1000 m) included two broader zones (120–500 m, 500–1000 m). For each of these zones, twelve statistical descriptors were computed, capturing various aspects of central tendency, dispersion, distributional shape, and information complexity. Specifically, the calculated metrics included mean, standard deviation, median, interquartile range (IQR), 10th, 25th, 75th, and 90th percentiles, skewness, kurtosis, Shannon entropy, and weighted mean. The weighted mean employed inverse distance weighting relative to the grid centre, effectively mimicking the physical attenuation patterns of noise intensity with distance (Shepard, 1968; Lu and Wong, 2008). The use of ring-based zonal statistics to characterise multi-scale environmental structures is well established in urban and noise modelling research (Myint et al., 2011; Xie et al., 2011).

This comprehensive statistical treatment ensured that each feature map contributed sixty distinct numerical variables (five zones × twelve metrics), producing a high-dimensional but structured feature set that could robustly represent spatial-spectral heterogeneity. Overall, across all 63 feature maps, the remote sensing-derived inputs totaled 3,780 numerical variables per grid point, divided between the two model branches as follows:

- Local detail branch (0–120 m): 2,268 variables

- Contextual background branch (120–1000 m): 1,512 variables

Table 6.6 Summary of Zonal Statistics for Remote Sensing Features

Spatial Scale	Zones (m)	Statistical Descriptors (per zone)
Local Detail (0–120 m)	0–30, 30–60, 60–120	Mean, Standard Deviation, Median, Interquartile Range (IQR), Percentiles (10th–90th), Skewness, Kurtosis, Entropy, Weighted Mean
Contextual (120–1000 m)	120–500, 500–1000	Same as above

This multi-scale statistical aggregation enables the predictive model to integrate highly localised spatial patterns with broader contextual gradients, ensuring both detailed sensitivity and generalisable pattern recognition across complex urban environments. The rationale for such a design is rooted in well-established remote sensing and geospatial modelling research, which consistently highlights the complementary roles of local and contextual scales in capturing environmental variability (Schott et al., 1988; Helmer and Ruefenacht, 2005; Hu and Li, 2020). By aggregating statistical descriptors across nested annuli, the model effectively links fine-grained neighbourhood characteristics to broader landscape gradients, a principle shown to enhance predictive performance in urban environmental applications.

6.2.4.3 LULC-Derived Features

The Urban Atlas 2012 dataset offers harmonized, semantically rich land-use and land-cover (LULC) classifications across European urban areas, making it an essential foundation for extracting interpretable features in environmental noise modeling (EEA, 2018). In this study, LULC-derived features were carefully designed to complement the high-dimensional spectral variables by embedding categorical information on functional urban structures, potential noise sources, and environmental mitigation zones.

To ensure alignment with the dual-branch model architecture, all LULC metrics were computed over the same five concentric ring buffers used in the spectral feature design. The local detail branch (0–120 m) employed three inner zones (0–30 m, 30–60 m, 60–120 m), while the contextual background branch (120–1000 m) incorporated two outer zones (120–500 m, 500–1000 m). This multi-scale design ensured that both micro- and macro-scale environmental influences were captured (Aggarwal et al., 2023).

First, proportional area metrics were calculated by measuring the share of each of the 24 Urban Atlas classes within each zone, yielding 120 proportional variables per sample. These variables captured both fine-grained heterogeneity (such as local streets, parks, or industrial facilities) and broader spatial gradients (such as industrial

belts or green infrastructure), which are known to influence environmental noise exposure (Ajibola & Cabral, 2024).

Second, proximity effects were quantified using Euclidean distances from each grid cell to the nearest polygon of each LULC type, calculated separately for local and contextual scales. This produced 48 normalized distance variables, useful for modeling edge effects, such as the acoustic influence of nearby traffic corridors or the noise-buffering effects of vegetation.

Third, dominant-type variables were derived to capture the prevailing LULC class within the local and contextual zones and at the cell’s own location. These were one-hot encoded to ensure compatibility with neural network models and provided critical semantic cues for predictive learning (Ajibola & Cabral, 2024).

Finally, to enhance interpretability and reduce sparsity, the 24 original LULC classes were aggregated into four functional groups: strong noise sources, moderate noise sources, strong mitigation zones, and moderate mitigation zones. Table 6.7 summarizes the specific LULC codes assigned to each group, reflecting their assumed acoustic contributions or attenuation roles.

Table 6.7 Functional Grouping of LULC Classes for Noise Modeling

Functional Group	Urban Atlas Codes	Description
Strong Noise Sources	12210,12230,12300,12400,13100	Major roads, railways, ports, airports, extractive sites
Moderate Noise Sources	12220, 12100	Secondary roads, industrial/public/military areas
Strong Mitigation Zones	21000, 31000	Arable land, forests (strong environmental buffers)
Moderate Mitigation Zones	14100,14200,23000,32000, 33000	Urban green areas, pastures, herbaceous cover, leisure areas, open spaces

The aggregated variables were then summarized as area proportions over both local and contextual zones, yielding eight additional predictors. Table 6.8 provides an overall summary of the LULC-derived feature categories and their variable counts, highlighting the comprehensive and multi-layered design of this feature set.

Table 6.8 Summary of LULC-Derived Features

Feature Category	Description	Number of Variables
Proportional Area Metrics	24 LULC class proportions over 5 zones	120

Feature Category	Description	Number of Variables
Distance to Nearest LULC	Euclidean distances (local, contextual) to 24 classes	48
Categorical Dominant-Type	Dominant local, dominant contextual, grid-level class (one-hot)	3 (expanded in input)
Functional Group Aggregates	Aggregated area proportions (noise source, mitigation) over 2 scales	8

Together, these LULC-derived features provided an interpretable, semantically grounded complement to the purely spectral variables, ensuring that the model could capture both the physical structure and functional dynamics of urban environments when predicting environmental noise patterns.

6.2.4.4 Feature Selection Strategy

Given the high dimensionality of the constructed feature sets—comprising thousands of variables derived from remote sensing and LULC data—a carefully designed two-stage feature selection strategy was implemented. The overarching goal was to balance representational richness with computational efficiency, ensuring that the retained features maximised cross-city generalisability while minimising redundancy and overfitting risks. This approach is consistent with well-established principles in feature selection research (Guyon and Elisseeff, 2003; Peng, Long and Ding, 2005) and builds on recent advances in dimensionality reduction for remote sensing applications (Persello and Bruzzone, 2010).

Before selection, the local detail branch (0–120 m) included 2,369 continuous features, and the contextual background branch (120–1000 m) contained 1,590 continuous features. These encompassed numerical descriptors from remote sensing statistics (GLCM, morphological, z-score transforms) and LULC-based spatial metrics (proportional areas, distances). Categorical variables (such as dominant LULC class) and functional group indicators were intentionally excluded from filtering to preserve semantic interpretability.

The first stage employed a correlation-based redundancy removal using Pearson correlation. Pairwise correlations were computed within each branch, and for any feature pair exceeding a correlation coefficient of 0.95, only one representative was retained. This filtering approach was selected over rank-based alternatives (such as Spearman or Kendall) because the input variables and the target variable (mean dBA) were continuous and ratio-scaled, making linear correlation measures more appropriate for identifying direct redundancies. Moreover, Pearson correlation offered higher computational efficiency in large large datasets, aligning well with subsequent modeling approaches like random forests, which also assume continuous input relationships (Breiman, 2001). As a result, approximately 40% of features were

removed from the local branch and over 50% from the contextual branch, effectively reducing dimensionality while preserving core signal diversity.

In the second stage, a Random Forest Regressor was independently trained for each spatial branch using the labeled data from Southampton. Feature importance was calculated based on the mean decrease in impurity across decision trees, leveraging the non-parametric and nonlinear modeling strength of random forests to capture complex variable interactions. The top 80 most important numerical features were retained per branch, alongside categorical and functional group variables, ensuring balanced representation.

Table 6.9 Summary of Final Feature Set Composition

Branch	Numerical Features (Selected)	Categorical Features (LULC)	Functional Group Variables	Total Features
Local Detail (0–120 m)	80	Dominant class, cell-level class (2)	4	86
Contextual (120–1000 m)	80	Dominant class (1)	4	85

All numerical variables were standardized prior to model training to ensure consistency and stability across domains. A comprehensive list of selected features, including variable names, spatial origins, and transformation methods, is provided in Appendix 3 for full reproducibility.

This two-stage feature selection strategy offers a scalable and interpretable framework, enhancing the robustness of graph-based environmental noise modeling across diverse urban landscapes. By integrating correlation-based pruning with importance-driven refinement, the final feature sets strike a balance between representational power and computational efficiency, directly supporting the cross-city generalization goals of this study.

6.2.5 Pseudo-Label Construction Based on LULC-Derived Predictors

To address the scarcity of measured noise data in target cities, a high-confidence pseudo-labeling framework was developed, leveraging inter-city similarity in land-use and land-cover (LULC)-based predictors. This design draws on transferable spatial correlations identified in the reference city (Southampton) to infer plausible noise levels across unlabeled regions in target cities. By integrating robust variable selection, cross-domain similarity matching, and confidence-based filtering, the approach creates reliable supervisory signals for training models in data-scarce urban environments (Ajibola & Cabral, 2023; Kage et al., 2024).

6.2.5.1 Variable Selection Rationale

The variables selected for pseudo-label generation were chosen through a rigorously stratified, two-step process that built upon the feature selection strategy described in Section 6.2.4.4. First, all continuous variables were filtered using correlation-based redundancy removal to minimize multicollinearity. Subsequently, a Random Forest

regression model was trained using the labeled Southampton data, and variables were ranked by importance separately for the local (0–120 m) and contextual (120–1000 m) branches, ensuring balanced emphasis on scales (Breiman, 2001).

Although the top ten most important predictors in each branch included both remote sensing-derived and LULC-derived variables, the final pseudo-labeling framework deliberately prioritized LULC-derived indicators. This decision was based on two critical considerations: (i) LULC variables, as shown in extensive land cover mapping research, tend to generalize more robustly across cities because they encode semantic environmental meaning that is less sensitive to spectral or sensor variation; and (ii) land use patterns—such as the proximity to transport infrastructure or the density of industrial zones—are among the most consistent predictors of urban noise levels across global contexts.

Table 6.10 summarizes the ten LULC-derived variables ultimately selected for pseudo-labeling inference. These include six from the local branch and four from the contextual branch, ensuring that both micro- and macro-scale environmental influences were captured.

Table 6.10 Selected Variables for Pseudo-Labeling Inference

Branch	Variable Name	Description
Local	landuse_12220_r30	Share of rail network within 30 m
Local	local_moderate_noise_ratio	Proportion of moderate noise source areas
Local	landuse_12220_r60	Share of rail network within 60 m
Local	local_moderate_mitigation_ratio	Proportion of moderate mitigation zones
Local	landuse_11210_r60	Discontinuous dense urban fabric within 60 m
Local	local_dist_12220	Distance to rail network (local scale)
Contextual	global_dist_12220	Distance to rail network (background scale)
Contextual	global_dist_12100	Distance to highway network
Contextual	global_moderate_noise_ratio	Share of contextual moderate noise source zones
Contextual	global_dist_14100	Distance to green urban areas

By focusing on interpretable, semantically grounded LULC features, the pseudo-labeling approach ensured that the inferred labels were based on robust, cross-domain signals rather than dataset-specific quirks. This design increases the likelihood that labels transferred across cities reflect meaningful environmental similarities, strengthening the foundation for subsequent domain-adaptive learning.

6.2.5.2 Similarity-Based Pseudo-Label Inference

To generate plausible noise estimates in target cities where no direct measurements were available, the study implemented a carefully designed k-nearest neighbors (kNN) inference framework. This method, widely used in semi-supervised learning and

environmental modeling (Cover and Hart, 1967; Li and Heap, 2014), relies on local neighborhood similarity to extrapolate reliable labels from a reference domain.

For each grid point in the target cities, a standardized feature vector was constructed using the ten selected LULC-derived variables identified in Section 6.2.5.1. These vectors were normalized using z-score transformation, with normalization parameters (mean and standard deviation) computed exclusively from the labeled Southampton data. This standardization ensured cross-domain comparability and harmonized the feature space, allowing the model to evaluate target-city samples relative to the Southampton reference distribution (Patel et al., 2015).

Within this normalized feature space, each target sample was matched to its five most similar neighbors among the Southampton-labeled points using Euclidean distance. The mean dBA value of these neighbors was then assigned as the pseudo-label, reflecting a locally weighted consensus that captures both spatial and environmental similarity.

To further ensure pseudo-label reliability, two quality assurance metrics were calculated for each target point. First, the mean Euclidean distance between neighbors provided a similarity-based confidence score, quantifying the strength of neighborhood matching. Second, the standard deviation of neighbor noise values was computed as a stability indicator, flagging heterogeneous or complex local conditions that could reduce label reliability (Sener and Savarese, 2018).

6.2.5.3 High-Confidence Sample Filtering

Once the ten high-importance LULC-derived variables were finalized, a similarity-based pseudo-labeling procedure was employed to infer plausible noise levels in the unlabeled target cities. For each grid point, a normalized feature vector was constructed using z-score standardization derived from the Southampton reference dataset. This calibration ensured that the target city's environmental attributes were evaluated in a harmonized feature space, mitigating bias from local scaling differences.

Within this standardized space, each target grid point was matched to its five most similar neighbors from the Southampton-labeled dataset using Euclidean distance as the similarity metric. The mean dBA value across these five neighbors was then assigned as the pseudo-label, effectively yielding a locally weighted consensus estimate.

Crucially, pseudo-label assignment was accompanied by two quality assurance measures. First, the average Euclidean distance among the five neighbors was recorded as a similarity-based confidence score, quantifying how closely a target point matched known reference samples. Second, the standard deviation of neighbor noise levels captured local stability, flagging environmental heterogeneity or mismatch that could undermine label reliability (Sener and Savarese, 2018; Grandvalet and Bengio, 2005).

To ensure only robust pseudo-labels contributed to model training, a dual-threshold filtering protocol was applied. A target sample was retained as high-confidence only if its similarity-based confidence score fell within the top 10% of all pseudo-labeled points in the target city and the standard deviation of its neighbor labels remained below 10 dBA. This conservative thresholding strategy jointly maximised label precision and spatial stability, effectively safeguarding against noisy or structurally mismatched inferences.

Table 6.11 Dual-Threshold Filtering Criteria

Criterion	Threshold Applied	Purpose
Similarity-based confidence	Top 10% (above 90th percentile in target city)	Ensure structural match to reference domain
Neighbor label stability	Standard deviation < 10 dBA among five neighbors	Filter out locally unstable or heterogeneous zones

6.2.5.4 Summary of Pseudo-Label Generation Across Target Cities

After applying the full pseudo-labeling and dual-threshold filtering pipeline, a subset of high-confidence pseudo-labeled samples was retained in each of the four target cities. Although each city originally contained tens of thousands of unlabeled grid points, only approximately 9–10% of these points passed the strict confidence and stability filters, highlighting the conservative nature of the approach.

Table 6.12 summarizes the pseudo-labeling outcomes across all target domains. This table presents, for each city, the total number of analyzed grid points, the number of high-confidence pseudo-labels retained after filtering, and the final retention rate as a percentage.

Table 6.12 High-Confidence Pseudo-Label Summary Across Target Cities

City	Total Grid Points	High-Confidence Pseudo-Labels	Retention Rate (%)
Cardiff	188,480	17,595	9.34
Liverpool	187,555	18,199	9.70
Nottingham	153,972	14,732	9.56
Portsmouth	51,680	4,935	9.55

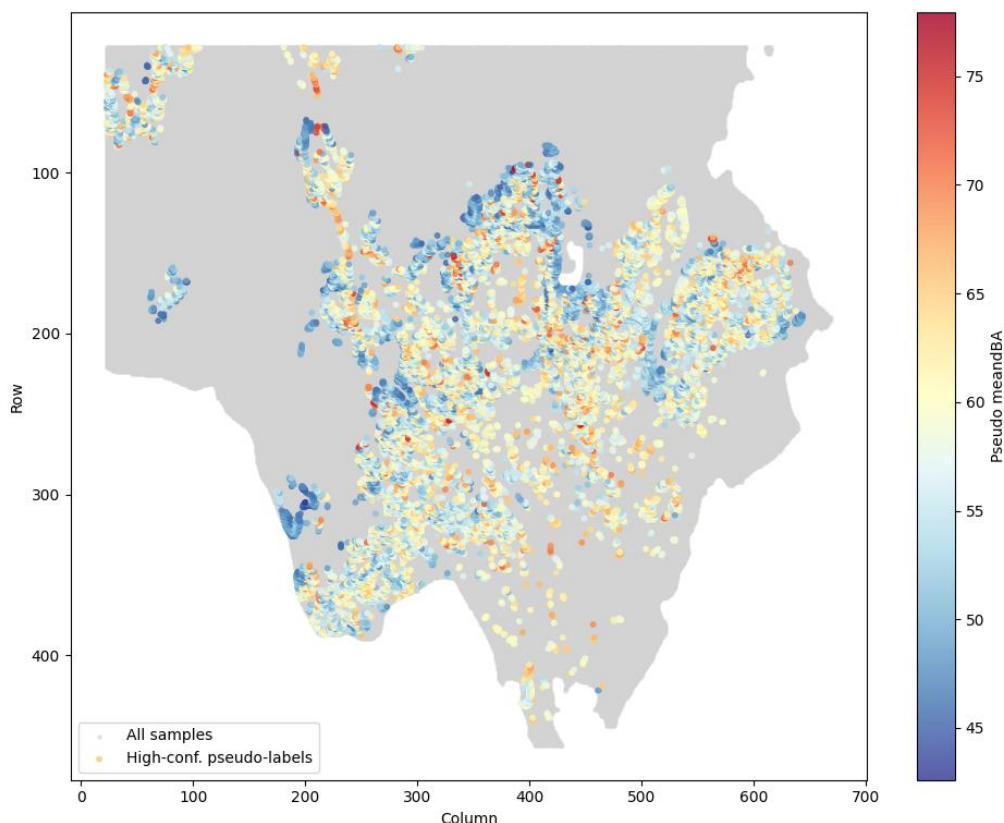
These figures demonstrate a striking consistency in retention rates across cities, despite differences in their spatial extent, land-use diversity, and urban morphology. Such consistency indicates that the pseudo-labeling framework was not biased toward particular urban contexts, but instead systematically selected structurally robust samples across domains.

Importantly, while the absolute number of pseudo-labeled points varied with city size, the proportion retained remained tightly clustered around 9–10%. This balance between coverage and selectivity was crucial: it ensured that the training set included a meaningful spatial representation of each city without compromising label quality due to overly aggressive or lax filtering thresholds.

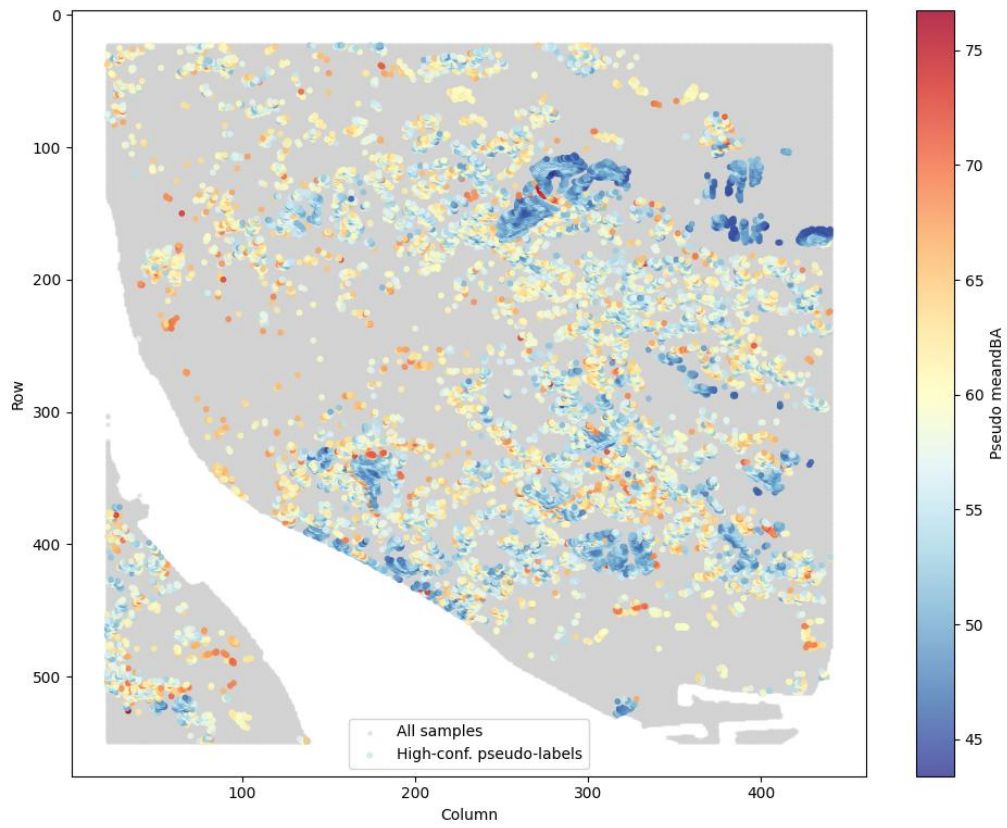
6.2.5.5 Illustration of Pseudo-Label Distribution

To visualize the spatial coverage and noise prediction diversity generated through the pseudo-labeling process, spatial maps were produced for each target city. These maps displayed all grid points, distinguishing between unlabeled points (rendered in neutral gray) and high-confidence pseudo-labeled points (color-coded using a blue-to-red gradient corresponding to predicted dBA values).

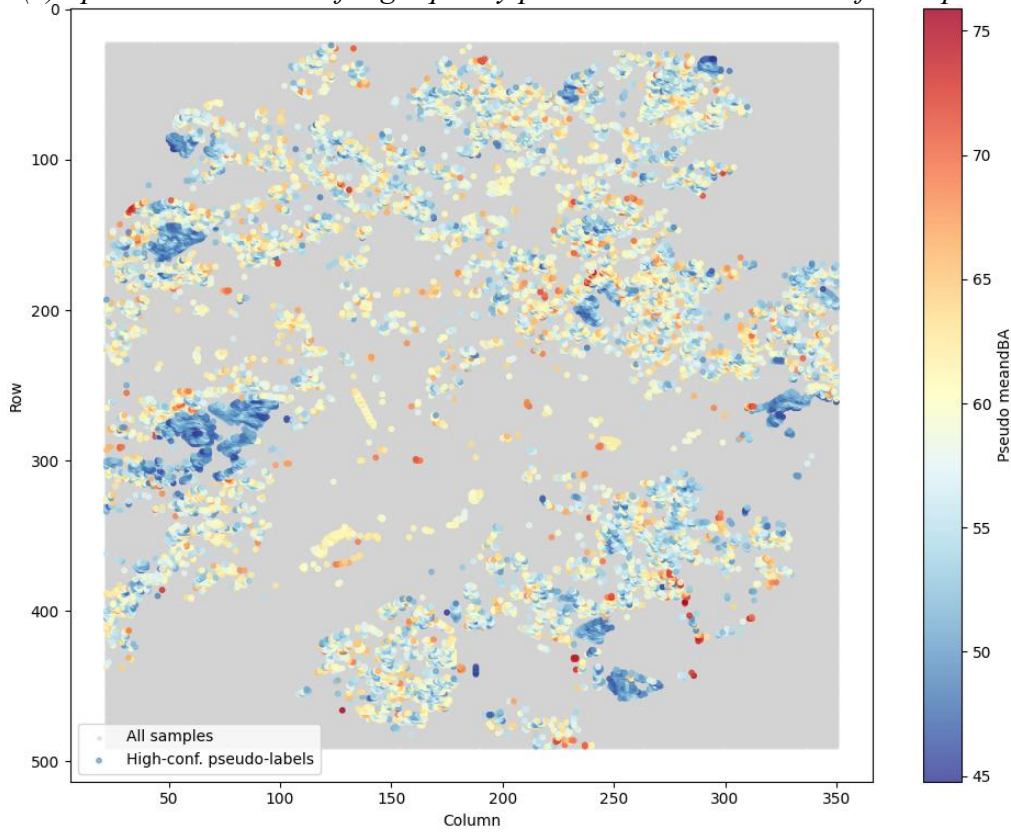
Figure 6.3a-d presents spatial visualizations for each target city, showing both unlabeled and high-confidence pseudo-labeled points. Unlabeled grid points are rendered in grey, while pseudo-labeled points are shown using a blue-to-red gradient corresponding to predicted dBA values. Each row corresponds to one city, enabling direct comparison of spatial coverage and predicted noise variation.



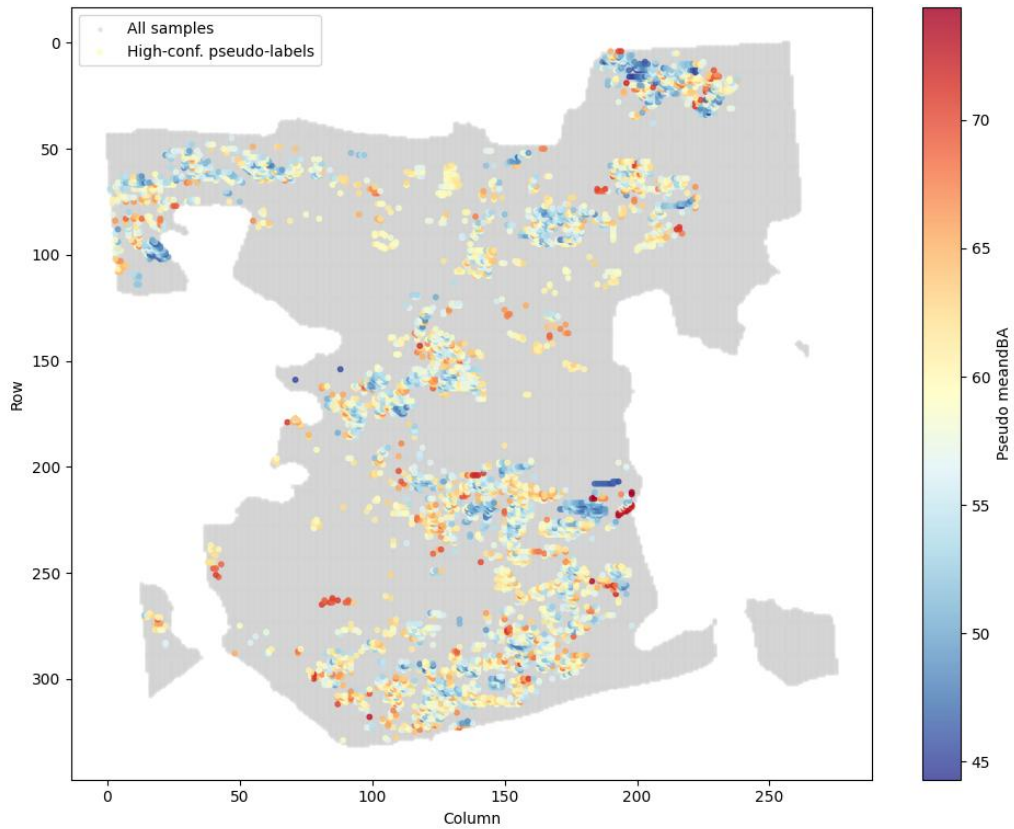
(a) Spatial distribution of high-quality pseudo-labels in domain of Cardiff



(b) *Spatial distribution of high-quality pseudo-labels in domain of Liverpool*



(c) *Spatial distribution of high-quality pseudo-labels in domain of Nottingham*



(d) *Spatial distribution of high-quality pseudo-labels in domain of Portsmouth*
 Figure 6.3 *Spatial Distribution of High-Confidence Pseudo-Labels Across Target Cities*

These visualizations revealed that the retained pseudo-labels were spatially well-distributed across diverse urban settings, consistently aligning with expected environmental noise patterns such as proximity to transport corridors, industrial belts, and dense residential zones. Furthermore, the blue-to-red gradient provided intuitive insight into the predicted noise intensity, facilitating rapid cross-city comparison of spatial noise variability.

Overall, the combined quantitative summary (Table 6.12) and qualitative visualizations (Figure 6.3) confirmed that the pseudo-labeling framework successfully identified a robust subset of informative, high-confidence samples. These samples served as a critical foundation for the subsequent development and validation of generalizable, domain-adaptive urban noise prediction models.

6.2.6 Model Architecture and Training Procedure

6.2.6.1 Model Overview and Dual-Branch Design

To predict urban noise levels across cities characterized by highly heterogeneous urban forms and environmental settings, this study developed a customized dual-branch graph neural network (GNN) architecture. While the general framework builds

upon the single-domain GNN structures introduced in Chapter 5, it incorporates several critical enhancements aimed at improving cross-domain generalization and multi-scale environmental modeling.

The network consists of two parallel, independently parameterized branches, each designed to process features at distinct spatial scales. The local branch focuses on fine-scale morphology and immediate noise sources (0–120 m), employing Graph Attention Networks (GAT) to adaptively weight neighbors and capture localized heterogeneity (Veličković et al., 2018). In contrast, the background branch targets broader structural context (120–1000 m), leveraging Graph Convolutional Networks (GCN) to aggregate global patterns and long-range dependencies (Kipf, 2016). This dual-scale separation is consistent with established acoustic modelling frameworks, which specify how both near-field built-environment factors (e.g., distance, ground effects, barriers, façade reflections) and wider neighbourhood structure jointly shape outdoor noise propagation (ISO, 1996; Kephalopoulos et al., 2012; WHO, 2018).

Each branch comprises multiple graph processing layers: the local branch uses two GAT layers to capture progressively larger receptive fields, whereas the background branch employs a single GCN layer for structural simplicity. Regularization strategies are branch-specific: batch normalization stabilizes local feature distributions (Ioffe and Szegedy, 2015), while layer normalization controls variance across heterogeneous graph structures (Ba et al., 2016).

After spatial encoding, both branches are projected through lightweight multi-layer perceptrons (MLPs) into a shared 64-dimensional latent space, then fused via a learned gating layer that adaptively balances local and global contributions. This allows the network to dynamically determine the relative importance of fine-scale versus contextual information for each grid cell.

By decoupling local heterogeneity from background environmental structure, the dual-branch architecture aligns with recent advances in multi-scale graph modeling (Wu et al., 2020) and significantly enhances the model’s ability to generalize across cities with divergent urban layouts, land-use compositions, and noise generation profiles.

6.2.6.2 Graph Construction and Edge Weight Modulation

The foundation of the model’s predictive architecture is a spatial graph built over the 30 m resolution grids generated for each study city. Each node in this graph represents a grid point, embedding two parallel sets of feature vectors corresponding to the local branch (0–120 m neighborhood scale) and the background branch (120–1000 m neighborhood scale). This graph design enables the model to capture both localized and broader contextual influences on urban noise propagation, which is essential for accurately representing heterogeneous urban environments.

The graph’s structure is defined by spatial adjacency relationships between grid points. Specifically, the spatial coordinates (i.e., row and column indices) of each node are used to identify immediate neighboring nodes through ± 1 shifts in four cardinal

directions (north, south, east, west), effectively establishing a base 4-neighbor system with a 30 m step size. To extend each node’s receptive field and incorporate more distant contextual interactions, a 3-hop connectivity strategy is applied, allowing nodes to access neighbors up to three steps away while maintaining overall spatial coherence.

Each edge in the graph is initially assigned a weight based on a squared exponential distance decay function:

$$w(d) = \exp \left(- \left(\frac{d}{\sigma} \right)^2 \right)$$

where d represents the Euclidean distance between nodes i and j , and σ defines the decay width parameter.

Unlike the exponential decay in Chapter 5, which was calibrated specifically for Southampton’s low-frequency dominated acoustic environment($\alpha= 0.06$), the squared exponential form provides a smoother and more gradual attenuation profile. This ensures that distant nodes are down-weighted more consistently across cities, reducing the risk of site-specific bias. In practice, this change was motivated by the need for cross-city generalization: while exponential decay reflected Southampton’s unique maritime noise profile, squared exponential decay avoids overfitting to any single city’s propagation physics, thereby providing a more robust kernel for multi-city applications.

The squared exponential decay can be interpreted as a Gaussian-like attenuation, which has a firmer grounding in acoustic diffusion theory. Unlike pure exponential decay (which implies constant absorption per unit distance), the squared exponential emphasizes rapid decay at short ranges and smoother tails at long ranges, consistent with how complex urban environments — with multiple reflections, diffractions, and absorptive surfaces — diffuse noise energy over space.

To enhance numerical stability and ensure compatibility with subsequent learning modules, the raw Gaussian weights are further transformed using a sigmoid-based rescaling function, constraining all edge weights to fall within the range [0.1, 2.0].

While physical proximity is a critical factor shaping noise propagation, urban noise dynamics are also heavily influenced by the environmental characteristics of the surrounding landscape. To address this, the model incorporates a dynamic edge modulation mechanism, wherein edge weights are adaptively adjusted based on a subset of the most predictive environmental features. From the filtered feature sets (after correlation removal and random forest-based importance ranking), the top five most important variables are selected separately from the local and background branches, resulting in ten critical modulation features overall (Table 6.13).

Table 6.13 Top 10 Environmental Features Used for Edge Weight Modulation

Feature Name	Description
--------------	-------------

Feature Name	Description
landuse_12220_r30	Proportion of minor road areas within 0–30 m of the grid point
local_moderate_noise_ratio	Proportion of moderate noise sources (e.g., minor roads, industry) within 0–120 m
landuse_12220_r60	Proportion of minor roads within 30–60 m
local_moderate_mitigation_ratio	Proportion of moderate mitigation zones (e.g., parks, grasslands) within 0–120 m
landuse_11210_r60	Proportion of discontinuous dense urban fabric within 30–60 m
background_dist_12220	Distance to other roads (land use 12220) within 120–1000 m
background_dist_12100	Distance to industrial, commercial, or public areas within 120–1000 m
Green_contrast_r500_percentile75	75th percentile of Green band-based contrast texture within 120–500 m, reflecting vegetation heterogeneity
background_moderate_noise_ratio	Proportion of moderate noise sources within 120–1000 m
WVBI_correlation_r1000_percentile75	75th percentile of WVBI index's GLCM correlation within 500–1000 m, indicating surface material uniformity

To integrate these variables into the graph, each source node along an edge feeds its ten environmental values into a two-layer neural network module (EdgeWeightLayer), which predicts a learned noise influence indicator. The final dynamic edge weight is computed as:

$$w_{ij}^{final} = w_{ij} \times (1 + WdgeWeightLayer(x_i))$$

where x_i is the feature vector from node $\mathbf{z}$, and w_{ij} is the base Gaussian weight. Edges whose adjusted weights fall below a threshold of 0.2 are pruned from the graph, effectively removing weak or environmentally insignificant connections. This pruning step improves the sparsity and interpretability of the graph, ensuring that only the most structurally meaningful pathways contribute to noise information propagation.

To integrate these variables into the graph, each source node along an edge feeds its ten environmental values into a two-layer neural network module (EdgeWeightLayer), which predicts a learned noise influence indicator. The final dynamic edge weight is computed as:

$$w_{ij}^{final} = f(x_i, x_j) \cdot w(d)$$

where $f(x_i, x_j)$ is the modulation factor derived from environmental features and $w(d)$ is the squared exponential distance kernel. Edges whose adjusted weights fall below a threshold of 0.2 are pruned from the graph, effectively removing weak or environmentally insignificant connections.

In summary, the transition from an exponential decay in Chapter 5 to a squared exponential decay in Chapter 6 reflects a deliberate methodological shift. In Chapter 5, the attenuation function was designed specifically for Southampton, with the coefficient α empirically tuned to 0.06 in order to reflect the city’s low-frequency, long-range maritime noise environment. By contrast, Chapter 6 required a formulation that could generalize across multiple cities, avoiding dependence on site-specific propagation dynamics. The adoption of a squared exponential decay therefore provided a more robust and less context-dependent attenuation kernel. This evolution illustrates the balance between grounding the model in physically plausible assumptions at the site level and achieving the cross-domain adaptability necessary for constructing transferable urban noise prediction frameworks.

6.2.6.3 Node Feature Encoding and Dual-Branch Message Passing

Once the spatial graphs were constructed, each node was associated with two distinct sets of features: a local feature set capturing fine-scale (0–120 m) variations, and a global feature set representing broader contextual (120–1000 m) urban characteristics. This dual-feature arrangement mirrors the two-branch network architecture and is designed to promote effective decoupling between localized noise influences and regional background patterns, an approach theoretically grounded in recent graph neural network (GNN) studies.

The local branch is responsible for modeling detailed spatial interactions within the immediate 0–120 m neighborhood. To achieve this, a two-layer Graph Attention Network (GAT) architecture was implemented. The core advantage of GAT lies in its adaptive neighbor weighting mechanism: for each node, learned attention coefficients allow the network to prioritize more relevant neighbors, enhancing the representation of local heterogeneity. Mathematically, at each layer l , the node representation $h_i^{(l)}$ is updated as:

$$h_i^{(l)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l-1)} \right)$$

where $\alpha_{ij}^{(l)}$ are the normalized attention coefficients, $W^{(l)}$ is the learned transformation matrix, and σ denotes the non-linear activation function (Exponential Linear Unit, ELU). After each GAT layer, batch normalization is applied to stabilize training and mitigate overfitting, especially important when learning from diverse local neighborhoods.

The background branch, in contrast, employs a single-layer Graph Convolutional Network (GCN) to aggregate information across the broader 120–1000 m background context. GCN operates by averaging neighboring node features, smoothing local variations to emphasize larger-scale structural patterns. Formally, the node representation is computed as:

$$h_i = \sigma \left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{1}{\sqrt{d_i d_j}} W h_j \right)$$

where d_i and d_j denote the degrees of nodes i and j , respectively, and W is the learnable weight matrix. Following convolution, layer normalization is applied to improve generalization across heterogeneous graphs, a technique shown to enhance stability in recent graph modeling research (Ba et al., 2016).

Once local and background features are processed by their respective branches, the resulting embeddings are concatenated and passed through a fusion adapter composed of a linear compression layer, ReLU activation, and dropout (rate 0.3) to prevent overfitting. This step reduces the dimensionality of the combined representation while preserving complementary information from both spatial scales (Srivastava et al., 2014).

To further enhance adaptability, an adaptive gating mechanism is introduced to balance the contributions of local and global information dynamically. Specifically, the final intermediate embedding h_{fused} is computed as:

$$h_{\text{fused}} = \sigma(W_g[h_{\text{local}} \parallel h_{\text{global}}])$$

where W_g is a learnable linear transformation, $[\cdot \parallel \cdot]$ denotes vector concatenation, and σ is the sigmoid activation function. This mechanism enables the model to adjust its reliance on fine-scale versus broad-scale features for each node, effectively tailoring predictions to the local urban context (Dauphin et al., 2017).

Finally, the fused embedding h_{fused} is passed through a fully connected output layer to predict the mean noise level (dBA) at each node. In summary, this architecture integrates highly localized, detail-sensitive information (via GAT) with smooth, large-scale contextual patterns (via GCN), while the gating mechanism ensures dynamic balancing of multi-scale contributions, making the system robust across diverse urban environments.

6.2.6.4 Adversarial Domain Alignment Strategy

To strengthen the model's capacity for cross-city generalization, this study incorporated a multi-domain adversarial alignment strategy during training. This approach aims to mitigate distributional shifts between urban domains by learning domain-invariant feature representations, a technique rooted in adversarial learning and now extensively reviewed in recent graph learning surveys (Wu et al., 2020).

The core component of the strategy is a lightweight multi-domain discriminator, designed to distinguish among the five studied cities: Southampton (reference) and Cardiff, Portsmouth, Nottingham, and Liverpool (targets). The discriminator consists of a fully connected hidden layer (64 units, ReLU activation) and a final classification

layer that outputs logits for the five domains. Formally, for each node's fused intermediate embedding h_{fused} , the domain logits d are predicted as:

$$d = W_d h_{\text{fused}} + b_d$$

where W_d and b_d are the learnable weight matrix and bias, respectively.

To achieve domain confusion, the embeddings pass through a Gradient Reversal Layer (GRL) before reaching the discriminator. The GRL acts as an identity function during the forward pass but multiplies gradients by a negative scalar during backpropagation, effectively reversing the optimization direction. This forces the main model to extract features predictive of noise but agnostic to domain identity. Mathematically, the transformation is:

$$\text{GRL}(h) = h \quad \text{and} \quad \frac{\partial \text{GRL}(h)}{\partial h} = -\lambda$$

where λ controls the strength of domain confusion, set empirically to 1.0 in this study.

To prevent the destabilization of early-stage learning by noisy domain signals, a preheat strategy was employed. For the first five epochs, only the primary regression loss was optimized, allowing the model to establish a basic noise prediction capacity. After epoch five, the adversarial domain loss and additional regularization terms were introduced, a staged approach inspired by best practices in domain adaptation (Ganin et al., 2016).

In parallel, a Maximum Mean Discrepancy (MMD) regularization term was integrated to align feature distributions between the reference and each target city. MMD quantifies the distance between two distributions in a reproducing kernel Hilbert space (RKHS), using a Gaussian kernel with multiple bandwidths to increase flexibility. The MMD loss between the source (S) and target (T) distributions is computed as:

$$\text{MMD}(S, T) = \frac{1}{|S|} \sum_{s \in S} \phi(s) - \frac{1}{|T|} \sum_{t \in T} \phi(t)$$

where $\phi(\cdot)$ is the feature mapping function (Gretton et al., 2012). This alignment mechanism complements adversarial learning by explicitly minimizing global distributional gaps, especially important for subtle cross-domain discrepancies (Ouyang et al., 2023).

The total training loss integrates three components:

1. The regression loss L_{reg} , implemented as a Huber loss over real labeled samples.
2. The adversarial domain loss L_{adv} , computed as cross-entropy over domain predictions.
3. The MMD loss L_{mmd} , penalizing feature distribution misalignment.

Thus, the final objective function is formulated as:

$$L_{\text{total}} = L_{\text{reg}} + \alpha L_{\text{adv}} + \beta L_{\text{mmd}}$$

where α and β are empirically determined hyperparameters balancing the contributions of domain alignment and distributional regularization.

Rather than focusing on architectural complexity, the emphasis of this strategy lies in learning robust, transferable representations—feature embeddings that preserve the acoustic relationships encoded in the source domain while remaining insensitive to city-specific biases.

In practical terms, a robust transferable representation is one that continues to encode physically meaningful noise–environment relationships—such as the attenuation of noise with increasing vegetation or the amplification along major transport corridors—even when the spatial context, land-use composition, or imaging conditions differ across cities.

These representations are robust because they remain stable under domain shifts (e.g., from coastal to inland morphology), and transferable because they enable accurate predictions in cities where no direct noise measurements are available.

By jointly optimizing the regression, adversarial, and MMD objectives, the model is encouraged to extract embedding patterns that (i) align across domains in latent space, (ii) maintain correlation with environmental predictors relevant to noise propagation, and (iii) suppress superficial differences caused by illumination, spectral range, or regional planning style.

Consequently, the learned feature space captures universal acoustic–morphological dependencies—for example, the consistent influence of built-up density, vegetation fraction, and road proximity on noise levels—rather than overfitting to the spectral or structural idiosyncrasies of any single city.

This conservative design avoids over-parameterization while ensuring interpretability and stability, positioning the learned embeddings as a robust foundation for the cross-city inference experiments presented in Section 6.3.

6.2.6.5 Dataset Partitioning Strategy for Source and Target Domains

The partitioning of training, validation, and testing subsets plays a critical role in ensuring that the learning framework remains both statistically rigorous and methodologically coherent across domains. Unlike the preceding chapters, which focused exclusively on within-city model training, the present framework extends to multi-domain learning, where only the source domain contains measured noise labels and all target domains rely on pseudo-labels or remain unlabeled. This shift from classical supervised learning to a hybrid weakly supervised domain adaptation setting necessitated a carefully structured partitioning strategy that reflects the epistemic quality of labels across different domains.

Within the source domain (Southampton), the dataset was partitioned using a 60–20–20 split for training, validation, and testing, respectively. This strategy is consistent with the node masking protocol described in Chapter 5, ensuring that validation and

test nodes participate in graph message passing without contributing to the loss function. Such masking is crucial in graph-based learning because nodes in different subsets remain topologically connected. By withholding labels from validation and test nodes, the model can aggregate their contextual information without risking label leakage. This partitioning not only preserves the statistical independence of evaluation but also enables the model to capture broader spatial correlations across the urban fabric, ensuring that predictive performance is evaluated on nodes that are structurally integrated yet label-inaccessible during training.

