



Revealing the built environment impacts on truck emissions using interpretable machine learning

Tongtong Shi^a, Meiting Tu^{a,*}, Ye Li^a, Haobing Liu^a, Dominique Gruyer^b

^a The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, PR China

^b IFSTTAR, University Gustave Eiffel, 77420 Champs sur Marne, France

ARTICLE INFO

Keywords:

Urban freight transport
Truck emissions
Built environment
Nonlinear effects
Interpretable machine learning

ABSTRACT

Understanding the factors influencing truck emissions remains critical for sustainable urban freight transport development. However, ignoring spatiotemporal and policy heterogeneity may lead to inaccurate predictions for specific regions and misinterpretation of outcomes. This study develops a comprehensive framework to analyze the nonlinear effects of the built environment on heavy-duty diesel truck emissions, utilizing large-scale GPS data from Shanghai, China. We introduce an interpretable predictive model that integrates random effects with a light gradient boosting machine to account for spatiotemporal and policy influences. The results show that proposed model outperforms baseline by 15%–20%, with an improvement exceeding 17% in the more complex tasks of localized predictions in central urban areas. Land use and road design factors contribute 72.26% to truck emissions, with industrial land density as the primary driver. Furthermore, the relationship between these factors and pollution emissions exhibits pronounced non-linearity, with threshold effects that vary under various policy restrictions.

1. Introduction

As the dominant mode of road freight transport, trucks account for a substantial proportion of urban freight activity while also being a major source of pollutant emissions (Zhang et al., 2025). Heavy-duty diesel trucks (HDDTs), despite constituting a small fraction of registered vehicles, are a significant source of global greenhouse gas emissions (Cheng et al., 2022; Liu et al., 2021). On-road diesel vehicles contribute approximately 20% of global anthropogenic nitrogen oxide (NO_x) emissions (Anenberg et al., 2017), with heavy-duty trucks alone accounting for an estimated approximately 20 million tons annually (Stohl et al., 2015). This disproportionate impact is evident in countries like China, where HDDTs represent merely 3% of the vehicle fleet while they are responsible for nearly 74% of total NO_x emissions and 52.4% of particulate matter (PM2.5) emissions from motor vehicles (Cheng et al., 2023). Similarly, in California, HDDTs comprise only 7% of registered vehicles but generate one-third of all road transportation NO_x emissions (Ramirez-Ibarra and Saphores, 2023). To effectively mitigate diesel truck emissions, it is crucial to thoroughly understand the key influencing factors (W. Li et al., 2024). Therefore, it is imperative for policymakers to elucidate the complex interrelationships between HDDT emissions distribution and these critical factors to develop targeted and efficacious intervention strategies.

Previous research on HDDT emissions has primarily focused on several traditional factors, including engine characteristics, vehicle load, driving conditions, and after-treatment system efficiency (Chen et al., 2018; Tang et al., 2022). Recently, an increasing body of

* Corresponding author at: School of Transportation Engineering, Tongji University, Add. 4800 Cao'an Road, Shanghai 201804, PR China.

E-mail addresses: tongtongshi@tongji.edu.cn (T. Shi), meitingtu@tongji.edu.cn (M. Tu), jamesli@tongji.edu.cn (Y. Li), liuhaobing@tongji.edu.cn (H. Liu), dominique.gruyer@univ-eiffel.fr (D. Gruyer).

