Transportation Research Part D 156 (2026) 105378

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Cross-city traffic noise modeling and inequality analysis via AlphaEarth Geospatial Foundation Models

Yan Zhang[a,b], Quan Qin[c], Hairu Nie[b], Mei-Po Kwan[a,b,*], Sijia He[d], Entong Ke[c]

[a] *Department of Geography and Resource Management, Wong Foo Yuan Building, The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China*
[b] *Institute of Space and Earth Information Science, Fok Ying Tung Remote Sensing Science Building, The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China*
[c] *School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China*
[d] *National Engineering Research Center of Geographic Information System, School of Geography and Information Engineering, China University of Geosciences, Wuhan, China*

A R T I C L E   I N F O

Keywords:
Geospatial foundation models
Traffic noise exposure
Remote sensing
Environmental justice
Multi-sensor fusion
AlphaEarth

A B S T R A C T

Traffic noise is a widespread environmental problem in modern cities. Traditional traffic noise modeling is limited to small-scale coverage by high sensor costs and complex simulation parameter settings. GeoAI-based methods (incorporating remote sensing and street view) can effectively address scalability challenges. However, those methods use RGB imagery and lack physical observation approaches (e.g., LiDAR), which often leads to insufficient characterization of complex urban environments (e.g., three-dimensional buildings, pavement materials, vegetation absorption). This study proposes a cross-city noise exposure framework via the AlphaEarth Foundation (AEF) model. Pre-trained on 3 billion multimodal observations (including SAR, and meteorology), AEF mitigates the reliance on extensive ground-truth datasets through capturing complex coupling between morphology, 3D structures, and acoustic characteristics. Our research findings demonstrate that: (1) The model achieves a coefficient of determination (R^2) greater than 0.84 for noise estimation. (2) The traffic noise exposure transfer computations across multiple global cities reveal exposure inequalities.

1. Introduction

Noise pollution ranks as the third most significant pollution source affecting human society, following air and water pollution. As a widespread yet underestimated environmental health hazard in high-density cities, it poses substantial risks to public health. The World Health Organization (WHO) has identified noise pollution as a significant threat to human health and well-being. Previous studies have confirmed direct links between noise exposure and cardiovascular diseases (Münzel et al., 2018, 2021), cognitive decline in children (Clark et al., 2012; Thompson et al., 2022; Welch et al., 2023), deteriorating mental health conditions and sleep disorders

* Corresponding author at: Institute of Space and Earth Information Science, Fok Ying Tung Remote Sensing Science Building, The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China

E-mail addresses: yanzhang@cuhk.edu.hk, sggzhang@whu.edu.cn (Y. Zhang), qinquan@whu.edu.cn (Q. Qin), hairunie2261@gmail.com (H. Nie), mpkwan@cuhk.edu.hk (M.-P. Kwan), hesj@cug.edu.cn (S. He), entongke@whu.edu.cn (E. Ke).

https://doi.org/10.1016/j.trd.2026.105378
Received 2 January 2026; Received in revised form 30 March 2026; Accepted 6 April 2026
Available online 19 April 2026
1361-9209/

(Dzhambov & Lercher, 2019; Hahad et al., 2025).

With accelerating global urbanization, an estimated 68% of the world's population will reside in cities by 2050 (Zhang et al., 2021). The spatial extent and intensity of noise pollution are expanding, particularly in rapidly growing megacities in the Global South, where regulatory frameworks and monitoring infrastructure remain inadequate (Li et al., 2022). Noise pollution encompasses multiple categories, among which traffic noise represents one of the most impactful sources. Traffic noise pollution in high-density cities is influenced by factors such as jobs-housing imbalance, transport system design, and variations in the adoption of sound-insulating materials. This creates considerable health impacts on the human population. Estimating the urban soundscape accurately is essential for assessing exposure levels and inequality, particularly for protecting vulnerable populations such as children, older adults, and those with pre-existing health conditions (Tang and Li, 2024; Wang et al., 2022, 2024). However, comprehensive traffic noise mapping at the urban scale remains a substantial practical challenge (Wei et al., 2016; Yang et al., 2020).

Existing research has primarily relied on three methodological frameworks, each presenting fundamental limitations. The first approach employs simulation models based on physical factors. This method requires extensive data support for traffic noise simulation, including multi-source data such as traffic flow information, road surface characteristics, and road hierarchies. Consequently, accurate estimations need difficult data pre-processing procedures (Kang et al., 2016; Bello et al., 2019; Korpilo et al., 2024). For a 10-square-kilometer urban area, data acquisition costs typically exceed tens of thousands of US dollars, while computational complexity constrains scalability to city-wide applications. The second approach utilizes dense sensor networks (Li et al., 2025). Although this method requires minimal data pre-processing, it demands thousands of monitoring stations, incurring substantial annual costs for equipment maintenance and data management (Wei et al., 2016; Wang et al., 2025). This approach, covering only selected critical zones and generating significant spatial coverage gaps, proves financially infeasible for most municipalities (Leach et al., 2024; Song et al., 2024). Furthermore, both of the mentioned methods suffer from severe data justice issues. Noise exposure assessments remain concentrated in affluent cities of the Global North, while 90% of urban population growth occurs in Asia and Africa, where environmental monitoring is sparse or absent (Hammer et al., 2014; Casey et al., 2017). This geographic disparity in data availability limits the development of noise reduction strategies in the communities also affected by other environmental health risks.

The third approach employs Land Use Regression (LUR) models. This statistical method uses regression frameworks to estimate noise levels based on factors such as land use types, road traffic density, and building height (Wang et al., 2024; Kumar et al., 2025). However, LUR models consider relatively limited factors and cannot capture nonlinear physical interactions (e.g., sound reflections from terrain and sound absorption by vegetation). They exhibit poor transferability across cities and require extensive local calibration datasets (Zhang et al., 2022).

The advancement of geospatial artificial intelligence (GeoAI) and machine learning methods (Janowicz et al., 2020; Lan and Cai, 2021; Liu and Biljecki, 2022; Yin et al., 2020), as well as multi-source geospatial big data provide new pathways for understanding and modeling urban noise. Widely available remote sensing imagery (Chen et al., 2025; Zhao et al., 2026), street view imagery (Huang et al., 2024; Zhuang et al., 2024), and noise complaint data (Zhang et al., 2024) offer some new solutions. Compared to the three introduced approaches, methods based on geotagged images (Street View Image or Aerial Image) can better balance spatial coverage and modeling accuracy (Zhang et al., 2026; Zhang and Kwan, 2025), demonstrating substantial practical application potential. Nevertheless, these emerging methods still face limitations. The image data employed typically contains only visible RGB bands, constraining the model's understanding of the physical mechanisms underlying noise generation (Lu et al., 2025).

Accurate urban noise modeling is fundamentally a multi-physical factors (or multiphysics) coupling problem involving complex processes such as source characteristics (e.g., vehicle traffic speed) (Lan and Cai, 2021), sound wave propagation (e.g., canyon effects produced by high-density skyscrapers) (Yuan et al., 2019; Benocci et al., 2020), environmental absorption, and meteorological conditions. It is not merely a visual pattern recognition task. While existing RGB image-based vision foundation models (e.g., vision transformers model pretrained on remote sensing imagery) (Siméoni et al., 2025) can capture visual features such as building morphology and road layouts in noise modeling tasks, they cannot directly perceive the critical physical processes of acoustic

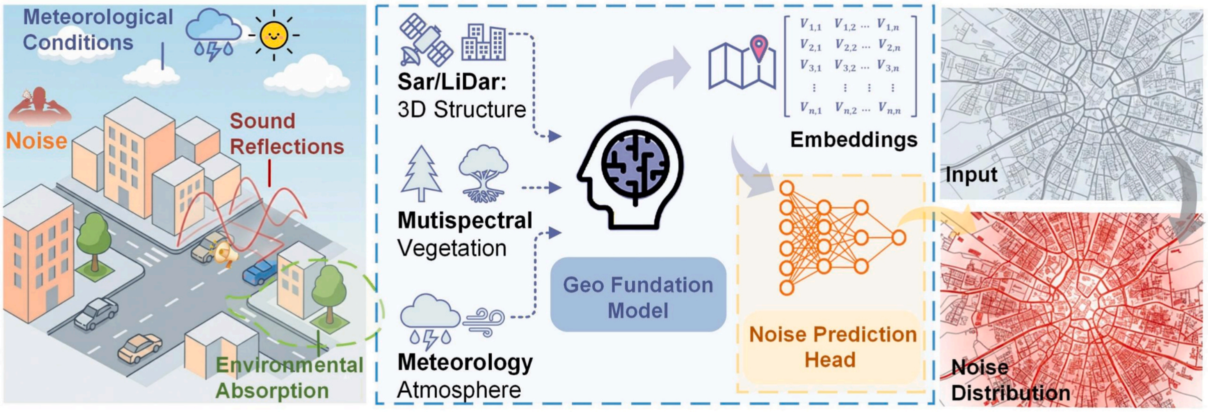

Fig. 1. Diagram illustrating the noise estimation research gap. This reveals the limitations of traditional remote sensing vision algorithms, where RGB visual information alone is insufficient for complex physical noise modeling tasks.

propagation mentioned above (Fig. 1).