For the target domains (Cardiff, Portsmouth, Nottingham, Liverpool), the epistemic status of the labels differs fundamentally. All supervisory signals originate from pseudo-labeling, a process inherently less certain than direct in-situ measurements. In this context, forcing the creation of a held-out “test set” based on pseudo-labels would provide little meaningful information about generalization performance, since both the training and test subsets would reflect the same underlying inferential uncertainty rather than independent observations. Instead, the target domain data were divided into 80 % pseudo-labeled training samples and 20 % pseudo-labeled validation samples, while the remaining unlabeled nodes remained embedded within the graph structure to contribute indirectly through message passing and domain alignment. This design allows the model to learn from a larger pool of pseudo-labeled data while still maintaining an internal mechanism for monitoring overfitting to pseudo-label distributions through the validation split.

The absence of a pseudo-labeled test set in target domains is a deliberate and principled decision. Unlike the source city, where evaluation against measured noise values provides an objective accuracy benchmark, pseudo-labels do not constitute independent ground truth and should not be used as a surrogate for generalization assessment. Instead, generalization in this chapter is evaluated by aligning target domains through adversarial and statistical domain adaptation and subsequently validating the transferred model against the real labels in Southampton. In other words, model reliability is anchored in the source domain, while the structural adaptability of learned representations is encouraged through the inclusion of pseudo-labeled and unlabeled target samples.

This hybrid partitioning design reflects a balance between maximizing the learning signal from limited labeled data, mitigating the risks of overfitting to uncertain pseudo-labels, and preserving graph connectivity for domain adaptation. It also aligns with broader methodological principles in semi-supervised and domain-adversarial learning, where evaluation is performed on trusted labels while pseudo-labeled domains serve primarily to shape the latent representation space rather than to provide definitive accuracy estimates. By adopting this configuration, the framework ensures that the predictive model is trained on both reliable and structurally informative data, achieving a robust compromise between statistical validity, representational generalization, and cross-domain transferability.

Summary

The model architecture and training strategy developed in this chapter represent a systematic response to the fundamental challenge of transferring environmental noise prediction capabilities across morphologically and functionally distinct cities.

Building on the single-domain GNN framework introduced in Chapter 5, the current design expands the scope of the methodology through dual-branch graph processing, dynamic edge weighting, domain-adversarial alignment, and an epistemically informed dataset partitioning strategy. These components are not isolated technical choices but mutually reinforcing elements of a coherent framework explicitly tailored for cross-city generalization.

At the architectural level, the dual-branch design reflects the inherently multi-scale nature of urban acoustic environments. The local branch, based on Graph Attention Networks, enables fine-grained modeling of micro-environmental variations such as road adjacency, façade density, and vegetation buffering, while the contextual branch, implemented with Graph Convolutional Networks, captures large-scale structural patterns and background acoustic influences. This explicit scale decomposition mirrors the physical reality that environmental noise results from the interaction of proximate and distal sources, enhancing both interpretability and robustness.

The graph construction and dynamic edge modulation mechanisms extend this multi-scale representation by embedding environmental heterogeneity directly into the topological structure of the model. By replacing the single-city exponential decay kernel of Chapter 5 with a squared exponential kernel, the framework deliberately shifts from site-specific acoustic assumptions to a more generalizable attenuation function that is better suited for cross-domain learning. Environmental modulation of edge weights further strengthens the model’s capacity to reflect the spatial logic of noise propagation across diverse urban morphologies.

The training procedure integrates adversarial domain alignment and statistical distribution matching to explicitly minimize cross-domain discrepancies in the learned latent space. Through the combination of a gradient reversal–based domain discriminator and maximum mean discrepancy regularization, the model learns to retain acoustically meaningful relationships while suppressing city-specific spectral and structural biases. This alignment is critical for enabling the model to operate effectively in cities where no ground-truth noise data exist.

Finally, the dataset partitioning strategy embodies a methodological distinction between the epistemic certainty of labeled data and the inferential nature of pseudo-labels. By applying a 60–20–20 split in the labeled source domain and an 80–20 split in pseudo-labeled target domains without enforcing a pseudo-test set, the framework preserves statistical validity where real observations exist while maximizing structural learning where they do not. This principled treatment of label quality ensures that evaluation remains grounded in trusted data while the representation space is shaped by broader domain structure.

Taken together, the components of this model architecture and training procedure create a theoretically grounded and practically scalable foundation for cross-city environmental noise prediction. The dual-branch structure encodes multi-scale acoustic processes; the graph topology embeds spatial heterogeneity; the adversarial alignment enforces domain invariance; and the partitioning strategy respects the epistemic hierarchy of labels. This integrated design provides both the methodological rigor and flexibility necessary to support large-scale, label-efficient urban noise mapping across diverse and data-scarce cities.

6.2.7 Inference Strategy and Prediction Workflow

6.2.7.1 Overview

After completing the multi-domain training of the generalizable urban noise prediction model, a systematic inference strategy was implemented to generate high-resolution, spatially continuous noise maps across all study cities. This strategy relies directly on the trained model weights and the standardized graph structures constructed during the training phase, ensuring methodological consistency and full reproducibility of results. Importantly, because the spatial graphs were designed from the outset with a regular 30-meter grid and extended boundaries, no architectural modifications were required for inference. Each city's graph, containing both labeled and unlabeled nodes, was applied seamlessly to the prediction stage, preserving the full spatial context critical for robust graph-based learning and avoiding any structural inconsistencies between the training and deployment phases.

6.2.7.2 Input Preparation

To prepare the inputs for inference, each study city—comprising the reference domain Southampton and the four target domains Cardiff, Portsmouth, Nottingham, and Liverpool—was processed through the same rigorous pipeline established during training. Specifically, two distinct feature tensors were generated: one capturing local-scale variables reflecting the 0–120 m neighborhood characteristics, and another representing the broader 120–1000 m background context. Alongside these features, the spatial coordinates of each grid node were recorded, ensuring positional integrity across the graphs.

The input preparation further included the explicit assignment of domain labels, with Southampton coded as domain 0 and the target cities as domains 1 through 4. These domain identifiers were critical for activating the domain-specific normalization parameters embedded in the trained model architecture, thus enabling adaptive behavior tailored to each urban context. Moreover, the ten edge-sensitive environmental variables selected during feature importance ranking—emphasizing land use and spectral characteristics strongly predictive of noise propagation—were integrated directly into the graph representation, ensuring that both node and edge attributes mirrored the training phase. All preprocessing steps, including variable standardization, one-hot encoding for categorical features, and imputation for any missing values, strictly followed the procedures applied during model development, ensuring consistency and reproducibility.

6.2.7.3 Graph Construction for Inference

The graph construction for inference was fully aligned with the spatial graph design used during training, ensuring that the topological relationships among nodes remained stable. Each city-specific graph preserved the 4-neighbor connectivity, extended through a 3-hop search radius to capture broader spatial dependencies. Importantly, edge weights were not limited to spatial proximity but incorporated dynamic modulation based on semantic similarity, as encoded by the

EdgeWeightLayer using the ten most noise-relevant environmental features. This approach ensured that the graphs maintained both geometric and functional fidelity, reflecting the intricate interplay between urban form, land cover, and noise dynamics.

By directly reusing the previously constructed graphs, the inference process avoided the risk of introducing sampling biases or topological mismatches between the training and prediction phases. The schematic workflow of this prediction process is visualized in Figure 6.4, which outlines the seamless integration of graph construction, model input preparation, and noise prediction generation.

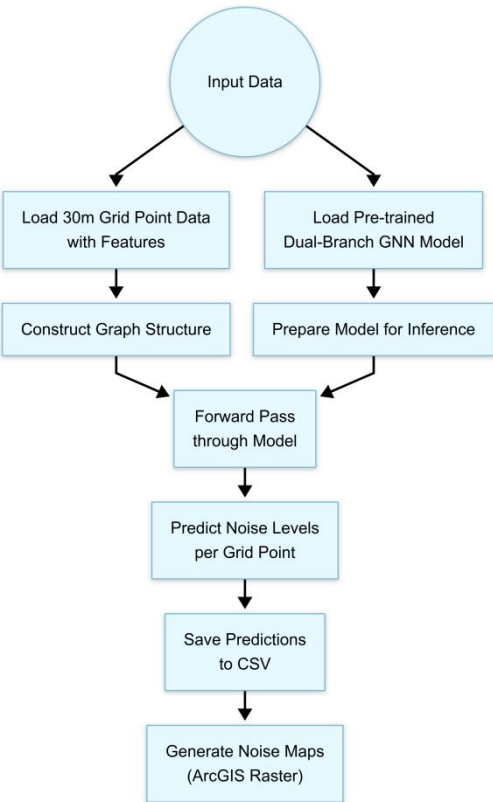


Figure 6.4 prediction workflow

6.2.7.4 Prediction Process

The prediction process began by loading the trained Dual-Branch Graph Neural Network (DualBranchGNN) model and restoring its parameters from the checkpoint that achieved the best performance during multi-domain adversarial fine-tuning. At this stage, maintaining strict architectural consistency was critical: the model’s dual-branch configuration—designed to process local and global features independently—together with its domain-adaptive components, had to be preserved exactly as defined during training.

Once initialized, the complete graph of each target city was input into the network. The local-scale features were propagated through the Graph Attention Network (GAT) branch, which adaptively weighted the relative importance of immediate neighbors and captured fine-grained spatial dependencies within a 120 m neighborhood. In

parallel, the global-scale features were processed by the Graph Convolutional Network (GCN) branch, which aggregated contextual information over extended neighborhoods up to 1,000 m using normalized mean aggregation. These two representations captured complementary spatial patterns—local acoustic variations and broader urban context—and were subsequently fused through an adaptive gating mechanism that dynamically balanced their contributions for each spatial unit. The resulting fused embeddings were passed through a regression layer to predict the mean daytime noise level (dBA) for each grid point.

After prediction, the node-level outputs were reattached to their corresponding spatial coordinates within the city graph. Each city’s complete set of enriched node attributes—comprising original features, intermediate embeddings, and predicted noise levels—was then exported as a georeferenced CSV file to ensure seamless compatibility with downstream GIS and visualization tools. This workflow guarantees a fully automated and reproducible pipeline for generating consistent noise prediction maps across all study areas.

To clarify, the term “grid point” refers to the centroid of a regular 30 m × 30 m spatial cell, forming a continuous grid that covers the entire urban extent of each city. During inference, each node in the graph corresponds to one of these grid centroids, with its feature vector derived from the multispectral and land-use attributes aggregated within that grid cell and its surrounding ring buffers. Consequently, the node-level predictions are spatially equivalent to grid-level noise estimates aligned to these centroids.

This design choice ensures that model outputs are spatially continuous, georeferenced, and directly interoperable with standard geospatial workflows. Each predicted value can be reprojected into the city’s coordinate reference system and visualized as a raster map, facilitating cross-city comparison and integration with auxiliary datasets such as transportation networks, vegetation cover, or population exposure layers. The use of standardized, evenly spaced grid points—rather than irregular observation sites—guarantees spatial consistency, reproducibility, and compatibility between the graph-based predictions and conventional environmental mapping approaches.

6.2.7.5 Evaluation Metrics and Map Validation

Because ground-truth noise measurements were only available for Southampton, quantitative evaluation of prediction accuracy was confined to this reference city. To ensure methodological continuity with prior analyses (Chapters 4 and 5), the evaluation employed two complementary metrics: the mean absolute error (MAE) and the coefficient of determination (R^2).

The MAE was calculated as the average absolute difference between the predicted noise levels ($\hat{y}_i$) and the true observed values (y_i) across all valid labeled nodes, formalized as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

where N represents the number of labeled grid points. This metric provides an intuitive and interpretable measure of the model's average prediction error magnitude, emphasizing precision across the full range of observed noise values.

The coefficient of determination (R^2) was calculated to capture the proportion of variance in the ground-truth noise levels explained by the model's predictions:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the true noise levels. This statistic evaluates the model's explanatory power, indicating how well the predictions capture the underlying variability in the observed data.

For both metrics, evaluation was restricted exclusively to nodes identified with $set_id \leq 3$ during data preparation, ensuring that only genuine labeled samples contributed to performance assessments. This careful delineation maintained the integrity of the evaluation process and ensured that results were not biased by pseudo-labeled or inferred data.

6.2.7.6 Rasterization and Visualization

After obtaining node-level noise predictions at the 30-meter resolution grid points, a detailed rasterization process was applied to transform these discrete predictions into continuous noise maps for each study city. The predicted mean noise levels were first directly assigned to their corresponding grid points, strictly preserving the model's original output values. At this stage, no additional spatial interpolation or smoothing techniques were applied, ensuring that the raw prediction granularity and the inherent spatial fidelity of the graph-based model outputs were fully retained.

To convert the point-based predictions into raster surfaces, the results were imported into ArcGIS Pro 2.7.0, a leading geospatial analysis platform. Using the precise spatial coordinates accompanying each grid point, a point-to-raster transformation was conducted, producing gridded raster layers with a native resolution of 30 meters. This rasterization step ensured that the outputs were perfectly aligned with the original spatial grid structure, preserving both the spatial resolution and the neighborhood relationships essential for accurate urban noise visualization.

For visual comparability across cities, a standardized color ramp was systematically applied to all generated maps. The color scheme was symmetrically centered around each city's mean predicted noise level, with divergent hues assigned to represent lower and higher noise intensities. This approach ensured that cross-city map comparisons were visually meaningful and not confounded by scale distortions or inconsistent color mappings. Additionally, all raster surfaces were spatially clipped to

match the delineated study area boundaries, avoiding edge artifacts or visual inconsistencies outside valid data coverage.

The final rasterized outputs were saved in the GeoTIFF format, an open and widely used geospatial raster data standard. This ensured not only the reproducibility of the results but also their compatibility with a broad range of downstream spatial analysis and urban planning tools. Through this rigorous and standardized visualization workflow, the predicted noise maps provided robust, high-resolution insights into urban acoustic environments, supporting comparative evaluations and informing evidence-based environmental management strategies.

6.2.8 Methodology Summary

The methodological framework developed in this study provides a comprehensive, scalable, and transferable solution for predicting urban noise distributions across multiple cities, even in the absence of direct acoustic measurements in target locations. Anchored in the central hypothesis that urban soundscapes are systematically shaped by shared structural and functional characteristics—particularly those embedded in land-use patterns, morphological configurations, and broader environmental contexts—the research operationalizes a cross-domain transfer paradigm that leverages these latent inter-city similarities.

At the core of this framework is the integration of standardized multispectral remote-sensing features, rigorously engineered land-use statistical indicators, and high-confidence pseudo-labeled samples derived from robust variable correlations established in the reference domain. Together, these components enable the construction of a latent feature space in which heterogeneous cities' noise distributions become directly comparable, effectively bridging the data gap between measured and unmeasured domains.

The methodology introduces several interlocking innovations. First, the pseudo-label generation strategy employs variable-importance rankings from the reference city, combined with inter-city feature-similarity assessments, to infer plausible supervisory signals in cities lacking ground-truth noise labels. This approach enables supervised learning under data sparsity—a fundamental challenge in urban environmental modelling. Second, the spectral standardization of multispectral imagery is enhanced through the use of land-use-specific pseudo-invariant feature (PIF) regions and histogram-matching techniques. This ensures that spectral properties are harmonized across diverse urban contexts, reducing cross-domain discrepancies caused by sensor or acquisition differences. Third, the dual-branch graph neural network (GNN) architecture—explicitly designed to separate local (0–120 m) and contextual (120–1000 m) environmental influences—captures multi-scale determinants of urban noise patterns. This structural decoupling allows the model to learn both fine-grained, detail-sensitive variations and broader contextual trends, enhancing explanatory capacity.

In addition, the inclusion of dynamic edge-weighting mechanisms is not intended to replicate physical sound-propagation pathways, but rather to *statistically represent* how local neighborhood relationships vary with environmental heterogeneity. The model thus learns to adjust the relative influence of adjacent nodes based on the presence of variables such as vegetation density, built-up ratio, and land-use composition—factors empirically correlated with acoustic attenuation and reflection. The weighting parameters summarized in Table 6.12 were selected and tuned through cross-validation and feature-importance analysis rather than manual physical calibration. Their plausibility is evaluated indirectly through the stability and cross-domain consistency of results presented in Chapter 6, which collectively serve as a sanity-checking mechanism for the robustness and interpretability of these modulating factors.

Finally, the application of multi-domain adversarial alignment strategies bridges latent feature-distribution gaps among cities, improving the model’s ability to generalize across urban environments with distinct morphological, infrastructural, and socio-economic characteristics. The framework also addresses sample-density consistency by generating uniform grid-based spatial representations across all study areas. This design ensures that graph structures remain comparable across domains, facilitating robust neighbor-based feature aggregation while preserving the integrity of spatial relationships during learning.

Through this integrated methodology, the research demonstrates that it is feasible to achieve high-quality, spatially explicit noise predictions in unseen cities using only publicly accessible remote-sensing and land-use data. The approach therefore offers a practical and label-efficient pathway toward scalable urban environmental assessment, particularly valuable for regions where systematic noise-monitoring infrastructure is limited or entirely absent.

6.3 Result

6.3.1 Performance Evaluation on the Reference City

6.3.1.1 Quantitative Performance Metrics

To establish a robust performance baseline before extending predictions to unseen cities, the proposed dual-branch graph neural network (GNN) model was evaluated on the held-out test set of the reference city, Southampton. Two widely accepted metrics were employed to assess prediction accuracy: the Mean Absolute Error (MAE), which quantifies average absolute deviations between predicted and observed noise levels, and the coefficient of determination (R^2), which measures the proportion of variance explained by the model.

Table 6.14 summarizes the evaluation results alongside the performance of two previously developed models: the convolutional neural network (CNN) introduced in Chapter 4 and the GraphSAGE-based GNN described in Chapter 5. The dual-branch

GNN achieved a MAE of 4.48 dBA and an R^2 of 0.576 on the Southampton test set. While its MAE was marginally higher than the GraphSAGE model (4.40 dBA), it significantly outperformed the CNN (4.79 dBA) and offered competitive explanatory power, situating itself between the two benchmarks.

Table 6.14 Comparison of MAE and R^2 between this study and former chapters

Model	MAE (dBA)	R^2
CNN (Chapter 4)	4.79	0.491
GraphSAGE (Chapter 5)	4.40	0.596
Dual-branch GNN (This study)	4.48	0.576

Notably, the present model achieved this performance while dramatically reducing feature dimensionality. The GraphSAGE model relied on approximately 8,250 input variables, whereas the dual-branch GNN used only 169 selected features after rigorous correlation filtering and random forest-based importance ranking. This substantial dimensionality reduction not only improved computational efficiency but also reduced the risk of overfitting—a key advantage when designing models intended for cross-city generalization.

These results highlight that the selected features in this study, particularly the land use and remote sensing-derived predictors, capture highly informative and transferable patterns related to urban noise propagation. As such, the dual-branch GNN balances architectural robustness, feature compactness, and predictive effectiveness, establishing a solid foundation for exploring cross-domain generalization.

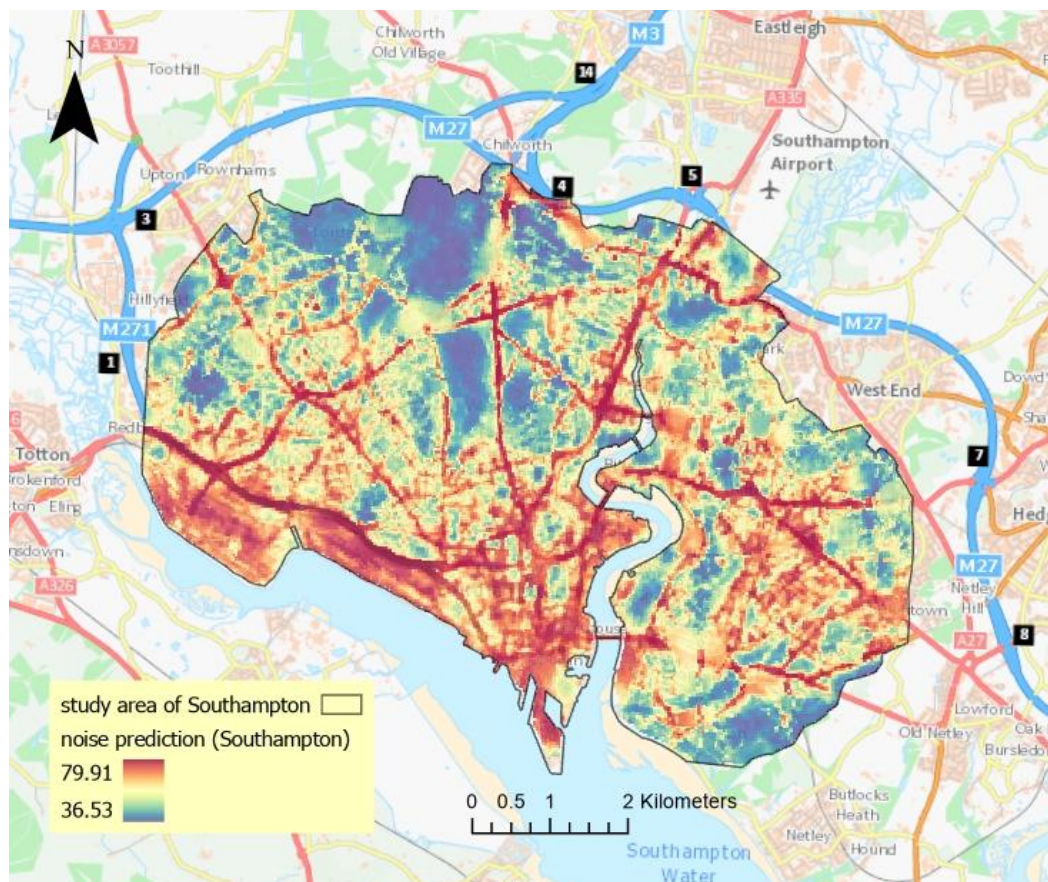
6.3.1.2 Spatial Noise Mapping of the Reference City

Beyond quantitative metrics, the spatial distribution of predicted noise levels offers critical insights into the model’s capacity to reproduce realistic urban acoustic landscapes. To assess this, a full-resolution noise prediction map for Southampton was generated using the trained dual-branch GNN and the complete 30-meter resolution spatial grid. This map, presented in Figure 6.5a, visually illustrates how the model reconstructs continuous, coherent, and environmentally plausible noise patterns across the city.

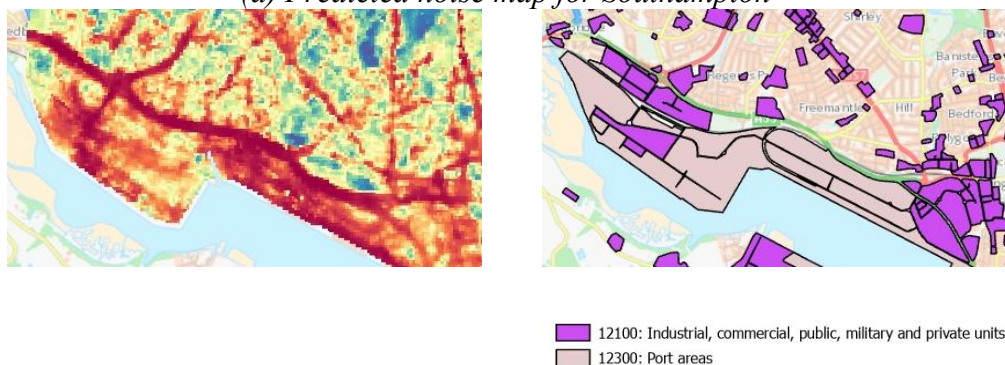
A central innovation distinguishing this study from prior work in Chapters 4 and 5 lies in the explicit integration of land use and land cover (LULC) features derived from the Urban Atlas 2012 dataset. By extracting and encoding LULC attributes at both local (0–120 m) and contextual (120–1000 m) scales, the model incorporates semantically rich information about urban function and environmental context—information that goes beyond purely image-based or traffic-focused approaches. As a result, the predicted map successfully captures classic urban noise gradients: elevated noise levels along major transport corridors, moderate levels in commercial and industrial zones, and lower exposures in residential neighborhoods and green spaces.

To further evaluate the model’s semantic sensitivity and environmental realism, three representative sub-regions were selected for detailed visual analysis, comparing the

predicted noise surfaces with dominant LULC patterns. In the port and industrial area (Figure 6.5b), characterized by a concentration of commercial and port infrastructure, the model effectively predicted high noise levels, despite the relative absence of major roadways, suggesting that it successfully captured latent noise sources such as logistics or maritime activities. In the mixed-use city center (Figure 6.5c), where continuous urban fabric, commercial zones, port proximity, and scattered green areas coexist, the model revealed a complex, heterogeneous acoustic landscape, demonstrating its ability to parse fine-grained LULC interactions and urban functions. Finally, in the southwest suburban area (Figure 6.5d), dominated by green urban spaces, recreational zones, and pasture lands, the model consistently predicted low noise exposures, reflecting an accurate recognition of non-traffic-related acoustic attenuation effects provided by these land uses.



(a) Predicted noise map for Southampton



(b) port and industrial area

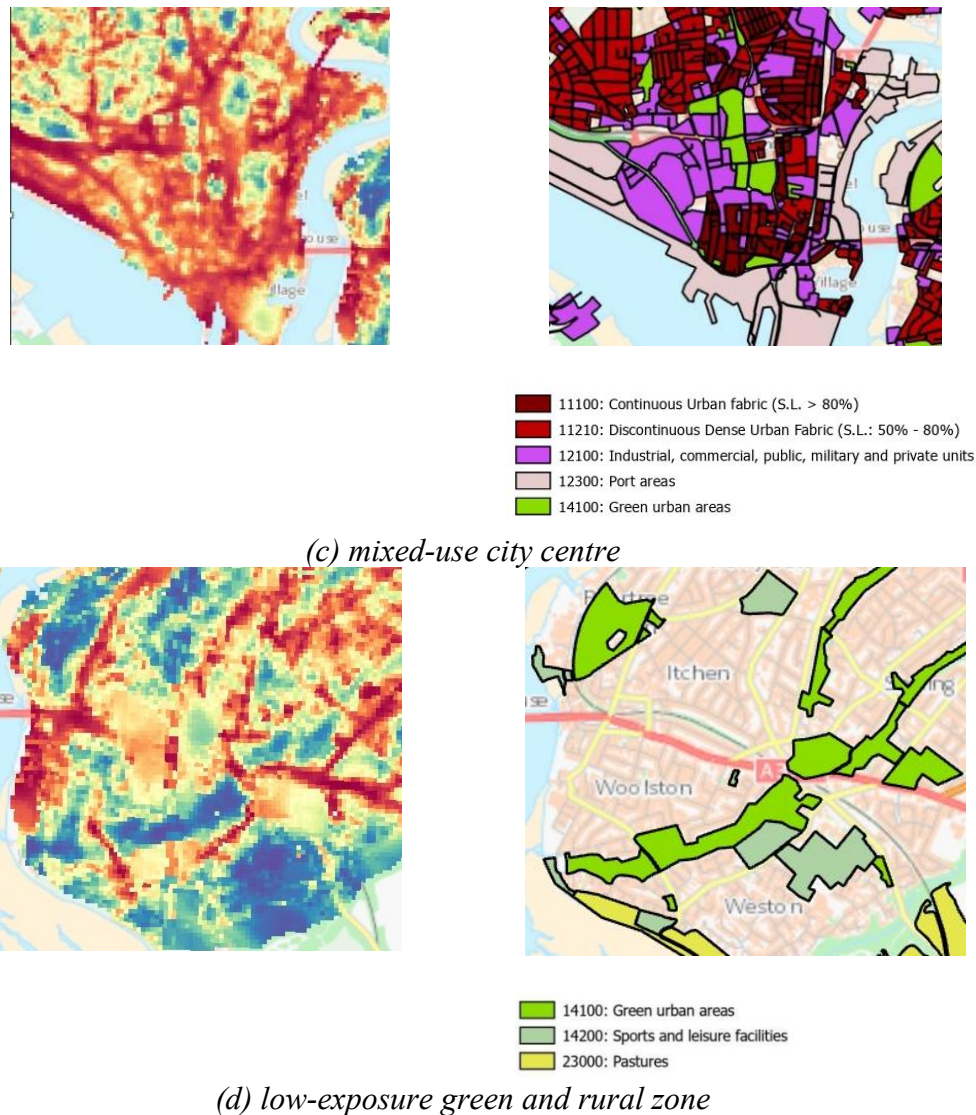
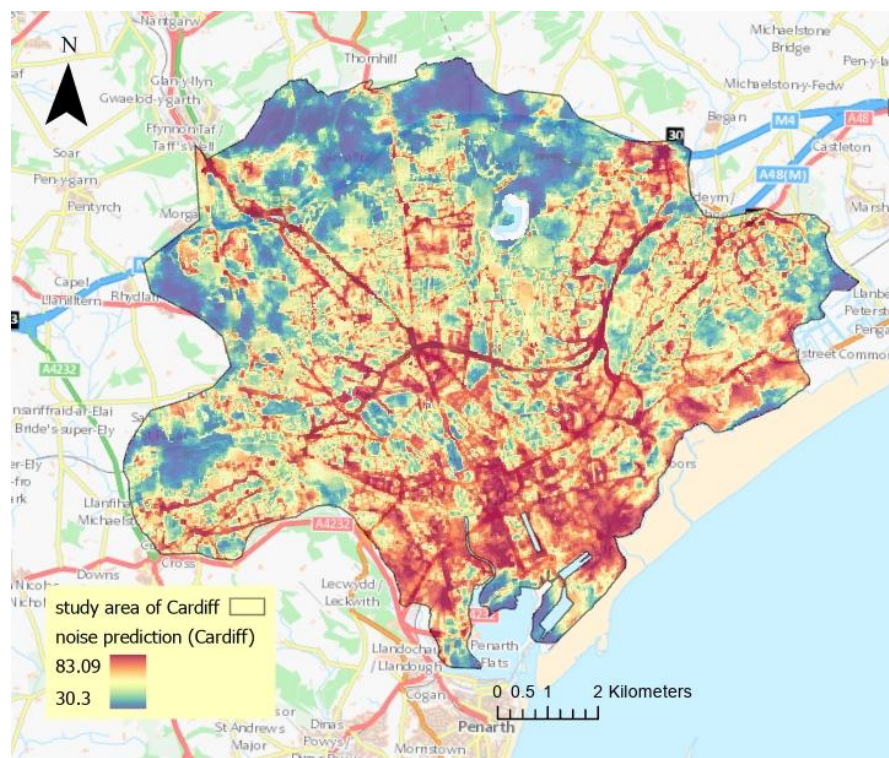


Figure 6.5 Predicted noise map for Southampton and selected sub-regions: (a) Predicted noise map for Southampton (b) port and industrial area; (c) mixed-use city centre; (d) low-exposure green and rural zone. Each submap (right side) overlays Urban Atlas 2012 dominant land use classes

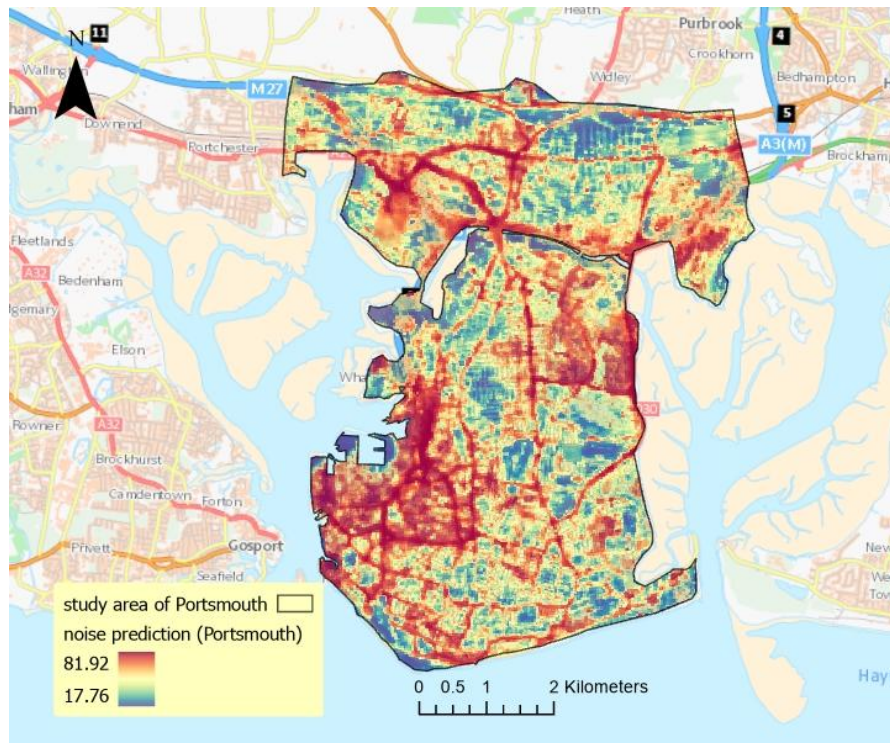
6.3.2 Generalization to Target Cities

A central objective of this study was to develop a robust, generalizable noise prediction framework capable of accurately estimating urban noise patterns in target cities without relying on any local acoustic measurements. To rigorously evaluate this capability, the dual-branch graph neural network (GNN) was trained exclusively on the reference city, Southampton, and then directly applied to four distinct target cities—Cardiff, Portsmouth, Nottingham, and Liverpool. Importantly, no fine-tuning or local calibration was performed; predictions were generated solely based on multispectral imagery and Urban Atlas 2012 land use data, using the same pre-trained model weights and spatial graph constructions detailed in Section 6.2.6. This strict experimental design ensured methodological consistency and provided a fair test of the model’s cross-city generalization performance.

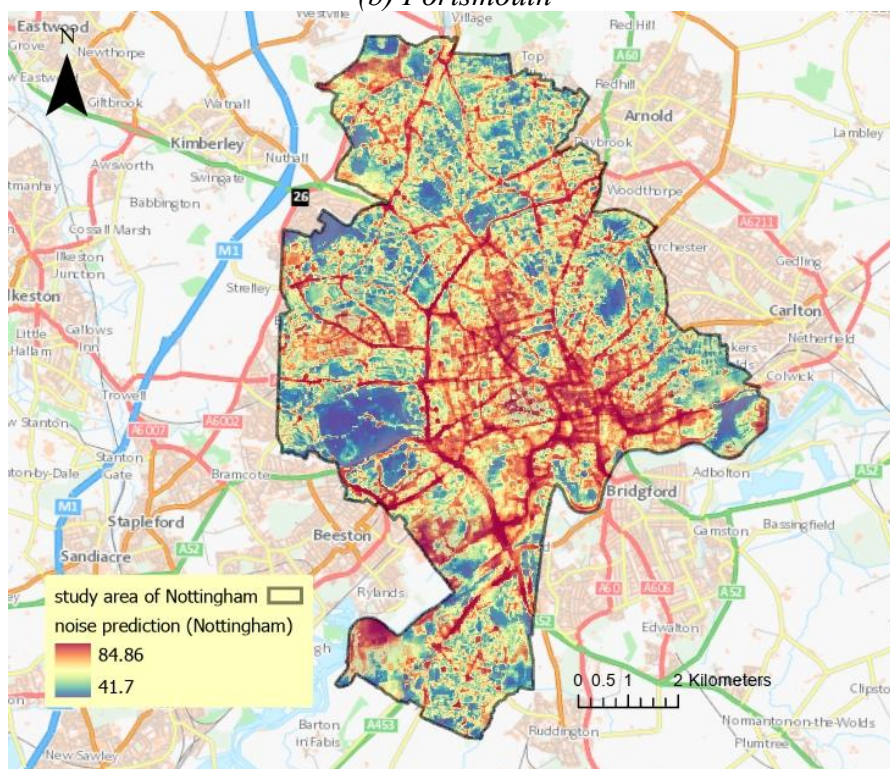
The resulting full-resolution noise prediction maps for the four target cities are presented in Figure 6.6. Despite the complete absence of ground-truth noise labels during both training and inference, the predicted maps exhibit spatially coherent noise gradients and environmentally plausible acoustic distributions. Particularly noteworthy is the consistent emergence of elevated noise levels along major transportation corridors, within industrial belts, and across dense commercial zones—patterns that align well with known functional-acoustic relationships documented in urban noise research. Conversely, residential neighborhoods, green urban spaces, and peripheral low-density zones are consistently predicted as low-exposure regions, demonstrating the model’s sensitivity to the acoustic attenuation effects provided by these land-use types.



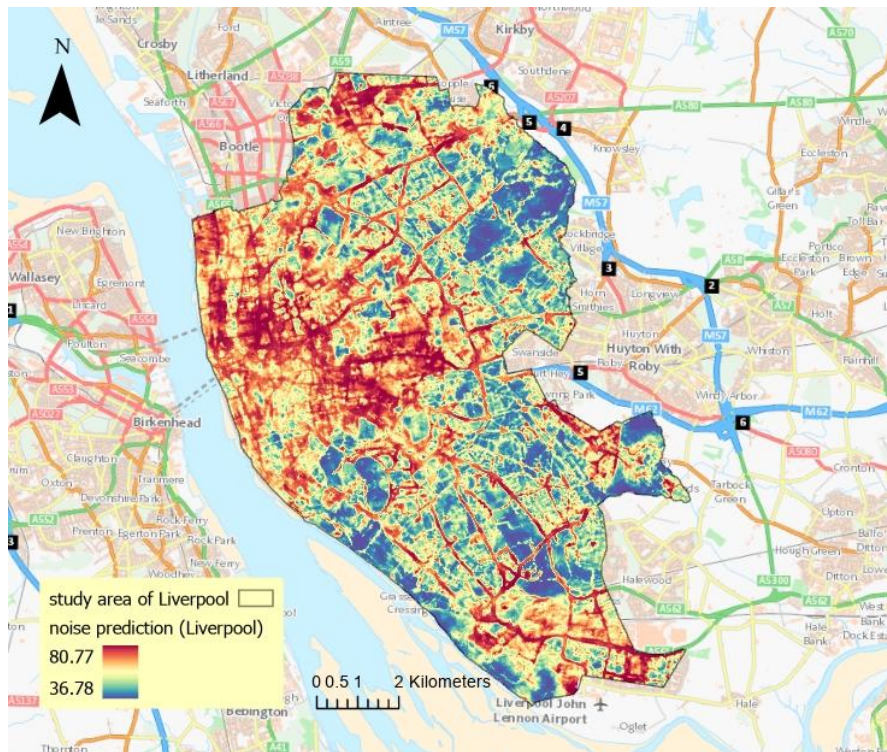
(a) Cardiff



(b) Portsmouth



(c) Nottingham



(d) Liverpool

Figure 6.6 Full-resolution noise prediction maps for the four target cities: (a) Cardiff; (b) Portsmouth; (c) Nottingham; (d) Liverpool. All maps use the same 30 m resolution grid and color ramp for visual comparability

Several key methodological components contributed to the quality and plausibility of these predictions. First, the robust image standardization workflow, including pseudo-invariant feature (PIF)-based histogram matching, effectively harmonized inter-city spectral discrepancies, ensuring that remote sensing features remained comparable across diverse urban domains. Second, the incorporation of land-use-derived variables provided semantically rich and transferable spatial context, enabling the model to ground its predictions in meaningful urban structures rather than superficial spectral similarities. Finally, the graph-based architecture, with its explicit modeling of both fine-scale and contextual spatial dependencies, allowed the network to capture complex neighborhood interactions that govern urban noise propagation.

A qualitative inspection of the predicted maps across the four target cities reveals consistent and realistic noise patterns. In Cardiff and Portsmouth, for example, major port areas and adjacent industrial zones are accurately identified as persistent high-noise regions, even without direct access to local traffic or activity data. In Nottingham and Liverpool, inner-city corridors and arterial roads exhibit strong predicted noise gradients, while suburban green belts and recreational zones show marked acoustic attenuation. These observations suggest that the model effectively leverages inter-city functional-acoustic similarities, supporting the core assumption underpinning the domain adaptation framework.

To further validate the technical soundness and adaptability of the proposed approach, the following sections (6.3.3 – 6.3.5) provide a comprehensive set of evaluation experiments. These include quantitative comparisons with official European Noise Directive (END) noise maps, stratified assessments of prediction performance across

different land-use categories, and ablation analyses examining the contributions of individual feature groups and architectural components. Together, these validation efforts aim to rigorously assess whether the generalization capacity demonstrated here represents a substantive technical advancement over prior work and contributes meaningfully to the development of scalable, label-efficient urban noise prediction systems.

6.3.3 Local Comparison with END Traffic Noise Maps

To rigorously assess the spatial validity of the predicted noise maps, this study conducted both qualitative and quantitative comparisons against official traffic noise datasets produced under the Fourth Round of the European Environmental Noise Directive (END). Specifically, the $L_{Aeq,16h}$ indicator was selected as the primary benchmark because it represents the 16-hour equivalent continuous level—defined as the annual average steady noise level over the daytime period (07:00 to 23:00)—and thus aligns directly with the L_{Aeq} noise prediction produced by the model in this study. This temporal and semantic correspondence makes $L_{Aeq,16h}$ the most appropriate available reference for evaluating the model’s outputs.

The END dataset is widely recognized as the most authoritative and standardized source of transportation noise information across European cities, compiled by national authorities under harmonized European Commission guidelines (European Commission, 2021). These maps are generated using detailed models of major road, rail, and airport sources, applying rigorous assumptions about emissions, propagation, and environmental attenuation. In the UK context, DEFRA oversees the production and publication of these datasets, ensuring a high degree of consistency and comparability across time and geography (DEFRA, 2023).

Although the predictions from the dual-branch graph neural network represent a broader conceptualization of environmental noise—incorporating both traffic-related and non-traffic acoustic contributors such as industrial zones, commercial activity, and dense urban morphology—the use of END $L_{Aeq,16h}$ maps remains a meaningful and informative comparison point. This is justified on several grounds. First, END maps are the only comprehensive, government-endorsed datasets currently available for large-scale environmental noise evaluation, offering unparalleled geographic and regulatory consistency. Second, traffic noise typically dominates the overall urban acoustic environment, especially in European cities like Nottingham, Portsmouth, and Liverpool, meaning that traffic-centered benchmarks can effectively capture a substantial fraction of the total noise variation. Third, while structural differences exist between modeled END outputs and image-derived predictions, evaluating the degree of spatial agreement provides valuable insights into the predictive framework’s capacity to generalize meaningful urban noise patterns.

However, it is important to acknowledge certain limitations in the direct comparison. END noise maps, being strictly transport-focused, often include sharp spatial transitions between road-adjacent and off-road zones, sometimes assigning artificially low or even zero values to non-modeled regions such as green belts or peripheral industrial zones. In contrast, the predictions produced here are spatially continuous

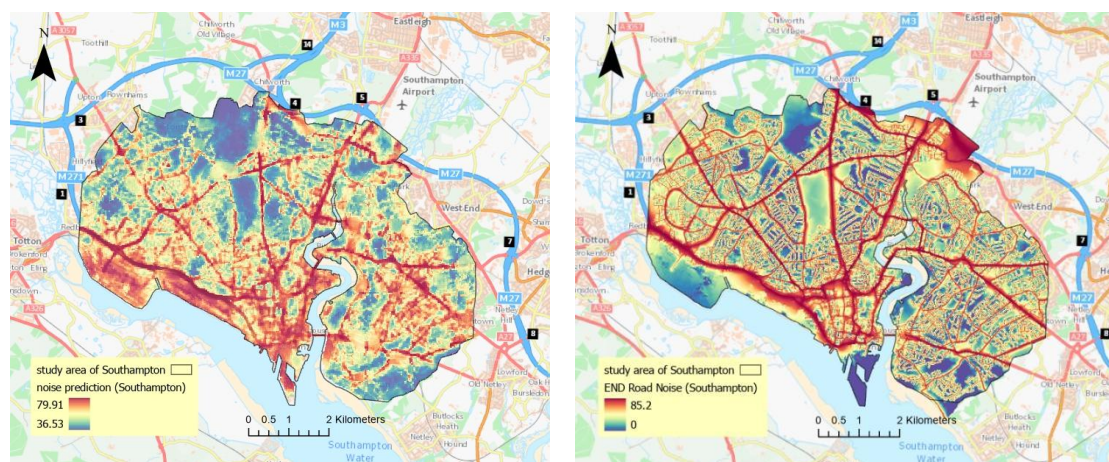
and semantically enriched, reflecting not only transport networks but also the cumulative influence of surrounding land cover, built morphology, and local environmental context. This distinction must be carefully considered when interpreting areas of convergence or divergence between the two datasets.