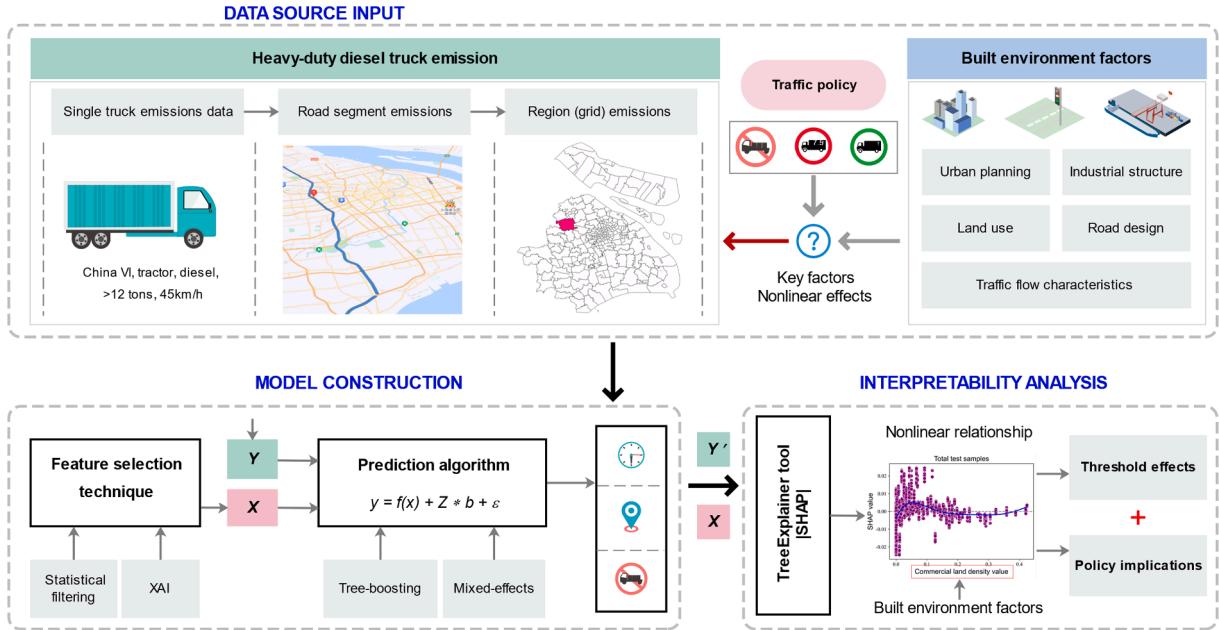


Fig. 1. Illustration of the analysis process for factors influencing freight truck emissions.

evidence suggests that characteristics of the built environment—such as land use, urban density, and transportation infrastructure—can exert a substantial influence on passenger vehicle emissions (Feng et al., 2022a; Yang and Cao, 2018; Zhao et al., 2023). However, the majority of this research has primarily targeted passenger transport and carbon dioxide emissions, while urban freight planning has received comparatively limited attention (Peng et al., 2024). Due to the difficulty in obtaining precise data, the exact relationship between these built environment factors and freight truck emissions remains unclear (Cheng et al., 2023).

Furthermore, most studies assume a linear relationship between these factors and traffic emissions (Cheng et al., 2023; Jia, 2021), but this oversimplifies the complex, nonlinear dynamics of urban traffic systems (Wu et al., 2023; Ding et al., 2022). Machine learning (ML) techniques, particularly tree-boosting algorithms, have demonstrated superior predictive performance, with boosted decision trees widely regarded as one of the most effective off-the-shelf nonlinear learning methods (Sigrist, 2023a). However, several important yet unresolved issues persist: a) *How to balance predictive performance, critical information retention, and interpretability in feature selection?* High feature correlation within tree-boosting models complicates the identification of primary decision drivers at split nodes because similar features exhibit complex substitution and interaction effects (Lundberg et al., 2020). Models that fail to account for these effects risk producing predictions that are flawed or misleading (Houssein et al., 2023); b) *How to address spatial-temporal heterogeneity in traffic data using tree-boosting models?* The models rely on the assumption of conditional independence among samples, which limits their applicability to spatial and temporal data. This assumption often results in discontinuous predictions, especially for spatially correlated traffic patterns (Wu et al., 2023). Ignoring the heterogeneity in temporal and spatial dimensions can lead to predictions that are inaccurate or even irrelevant for specific regions (Sigrist, 2023a); c) *How can predictive models incorporate the effects of regulatory policies on freight truck emissions?* Many cities impose temporal restrictions on freight trucks to reduce emissions during peak hours and enforce spatial regulations to restrict their access to certain zones (Lyons et al., 2017; Yang et al., 2023). However, few studies directly analyze how these regulatory policies influence freight truck emissions.

To bridge this gap, this study aims to investigate the nonlinear relationship between the built environment and freight truck emissions (as illustrated in Fig. 1), while accounting for the temporal and spatial heterogeneity of policy effects. Specifically, we achieve this through three key contributions:

- 1) We develop a novel framework that integrates feature selection, predictive modeling, and interpretability analysis. Unlike traditional approaches that treat feature selection as a static preprocessing step, our framework introduces a feature weight updating strategy, dynamically combining statistical filtering with Explainable Artificial Intelligence (XAI) (van Zyl et al., 2024) to identify the optimal feature set. This dynamic process achieves a balanced trade-off between predictive accuracy, feature interpretability, and the dimensionality reduction required for high-dimensional datasets.
- 2) We propose a hybrid modeling approach, termed ST-MELightGB, which integrates mixed-effects models with light gradient boosting machine (LightGBM) (Ke et al., 2017). This approach leverages the spatiotemporal decomposition of heavy truck

restriction policies to cluster units sharing similar random effects, thereby overcoming the conditional independence assumption inherent in tree-based methods. This design paradigm especially enhances localized forecasting accuracy, providing reliable predictive outcomes for region-specific or time-sensitive research applications.