The emergence of geospatial foundation models (GeoFMs) represents a paradigm shift in Earth observation analysis, providing novel capabilities for addressing environmental modeling challenges with sparse data sources (Mai et al., 2024; Janowicz et al., 2025). Unlike traditional deep learning models trained on task-specific labeled datasets or solely optical imagery, GeoFMs leverage self-supervised pretraining on massive multi-modal earth observation data. And it could easily be generalized to other tasks (such as land use classification, land surface temperature estimation, cropland and greenspace identification) (Zhang et al., 2023) based on zero-shot or few-shot learning. The newly released Google AlphaEarth Foundation (AEF) Model (Brown et al., 2025) was pretrained on 3 billion observations encompassing nine sensor modalities (Sentinel-2 multispectral, Sentinel-1 SAR, LiDAR, MODIS, ERA5 meteorological reanalysis, etc.) and one unstructured text source (Wikipedia), effectively capturing the physical coupling processes between human society and the urban built environment (Lian et al., 2026).

Through implicitly encoding these cross-modal dependencies, the geographic contrastive learning pretraining process maximizes information similarity between co-located sensor observations (Huang et al., 2024; Qin et al., 2025). GeoFMs such as AlphaEarth could generate a 64-dimensional embedding vector for each 10×10-meter grid cell on the Earth's surface. This vector could transform Earth observation (EO) knowledge into a quantifiable "semantic space," enabling cross-regional comparisons (across cities or even countries) based on minimal local validation data and facilitating transfer learning to unseen cities. This capability cannot be achieved by vision foundation models (e.g., DINO) (Siméoni et al., 2025) that rely solely on RGB imagery and lack domain-specific prior physical knowledge.

Urban traffic noise arises from diverse factors. The multi-modal, multiphysics-driven AEF integrates sensor data from Sentinel-2 multispectral, SAR, LiDAR, and ERA5 meteorological sources, which could help us establish implicit mapping relationships between sensor observations and acoustic environments, respectively. The AEF could capture the street canyon sound reflections from SAR texture, vegetation effects from multispectral data, and atmospheric conditions from weather observations. Unlike RGB-based visual models that learn location-specific patterns, this approach encodes generalizable physical principles governing sound propagation. These Multiphysics-based rules enable robust cross-city transfer through few-shot learning.

In summary, this study advances Sustainable Development Goal 11 (Sustainable Cities and Communities) by providing scalable tools for noise pollution assessment in data-limited environments. By reducing sensor deployment costs, this approach enables city-wide traffic noise mapping previously infeasible in the Global South. The few-shot transfer learning paradigm is generalizable to other urban environmental health challenges where sparse ground-truth data constrain traditional modeling methods (e.g., air quality, urban heat islands, flooding).

2. Methodology

To achieve accurate regression of noise levels, the entire model adopts a fine-tuning paradigm of "pre-trained backbone + trainable head." We construct a lightweight task-specific prediction head on top of the pre-trained geospatial foundation model as shown in

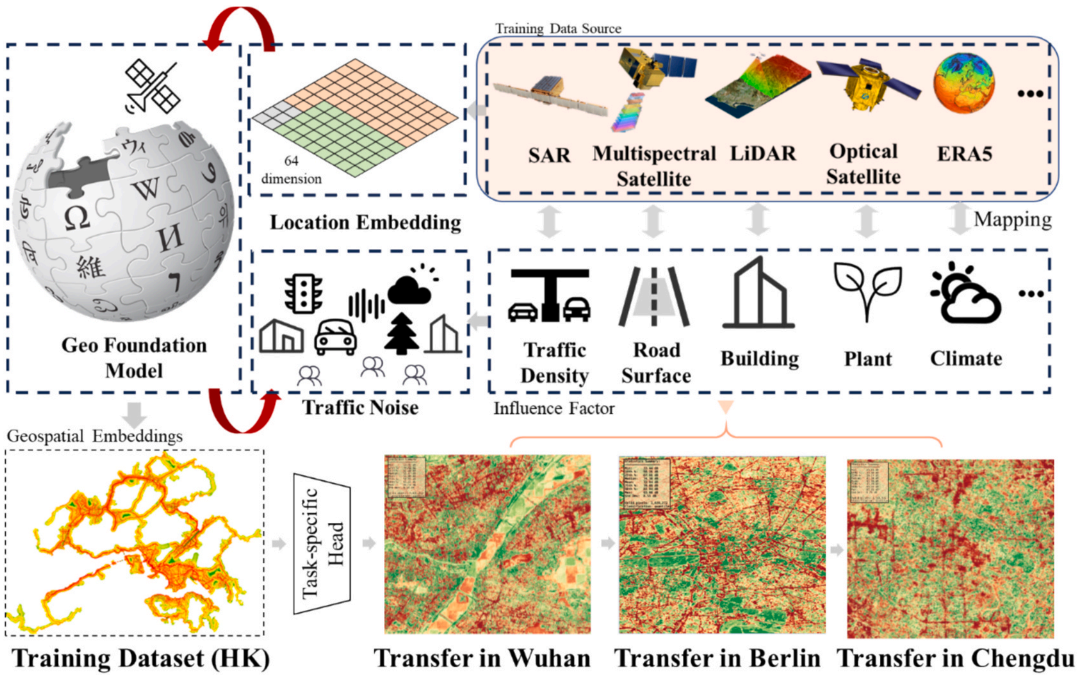

Fig. 2. Schematic diagram of the research framework. This study adopts a two-stage approach: Stage 1 leverages the geospatial foundation model to extract physics-informed embeddings, while Stage 2 applies a task-specific regression head for downstream road traffic noise exposure prediction.

Fig. 2. This head was designed to map the general, high-dimensional geospatial features extracted by the foundation model to the target noise values efficiently. The foundation model serves as a frozen feature extractor, and its output provides us with a feature vector rich in information from multiple physical observation factors. During the training phase, we keep the parameters of the foundation model fixed and update only the parameters of the task-specific head.

This paradigm allows us to fully leverage the powerful generalized representations learned by the foundation model from massive multimodal data, treating them as transferable geographic prior knowledge. By optimizing a small number of parameters in the head, we significantly improve training efficiency and effectively prevent potential overfitting on limited ground-truth data. Acting as a dedicated regression adapter, this head enables the same foundation model to be flexibly and cost-effectively adapted to the specific downstream task of urban traffic noise prediction.

2.1. Noise exposure dataset and research areas

Google AlphaEarth Foundation Model has demonstrated strong performance in multiple geoscience-related classification and regression tasks. Yet, its application in social computing, particularly in noise modeling, has not been widely deployed. One key limitation is the lack of large-scale, high-quality noise observation data. While massive Earth observation data are continuously collected globally, high-quality noise labels remain scarce, given the substantial effort required for physical measurements and observations. We use the traffic noise exposure map constructed for Hong Kong by Chan et al. (2023) based on the Road Traffic Noise Assessment Method for Hong Kong (RONOSS-HK) as our training dataset. Released in August 2023, this dataset is a high-resolution road traffic noise map at a 4-meter grid scale and was constructed for simulating Hong Kong's traffic noise for 2020. The reliability and accuracy of the RONOSS-HK dataset have been validated against extensive field measurements and modeled using the Road Traffic Noise Assessment Method for Hong Kong, demonstrating high spatial fidelity and strong agreement with measured traffic noise levels across diverse urban locations (Chan et al., 2023). Summary statistics indicate that the dataset captures realistic temporal and spatial variations of traffic noise, supporting its use as a robust training dataset for AlphaEarth-based modeling. The dataset, representing L_{10} values at a 4-meter grid resolution, was obtained from the Hong Kong Government Geoportal (https://portal.csdi.gov.hk/geoportal/#metadataInfoPanel), which provides standardized access to official environmental and geospatial data, ensuring reproducibility and transparency for research applications. The dataset represents noise levels as L_{10} (1 hour) in dBA, which indicates the noise level exceeded 10% of the time during one-hour traffic peak periods. This is because traffic noise fluctuates (high when vehicles pass, low during idle periods), and L_{10} represents the occurrence frequency of higher noise levels, better reflecting the impact of peak noise compared to the average value Leq. L_{110} is traditionally used to evaluate traffic noise in Commonwealth countries and regions. As shown in Fig. 3, our study area's traffic noise distribution is primarily concentrated along both sides of roads, with decibel values ranging from 0 to 97 dBA.

Google AlphaEarth Foundations is an embedding field model containing 64-dimensional vector representations of Earth's surface in a grid format, with a minimum grid size of 10 meters × 10 meters, encompassing global annual embedding field layers from 2017 to 2024 as analyzable datasets. These embeddings were accessed and extracted via the Google Earth Engine (GEE) platform, which standardizes the retrieval of multi-source Earth observation data and ensures reproducibility and scalability for research applications. To achieve temporal matching, we used the 2020 embedding field layer that corresponds to the same time period as the noise dataset. The embeddings integrate information from multiple sensor modalities, including Sentinel-2 multispectral, Sentinel-1 SAR, LiDAR, and ERA5 meteorology, providing rich spatial and temporal context for each grid cell. It provides highly generalizable geospatial representations by fusing spatial, temporal, and measurement contexts from multiple sources, enabling accurate and efficient generation of mapping and monitoring systems from local to global scales. The 64-dimensional vector serves as a comprehensive representation of multiple features of a given parcel, implicitly containing rich information for downstream tasks. Comparable to many

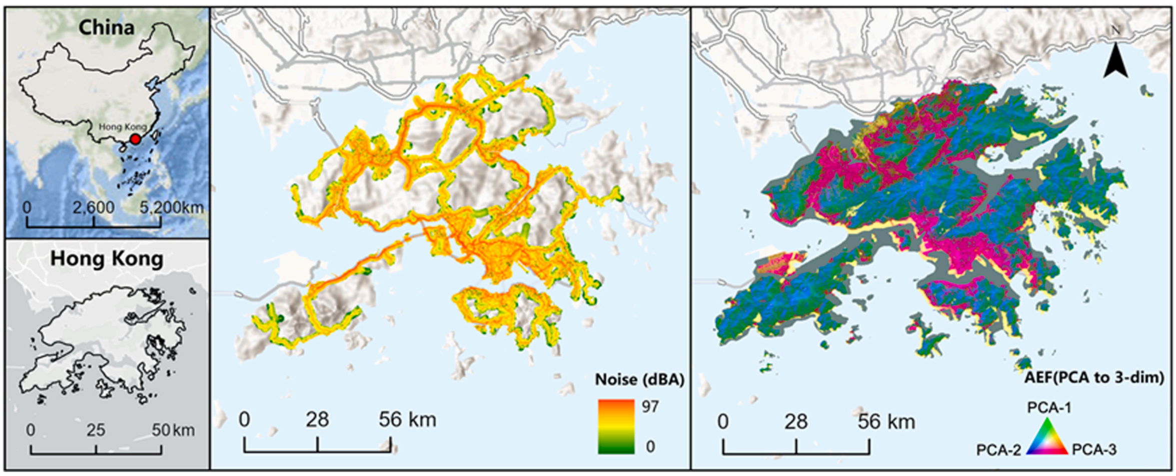

Fig. 3. Study area and research data.

previous single-purpose models, foundation models require minimal fine-tuning and only very limited (task-specific) training data (Huang et al., 2023; Janowicz et al., 2025).