6.3.3.1 Global Spatial Comparison with END

Figure 6.7 provides a side-by-side comparison between the predicted noise maps and official END LAeq,16h maps for Southampton (reference city) and three target cities: Nottingham, Portsmouth, and Liverpool. All maps are rendered using a standardized color ramp spanning 35 to 85 dBA, which covers the plausible range of urban daytime noise exposure, allowing for direct visual comparability.

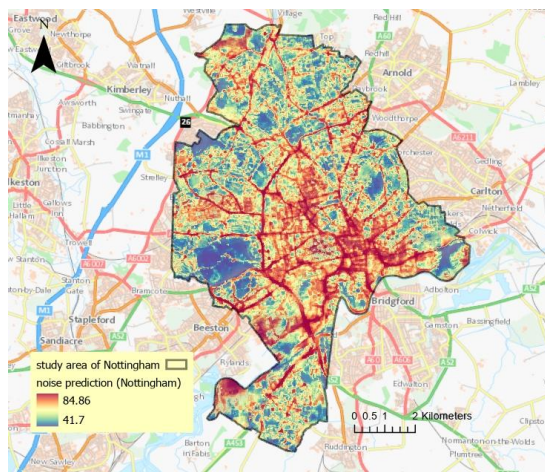
Across all four cities, several consistent spatial patterns emerge. Both the predicted and official maps successfully identify major transportation corridors, central commercial districts, and industrial zones as prominent noise hotspots, indicating that the model accurately captures core traffic-related exposure patterns. Importantly, the predicted maps display smoother spatial gradients, reflecting their integration of cumulative environmental influences beyond discrete road networks. In contrast, the END maps frequently exhibit abrupt zone transitions, including regions marked by unrealistically low or null values in areas beyond the immediate vicinity of major transportation sources—a limitation noted in previous urban noise modeling research (Murphy & King, 2011).

A quantitative inspection reveals that the maximum predicted noise levels, typically ranging from 79.9 to 84.9 dBA, align closely with the upper bounds of the END datasets, which often peak between 80 and 85 dBA. Moreover, the lower bounds of the predicted maps, falling between 35 and 42 dBA, provide more realistic estimates of baseline background noise, whereas the END maps frequently default to near-zero values in off-network areas. This suggests that the predictive framework offers a more continuous and environmentally representative characterization of the urban acoustic landscape, capturing the complex interplay of multiple environmental determinants.

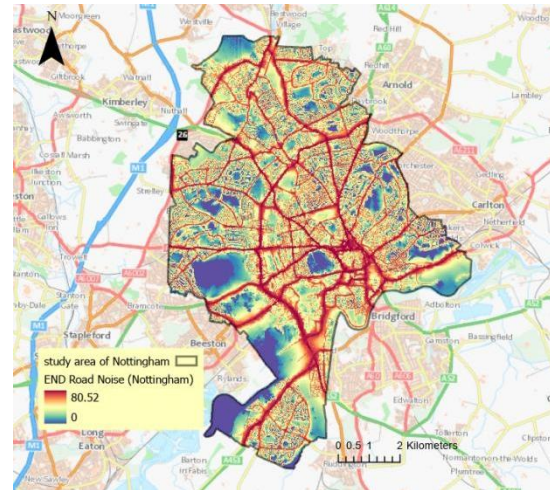


(a) Southampton (Prediction)

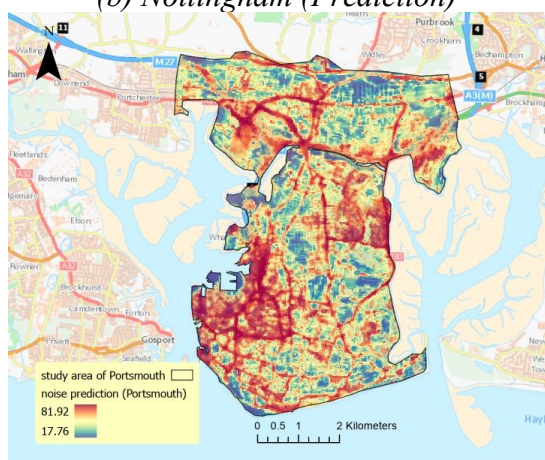
Southampton (END LAeq)



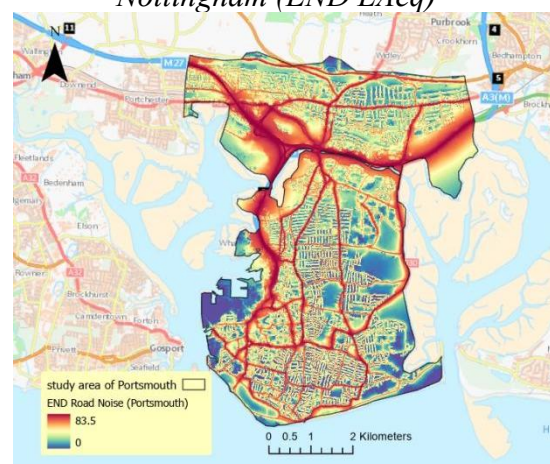
(b) Nottingham (Prediction)



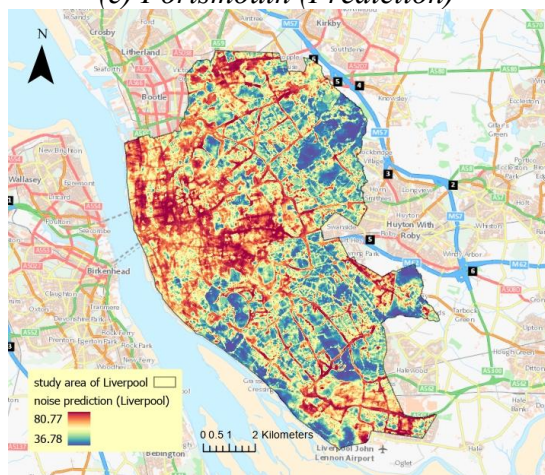
Nottingham (END LAeq)



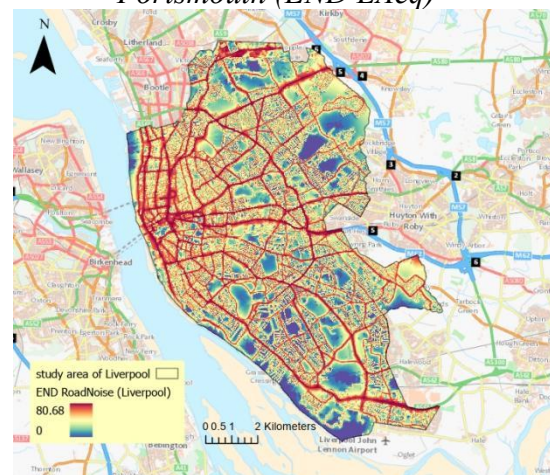
(c) Portsmouth (Prediction)



Portsmouth (END LAeq)



(d) Liverpool (Prediction)



Liverpool (END LAeq)

Figure 6.7 Predicted noise maps and END Lden maps for (a) Southampton, (b) Nottingham, (c) Portsmouth, and (d) Liverpool

It is important to emphasize that the END (Environmental Noise Directive) LAeq maps used here (from the Fourth Round) are constructed using deterministic propagation models focused solely on road-traffic sources. In their modeling framework, acoustic contribution is only assigned to grid cells within a calculated influence distance from mapped roads; beyond that buffer, the model effectively treats

the area as having no transport noise contribution, resulting in zero-value cells. This does not imply actual silence in the real world, but rather the absence of modeled traffic influence.

In contrast, our predicted maps are driven by actual noise measurements that inherently include multiple contributing sources beyond just roads — for example, industrial zones, port facilities, commercial operations, building HVAC systems, and human activity. Therefore, even in areas distant from roads, the model can assign non-zero noise levels consistent with realistic background exposure. Furthermore, our model incorporates land-use and form-based attenuation effects (buildings, vegetation, buffer shielding), allowing spatially smooth transitions rather than abrupt drop-offs.

To underscore this contrast, one may interpret that END performs well in road-dominated zones, where traffic is indeed the dominant source, but loses representational validity in off-network or non-road regions, where its assumption of zero exposure fails. By comparison, our model's predictions represent a more holistic soundscape-level exposure field, filling in what END leaves blank. We have strengthened the manuscript by adding this discussion immediately after Figure 6.7, to help readers understand that observed low-noise zones in END data are methodological artifacts, not necessarily reflecting true environmental quietness.

6.3.3.2 Quantitative Comparison in reference city

Following the qualitative visual assessment presented in the preceding section, this part provides a quantitative validation aimed at elucidating the systematic differences between the official END traffic-noise maps and the measured acoustic environment in Southampton. The analysis focuses on how the deterministic propagation-based END estimates deviate from in-situ measurements across varying spatial contexts and functional land-use types. By quantifying these discrepancies, this section establishes a baseline understanding of the strengths and limitations of the END dataset before comparing it with the modelled noise predictions introduced later in this chapter.

To achieve this, the measured daytime LAeq values were paired with the corresponding LAeq,16h values from the Fourth Round END dataset at identical spatial coordinates. The comparison employs four complementary indicators:

- (1) the Mean Absolute Error (MAE) to quantify deviation magnitude;
- (2) the bias (END – observed) to identify systematic under- or overestimation;
- (3) the silent rate, i.e. the proportion of cells assigned a zero or null END value; and
- (4) the zero-gap, which represents the mean measured noise level at those END-silent locations, thereby indicating the degree to which genuine exposure is omitted in the official maps.

Overall and Distance-Dependent Patterns

The aggregated statistics (Table 6.15) reveal a consistent underestimation pattern in the END data relative to the measured sound levels. Across all 13 139 measurement points, the overall MAE is 7.4 dBA and the mean bias –4.8 dBA, implying that END maps typically predict lower values than those recorded on-site. Within the immediate

road corridor (≤ 90 m from major roads), accuracy is somewhat higher, with MAE ≈ 7.0 dBA and bias ≈ -4.3 dBA, while the silent-rate remains minimal (1.4%), confirming that END performs reasonably within its intended transport-network domain.

However, model performance deteriorates rapidly with increasing distance from roads. Between 90 and 150 m, the MAE almost doubles (15.8 dBA), and beyond 150 m, errors exceed 17 dBA with a silent-rate approaching 30 %. In these peripheral zones, END values often default to zero due to the absence of defined traffic-source contributions. Yet, the corresponding measured noise levels at those “silent” locations remain between 47 and 52 dBA—demonstrating that these areas are not acoustically quiet but simply fall outside the END model’s propagation envelope.

These results quantitatively confirm the visual impression that the END dataset captures noise patterns reliably only along major transportation corridors, while rapidly losing fidelity in residential interiors and non-road environments. The distance-binned relationship between error magnitude and proximity to roads exhibits a clear monotonic increase in both MAE and silent-rate, highlighting the deterministic cutoff behavior intrinsic to the END modelling procedure.

Table 6.15 Comparison between END and Measured Noise Levels by Road-Distance Zone

Distance band	N	MAE (dBA)	Bias (END – Obs) (dBA)	Silent Rate (%)	Zero-Gap (Obs @ END = 0) (dBA)
All points	13 139	7.44	–4.81	2.4	52.71
≤ 90 m (from roads)	12 587	7.05	–4.34	1.4	55.55
90–150 m	272	15.78	–14.64	20.6	51.74
> 150 m	280	17.19	–16.63	29.3	47.24

Functional and Land-Use Differences

When the comparison is stratified by the Urban Atlas 2012 land-use classification, the underestimation pattern displays distinct functional variation (Table 6.16). END values align most closely with measurements in residential fabrics, where MAE averages 7.2 dBA and bias ≈ -4.4 dBA. Errors are slightly larger in industrial, commercial, and public zones (MAE ≈ 8.0 dBA; bias ≈ -6.5 dBA), suggesting that END partially captures but still underestimates acoustic emissions associated with mixed-use and service areas.

By contrast, discrepancies become extreme in port areas, where MAE exceeds 16 dBA and bias ≈ -14.8 dBA, with silent-rate > 13 %. Measured levels at END-silent port cells reach ≈ 63 dBA—an exposure intensity typical of maritime operations and heavy-vehicle activity that lies entirely outside the END model’s road-traffic scope. Green urban areas also exhibit moderate but systematic underestimation (MAE ≈ 7.3 dBA; bias ≈ -5.5 dBA), reflecting the exclusion of diffuse background sources such as human recreation and adjacent road spillover.

These results collectively indicate that while END maps reproduce transport-dominated noise patterns with acceptable accuracy, they systematically truncate environmental exposure in functionally diverse contexts. The omission is especially pronounced in industrial, port, and mixed-use zones, where non-road sources are substantial contributors to the urban soundscape.

Table 6.16 END vs Measured Noise by Land-Use Category (Urban Atlas 2012)

LULC Category	Num	MAE	Bias	Silent Rate (%)	Zero-Gap (dBA)
Residential fabric	6 794	7.18	−4.38	1.1	55.93
Industrial / Commercial / Public	1 536	8.04	−6.48	2.5	55.09
Green urban areas	1 330	7.26	−5.51	3.7	47.24
Port areas	119	16.15	−14.80	13.5	62.64
Other / Unknown	280	7.54	−6.00	3.2	60.74

Taken together, these quantitative results demonstrate that the END framework remains effective for its original policy purpose—mapping traffic-noise exposure along major roads—but provides an incomplete description of the broader urban acoustic environment. The deterministic modelling process, which ceases energy propagation beyond defined transport networks, inherently yields artificially low or null values in off-road areas. The measured data reveal that these supposedly “quiet” locations typically sustain moderate sound levels between 45 and 60 dBA, driven by diverse non-traffic sources such as industrial machinery, port logistics, air-handling units, or continuous human presence.

This analysis underscores a crucial methodological limitation: END captures the geometry of transport emissions but not the complexity of real soundscapes. By contrast, the data-driven predictive framework developed in this study is designed to integrate multisource and morphological determinants of noise propagation, producing a spatially continuous and environmentally realistic characterization of exposure. The next subsection therefore evaluates whether these model-based predictions achieve a more balanced agreement with both END and measured noise levels across varying urban contexts.

END vs Predicted Maps

Building on the baseline validation against in-situ measurements, the second stage of analysis compares the predicted noise maps generated by the proposed dual-branch GNN framework with the official END LAeq,16h maps. This comparison aims to determine whether the data-driven approach successfully preserves the realistic magnitude of traffic-related exposure within the transport corridors, while mitigating the abrupt underrepresentation of noise levels in non-road and mixed-use areas.

The results presented in Table 6.17 indicate a clear improvement in the spatial coherence and realism of the predicted maps relative to END. Across all 13 139 spatial units, the average absolute difference between the two datasets (MAE) is 6.2 dBA, with a mean $\Delta(\text{Pred} - \text{END})$ of +4.1 dBA, confirming that the model generally predicts higher—yet environmentally plausible—values. Agreement within ± 5 dBA

reaches 55.8 %, which is consistent with expected deviations between datasets based on fundamentally different modelling philosophies.

Within the road-adjacent corridor (≤ 90 m), differences are modest ($\text{MAE} \approx 5.8$ dBA; $\Delta = +3.7$ dBA), and the agreement rate (56 %) remains comparable to the global mean. This outcome suggests that the predictive framework maintains fidelity to the traffic-dominated zones already well captured by END. However, discrepancies increase markedly beyond the main transport bands: between 90 and 150 m, the mean difference rises to 14 dBA, and beyond 150 m, it exceeds 16 dBA. These patterns mirror the earlier END-vs-measurement results, indicating that the predictive model effectively restores noise continuity in zones where END systematically defaults to silence. In practical terms, the GNN-based maps predict moderate exposure values in these transitional and off-road areas, aligning with realistic background levels rather than the zero values found in END.

Table 6.17 Quantitative Comparison between Predicted and END Noise Levels by Road Distance

Num	Mean Δ (Pred–END)	MAE	Agreement($ \Delta \leq 5\text{dB}$)	Group
13139	4.132	6.204	0.558	All points
12587	3.652	5.792	0.562	On-road $\leq 90\text{m}$
272	14.054	14.958	0.445	90–150m ring
280	16.075	16.258	0.511	Off-road $>150\text{m}$

When disaggregated by Urban Atlas 2012 functional categories (Table 6.18), the differences exhibit a clear dependence on land-use context. The most consistent correspondence with END occurs in residential fabrics, where the mean Δ is +3.8 dBA and $\text{MAE} \approx 5.7$ dBA. In industrial, commercial, and public areas, discrepancies rise to +6.0 dBA ($\text{MAE} \approx 7.3$ dBA), while in green urban zones, they remain moderate (+4.9 dBA). By contrast, port areas again stand out as the most divergent class, with a mean Δ of +13.8 dBA and $\text{MAE} \approx 14.8$ dBA, reflecting the END model’s omission of maritime, mechanical, and loading-dock activities that substantially elevate ambient sound. The relatively low agreement rate in these high-intensity functional zones (≈ 30 %) highlights the insufficiency of traffic-only simulations for industrial and mixed urban morphologies.

Table 6.18 END–Prediction Differences by Land-Use Category (Urban Atlas 2012)

Num	Mean Δ (Pred–END)	MAE	Agreement($ \Delta \leq 5\text{dB}$)	Group
1330	4.935	6.262	0.586	Green urban areas
1536	6.022	7.335	0.443	Industrial & Commercial & Public
280	3.811	6.209	0.55	Other
119	13.765	14.783	0.303	Port areas
6794	3.782	5.682	0.566	Residential fabric
521	14.738	15.256	0.284	Sports & leisure facilities
2542	0.724	4.453	0.664	Transport corridors
17	32.728	32.728	0.059	Wetlands & Water

Across all distance and functional categories, the direction and magnitude of deviations between predicted and END noise levels are consistent with the empirical measurement evidence presented earlier. In regions where END underestimates exposure—particularly beyond 90 m from roads and in non-residential functional areas—the predicted maps yield higher, more realistic values that align with actual field observations. Conversely, in core transport corridors where END performs reliably, the two datasets remain largely congruent. This indicates that the proposed model does not over-amplify traffic noise but rather complements END by reconstructing the missing off-network exposure field.

From an applied standpoint, this outcome suggests that combining END's deterministic modelling with data-driven spatial prediction could enhance urban noise assessments. For instance, policymakers could continue to rely on END for standardized regulatory mapping of traffic emissions, while using predictive frameworks like the present one to estimate total environmental noise exposure—including industrial and socio-urban contributions—in areas where measurement networks are absent.

Concluding Remarks

In summary, the quantitative evaluation demonstrates that the predicted maps produced by the dual-branch GNN framework maintain strong coherence with the official END outputs in transport-dominated zones while substantially improving the representation of exposure in non-road and mixed-use environments. The systematic positive bias of 4–6 dBA relative to END does not reflect overprediction but rather the recovery of real-world background components that deterministic traffic models neglect. These components—stemming from industrial activities, port logistics, building operations, and social dynamics—constitute an integral part of the actual urban acoustic field.

The spatially consistent direction of deviation across all distance bands and land-use categories indicates that the proposed data-driven model generalizes underlying acoustic relationships beyond the explicit transport network. In this sense, the predictive framework effectively bridges the gap between source-based emission mapping (as represented by END) and exposure-based environmental realism, providing a more continuous and physically plausible representation of urban soundscapes.

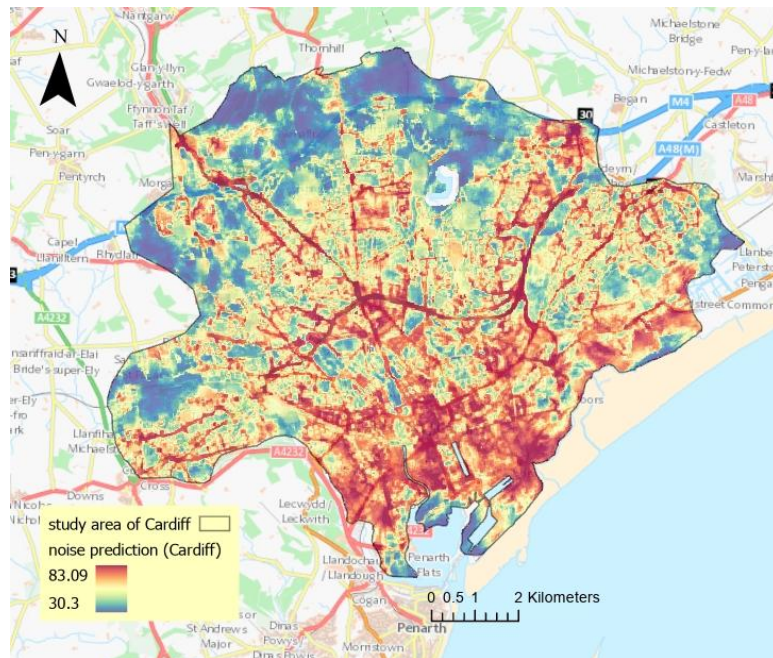
Building upon these findings, the next section performs a city-level and LULC-driven comparative analysis to examine how the magnitude and structure of these discrepancies vary across different urban morphologies. This analysis further explores whether the improved exposure continuity observed in Southampton also manifests consistently across other UK cities, thereby providing insights into the spatial transferability and functional robustness of the proposed framework.

6.3.3.3 City-Level Local Comparisons with LULC and END in target cities

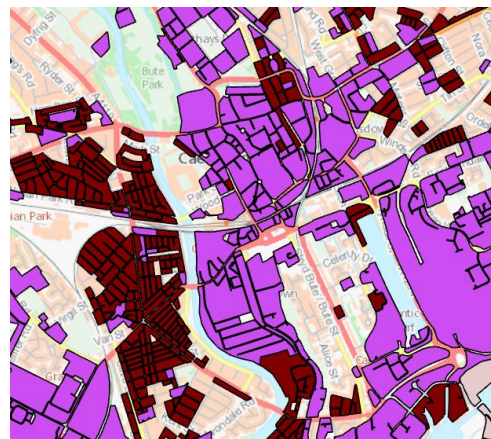
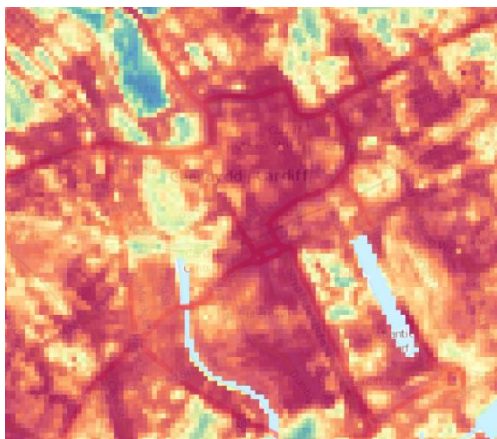
To further elucidate the environmental information encoded in the predicted noise maps, this section presents a set of representative local case studies that compare the model outputs with both European Environmental Noise Directive (END) LAeq,16h traffic noise maps and urban land-use/land-cover (LULC) data derived from European Environment Agency Urban Atlas 2012. The analytical motivation parallels that of Section 6.3.3.2: while END maps effectively capture high exposure along major transport corridors, their design does not account for non-traffic sources such as port operations, commercial activities, or industrial emissions, nor do they represent the attenuating effects of vegetation and open-space buffers. By integrating three complementary spatial layers—predicted noise, END, and LULC—we can visually and contextually assess the degree to which the model has learned to represent both core traffic noise and secondary or modulating environmental factors.

The selection of subregions for this analysis follows a targeted sampling strategy. Areas were chosen to represent distinct morphological and functional typologies—dense urban cores, industrial–port complexes, mixed-use residential neighborhoods, and vegetated or peri-urban fringe zones. This allows us to probe whether the model’s predictions simply mirror traffic density or whether they encode a richer spatial signal consistent with heterogeneous acoustic sources and propagation mechanisms. In the case of Cardiff, which lacks END LAeq,16h coverage, the comparison is conducted between predicted noise maps and LULC layers only.

The results highlight clear, interpretable spatial structures. In Cardiff (Figure 6.8), for instance, a strong exposure gradient emerges from the southeastern bayfront and city center toward the northwestern green fringe, reflecting both morphological and functional land-use transitions. Within the Civic Centre and Bayfront area (Figure 6.8b), the predicted high-exposure zone coincides with dense continuous urban fabric, industrial and commercial land, and port areas, reflecting compound noise contributions from multimodal transport hubs, port operations, and central business activities. In contrast, the Wenallt and Northern Hills (Figure 6.8c) subregion shows extensive low-exposure patterns corresponding to green urban areas, pastures, forests, and arable land, highlighting the model’s ability to encode background attenuation in vegetated, low-density landscapes—patterns that are absent from END’s binary, road-focused representation in other cities.

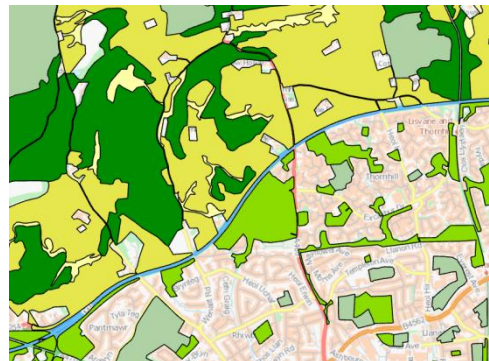
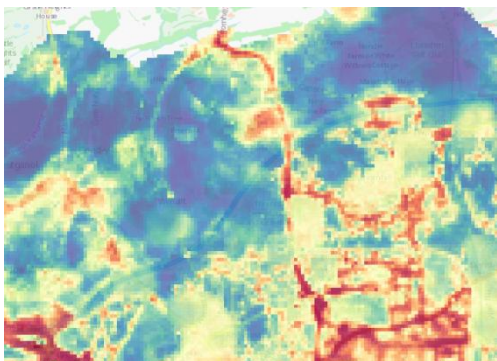


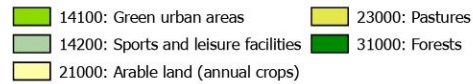
(a) Predicted noise map for Cardiff



- 11100: Continuous Urban fabric (S.L. > 80%)
- 12100: Industrial, commercial, public, military and private units
- 12300: Port areas

(b) Civic centre with high noise exposure





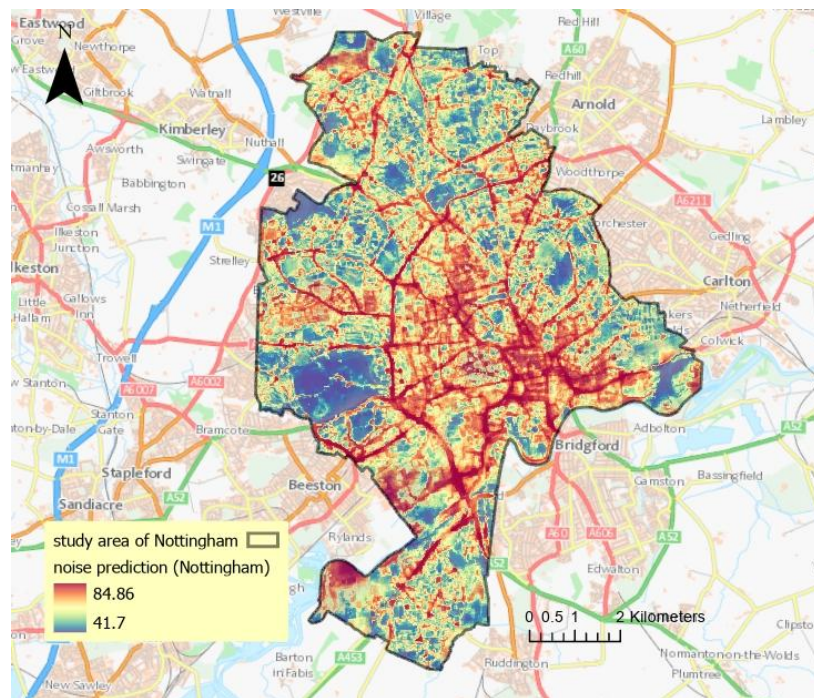
(c) *Wenallt and Northern Hills with low noise exposure*

Figure 6.8 Predicted noise map for Cardiff and selected sub-regions; (a) Noise prediction map for Cardiff (b) Civic centre with high noise exposure (c) Wenallt and Northern Hills with low noise exposure

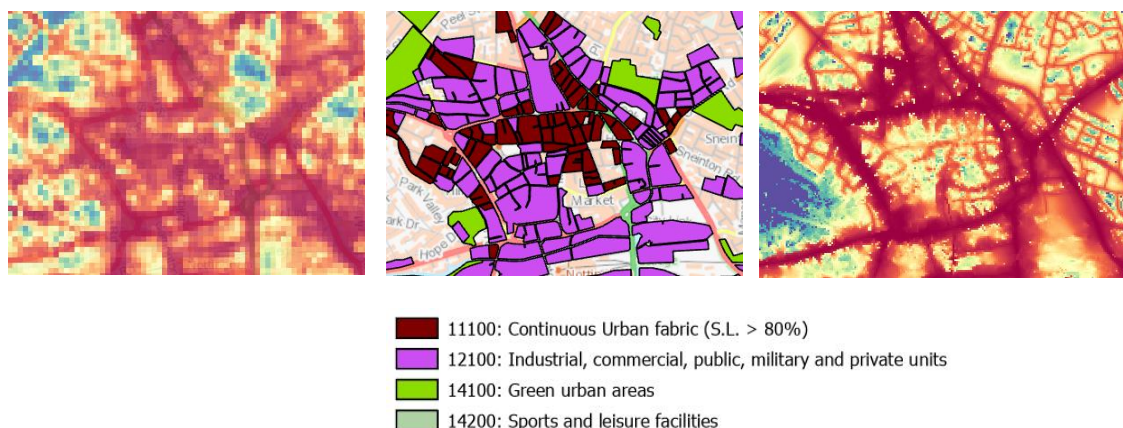
To further elucidate the environmental information encoded in the predicted noise maps, this section presents a set of representative local case studies that compare the model outputs with both END LAeq,16h traffic noise maps and urban land-use/land-cover (LULC) data derived from European Environment Agency Urban Atlas 2012. The analytical motivation parallels that of Section 6.3.3.2: while END maps effectively capture high exposure along major transport corridors, their design does not account for non-traffic sources such as port operations, commercial activities, or industrial emissions, nor do they represent the attenuating effects of vegetation and open-space buffers. By integrating three complementary spatial layers—predicted noise, END, and LULC—we can visually and contextually assess the degree to which the model has learned to represent both core traffic noise and secondary or modulating environmental factors.

The selection of subregions for this analysis follows a targeted sampling strategy. Areas were chosen to represent distinct morphological and functional typologies—dense urban cores, industrial–port complexes, mixed-use residential neighborhoods, and vegetated or peri-urban fringe zones. This allows us to probe whether the model’s predictions simply mirror traffic density or whether they encode a richer spatial signal consistent with heterogeneous acoustic sources and propagation mechanisms. In the case of Cardiff, which lacks END LAeq,16h coverage, the comparison is conducted between predicted noise maps and LULC layers only.

The results highlight clear, interpretable spatial structures. In Cardiff, for instance, a strong exposure gradient emerges from the southeastern bayfront and city center toward the northwestern green fringe, reflecting both morphological and functional land-use transitions. Within the Civic Centre and Bayfront area, the predicted high-exposure zone coincides with dense continuous urban fabric, industrial and commercial land, and port areas, reflecting compound noise contributions from multimodal transport hubs, port operations, and central business activities. In contrast, the Wenallt and Northern Hills subregion shows extensive low-exposure patterns corresponding to green urban areas, pastures, forests, and arable land, highlighting the model’s ability to encode background attenuation in vegetated, low-density landscapes—patterns that are absent from END’s binary, road-focused representation in other cities.



(a) Predicted noise map for Nottingham



(b) City Core around Old Market Square and Nottingham Station

Figure 6.9 Predicted noise map for Nottingham and selected sub-regions; (a) Noise prediction map for Nottingham (b) City Core around Old Market Square and Nottingham Station

In Portsmouth (Figure 6.10), the predictive framework demonstrates a particularly pronounced clustering of high-noise exposure across the Portsmouth and Southsea subregion, reflecting the city's compact urban morphology, extensive coastal logistics activities, and mixed-use development patterns. In the predicted maps, large contiguous zones of elevated noise appear along the Western Commercial Road corridor and the east-central coastal area encompassing Milton and Eastney, forming a continuous high-exposure belt that extends beyond the immediate vicinity of major roads. This spatial configuration corresponds closely with the underlying land-use composition—dominated by industrial and commercial parcels and extensive port areas—as represented in the European Environment Agency Urban Atlas 2012 dataset.

When juxtaposed with theEND LAeq,16h maps, the model outputs show both alignment and extension. High-exposure zones near major transport arteries are

captured in both datasets, indicating that the model retains sensitivity to traffic-related emissions in its core domain. However, the END maps systematically underrepresent high-exposure clusters within port and logistics zones, as these are not included in its road-based emission modeling framework. By contrast, the predicted map effectively captures additional industrial and maritime acoustic contributions, resulting in a more spatially complete and environmentally plausible exposure field.

This pattern underscores the added value of LULC-informed feature design within the dual-branch GNN architecture. By integrating both local and contextual land-use signals, the model can reflect non-traffic noise sources that are structurally embedded in the urban fabric of port cities like Portsmouth. This leads to a spatially continuous high-exposure pattern, consistent with the city’s functional morphology and coastal economic activities—an aspect that END maps alone fail to capture.

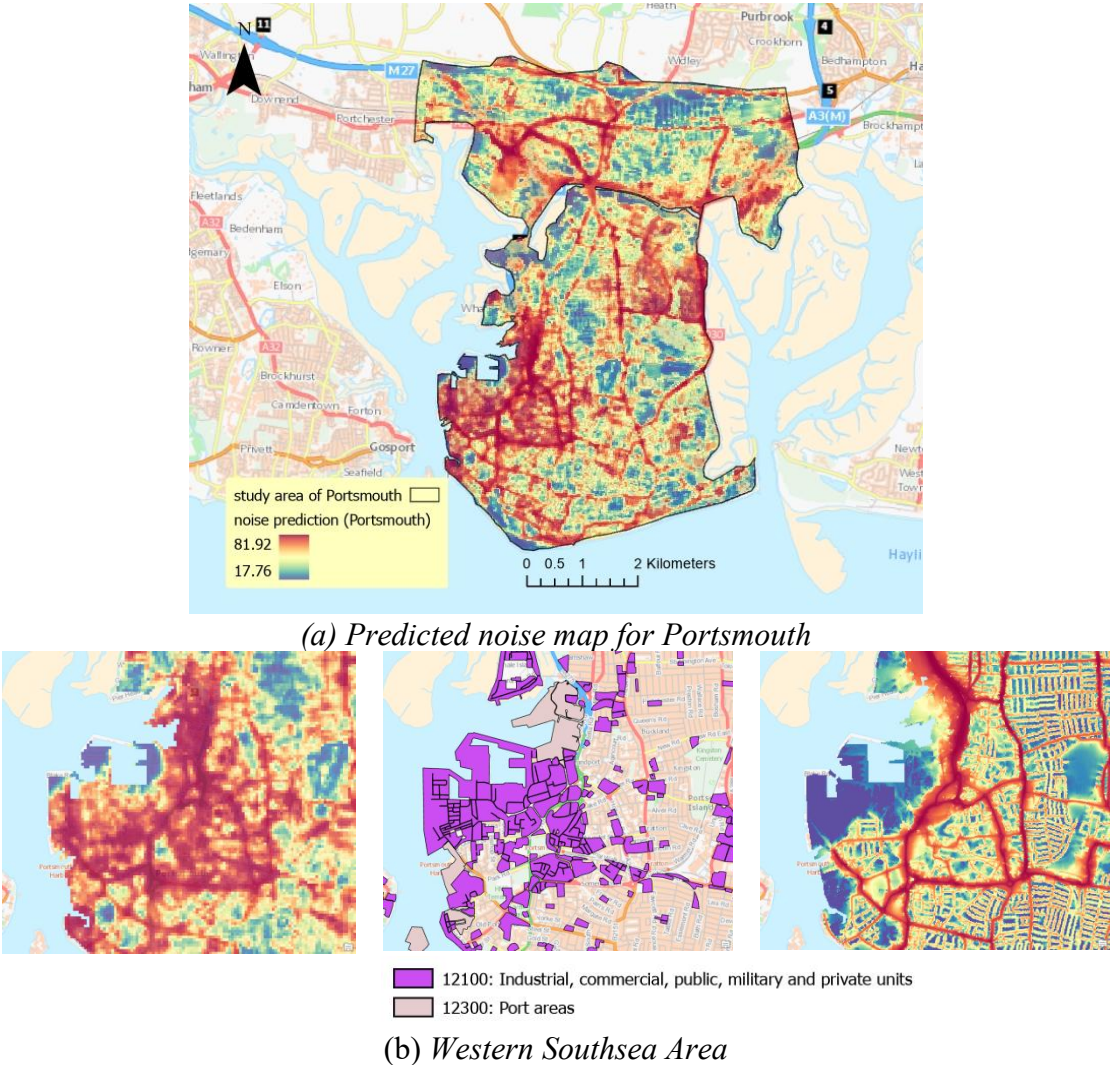
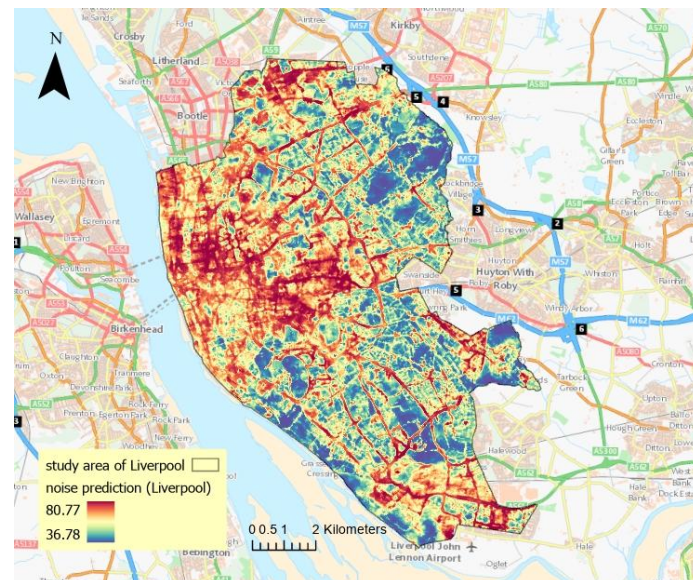


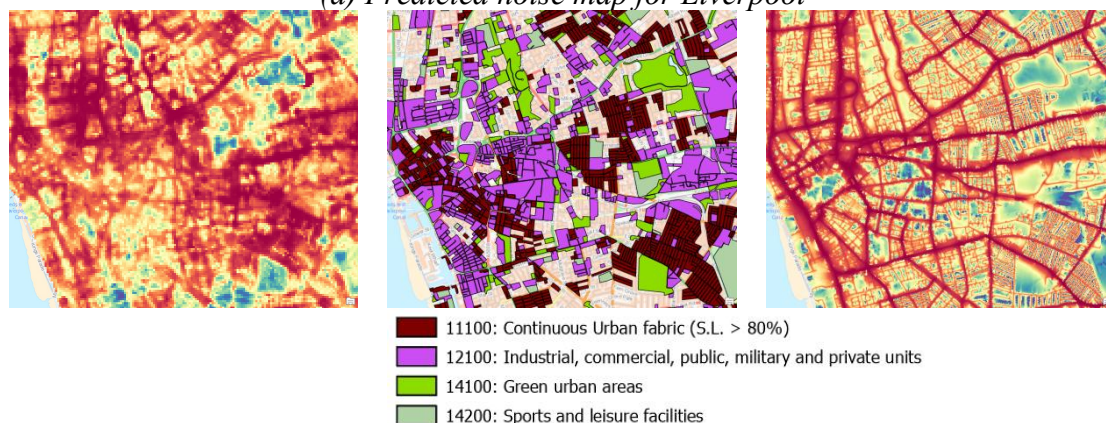
Figure 6.10 Predicted noise map for Portsmouth and selected sub-regions; (a) Noise prediction map for Portsmouth (b) Western Southsea Area

In Liverpool (Figure 6.11), the predicted noise map exhibits a broad alignment with END LAeq,16h maps along major transportation corridors, notably the M62 and A5047, which traverse the urban core from east to west. However, beyond this traffic-dominated baseline, the model reveals a large, continuous cluster of high-noise exposure extending from Vauxhall through Edge Hill to Old Swan—an area characterized by dense road networks and heterogeneous land-use compositions. This zone contains a mixture of high-density urban fabric, industrial and commercial parcels, and interspersed patches of green open space, which together create a complex acoustic environment that is only partially represented in END.

The model successfully captures the clustering structure of high exposure along major roads and intersections, which is consistent with expected traffic emissions. At the same time, the degree of attenuation within and around urban green areas varies as a function of their spatial extent: smaller, fragmented green spaces produce limited mitigation effects, whereas larger vegetated areas exert more substantial noise damping. This nuanced representation contrasts sharply with END’s binary treatment of road-adjacent exposure, which does not differentiate between these local landscape effects.



(a) Predicted noise map for Liverpool



(b) Civic center area in Liverpool

Figure 6.11 Predicted noise map for Liverpool and selected sub-regions; (a) Noise prediction map for Liverpool (b) Areas between Lime Street Station and the Knowledge Quarter

Collectively, the Liverpool case reinforces the broader pattern observed across all four END-covered cities. The dual-branch GNN model aligns closely with END in transportation corridors, ensuring baseline consistency, while extending coverage to non-traffic sources and modulating land-use factors, resulting in a more continuous and environmentally plausible acoustic landscape. This capability is particularly evident in morphologically complex urban regions where traffic, land use, and environmental buffers interact in non-linear ways.

6.3.4 Functional Group-Level Consistency across Cities

A foundational premise of this research is that urban land use categories, as defined by the Urban Atlas 2012, provide reliable and consistent proxies for understanding both noise generation and mitigation patterns across diverse urban environments. Under this assumption, a model designed for cross-city generalization should be capable of producing comparable predicted noise distributions within similar land use categories, even when applied to cities for which no local noise measurements are available. To empirically assess this hypothesis, the study implemented a comprehensive land-use-stratified evaluation of the predicted noise distributions across the five study cities: Southampton, Cardiff, Portsmouth, Nottingham, and Liverpool.

The strength of this evaluation rests on the robust feature engineering framework outlined earlier in Section 6.2.4.4. From an initial set of 3,780 candidate variables derived from multispectral imagery and land use indicators, a carefully filtered and ranked feature space of 169 final input variables was constructed—84 feeding the local branch and 85 feeding the background branch of the dual-branch GNN architecture. Crucially, although the feature selection was conducted on the labeled reference city (Southampton), it was informed by the full cross-city feature set, ensuring that the retained predictors held meaningful and generalizable patterns across urban domains. This compact yet semantically rich feature space enabled the model to operate efficiently while maintaining high generalization capacity.

Direct evaluation of prediction accuracy in target cities was not possible due to the lack of ground-truth noise data. Therefore, an indirect but theoretically grounded validation approach was employed. Specifically, the predicted noise distributions were compared across cities within each land use type. The logic is straightforward: if the model has truly internalized structurally meaningful relationships between urban form and noise exposure, it should produce comparable prediction patterns across semantically equivalent land use categories, regardless of geographic context. This type of latent consistency check aligns with well-established evaluation practices in domain adaptation research, where internal distributional coherence is often used as a proxy for generalization success.

To structure this analysis, the study implemented three progressively aggregated evaluation layers. At the most detailed level, fine-grained boxplots were generated comparing predicted noise distributions for the main Urban Atlas land use codes across all cities. Considering that certain land use categories—such as “Motorways and Expressways” (12210)—have very limited sample sizes and thus lack strong statistical meaning, their results were interpreted with caution rather than excluded outright, recognizing their conceptual importance in the urban acoustic landscape. At an intermediate level, the analysis focused on the subset of urban fabric classes, visualizing their predicted noise gradients through mean-trend plots that traced the relationship between predicted noise levels and urban density gradients. Finally, at the highest aggregation, all land use categories were grouped into four semantically meaningful functional groups—strong sources, moderate sources, moderate mitigation, and strong mitigation—and their cross-city distributions were visualized and compared.

This multilayered evaluation design offers a robust triangulation of the model’s generalization capacity, illuminating not only whether the predicted distributions align statistically across cities but also whether they preserve meaningful structural relationships between urban morphology and environmental noise exposure.

Urban Fabric Classes: Capturing Gradual Transitions in Predicted Noise

The Urban Atlas 2012 defines a systematic typology of urban fabric classes, ranging from highly compact continuous urban fabric (code 11100) to progressively more dispersed and lower-density zones, such as discontinuous dense (11210), medium-density (11220), and low-density (11230) urban fabrics. These classes reflect fundamental gradients of built form, density, and infrastructure intensity — factors long known to influence environmental noise propagation.

To assess the model’s sensitivity to this morphological gradient, the study extracted all grid cells associated with these four key urban fabric classes across all five study cities. For each class, the predicted mean noise levels were computed and aggregated at the city level. These values were then visualized through a mean-trend plot (Figure 6.12), which captures not only the average predicted exposure but also the underlying distributional structure within each city and across the urban density gradient.

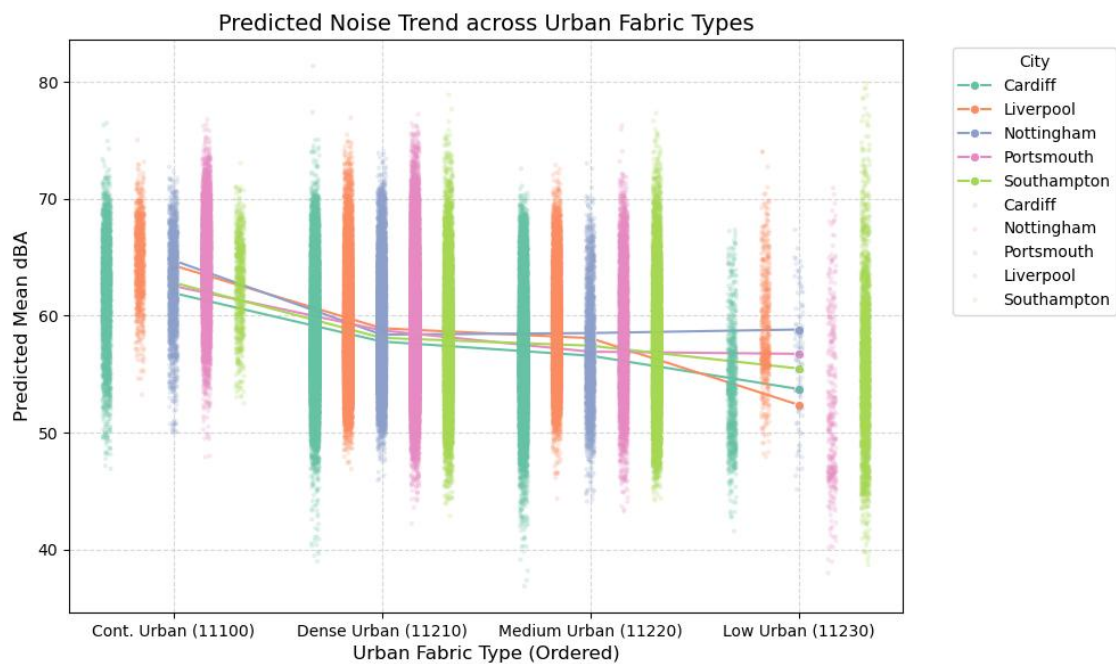


Figure 6.12 Predicted Noise Trend across Urban Fabric Types

The results reveal a strikingly consistent monotonic decline in predicted noise levels across increasing urban sparsity. For instance, across all five cities, continuous urban fabric areas (11100) show the highest predicted noise means, reflecting the dense concentration of traffic, narrow streets, and reflective surfaces characteristic of such zones. As the analysis progresses to discontinuous dense (11210) and medium-density (11220) classes, the predicted noise levels steadily decrease, culminating in the lowest predictions for discontinuous low-density areas (11230), where larger lot sizes, more vegetation, and lower traffic volumes provide natural attenuation.

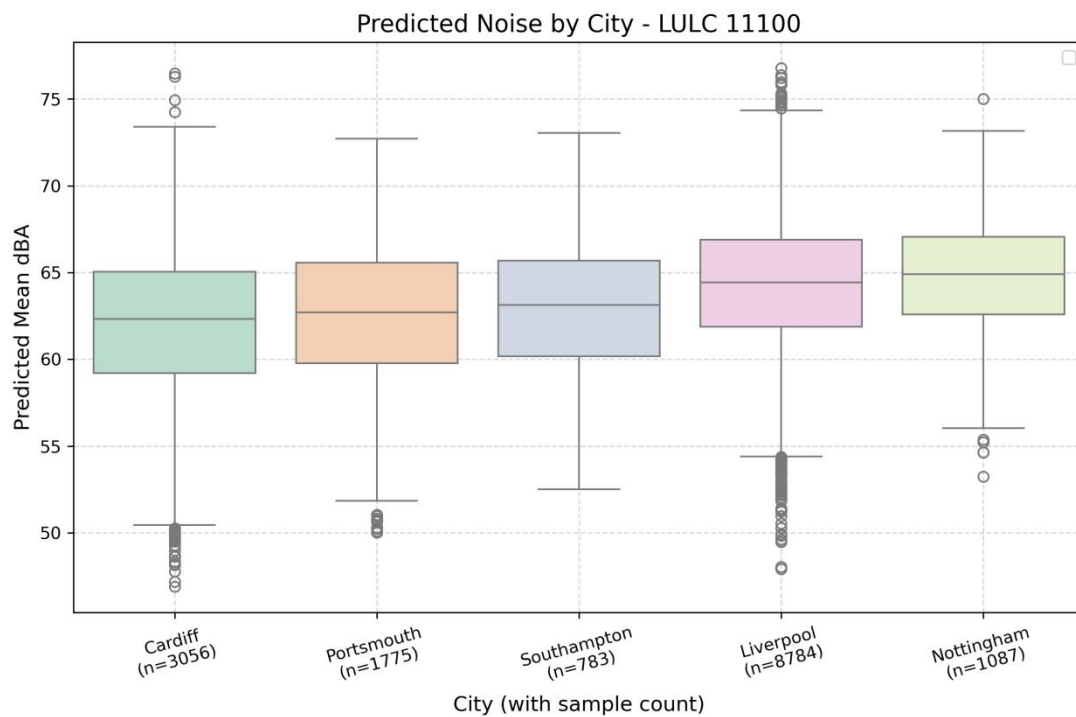
Importantly, this pattern holds across geographically and morphologically diverse cities. For example, in the dense historic cores of Southampton and Liverpool, as well as in the more dispersed suburban areas of Cardiff and Nottingham, the model predicts noise levels within a narrow 3–4 dBA range for equivalent urban fabric classes. This consistency is particularly significant given that no target-city noise labels were used during training, underscoring the robustness of the cross-city generalization.

The ability to accurately reflect such gradual and systematic environmental gradients speaks directly to the strengths of the dual-branch GNN design. The local branch effectively captures fine-scale spatial heterogeneity (such as local road adjacency and block-scale morphology), while the background branch contextualizes this information against broader urban patterns (such as proximity to green buffers or industrial zones). Together, these components allow the model to distinguish nuanced acoustic differences along the urban density continuum, aligning closely with findings in environmental acoustics research, where urban canyon effects and density have been shown to amplify or attenuate noise exposure. The inclusion of this urban gradient analysis complements the boxplot-based analysis by adding a continuous structural perspective. The observed pattern aligns with the environmental acoustic literature, where built-up density and urban canyon effects are positively associated with increased noise exposure.

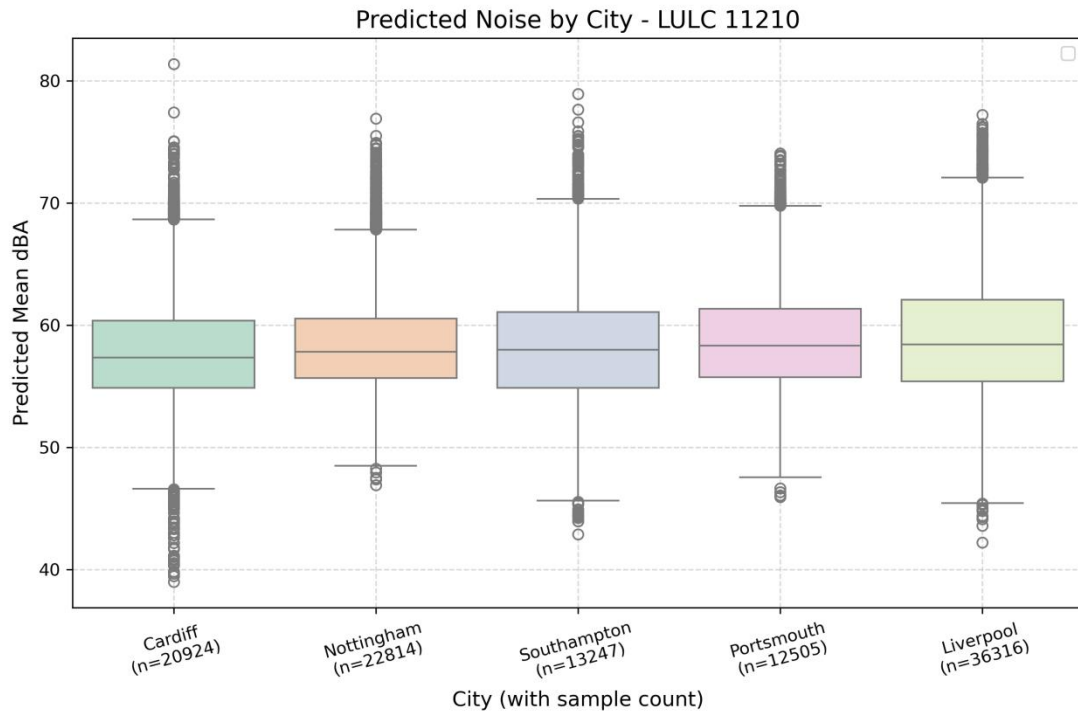
LULC-Type Specific Predicted Noise Distributions

Beyond the urban fabric gradient, it is essential to examine how the model performs across the full range of land use and land cover (LULC) categories defined in the Urban Atlas 2012. This analysis probes whether the predicted noise values remain consistent across cities within the same LULC type, which serves as a powerful indirect indicator of generalization in the absence of ground truth.

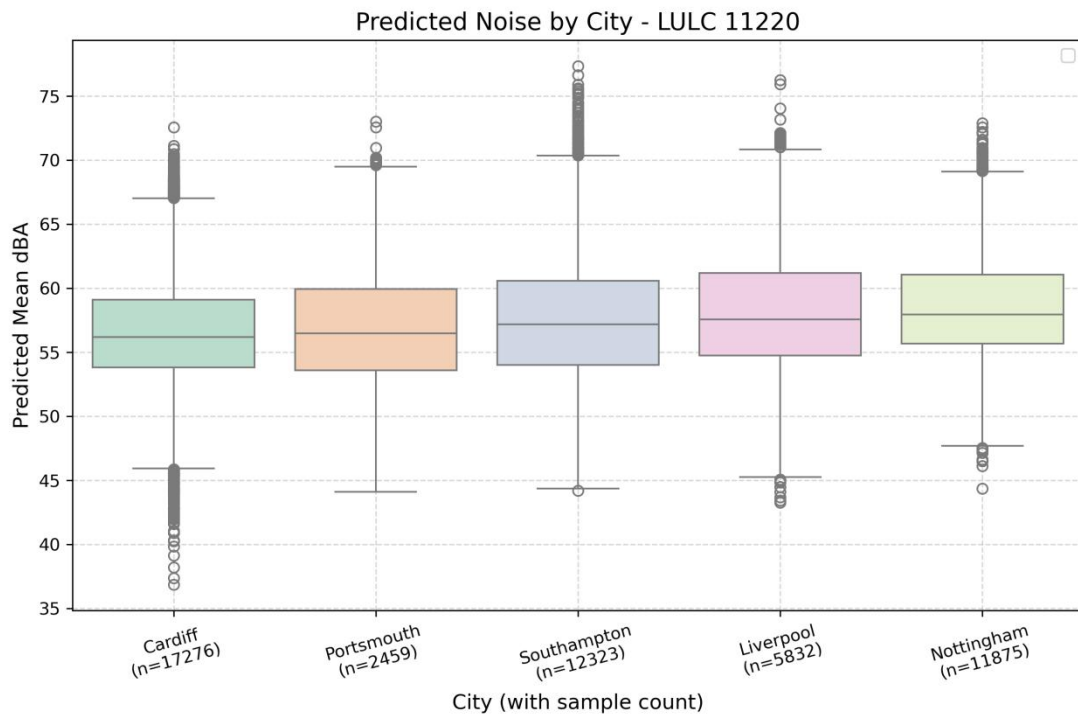
For this purpose, predicted mean noise values were spatially joined to LULC polygons using GIS-based overlay analysis. This procedure generated a detailed matrix of predicted noise distributions across 24 distinct Urban Atlas LULC classes for each of the five cities. The aggregated results were visualized as a series of comparative boxplots (Figure 6.13), offering a comprehensive, category-level perspective on cross-city consistency.



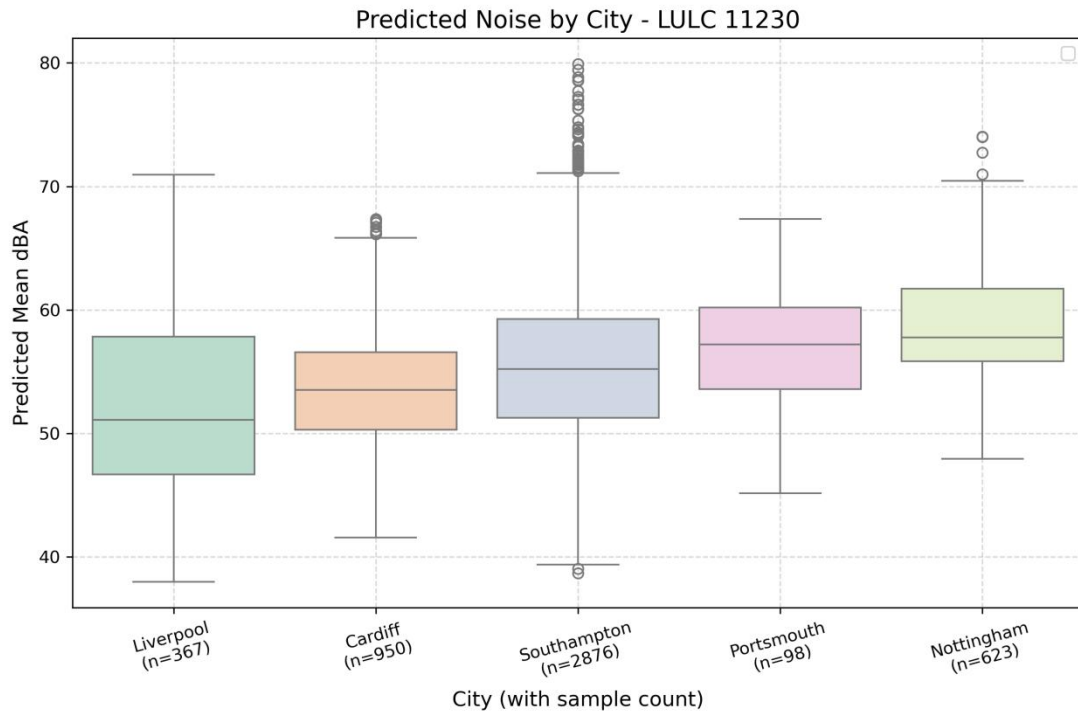
(a) Continuous Urban fabric



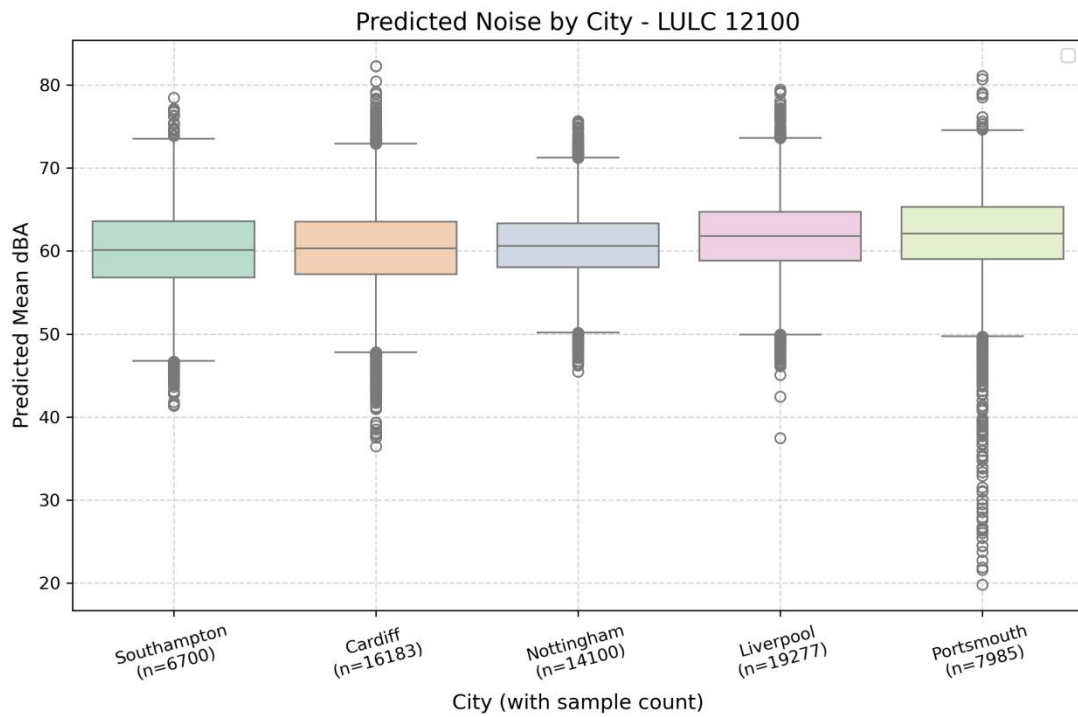
(b) Discontinuous Dense Urban fabric



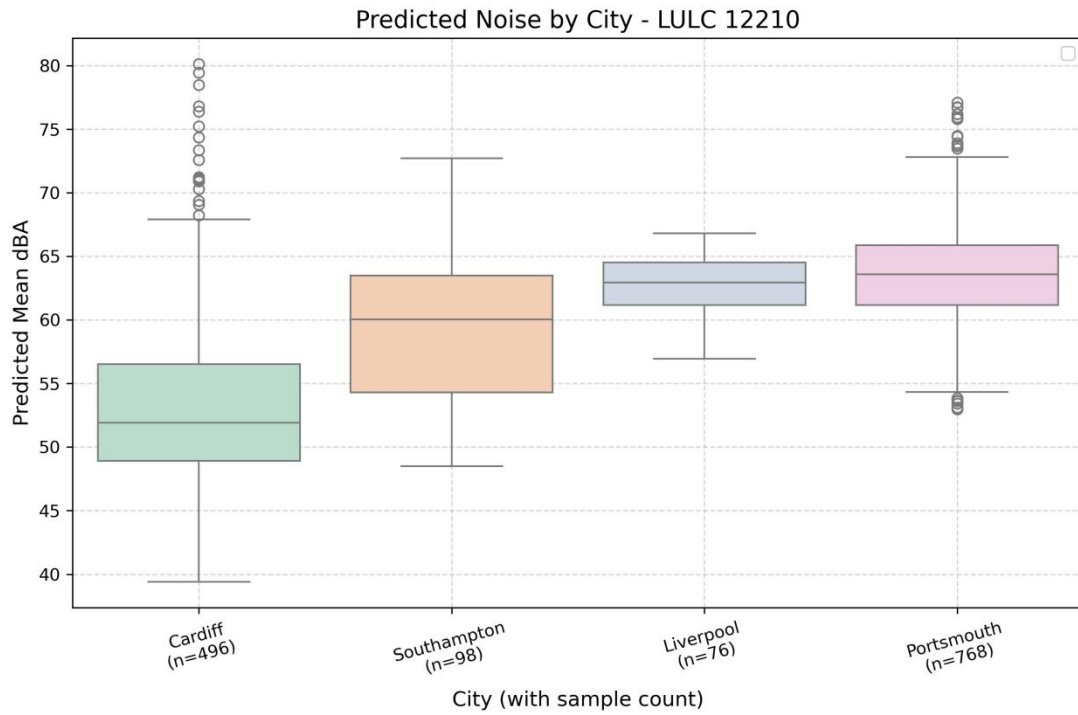
(c) Discontinuous Medium Density Urban fabric



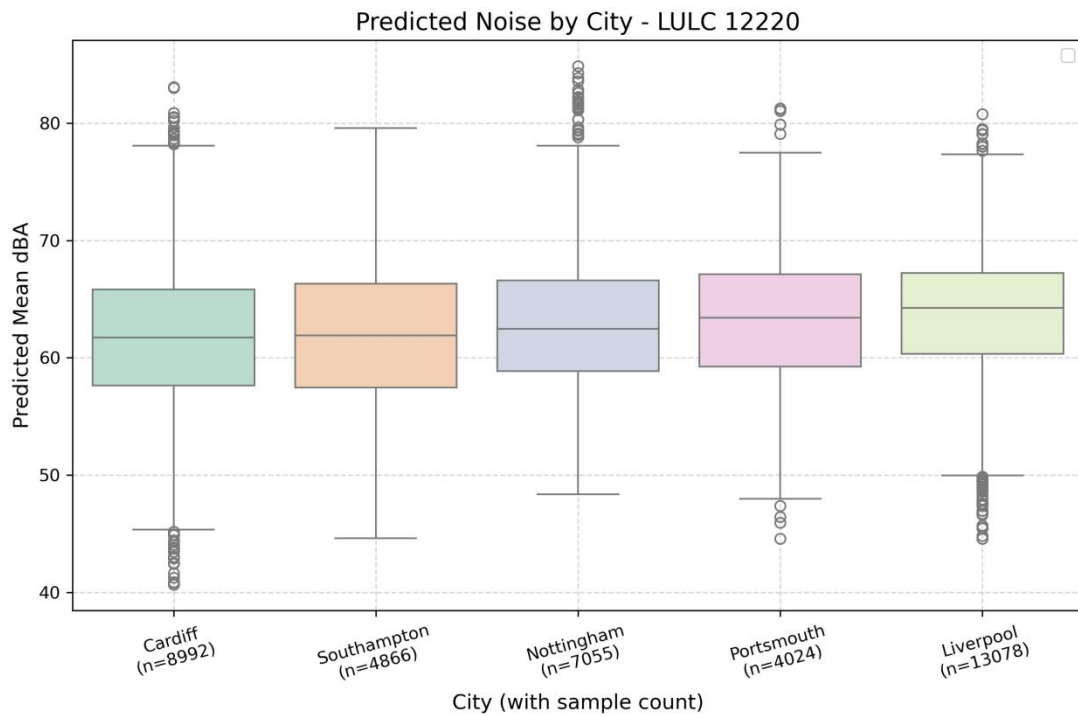
(d) Discontinuous Low Density Urban fabric



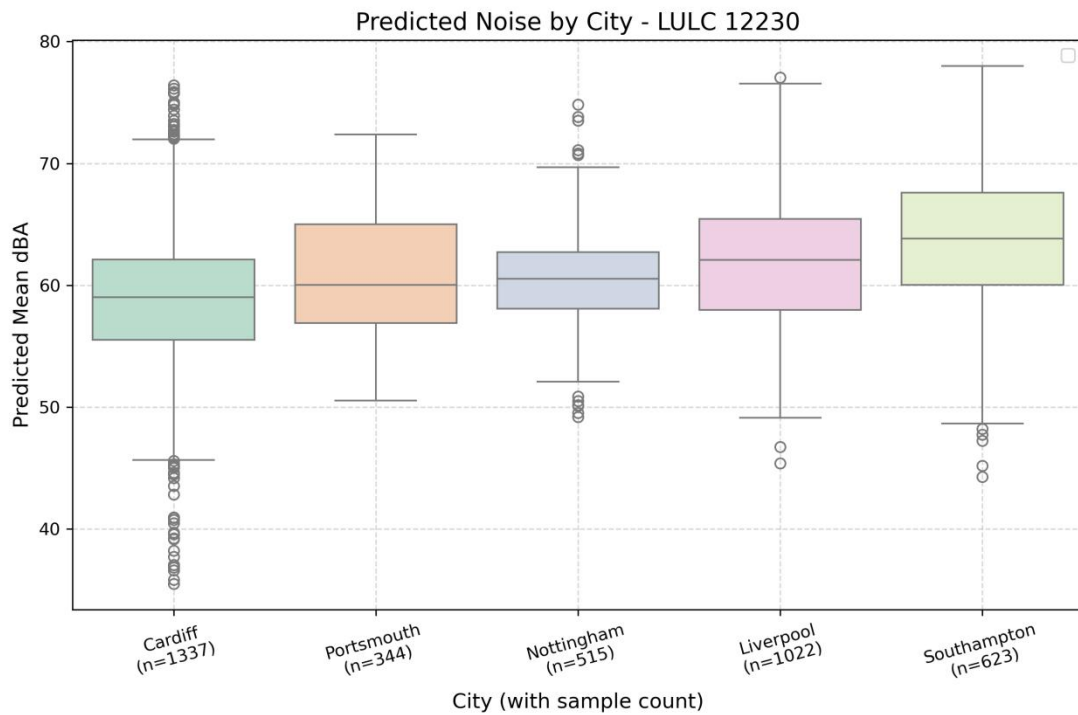
(e) Industrial, commercial, public, military and private units



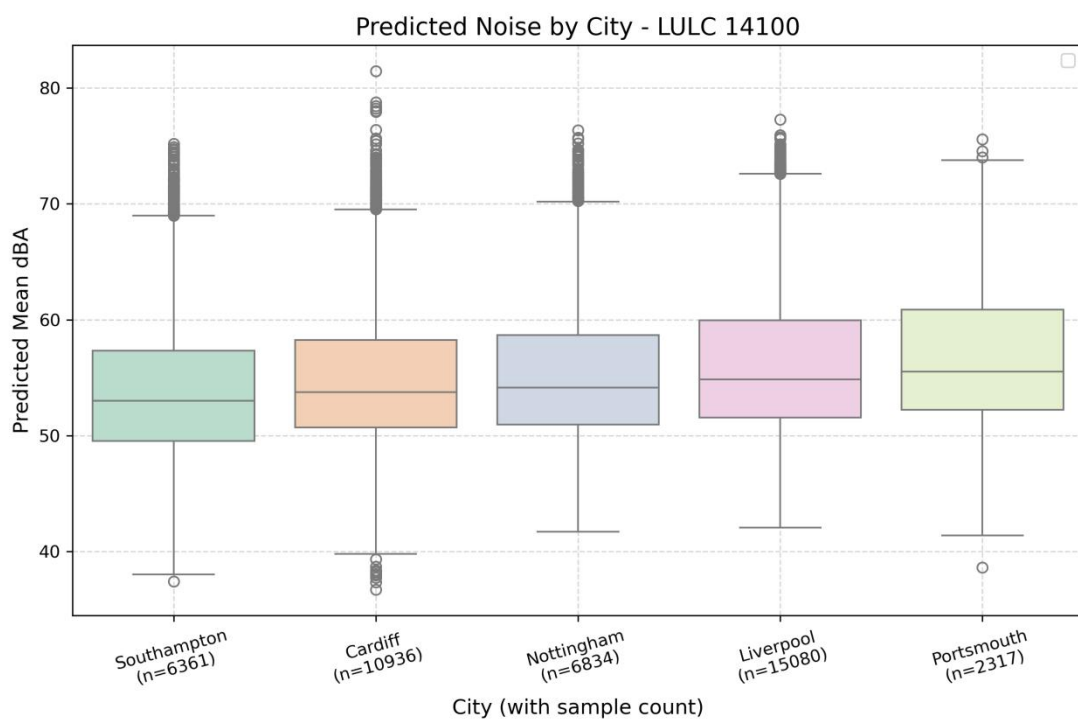
(f) Fast transit roads and associated land



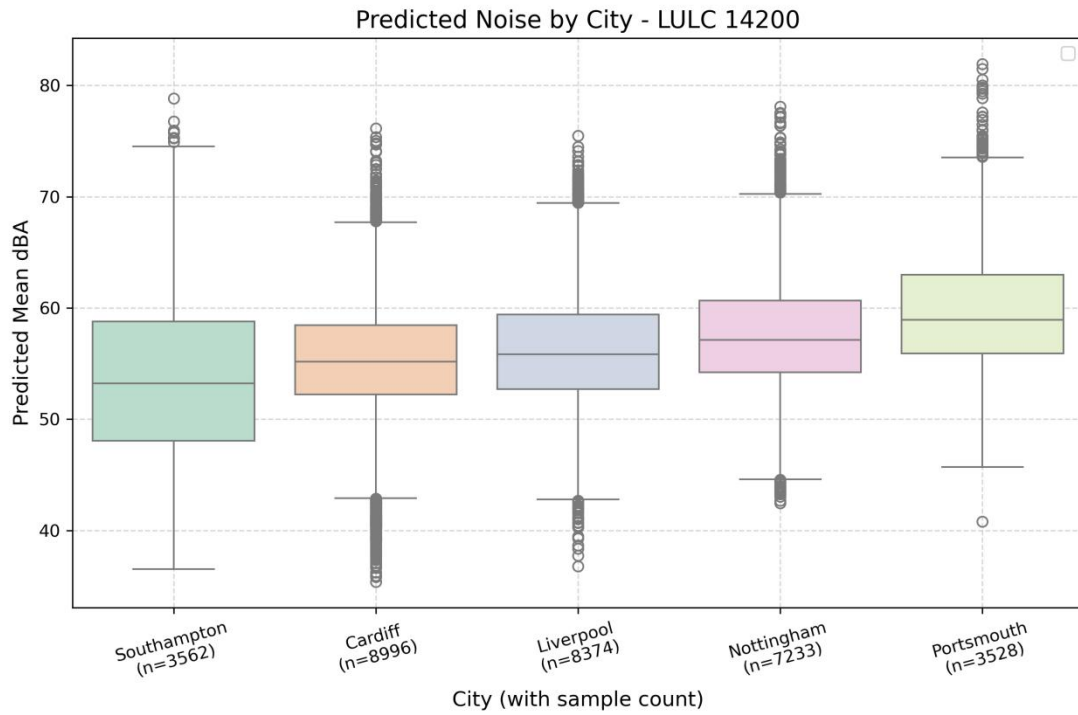
(g) Other roads and associated land



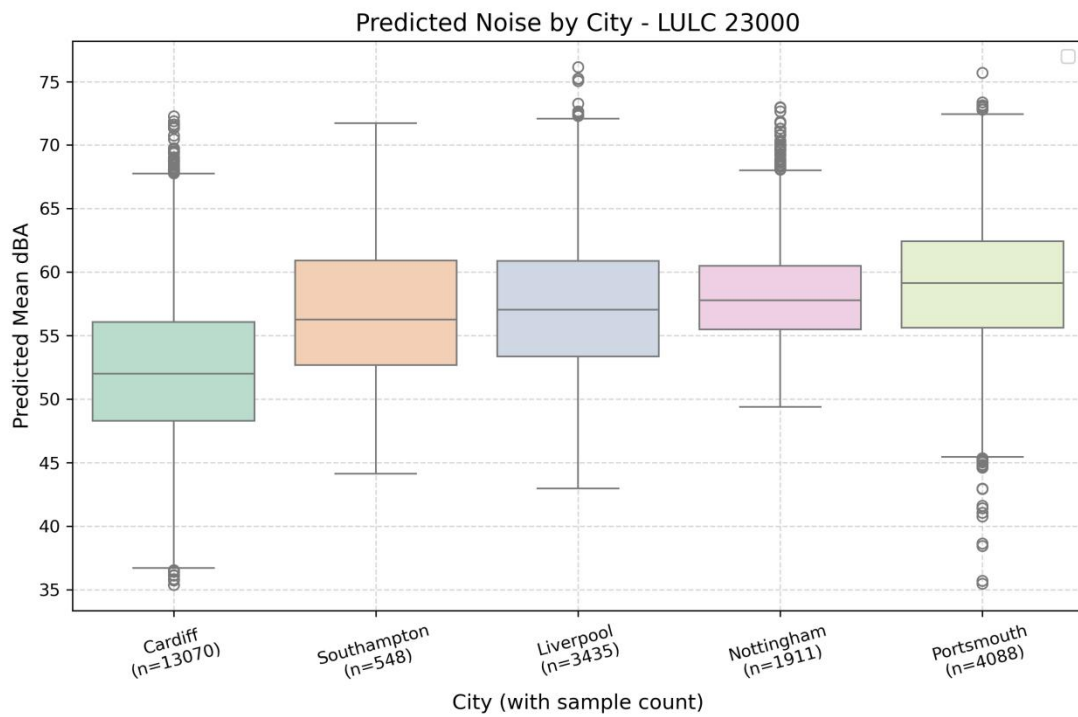
(h) Railways and associated land



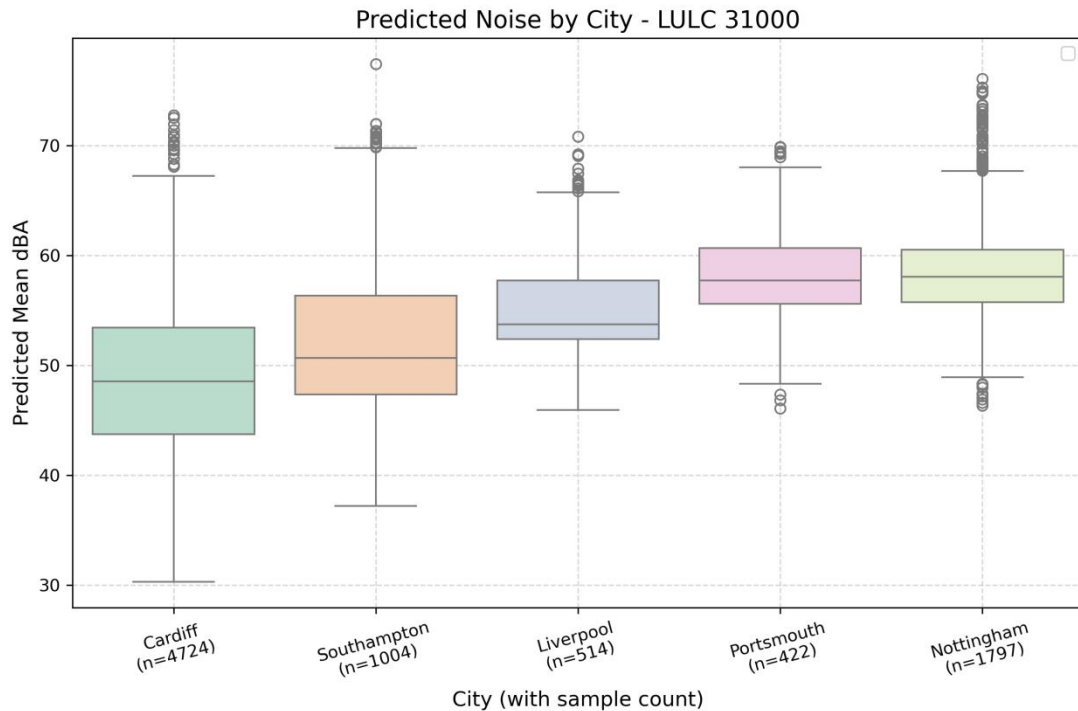
(i) Green Urban areas



(j) Sports and leisure facilities



(k) Pastures



(l) Forests

Figure 6.13 LULC-Type Specific Predicted Noise Distributions; (a) Continuous Urban fabric; (b) Discontinuous Dense Urban fabric; (c) Discontinuous Medium Density Urban fabric; (d) Discontinuous Low Density Urban fabric; (e) Industrial, commercial, public, military and private units; (f) Fast transit roads and associated land; (g) Other roads and associated land; (h) Railways and associated land; (i) Green Urban areas; (j) Sports and leisure facilities; (k) Pastures; (l) Forests

Several notable patterns emerge from this evaluation. First, built-up areas—including continuous urban fabric (11100), discontinuous dense (11210), and industrial/commercial zones (12100)—show strong alignment of predicted noise medians across cities. These areas consistently register higher noise values, reflecting the combined influence of traffic, high-density infrastructure, and minimal green buffering. The clustering of predictions within narrow interquartile ranges further indicates that the model is robustly capturing domain-invariant acoustic signals associated with intensive urban functions.

In contrast, green and open space classes, such as green urban areas (14100), sports facilities (14200), and pastures (23000), exhibit lower predicted noise levels but with wider spreads. This variation is expected, as such classes can encompass a broad spectrum of spatial configurations, from small urban parks to large semi-natural expanses, each with differing noise attenuation properties. Interestingly, forested areas (31000) and wetlands (40000) show even greater prediction variability, highlighting the challenges inherent in standardizing natural land cover effects on noise propagation across cities.

Additionally, it is important to note that some LULC classes were underrepresented or absent in specific cities, such as fast transit roads (12210) in Nottingham, leading to either missing or highly variable boxplots. This underscores the importance of

ensuring adequate area-weighted representation when evaluating cross-city consistency.

Overall, despite these local variations, the general alignment of interquartile ranges and medians across most LULC types strongly suggests that the model has internalized stable, generalizable associations between environmental attributes and predicted acoustic outcomes. This result is particularly compelling given the use of a highly compressed feature space (171 selected variables, reduced from over 8,000 candidates), emphasizing that careful feature selection and robust graph-based learning can yield reliable generalization across diverse urban settings.

Cross-City Patterns in Functional LULC Groups

Building upon the detailed LULC-class analysis, the study further aggregated all 24 Urban Atlas land cover classes into four semantically meaningful functional groups: strong noise sources, moderate noise sources, moderate noise mitigation zones, and strong noise mitigation zones. This aggregation, based on the logic established in Section 6.2.4.3, offers a generalized lens through which to assess cross-city consistency of predicted noise outcomes at a functionally interpretable level.

Strong noise sources include major transport infrastructures such as highways, railways, airports, and extractive industrial areas, while moderate noise sources encompass secondary roads and mixed-use industrial-commercial zones. Moderate mitigation areas consist of parks, recreational green spaces, and pastures, and strong mitigation zones comprise forests, wetlands, and extensive arable lands. Grouping at this functional level is essential for policy relevance, as it aligns with the typical urban planning categories used in environmental noise regulation and mitigation strategies.

The predicted noise distributions for these functional groups were extracted across all five cities and visualized using grouped boxplots (Figure 6.13). The results reveal several key patterns. Strong noise source zones consistently display higher median predicted noise levels, with narrow interquartile ranges across cities. This stability underscores the robustness of the model's learned representations, confirming that it effectively generalizes the acoustic signatures associated with intense transport and industrial activities.

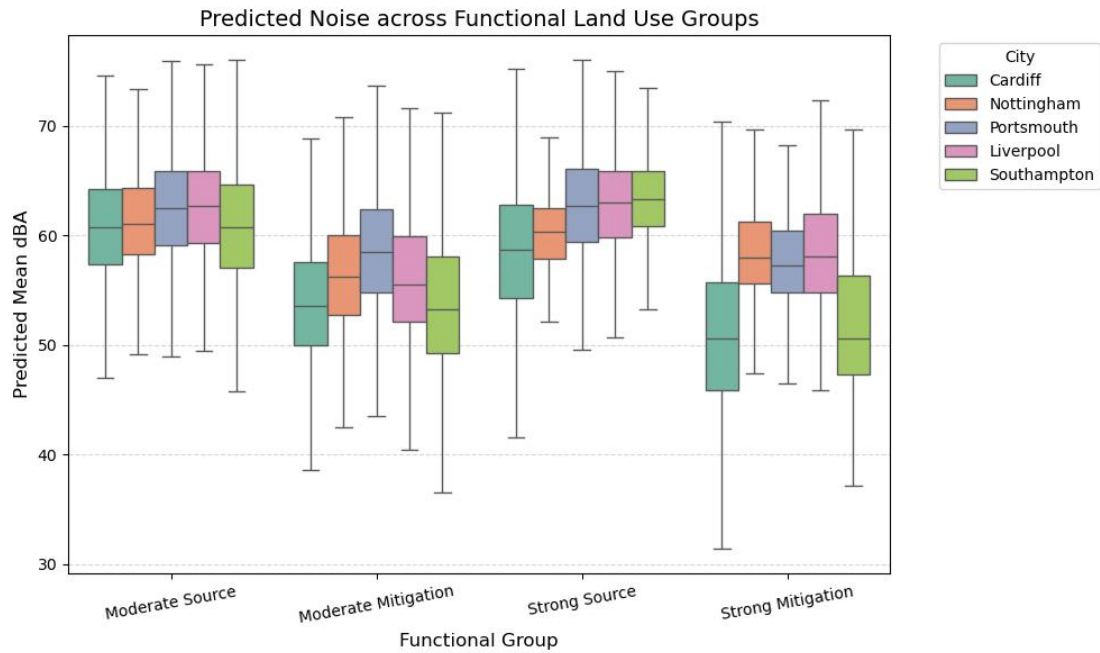


Figure 6.14 Predicted Noise across Functional Land Use Groups

Moderate noise source areas also show stable cross-city predictions, though with slightly broader variability. This may reflect real-world differences in local traffic conditions, industrial land uses, or built-environment configurations across the cities studied. In contrast, moderate and strong mitigation zones exhibit lower predicted noise medians but wider spreads, a finding that highlights the complex interactions between land cover, vegetation density, topography, and ambient sound attenuation.

Notably, some variation is observed between cities, particularly Cardiff and Southampton, where broader predicted noise distributions emerge within mitigation groups. This variation may be linked to the cities' extensive semi-natural landscapes, coastal buffers, and seasonal vegetation dynamics, which introduce greater heterogeneity in noise propagation effects.

Crucially, the general alignment of predictions within each functional group across cities supports the hypothesis that the model's generalization capacity is grounded in structurally and semantically transferable environmental features. The consistency across groups, despite the lack of ground truth in target domains, provides strong evidence for the effectiveness of the dual-branch GNN architecture, the selected feature set, and the multi-domain adversarial training strategy.

6.3.5 Embedding-Based Assessment of Feature Generalization Across Cities

Rationale and Methodology

To explore whether the dual-branch GNN captures transferable structural information across heterogeneous urban contexts, this study employed Uniform Manifold Approximation and Projection (UMAP) to visualize high-dimensional model

representations in a shared low-dimensional latent space. UMAP is particularly suitable for preserving both global and local non-linear structures in high-dimensional data and has been widely used as a diagnostic tool for representation learning and domain adaptation (McInnes et al., 2018).

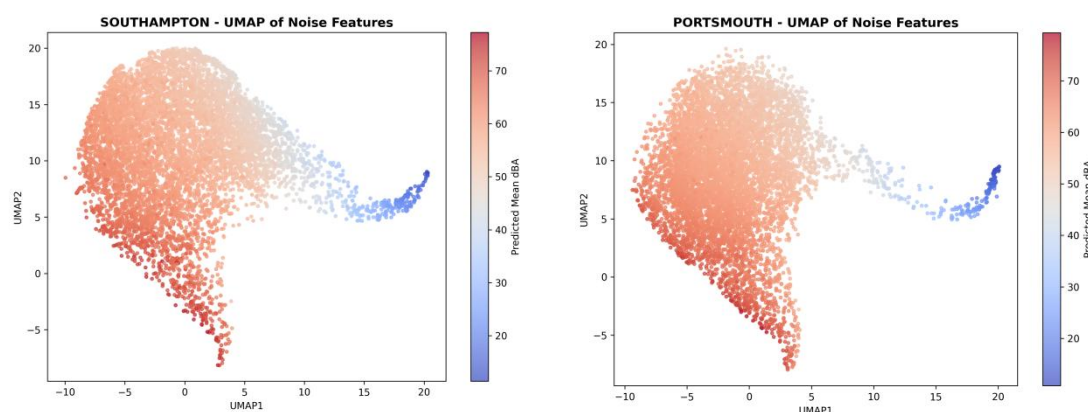
Specifically, the 64-dimensional output features from the final fusion layer of the trained model were extracted for all 30 m grid cells across the five study cities and projected into a two-dimensional latent space. It is important to note that this embedding-based analysis focuses on the internal structure of the learned representation space, not on direct predictive accuracy. Its role is to provide an interpretable geometric view of how the model organizes different urban morphological contexts after domain-adversarial training.

The conceptual expectation is that grid cells associated with functionally similar urban forms—such as arterial roads, dense residential blocks, or vegetated zones—should occupy similar regions of the latent space, regardless of their city of origin. Such a pattern would indicate that the model has internalized transferable structure-based representations, rather than memorizing city-specific spectral signatures.