2.2. Traffic noise prediction model architecture

We model point-level urban noise exposure as a supervised regression problem. For each location p, a geospatial foundation model provides a 64-dimensional embedding vector $x_p \in R^{64}$ derived from the geospatial context. This geospatial foundation model provides 10-meter × 10-meter grid cells covering every inch of the Earth's surface, employing self-supervised learning algorithms including contrastive learning and masked prediction techniques. The model takes multi-source Earth observation data as input and outputs high-dimensional vector representations that encapsulate the region's characteristics. These learned representation vectors (AlphaEarth Foundations satellite embeddings), pre-trained on multi-source Earth observation data including optical, SAR, and land cover imagery, encode rich geospatial context and have demonstrated high accuracy in tasks such as population prediction and air quality forecasting (Alvarez et al., 2025). Our predictor $f_\theta(\cdot)$ maps x_p to a continuous noise exposure estimate $\hat{y}_p \in R$ (in dBA).

In this research, we employed the Multi-Layer Perceptron-based specific task head to help us predict the distribution of traffic noise. The network is a multilayer perceptron with residual refinement and channel attention. First, an entry layer projects the input embedding into a higher-dimensional feature space using a fully connected layer 64 →512, followed by Batch Normalization and a non-linear activation (e.g., ReLU/SiLU). Feature refinement is then performed by two consecutive Residual-SE blocks at a width of 512. Each Residual-SE block consists of two 512 →512 linear transformations with Batch Normalization and activation (after the first transformation), combined with an identity skip connection. The Residual-SE blocks are designed to facilitate gradient flow, reduce training instability, and allow the network to learn complex nonlinear feature interactions, enhancing the model's ability to capture subtle patterns in geospatial embeddings, following the design of SE-ResNet (Hu et al., 2018). To adaptively recalibrate feature channels, we incorporate a Squeeze-and-Excitation (SE) module within each block: a bottleneck gating network 512→32→512 produces channel-wise weights via a sigmoid function, which are applied multiplicatively to the block features before residual addition. This design encourages the model to emphasize noise-relevant components of the learned geospatial representation while maintaining stable optimization through residual learning.

After the 512-width stage, the representation is compressed via a linear projection 512→256. Two additional Residual-SE blocks are applied at width 256, using an SE bottleneck 256→16→256. The features are then further projected to 128 dimensions (256→128) and mapped to the final scalar prediction using a linear regression head (128→1). This architecture contains 1,607,690 trainable parameters. The model is trained end-to-end to minimize a regression loss between $\hat{y}_p$ and the observed noise exposure y_p. We use the mean squared error (MSE), defined as $\mathrm{MSE} = \frac{1}{n}\sum i = 1^n\left(y_i - \hat{y}_1\right)^2$, as the loss function, following standard practice in urban noise prediction (Wei et al., 2016). Several hyperparameters associated with the prediction head, including the learning rate, batch size, and hidden dimension, were determined based on empirical experience and preliminary experiments to ensure stable convergence and effective learning. It enabled direct prediction of point-level noise exposure from foundation-model embeddings.

2.3. Spatial concentration and inequality assessment of traffic noise exposure

Given that our method can generate high-resolution traffic noise grids, it enables us to quantify the differential noise burden across different social groups and urban residents from an environmental justice perspective, aligning with the specific requirements of the United Nations Sustainable Development Goal 11 (Sustainable Cities and Communities) regarding "reducing adverse environmental impacts in cities." This study employs the 2020 high-resolution population density gridded product from the WorldPop project. This dataset integrates multi-source remote sensing imagery, land use types, nighttime light data, and census statistics, utilizing machine learning methods such as random forests to disaggregate national or administrative-level census data to the grid scale. Compared to traditional administrative-level statistical data, it can more finely characterize the spatial heterogeneity of microscale population distribution within cities. This allows us to calculate population-weighted exposure indicators rather than traditional "area-centric" metrics. Additionally, since we are modeling road traffic noise, we utilize OpenStreetMap road network data for extraction. To systematically characterize the distance-decay features of road traffic noise impacts and their differential effects on residential populations at various distances, we constructed multi-scale road buffer zones based on road centerline vectors (buffer distances: 50 m, 100 m, 150 m, 200 m, 250 m, 300 m, etc.). These buffer zone vectors serve both as spatial analysis units for zonal statistics and as sub-region masks to define noise impact zones at different distance scales, supporting subsequent distance-stratified comparative analysis of exposure inequality.

To comprehensively reveal the inequality in the spatial distribution of road traffic noise burden in urban spaces, we introduce the Noise Burden Concentration Gini Coefficient as a core evaluation indicator. This metric quantifies the geographic concentration of the coupled "population × noise" burden from a spatial dimension, effectively identifying spatial hotspots of noise impacts and providing a quantitative basis for prioritizing spatial interventions in urban noise governance. We employ the Gini Coefficient as the measurement indicator, which characterizes the concentration or dispersion of resource allocation through the Lorenz Curve.

Specifically, for the i-th grid cell (both Google AlphaEarth data and WorldPop data exist in grid format), we denote the noise exposure as L_i (in dBA; this study adopts L_{10} as the noise exposure indicator, representing the noise level exceeded 10% of the time, which better reflects the impact of high-noise events on residents) and the population as P_i (in persons). We define the Population-Weighted Exposure (PWE) indicator: $\mathrm{PWE} = \frac{\sum_{i=1}^{N} P_i L_i}{\sum_{i=1}^{N} P_i}$, where N is the total number of valid grid cells. This indicator eliminates the neglect of spatial population density distribution inherent in traditional area-based averages through population weighting, providing a more realistic reflection of actual noise exposure levels experienced by urban resident populations. Furthermore, considering that environmental justice research requires simultaneous attention to the coupled effects of exposure intensity and the scale of affected populations, we introduce the concept of Noise Burden, defined as the product of population and noise exposure: $B_i = P_i \times L_i$, with units denoted as person·dB. This indicator comprehensively reflects the dual information of "how many people" and "how much noise they bear" in a specific area, effectively identifying dual hotspot areas of "high population density + high noise exposure." Based on this, we calculate the total noise burden at the city or study area scale: $B_{total} = \sum_{i=1}^{N} B_i = \sum_{i=1}^{N} P_i L_i$, which represents the absolute magnitude of overall noise impact across the entire city, where N is the total number of valid grid cells.

To calculate the distributional inequality of road traffic noise burden in urban space, we first compute the noise burden $B_i = P_i \times L_i$ for all N valid grid cells, then sort these grid cells in ascending order by noise burden value B_i. It generated the ordered sequence $B_{(1)}, B_{(2)}, \cdots, B_{(N)}$, where $B_{(k)}$ represents the noise burden value of the k-th grid cell after sorting.

Subsequently, we construct the Lorenz curve coordinates. We define the cumulative grid cell proportion $x'k$ as the horizontal coordinate of the Lorenz curve: $x'k = \frac{k}{N}, \quad k = 1,2,\cdots,N$. This coordinate represents the proportion of the first k grid cells (after sorting) relative to the total number of cells, with a range of [0, 1]. We define the cumulative noise burden proportion $y'k$ as the vertical coordinate of the Lorenz curve: $y'k = \frac{\sum i=1^{k} B(i)}{\sum i=1^{N} B(i)}, k = 1,2,\cdots,N$. This coordinate represents the proportion of the cumulative noise burden of the first k grid cells (after sorting) relative to the total burden, also with a range of [0, 1]. Connecting the point set forms the Lorenz curve, which starts at (0, 0) and ends at (1, 1).

In Step 3, we calculate the area under the Lorenz curve. Using the Trapezoidal Rule for numerical integration, we calculate the area A' between the Lorenz curve and the horizontal axis:$A' = \sum_{k=1}^{N} \frac{1}{2} (x'k - x'k-1)(y'k + y'k-1)$, where $x'_0 = 0$ and $y'_0 = 0$. In the ideal case of perfect equal distribution (i.e., all grid cells have exactly the same noise burden), the Lorenz curve coincides with the diagonal line $y' = x'$, and the area under the curve $A' = 0.5$.