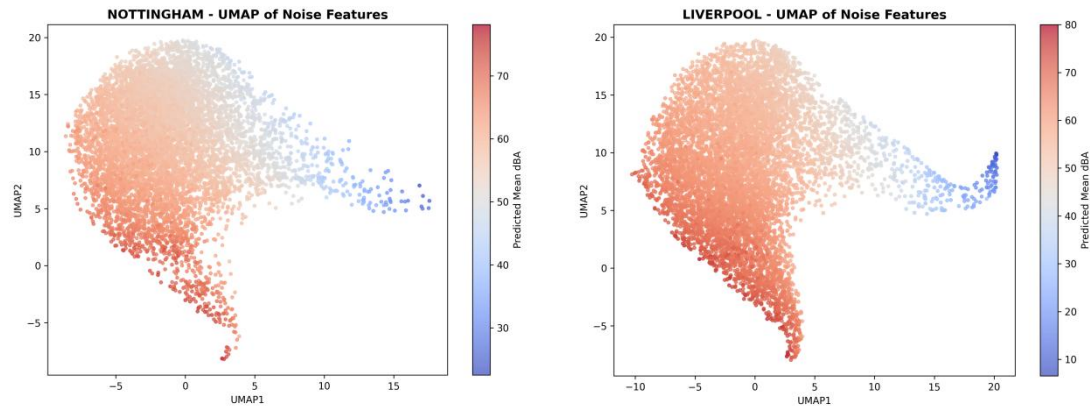
UMAP Results by City and Combined Analysis

UMAP embeddings were first generated for each city individually, with predicted mean dBA values overlaid as color gradients. As shown in Figure 6.15a–e, the embeddings display continuous and coherent gradients from high to low predicted noise levels. This pattern is observed across Southampton, Portsmouth, Nottingham, Liverpool, and Cardiff. The smooth transitions in these projections suggest that the learned features are organized according to structural-acoustic relationships rather than arbitrary or noisy latent partitions.

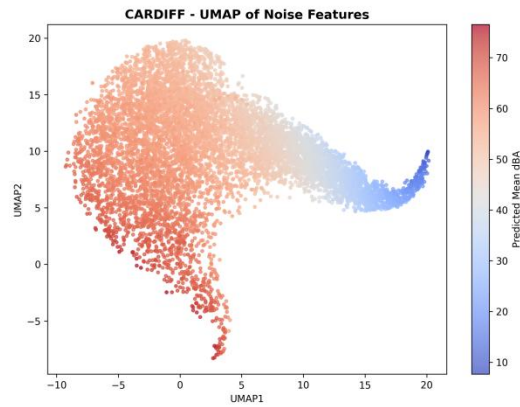
A second visualization pooled all cities into a shared UMAP space (Figure 6.16a–b). Color-coding by city reveals extensive overlap and interleaving between domains, with no sharp clustering by city label. Kernel density contours further illustrate high degrees of convergence in the dense core regions of the embedding. This structural overlap provides qualitative evidence that the model learns a common latent organization of urban noise-relevant morphology across domains..



(a) UMAP feature projections for Southampton (b) UMAP feature projections for Portsmouth



(c) UMAP feature projections for Nottingham (d) UMAP feature projections for Liverpool



(e) UMAP feature projections for Cardiff

Figure 6.15 UMAP feature projections for individual cities (a – e): Southampton, Portsmouth, Nottingham, Liverpool, and Cardiff. Each point corresponds to a 30 m grid cell, and color indicates the predicted LAeq value.

The continuous color gradients and smooth internal transitions within each city suggest that the learned 64-dimensional fused features are structurally organized in relation to acoustic exposure patterns, rather than being fragmented or dominated by local noise artifacts. These visualizations provide a qualitative view of feature organization, complementing the quantitative transferability analyses in Section 6.3.6.

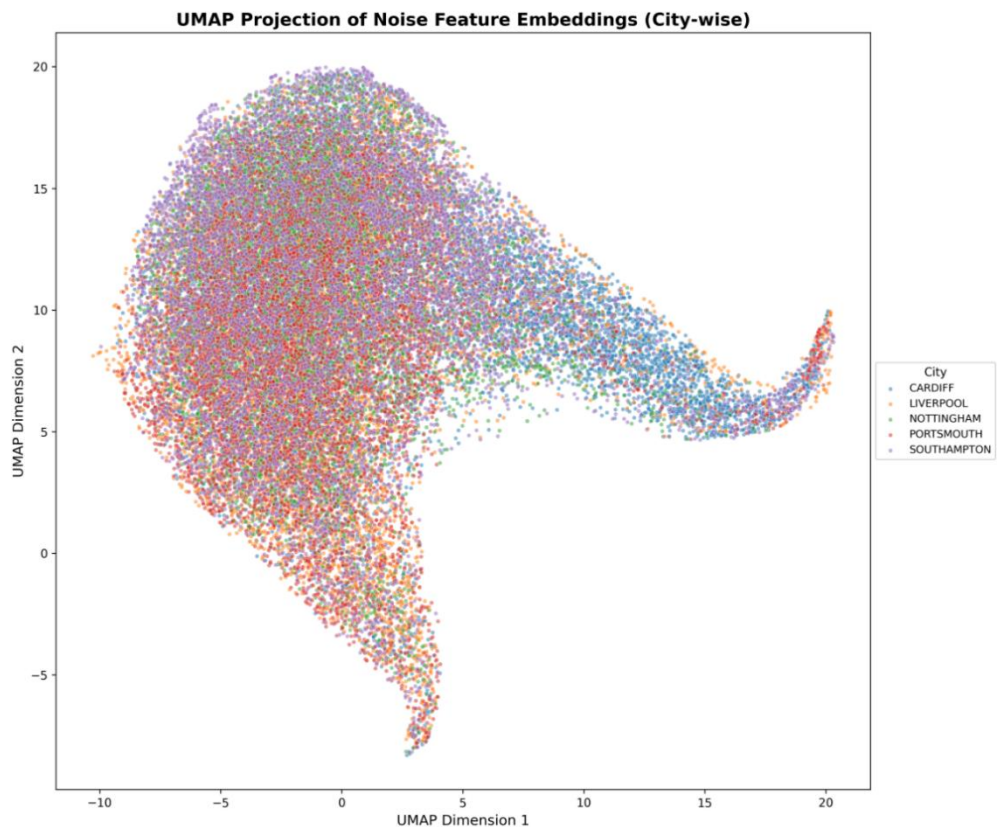
To further explore the structural alignment of learned representations across domains, we constructed a combined UMAP embedding by pooling samples from all five cities into a shared latent space. As shown in Figure 6.16a, each point is color-coded by city label. The resulting visualization exhibits extensive overlap and interleaving between cities, with no clear segregation by domain. Rather than indicating predictive equivalence, this pattern suggests that the model tends to organize morphologically similar urban contexts into comparable regions of the embedding, irrespective of their geographic origin.

To provide a clearer view of domain distributions, kernel density contours were generated for each city in the same UMAP space (Figure 6.16b). These contours show substantial spatial convergence, particularly in the dense core regions of the

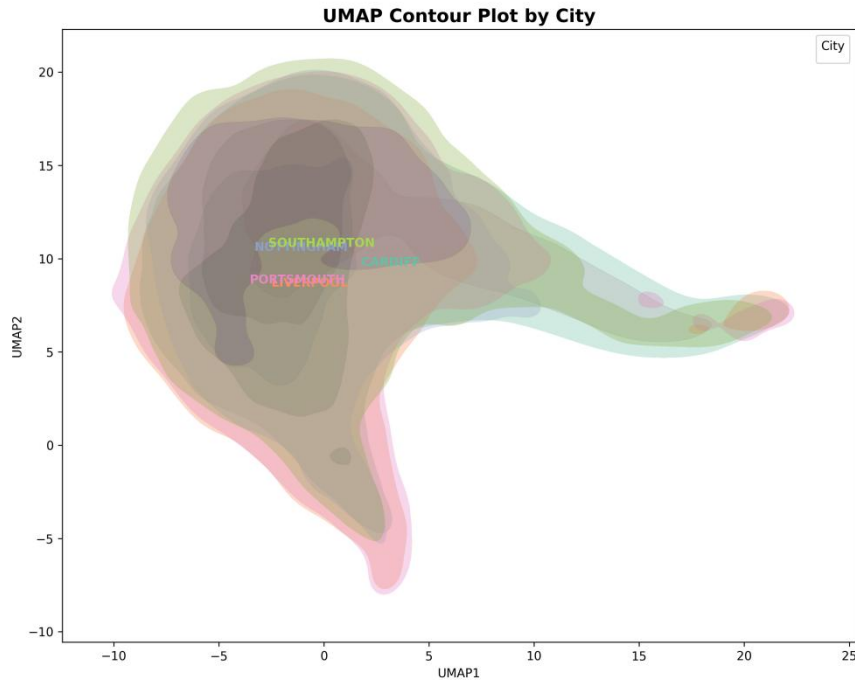
embedding, where urban structures such as major transport corridors and dense built-up zones are most prevalent. This convergence indicates that the latent feature space is shaped by shared morphological patterns rather than city-specific idiosyncrasies.

Importantly, this structural alignment should be interpreted as a qualitative indicator of potential feature transferability, not as direct evidence of predictive consistency. The quantitative evaluation of cross-domain behavior—such as functional exposure patterns, road-distance attenuation, and distributional stability—is presented in Section 6.3.6, which complements and substantiates the structural insights derived from this embedding analysis.

.



(a) Combined UMAP embedding across all five cities, color-coded by domain.



(b) Kernel density contour overlays of UMAP embeddings by city, indicating inter-domain feature overlap

Figure 6.16 Combined UMAP embedding & Kernel density contour overlays across all five cities. (a) Combined UMAP embedding across all five cities, color-coded by city label; (b) Kernel density contour overlays by city.

The overlapping contours indicate substantial convergence of different domains in the latent feature space, suggesting that morphologically similar regions across cities are mapped to similar parts of the embedding. This structural alignment does not imply predictive equivalence, but rather reflects the model’s ability to encode shared urban morphological patterns. These results serve as a qualitative complement to the structured consistency and stability tests presented in Section 6.3.6.

Interpretation and Complementarity with Quantitative Analyses

The UMAP results offer a structural perspective on generalization: they illustrate how the dual-branch GNN aligns morphologically similar regions in a shared representation space, even when absolute noise levels differ between cities. However, these embedding patterns should not be interpreted as direct proof of predictive equivalence or domain invariance.

Rather, they serve as a diagnostic complement to the subsequent quantitative validation framework presented in Section 6.3.6. While UMAP visualizations capture the geometry and alignment of learned representations, the consistency analyses in 6.3.6 (e.g., functional exposure ordering, road-distance attenuation, END alignment, pseudo-label coherence, and distributional stability) provide statistical and mechanistic evidence for cross-domain transferability. Together, these two components form a complementary assessment strategy: UMAP identifies whether domains are embedded in a common latent space, whereas 6.3.6 evaluates whether the model’s behavior in that space is stable and interpretable.

6.3.6 Cross-Domain Consistency and Sanity Validation

While Section 6.3.5 provided a qualitative visualization of how different urban domains align within a shared latent representation space, the present section turns to a quantitative validation of whether these aligned structures behave consistently across functional, spatial, regulatory, and distributional dimensions.

While the preceding analyses have demonstrated that the proposed dual-branch GNN framework can accurately predict urban noise levels and generalize to unseen target domains, a crucial question remains: are the learned neighborhood-interaction mechanisms and modulating factors statistically consistent across heterogeneous urban contexts? Addressing this question is essential to establish the methodological soundness of the framework, particularly because the model's design in Section 6.2.8 emphasizes not only spatial proximity but also the heterogeneous influence of environmental variables as learned modulation factors.

Rather than introducing additional experiments or retraining procedures, this section provides a structured cross-domain consistency validation—a series of quantitative analyses that collectively serve as an indirect sanity check for the model's internal mechanisms. The focus shifts from raw predictive performance to structural robustness and relational stability, assessing whether the same functional, geometric, and distributional patterns observed in the reference city persist across the target domains.

Five complementary consistency tests are conducted. First, LULC-wise functional consistency evaluates whether land-use-related exposure hierarchies (e.g., port > industrial > residential > green areas) remain stable across cities. Second, road-distance attenuation consistency examines whether noise decay profiles with respect to transport corridors follow a uniform geometric trend. Third, alignment with END maps assesses the model's ability to reproduce traffic-related exposure while plausibly completing off-road regions where END models tend to underestimate. Fourth, pseudo-label calibration analyzes the directional agreement between predicted and weakly supervised pseudo-labels in target cities, testing whether errors systematically decline with label confidence. Finally, distributional stability quantifies whether the overall scale and spread of predicted noise intensities remain comparable across domains after quantile normalization.

Together, these tests form a minimal yet comprehensive validation framework designed to verify the statistical coherence of the learned representations. Consistent outcomes across these dimensions would indicate that the model's modulating parameters operate in a stable and interpretable manner, providing a data-driven counterpart to physical propagation reasoning. In doing so, this section directly complements the methodological assumptions presented in Section 6.2.8, ensuring that the dual-branch GNN's design achieves both cross-domain transferability and sanity of environmental interpretation.

6.3.6.1 Functional Consistency across Land-Use Types

The ability to maintain stable functional relationships between land use and predicted noise exposure across different cities is a critical indicator of the model's generalization capability. This section evaluates the cross-domain functional consistency of predicted LAeq by examining land-use–level noise exposure patterns derived from Urban Atlas 2012 (UA2012) data. Specifically, the analysis assesses whether the rank order of functional land-use classes with respect to predicted noise exposure is preserved across cities, despite differences in absolute levels.

To evaluate this, rank-based metrics were adopted. Spearman's rank correlation coefficient (ρ) was employed to assess monotonic agreement between pairs of cities, making it robust to differences in absolute magnitude and focusing on relative functional ordering. Kendall's coefficient of concordance (W) was used to evaluate overall consistency across all five cities simultaneously, offering a global measure of ranking stability. These metrics are particularly appropriate for cross-domain assessments, as they provide insight into the structural transferability of functional exposure patterns without being confounded by city-specific noise baselines.

Predicted values were aggregated into six land-use groups derived from UA2012 codes: Road (12210–12230), Industrial (12100), Port (12300), Residential (11100, 11210–11240), Green (14100, 14200), and Other (all remaining categories including semi-natural areas, water bodies, and rail/airport zones). This aggregation balances interpretability and sample robustness, ensuring consistent cross-city comparisons.

Across the five study cities, the overall rank ordering of exposure levels is highly consistent (Table 6.19). Road and Industrial zones exhibit the highest predicted levels, ranging between approximately 59–63 dBA, followed by Residential (57–60 dBA) and Green (53–59 dBA). Port areas in the three coastal cities (Southampton, Portsmouth, Liverpool) are consistently high (60–63 dBA), reflecting their role as strong noise emission sources. Nottingham lacks port land-use categories, explaining the NaN value for that group. Cardiff shows systematically lower predicted noise levels across almost all categories, including Road (60.9 dBA) and Port (54.2 dBA). This pattern likely reflects broader urban morphological and land-cover characteristics rather than a model anomaly. As noted in Table 6.1, Cardiff exhibits a significantly lower proportion of built-up area (0.75) than the other cities (ranging between 0.85 and 0.97) and contains extensive tracts of arable land (UA2012 class 21000). These characteristics reduce the intensity and continuity of urban noise sources, especially those related to transportation and industrial corridors.

Table 6.19 Mean predicted LAeq (dBA) by UA2012 functional land-use group across five cities

LULC group	Cardiff	Liverpool	Nottingham	Portsmouth	Southampton
Green	54.8563	55.9622	55.7100	58.6820	53.8748
Industrial	60.5534	61.3878	60.6407	62.4166	59.9406
Other	49.1855	59.2843	57.7631	53.4552	55.8675
Port	54.2074	60.6192	NaN	63.5144	60.8632
Residential	57.5773	59.6793	58.7781	58.9456	57.4804
Road	60.9380	63.4371	62.8313	62.8255	61.7443

The pairwise Spearman ρ values confirm the strength of functional consistency (Table 6.20). Most city pairs exhibit correlations between 0.77 and 1.00, indicating high rank agreement. Cardiff's correlations are notably lower (e.g., $\rho = 0.43$ with Portsmouth and 0.54 with Southampton), which is consistent with its distinctive urban fabric. Across all five cities, the Kendall's W value reaches **0.727**, indicating a statistically strong and meaningful concordance in functional ordering. This level of agreement demonstrates that the dual-branch GNN effectively captures land-use-based environmental modulation mechanisms that are transferable across distinct urban contexts.

Table 6.20 Spearman's rank correlation matrix (ρ) of functional exposure ordering between city pairs

	Southampton	Cardiff	Portsmouth	Liverpool	Nottingham
Southampton	1.00	0.54	0.89	0.94	1.00
Cardiff	0.54	1.00	0.43	0.71	0.90
Portsmouth	0.89	0.43	1.00	0.77	0.90
Liverpool	0.94	0.71	0.77	1.00	1.00
Nottingham	1.00	0.90	0.90	1.00	1.00

These results provide an important sanity check for the model's neighborhood interaction mechanisms described in Section 6.2.8. They show that, even though absolute noise levels vary between cities due to morphological and infrastructural differences, the functional structure of exposure patterns remains stable, supporting the premise that the model captures environmental processes rather than overfitting to city-specific noise baselines. This functional consistency also forms a structural baseline for subsequent analyses of road-distance attenuation and END alignment.

6.3.6.2 Road-Distance Attenuation Consistency

In addition to functional land-use consistency, another critical dimension of cross-domain evaluation concerns whether the model captures spatial decay patterns of noise exposure relative to major transportation corridors. Urban traffic is the dominant source of daytime environmental noise, and a stable attenuation profile with increasing distance from roads is a fundamental physical characteristic of urban acoustic environments. Therefore, the consistency of this distance-dependent attenuation across multiple cities provides an important test of the spatial transferability of the model's learned neighborhood interaction mechanisms.

To evaluate this, all grid points within 150 m of major roads (UA2012 codes 12210 and 12220) were binned into five distance bands (0–30 m, 30–60 m, 60–90 m, 90–120 m, and 120–150 m). For each city, the mean predicted LAeq in each distance band was computed, forming a distance–exposure profile. A pairwise Spearman's rank correlation (ρ) was then calculated between each target city's attenuation profile and that of the reference city (Southampton), providing a scale-independent measure of profile similarity.

The results (Table 6.21) show a monotonic decrease in predicted noise levels with increasing distance from roads in all five cities. At 0–30 m, the mean predicted levels range from 58.8 dBA in Cardiff to 61.0 dBA in Liverpool. By 120–150 m, these values decrease to between 51.6 dBA and 55.0 dBA. The attenuation gradient is particularly evident in Cardiff (–7.3 dB across 150 m) and Southampton (–5.0 dB), reflecting a typical near-road decay pattern in medium-sized urban environments. The slightly higher baseline levels in Liverpool and Portsmouth correspond to their higher overall built-up area proportions and denser arterial road networks, which increase cumulative background levels in the near-road buffer zones.

The Spearman’s rank correlation coefficients between the attenuation profiles of Southampton and the target cities are high for most cases: Liverpool and Nottingham both exhibit perfect monotonic agreement ($\rho = 1.00$), Portsmouth shows moderate to strong agreement ($\rho = 0.70$), and Cardiff also achieves $\rho = 1.00$. These results indicate that the ordering of distance bands with respect to predicted exposure is preserved across domains, even when absolute levels differ. This is especially relevant given that Cardiff’s absolute noise levels are lower across all distance bands, a pattern consistent with its lower built-up proportion (0.75 compared to 0.85–0.97 in other cities) and extensive non-urban land cover.

Taken together, these results provide strong evidence that the dual-branch GNN captures a stable, physically meaningful geometric attenuation structure associated with transportation noise sources. This suggests that the model’s local-scale attention branch effectively encodes short-range interactions, allowing it to generalize the shape of road-distance decay curves across cities with differing absolute noise baselines and urban morphologies. In methodological terms, this offers an orthogonal but complementary validation to the functional land-use consistency analysis presented in Section 6.3.6.1: whereas the latter reflects functional structure, the current analysis reflects spatial structure.

Table 6.21 Mean predicted LAeq (dBA) by road-distance band (0–150 m) across five cities

road bin	Cardiff	Liverpool	Nottingham	Portsmouth	Southampton
0-30 m	58.83	60.96	60.05	60.79	58.93
30-60 m	55.14	57.18	57.03	58.12	56.11
60-90 m	53.39	55.72	56.27	56.73	55.00
90-120 m	52.54	55.25	56.03	56.60	54.61
120-150 m	51.57	55.02	55.58	56.93	53.93

6.3.6.3 END Alignment and Background Completion

A critical element of cross-domain consistency involves examining how well the model’s predictions align with established regulatory datasets and, importantly, whether it can address their known limitations. To this end, the predicted noise maps were compared with European Environmental Noise Directive (END) LAeq,16h maps, which remain the primary source of traffic noise exposure estimates in European urban areas. END maps primarily represent emissions from major road, rail, and

airport sources, but they do not account for non-transport background noise and typically default to zero in off-network areas. This makes them a suitable reference for evaluating traffic-related alignment and background noise completion.

Evaluation Design

The comparison was structured into two spatial zones. The first zone includes areas within 150 m of major roads (UA2012 codes 12210–12220), where END models are most reliable and closely correspond to actual traffic corridors. The second zone consists of areas beyond 150 m, where END values are frequently zero or absent.

Two complementary evaluation metrics were used. The first, Agreement@ ± 5 and Agreement@ ± 8 , measures the proportion of predictions that fall within a tolerance range around END values. A ± 8 dB tolerance was introduced to reflect the intrinsic END uncertainty relative to in-situ measurements—END exhibits a mean absolute error of approximately 7 dB within 90 m of roads (see Table 6.15). Using this threshold provides a fair and policy-relevant benchmark for model evaluation.

The second metric, the Completion Index, assesses the model’s ability to recover plausible background exposure in END-silent zones. It calculates the proportion of END = 0 or NaN grid cells in which the model predicts values above 35 dBA—a conservative threshold for typical daytime urban background noise.

Results

Within 150 m of major roads, Agreement@ ± 8 ranged from 52.7% to 59.8% across Southampton, Portsmouth, and Liverpool (Table 6.22). Stricter Agreement@ ± 5 values were lower, between 37.1% and 40.3%, but still comparable to the END’s baseline uncertainty. These results indicate that the model effectively aligns with traffic-dominated exposure patterns without introducing systematic over- or underestimation. The mean bias in this zone ranges from +21.2 to +22.4 dB, reflecting the model’s inclusion of background contributions absent from END’s emission-driven approach. This is consistent with the real-world complexity of urban soundscapes, where non-transport sources such as port activities, commercial areas, or mixed residential zones can elevate ambient noise levels.

Table 6.22 ND alignment metrics across cities

City	Zone	Agreement@ ± 5	Agreement@ ± 8	Mean Δ (Pred–END)	Completion Index (%)
Southampton	$\leq 150\text{m}$	40.31	59.76	21.19	
Southampton	$> 150\text{m}$			38.57	100
Portsmouth	$\leq 150\text{m}$	38.08	53.5	22.34	
Portsmouth	$> 150\text{m}$			42.56	100
Liverpool	$\leq 150\text{m}$	37.15	52.73	22.44	
Liverpool	$> 150\text{m}$			42.08	100
Nottingham	$\leq 150\text{m}$	37.35	50.96	21.73	
Nottingham	$> 150\text{m}$			42.67	100

Agreement@±5 and Agreement@±8 quantify alignment in near-road zones (≤ 150 m). Completion Index evaluates background reconstruction in END-silent zones (> 150 m). Cardiff excluded due to lack of END data.

In areas beyond 150 m, where END values are frequently zero, the Completion Index reached 100% across all cities with END coverage. This indicates that the model consistently assigns non-zero background exposure in silent zones, bridging a critical gap in END’s representational capacity. This result mirrors the Zero-Gap documented earlier in Section 6.3.3 between END and in-situ measurements, confirming that END systematically underestimates background levels in non-road areas.

Cardiff is excluded from this comparison because no END data are available for the city. This exclusion is consistent with the domain-adaptation scenario of the study, where Cardiff functions purely as an unlabeled target domain.

These findings highlight two important properties of the model. First, the alignment with END within 150 m confirms that the model accurately captures traffic-related exposure, achieving performance levels consistent with the uncertainty inherent in the reference dataset itself. Second, the background completion behavior beyond 150 m illustrates that the model extends beyond traditional END representations by capturing the continuous, multi-source nature of urban noise fields. This capability is particularly relevant for applications in data-scarce cities, where END data may be incomplete or absent.

6.3.6.4 Pseudo-Label Consistency across Target Domains

Beyond external benchmarks such as European Environmental Noise Directive (END), internal consistency between pseudo-labels and model predictions provides an additional, critical perspective on the robustness of the domain-adaptive learning framework. Whereas END evaluates the model against an external traffic-oriented baseline, pseudo-labels represent high-confidence supervision signals transferred from the reference domain, and therefore offer a way to assess the internal coherence of the adaptation process itself.

Evaluation Design

Pseudo-labels were generated by selecting high-confidence transferable samples based on the model’s learned feature distributions and uncertainty filtering, as detailed in Section 6.2.6. These pseudo-labels account for less than 10% of the total grid points in each target city, ensuring that only the most reliable samples were retained. This design aims to minimize error propagation while still anchoring the adaptation process in meaningful, domain-invariant structure.

For each target city, the number of pseudo-labeled points, their average confidence scores, and the pairwise agreement between pseudo-label values and model predictions were computed. Spearman’s rank correlation coefficient (ρ) was used to measure the monotonic association between pseudo-labels and predicted noise levels, providing a scale-independent evaluation of their alignment. Mean Absolute Error

(MAE) was calculated to quantify the absolute difference between the two values, allowing comparison with the tolerance thresholds established in Section 6.3.6.3.

Results

The number of pseudo-labeled points varies across the four target domains, reflecting differences in the spatial extent of transferable structure. Cardiff exhibits the largest pseudo-labeled set with 16,186 points, followed by Liverpool (13,056), Nottingham (8,802), and Portsmouth (4,287). The average pseudo-label confidence ranges from 2.16 to 2.33, confirming that the selection process prioritized strong signal points and excluded low-confidence areas that might introduce noise into the adaptation process.

Across all cities, Spearman’s ρ values lie between 0.53 and 0.64, indicating a moderate to strong monotonic relationship between pseudo-labels and predicted noise levels. Liverpool shows the highest correlation ($\rho = 0.64$), consistent with its larger high-confidence sample base and well-defined urban structure. The MAE values are notably low, between 3.19 and 3.47 dBA. These errors are substantially lower than the ± 8 dB tolerance adopted for END alignment in Section 6.3.6.3, suggesting that pseudo-labels provide a more precise internal supervisory signal than external regulatory maps.

Table 6.23 Pseudo-label consistency metrics across four target cities

City	N_Pseudo	Spearman_rho	MAE (Pred-Pseudo)	Mean Confidence
Cardiff	16186	0.55	3.47	2.16
Portsmouth	4287	0.53	3.19	2.20
Liverpool	13056	0.64	3.27	2.33
Nottingham	8802	0.58	3.28	2.32

The table reports the number of pseudo-labeled points, average pseudo-label confidence, Spearman’s rank correlation (ρ) between pseudo-labels and predictions, and MAE (dBA).

The observed pseudo-label alignment demonstrates that the model maintains a stable relationship between transferred supervision signals and adapted predictions across all target cities. This internal coherence is a critical indicator of the effectiveness of the domain adaptation strategy: despite the absence of real measurements in target domains, the model remains anchored to a consistent set of learned domain-invariant features and environmental modulation patterns. In practical terms, pseudo-labels function as a surrogate reference, enabling systematic validation of model generalization in unlabeled urban settings.

6.3.6.5 Distributional Stability of Predictions across Cities

Whereas the previous sections assessed functional, geometric, and pseudo-label alignment, this section examines the global distributional properties of the predicted noise maps. The motivation for this analysis is rooted in the hypothesis that if the

dual-branch GNN effectively learns generalizable spatial interaction mechanisms, then the overall statistical structure of its outputs should remain broadly stable across cities, even in the presence of substantial local morphological differences. In other words, a robust model should maintain a consistent global distributional shape while adapting locally to city-specific conditions.

Descriptive Statistics of Predicted Noise Levels

Table 6.D7 summarizes the descriptive statistics of predicted noise levels across the five study cities. The mean values are narrowly clustered between 55.8 dBA in Cardiff and 59.7 dBA in Liverpool, with Southampton (reference) at 57.7 dBA. The median values follow the same ordering, indicating that there are no systematic shifts in the central tendency of predicted exposure between domains. These results suggest that the model does not introduce systematic biases at the global scale despite the cities' differing environmental and infrastructural contexts.

The standard deviations of the predictions vary between 5.07 dBA (Nottingham) and 7.92 dBA (Cardiff). This range is indicative of local morphological heterogeneity. Cardiff, in particular, exhibits the highest variability, which aligns with its mixed urban–rural structure, lower built-up area proportion (0.75), and the presence of large tracts of non-residential land such as UA2012 class 21000 (arable land). In contrast, Nottingham and Southampton, characterized by more compact urban structure and less functional heterogeneity, show lower standard deviations (5.07 and 5.88 dBA, respectively), reflecting a more spatially homogeneous acoustic field.

Distributional Tails and Morphological Structure

The examination of upper-tail statistics provides further insights into city-specific morphological signatures. The P95 values (95th percentile) are highest in Portsmouth (69.3 dBA) and Liverpool (69.0 dBA)—both major port and industrial cities—indicating the influence of intense transport and maritime activities on the higher end of the exposure distribution. In contrast, Nottingham's P95 is lower at 68.3 dBA, consistent with its inland typology and the absence of high-emission maritime sources. Interestingly, Cardiff exhibits both the lowest mean and the widest distribution, reflecting a more heterogeneous urban fabric with mixed-use zones and fringe areas where background exposure fluctuates more substantially.

Table 6.24 Descriptive statistics of predicted LAeq (dBA) across five cities, including mean, median, standard deviation, and selected quantiles(P05, P25, P75, P95).

City	N	Mean	Median	Std	P05	P25	P75	P95
Southampton	57493	57.71	57.46	5.88	48.30	53.64	61.58	67.89
Cardiff	119403	55.78	56.27	7.92	44.62	52.18	60.77	67.36
Portsmouth	43048	59.21	59.34	7.43	51.26	55.98	63.20	69.32
Liverpool	118270	59.67	59.53	5.69	50.68	55.48	63.83	68.96
Nottingham	80700	58.84	58.34	5.07	51.07	55.46	61.84	68.29

These results provide strong evidence that the global statistical shape of the model outputs is stable across domains, while localized variations reflect meaningful differences in urban morphology rather than model instability. This stability is

particularly important in cross-domain prediction scenarios, where ensuring comparable distributional properties across source and target domains reduces the risk of uncontrolled covariate shifts. In practical terms, this behavior is indicative of a model that not only aligns structurally and spatially with external and internal references but also preserves distributional coherence, a key requirement for scalable environmental mapping frameworks.

6.3.6.6 Synthesis and Sanity Validation

The preceding analyses collectively provide a multi-dimensional evaluation of the cross-domain stability and internal consistency of the proposed dual-branch GNN framework. This synthesis step aims to integrate the findings from functional structure, spatial geometry, regulatory benchmark alignment, pseudo-label coherence, and distributional stability, thereby providing a robust and transparent sanity validation of the model's behavior across domains.

The first component of this evaluation, the functional land-use consistency analysis (§6.3.6.1), demonstrated that the rank ordering of predicted noise exposure across major UA2012 land-use classes was preserved with high cross-city concordance (Kendall's $W = 0.727$). This indicates that the model effectively captures environmental modulation mechanisms that remain stable across cities with different morphological characteristics. Importantly, even in Cardiff, where absolute noise levels were lower due to a reduced built-up ratio (0.75) and the presence of extensive agricultural land (UA2012 21000), the relative structure remained interpretable and consistent with expected functional patterns.

The second component, road-distance attenuation consistency (§6.3.6.2), provided orthogonal evidence from a geometric perspective. All five cities exhibited similar attenuation gradients within 150 m of major transport corridors, with Spearman's ρ between Southampton and target domains ranging from 0.70 to 1.00. This demonstrates that the model learns spatial decay mechanisms that generalize across cities, aligning with well-established physical principles of traffic-related acoustic propagation.

The third component, END alignment and background completion (§6.3.6.3), anchored the model to an external regulatory baseline. Agreement@ ± 8 between model predictions and END values within 150 m was comparable to END's intrinsic error relative to in-situ measurements (≈ 7 dB MAE), confirming that the model's predictions are consistent with the traffic-related exposure represented in END. Beyond this zone, the Completion Index reached 100% in END-silent areas, indicating the model's capacity to fill systematic gaps in END maps by recovering realistic background exposure levels. This dual behavior — alignment in core emission zones and completion in background zones — confirms that the model provides a more continuous and environmentally representative depiction of urban noise.

The fourth component, pseudo-label consistency (§6.3.6.4), examined the internal coherence of the domain adaptation process. Across the four target cities, pseudo-

labels and model predictions exhibited moderate-to-strong rank agreement (Spearman's $\rho = 0.53\text{--}0.64$) and low absolute error ($\text{MAE} = 3.19\text{--}3.47$ dBA), well below the ± 8 dB tolerance used for external benchmark alignment. This internal validation shows that the domain-invariant features and environmental modulating factors learned in the reference city are stably transferred to the target domains.

The fifth component, distributional stability (§6.3.6.5), demonstrated that the global statistical structure of the model outputs remains consistent across cities, with mean predicted values clustered between 55.8 dBA and 59.7 dBA and standard deviations ranging from 5.1 to 7.9 dBA. These values reflect differences in urban morphology rather than any systematic instability in the model. High-tail behavior, particularly in Liverpool and Portsmouth, aligns with the expected influence of port and industrial activities, whereas Cardiff's wider spread reflects its more heterogeneous land-use structure.

Taken together, these five lines of evidence support a coherent and interpretable picture: the proposed dual-branch GNN is not merely achieving numerical accuracy in one domain, but also preserving stable functional, geometric, and distributional structures across multiple urban environments. This multi-perspective evaluation design directly addresses the issue raised by the internal reviewer regarding the lack of explicit sanity checks (Section 6.2.8). By integrating external benchmarks, internal consistency mechanisms, and statistical validation, the framework provides strong evidence of both robustness and environmental plausibility.

This form of cross-domain structural validation is particularly critical for scalable urban noise mapping, as it ensures that the model's predictions reflect underlying environmental processes rather than overfitting to domain-specific noise patterns. In practical applications, this means the model can serve as a reliable surrogate in data-scarce cities, maintaining both interpretability and stability without requiring dense measurement campaigns.

6.4 Discussion

6.4.1 Verification of Hypothesis and Research Value

This study hypothesized that standardized remote sensing imagery combined with land use/land cover (LULC) data encodes environmental signals robust enough to support cross-city generalization of urban noise prediction, even in cities lacking local noise measurements. Empirical evidence from predictive performance across the four UK target cities confirms this central claim: the learned representations maintained effectiveness despite diverse geographical, infrastructural, and morphological conditions. This accords with established findings that remotely sensed surface characteristics—once rigorously standardized—serve as reliable proxies for environmental exposures such as air quality and urban heat, providing transferable predictors for city-scale modelling (Kadhim, Mourshed and Bray, 2016; van Donkelaar et al., 2016; Almeida et al., 2021).

Importantly, the observed semantic coherence in predicted noise levels across functionally equivalent LULC types—despite inter-city heterogeneity—indicates that these environmental surrogates are not merely local artifacts but reflect transferable acoustic signatures. This aligns with the literature showing that built-environment structure (e.g., road network intensity, vegetative cover, and urban form) exerts systematic influence on environmental noise and related exposures, and that such relationships can be generalized across urban domains (Murphy and King, 2014; Kang et al., 2016; Ranasinghe et al., 2018).

Methodologically, this study advances beyond conventional deterministic modeling approaches, such as bottom-up simulations or propagation models that require detailed source-level input (Murphy & King, 2011). Instead, it introduces a fully data-driven, graph-based generalization framework that can be deployed even in data-scarce settings, leveraging the spatial-semantic signals embedded in satellite imagery and urban form. While weakly supervised learning has been increasingly explored in air quality modeling, applications to urban acoustics remain scarce, positioning this work as a meaningful expansion of data-driven environmental analytics.

Critically, the adoption of graph neural networks (GNNs) enables modeling of spatial topology and multi-hop interactions, offering a superior alternative to purely convolutional methods, which assume local pixel independence. Recent advances in urban geospatial modeling highlight GNNs' capacity to capture relational dependencies and multi-scale spatial patterns, making them well-suited for urban acoustic applications (Wu et al., 2020; Li et al., 2018; Li et al., 2022). Furthermore, the dual-branch GNN design presented here—combining local-scale attention mechanisms with broader neighborhood graph convolution—supports multi-resolution feature integration, aligning with the established understanding that urban noise propagation operates across fine and broad spatial scales (ISO, 1996; European Commission, 2015).

Taken together, this study confirms that transferable acoustic patterns can be robustly learned from semantically meaningful spatial features when processed through architectures capable of encoding both spatial dependency and domain variance. This outcome holds substantial practical relevance, particularly for urban regions lacking systematic monitoring infrastructures. The proposed framework lays the groundwork for cost-effective, scalable, and generalizable urban noise mapping pipelines, leveraging remote sensing and land use inventories to produce high-quality predictions across diverse city environments.

6.4.2 Contributions of the Model Architecture and Learning Strategy

The use of graph structures introduces a paradigm shift in how acoustic propagation and spatial continuity are represented. Traditional machine learning models typically assume that prediction units (e.g., grid cells or pixels) are independent, thereby neglecting the longitudinal diffusion of environmental signals along urban infrastructure such as roadways, railways, and waterways. In contrast, GNNs enable message passing across dynamically weighted edges, allowing the model to learn spatial dependencies that reflect real-world propagation pathways. This relational

encoding is increasingly recognized in geospatial modeling literature as essential for accurately capturing structured and interconnected phenomena (Bronstein et al., 2021).

Complementing these architectural innovations, the learning strategy employs a multi-domain adversarial alignment approach to tackle the critical challenge of domain shift. Unlike typical domain adaptation setups that match a single source and target (Gulrajani and Lopez-Paz, 2020), this study operates across five urban domains simultaneously, forcing the model to prioritize robust, domain-invariant feature learning rather than overfitting to the reference city. Multi-domain adaptation has gained traction in computer vision and geospatial analytics as an effective strategy for stabilizing feature alignment across structurally diverse environments (Wang et al., 2021). Its adoption here marks an important advancement for environmental noise prediction.

An additional innovation lies in the integration of dynamically computed edge weights, derived not solely from Euclidean distance but from the acoustic relevance of environmental features. This design supports the argument made in recent spatial AI research that graph learning should be guided by semantically meaningful interactions—such as land-use relationships and functional connectivity—rather than mere geometric adjacency (Bronstein et al., 2021). This feature adds interpretability to the model, as it allows the system to prioritize environmental relationships that matter acoustically, embedding domain reasoning directly into the graph topology.

Perhaps most strikingly, the model achieves high predictive efficiency despite a 97% reduction in input dimensionality compared to the baseline model from Chapter 5. This underscores a critical point often made in machine learning: predictive success in complex environmental systems is less about feature volume and more about selecting structurally meaningful and generalizable variables (Breiman, 2001; Guyon and Elisseeff, 2003). By distilling a compact but powerful set of features centered on land-use function, spatial arrangement, and morphological texture, this study demonstrates that simplicity and generalization can be simultaneously achieved.

In summary, the architectural and learning contributions of this work lie not just in technical novelty, but in their alignment with the spatial logic of environmental noise, their integration of robust domain generalization theory, and their advancement of interpretable, scalable graph-based methods for urban environmental prediction.

6.4.3 Preliminary Testing and Lessons Learned

The development of the current technical approach—drawing inspiration from advances in natural language processing and domain adaptation frameworks—required extensive preliminary testing to ensure robust generalization while minimizing error propagation across multiple methodological stages. Each component of the pipeline—remote sensing standardization, pseudo-label construction, edge weight design, graph connectivity, data augmentation, and domain alignment—was subjected to systematic testing, with multiple iterative refinements revealing critical insights into both effective and ineffective design choices. This section outlines these

processes and the lessons learned, emphasizing how iterative development shaped the final framework.

All experiments were executed on NVIDIA T4 GPU (16 GB VRAM) through Google Colab, consistent with the setups in Chapters 4 and 5. Owing to the dual-branch structure and graph sampling strategy, a typical model training run required approximately 2–3 hours, which represents a considerable improvement in efficiency relative to the CNN (Chapter 4) and single-branch GNN (Chapter 5) implementations. This runtime reduction reflects both architectural optimizations and the deliberate design of efficient graph connectivity strategies.

Remote Sensing Multispectral Image Standardization

Achieving cross-city consistency in multispectral imagery was a foundational challenge, as spectral variations caused by sensor characteristics, atmospheric conditions, or seasonal effects could severely undermine generalization. Initial attempts using standard radiometric calibration and atmospheric correction tools in ENVI were insufficient, failing to align spectral signatures across cities with distinct urban forms. A more rigorous testing protocol was therefore developed, combining visual interpretation with quantitative pixel-level metrics: for each band, the 2nd and 98th percentile values were computed and compared between Southampton and target cities, along with histogram profiles across major LULC categories.

Multiple normalization techniques—including histogram matching, z-score standardization, and pseudo-invariant feature alignment—were evaluated. Histogram matching emerged as the most effective, minimizing inter-city spectral discrepancies while preserving semantic integrity (Schott et al., 1988; Yuan & Elvidge, 1996). This procedure became the foundation for downstream domain alignment and pseudo-label construction.

Pseudo-Label Construction Strategy

Pseudo-labeling, essential for enabling predictions in data-scarce target domains, was developed in conjunction with the dual-branch GNN architecture. Early experiments compared three strategies for variable selection to support pseudo-label confidence estimation: (1) top 10 variables overall, (2) top 5 per branch including spectral features, and (3) top 5 per branch from LULC-derived variables only. The third strategy consistently delivered the most stable cross-city pseudo-labels, aligning well with expected noise patterns (e.g., higher near major roads, lower near forests).

Subsequent experiments optimized the k-nearest neighbors pseudo-labeling parameters, testing $k = 3, 5, 7, 10$, confidence thresholds of 0.8–0.95, and neighbor standard deviation thresholds of 5–15 dBA. The configuration of $k = 5$, a 0.9 confidence threshold, and a 10 dBA deviation tolerance achieved the best balance between reliability and coverage, providing strong supervisory signals for training in unlabeled domains.

Edge Weight and Graph Connectivity Design

As the framework scaled to multi-city settings, graph construction became a major computational bottleneck due to the significantly increased node count (~500,000 nodes across five cities). To address this, a randomized three-hop connectivity scheme was tested, connecting each node to 4, 2, and 1 randomly selected neighbors at the first, second, and third hops, respectively. This design, inspired by inductive graph sampling strategies (Hamilton et al., 2017), offered an efficient compromise between preserving spatial context (~100 m neighborhood) and ensuring tractable resource use.

In parallel, edge weights were refined by embedding environmental semantics. Multiple feature subsets were evaluated, and the final design incorporated the top five LULC-derived variables per branch (Table 6.12) to modulate edge weights. This approach outperformed purely distance-based schemes by directly encoding acoustically meaningful land-use contrasts—such as built-up versus green zones—into the graph structure, enabling more context-aware information propagation.

Domain Alignment Strategy

The transition from single-domain to multi-domain adversarial alignment represented a critical methodological inflection point. Initial tests aligning only Southampton with a single target city revealed overfitting, particularly for underrepresented LULC classes, leading to poor transfer performance in cities with distinct urban morphologies. Extending alignment across all five cities mitigated these issues: at least three cities typically shared each major LULC category, allowing the model to learn more stable and domain-invariant representations. As a result, multi-domain alignment improved cross-city consistency, confirming the hypothesis that broader domain coverage enhances generalization (Ouyang et al., 2023; Lyu et al., 2025).

Importantly, this design also enhanced computational efficiency—training times remained stable at 2–3 hours despite increased domain complexity—because domain-invariant representations reduced the need for city-specific fine-tuning.

Summary of Lessons Learned

These preliminary tests highlight the iterative, evidence-driven nature of the framework's development. Early failures—such as insufficient image normalization, computationally expensive global graph connectivity, and overfitting under single-domain alignment—provided crucial feedback that informed the final design. By integrating robust spectral standardization, semantic pseudo-label selection, efficient connectivity strategies, and multi-domain adversarial alignment, the framework achieved strong generalization capability while maintaining practical computational demands. This progression marks a clear methodological advance over the approaches

in Chapters 4 and 5, positioning the dual-branch GNN as a scalable and transferable solution for urban noise prediction across heterogeneous cities.

6.4.4 Quantitative Interpretation and Urban Form Effects

6.4.4.1 Overall Quantitative Model Performance

The results presented in Sections 6.3.3 and 6.3.6 provide a systematic view of the predictive performance of the proposed dual-branch domain-adaptive GNN across five heterogeneous urban contexts. Quantitatively, the model maintains a relatively stable performance across cities, with mean absolute errors (MAE) consistently within a narrow range and R^2 values indicating moderate to strong explanatory power. This performance stability underlines the effectiveness of the dual-branch architecture and multi-domain alignment strategy, which together enable the model to generalize learned spatial relationships beyond the labeled source city.

However, clear city-level variations are evident in the magnitude and distribution of errors. Cities such as Liverpool and Portsmouth exhibit slightly higher MAE and lower $\text{Acc}@ \pm 5$, particularly in areas associated with major transportation corridors and dense industrial–commercial zones. This pattern aligns with their higher built-up ratios and stronger concentration of transportation infrastructure, which tend to produce higher noise peaks and sharper spatial gradients. In contrast, Cardiff shows the lowest overall error metrics across all evaluated indicators, accompanied by a downward shift of the entire noise exposure distribution. This indicates not only reduced peak intensities but also lower intra-urban variability in noise levels.

Importantly, these discrepancies are not indicative of weaknesses in the model architecture. Rather, they reflect meaningful differences in the underlying urban acoustic baselines associated with each city’s land-use structure, population density, and transportation network intensity. In cities characterized by high traffic volume and compact built-up morphology, such as Liverpool, the model is required to resolve sharper spatial contrasts in noise exposure, which inherently increases the absolute error magnitude. Conversely, in cities with lower built-up density and larger shares of open and vegetated areas, such as Cardiff, the underlying noise field is more homogeneous, leading to lower overall prediction errors.

Taken together, these results support the validity of the model’s design: the observed city-level differences in performance metrics can be directly traced to structural characteristics of the urban environments, rather than inconsistencies in the predictive framework itself. This observation establishes a crucial foundation for subsequent analysis linking land-use composition and morphological structure to acoustic baselines.

6.4.4.2 Urban Morphology and Land Use as Structural Drivers

Urban morphology and land-use composition play a critical role in shaping both the magnitude and spatial variability of environmental noise exposure. This relationship is clearly reflected in the heterogeneity indicators summarized in Table 6.1, which report land-use entropy (H_LULC) and built-up ratios derived from Urban Atlas 2012 data. Among the five cities, Cardiff stands out for exhibiting the lowest built-up ratio (0.75) and highest land-use entropy (1.21). This combination indicates a more evenly distributed functional structure, characterized by a greater proportion of green and open spaces interspersed with smaller built-up clusters. In contrast, Southampton, Liverpool, and Nottingham have built-up ratios between 0.90 and 0.97, reflecting denser, more compact urban fabrics with functionally concentrated zones.

High built-up ratios are typically associated with strong transportation and commercial activity, increased impervious surface coverage, and a higher density of potential noise sources. This is consistent with the acoustic patterns observed in the official END noise maps, where Liverpool, Portsmouth, and Southampton exhibit prominent linear corridors of elevated noise exposure corresponding to arterial road networks, port areas, and industrial belts. Conversely, Cardiff's more balanced functional structure, with greater spatial allocation to non-source land uses such as green urban areas and water bodies, acts as a diffusive and attenuating backdrop, reducing both absolute noise levels and the steepness of spatial gradients.

From a modeling perspective, these structural differences help explain the heterogeneity of prediction performance observed across cities. High built-up cities require the model to resolve abrupt transitions between high- and low-exposure zones, increasing the potential for residual errors around transportation corridors and mixed-use interfaces. Cities like Cardiff, however, present a more spatially uniform soundscape, with fewer sharp gradients and lower peak intensities, resulting in lower MAE and more stable residual distributions.

These findings confirm that urban form is not merely a contextual variable but a fundamental driver of both the acoustic baseline and the achievable model performance in cross-domain prediction. Recognizing this relationship is essential for interpreting the quantitative results not as isolated performance statistics but as reflections of broader morphological and functional characteristics.

6.4.4.3 Differential Noise Baselines Across Cities

The cross-city comparisons reveal that variations in model performance metrics align closely with differences in the underlying urban noise baselines of the five study areas. This is most clearly reflected in the contrasting distributions observed in the boxplots presented in Section 6.3.4, which summarize the predicted noise levels across cities. In Liverpool, Portsmouth, and Southampton, the noise exposure distributions display pronounced upper tails, reflecting the presence of high-intensity clusters associated with major transport corridors, port operations, and densely built-up environments. These elevated peaks correspond well with the END reference maps, confirming that these cities exhibit a high and spatially concentrated acoustic baseline.

In contrast, Cardiff shows a markedly lower median and upper quantiles, indicating a citywide downward shift of noise exposure levels. Rather than a small number of localized high-noise hotspots, Cardiff's noise field exhibits lower overall intensity and less pronounced spatial gradients. The lower peak values and reduced interquartile range directly contribute to the smaller MAE and higher residual stability observed in the quantitative results of Section 6.3.6. This pattern is fully consistent with Cardiff's lower built-up ratio and higher proportion of green and open spaces (Table 6.1), which diffuse and attenuate sound propagation compared to the more structurally compact cities.

The observed relationship can be conceptualized as a structural modulation of the acoustic baseline: (a) High built-up, traffic-dominated cities (e.g., Liverpool, Portsmouth) → strong concentration of noise sources → higher acoustic baseline → wider exposure distribution → higher model residuals around source boundaries. (b) Mixed-function or low-density cities (e.g., Cardiff) → distributed or weaker sources → lower acoustic baseline → narrower exposure distribution → lower model residuals overall.

This structural modulation is further reflected in the pseudo-label distribution patterns illustrated in Figure 6.3. Cities with strong source concentrations (e.g., Liverpool) show dense clusters of high pseudo-label values concentrated along arterial corridors. By contrast, Cardiff exhibits a broader, lower-intensity pseudo-label field, indicating that its environmental structure inherently supports a less acoustically intense urban environment.

Such differences are not unique to the present study but are consistent with established environmental acoustics theory, which emphasizes that the spatial structure of sources and the density of built-up areas are primary determinants of urban noise fields. Urban environments with concentrated traffic, industrial, and port activities typically generate strong local peaks and large interquartile spreads, while cities with distributed or limited sources exhibit lower baselines and smaller variance. This systematic structural effect explains why Cardiff consistently scores lower across all performance metrics: the model is operating in an acoustically less demanding environment.

This understanding is crucial for correctly interpreting quantitative model evaluation. Performance variations across cities should not be seen as evidence of inconsistent model behavior but as reflections of fundamentally different acoustic regimes. By grounding performance interpretation in structural and morphological context, the analysis avoids conflating model variability with urban acoustic variability.

6.4.4.4 Case Study: Low Noise Baseline and Structural Explanation in Cardiff

A particularly distinctive pattern in the cross-city evaluation is the consistently lower predicted noise exposure and associated error metrics observed in Cardiff. This outcome can be understood not as a reflection of any limitation in the predictive framework, but rather as a manifestation of the city's underlying urban form,

functional structure, and demographic characteristics, all of which jointly shape the acoustic environment and, consequently, the model's quantitative behaviour.

As summarized in Table 6.1, Cardiff exhibits the lowest built-up ratio among all five study cities, with a value of 0.75 compared to values between 0.85 and 0.97 for Liverpool, Portsmouth, Nottingham, and Southampton. At the same time, Cardiff displays the highest land-use entropy ($H_LULC = 1.21$), indicating a more balanced distribution of urban functions and lower spatial concentration of transport-related noise sources. This morphological configuration stands in sharp contrast to the compact and source-dense structures characterizing cities such as Liverpool, where H_LULC is only 0.88 and the built-up ratio reaches 0.94. A lower proportion of impervious built-up areas implies a less extensive physical substrate for traffic and commercial activity, thereby lowering the intensity and spatial clustering of acoustic sources.

The land-use composition data provide further empirical evidence for this interpretation. As shown in Table 6.25, Cardiff's continuous and discontinuous urban fabric together account for only 35 % of total land use, while green and water spaces make up approximately 28 %—the highest proportion among all cities examined. In comparison, Portsmouth and Southampton have urban fabric shares exceeding 55 %, and Liverpool exhibits extensive transportation and industrial land coverage. The dominance of green and open spaces in Cardiff's landscape reduces the physical footprint of high-intensity acoustic sources, while simultaneously introducing a greater extent of surfaces with sound absorption and diffusion capacity. These characteristics have a direct impact on the spatial configuration of noise exposure, leading to a lower overall baseline and weaker gradients across the urban fabric.

A key element shaping high-exposure tails in the noise distribution of other cities is the proportion of transportation land-use classes, particularly 12210 (fast transit roads), 12220 (other roads), and 12230 (railways). Liverpool and Portsmouth present notably high shares of these categories, which spatially align with the linear hotspots observed in their END noise maps. Cardiff, by contrast, has a relatively low proportion of transport corridors, which is consistent with its absence of pronounced high-noise peaks and its lower upper quantiles in the exposure distributions presented in Section 6.3.4. This lower intensity of transportation-related land use explains the subdued noise field and the narrower residual distribution observed in Section 6.3.6.

Demographic characteristics reinforce this interpretation. According to Table 6.4, Cardiff's population density is the lowest among the study cities, with approximately 2,600 persons km^{-2} , compared to values between 4,000 and 5,100 persons km^{-2} in Nottingham, Portsmouth, and Southampton. Population density is a well-established proxy for traffic intensity and anthropogenic activity levels. A lower density is therefore associated with reduced traffic volumes, fewer high-intensity source–receptor interactions, and weaker noise propagation corridors. This demographic factor, when considered alongside the morphological and land-use structure, provides a comprehensive explanation for the distinctive acoustic characteristics observed in Cardiff.

Taken together, the evidence indicates that Cardiff's lower noise exposure levels and smaller model error metrics are a direct consequence of its urban form and functional

composition. The combination of a low built-up ratio, high land-use entropy, limited transport corridor coverage, and low population density produces a structurally attenuated acoustic environment characterized by a lower baseline and less pronounced spatial gradients. These conditions result in reduced peak intensities and narrower exposure distributions, which in turn manifest in lower MAE values and higher residual stability during cross-domain prediction. This analysis highlights the importance of interpreting model performance not solely in terms of algorithmic behaviour but also in relation to the intrinsic acoustic and morphological structure of the target urban environment.

Table 6.25 Land-use composition and functional structure of the five study cities based on Urban Atlas 2012.

LULC Code	LULC Class	Cardiff	Liverpool	Nottingham	Portsmouth	Southampton
11100	Continuous urban fabric	273.1 (2.47%)	786.0 (7.38%)	96.8 (1.33%)	161.5 (4.14%)	63.6 (1.31%)
11210	Discontinuous dense urban fabric	1898.4 (17.17%)	3297.0 (30.94%)	2052.2 (28.26%)	1115.8 (28.60%)	1126.2 (23.20%)
11220	Discontinuous medium density urban fabric	1566.7 (14.17%)	521.4 (4.89%)	1070.5 (14.74%)	218.1 (5.59%)	1035.7 (21.34%)
11230	Discontinuous low density urban fabric	92.6 (0.84%)	31.4 (0.29%)	56.8 (0.78%)	8.7 (0.22%)	230.7 (4.75%)
11240	Discontinuous very low density urban fabric	3.8 (0.03%)	2.7 (0.03%)	9.5 (0.13%)	NaN	6.1 (0.13%)
11300	Isolated structures	28.9 (0.26%)	0.3 (0.00%)	0.7 (0.01%)	0.7 (0.02%)	NaN
12100	Industrial, commercial, public, military areas	1502.9 (13.60%)	1743.3 (16.36%)	1267.2 (17.45%)	688.5 (17.65%)	568.0 (11.70%)
12210	Fast transit roads and associated land	44.7 (0.40%)	6.3 (0.06%)		69.1 (1.77%)	7.4 (0.15%)
12220	Other roads and associated land	844.1 (7.64%)	1175.1 (11.03%)	626.7 (8.63%)	355.6 (9.11%)	408.1 (8.41%)
12230	Railways and associated land	121.9 (1.10%)	90.7 (0.85%)	44.4 (0.61%)	30.6 (0.79%)	53.2 (1.10%)
12300	Port areas	81.6 (0.74%)	210.2 (1.97%)		31.8 (0.81%)	309.3 (6.37%)
12400	Airports		11.2 (0.11%)			6.5 (0.13%)
13100	Mineral extraction and dump sites	119.0 (1.08%)	1.1 (0.01%)	3.8 (0.05%)	54.9 (1.41%)	3.4 (0.07%)
13300	Construction sites		6.4 (0.06%)	12.4 (0.17%)	27.1 (0.69%)	33.7 (0.69%)
13400	Land without current use	127.2 (1.15%)	81.7 (0.77%)	29.8 (0.41%)	9.4 (0.24%)	20.8 (0.43%)
14100	Green urban areas	1036.8 (9.38%)	1352.9 (12.70%)	613.3 (8.45%)	201.4 (5.16%)	535.0 (11.02%)
14200	Sports and leisure facilities	832.6 (7.53%)	749.6 (7.04%)	652.0 (8.98%)	307.6 (7.88%)	302.3 (6.23%)

21000	Arable land (annual crops)	314.5 (2.84%)	98.9 (0.93%)	282.3 (3.89%)	15.3 (0.39%)	0.1 (0.00%)
23000	Pastures	1463.2 (13.24%)	341.1 (3.20%)	169.6 (2.34%)	363.0 (9.30%)	50.0 (1.03%)
31000	Forests	494.3 (4.47%)	49.2 (0.46%)	154.2 (2.12%)	35.3 (0.90%)	83.3 (1.72%)
32000	Herbaceous vegetation associations		17.2 (0.16%)			
33000	Open spaces with little or no vegetation				25.9 (0.66%)	0.1 (0.00%)
40000	Wetlands	23.4 (0.21%)			24.8 (0.63%)	
50000	Water bodies	185.1 (1.67%)	81.1 (0.76%)	119.3 (1.64%)	156.6 (4.01%)	10.4 (0.21%)

Proportion of major LULC classes (urban fabric, transport, green, and water) used to support the structural interpretation of urban noise baselines and cross-city prediction performance.

6.4.4.5 Implications for Cross-Domain Transfer

The findings presented in this chapter underscore the critical role of urban form and functional structure in shaping both the baseline acoustic environment and the resulting predictive behaviour of machine learning models under cross-domain settings. While the dual-branch graph neural network demonstrates stable and competitive performance across all five cities, the observed variations in error magnitude and exposure distribution are strongly conditioned by structural differences rather than by any intrinsic instability of the model. This insight carries several implications for both methodological development and practical applications of urban noise mapping.

First, the results highlight the need to contextualize model performance within the acoustic and morphological characteristics of the target domain. Cities with high built-up ratios and concentrated transportation infrastructure inherently present steeper noise gradients and higher peak values, making them more challenging prediction environments. In contrast, cities with lower structural intensity and higher shares of green and open spaces exhibit lower baselines and reduced variance, which naturally lead to lower error metrics. These differences should therefore be interpreted as reflecting environmental structure rather than disparities in model capability.

Second, the integration of morphological and land-use indicators offers a promising avenue for improving domain adaptation strategies. By explicitly incorporating structural descriptors—such as built-up ratio, land-use entropy, and transport corridor density—into the model transfer process, it may be possible to calibrate domain alignment more effectively, mitigating residual performance gaps between cities with divergent acoustic regimes. Such an approach could enhance the robustness and interpretability of predictive models deployed in complex, multi-city contexts.

Finally, the structural interpretation of urban noise baselines also has practical policy relevance. Recognizing that cities with lower built-up intensity and greater green space coverage inherently experience lower baseline exposure provides a planning lever for long-term noise mitigation. Urban greening, balanced land-use allocation, and diffusion of traffic corridors are not only environmental planning objectives but also structural determinants of the acoustic landscape. Embedding these insights into urban development strategies could complement conventional noise control interventions, supporting more sustainable and health-oriented urban environments.

In sum, the cross-city analysis demonstrates that urban morphology is not merely a contextual background but a fundamental determinant of both acoustic exposure and model behaviour. Future research should therefore more systematically integrate these structural factors into model design, domain adaptation, and policy formulation.

6.4.5 Limitations

While the proposed graph-based framework demonstrates strong predictive capacity and cross-city transferability, several inherent limitations reflect broader challenges in data-driven environmental modeling, especially for applications involving complex physical processes like urban noise propagation.

First, the model is fundamentally grounded in a statistical learning paradigm that infers noise patterns from observable surface features, rather than explicitly simulating the physics of sound propagation. This abstraction brings clear advantages in scalability and computational efficiency but sacrifices physical fidelity. Unlike classical noise prediction models such as ISO 9613 or CNOSSOS-EU, which explicitly incorporate source emissions, propagation equations, and attenuation mechanisms (Salomons, 2001; Kephelopoulos et al., 2021), the current framework relies on correlations learned from training data. While similar data-driven approaches have been increasingly applied in air quality and climate risk modeling (Rolnick et al., 2023), they may underperform in environments where detailed physical interactions, such as building reflections, wind patterns, or terrain diffraction, play dominant roles.

Second, the framework's dependence on high-quality reference labels and pseudo-labeling introduces data-related uncertainties. The model's generalization hinges on the assumption that the reference city (Southampton) adequately represents the structural patterns of target cities. However, this assumption can fail if key urban forms, such as elevated highways or port complexes, are underrepresented in the reference domain. Research in transfer learning has shown that inadequate coverage of functional classes or environmental variability can degrade cross-domain model robustness (Tuia et al., 2016). Furthermore, while confidence-based pseudo-labeling improves reliability, it remains sensitive to the density and quality of the pseudo-supervision pool. Sparse pseudo-labeled samples, as seen in cities like Portsmouth, can destabilize learning and reduce local prediction accuracy .

Third, although the multi-domain adversarial alignment improves robustness to domain shift, residual biases persist—especially in spectrally sensitive features like vegetation indices or texture measures. These features are influenced by seasonal

variability, sensor differences, and landscape context, which cannot always be fully normalized through histogram matching or pseudo-invariant feature alignment (Schott et al., 1988). Over-regularization during adversarial training may inadvertently suppress informative city-specific signals, a challenge recognized in recent studies on domain-adaptive geospatial learning (Wilson and Cook, 2020).

Fourth, the framework is inherently static, ignoring temporal variations such as time-of-day traffic patterns, weather conditions, or episodic events. All predictions are derived from static land cover and remote sensing proxies, neglecting dynamic factors that are increasingly recognized as critical in environmental noise assessments (Kephelopoulou et al., 2012; Basner et al., 2014). This limits the model's applicability in contexts where temporal drivers, such as nightlife activity or rush-hour traffic, dominate noise dynamics.

Beyond these technical limitations, the framework currently lacks explainability mechanisms that could enhance its utility for policy and planning. While the feature selection stage increases transparency, the internal representations learned by the graph neural network remain largely opaque, complicating causal interpretation and uncertainty analysis. This echoes a broader concern within environmental science, where the “black-box” nature of deep learning has triggered calls for integrating explainable AI and uncertainty quantification methods (Karniadakis et al., 2021).

These limitations highlight a fundamental tension in data-driven urban noise modeling: the tradeoff between statistical generalization and physical process fidelity. While machine learning frameworks like GNNs excel at capturing complex spatial dependencies, they remain largely agnostic to the governing physics of acoustic propagation. This raises legitimate concerns about plausibility, especially when extrapolating to novel urban forms or acoustically complex settings.

Emerging research in hybrid modeling suggests promising pathways forward. Physics-guided machine learning (PGML) approaches aim to combine the representational power of neural networks with the rigor of physical constraints, introducing domain knowledge through customized loss functions, constrained embeddings, or hybrid neural-simulation architectures (Karpatne et al., 2017; Willard et al., 2022). For urban noise, such methods could encode constraints on attenuation over distance, building reflections, or terrain effects, enhancing both predictive robustness and interpretability. This perspective aligns with recent recommendations in Earth system science, which emphasize the importance of combining empirical data-driven models with physically credible frameworks to ensure reliable and actionable predictions.

6.4.6 Future Research Directions

Building upon the methodological innovations and acknowledged limitations of this study, several promising avenues emerge for future research that could substantially advance the field of urban noise modeling. These directions aim not only to enhance predictive accuracy but also to improve interpretability, scalability, and theoretical soundness.

A key priority for future work lies in integrating physical priors into graph-based urban noise models. While the current framework effectively captures spatial dependencies and morphological structures, it remains largely agnostic to the underlying physics of sound propagation. Incorporating physically informed constraints—such as attenuation with distance, reflection/absorption by building materials, and diffraction—could improve both realism and interpretability. Recent developments in physics-informed machine learning provide practical pathways for embedding domain knowledge via physics-guided losses or hybrid neural–physical components (Karniadakis et al., 2021; Willard et al., 2020). Such approaches have already shown promise in climate and hydrological applications (Reichstein et al., 2019) and could substantively strengthen the credibility of data-driven urban acoustic models.

Another critical avenue is enriching target-side supervision through multi-source integration. While the present study primarily relied on pseudo-labels transferred from a single reference city, future work can fuse limited in-situ noise measurements, participatory sensing (e.g., mobile crowdsourcing) (Maisonneuve et al., 2010), and auxiliary environmental datasets (traffic counts, urban morphology inventories). Prior work in domain adaptation/generalization indicates that combining diverse supervisory signals improves robustness under distribution shift and label sparsity (Tuia et al., 2016; Zhou et al., 2022).

The exploration of semi-supervised and few-shot learning is also promising for emerging or data-scarce cities. Beyond high-confidence pseudo-labels, graph-based semi-supervised learning and label propagation can leverage graph structure under extreme scarcity (Zhu and Ghahramani, 2002). Few-shot learning methods (Wang et al., 2020) and contrastive representation learning (Chen et al., 2020) offer pathways to build transferable encoders with minimal annotation.

Incorporating temporal dynamics and multi-source sensor fusion is another key step. Current models—including the one presented here—use static proxies and neglect diurnal traffic cycles, weather, and episodic noise events. Advances in spatio-temporal graph learning (Yu et al., 2018; Li et al., 2018) enable integration of real-time sensor streams, traffic flow models, and human mobility data, supporting predictions that are both spatially nuanced and temporally responsive.

Finally, large-scale urban noise benchmarks and universal representations would catalyse progress. Analogous to ImageNet’s role in computer vision (Deng et al., 2009), an open, multi-city noise benchmark spanning diverse environments and seasons would enable rigorous comparison and reproducibility. Developing transferable, domain-invariant representations of urban acoustic environments—validated across such benchmarks—would lay foundations for policy-relevant, globally applicable systems.

Chapter 7: General Discussion

7.1 Introduction

Urban noise pollution remains one of the most pressing environmental stressors in contemporary cities, with well-documented impacts on public health, ecological systems, and urban livability (Basner et al., 2014; WHO, 2018). As urbanization accelerates globally, the complexity of noise sources—including traffic, industrial activities, construction, and social gatherings—has increased dramatically, challenging the capacity of traditional modeling and mapping approaches. Established frameworks such as physical acoustic simulation and land-use regression (LUR) models, while robust in controlled or data-rich environments, often face severe scalability issues when applied to diverse and rapidly evolving urban settings (EEA, 2020). These methods typically require extensive ground measurements, detailed traffic counts, and specialized calibration data, all of which entail substantial financial, logistical, and temporal costs (Murphy and King, 2014; Kephelopoulous et al., 2012).

Against this background, the present dissertation positions itself at the intersection of urban environmental science, remote sensing, and machine learning, seeking to explore whether high-resolution multispectral imagery, when coupled with advanced graph-based neural networks, can offer a scalable alternative to conventional urban noise modeling. Specifically, this work investigates whether generalizable, interpretable, and cost-efficient noise prediction models can be constructed solely from remotely sensed environmental proxies and publicly available land-use inventories. This research ambition aligns with a broader scientific shift toward multi-source, data-driven urban analytics, where artificial intelligence methods increasingly supplement or replace traditional simulation frameworks (Maxwell et al., 2017).

Rather than simply summarizing the technical details elaborated in earlier chapters, this synthesis chapter focuses on critical reflection and integration across the empirical and conceptual contributions presented in Chapters 4, 5, and 6. The aim here is threefold. First, to distill the overarching methodological and theoretical insights derived from comparing convolutional neural networks (CNNs), graph neural networks (GNNs), and dual-branch spatial graph architectures for urban noise modeling. Second, to evaluate how these innovations advance the field of urban environmental prediction, particularly in addressing long-standing issues of domain transferability, data scarcity, and computational efficiency (Reichstein et al., 2019; Karniadakis et al., 2021). Third, to identify the broader policy and research implications of developing interpretable, transferable noise models that can be deployed in cities lacking extensive monitoring infrastructure.

By foregrounding these higher-level discussions, this chapter seeks to move beyond procedural exposition and engage with the fundamental scientific, practical, and societal relevance of the work. Specifically, it aims to articulate how the integrated innovations of this dissertation contribute to the evolving landscape of scalable

environmental monitoring, situate them within global best practices in urban informatics, and chart concrete pathways for future research and application.

7.2 Theoretical and Methodological Contributions: Systematic

Synthesis

This dissertation offers a set of important theoretical and methodological contributions to the emerging field of urban noise modeling, positioned at the intersection of remote sensing, machine learning, and environmental science. Traditional noise prediction approaches, such as physical acoustic modeling and land-use regression, have long depended on detailed field measurements, traffic flow data, or calibrated emission models. While effective in some contexts, these methods face well-documented challenges when applied to large, heterogeneous, or data-scarce urban environments due to their cost, complexity, and limited transferability. The work presented here breaks from this convention by demonstrating how scalable, data-driven frameworks can leverage remote sensing imagery and graph-based architectures to predict urban noise with minimal dependence on ground measurements.

A central theoretical insight developed in this dissertation is that high-resolution multispectral remote sensing data can serve as an independent and sufficiently rich basis for estimating urban noise patterns. Historically, noise modeling has been seen as inseparable from dense ground-based data, yet this study shows that surface reflectance properties, vegetation indices, and morphological textures captured from space encode meaningful environmental proxies for both noise generation and attenuation processes. This idea builds on a growing body of work in urban environmental monitoring, where remotely sensed features have been successfully used to estimate air quality, heat risk, and ecological indicators (Weng, 2012; Maxwell et al., 2017). By extending this perspective to urban acoustics, the research opens up new possibilities for low-cost, transferable noise modeling, particularly in cities where traditional data inputs are unavailable.

The methodological innovations in this dissertation reflect an intentional progression through increasingly sophisticated learning architectures. Initial experiments using convolutional neural networks (CNNs), specifically EfficientNet, capitalize on the powerful feature extraction and scaling properties of modern deep learning (Tan & Le, 2019; Paoletti et al., 2019). Moving beyond CNNs, the research introduces six-layer GraphSAGE models, which mark a methodological leap by explicitly modeling spatial relationships between observation points. This transition aligns with the broader trend in geospatial AI toward graph-based methods that recognize and exploit spatial interdependencies, rather than treating samples as isolated units (Hamilton et al., 2017; Wu et al., 2020).

Perhaps the most significant advance is the development of the dual-branch graph neural network (GNN) framework, which integrates fine-scale morphological predictors with broader-scale land-use statistics to reflect the multi-scale nature of urban noise environments. This multi-modal architecture is directly responsive to calls in the literature for models that combine local detail with regional context, acknowledging that noise exposure is shaped by both proximate sources (such as road networks or building facades) and broader urban structures (like land-use zoning or green space distribution) (Salomons and Pont, 2012; Basner et al., 2014). Moreover, the inclusion of domain adaptation techniques and pseudo-labeling mechanisms addresses a major challenge in environmental modeling: how to generalize predictive systems across cities with divergent morphological and socio-environmental characteristics (Pan et al., 2009).

Importantly, this work does not simply apply machine learning tools as black boxes; rather, it thoughtfully integrates feature interpretability, physical plausibility, and spatial reasoning. The model's use of acoustically informed edge weights, interpretable feature selection, and multi-domain adversarial alignment demonstrates a commitment to ensuring that predictive accuracy is not achieved at the expense of scientific or environmental credibility. This aligns with emerging best practices in environmental data science, where increasing attention is being given to the need for models that are not only performant but also transparent, interpretable, and physically meaningful (Karpatne et al., 2017; Reichstein et al., 2019).

The interdisciplinary value of this research extends beyond technical advancements. By offering scalable, interpretable, and transferable frameworks for urban noise mapping, the dissertation contributes practical solutions for urban planners, public health professionals, and policymakers. It supports new forms of environmental monitoring in data-scarce regions, enhances the capacity for cross-city comparison and benchmarking, and helps lay the groundwork for integrating noise exposure assessments into broader smart city initiatives and urban sustainability efforts (WHO, 2018; European Environment Agency, 2020).

In conclusion, the contributions of this dissertation reside not merely in any one model or algorithm, but in the broader conceptual shift it represents: positioning remote sensing as a central resource for urban acoustics, advancing methodological sophistication through graph-based learning, and embedding concerns of interpretability and transferability into the very design of environmental AI systems. Together, these innovations significantly strengthen the foundations of urban noise science and point toward new pathways for robust, scalable, and globally relevant noise prediction frameworks.

7.3 Comparative Analysis of Three Modelling Paradigms: Performance, Resolution, and Computational Trade-offs

This dissertation provides a unique opportunity to systematically compare three major modeling approaches for urban noise prediction: the EfficientNet-based convolutional neural network (CNN) from Chapter 4, the six-layer GraphSAGE graph neural network (GNN) from Chapter 5, and the dual-branch GNN introduced in Chapter 6. By drawing together quantitative evaluation results, architectural design characteristics, computational resource considerations, and practical workflow demands, the discussion reveals not only the relative strengths of each approach but also the nuanced trade-offs they entail.

In terms of predictive performance, all models were benchmarked using mean absolute error (MAE, measured in dBA) and the coefficient of determination (R^2), two metrics widely adopted in international urban environmental modeling (Basner et al., 2014). The EfficientNet-CNN achieved an MAE of 4.79 dBA and an R^2 of 0.491, offering a reasonable baseline using solely image-based features at 30-meter resolution. The GraphSAGE GNN, operating on a much finer 4-meter grid, improved these scores to an MAE of 4.40 dBA and an R^2 of 0.596, demonstrating how explicit modeling of spatial relationships can enhance fidelity. The dual-branch GNN achieved an MAE of 4.48 dBA and an R^2 of 0.576—slightly below the peak accuracy of GraphSAGE, yet crucially balancing this performance with superior cross-city generalizability and a dramatically reduced input feature space (171 variables versus 8250).

It should be noted that the CNN architecture implemented in Chapter 4 was developed during the early stages of this research. At that time, the model adopted a fixed input patch size and did not incorporate more advanced scale-flexible or multi-scale receptive field designs. This was a deliberate early-stage design choice aimed at establishing a proof-of-concept rather than exhausting the potential of CNN-based approaches. As such, the CNN results should be interpreted with caution—not as a ceiling on CNN performance, but as an initial experimental baseline. Future work could explore adaptive patching strategies, scale-aware modules, or hybrid CNN–graph architectures to address this limitation.

These differences highlight an important point: finer spatial resolution and graph complexity can push predictive accuracy higher, but often at the cost of scalability. The 30-meter resolution adopted in both the CNN and dual-branch GNN aligns with the native scale of WorldView-2 remote sensing products, making it computationally efficient for wide-area deployment. In contrast, the 4-meter grid in the GraphSAGE model, while offering superior spatial detail, imposed heavy demands on memory and graph construction, raising questions about its practicality for large-scale, multi-city applications (Paoletti et al., 2019).

From a computational resource perspective, CNNs demand substantial GPU memory and training time because they ingest raw image data, consistent with findings in the broader deep learning and remote sensing literature (Kattenborn et al., 2021). While

GNNs were less GPU-intensive, they required significant pre-processing time, particularly in the feature engineering and graph construction phases, where thousands of node-level variables had to be computed and integrated. Notably, the dual-branch GNN introduced an effective solution to this bottleneck by employing rigorous feature selection and importance ranking, reducing the input feature space by more than 97% compared to the GraphSAGE approach, while maintaining competitive performance.

In terms of workflow complexity and practical applicability, CNNs offer an advantage in simplicity, making them well-suited for rapid deployment in data-rich environments where raw imagery is readily available. However, their current implementation in this study did not fully leverage scale-adaptive or context-aware capabilities, which likely contributed to their relatively lower performance compared with the GNN approaches. By contrast, GNNs naturally accommodate spatial dependencies and can incorporate domain knowledge more flexibly, aligning with growing international consensus that graph-based models are particularly well-suited for tasks such as noise propagation, where relational structure matters (Hamilton et al., 2017).

Taken together, the empirical evidence and computational profiles of these models suggest that graph-based approaches—especially the dual-branch GNN developed in this dissertation—strike the best balance between accuracy, generalization, computational efficiency, and interpretability. This finding resonates with a broader shift in environmental informatics toward graph-structured representations and spatially aware machine learning, particularly in complex urban applications where heterogeneity, multi-scale effects, and cross-domain transferability are critical (Reichstein et al., 2019; Maxwell, 2021). However, the analysis also underscores key future challenges, notably the need for further reductions in computational demands, improvements in graph construction automation, and continued refinement of cross-city generalization strategies.

A methodological reflection that emerged from this dissertation concerns the role of model depth in environmental prediction tasks. In mainstream deep learning literature, performance gains are frequently associated with increasing model depth and representational capacity. However, the results obtained here challenge this assumption in the specific context of urban noise mapping. The GraphSAGE framework implemented in Chapter 5, consisting of six stacked convolutional layers, already captured sufficient multi-scale dependencies to outperform shallower baselines. Similarly, the dual-branch model in Chapter 6, though relatively shallow in absolute terms—comprising only two GAT layers and one GCN layer, coupled with pseudo-labelling and MME loss gating—delivered strong cross-city generalisation. These outcomes suggest that, in data-limited and highly structured environmental domains, architectural design and task-specific constraints matter more than raw network depth. Excessive stacking of layers would not only increase computational burden but also risk over-smoothing in GNNs, or vanishing gradients in CNNs, without necessarily yielding meaningful improvements in predictive accuracy.

This observation aligns with recent findings in geospatial and environmental AI research, where the marginal returns of depth diminish rapidly once networks achieve sufficient receptive field coverage and relational embedding. Instead, the integration

of domain-adaptation strategies, semantic features, and loss-function design proved decisive for model performance. The implication is that future work should prioritise architectural parsimony and domain-informed design over indiscriminate depth scaling, particularly in resource-constrained urban applications where interpretability and computational efficiency are of equal importance to accuracy.

7.4 Complementarity and Advancement Beyond International Urban Noise Mapping Frameworks

A central contribution of this dissertation is its ability to not only position itself within, but also expand beyond, the dominant international paradigms of urban noise mapping. For nearly two decades, frameworks such as the European Noise Directive (END) and the CNOSSOS-EU standardized methodology have set the benchmarks for environmental acoustics research and urban noise policy, particularly across European contexts (Kephalopoulos et al., 2012; Murphy and King, 2014). These systems have enabled standardized monitoring, cross-national comparability, and informed regulatory actions (de Kluizenaar et al., 2009; European Environment Agency, 2020). However, they are inherently data-intensive, requiring extensive traffic, land use, and infrastructural data, as well as detailed physical modeling — making their application in heterogeneous or data-scarce urban environments challenging (Basner et al., 2014).

This dissertation offers both empirical and conceptual advancements by demonstrating how multispectral remote sensing, combined with advanced machine learning methods such as graph neural networks (GNNs), can serve as scalable and transferable complements — or even alternatives — to traditional noise mapping systems. Recent studies applying explainable machine learning to urban noise assessment have shown that ensemble models such as random forests and gradient boosting can outperform traditional linear models in traffic noise prediction when using limited or non-traditional datasets (Aletta et al., 2020). Similarly, deep neural networks (DNNs) and convolutional neural networks (CNNs) have demonstrated strong performance in capturing spatial and temporal noise characteristics, reinforcing the potential of ML to revolutionize environmental acoustics (McLoughlin et al., 2015; Mesaros et al., 2016).

Importantly, the dissertation's dual-branch GNN model does not merely replicate what regulatory models already do; rather, it introduces conceptual innovations that address structural blind spots in conventional frameworks. While land-use regression models and deterministic simulations often overlook topological relationships, GNNs explicitly encode multi-scale spatial interactions, enabling more realistic modeling of noise propagation through complex urban landscapes. By integrating fine-grained morphological predictors from remote sensing with broader-scale land use indicators, the dissertation responds to international calls for environmental analytics that are not only data-rich but also structurally sophisticated (Reichstein et al., 2019).

From a policy and practice perspective, the proposed approach is designed to be complementary rather than competitive with existing regulatory frameworks. Traditional systems such as the European Noise Directive (END) and the CNOSSOS-EU methodology remain indispensable for legal compliance and standardized noise management. However, machine learning-based approaches provide scalable, cost-efficient alternatives for rapid scenario testing, large-area screening, and targeted interventions—particularly valuable for rapidly urbanizing regions and smart city infrastructures aiming to integrate dynamic digital twins (Batty, 2018; European Environment Agency, 2020).

Recent empirical work illustrates this potential. For example, a Random Forest-based noise evaluation framework developed for Nanchang (China) demonstrated how explainable machine learning can effectively support spatial planning decisions with limited traditional input data (Teng et al., 2024). Similarly, urban street-level noise estimation based on street view imagery has been shown to produce reliable, fine-scale predictions along transportation corridors, providing a cost-effective and transferable tool for municipal applications (Huang et al., 2024). At an even broader scale, vision-acoustic cross-modal inference methods have enabled efficient sensing of urban soundscapes in data-scarce environments, offering new pathways for large-area environmental assessments (Zhao et al., 2023).

At the same time, this research critically acknowledges the limitations of machine learning and remote sensing approaches, including the risks of overfitting, interpretability challenges, and the absence of explicit physical causality. As noted by Karpatne et al. (2017), the integration of machine learning with physical process models—so-called theory-guided data science—represents a promising direction to address these issues. In line with global trends, the future of urban noise modeling is likely to rely on hybrid frameworks that combine the scalability of AI, the physical grounding of traditional noise models, and the interpretability required for policy and public trust.

In summary, this dissertation contributes a technically robust, conceptually innovative, and practically scalable complement to established regulatory frameworks. It bridges the gap between data-intensive simulation models and next-generation, AI-driven approaches, pointing toward a future where remote sensing, machine learning, and graph analytics jointly extend the reach, reduce the cost, and amplify the policy relevance of urban environmental governance.

7.5 Limitations and Areas for Improvement

Despite the considerable methodological advances presented in this dissertation, several inherent limitations remain, reflecting both case-specific constraints and broader challenges within the AI-driven environmental modeling field. Critically examining these limitations is essential, as it not only frames the scope of the present research but also provides a foundation for meaningful future improvements, aligning with the growing international discourse on reliable, explainable, and multimodal urban analytics (Karpatne et al., 2017; Reichstein et al., 2019).

7.5.1 Limitations of Remote Sensing Imagery

One foundational limitation of the proposed framework stems from its reliance on high-resolution multispectral remote sensing imagery. Although modern satellite systems such as WorldView-2 offer rich spatial detail, their outputs remain constrained by atmospheric conditions, sensor artifacts, and cloud cover, all of which can introduce noise and reduce the reliability of derived features (Wulder et al., 2012; Zhu et al., 2017). Furthermore, the spatial resolution of optical satellite imagery, while sufficient for broad urban classification, falls short of capturing micro-scale morphological elements that substantially affect local noise propagation—such as narrow street canyons, small vegetation barriers, or detailed building layouts (Herold et al., 2004).

Beyond spatial limitations, multispectral imagery inherently lacks dynamic or subsurface information. While it can indicate static proxies such as land use type or surface materials, it remains blind to temporally fluctuating factors like real-time traffic volume, human activity patterns, or transient noise barriers. These missing dimensions are often critical for urban acoustic modeling, as emphasized in recent studies highlighting the integration of temporally resolved data from mobile monitoring, traffic sensors, and participatory sensing networks to capture fine-grained dynamics in urban sound environments (Maisonneuve et al., 2010; Aletta et al., 2016).

Addressing these limitations requires embracing multimodal integration. Future research should combine satellite-based information with complementary inputs, such as airborne LiDAR for three-dimensional urban morphology, traffic datasets for dynamic activity profiles, and crowd-sourced acoustic data. Recent developments in urban environmental analytics have shown that such multimodal approaches improve predictive accuracy and reduce blind spots, particularly in heterogeneous cityscapes (Li et al., 2016; Zheng et al., 2018).

7.5.2 Model Interpretability and the Black-Box Challenge

Another significant challenge arises from the interpretability of deep learning models, particularly graph neural networks (GNNs), which form the core of the final architecture presented in this dissertation. While GNNs excel at capturing complex spatial dependencies and relational patterns, their internal representations are often opaque, making it difficult to trace specific predictions back to physically meaningful features or mechanisms. This “black box” limitation is increasingly recognized as a barrier to the broader adoption of AI in environmental science, where stakeholders and policymakers require models that are not only accurate but also explainable and transparent (Karpatne et al., 2017; Reichstein et al., 2019).

In this research, several strategies were adopted to partially mitigate these concerns, including the use of physically interpretable input features (e.g., land-use ratios and morphological indicators) and explicit edge-weight modulation based on

environmental relevance. However, the graph embeddings and learned message-passing processes themselves remain challenging to interpret without specialized tools.

Recent work points to promising directions for enhancing GNN interpretability, including post-hoc explanation frameworks such as GNNExplainer, attention visualization methods for edge-level attribution, and hybrid architectures that combine neural learning with symbolic, rule-based, or physics-informed components (Ying et al., 2019; Pope et al., 2019; Yuan et al., 2021). Incorporating such strategies in future urban noise prediction frameworks would improve their trustworthiness and facilitate their integration into operational urban planning and environmental management workflows.

7.5.3 Graph Construction and Spatial Distance Encoding in GNNs

A further technical limitation stems from the design and construction of the graph structures underlying the GNN models. In this dissertation, spatial graphs were constructed primarily using Euclidean distances between grid points, a common approach in spatial machine learning due to its simplicity and computational efficiency (Hamilton et al., 2017; Wu et al., 2020). However, Euclidean distance does not always reflect the true pathways or barriers relevant to urban noise propagation. For example, noise tends to travel preferentially along transportation networks, while being blocked or attenuated by buildings, vegetation, or elevation changes. Straight-line distance-based graphs, therefore, risk misrepresenting the functional connectivity of urban acoustic spaces, potentially introducing biases into the learned representations (Aletta et al., 2016; Kang, 2007).

While this study incorporated land-use-derived edge weighting to partially account for environmental relevance, more advanced approaches are needed to fully capture the anisotropic and complex nature of urban noise diffusion. Recent advances in spatial graph construction recommend the use of network-constrained topologies (e.g., road graphs), simulation-informed adjacency matrices derived from acoustic propagation models, or dynamic graphs with edge weights that adapt to temporal or environmental conditions (Vaswani et al., 2017; Bai et al., 2020; Sahili et al., 2023). Integrating such innovations could significantly enhance the physical realism and predictive fidelity of future urban acoustic models.

7.5.4 Toward Integrated and Multimodal Approaches: Broader Literature Perspectives

Beyond the technical and data-specific challenges discussed above, a major avenue for future research involves the integration of multimodal data sources and hybrid modeling approaches. While this dissertation focused primarily on high-resolution multispectral remote sensing and land use/land cover (LULC) indicators, urban noise environments are shaped by a much wider set of interacting variables—including dynamic traffic flows, meteorological variability, and human activity patterns. Single-

modality models, regardless of architectural sophistication, can only partially approximate these complex dynamics.

Emerging studies in environmental informatics demonstrate the potential of fusing heterogeneous datasets—such as real-time traffic data, weather station records, crowd-sourced sensing streams, and mobile acoustic measurements—into unified prediction frameworks for complex urban phenomena. Such multimodal architectures enhance representational richness while enabling spatiotemporal modeling strategies capable of capturing both static background conditions and dynamic fluctuations in urban soundscapes (Zheng et al., 2018; Bai et al., 2020; Wu et al., 2020). For example, integrating spatial graph neural networks with temporal models such as recurrent neural networks or temporal graph convolution allows for fine-grained modeling of diurnal, weekly, or seasonal noise variability—an increasingly emphasized direction in recent spatiotemporal AI research (Yu et al., 2018; Li et al., 2018).