Finally, we calculate the Noise Burden Concentration Gini Coefficient: $G = 1 - 2A'$. When the noise burden is perfectly equally distributed, $A' = 0.5$ and $G = 0$; when the noise burden is completely concentrated in a single grid cell, $A' \approx 0$ and $G \approx 1$. Therefore, the theoretical range of G is [0, 1], with higher values indicating greater spatial concentration of noise burden.

The key characteristic of this metric lies in its use of spatial unit proportion (rather than population proportion) as the horizontal coordinate and noise burden value (rather than pure noise exposure value) as the sorting criterion. It makes this indicator more sensitive to dual hotspot areas with "high population density and high noise exposure." Specifically, when this value is high, it indicates that urban noise burden exhibits significant spatial clustering characteristics, meaning that a small number of areas (such as high-density residential areas along arterial roads) bear the vast majority of the noise burden. These areas often represent concentrated manifestations of environmental injustice and should become priority target areas for noise reduction policies and health intervention measures. Conversely, if this value is low, it indicates that the noise burden is relatively dispersed in space and the burden borne by different areas is relatively balanced. In this case, comprehensive city-wide noise control strategies may be needed rather than targeted interventions focused on specific hotspots (i.e., identifying "where is affected").

3. Results

3.1. Foundation model for noise prediction and performance evaluation

We extracted Google AlphaEarth Foundation Model data grids for the Hong Kong region corresponding to the temporal period

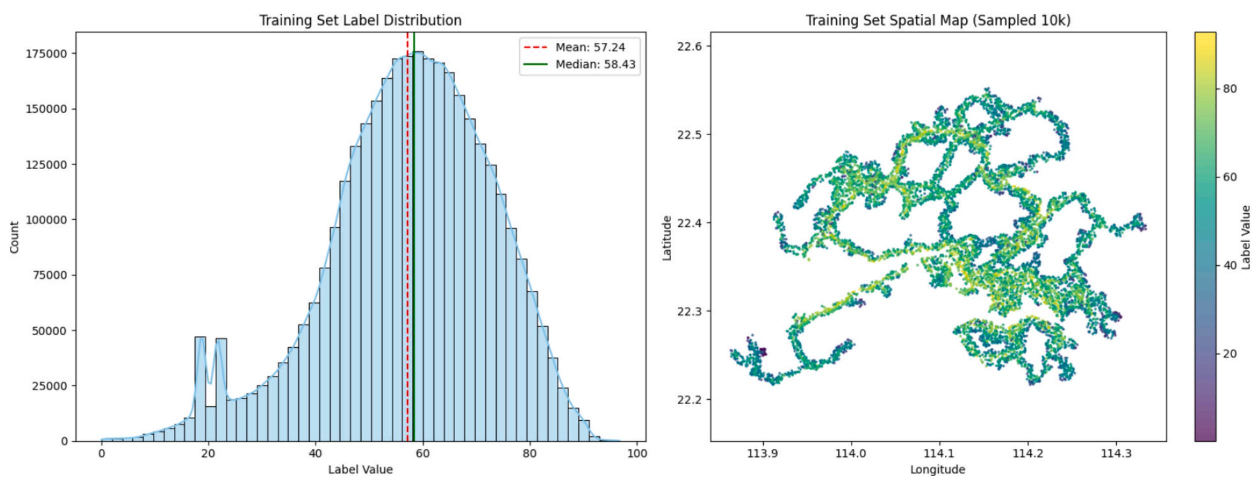

Fig. 4. Noise label density distribution and spatial distribution patterns across the Hong Kong study area.

(2020) of the noise dataset, establishing spatial linkages between the foundation model embeddings and noise grids. The multimodal embeddings enable the model to capture physical processes, including sound reflection, attenuation through vegetation, and meteorological effects, without explicitly modeling the physics. We employed 5-fold cross-validation on the Hong Kong dataset to evaluate model robustness and generalization. R^2, RMSE, and MAE were calculated for each fold. This process yielded a total of 1,603,200 training samples with a label mean of 57.86 ± 9.15 dBA. The dataset was partitioned into a training set (70%, 1,122,240 samples), a validation set (15%, 240,480 samples), and a test set (15%, 240,480 samples) to ensure robust model evaluation and prevent overfitting.

Fig. 4 illustrates the density distribution and spatial distribution patterns of noise labels across the study area. The samples are uniformly distributed throughout the study area, encompassing diverse noise exposure scenarios including metropolitan core areas, suburban residential zones, and countryside environments. This comprehensive spatial coverage ensures that the model learns generalizable patterns across varying urban morphologies and traffic conditions. The label distribution exhibits a roughly normal distribution centered around the mean.

We constructed a specialized Noise Exposure Estimation Network as described in the method Section to capture complex nonlinear relationships between geospatial embeddings and noise levels. The model was trained for 150 epochs using the Adam optimizer with an initial learning rate of 0.001. We employed a batch size of 512 to balance computational efficiency and gradient stability. To enhance model robustness, we applied several data augmentation strategies during training. Random spatial perturbations were introduced by slightly shifting grid cell locations to account for spatial uncertainty in foundation model embeddings. Additionally, we incorporated dropout (rate = 0.3) in the fully connected layers and employed L2 regularization (weight decay = 1e-4) to improve generalization.

As shown in Table 1, the model demonstrates consistent and robust performance across training, validation, and test sets, indicating strong generalization capability without significant overfitting. The model achieves R^2 values exceeding 0.84 across all three sets, with the test set R^2 of 0.8451 demonstrating that the model explains over 84% of the spatial variance in urban noise levels. The Root Mean Square Error (RMSE) values range from 5.73 to 6.15 dBA. The Mean Absolute Error (MAE) remains below 3 dBA across all sets (2.67-2.92 dBA), which is well within the acceptable range for urban noise mapping applications, as measurement uncertainties in field noise monitoring typically range from 2-5 dBA (Wei et al., 2016).

Fig. 5 presents scatter plots comparing true values and predicted values for the training, validation, and test sets, along with the spatial distribution of Mean Absolute Error (MAE). The strong linear correlation between predicted and observed noise levels across all datasets confirms the model's ability to capture the underlying physical relationships between urban built environment characteristics (encoded in the foundation model embeddings) and acoustic exposure.

Within the study area (Hong Kong), the model successfully explains over 80% of spatial noise variance with MAE below 3 dBA, meeting the requirements for practical urban noise modeling applications. The spatial distribution of prediction errors reveals that higher MAE values are primarily concentrated in transitional zones between high-traffic arterial roads and low-noise residential areas, where sharp noise gradients challenge model interpolation. Error analysis by noise level categories shows that the model performs best in the moderate noise range (55-75 dBA), where the majority of urban areas fall, with slightly degraded performance in extremely quiet (<45 dBA) and extremely noisy (>85 dBA) regions due to limited training samples in these extremes. This suggests that future improvements could benefit from stratified sampling or synthetic data augmentation techniques to enhance the representation of tail distribution scenarios.

3.2. *Cross-city transfer and generalization assessment*

Although the model has demonstrated high usability in the Hong Kong region, its reusability in other cities remains uncertain. To evaluate cross-domain generalization capability, we first selected Shenzhen. The city with a similar spatial morphology and regional context to Hong Kong, as the transfer learning experimental scenario. We input AlphaEarth TIFF data from Shenzhen's metropolitan area into the Hong Kong-trained model for zero-shot inference (without any retraining or parameter updates), outputting traffic noise exposure predictions for spatial points within the study region. As shown in Fig. 6, the transfer results exhibit strong spatial plausibility. The predicted high values are primarily distributed along both sides of roads (we only take the prediction value along with the road into consideration), while lower predicted values are concentrated in green spaces and other non-impervious surfaces and non-hardened road areas, indicating that the model can effectively characterize noise exposure conditions in neighboring regions. Simultaneously, the frequency distribution of decibel values is mainly concentrated around 50 dBA and 70 dBA, with a mean value of approximately 61.66 dBA.

Furthermore, we conducted extrapolation tests across a broader range of geographic locations and city types, including inland cities in China such as Wuhan and Chengdu, as well as several cities in other countries (e.g., Paris, London, Toronto, Tokyo, and Boston).

Table 1
Model Performance Metrics Across Different Dataset Splits.

Metrics	Training Set	Validation Set	Test Set
Loss	2.26	2.49	2.50
R^2	0.87	0.85	0.85
MAE dBA	2.67	2.91	2.92
RMSE dBA	5.73	6.12	6.15