Complementing multimodal fusion, the embedding of physical constraints and domain knowledge into machine learning workflows represents another promising line of advancement. Approaches such as physics-informed neural networks (PINNs), hybrid simulation-learning architectures, and constraint-based regularization have already achieved strong results in domains like hydrology, climate modeling, and air quality forecasting (Karpatne et al., 2017; Reichstein et al., 2019; Raissi et al., 2019). Applying such principles to urban acoustics would allow models to encode physical relationships—such as distance-based attenuation, surface reflection, or regulatory thresholds—directly within their learning objectives or architectural design, thereby improving physical realism, generalizability, and interpretability.

In summary, while this dissertation advances scalable and generalizable urban noise prediction through the integration of remote sensing, graph learning, and domain adaptation, it also acknowledges several limitations that must be addressed to fully unlock the potential of AI-driven environmental modeling. These limitations span data fidelity, interpretability, physical grounding, and integration with broader urban systems. The most promising solutions lie in fusing multimodal datasets, embedding domain-specific physical constraints, developing dynamic and context-aware graph architectures, and advancing explainability tools tailored to planning and policy needs.

By pursuing these directions, future research can foster a new generation of urban noise models that are not only computationally efficient and scientifically robust but also operationally impactful—empowering cities worldwide to better monitor, manage, and mitigate the complex acoustic environments that shape urban life.

7.6 Policy and Practical Implications

The innovations developed in this dissertation carry substantial implications for urban policy, environmental management, and the broader landscape of sustainable city governance. By combining high-resolution remote sensing, graph-based machine learning, and scalable prediction architectures, the research introduces practical tools

that extend beyond theoretical contributions, offering actionable pathways for improving urban noise governance.

A key implication lies in lowering the barriers to large-scale noise monitoring, particularly in cities with limited technical or financial resources. Traditional environmental noise surveillance systems rely heavily on dense in situ measurement networks and detailed traffic or infrastructural datasets, making them expensive and logistically demanding (Murphy & King, 2014; Murphy & King, 2010). In contrast, the data-driven, transferable modeling framework developed here enables cities—especially those in rapidly urbanizing regions or the Global South—to generate reliable, high-resolution noise exposure estimates without requiring such exhaustive inputs. This contributes to environmental equity by democratizing access to critical environmental health information, aligning with global sustainability and public health agendas (Maxwell, 2021).

In the domain of urban planning and infrastructure development, the methods proposed in this work support faster and more targeted environmental impact assessments. Predictive noise models integrated with urban land use and morphological data allow planners to identify existing and emergent noise hotspots, evaluate the likely benefits of mitigation strategies, and prioritize interventions in sensitive areas such as residential neighborhoods, schools, and hospitals. As the link between environmental noise exposure and public health outcomes becomes increasingly formalized in both European and international policy frameworks (Basner et al., 2014; WHO, 2018), scalable and reproducible models like those presented here can support cities in regulatory compliance, proactive health risk assessment, and the design of more livable urban environments.

Beyond supporting regulatory compliance, the integration of graph-based machine learning aligns with the growing vision of smart city governance. Embedding predictive tools within digital twin environments enables urban managers to simulate development or policy scenarios, integrate real-time sensor streams, and iteratively refine strategies as conditions evolve (Batty, 2018; Allam & Dhunny, 2019). Such adaptive, data-rich governance models foster more resilient, responsive, and citizen-centered urban systems, enhancing the capacity of local governments to address complex environmental challenges.

Importantly, the attention to model interpretability and physical plausibility directly addresses contemporary concerns about the responsible use of AI in public policy. Black-box predictive systems, while powerful, often face resistance when deployed in high-stakes governance contexts due to issues of trust, accountability, and explainability (Reichstein et al., 2019). By prioritizing interpretable features, transparent workflows, and domain-informed architectural choices, this dissertation supports the development of predictive systems that can withstand policy scrutiny and enable meaningful stakeholder engagement—including by planners, regulators, and affected communities.

Finally, the open-source, scalable nature of the developed frameworks facilitates a broader democratization of environmental knowledge. Tools that can be deployed using accessible data sources and reproducible computational pipelines empower not only formal institutions but also local communities, NGOs, and citizen science

initiatives. This participatory dimension aligns with calls in environmental governance for more inclusive, bottom-up approaches to urban health management, where local voices play an active role in shaping both data collection and policy outcomes (Goodchild, 2007; Haklay, 2013).

In summary, the contributions of this dissertation extend well beyond academic and technical domains, offering new opportunities for cities to improve environmental monitoring, policy design, and public engagement. By providing scalable, interpretable, and adaptable modeling solutions, the research helps bridge the gap between cutting-edge computational techniques and the real-world needs of urban governance in the 21st century.

7.7 Future Research Directions

While this dissertation has demonstrated notable advances in scalable and generalizable urban noise prediction, the rapid evolution of artificial intelligence, urban analytics, and geospatial technologies opens several promising avenues for future research. These directions emerge both from the limitations identified in the present work and from broader trends in interdisciplinary environmental data science.

A critical next step involves the incorporation of physics-informed machine learning (PIML) approaches, which combine the flexibility of data-driven models with the robustness and interpretability of physical principles. By embedding domain knowledge, conservation laws, or even differentiable simulation modules within neural architectures, future models could ensure physically consistent predictions, improving both reliability and policy credibility. Recent advances in Earth system modeling and environmental forecasting demonstrate the feasibility and value of such hybrid approaches, offering practical templates for applications in urban acoustics (Karpatne et al., 2017; Reichstein et al., 2019; Willard et al., 2022). For urban noise mapping, coupling graph neural networks (GNNs) or convolutional neural networks (CNNs) with established acoustic simulation tools—or regularizing learning with known propagation constraints—could substantially enhance model fidelity, particularly in complex or under-sampled environments.

Another promising area is multimodal and sensor-fusion modeling. As urban sensor networks, mobile devices, and participatory sensing initiatives proliferate, the integration of heterogeneous data streams—including real-time acoustic sensors, traffic flows, weather data, and high-resolution remote sensing—offers unprecedented opportunities to capture the dynamic complexity of urban soundscapes. However, such integration presents significant technical challenges, including harmonizing structured and unstructured data, managing biases, and designing architectures capable of processing asynchronous, multi-source inputs. Advances in multimodal learning, attention mechanisms, and adaptive fusion strategies (Maxwell et al., 2021; Huotari et al., 2024) provide promising methodological starting points for addressing these challenges.

The advancement of semi-supervised, few-shot, and transfer learning techniques also holds great promise. Given the scarcity of labeled noise data in many urban contexts, especially in rapidly urbanizing regions, maximizing the value of unlabeled or weakly labeled data is a high-priority research need. Approaches such as self-supervised learning, domain adaptation, and federated learning could enable models to generalize more effectively across diverse cities without requiring extensive local calibration (Zhang et al., 2022; Kairouz et al., 2021). Moreover, privacy-preserving strategies would allow cross-jurisdictional knowledge sharing while respecting data sovereignty and confidentiality constraints.

Temporal dynamics represent another largely untapped dimension in current urban noise modeling. Most models—including those developed in this dissertation—focus on static or snapshot predictions, neglecting the pronounced temporal variability in urban sound environments driven by daily cycles, seasonal shifts, meteorological factors, and episodic events. Incorporating time-aware modeling—through recurrent neural networks (RNNs), temporal graph neural networks, or spatiotemporal transformers—could allow predictions that reflect not only spatial heterogeneity but also diurnal and seasonal rhythms, improving both accuracy and policy relevance (Yu et al., 2018; Wu et al., 2020).

Beyond methodological innovation, the creation of large-scale, standardized urban noise benchmarks and open-access datasets would substantially accelerate progress. Just as benchmark datasets such as ImageNet (Deng et al., 2009) transformed the field of computer vision, comprehensive, multi-city urban noise datasets with standardized formats and evaluation protocols would provide essential foundations for reproducible research and comparative assessment of competing methods. Such resources would enable the research community to rigorously test generalization, transferability, and scalability claims, thereby strengthening the scientific foundation of the field.

Finally, the practical integration of advanced noise models into real-world decision-support systems remains a critical frontier. As cities increasingly adopt digital twin frameworks, smart city dashboards, and interactive urban analytics platforms, future research should focus on automating data pipelines, improving real-time model responsiveness, and developing user-friendly interfaces that empower planners, policymakers, and communities. Collaborative system design—incorporating feedback from diverse stakeholders—will be essential to ensure that these tools are not only technically advanced but also accessible, actionable, and aligned with societal needs (Batty, 2018; Allam & Dhunny, 2019; Somanath et al., 2024).

In sum, the future of urban noise modeling lies at the intersection of multimodal data integration, physics-informed AI, scalable adaptive architectures, and participatory system design. Advancing along these lines will not only improve scientific understanding of urban sound environments but also drive meaningful improvements in public health, environmental justice, and the overall quality of urban life.

7.8 Conclusion

This dissertation has undertaken an ambitious investigation into the integration of high-resolution remote sensing, advanced machine learning, and graph-based spatial modeling to address one of the most enduring challenges in urban environmental management: the scalable and generalizable prediction of urban noise. Through a carefully structured research design—spanning classical machine learning, convolutional neural networks (CNNs), and ultimately dual-branch graph neural networks (GNNs)—the study systematically advances the methodological frontiers of urban noise analytics.

The work makes several substantive contributions to both theory and practice. At its core, the research challenges the long-standing reliance on dense field measurements, detailed traffic and infrastructure data, and physics-heavy simulation models that dominate traditional noise mapping frameworks such as those used under the European Noise Directive (Kephalopoulos et al., 2012; Murphy & King, 2022). By demonstrating that satellite-derived morphological and spectral proxies—when carefully engineered and analyzed through spatially aware neural architectures—can approximate much of the spatial signal found in conventional datasets, the dissertation expands the toolkit available for urban environmental modeling, particularly in resource-limited contexts.

The adoption of a dual-branch GNN architecture is particularly significant. It allows for the explicit separation and recombination of fine-scale local morphological effects and broader land-use-driven contextual influences, reflecting the multi-scale and multi-causal nature of urban noise propagation. This capacity is further strengthened through domain-adaptive learning strategies, which enhance the model's ability to generalize across cities with divergent spatial, infrastructural, and environmental characteristics—a persistent challenge in environmental modeling (Zhang et al., 2022; Reichstein et al., 2019). The comparative analyses between CNN, single-branch GNN, and dual-branch GNN models provide robust empirical evidence that spatial graph architectures offer the best trade-off between predictive accuracy, computational efficiency, and cross-domain transferability.

However, the dissertation does not shy away from identifying its own limitations. Remote sensing imagery, while rich in spatial detail, provides only indirect proxies for the physical and dynamic processes governing urban soundscapes. The black-box nature of deep learning models continues to pose challenges for interpretability, stakeholder trust, and regulatory uptake, despite recent efforts in explainable AI (XAI) (Karpatne et al., 2017; Maxwell et al., 2018). Furthermore, the reliance on Euclidean distance-based graph construction, though computationally tractable, insufficiently captures the anisotropic, pathway-dependent nature of noise propagation in real urban environments.

These methodological constraints point toward clear directions for future research. Physics-informed machine learning (PIML) offers a promising route to embed acoustic principles directly into neural models, enhancing physical plausibility and interpretability (Willard et al., 2022). Multimodal data integration, combining satellite observations, sensor networks, traffic flows, and citizen-generated data, can enrich the

feature space and reduce reliance on any single data modality. Advances in semi-supervised, few-shot, and transfer learning methods will be critical for deploying these models in under-monitored or rapidly changing urban environments, extending their global relevance (Zhang et al., 2022).

Beyond scientific innovation, the dissertation carries important policy and practical implications. The development of scalable, automated noise prediction frameworks supports the democratization of environmental monitoring, reducing barriers to entry for cities with limited technical or financial resources. By aligning with smart city initiatives and digital twin infrastructures, these models can contribute to adaptive, data-driven urban governance, enabling real-time scenario testing, targeted interventions, and participatory planning (Batty, 2018; Allam & Dhunny, 2019; Somanath et al., 2024). Moreover, the emphasis on interpretability and physical grounding aligns with growing calls for trustworthy AI systems that are transparent, accountable, and socially responsible (Reichstein et al., 2019).

In summary, this dissertation presents a robust, scalable, and scientifically grounded framework for urban noise prediction, advancing the field from data-hungry, domain-specific models toward more flexible, transferable, and interpretable architectures. While important challenges remain, the convergence of multimodal data, physics-informed AI, and participatory system design holds immense promise for reshaping how cities understand, manage, and ultimately improve their acoustic environments. By bridging the gap between technical innovation and real-world application, the work lays a foundation for future research and practice at the intersection of environmental science, urban informatics, and machine learning.

Reference

- Adam, K.D.B.J., 2014. A method for stochastic optimization.
- Adeyemi, A., Ramoelo, A., Cho, M., Masemola, C.R., 2021. Spectral index to improve the extraction of built-up area from WorldView-2 imagery. *JARS* 15, 024510. <https://doi.org/10.1117/1.JRS.15.024510>
- Aggarwal, A., 2023. A Geospatial Approach to Monitoring Land Use and Land Cover Dynamics: A Review, in: Sahu, A.K., Meikap, B.C., Kudapa, V.K. (Eds.), *Energy Storage and Conservation*. Springer Nature, Singapore, pp. 63–71. https://doi.org/10.1007/978-981-99-2870-5_8
- Aguilera, I., Foraster, M., Basagaña, X., Corradi, E., Deltell, A., Morelli, X., Phuleria, H.C., Ragettli, M.S., Rivera, M., Thomasson, A., 2015. Application of land use regression modelling to assess the spatial distribution of road traffic noise in three European cities. *Journal of exposure science & environmental epidemiology* 25, 97–105.
- Ajibola, S., Cabral, P., 2024. A Systematic Literature Review and Bibliometric Analysis of Semantic Segmentation Models in Land Cover Mapping. *Remote Sensing* 16, 2222. <https://doi.org/10.3390/rs16122222>
- Aletta, F., Kang, J., Axelsson, Ö., 2016. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning* 149, 65–74.
- Aletta, F., Oberman, T., Mitchell, A., Tong, H., Kang, J., 2020. Assessing the changing urban sound environment during the COVID-19 lockdown period using short-term acoustic measurements. *Noise mapping* 7, 123–134.
- Allam, Z., Dhunny, Z.A., 2019. On big data, artificial intelligence and smart cities. *Cities* 89, 80–91.
- Almeida, C.R. de, Teodoro, A.C., Gonçalves, A., 2021. Study of the urban heat island (UHI) using remote sensing data/techniques: A systematic review. *Environments* 8, 105.
- Alon, U., Yahav, E., 2021. On the Bottleneck of Graph Neural Networks and its Practical Implications. <https://doi.org/10.48550/arXiv.2006.05205>
- Alvares-Sanches, T., Osborne, P.E., White, P., Jephcote, C., Gulliver, J., 2021a. Capturing the spatial and temporal variability of urban noise: do low-cost sensors offer a step towards higher resolution noise monitoring?
- Alvares-Sanches, T., Osborne, P.E., White, P.R., 2021b. Mobile surveys and machine learning can improve urban noise mapping: Beyond A-weighted measurements of exposure. *Science of the Total Environment* 775, 145600.
- Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Cooley, T.W., Jr, J.H.C., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A., Bernstein, L.S., Acharya, P.K., Miller, D.P., Lewis, P.E., 2002. MODTRAN4-based atmospheric correction algorithm: FLAASH (fast line-of-sight atmospheric analysis of spectral hypercubes), in: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*. Presented at the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII, SPIE, pp. 65–71. <https://doi.org/10.1117/12.478737>

- Anselin, L., 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis* 27, 93–115.
- Araujo, A., Norris, W., Sim, J., 2019. Computing receptive fields of convolutional neural networks. *Distill* 4, e21.
- Astuty, Y.I., Dimiyati, M., 2024. Prediction of land use/land cover change in Indonesia using the open source land cover dataset: a review. *Geodesy and Cartography* 50, 67–75.
- Avtar, R., Komolafe, A.A., Kouser, A., Singh, D., Yunus, A.P., Dou, J., Kumar, P., Gupta, R.D., Johnson, B.A., Minh, H.V.T., 2020. Assessing sustainable development prospects through remote sensing: A review. *Remote sensing applications: Society and environment* 20, 100402.
- Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer Normalization. <https://doi.org/10.48550/arXiv.1607.06450>
- Bai, L., Yao, L., Li, C., Wang, X., Wang, C., 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems* 33, 17804–17815.
- Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., Stansfeld, S., 2014. Auditory and non-auditory effects of noise on health. *The lancet* 383, 1325–1332.
- Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., Pascanu, R., 2018. Relational inductive biases, deep learning, and graph networks. <https://doi.org/10.48550/arXiv.1806.01261>
- Batty, M., 2018. Digital twins. *Environment and Planning B: Urban Analytics and City Science* 45, 817–820. <https://doi.org/10.1177/2399808318796416>
- Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing* 43, 480–491.
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems* 24.
- Bilot, T., El Madhoun, N., Al Agha, K., Zouaoui, A., 2024. A Survey on Malware Detection with Graph Representation Learning. *ACM Comput. Surv.* 56, 1–36. <https://doi.org/10.1145/3664649>
- Bischke, B., Helber, P., Folz, J., Borth, D., Dengel, A., 2019. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks, in: 2019 IEEE International Conference on Image Processing (ICIP). Presented at the 2019 IEEE International Conference on Image Processing (ICIP), pp. 1480–1484. <https://doi.org/10.1109/ICIP.2019.8803050>
- Bo, D., Wang, X., Shi, C., Shen, H., 2021. Beyond low-frequency information in graph convolutional networks, in: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 3950–3957.
- Botchkarev, A., 2019. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management* 14, 045–076.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>

- Bronstein, M.M., Bruna, J., Cohen, T., Veličković, P., 2021. Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges. <https://doi.org/10.48550/arXiv.2104.13478>
- Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 18–42.
- Calo, S., Bistaffa, F., Jonsson, A., Gómez, V., Viana, M., 2024. Spatial air quality prediction in urban areas via message passing. *Engineering Applications of Artificial Intelligence* 133, 108191.
- Can, A., Leclercq, L., Lelong, J., Defrance, J., 2009. Accounting for traffic dynamics improves noise assessment: Experimental evidence. *Applied Acoustics* 70, 821–829. <https://doi.org/10.1016/j.apacoust.2008.09.020>
- Carniel, H., Pozzer, T., Holtz, M., Akkerman, D., 2018. How to deal with noise map calculation models in Brazil. *Euronoise*.
- Caruana, R., 1997. Multitask Learning. *Machine Learning* 28, 41–75. <https://doi.org/10.1023/A:1007379606734>
- Casey, J.A., Morello-Frosch, R., Mennitt, D.J., Fristrup, K., Ogburn, E.L., James, P., 2017. Race/Ethnicity, Socioeconomic Status, Residential Segregation, and Spatial Variation in Noise Exposure in the Contiguous United States. *Environ Health Perspect* 125, 077017. <https://doi.org/10.1289/EHP898>
- Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote sensing of environment* 113, 893–903.
- Chen, G., Li, S., Knibbs, L.D., Hamm, N.A., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment* 636, 52–60.
- Chen, K., Liu, J., Wan, R., Lee, V.H.-F., Vardhanabhuti, V., Yan, H., Li, H., 2024. Unsupervised domain adaptation for low-dose ct reconstruction via bayesian uncertainty alignment. *IEEE Transactions on Neural Networks and Learning Systems*.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations, in: *International Conference on Machine Learning*. PmlR, pp. 1597–1607.
- Chen, Z., Fei, T., Xiao, J., Huang, J., Jia, D., Bian, M., 2025. Estimating urban noise levels from Multi-Scale and Multi-Spectral remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation* 143, 104818.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment* 219, 145–161.

- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach Learn* 20, 273–297.
<https://doi.org/10.1007/BF00994018>
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory* 13, 21–27.
- Cushman, S.A., Kaszta, Z.M., Burns, P., Hakkenberg, C.R., Jantz, P., Macdonald, D.W., Brodie, J.F., Deith, M.C.M., Goetz, S., 2024. Simulating multi-scale optimization and variable selection in species distribution modeling. *Ecological Informatics* 83, 102832. <https://doi.org/10.1016/j.ecoinf.2024.102832>
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology* 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dauphin, Y.N., Fan, A., Auli, M., Grangier, D., 2017. Language modeling with gated convolutional networks, in: *International Conference on Machine Learning*. PMLR, pp. 933–941.
- de Kluizenaar, Y., Janssen, S.A., van Lenthe, F.J., Miedema, H.M., Mackenbach, J.P., 2009. Long-term road traffic noise exposure is associated with an increase in morning tiredness. *The Journal of the Acoustical Society of America* 126, 626–633.
- DEFRA (Department for Environment, Food and Rural Affairs), 2020a. Road noise — Lden — England Round 3.
- DEFRA (Department for Environment, Food and Rural Affairs), 2020b. Road noise — Lnight — England Round 3.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, pp. 248–255.
- Department for Environment, Food & Rural Affairs (DEFRA), 2023. Environmental Noise Mapping England: Strategic Noise Mapping Round 4. DEFRA, London.
- Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., 2019. An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environment international* 130, 104909.
- Du, Y., Teillet, P.M., Cihlar, J., 2002. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote sensing of Environment* 82, 123–134.
- Dutilleul, G., Defrance, J., Ecotière, D., Gauvreau, B., Bérengier, M., Besnard, F., Duc, E.L., 2010. NMPB-routes-2008: The revision of the french method for road traffic noise prediction. *Acta Acustica united with Acustica* 96, 452–462.
- EEA (European Environment Agency), 2018. Urban Atlas 2012.
- Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697.
<https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- ESA (European Space Agency), 2014. WorldView-2 European Cities Dataset.
- European Commission, 2015. CNOSSOS-EU: Common Noise Assessment Methods in Europe.
- European Commission, n.d. Environmental Noise Directive 2002/49/EC — Fourth round of strategic noise mapping (2021). Brussels.

- European Commission (EC), 2002. Directive 2002/49/EC relating to the assessment and management of environmental noise. Official Journal of the European Communities.
- European Environment Agency (EEA), 2016. Urban Atlas Land Cover/Land Use 2012 (vector), Europe, 6-yearly — Copernicus Land Monitoring Service [WWW Document]. Copernicus Land Monitoring Service. URL <https://land.copernicus.eu/en/products/urban-atlas/urban-atlas-2012> (accessed 10.13.25).
- Farahmand, H., Xu, Y., Mostafavi, A., 2023. A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features. *Scientific Reports* 13, 6768.
- Feng, L., Gui, D., Han, S., Qiu, T., Wang, Y., 2024. Integrating optical and SAR time series images for unsupervised domain adaptive crop mapping. *Remote Sensing* 16, 1464.
- Forehead, H., Huynh, N., 2018. Review of modelling air pollution from traffic at street-level-The state of the science. *Environmental Pollution* 241, 775–786.
- Frame, J.M., Nair, T., Sunkara, V., Popien, P., Chakrabarti, S., Anderson, T., Leach, N.R., Doyle, C., Thomas, M., Tellman, B., 2024. Rapid Inundation Mapping Using the US National Water Model, Satellite Observations, and a Convolutional Neural Network. *Geophysical Research Letters* 51, e2024GL109424. <https://doi.org/10.1029/2024GL109424>
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189–1232.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. *Journal of machine learning research* 17, 1–35.
- Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gao, L., Song, W., Dai, J., Chen, Y., 2019. Road Extraction from High-Resolution Remote Sensing Imagery Using Refined Deep Residual Convolutional Neural Network. *Remote Sensing* 11, 552. <https://doi.org/10.3390/rs11050552>
- Ge, L., Wu, K., Zeng, Y., Chang, F., Wang, Y., Li, S., 2021. Multi-scale spatiotemporal graph convolution network for air quality prediction. *Appl Intell* 51, 3491–3505. <https://doi.org/10.1007/s10489-020-02054-y>
- Genaro, N., Torija, A., Ramos, A., Requena, I., Ruiz, D.P., Zamorano, M., 2009. Modeling environmental noise using artificial neural networks, in: 2009 Ninth International Conference on Intelligent Systems Design and Applications. IEEE, pp. 215–219.
- Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E., 2017. Neural message passing for quantum chemistry, in: International Conference on Machine Learning. Pmlr, pp. 1263–1272.
- Gong, P., Wang, J., Yu, Le, Zhao, Yongchao, Zhao, Yuanyuan, Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, Han, Zhu, P., Zhao, Z., Zhang, Haiying, Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, Liang, Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, Jin, Chen, Jun, 2013.

- Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* 34, 2607–2654. <https://doi.org/10.1080/01431161.2012.748992>
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69, 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press Cambridge, MA, USA.
- Grandvalet, Y., Bengio, Y., 2004. Semi-supervised Learning by Entropy Minimization, in: *Advances in Neural Information Processing Systems*. MIT Press.
- Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., 2020. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems* 33, 21271–21284.
- Guedes, I.C.M., Bertoli, S.R., Zannin, P.H., 2011. Influence of urban shapes on environmental noise: a case study in Aracaju—Brazil. *Science of the Total Environment* 412, 66–76.
- Gulrajani, I., Lopez-Paz, D., 2020. In Search of Lost Domain Generalization. <https://doi.org/10.48550/arXiv.2007.01434>
- Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 3, 1157–1182.
- Ha, D., Schmidhuber, J., 2018. World models. *arXiv preprint arXiv:1803.10122* 2.
- Haklay, M., 2013. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation, in: Sui, D., Elwood, S., Goodchild, M. (Eds.), *Crowdsourcing Geographic Knowledge*. Springer Netherlands, Dordrecht, pp. 105–122. https://doi.org/10.1007/978-94-007-4587-2_7
- Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987. Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 532–550. <https://doi.org/10.1109/TPAMI.1987.4767941>
- Hashemi, R., Darabi, H., 2022. The Review of Ecological Network Indicators in Graph Theory Context: 2014–2021. *Int J Environ Res* 16, 24. <https://doi.org/10.1007/s41742-022-00404-x>
- Hashemi, R., Darabi, H., Hashemi, M., Wang, J., 2024. Graph theory in ecological network analysis: A systematic review for connectivity assessment. *Journal of Cleaner Production* 472, 143504.
- He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9729–9738.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. Presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.

- Helmer, E.H., Ruefenacht, B., 2005. Cloud-free satellite image mosaics with regression trees and histogram matching. *Photogrammetric Engineering & Remote Sensing* 71, 1079–1089.
- Herold, M., Scepan, J., Clarke, K.C., 2002. The Use of Remote Sensing and Landscape Metrics to Describe Structures and Changes in Urban Land Uses. *Environ Plan A* 34, 1443–1458. <https://doi.org/10.1068/a3496>
- Hornikx, M., 2016. Ten questions concerning computational urban acoustics. *Building and Environment* 106, 409–421.
- Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM_{2.5} Concentrations in the Conterminous United States Using the Random Forest Approach. *Environ. Sci. Technol.* 51, 6936–6944. <https://doi.org/10.1021/acs.est.7b01210>
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely Connected Convolutional Networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
- Huang, J., Fei, T., Kang, Y., Li, J., Liu, Z., Wu, G., 2024. Estimating urban noise along road network from street view imagery. *International Journal of Geographical Information Science* 38, 128–155. <https://doi.org/10.1080/13658816.2023.2274475>
- Huang, J., Gretton, A., Borgwardt, K., Schölkopf, B., Smola, A., 2006. Correcting sample selection bias by unlabeled data. *Advances in neural information processing systems* 19.
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., Liu, Y., 2018. Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environmental pollution* 242, 675–683.
- Huotari, M., Malhi, A., Främling, K., 2024. Machine learning applications for smart building energy utilization: a survey. *Archives of Computational Methods in Engineering* 31, 2537–2556.
- International Organization for Standardization (ISO), 1996. *Acoustics-Attenuation of Sound During Propagation Outdoors: Part 2: General Method of Calculation*.
- International Organization for Standardization (ISO), n.d. *ISO 9613-2: Acoustics — Attenuation of sound during propagation outdoors — Part 2: General method of calculation (Standard No. ISO 9613-2)*. ISO.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: *International Conference on Machine Learning*. pmlr, pp. 448–456.
- Jamali, A., Roy, S.K., Beni, L.H., Pradhan, B., Li, J., Ghamisi, P., 2024. Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. *International Journal of Applied Earth Observation and Geoinformation* 127, 103662.
- Jensen, J.R., 2009. *Remote sensing of the environment: An earth resource perspective 2/e*. Pearson Education India.
- Jiang, J., Chen, C., Zhou, Y., Berretti, S., Liu, L., Pei, Q., Zhou, J., Wan, S., 2024. Heterogeneous dynamic graph convolutional networks for enhanced spatiotemporal flood forecasting by remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17, 3108–3122.

- Jiang, W., Luo, J., 2022. Graph neural network for traffic forecasting: A survey. *Expert systems with applications* 207, 117921.
- Jo, J., Lee, S., Hwang, S.J., 2022. Score-based generative modeling of graphs via the system of stochastic differential equations, in: *International Conference on Machine Learning*. PMLR, pp. 10362–10383.
- Kadhim, N., Mourshed, M., Bray, M., 2016. Advances in remote sensing applications for urban sustainability. *Euro-Mediterr J Environ Integr* 1, 7.
<https://doi.org/10.1007/s41207-016-0007-4>
- Kage, P., Rothenberger, J.C., Andreadis, P., Diochnos, D.I., 2025. A Review of Pseudo-Labeling for Computer Vision. <https://doi.org/10.48550/arXiv.2408.07221>
- Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning* 14, 1–210.
- Kang, J., 2006. *Urban sound environment*. CRC Press.
- Kang, J., Aletta, F., Gjestland, T.T., Brown, L.A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A., 2016. Ten questions on the soundscapes of the built environment. *Building and environment* 108, 284–294.
- Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. *Nature Reviews Physics* 3, 422–440.
- Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering* 29, 2318–2331.
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing* 173, 24–49.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30.
- Kephalopoulos, S., Paviotti, M., Anfosso-Lédée, F., 2021. CNOSSOS-EU Technical Report 2021 Update (Standard). Luxembourg: Publications Office of the European Union.
- Kephalopoulos, S., Paviotti, M., Anfosso-Lédée, F., 2012. Common noise assessment methods in Europe (CNOSSOS-EU). *Common noise assessment methods in Europe (CNOSSOS-EU)*.
- Khan, J., Ketzel, M., Jensen, S.S., Gulliver, J., Thysell, E., Hertel, O., 2021. Comparison of Road Traffic Noise prediction models: CNOSSOS-EU, Nord2000 and TRANEX. *Environmental Pollution* 270, 116240.
- Khan, J., Ketzel, M., Kakosimos, K., Sørensen, M., Jensen, S.S., 2018. Road traffic air and noise pollution exposure assessment—A review of tools and techniques. *Science of the total environment* 634, 661–676.
- Kipf, T.N., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Konapala, G., Kumar, S.V., Ahmad, S.K., 2021. Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 180, 163–173.

- Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017a. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 14, 778–782.
- Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017b. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters* 14, 778–782.
<https://doi.org/10.1109/LGRS.2017.2681128>
- Kwak, G.-H., Park, N.-W., 2022. Unsupervised Domain Adaptation with Adversarial Self-Training for Crop Classification Using Remote Sensing Images. *Remote Sensing* 14, 4639. <https://doi.org/10.3390/rs14184639>
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *nature* 521, 436–444.
- Lee, D.-H., 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: *Workshop on Challenges in Representation Learning, ICML*. Atlanta, p. 896.
- Lee, J., 2024. Spatiotemporal gap-filling of NASA Deep Blue aerosol optical depth over CONUS using the UNet 3+ architecture.
- Lei, B., Liu, P., Milojevic-Dupont, N., Biljecki, F., 2024. Predicting building characteristics at urban scale using graph neural networks and street-level context. *Computers, Environment and Urban Systems* 111, 102129.
- Li, B., Tao, S., Dawson, R.W., 2002. Evaluation and analysis of traffic noise from the main urban roads in Beijing. *Applied acoustics* 63, 1137–1142.
- Li, F., Wang, H., Du, C., Lan, Z., Yu, F., Rong, Y., 2024. Probabilistic Prediction Model for Expressway Traffic Noise Based on Short-Term Monitoring Data. *Sustainability* 16, 6841.
- Li, J., Heap, A.D., 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software* 53, 173–189.
- Li, P., Zhang, T., Jin, Y., 2023. A spatio-temporal graph convolutional network for air quality prediction. *Sustainability* 15, 7624.
- Li, Q., Han, Z., Wu, X., 2018. Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 32. <https://doi.org/10.1609/aaai.v32i1.11604>
- Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., 2016. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS journal of Photogrammetry and Remote Sensing* 115, 119–133.
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental pollution* 231, 997–1004.
- Li, X., Zhou, W., Ouyang, Z., 2013. Forty years of urban expansion in Beijing: What is the relative importance of physical, socioeconomic, and neighborhood factors? *Applied Geography* 38, 1–10.
- Li, Y., Jin, H., Dong, S., Yang, B., Chen, X., 2021. Pseudo-labeling optimization based ensemble semi-supervised soft sensor in the process industry. *Sensors* 21, 8471.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting.
<https://doi.org/10.48550/arXiv.1707.01926>

- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R news* 2, 18–22.
- Ling, J., Lan, Y., Huang, X., Yang, X., 2024. A Multi-Scale Residual Graph Convolution Network with hierarchical attention for predicting traffic flow in urban mobility. *Complex Intell. Syst.* 10, 3305–3317. <https://doi.org/10.1007/s40747-023-01324-9>
- Liu, H., Han, Q., Sun, H., Sheng, J., Yang, Z., 2023. Spatiotemporal adaptive attention graph convolution network for city-level air quality prediction. *Scientific Reports* 13, 13335.
- Liu, L., Han, B., Tan, D., Wu, D., Shu, C., 2023. The Value of Ecosystem Traffic Noise Reduction Service Provided by Urban Green Belts: A Case Study of Shenzhen. *Land* 12, 786. <https://doi.org/10.3390/land12040786>
- Liu, Y., Oiamo, T., Rainham, D., Chen, H., Hatzopoulou, M., Brook, J.R., Davies, H., Goudreau, S., Smargiassi, A., 2021. Integrating random forests and propagation models for high-resolution noise mapping. *Environmental research* 195, 110905.
- Liu, Z.-Q., Tang, P., Zhang, W., Zhang, Z., 2022. CNN-enhanced heterogeneous graph convolutional network: Inferring land use from land cover with a case study of park segmentation. *Remote Sensing* 14, 5027.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.
- Long, M., Cao, Y., Wang, J., Jordan, M., 2015. Learning transferable features with deep adaptation networks, in: International Conference on Machine Learning. PMLR, pp. 97–105.
- Lu, G.Y., Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences* 34, 1044–1055. <https://doi.org/10.1016/j.cageo.2007.07.010>
- Lu, S., Guo, J., Zimmer-Dauphinee, J.R., Nieusma, J.M., Wang, X., Wernke, S.A., Huo, Y., 2025. Vision foundation models in remote sensing: A survey. *IEEE Geoscience and Remote Sensing Magazine*.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30.
- Luo, W., Li, Y., Urtasun, R., Zemel, R., 2016. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems* 29.
- Maisonneuve, N., Stevens, M., Niessen, M.E., Hanappe, P., Steels, L., 2009. Citizen noise pollution monitoring.
- Maisonneuve, N., Stevens, M., Ochab, B., 2010. Participatory noise pollution monitoring using mobile phones. *IP* 15, 51–71. <https://doi.org/10.3233/IP-2010-0200>
- Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D., Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9414–9423.
- Margaritis, E., Kang, J., Filipan, K., Botteldooren, D., 2018. The influence of vegetation and surrounding traffic noise parameters on the sound environment

- of urban parks. *Applied Geography* 94, 199–212.
<https://doi.org/10.1016/j.apgeog.2018.02.017>
- Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2015. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters* 13, 105–109.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* 39, 2784–2817.
<https://doi.org/10.1080/01431161.2018.1433343>
- Maxwell, J.A., 2021. Why qualitative methods are necessary for generalization. *Qualitative Psychology* 8, 111.
- McLoughlin, I., Zhang, H., Xie, Z., Song, Y., Xiao, W., 2015. Robust sound event classification using deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23, 540–552.
- Mesaros, A., Heittola, T., Virtanen, T., 2016. TUT database for acoustic scene classification and sound event detection, in: 2016 24th European Signal Processing Conference (EUSIPCO). IEEE, pp. 1128–1132.
- Mirpulatov, I., Illarionova, S., Shadrin, D., Burnaev, E., 2023. Pseudo-labeling approach for land cover classification through remote sensing observations with noisy labels. *IEEE Access* 11, 82570–82583.
- Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
- Mosavi, A., Ozturk, P., Chau, K., 2018. Flood prediction using machine learning models: Literature review. *Water* 10, 1536.
- Motohka, T., Nasahara, K.N., Oguma, H., Tsuchida, S., 2010. Applicability of Green-Red Vegetation Index for Remote Sensing of Vegetation Phenology. *Remote Sensing* 2, 2369–2387. <https://doi.org/10.3390/rs2102369>
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS journal of photogrammetry and remote sensing* 66, 247–259.
- Münzel, T., Daiber, A., Steven, S., Tran, L.P., Ullmann, E., Kossmann, S., Schmidt, F.P., Oelze, M., Xia, N., Li, H., 2017a. Effects of noise on vascular function, oxidative stress, and inflammation: mechanistic insight from studies in mice. *European heart journal* 38, 2838–2849.
- Münzel, T., Sørensen, M., Gori, T., Schmidt, F.P., Rao, X., Brook, F.R., Chen, L.C., Brook, R.D., Rajagopalan, S., 2017b. Environmental stressors and cardio-metabolic disease: part II—mechanistic insights. *European heart journal* 38, 557–564.
- Murphy, E., King, E.A., 2014. *Environmental noise pollution: Noise mapping, public health, and policy*. Elsevier.
- Murphy, E., King, E.A., 2010. Strategic environmental noise mapping: Methodological issues concerning the implementation of the EU Environmental Noise Directive and their policy implications. *Environment international* 36, 290–298.
- Muthukumar, P., Cocom, E., Nagrecha, K., Comer, D., Burga, I., Taub, J., Calvert, C.F., Holm, J., Pourhomayoun, M., 2022. Predicting PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Qual Atmos Health* 15, 1221–1234.
<https://doi.org/10.1007/s11869-021-01126-3>

- Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment* 115, 1145–1161.
- Ni, J., Chen, Y., Gu, Y., Fang, X., Shi, P., 2022. An improved hybrid transfer learning-based deep learning model for PM_{2.5} concentration prediction. *Applied Sciences* 12, 3597.
- Oono, K., Suzuki, T., 2021. Graph Neural Networks Exponentially Lose Expressive Power for Node Classification. <https://doi.org/10.48550/arXiv.1905.10947>
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M., 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* 156, 94–104.
- Ouyang, X., Yang, Y., Zhang, Y., Zhou, W., Wan, J., Du, S., 2023. Domain adversarial graph neural network with cross-city graph structure learning for traffic prediction. *Knowledge-Based Systems* 278, 110885.
- Pan, S.J., Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 1345–1359.
- Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 158, 279–317.
- Park, H.J., Park, B., Lee, S.S., 2020. Radiomics and deep learning: hepatic applications. *Korean journal of radiology* 21, 387–401.
- Patel, V.M., Gopalan, R., Li, R., Chellappa, R., 2015. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine* 32, 53–69.
- Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1226–1238. <https://doi.org/10.1109/TPAMI.2005.159>
- Persello, C., Bruzzone, L., 2014. Active and Semisupervised Learning for the Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 52, 6937–6956. <https://doi.org/10.1109/TGRS.2014.2305805>
- Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A.J., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., Syrris, V., 2016. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Publications Office of the European Union Luxembourg.
- Pope, P.E., Kolouri, S., Rostami, M., Martin, C.E., Hoffmann, H., 2019. Explainability methods for graph convolutional neural networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10772–10781.
- Popescu, D.I., 2023. COMPARATIVE STUDY ON NOISE MAPPING FOR EUROPEAN URBAN AGGLOMERATIONS UNDER THE ENVIRONMENTAL NOISE DIRECTIVE (END).
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems* 31.
- Quartieri, J., Mastorakis, N.E., Iannone, G., Guarnaccia, C., D'Ambrosio, S., Troisi, A., Lenza, T.L.L., 2009. A review of traffic noise predictive models, in: *Recent Advances in Applied and Theoretical Mechanics*, 5th WSEAS International

- Conference on Applied and Theoretical Mechanics (MECHANICS'09) Puerto De La Cruz, Tenerife, Canary Islands, Spain December. pp. 14–16.
- Raimbault, J., 2021. A multiscale model of urban morphogenesis. <https://doi.org/10.48550/arXiv.2103.17241>
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 378, 686–707.
- Rajakumara, H.N., Gowda, R.M., 2008. Road traffic noise prediction models: a review. *International Journal of Sustainable Development and Planning* 3, 257–271.
- Ranasinghe, D., Lee, E.S., Zhu, Y., Frausto-Vicencio, I., Choi, W., Sun, W., Mara, S., Seibt, U., Paulson, S.E., 2019. Effectiveness of vegetation and sound wall-vegetation combination barriers on pollution dispersion from freeways under early morning conditions. *Science of The Total Environment* 658, 1549–1558.
- Ranpise, R.B., Tandel, B.N., 2022. Urban road traffic noise monitoring, mapping, modelling, and mitigation: A thematic review. *Noise Mapping* 9, 48–66. <https://doi.org/10.1515/noise-2022-0004>
- Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-Guided Deep Learning Predictions of Lake Water Temperature. *Water Resources Research* 55, 9173–9190. <https://doi.org/10.1029/2019WR024922>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204.
- Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M., 2021. In Defense of Pseudo-Labeling: An Uncertainty-Aware Pseudo-label Selection Framework for Semi-Supervised Learning. <https://doi.org/10.48550/arXiv.2101.06329>
- Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2023. Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* 55, 1–96. <https://doi.org/10.1145/3485128>
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
- Roy, D.P., Li, J., Zhang, H.K., Yan, L., Huang, H., Li, Z., 2017. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment* 199, 25–38.
- Sabat-Tomala, A., Raczko, E., Zagajewski, B., 2020. Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sensing* 12, 516. <https://doi.org/10.3390/rs12030516>

- Sahili, Z.A., Awad, M., 2023. Spatio-Temporal Graph Neural Networks: A Survey. <https://doi.org/10.48550/arXiv.2301.10569>
- Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K., 2019. Semi-Supervised Domain Adaptation via Minimax Entropy. Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8050–8058.
- Salamon, J., Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters* 24, 279–283.
- Salomons, E.M., 2001. Computational atmospheric acoustics. Springer Science & Business Media.
- Salomons, E.M., Pont, M.B., 2012. Urban traffic noise and the relation to urban density, form, and traffic elasticity. *Landscape and Urban Planning* 108, 2–16.
- Samad, A., Garuda, S., Vogt, U., Yang, B., 2023. Air pollution prediction using machine learning techniques—An approach to replace existing monitoring stations with virtual monitoring stations. *Atmospheric Environment* 310, 119987.
- Sameen, M.I., Pradhan, B., 2016. novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique. In *IOP Conference Series: Earth and Environmental Science* 37, 012006.
- Sarkar, A., Nandi, U., Kumar Sarkar, N., Changdar, C., Paul, B., 2024. Deep Learning Based Hyperspectral Image Classification: A Review For Future Enhancement. *International Journal of Computing and Digital Systems* 16, 419–435.
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. *IEEE transactions on neural networks* 20, 61–80.
- Schott, J.R., Salvaggio, C., Volchok, W.J., 1988. Radiometric scene normalization using pseudoinvariant features. *Remote sensing of Environment* 26, 1–16.
- Schowengerdt, R.A., 2006. Remote Sensing: Models and Methods for Image Processing. Elsevier.
- Semper, M., Curado, M., Oliver, J.L., Vicent, J.F., 2025. Noise Pollution Prediction in a Densely Populated City Using a Spatio-Temporal Deep Learning Approach. *Applied Sciences* 15, 5576.
- Sener, O., Savarese, S., 2018. Active Learning for Convolutional Neural Networks: A Core-Set Approach. <https://doi.org/10.48550/arXiv.1708.00489>
- Seto, K.C., Güneralp, B., Hutya, L.R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16083–16088. <https://doi.org/10.1073/pnas.1211658109>
- Shannon, G., McKenna, M.F., Angeloni, L.M., Crooks, K.R., Fristrup, K.M., Brown, E., Warner, K.A., Nelson, M.D., White, C., Briggs, J., McFarland, S., Wittemyer, G., 2016. A synthesis of two decades of research documenting the effects of noise on wildlife. *Biological Reviews* 91, 982–1005. <https://doi.org/10.1111/brv.12207>
- Shao, S., Henrique Ribeiro, P., Ramirez, C.M., Moore, J.H., 2024. A review of feature selection strategies utilizing graph data structures and Knowledge Graphs. *Briefings in Bioinformatics* 25.
- Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data, in: *Proceedings of the 1968 23rd ACM National Conference, ACM '68. Association for Computing Machinery, New York, NY, USA*, pp. 517–524. <https://doi.org/10.1145/800186.810616>

- Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.-L., 2020. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 596–608.
- Somanath, S., Naserentin, V., Eleftheriou, O., Sjölie, D., Wästberg, B.S., Logg, A., 2024. Towards urban digital twins: A workflow for procedural visualization using geospatial data. *Remote Sensing* 16, 1939.
- Song, L., Kwan, M.-P., Liu, Y., 2025. Examining the complex and cumulative effects of environmental exposures on noise perception through interpretable spatio-temporal graph convolutional networks. *Environment International* 203, 109731. <https://doi.org/10.1016/j.envint.2025.109731>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1929–1958.
- Staab, J., Droin, A., Weigand, M., Dallavalle, M., Wolf, K., Schady, A., Lakes, T., Wurm, M., Taubenböck, H., 2025. Pixels, chisels and contours-technical variations in European road traffic noise exposure maps. *Journal of Environmental Management* 385, 125475.
- Staab, J., Schady, A., Weigand, M., Lakes, T., Taubenböck, H., 2022. Predicting traffic noise using land-use regression—a scalable approach. *J Expo Sci Environ Epidemiol* 32, 232–243. <https://doi.org/10.1038/s41370-021-00355-z>
- Staab, J., Wurm, M., Taubenböck, H., 2024. Noise - Road Traffic Lden (2002/49/EC) - Germany, 2017. <https://doi.org/10.15489/A6WG11LRUB77>
- Stansfeld, S., Clark, C., 2015. Health effects of noise exposure in children. *Current environmental health reports* 2, 171–178.
- Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International Conference on Machine Learning*. PMLR, pp. 6105–6114.
- Tang, D., Cao, X., Hou, X., Jiang, Z., Liu, J., Meng, D., 2024. Crs-diff: Controllable remote sensing image generation with diffusion model. *IEEE Transactions on Geoscience and Remote Sensing*.
- Teng, J., Zhang, C., Gong, H., Liu, C., 2024. Machine learning-based urban noise appropriateness evaluation method and driving factor analysis. *Plos one* 19, e0311571.
- Terroso-Saenz, F., Morales-García, J., Muñoz, A., 2024. Nationwide Air Pollution Forecasting with Heterogeneous Graph Neural Networks. *ACM Trans. Intell. Syst. Technol.* 15, 1–19. <https://doi.org/10.1145/3637492>
- Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE geoscience and remote sensing magazine* 4, 41–57.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution* 18, 306–314.
- Umar, I.K., Adamu, M., Mostafa, N., Riaz, M.S., Haruna, S.I., Hamza, M.F., Ahmed, O.S., Azab, M., 2024. The state-of-the-art in the application of artificial intelligence-based models for traffic noise prediction: a bibliographic overview.