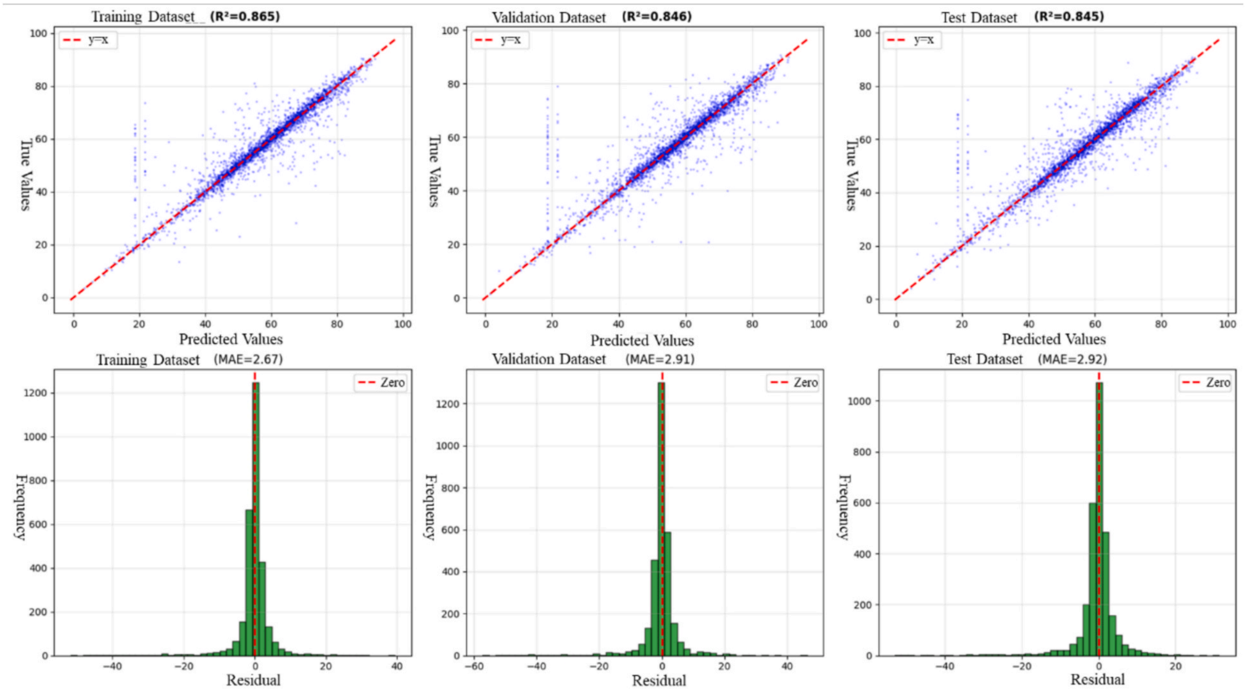

Fig. 5. Scatter plots of true vs. predicted noise values and spatial distribution of MAE across training, validation, and test sets.

Taking Wuhan as an example (Fig. 7), the predicted value distribution exhibits an overall unimodal characteristic (focused around 60 dBA), but the spatial continuity of high-noise bands near roads is relatively weak, and the high-value corridors formed along roads (shown in red) are less visually distinct compared to the Shenzhen case. This suggests that the model may experience cross-domain degradation in cities with greater geographic distance and more pronounced differences in urban morphology and acoustic conditions, necessitating subsequent few-shot calibration or stronger domain adaptation strategies to further enhance robustness.

Fig. 8 presents two typical high-noise exposure areas. Through comparative observations of street-view and remote sensing imagery, we identified several key factors: (1) multi-level sound sources generated by transportation infrastructure such as elevated viaducts; (2) differential acoustic absorption by varying land cover types; (3) sound wave reflections caused by reflective building facades such as glass curtain walls; (4) acoustic canyon effects formed by the three-dimensional built environment. The spatial superposition of these elements renders traditional noise models based on simplified assumptions inadequate, highlighting the necessity of employing deep learning to capture complex nonlinear relationships.

3.3. *Cross-city traffic noise exposure and inequality assessment*

To overcome the limitations of traditional area-based average exposure indicators that neglect the spatial heterogeneity of population distribution, this study employs Population-Weighted Mean Exposure (L_w) as a comprehensive indicator reflecting the actual exposure levels experienced by "ordinary residents." This metric calculates the weighted average of noise exposure values across grid cells according to their population counts, providing a more realistic characterization of the average noise levels borne by urban resident populations. The calculation formula is: $L_w = \frac{\sum_{i=1}^{N} P_i L_i}{\sum_{i=1}^{N} P_i}$.

We computed this metric for nine cities: Berlin, Paris, London, Toronto, Houston, Chengdu, Tokyo, Boston, and Shenzhen in Fig. 9. We constructed buffer zones at different distances from OSM roads to calculate noise exposure conditions. Taking the 150-meter buffer zone as an example, different cities exhibit significant differentiation patterns. Berlin's L_w is 72.6 dBA, Paris 67.9 dBA, London 74.4 dBA, Toronto 73.1 dBA, Houston 59.4 dBA, Chengdu 43.6 dBA, Tokyo 66.7 dBA, Boston 76.2 dBA, and Shenzhen 69.1 dBA. These values not only reflect the average noise levels actually borne by residents in each city but also reveal the deep connections among urban traffic patterns, built environment characteristics, and noise exposure.

From the absolute numerical distribution of L_w, the sample cities can be divided into four tiers: Very low exposure cities ($L_w < 50$ dBA) include only Chengdu (43.6 dBA), with this anomalously low value potentially related to the sampling area containing extensive park green spaces and low-traffic residential zones; Low exposure cities ($50 \leq L_w < 65$ dBA) include Houston (59.4 dBA), whose low exposure level is closely related to North American low-density, wide-road urban morphology; Moderate exposure cities ($65 \leq L_w < 73$

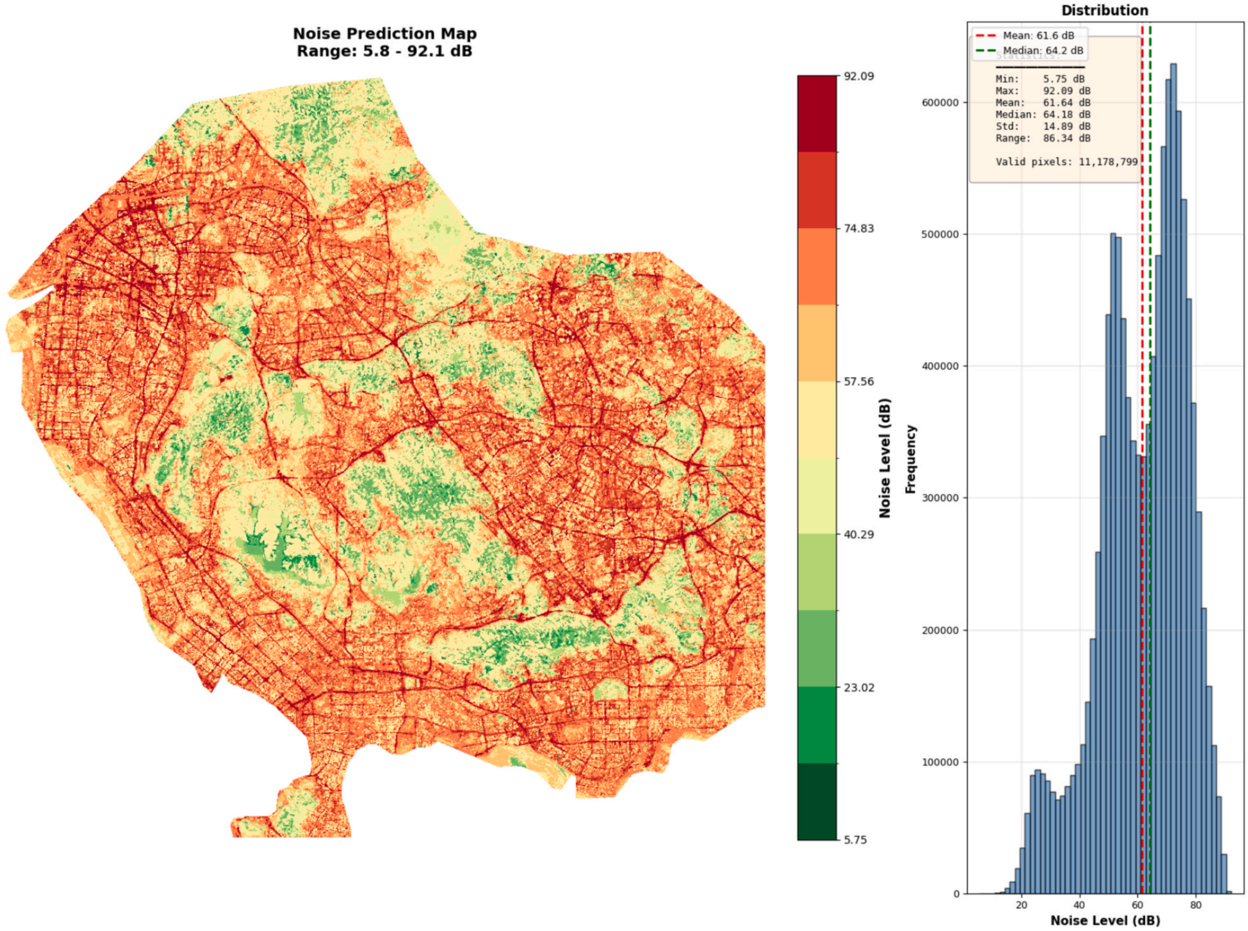

Fig. 6. Road traffic noise prediction results and decibel value frequency distribution for Shenzhen (adjacent to Hong Kong).

dBA) include Tokyo (66.7 dBA), Paris (67.9dBA) and Shenzhen (69.1 dBA), reflecting a balance between high-density development and modernized traffic management; High exposure cities ($L_w \geq 73$ dBA) include Berlin,Toronto, London, and Boston, with these cities' L_w values exceeding WHO-recommended daytime noise guideline values by approximately 20 dB.

Using the same 150-meter buffer zone as an example, the Noise Burden Concentration Gini Coefficient (G) shows significant differences across cities in Fig. 10. Berlin's G is 0.209, Paris 0.3006, London 0.406, Toronto 0.409, Houston 0.467, Chengdu 0.169, Tokyo 0.288, Boston 0.571, and Shenzhen 0.428. Notably, the sampling areas in this study are all metropolitan core areas of respective cities with similar spatial coverage, meaning that the above G value differences primarily reflect essential differences in the internal distribution patterns of noise burden within cities.

From the numerical distribution of G, sample cities can be divided into three tiers: Low concentration cities (G $<$ 0.3) include Chengdu (0.169), Berlin (0.209), and Tokyo (0.288), indicating that noise burden in these cities exhibits relatively homogeneous spatial distribution; Moderate concentration cities ($0.3 \leq G < 0.45$) include London (0.406), Paris (0.3006), Toronto (0.409), and Shenzhen (0.428), where noise burden begins to show some degree of spatial concentration tendency; High concentration cities ($G \geq 0.45$) include Houston (0.467) and Boston (0.571), indicating that noise burden is highly concentrated in a small number of grid cells, representing significant environmental injustice issues.

Of particular concern is Boston (Fig. 11), which has both a high L_w of 76.2 dBA and the highest noise burden concentration (G = 0.571). This dual characteristic of “high exposure + high concentration” indicates that Boston not only suffers from severe overall noise problems but also exhibits extremely unbalanced noise burden concentrated in a few areas. The differences in these distribution patterns reflect the deep coupling relationships among urban morphology, traffic network structure, and population distribution patterns, providing quantitative evidence for understanding urban environmental justice issues and offering a scientific basis for formulating differentiated, targeted noise governance strategies.

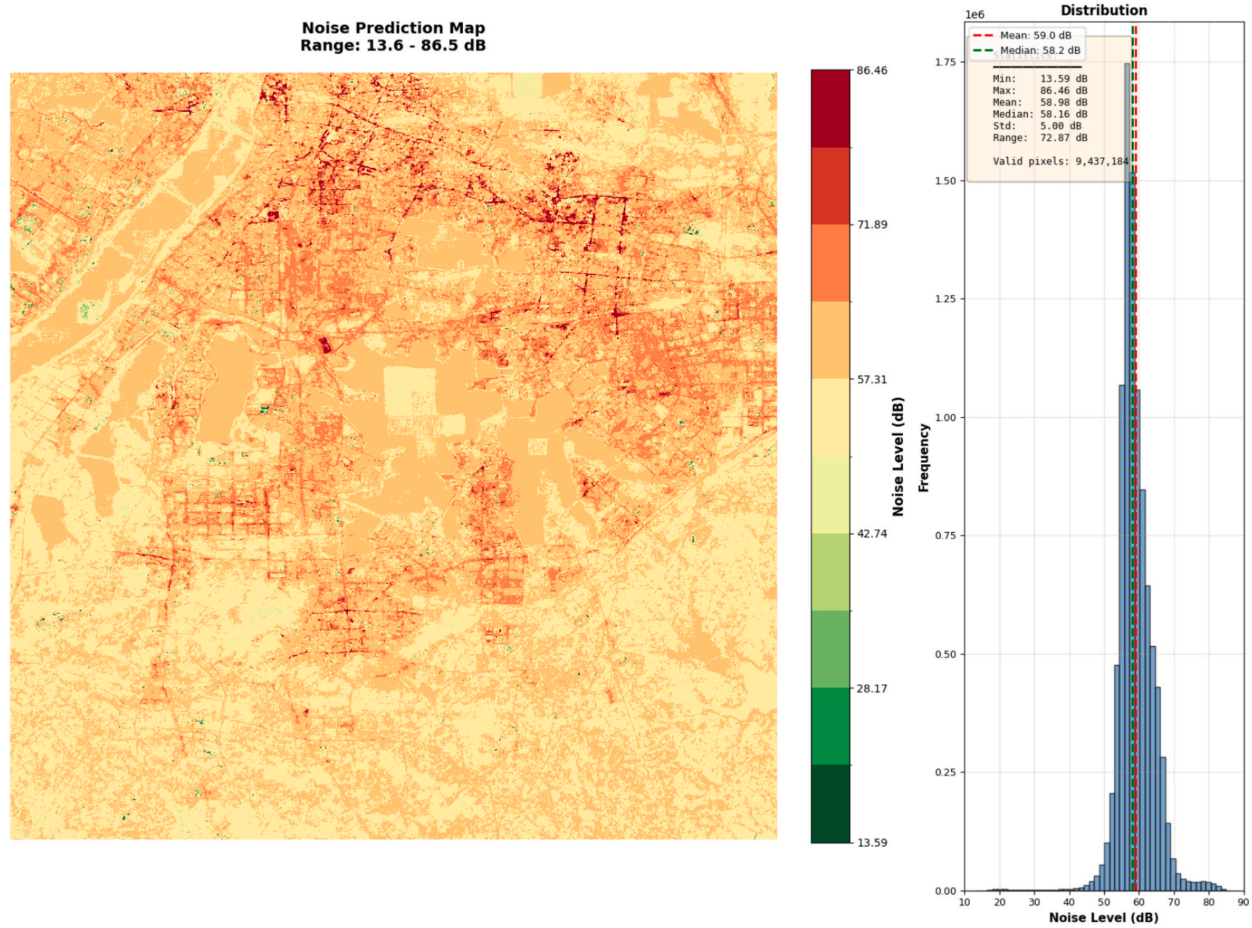

Fig. 7. Road traffic noise prediction results and decibel value frequency distribution for Wuhan.

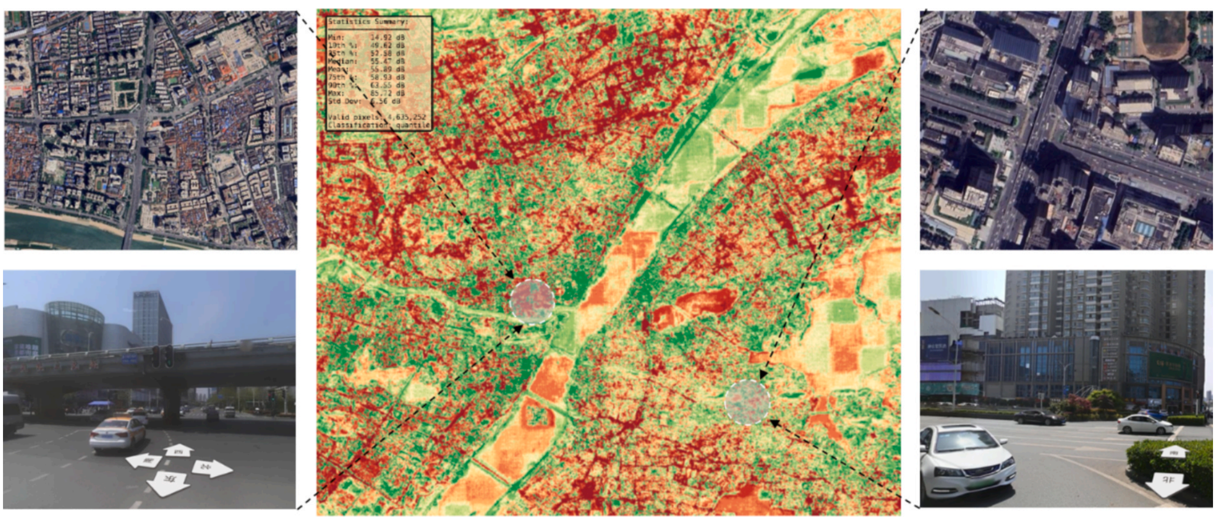

Fig. 8. Urban traffic noise modeling faces complex and diverse spatial scenarios.

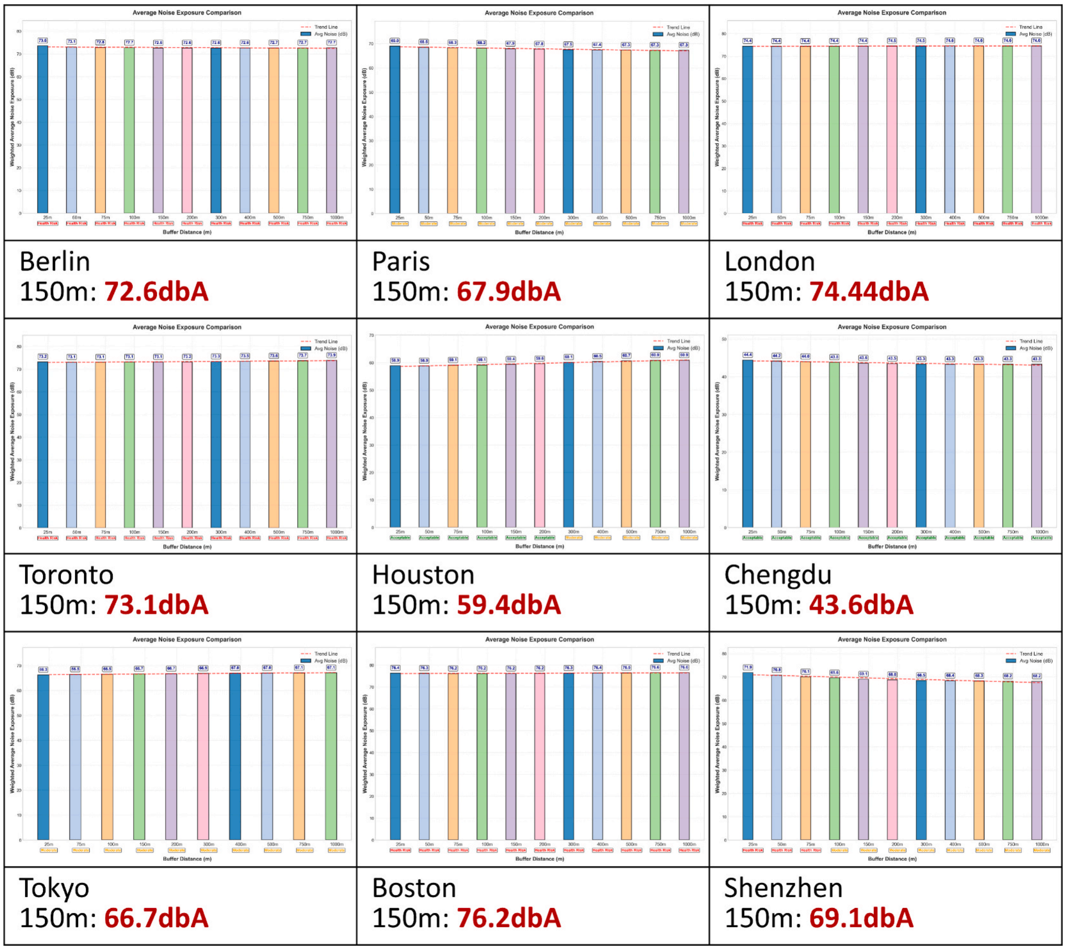

Fig. 9. Population-weighted mean noise exposure across cities under different road buffer distances.