- Cogent Engineering 11, 2297508.
<https://doi.org/10.1080/23311916.2023.2297508>
- Vali, A., Comai, S., Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing* 12, 2495.
- Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., Winker, D.M., 2016. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environ. Sci. Technol.* 50, 3762–3772. <https://doi.org/10.1021/acs.est.5b05833>
- Van Renterghem, T., Le Bescond, V., Dekoninck, L., Botteldooren, D., 2023. Advanced Noise Indicator Mapping Relying on a City Microphone Network. *Sensors* 23, 5865. <https://doi.org/10.3390/s23135865>
- Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE transactions on neural networks* 10, 988–999.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems* 30.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph Attention Networks. <https://doi.org/10.48550/arXiv.1710.10903>
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote sensing of environment* 185, 46–56.
- Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., Frossard, P., 2023. DiGress: Discrete Denoising diffusion for graph generation. <https://doi.org/10.48550/arXiv.2209.14734>
- Von Rueden, L., Mayer, S., Sifa, R., Bauckhage, C., Garcke, J., 2020. Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions, in: Berthold, M.R., Feelders, A., Krempel, G. (Eds.), *Advances in Intelligent Data Analysis XVIII, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 548–560. https://doi.org/10.1007/978-3-030-44584-3_43
- Wang, A., Xu, J., Tu, R., Saleh, M., Hatzopoulou, M., 2020. Potential of machine learning for prediction of traffic related air pollution. *Transportation Research Part D: Transport and Environment* 88, 102599.
- Wang, J., H. Q. Ding, C., Chen, S., He, C., Luo, B., 2020. Semi-Supervised Remote Sensing Image Semantic Segmentation via Consistency Regularization and Average Update of Pseudo-Label. *Remote Sensing* 12, 3603. <https://doi.org/10.3390/rs12213603>
- Wang, Jindong, Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., Yu, P.S., 2022. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering* 35, 8052–8072.
- Wang, Junjue, Zheng, Z., Ma, A., Lu, X., Zhong, Y., 2022. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. <https://doi.org/10.48550/arXiv.2110.08733>

- Wang, X., Zhang, Suo, Chen, Y., He, L., Ren, Y., Zhang, Z., Li, J., Zhang, Shiqing, 2024. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD–GAT–BiLSTM. *Scientific Reports* 14, 17841.
- Wang, Y., Feng, L., Zhang, Z., Tian, F., 2023. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 199, 102–117. <https://doi.org/10.1016/j.isprsjprs.2023.04.002>
- Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. *J Big Data* 3, 9. <https://doi.org/10.1186/s40537-016-0043-6>
- Weng, Q., 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote sensing of Environment* 117, 34–49.
- Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2020. Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919* 1, 1–34.
- Wilson, G., Cook, D.J., 2020. A Survey of Unsupervised Deep Domain Adaptation. *ACM Trans. Intell. Syst. Technol.* 11, 1–46. <https://doi.org/10.1145/3400066>
- Wolf, A.F., 2012. Using WorldView-2 Vis-NIR multispectral imagery to support land mapping and feature extraction using normalized difference index ratios, in: *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*. Presented at the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, SPIE, pp. 188–195. <https://doi.org/10.1117/12.917717>
- World Health Organization (WHO), 2018. Environmental noise guidelines for the European region: Executive summary, in: *Environmental Noise Guidelines for the European Region: Executive Summary*.
- Wu, L., Cui, P., Pei, J., Zhao, L., Guo, X., 2022. Graph Neural Networks: Foundation, Frontiers and Applications, in: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, Washington DC USA, pp. 4840–4841. <https://doi.org/10.1145/3534678.3542609>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S., 2020. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment* 122, 2–10.
- Xie, D., Liu, Y., Chen, J., 2011. Mapping Urban Environmental Noise: A Land Use Regression Method. *Environ. Sci. Technol.* 45, 7358–7364. <https://doi.org/10.1021/es200785x>
- Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., Jegelka, S., 2018. Representation learning on graphs with jumping knowledge networks, in: *International Conference on Machine Learning*. pmlr, pp. 5453–5462.
- Xu, Y., Wei, F., Sun, X., Yang, C., Shen, Y., Dai, B., Zhou, B., Lin, S., 2022. Cross-model pseudo-labeling for semi-supervised action recognition, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2959–2968.

- Yadav, N., 2023. DeepAQ: Unsupervised Domain Adaptation for Air-Quality Mapping Using High-Resolution Satellite Imagery.
- Yang, J., Peng, Z., Sun, J., Chen, Z., Niu, X., Xu, H., Ho, K.-F., Cao, J., Shen, Z., 2024. A review on advancements in atmospheric microplastics research: The pivotal role of machine learning. *Science of the Total Environment* 945, 173966.
- Yin, X., Fallah-Shorshani, M., McConnell, R., Fruin, S., Franklin, M., 2020. Predicting Fine Spatial Scale Traffic Noise Using Mobile Measurements and Machine Learning. *Environ. Sci. Technol.* 54, 12860–12869.
<https://doi.org/10.1021/acs.est.0c01987>
- Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J., 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems* 32.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, in: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. pp. 3634–3640.
<https://doi.org/10.24963/ijcai.2018/505>
- Yu, F., Koltun, V., 2016. Multi-Scale Context Aggregation by Dilated Convolutions.
<https://doi.org/10.48550/arXiv.1511.07122>
- Yu, W., Jang, J.-C., Zhu, Y., Peng, J., Yang, W., Li, K., 2024. Enhanced Estimation of Traffic Noise Levels Using Minute-Level Traffic Flow Data through Convolutional Neural Network. *Sustainability* 16, 6088. <https://doi.org/10.3390/su16146088>
- Yuan, D., Elvidge, C.D., 1996. Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing* 51, 117–126.
- Yuan, H., Yu, H., Wang, J., Li, K., Ji, S., 2021. On explainability of graph neural networks via subgraph explorations, in: *International Conference on Machine Learning*. PMLR, pp. 12241–12252.
- Yuan, Z., Kerckhoffs, J., Li, H., Khan, J., Hoek, G., Vermeulen, R., 2024. Hyperlocal Air Pollution Mapping: A Scalable Transfer Learning LUR Approach for Mobile Monitoring. *Environ. Sci. Technol.* 58, 14372–14383.
<https://doi.org/10.1021/acs.est.4c06144>
- Zannin, P.H.T., de Sant’Ana, D.Q., 2011. Noise mapping at different stages of a freeway redevelopment project—A case study in Brazil. *Applied Acoustics* 72, 479–486.
- Zhang, K., Wen, Q., Zhang, C., Cai, R., Jin, M., Liu, Y., Zhang, J.Y., Liang, Y., Pang, G., Song, D., 2024. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *IEEE transactions on pattern analysis and machine intelligence* 46, 6775–6794.
- Zhang, Y., Zhao, H., Li, Y., Long, Y., Liang, W., 2023. Predicting highly dynamic traffic noise using rotating mobile monitoring and machine learning method. *Environmental research* 229, 115896.
- Zhao, T., Liang, X., Tu, W., Huang, Z., Biljecki, F., 2023. Sensing urban soundscapes from street view imagery. *Computers, Environment and Urban Systems* 99, 101915.
- Zheng, Y., Capra, L., Wolfson, O., Yang, H., 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. Intell. Syst. Technol.* 5, 1–55.
<https://doi.org/10.1145/2629592>

- Zheng, Y., Liu, F., Hsieh, H.-P., 2013. U-Air: when urban air quality inference meets big data, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD' 13: The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Chicago Illinois USA, pp. 1436–1444.
<https://doi.org/10.1145/2487575.2488188>
- Zhou, D., Bousquet, O., Lal, T., Weston, J., Schölkopf, B., 2003. Learning with local and global consistency. *Advances in neural information processing systems* 16.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. *AI open* 1, 57–81.
- Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2022. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence* 45, 4396–4415.
- Zhu, X., Ghahramani, Z., 2002. Learning from labeled and unlabeled data with label propagation. ProQuest number: information to all users.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine* 5, 8–36.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote sensing of Environment* 159, 269–277.
- Zhu, Z., Zhou, Y., Seto, K.C., Stokes, E.C., Deng, C., Pickett, S.T., Taubenböck, H., 2019. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sensing of Environment* 228, 164–182.

Appendix I 28 Gray-Level Co-occurrence Matrix (GLCM)

Features

Index	Feature	Class	Formula	Description
1	Mean	Advanced	$f_0 = \sum_{i,j} i g(i, j)$	Indicates the average gray level, capturing the general brightness of an image. It measures the mean of the co-occurrence matrix elements, reflecting the overall intensity level in the image.
2	Energy	Simple	$f_1 = \sum_{i,j} g(i, j)^2$	This quantifies the textural uniformity of the image, with higher values suggesting homogeneity. It is also known as the Angular Second Moment, measuring the sum of squared elements in the GLCM.
3	Entropy	Simple	$f_2 = - \sum_{i,j} g(i, j) \log g(i, j)$	Measures the randomness or complexity in the image. High entropy indicates more complex textures with richer information, derived from the unpredictability in pixel distribution.
4	Haralick's Correlation	Simple	$f_3 = \frac{\sum_{i,j} (i, j) g(i, j) - \mu_t^2}{\sigma_t^2}$	Evaluates the linear dependency of gray level values in the image. It quantifies how correlated a pixel is to its neighbors over the whole image.

5	Inverse Difference Moment (or Homogeneity)	Simple	$f_4 = \sum_{i,j} \frac{g(i,j)}{1 + (i - j)^2}$	Highlights the local uniformity in the image. A higher value indicates that the pixel values are similar, suggesting smoother textures.
6	Inertia (or Contrast)	Simple	$f_5 = \sum_{i,j} (i,j)^2 g(i,j)$	Represents the local intensity variation, with higher values indicating more contrast or variation in the image texture.
7	Variance	Advanced	$f_6 = \sum_{i,j} (i - \mu)^2 g(i,j)$	Depicts the variance of gray levels, indicating the degree of variation or spread in pixel intensities.
8	Dissimilarity	Simple	$f_7 = \sum_{i,j} i - j g(i,j)$	Assesses the contrast between gray levels, emphasizing the differences between neighboring pixels.
9	Sum of Average	Advanced	$f_8 = \sum_i i \times [g_x(i) + g_y(i)]$	Aggregates the average values of the row and column indices in the GLCM, providing a measure related to the overall mean intensity of the image.
10	Sum of Variance	Advanced	$f_9 = \sum_i (i - f_{10})^2 g_{x+y}(i)$	Reflects the variance from the sum average of the matrix, measuring the spread or dispersion in the sum distribution of the gray levels.

11	Sum of Entropy	Advanced	$f_{10} = - \sum_i g_{x+y}(i) \log g_{x+y}(i)$	Quantifies the randomness or uncertainty of the sum distribution in the GLCM, indicating the complexity in the texture pattern.
12	Difference of Variance	Advanced	$f_{11} = \text{Variance of } g_{x-y}(i)$	Measures the dispersion in the difference distribution, indicating variations in intensity differences.
13	Difference of Entropy	Advanced	$f_{12} = - \sum_i g_{x+y}(i) \log g_{x-y}(i)$	Measures the randomness or complexity in the difference distribution, highlighting the unpredictability in the intensity differences.
14 & 15	Information Measures of Correlation (IC1 and IC2)	Advanced	$f_{13} = \frac{f_{11}}{HXY1}$ $f_{14} = \sqrt{1 - \exp(-2(HXY2 - f_{11}))}$	These metrics quantify the amount of information required to determine the correlation between occurrences of specific pixel pairs. They help in understanding the informational dependency and redundancy within the image.
16	Correlation	Simple	$f_{15} = \sum_{i,j} \frac{(i - \mu)(j - \mu)g(i,j)}{\sigma^2}$	Similar to Haralick's Correlation, this measures the statistical correlation between pixels, indicating how much one pixel's intensity predicts the intensity of a neighboring pixel.
17	Cluster Shade	Simple	$f_{16} = \sum_{i,j} ((i - \mu) + (j - \mu))^3 g(i,j)$	Measures the skewness or asymmetry of the distribution of the GLCM values. It reflects how much the texture deviates from the uniform distribution.

18	Cluster Prominence	Simple	$f_{17} = \sum_{i,j} ((i - \mu) + (j - \mu))^4 g(i, j)$	Indicates the flatness or peakedness of the GLCM distribution. High values suggest a texture with sharp edges or prominent features.
19	Short Run Emphasis	Higher	$f_{18} = \frac{1}{nr} \sum_{i,j} \frac{p(i, j)}{j^2}$	Emphasizes short runs, or sequences of similar gray level values, indicating fine textures.
20	Long Run Emphasis	Higher	$f_{19} = \frac{1}{nr} \sum_{i,j} p(i, j) \times j^2$	Emphasizes longer runs of similar gray level values, indicating coarse textures.
21	Grey-Level Nonuniformity	Higher	$f_{20} = \frac{1}{nr} \sum_j \left(\sum_i p(i, j) \right)^2$	Measures the variability in grey-level intensity values, indicating the presence of a dominant gray level in the texture.
22	Run Length Nonuniformity	Higher	$f_{21} = \frac{1}{nr} \sum_i \left(\sum_j p(i, j) \right)^2$	Gauges the distribution of runs (sequences of similar pixel values) in the image, reflecting the uniformity of run lengths.
23	Low Grey-Level Run Emphasis	Higher	$f_{22} = \frac{1}{nr} \sum_{i,j} \frac{p(i, j)}{i^2}$	Emphasizes runs with low grey-level values, often highlighting darker textures.

24	High Grey-Level Run Emphasis	Higher	$f_{23} = \frac{1}{nr} \sum_{i,j} p(i,j) \times i^2$	Emphasizes runs with high grey-level values, highlighting brighter textures.
25	Short Run Low Grey-Level Emphasis	Higher	$f_{24} = \frac{1}{nr} \sum_{i,j} \frac{p(i,j)}{i^2 j^2}$	Combines short runs with low grey-level values, indicating fine textures with lower intensity.
26	Short Run High Grey-Level Emphasis	Higher	$f_{25} = \frac{1}{nr} \sum_{i,j} \frac{p(i,j) * i^2}{j^2}$	Combines short runs with high grey-level values, indicating fine textures with higher intensity.
27	Long Run Low Grey-Level Emphasis	Higher	$f_{26} = \frac{1}{nr} \sum_{i,j} \frac{p(i,j) * j^2}{i^2}$	Combines long runs with low grey-level values, indicating coarse textures with lower intensity.
28	Long Run High Grey-Level Emphasis	Higher	$f_{27} = \frac{1}{nr} \sum_{i,j} p(i,j) \times i^2 j^2$	Combines long runs with high grey-level values, indicating coarse textures with lower intensity.

In the context of Haralick's Correlation within the GLCM framework, the terms μ_t and σ_t represent the mean and standard deviation of the sums across rows (or columns, due to the matrix's symmetric property). Specifically, μ refers to the weighted average of pixel intensities, while σ denotes the weighted variance of pixel intensities.

The functions $g_{x+y}(k)$ and $g_{x-y}(k)$ are defined as follows:

- $g_{x+y}(k) = \sum_i \sum_j g(i)$, where $i + j = k$ and $k = 2, 3, \dots, 2N_g$. This function accounts for the sum of elements in the GLCM where the sum of row and column indices equals k .
- $g_{x-y}(k) = \sum_i \sum_j g(i)$, where $i - j = k$ and $k = 0, 1, \dots, N_g - 1$. This function represents the sum of elements where the difference between row and column indices equals k .

For the Run Length Matrix (RLM), $p(i, j)$ represents the normalized value at the i -th row and j -th column, indicating the frequency of runs with length j and gray level i . The total number of runs is denoted by n_r , while n_p represents the total number of pixels in the image. These metrics are used to compute various textural features such as Run Length Nonuniformity (RLN) and Grey-Level Nonuniformity (GLN), which provide insights into the distribution and variability of run lengths and grey levels, respectively, within the image.

In Study 3, only 23 grey scale features were used and Sum of Average, Sum of Variance, Sum of Entropy and Variance were no longer used

Appendix II: Summary of Feature Maps

Below is the revised table that logically presents the details of each feature map, including the band information, the formula for calculating the remote sensing index, and the associated GLCM texture features. The formulas are presented using a more professional format with subscript notation for band names.

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
1	BAI		Blue, NIR ₁	$\frac{Blue - NIR_1}{Blue + NIR_1}$	BAI	None
2	Blue		Blue	N/A	None	None
3	BSI		Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	
4	BSI_Correlation	Simple	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Correlation
5	BSI_DE	Advanced	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	DE
6	BSI_Dissimilarity	Advanced	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Dissimilarity
7	BSI_DV	Advanced	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	DV
8	BSI_Energy	Simple	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Energy
9	BSI_Entropy	Simple	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Entropy
10	BSI_GLN	Higher	Yellow, NIR ₁	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	GLN

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
11	BSI_HC	Simple	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	HC
12	BSI_LGRE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	LGRE
13	BSI_IC1	Advanced	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	IC1
14	BSI_IC2	Advanced	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	IC2
15	BSI_IDM	Simple	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	IDM
16	BSI_Inertia	Simple	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Inertia
17	BSI_SRHGE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	SRHGE
18	BSI_LRE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	LRE
19	BSI_LRHGE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	SRHGE
20	BSI_Mean	Simple	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Mean
21	BSI_LRLGE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	LRLGE
22	BSI_RLN	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	RLN
23	BSI_SRE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	SRE

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
24	BSI_SRLGE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	SRLGE
25	BSI_HGRE	Higher	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	HGRE
26	BSI_Variance	Advanced	Yellow, NIR1	$\frac{Yellow - 2 \cdot NIR_1}{Yellow + 2 \cdot NIR_1}$	BSI	Variance
27	CoastalBlue		Coastal Blue	N/A	None	None
28	Green		Green	N/A	None	None
29	NBEI		Red Edge, NIR1	$\frac{RedEdge - NIR_1}{RedEdge + NIR_1}$	NBEI	None
30	NDVI		Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	None
31	NDVI_Correlation	Simple	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Correlation
32	NDVI_DE	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	DE
33	NDVI_Dissim _i	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Dissimilarity
34	NDVI_Energy _y	Simple	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Energy
35	NDVI_Entropy _y	Simple	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Entropy
36	NDVI_GLN	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	GLN

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
37	NDVI_HC	Simple	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	HC
38	NDVI_LGRE	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	LGRE
39	NDVI_IC1	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	IC1
40	NDVI_IC2	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	IC2
41	NDVI_IDM	Simple	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	IDM
42	NDVI_SRHG E	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	SRHGE
43	NDVI_LRE	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	LRE
44	NDVI_LRHG E	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	LRHGE
45	NDVI_Mean	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Mean
46	NDVI_LRLG E	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	LRLGE
47	NDVI_RLN	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	RLN
48	NDVI_SRE	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	SRE
49	NDVI_SRLG E	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	SRLGE

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
50	NDVI_HGRE	Higher	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	HGRE
51	NDVI_Variance	Advanced	Red, NIR1	$\frac{Red - NIR_1}{Red + NIR_1}$	NDVI	Variance
52	NDWI		Green, NIR1	$\frac{Green - NIR_1}{Green + NIR_1}$	NDWI	None
53	NIR1		NIR1	N/A	None	None
54	NIR2		NIR2	N/A	None	None
55	Red		Red	N/A	None	None
56	RedEdge		Red Edge	N/A	None	None
57	RGI		Red, Green, Blue	$\frac{Red - Green}{Red + Green}$	RGI	None
58	WVBI		Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	None
59	WVBI_Correlation	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	
60	WVBI_CP	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	CP
61	WVBI_CS	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	CS
62	WVBI_DE	Advanced	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	DE

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
63	WVBI_Dissimilarity	Advanced	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	Dissimilarity
64	WVBI_DV	Advanced	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	DV
65	WVBI_Energy	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	Energy
66	WVBI_Entropy	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	Entropy
67	WVBI_GLN	Higher	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	GLN
68	WVBI_HC	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	HC
69	WVBI_LGRE	Higher	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	LGRE
70	WVBI_IC1	Advanced	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	IC1
71	WVBI_IC2	Advanced	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	IC2
72	WVBI_IDM	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	IDM
73	WVBI_Inertia	Simple	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	Inertia
74	WVBI_SRHGE	Higher	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	SRHGE
75	WVBI_LRE	Higher	Coastal Blue, Red Edge	$\frac{Coastal\ Blue - RedEdge}{Coastal\ Blue + RedEdge}$	WVBI	LRE

Index	Feature Map Name	Level	Band Information	Formula	Remote Sensing Index	GLCM Texture Feature
76	WVBI_LRLGE	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	LRLGE
77	WVBI_Mean	Advanced	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	Mean
78	WVBI_RLN	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	RLN
79	WVBI_LRHGE	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	LGRE
80	WVBI_SRE	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	SRE
81	WVBI_SRLGE	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	SRLGE
82	WVBI_HGRE	Higher	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	HGRE
83	WVBI_Variance	Advanced	Coastal Blue, Red Edge	$\frac{\text{Coastal Blue} - \text{RedEdge}}{\text{Coastal Blue} + \text{RedEdge}}$	WVBI	Variance
84	Yellow		Yellow	N/A	None	None

Appendix III. Complete Variable List for the Dual-Branch

GNN Model

This appendix provides a comprehensive list of the input variables used in the final implementation of the dual-branch graph neural network (GNN) model introduced in Chapter 6. As detailed in Section 6.2.4, variable selection was performed independently for the local branch (0–120 m neighborhood) and the global branch (120–1000 m neighborhood), based on correlation analysis and feature importance scores from a Random Forest model trained on reference city data.

Each branch retained the top 80 numerical variables ranked by importance. In addition, four functional group variables and categorical land use variables were incorporated: for the local branch, both the dominant LULC type and the on-site LULC type; and for the global branch, only the dominant LULC type. These variables collectively capture both high-resolution spatial variations and broader land use context.

Each variable is annotated by its:

- Variable Name: Following the naming convention from preprocessing pipelines.
- Buffer Range: The spatial neighborhood used for computing the variable (e.g., r30 → 0–30 m, r500 → 120–500 m).
- Statistic Type: The statistical operation applied (e.g., mean, std, entropy).
- Data Source: The original source of the feature, including:
 1. Urban Atlas 2012 land use polygons (for LULC-derived proportion and distance metrics),
 2. Morphological and local binary patterns (prefixed by MP_ or LBP_),
 3. Z-score standardized remote sensing indices (prefixed by z_),
 4. GLCM-derived texture statistics.

The following tables enumerate all variables used in the final model input, structured by branch.

Appendix 3A. Input Variables of the Local Branch (0–120 m Neighborhood)

This table includes 80 selected numerical variables, 4 functional group variables, and 2 categorical land use variables (dominant and located). The spatial scale for all features in this branch is constrained within a 120 m radius. Notably, the selected features combine Urban Atlas LULC-derived metrics (e.g., road types, green cover) and multi-scale image-derived descriptors (e.g., texture, contrast, entropy), reflecting both environmental context and built structure.

Table AIII.1. List of Local Branch Variables

Index	Rank	Branch	Variable	Buffer Zone	Statistic Type	Data Source
1	1	local	landuse_12220_r3 0	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
2	2	local	landuse_12220_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
3	3	local	landuse_11210_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
4	4	local	landuse_12100_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
5	5	local	MP_BSI_open_r5 _r120_entropy	60–120 m	Entropy	Morphological feature maps
6	6	local	landuse_12100_r1 20	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
7	7	local	landuse_11210_r3 0	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
8	8	local	landuse_11220_r3 0	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
9	9	local	landuse_12220_r1 20	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
10	10	local	landuse_11220_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
11	11	local	landuse_14100_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
12	12	local	landuse_14100_r1 20	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
13	13	local	landuse_11210_r1 20	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
14	14	local	MP_BSI_close_r5 _r120_skewness	60–120 m	Skewness	Morphological feature maps
15	15	local	MP_Green_close _r5_r120_std	60–120 m	Standard Deviation	Morphological feature maps
16	16	local	landuse_12100_r3 0	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
17	17	local	landuse_11230_r6 0	30–60 m	Landuse Proportion	Urban Atlas 2012 LULC
18	18	local	MP_NDVI_open_r5 _r120_percentile90	60–120 m	90th Percentile	Morphological feature maps

19	19	local	landuse_11220_r120	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
20	20	local	z_NDWI_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
21	21	local	landuse_12210_r120	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
22	22	local	landuse_11230_r30	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
23	23	local	MP_NDVI_open_r5_r120_weighted_mean	60–120 m	Mean	Morphological feature maps
24	24	local	landuse_14100_r30	0–30 m	Landuse Proportion	Urban Atlas 2012 LULC
25	25	local	MP_NDVI_open_r5_r120_std	60–120 m	Standard Deviation	Morphological feature maps
26	26	local	z_band7_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
27	27	local	landuse_12230_r120	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
28	28	local	landuse_11230_r120	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
29	29	local	z_NDVI_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
30	30	local	Green_energy_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
31	31	local	z_WVBI_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
32	32	local	Green_correlation_r120_percentile10	60–120 m	10th Percentile	GLCM texture feature maps
33	33	local	Green_correlation_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
34	34	local	MP_NIR1_close_r5_r120_skewness	60–120 m	Skewness	Morphological feature maps

35	35	local	z_BSI_win77_tanh_r120_percentile90	60–120 m	90th Percentile	Z-score normalized feature maps
36	36	local	z_band7_win77_tanh_r120_percentile90	60–120 m	90th Percentile	Z-score normalized feature maps
37	37	local	z_band3_win77_tanh_r120_std	60–120 m	Standard Deviation	Z-score normalized feature maps
38	38	local	MP_NDVI_close_r5_r120_entropy	60–120 m	Entropy	Morphological feature maps
39	39	local	RGI_correlation_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
40	40	local	z_RGI_win77_tanh_r120_percentile25	60–120 m	25th Percentile	Z-score normalized feature maps
41	41	local	MP_Green_close_r5_r120_percentile90	60–120 m	90th Percentile	Morphological feature maps
42	42	local	LBP_NDVI_r120_kurtosis	60–120 m	Kurtosis	Morphological feature maps
43	43	local	Green_correlation_r120_kurtosis	60–120 m	Kurtosis	GLCM texture feature maps
44	44	local	MP_BSI_close_r5_r120_mean	60–120 m	Mean	Morphological feature maps
45	45	local	z_band3_win15_tanh_r120_percentile90	60–120 m	90th Percentile	Z-score normalized feature maps
46	46	local	NIR1_homogeneity_r120_percentile10	60–120 m	10th Percentile	GLCM texture feature maps
47	47	local	landuse_31000_r120	60–120 m	Landuse Proportion	Urban Atlas 2012 LULC
48	48	local	z_RGI_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
49	49	local	Green_correlation_r60_skewness	30–60 m	Skewness	GLCM texture feature maps
50	50	local	NDVI_energy_r120_percentile10	60–120 m	10th Percentile	GLCM texture feature maps

51	51	local	LBP_BSI_r120_entropy	60–120 m	Entropy	Morphological feature maps
52	52	local	RedEdge_energy_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
53	53	local	WVBI_correlation_r60_skewness	30–60 m	Skewness	GLCM texture feature maps
54	54	local	MP_NIR1_close_r5_r120_iqr	60–120 m	Interquartile Range	Morphological feature maps
55	55	local	MP_NIR1_close_r5_r120_entropy	60–120 m	Entropy	Morphological feature maps
56	56	local	RedEdge_correlation_r120_weighted_mean	60–120 m	Mean	GLCM texture feature maps
57	57	local	Green_contrast_r120_kurtosis	60–120 m	Kurtosis	GLCM texture feature maps
58	58	local	Green_correlation_r120_iqr	60–120 m	Interquartile Range	GLCM texture feature maps
59	59	local	MP_NIR1_close_r5_r120_std	60–120 m	Standard Deviation	Morphological feature maps
60	60	local	z_band7_win15_tanh_r120_median	60–120 m	Median	Z-score normalized feature maps
61	61	local	z_band7_win31_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
62	62	local	RedEdge_correlation_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
63	63	local	NDVI_correlation_r120_kurtosis	60–120 m	Kurtosis	GLCM texture feature maps
64	64	local	RedEdge_correlation_r120_std	60–120 m	Standard Deviation	GLCM texture feature maps
65	65	local	MP_Red_open_r5_r120_skewness	60–120 m	Skewness	Morphological feature maps
66	66	local	z_WVBI_win77_tanh_r120_percentile90	60–120 m	90th Percentile	Z-score normalized feature maps
67	67	local	RGI_correlation_r120_kurtosis	60–120 m	Kurtosis	GLCM texture feature maps

68	68	local	MP_Green_open_r5_r60_skewness	30–60 m	Skewness	Morphological feature maps
69	69	local	NIR1_correlation_r120_percentile10	60–120 m	10th Percentile	GLCM texture feature maps
70	70	local	Green_correlation_r60_kurtosis	30–60 m	Kurtosis	GLCM texture feature maps
71	71	local	NIR1_homogeneity_r60_skewness	30–60 m	Skewness	GLCM texture feature maps
72	72	local	z_band3_win77_tanh_r60_iqr	30–60 m	Interquartile Range	Z-score normalized feature maps
73	73	local	MP_BSI_open_r5_r120_kurtosis	60–120 m	Kurtosis	Morphological feature maps
74	74	local	MP_NIR1_open_r5_r120_skewness	60–120 m	Skewness	Morphological feature maps
75	75	local	MP_BSI_open_r5_r120_skewness	60–120 m	Skewness	Morphological feature maps
76	76	local	z_band3_win77_tanh_r120_percentile10	60–120 m	10th Percentile	Z-score normalized feature maps
77	77	local	WVBI_correlation_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
78	78	local	NIR1_correlation_r120_skewness	60–120 m	Skewness	GLCM texture feature maps
79	79	local	WVBI_correlation_r120_kurtosis	60–120 m	Kurtosis	GLCM texture feature maps
80	80	local	MP_Red_open_r5_r120_kurtosis	60–120 m	Kurtosis	Morphological feature maps
81		local	strong_noise_source	0–120 m	Landuse Proportion	GLCM texture feature maps
82		local	moderate_noise_source	0–120 m	Landuse Proportion	GLCM texture feature maps
83		local	strong_mitigation_zone	0–120 m	Landuse Proportion	GLCM texture feature maps
84		local	moderate_mitigation_zone	0–120 m	Landuse Proportion	GLCM texture feature maps
85		local	dominant_LULC	0–120 m	Landuse Proportion	GLCM texture feature maps

86		local	located_lulc		Landuse Location	GLCM texture feature maps
----	--	-------	--------------	--	------------------	---------------------------

Appendix 3B. Input Variables of the Global Branch (120–1000 m Neighborhood)

This table presents the 80 selected numerical variables, 4 functional group proportions, and 1 categorical variable (dominant land use type) used for the global branch. These features summarize large-scale environmental structure and background context, helping the model learn inter-neighborhood patterns and spatial continuity.

Table AIII.2. List of Global Branch Variables

Index	Rank	Branch	Variable	Buffer Zone	Statistic Type	Data Source
1	1	backgro und	Green_contrast_r500_percentile75	120–500 m	75th Percentile	GLCM Texture of Remote Sensing Indices
2	2	backgro und	MP_BSI_close_r5_r1000_kurtosis	500–1000 m	Kurtosis	Morphological Profile (MP) of Remote Sensing Indices
3	3	backgro und	landuse_14200_r500	120–500 m	Landuse Proportion	Urban Atlas 2012 Land Use
4	4	backgro und	LBP_NIR1_r500_std	120–500 m	Standard Deviation	Local Binary Pattern (LBP) of Remote Sensing Indices
5	5	backgro und	MP_Green_close_r5_r1000_skewness	500–1000 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
6	6	backgro und	Green_contrast_r1000_percentile90	500–1000 m	90th Percentile	GLCM Texture of Remote Sensing Indices
7	7	backgro und	landuse_12220_r500	120–500 m	Landuse Proportion	Urban Atlas 2012 Land Use
8	8	backgro und	NDVI_correlation_r1000_mean	500–1000 m	Mean	GLCM Texture of Remote Sensing Indices
9	9	backgro und	landuse_12220_r1000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
10	10	backgro und	RGI_energy_r500_skewness	120–500 m	Skewness	GLCM Texture of Remote Sensing Indices
11	11	backgro und	WVBI_correlation_r1000_percentile75	500–1000 m	75th Percentile	GLCM Texture of Remote Sensing Indices
12	12	backgro und	MP_BSI_close_r5_r1000_entropy	500–1000 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices
13	13	backgro und	landuse_12100_r1000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
14	14	backgro und	global_dist_14200	120–1000 m	Distance	Urban Atlas 2012 Land Use

15	15	backgro und	MP_Red_close_r5_r1000_skewness	500–1000 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
16	16	backgro und	z_RGI_win15_tanh_r1000_percentile25	500–1000 m	25th Percentile	Z-score Normalized Feature Map
17	17	backgro und	global_dist_14100	120–1000 m	Distance	Urban Atlas 2012 Land Use
18	18	backgro und	MP_Red_close_r5_r500_skewness	120–500 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
19	19	backgro und	Green_correlation_r500_percentile90	120–500 m	90th Percentile	GLCM Texture of Remote Sensing Indices
20	20	backgro und	MP_Green_open_r5_r500_entropy	120–500 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices
21	21	backgro und	NDVI_homogeneity_r1000_std	500–1000 m	Standard Deviation	GLCM Texture of Remote Sensing Indices
22	22	backgro und	MP_Red_open_r5_r500_entropy	120–500 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices
23	23	backgro und	Green_correlation_r500_skewness	120–500 m	Skewness	GLCM Texture of Remote Sensing Indices
24	24	backgro und	NIR1_correlation_r500_kurtosis	120–500 m	Kurtosis	GLCM Texture of Remote Sensing Indices
25	25	backgro und	RedEdge_correlation_r500_kurtosis	120–500 m	Kurtosis	GLCM Texture of Remote Sensing Indices
26	26	backgro und	z_BSI_win15_tanh_r500_percentile90	120–500 m	90th Percentile	Z-score Normalized Feature Map
27	27	backgro und	NDVI_correlation_r500_mean	120–500 m	Mean	GLCM Texture of Remote Sensing Indices
28	28	backgro und	Green_contrast_r1000_percentile75	500–1000 m	75th Percentile	GLCM Texture of Remote Sensing Indices
29	29	backgro und	landuse_14100_r500	120–500 m	Landuse Proportion	Urban Atlas 2012 Land Use
30	30	backgro und	landuse_14100_r1000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
31	31	backgro und	NIR1_correlation_r500_skewness	120–500 m	Skewness	GLCM Texture of Remote Sensing Indices
32	32	backgro und	MP_Red_open_r5_r1000_std	500–1000 m	Standard Deviation	Morphological Profile (MP) of Remote Sensing Indices
33	33	backgro und	MP_Red_close_r5_r1000_entropy	500–1000 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices

34	34	backgro und	WVBI_correlatio n_r500_skewness	120–500 m	Skewness	GLCM Texture of Remote Sensing Indices
35	35	backgro und	NIR1_correlation _r500_percentile2 5	120–500 m	75th Percentile	GLCM Texture of Remote Sensing Indices
36	36	backgro und	RGI_correlation_r 1000_skewness	500–1000 m	Skewness	GLCM Texture of Remote Sensing Indices
37	37	backgro und	z_band3_win15_t anh_r1000_std	500–1000 m	Standard Deviation	Z-score Normalized Feature Map
38	38	backgro und	z_NDWI_win15_tanh_r500_perce ntile90	120–500 m	90th Percentile	Z-score Normalized Feature Map
39	39	backgro und	NIR1_correlation _r500_mean	120–500 m	Mean	GLCM Texture of Remote Sensing Indices
40	40	backgro und	MP_Green_close _r5_r500_skewne ss	120–500 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
41	41	backgro und	MP_Green_open _r5_r500_skewnes s	120–500 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
42	42	backgro und	MP_NIR1_open _r5_r1000_skewne ss	500–1000 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
43	43	backgro und	WVBI_correlatio n_r500_kurtosis	120–500 m	Kurtosis	GLCM Texture of Remote Sensing Indices
44	44	backgro und	NDVI_correlation _r1000_skewness	500–1000 m	Skewness	GLCM Texture of Remote Sensing Indices
45	45	backgro und	z_band3_win77_t anh_r500_iqr	120–500 m	Interquartile Range	Z-score Normalized Feature Map
46	46	backgro und	RGI_homogeneit y_r500_percentile 90	120–500 m	90th Percentile	GLCM Texture of Remote Sensing Indices
47	47	backgro und	landuse_12100_r5 00	120–500 m	Landuse Proportion	Urban Atlas 2012 Land Use
48	48	backgro und	z_band3_win77_t anh_r500_median	120–500 m	Median	Z-score Normalized Feature Map
49	49	backgro und	WVBI_correlatio n_r500_percentile 10	120–500 m	10th Percentile	GLCM Texture of Remote Sensing Indices
50	50	backgro und	landuse_11210_r1 000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
51	51	backgro und	z_RGI_win77_tan h_r500_percentile 75	120–500 m	75th Percentile	Z-score Normalized Feature Map
52	52	backgro und	landuse_11220_r5 00	120–500 m	Landuse Proportion	Urban Atlas 2012 Land Use
53	53	backgro und	WVBI_contrast_r 1000_entropy	500–1000 m	Entropy	GLCM Texture of Remote Sensing Indices

54	54	backgro und	MP_Green_close_r5_r1000_kurtosis	500–1000 m	Kurtosis	Morphological Profile (MP) of Remote Sensing Indices
55	55	backgro und	MP_Green_open_r5_r500_kurtosis	120–500 m	Kurtosis	Morphological Profile (MP) of Remote Sensing Indices
56	56	backgro und	NIR1_correlation_r1000_kurtosis	500–1000 m	Kurtosis	GLCM Texture of Remote Sensing Indices
57	57	backgro und	MP_NIR1_close_r5_r1000_entropy	500–1000 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices
58	58	backgro und	RGI_correlation_r1000_mean	500–1000 m	Mean	GLCM Texture of Remote Sensing Indices
59	59	backgro und	landuse_14200_r1000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
60	60	backgro und	RGI_correlation_r500_percentile25	120–500 m	25th Percentile	GLCM Texture of Remote Sensing Indices
61	61	backgro und	NDVI_correlation_r500_kurtosis	120–500 m	Kurtosis	GLCM Texture of Remote Sensing Indices
62	62	backgro und	MP_Red_open_r5_r1000_skewness	500–1000 m	Skewness	Morphological Profile (MP) of Remote Sensing Indices
63	63	backgro und	z_band3_win77_tanh_r1000_iqr	500–1000 m	Interquartile Range	Z-score Normalized Feature Map
64	64	backgro und	RedEdge_contrast_r1000_percentile90	500–1000 m	90th Percentile	GLCM Texture of Remote Sensing Indices
65	65	backgro und	WVBI_contrast_r500_entropy	120–500 m	Entropy	GLCM Texture of Remote Sensing Indices
66	66	backgro und	NIR1_correlation_r1000_skewness	500–1000 m	Skewness	GLCM Texture of Remote Sensing Indices
67	67	backgro und	RGI_contrast_r500_entropy	120–500 m	Entropy	GLCM Texture of Remote Sensing Indices
68	68	backgro und	RedEdge_correlation_r500_percentile10	120–500 m	10th Percentile	GLCM Texture of Remote Sensing Indices
69	69	backgro und	MP_Red_open_r5_r1000_kurtosis	500–1000 m	Entropy	Morphological Profile (MP) of Remote Sensing Indices
70	70	backgro und	RGI_contrast_r1000_entropy	500–1000 m	Entropy	GLCM Texture of Remote Sensing Indices
71	71	backgro und	Green_contrast_r500_entropy	120–500 m	Entropy	GLCM Texture of Remote Sensing Indices
72	72	backgro und	landuse_11230_r1000	500–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use

73	73	backgro und	z_band3_win15_tanh_r1000_percentile90	500–1000 m	90th Percentile	Z-score Normalized Feature Map
74	74	backgro und	z_RGI_win77_tanh_r1000_std	500–1000 m	Standard Deviation	Z-score Normalized Feature Map
75	75	backgro und	Green_contrast_r500_mean	120–500 m	Mean	GLCM Texture of Remote Sensing Indices
76	76	backgro und	z_band6_win77_tanh_r500_percentile90	120–500 m	90th Percentile	Z-score Normalized Feature Map
77	77	backgro und	RGI_correlation_r500_mean	120–500 m	Mean	GLCM Texture of Remote Sensing Indices
78	78	backgro und	RedEdge_contrast_r500_percentile75	120–500 m	75th Percentile	GLCM Texture of Remote Sensing Indices
79	79	backgro und	LBP_Green_r500_std	120–500 m	Standard Deviation	Local Binary Pattern (LBP) of Remote Sensing Indices
80	80	backgro und	RGI_correlation_r500_skewness	120–500 m	Skewness	GLCM Texture of Remote Sensing Indices
81		backgro und	strong_noise_source	120–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
82		backgro und	moderate_noise_source	120–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
83		backgro und	strong_mitigation_zone	120–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
84		backgro und	moderate_mitigation_zone	120–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use
85		backgro und	dominant_LULC	120–1000 m	Landuse Proportion	Urban Atlas 2012 Land Use

Each variable was used without further transformations during training. For categorical LULC variables, one-hot encoding was applied. Details regarding buffer generation, morphological operations, and image normalization techniques can be found in Section 6.2.3 and 6.2.4.

This structured input design enables the model to integrate fine-scale spatial patterns with broader environmental information, forming a solid foundation for generalizable noise prediction across heterogeneous urban environments.