Fig. 10. Noise burden concentration Gini coefficients across cities under different road buffer distances.

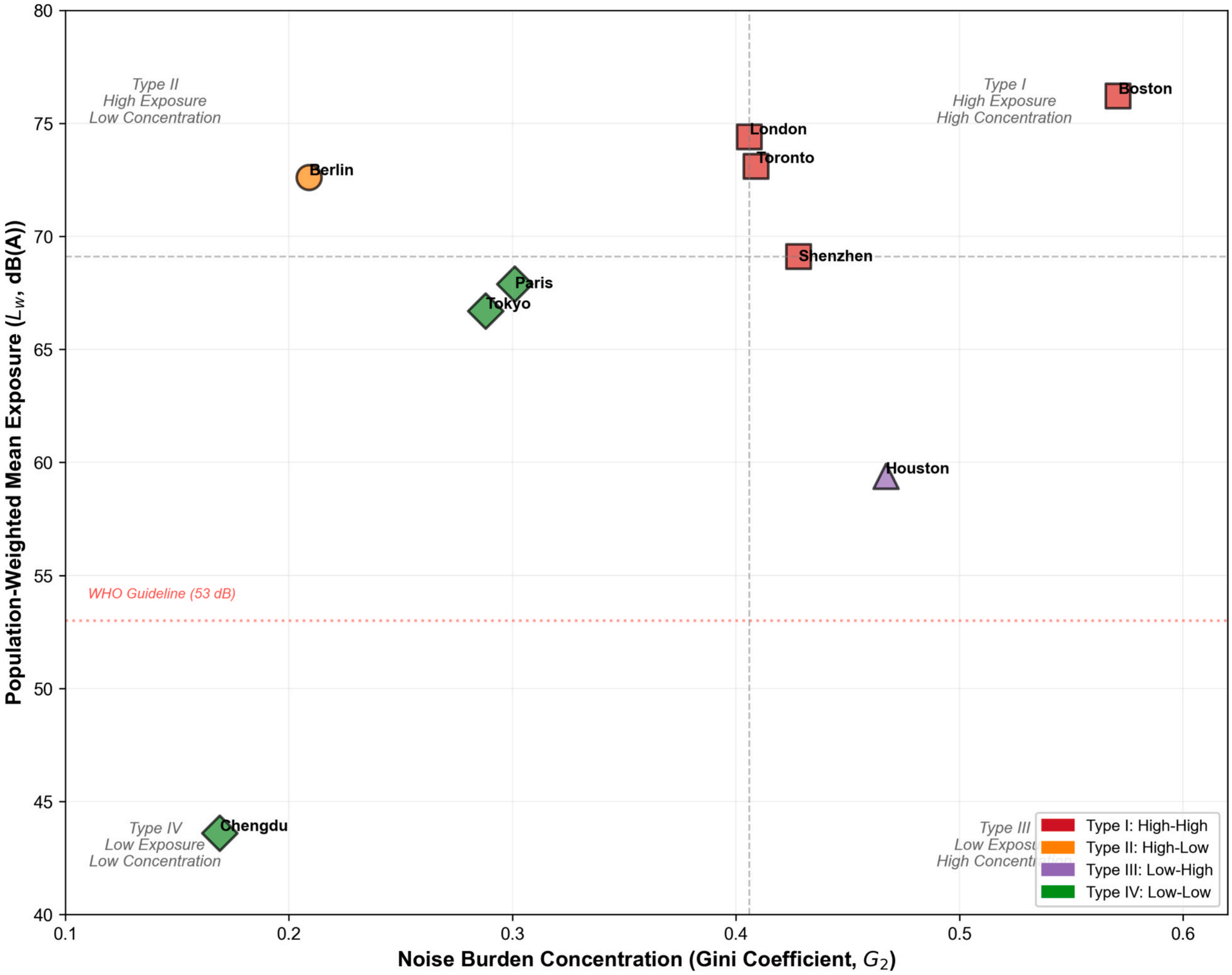

Fig. 11. Scatter Plot of Noise Burden Concentration and Population-Weighted Exposure.

4. Conclusion

Foundation models represent large-scale, highly generalizable AI systems pretrained on massive datasets that can be readily adapted to diverse downstream tasks across domain boundaries. They can be easily fine-tuned for numerous downstream applications or even directly applied to novel tasks through few-shot or zero-shot learning paradigms. Geospatial foundation models have demonstrated excellent performance in standardized remote sensing tasks such as land cover classification and crop yield prediction (Brown et al., 2025), while their application in complex urban environment modeling remains largely unexplored. Existing research has primarily focused on biophysical variables with direct spectral signatures (e.g., vegetation indices, land surface temperature), whereas noise pollution exhibits stronger socioeconomic correlations with human activities. Leveraging geospatial foundation models for urban environmental health research, particularly for complex coupled problems like noise, represents an entirely new frontier. Unlike vegetation greenness (which can be directly measured through the NDVI), noise is an emergent property of coupled urban systems, requiring integration of human activity patterns (traffic), built environment geometry (buildings), and natural attenuation features (vegetation, topography). Sufficient information cannot be obtained solely from RGB visual features, necessitating multi-modal physical observations.

This study advances multiphysics-based few-shot noise modeling by applying geospatial foundation models to the characterization of the urban acoustic environment. Moreover, it also validates the effectiveness of geospatial foundation models in environmental exposure applications. Through cross-city transfer learning spanning three continents with diverse urban structures and systematic inequality assessment, we identified disproportionate spatial distribution patterns of noise exposure across different cities, providing valuable insights for urban planning and noise mitigation strategies. Our framework demonstrates several key contributions:

First, we combined foundation model embeddings with task-specific heads to achieve noise prediction accuracy with $R^2 > 0.84$. This means we can leverage geospatial foundation models to address the critical data scarcity challenge in the Global South, where 90% of urban population growth occurs but environmental monitoring infrastructure remains inadequate. This paradigm enables city-wide noise mapping previously infeasible due to the prohibitive costs of sensor deployment or physics-based simulation. Second, zero-shot transfer experiments across nine global cities (Berlin, Paris, London, Toronto, Houston, Chengdu, Tokyo, Boston, Shenzhen) revealed spatially plausible noise predictions without local calibration. Near-distance transfers (Hong Kong → Shenzhen) exhibited strong spatial coherence, while far-distance applications showed performance degradation in cities with substantially different urban structures. It also highlighted the necessity of local data fine-tuning and the characteristic differences in acoustic environments across different cities. Third, population-weighted exposure analysis (L_w) and Gini coefficient assessment (G) provided a systematic quantification of noise burden inequality. We identified a spectrum ranging from relatively equitable cities (Chengdu: G=0.169, L_w=43.6 dBA) to severely inequitable contexts (Boston: G=0.571, L_w=76.2 dBA). It demonstrated how the coupled "high exposure + high concentration" pattern in certain cities reflects deep structural inequities in transportation planning and residential segregation.

Despite these advances, there are still several limitations in this research. First, ground-truth validation data are currently available only for Hong Kong, limiting our ability to rigorously quantify cross-city prediction accuracy. While spatial plausibility checks and physical reasoning support the transferred predictions, differences in urban characteristics across cities (such as building density, urban morphology, or traffic patterns) still limit the model's generalizability. Future work should explore few-shot calibration strategies that leverage limited local noise measurements (e.g., 100-500 samples per city) to adapt the pre-trained model.

Funding Acknowledgments

This research was supported by a grant from the Smart Traffic Fund (Grant no: PSRI/44/2208/PR). It was also supported by grants from the Hong Kong Research Grants Council (General Research Fund Grant no. 14605920, 14606922, 14603724; Collaborative Research Fund Grant no. C4023-20GF; Research Matching Grants RMG 8601219, 8601242, 3110151), a grant from the Research Committee on Research Sustainability of Major Research Grants Council Funding Schemes (3133235) of the Chinese University of Hong Kong (CUHK), a grant from the 1+1+1 CUHK-CUHK(SZ)-GDSTC Joint Collaboration Fund (4760974, 2025A0505000062), an Innovation and Technology Fund (ITF) (ITS/018/24SC), a grant from the Vice-Chancellor's One-off Discretionary Fund (Smart and Sustainable Cities: City of Commons) (4930787) of CUHK. Open Research Fund Program of Shenzhen Key Laboratory of Spatial Smart Sensing and Services (GEM-Lab-2024010), Open Fund of National Engineering Research Center of Geographic Information System (NERCGIS-202405), Open Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (24I04) and the grant from the National Natural Science Foundation of China (42501579). The funders had no role in the study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

CRediT authorship contribution statement

Yan Zhang: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Quan Qin:** Visualization, Validation, Data curation. **Hairu Nie:** Visualization, Validation. **Mei-Po Kwan:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Sijia He:** Visualization, Validation. **Entong Ke:** Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Alvarez, C.I., Ulloa Vaca, C.A., Echeverria Llumipanta, N.A., 2025. Machine learning for urban air quality prediction using Google AlphaEarth Foundations satellite embeddings: A case study of Quito, Ecuador. Remote Sens. (Basel) 17 (20), 3472.

Bello, J.P., Silva, C., Nov, O., et al., 2019. Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. Commun. ACM 62 (2), 68–77.

Benocci, R., Bisceglie, A., Angelini, F., et al., 2020. Influence of traffic noise from local and surrounding areas on high-rise buildings. Appl. Acoust. 166, 107362.

Brown C F, Kazmierski M R, Pasquarella V J, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv:2507.22291, 2025.

Casey, J.A., Morello-Frosch, R., Mennitt, D.J., et al., 2017. Race/ethnicity, socioeconomic status, residential segregation, and spatial variation in noise exposure in the contiguous United States. Environ. Health Perspect. 125 (7), 077017.

Clark, C., Crombie, R., Head, J., et al., 2012. Does traffic-related air pollution explain associations of aircraft and road traffic noise exposure on children's health and cognition? A secondary analysis of the United Kingdom sample from the RANCH project. Am. J. Epidemiol. 176 (4), 327–337.

Chan, K., Chan, E., Lai, J., Lee, B., Lee, C., 2023. The development of noise mapping in Hong Kong. International Institute of Acoustics and Vibration (IIAV), Prague, Czech Republic.

Chen, Z., Fei, T., Xiao, J., et al., 2025. Estimating urban noise levels from Multi-Scale and Multi-Spectral remote sensing imagery. Int. J. Appl. Earth Obs. Geoinf. 143, 104818.

Dzhambov, A.M., Lercher, P., 2019. Road traffic noise exposure and depression/anxiety: an updated systematic review and meta-analysis. Int. J. Environ. Res. Public Health 16 (21), 4134.

Hahad, O., Kuntic, M., Al-Kindi, S., et al., 2025. Noise and mental health: evidence, mechanisms, and consequences. J. Eposure Sci. Environ. Epidemiol. 35 (1), 16–23.

Hammer, M.S., Swinburn, T.K., Neitzel, R.L., 2014. Environmental noise pollution in the United States: developing an effective public health response. Environ. Health Perspect. 122 (2), 115–119.
Hu, J., Shen, L., Sun, G., 2018:. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 7132–7141.
Huang, J., Fei, T., Kang, Y., et al., 2024. Estimating urban noise along road network from street view imagery. Int. J. Geogr. Inf. Sci. 38 (1), 128–155.
Huang, W., Zhang, D., Mai, G., et al., 2023. Learning urban region representations with POIs and hierarchical graph infomax. ISPRS J. Photogramm. Remote Sens. 196, 134–145.
Janowicz, K., Gao, S., McKenzie, G., et al., 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. Int. J. Geogr. Inf. Sci. 34 (4), 625–636.
Janowicz, K., Mai, G., Huang, W., et al., 2025. GeoFM: how will geo-foundation models reshape spatial data science and GeoAI? Int. J. Geogr. Inf. Sci. 39 (9), 1849–1865.
Kang, J., Aletta, F., Gjestland, T.T., et al., 2016. Ten questions on the soundscapes of the built environment. Build. Environ. 108, 284–294.
Korpilo, S., Nyberg, E., Vierikko, K., et al., 2024. Landscape and soundscape quality promote stress recovery in nearby urban nature: A multisensory field experiment. Urban For. Urban Green. 95, 128286.
Kumar, A., Kolhe, N.P., 2025. Develop a land use regression model to predict noise levels in the urban residential neighbourhood of Bhopal. Development and Sustainability, India. Environment, pp. 1–27.
Leach, F.C.P., Yang, J., Shankar, V., et al., 2024. The impact of a low traffic neighbourhood intervention on urban noise measured with low-cost sensors in Oxford, UK. Transp. Res. Part D: Transp. Environ. 134, 104357.
Lian, A., Zhang, Y., Liu, Y., et al., 2026. Exploring sentiment dynamics and their driving factors in megacity residents' environmental complaints through deep learning and multimodal data. Appl. Geogr. 186, 103806.
Liu, P., Biljecki, F., 2022. A review of spatially-explicit GeoAI applications in Urban Geography. Int. J. Appl. Earth Obs. Geoinf. 112, 102936.
Lu, F., Cheng, S., Wang, P., 2025. GeoAI enabled urban computing: status and challenges. Ann. GIS 1–19.
Mai, G., Huang, W., Sun, J., et al., 2024. On the opportunities and challenges of foundation models for geoai (vision paper). ACM Trans. Spatial Algorithms Syst. 10 (2), 1–46.
Münzel, T., Schmidt, F.P., Steven, S., et al., 2018. Environmental noise and the cardiovascular system. J. Am. Coll. Cardiol. 71 (6), 688–697.
Münzel, T., Sørensen, M., Daiber, A., 2021. Transportation noise pollution and cardiovascular disease. Nat. Rev. Cardiol. 18 (9), 619–636.
Qin, Q., Ai, T., Xu, S., et al., 2025. Learning dual context aware POI representations for geographic mapping. Int. J. Appl. Earth Obs. Geoinf. 142, 104683.
Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., & Bojanowski, P. (2025). DINOv3. arXiv preprint arXiv:2508.10104. https://doi.org/10.48550/arXiv.2508.10104.
Song, L., Liu, D., Kwan, M.P., et al., 2024. Machine-based understanding of noise perception in urban environments using mobility-based sensing data. Comput. Environ. Urban Syst. 114, 102204.
Tang, M., Li, X., 2024. Growing disparities in transportation noise exposure across major US cities over time. Transp. Res. Part D: Transp. Environ. 136, 104430.
Thompson, R., Smith, R.B., Karim, Y.B., et al., 2022. Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. Environ. Int. 158, 106905.
Wang, H., Yan, X., Chen, J., et al., 2024. Urban noise exposure assessment based on principal component analysis of points of interest. Environ. Pollut. 342, 123134.
Wang, H., Wu, Z., Chen, J., et al., 2022. Evaluation of road traffic noise exposure considering differential crowd characteristics. Transp. Res. Part D: Transp. Environ. 105, 103250.
Wei, W., Van Renterghem, T., De Coensel, B., et al., 2016. Dynamic noise mapping: A map-based interpolation between noise measurements with high temporal resolution. Appl. Acoust. 101, 127–140.
Lan, Z., Cai, M., 2021. Dynamic traffic noise maps based on noise monitoring and traffic speed data. Transp. Res. Part D: Transp. Environ. 94, 102796.
Li, G, Huang, Z, Fei, T, et al., 2025. Listen to the Road: Acoustic Traffic Monitoring on Edge Platforms via Lightweight Noise Spectrogram Transformer (LNST)[J]. Pervasive and Mobile. Computing 102132.
Li, Y., Zhang, Y., Chen, M., 2022. Highway Vehicle Route Reconstruction Using Sparse and Noisy Communication Base Station Fingerprints. IEEE Sens. J. 22 (22), 22040–22052.
Wang, J., Wang, Z., Li, C., et al., 2025. Transforming Noise Control Stations into Soundscape Resource Monitoring Networks: An Adaptive Location Recommendation. Environ. Sci. Technol.
Welch, D., Shepherd, D., Dirks, K.N., et al., 2023. Health effects of transport noise. Transp. Rev. 43 (6), 1190–1210.
Yang, W., He, J., He, C., et al., 2020. Evaluation of urban traffic noise pollution based on noise maps. Transp. Res. Part D: Transp. Environ. 87, 102516.
Yin, X., Fallah-Shorshani, M., McConnell, R., et al., 2020. Predicting fine spatial scale traffic noise using mobile measurements and machine learning. Environ. Sci. Technol. 54 (20), 12860–12869.
Yuan, M., Yin, C., Sun, Y., et al., 2019. Examining the associations between urban built environment and noise pollution in high-density high-rise urban areas: A case study in Wuhan. China. Sustainable Cities and Society 50, 101678.
Zhang, Y., Chen, N., Du, W., et al., 2021. Multi-source sensor based urban habitat and resident health sensing: A case study of Wuhan. China. Building and Environment 198, 107883.
Zhang, Y., Zhang, F., Chen, N., 2022. Migratable urban street scene sensing method based on vision language pre-trained model. Int. J. Appl. Earth Obs. Geoinf. 113, 102989.
Zhang, Y., Liu, P., Biljecki, F., 2023. Knowledge and topology: A two layer spatially dependent graph neural networks to identify urban functions with time-series street view image. ISPRS J. Photogramm. Remote Sens. 198, 153–168.
Zhang, Y., Kwan, M.P., Ma, H., 2024. Sensing noise exposure and its inequality based on noise complaint data through vision-language hybrid method. Appl. Geogr. 171, 103369.
Zhang, Y, Ke, E, Kwan, M P, et al., 2026. Multi-frequency street-level urban noise modeling and mapping through street view and remote sensing image fusion[J]. Computers, Environment and Urban. Systems 126, 102401.
Zhang, Y., Kwan, M.P., 2025. Considering human mobility is essential for accurate environmental exposure assessment. Science Bulletin.
Zhuang, Y., Kang, Y., Fei, T., et al., 2024. From hearing to seeing: Linking auditory and visual place perceptions with soundscape-to-image generative artificial intelligence. Comput. Environ. Urban Syst. 110, 102122.
Yihan Zhao, Haoyue Jiao, Teng Fei, Jing Huang, Ningdong You, Spatial distribution prediction and scale effect analysis of urban daytime noise based on remote sensing images: a case study of Chengdu, Journal of Environmental Management, Volume 397, 2026, 128379, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2025.128379.