- 3) We extend the utility of predictive models by integrating the spatiotemporal impacts of regulatory policies on freight truck emissions. Specifically, employing the TreeExplainer tool, an implementation of the SHapley Additive exPlanations (SHAP) method, we identify and quantify the nonlinear influences of key factors under varying policy regulation scenarios. Taking Shanghai as a case study, we validate the effectiveness of our proposed framework and the ST-MELightGB model. Based on the experimental findings, we propose targeted policy recommendations to relevant authorities to address freight truck emissions effectively.

The remainder of this paper is organized as follows. [Section 2](#) reviews the relevant literature on the correlation between HDDTs and influencing factors, highlighting gaps in existing research. [Section 3](#) introduces the research methodology. [Section 4](#) provides a description of the data and relevant variables included in the model. [Section 5](#) presents the results in detail. Finally, [Section 6](#) summarizes the main findings, proposes policy recommendations, and discusses potential future research directions.

2. Literature review

2.1. The association between the built environment and traffic emissions

The built environment refers to the human-made surroundings that provide the setting for human activities, encompassing urban planning, land use, infrastructure, and other related elements. A growing body of research explores the relationship between the built environment and traffic emissions, drawing attention from planners and policymakers due to its potential to inform more sustainable urban design ([X. Li et al., 2022](#)) and reduce environmental pollution ([Peng et al., 2024](#)).

In terms of urban planning and land use, one foundational concept in this area is the “3Ds” framework—density, diversity, and design—proposed by ([Cervero and Kockelman, 1997](#)). This framework posits that higher population densities, diverse land uses, and pedestrian-friendly designs encourage sustainable travel behaviors, such as walking, cycling, and public transit use, which in turn lead to reduced vehicle emissions. Subsequent studies have expanded this framework to include factors like destination accessibility and distance to transit, collectively referred to as the “5Ds” ([Ewing and Cervero, 2010](#)). These studies consistently demonstrate that compact, well-connected urban forms correlate with lower vehicle miles traveled (VMT) and consequently reduced greenhouse gas emissions. Notably, an increased mix of land uses significantly lowers weekday emissions, underscoring the critical role of mixed-use planning in urban decarbonization ([Wu et al., 2023](#)). In terms of transportation infrastructure, a negative correlation between road density and motor vehicle emissions suggests that denser road networks enhance mobility and reduce emissions in small cities ([Feng et al., 2022a](#)). Additionally, the availability of public transit stops significantly affects private vehicle emissions; increasing transit stops within a half-mile radius correlates with decreased CO₂ emissions, particularly when the number of stops exceeds 20 ([Wu et al., 2019](#)).

Despite this substantial research on passenger vehicle emissions, there is a notable gap in understanding how the built environment influences truck emissions. Among the limited studies in this area, ([Cheng et al., 2023](#)) utilized the multiscale geographically weighted regression (MGWR) model to analyze the impacts of socioeconomic indicators (such as GDP per capita and urbanization rate) and road attributes (including road density and highway proportion) on NOx emissions from heavy-duty diesel trucks (HDDTs). Their findings indicate spatial heterogeneity in these influences. Similarly, ([Shi et al., 2019](#)) identified varied patterns of spatial inequality in CO₂ emissions across different administrative levels. Furthermore, [Peng et al., 2024](#) investigated urban freight traffic by analyzing built environment factors at a granular spatial resolution, revealing that the density of main roads, built-up areas, proportion of industrial land, and electricity consumption are key determinants of freight pollution.

However, the intersection of these factors with policy interventions adds another layer of complexity. Although extant literature has established correlations between traffic emissions and various elements of the built environment, a consensus remains elusive regarding the differential impacts of spatial configurations, temporal variations, and regulatory constraints.

2.2. Feature selection techniques in the transportation domain

While high-dimensional data offer a wealth of information regarding factors influencing traffic emissions, they simultaneously complicate knowledge discovery and correlation analysis due to the presence of redundant and irrelevant features (in this study, factors influencing traffic emissions), which may obfuscate underlying patterns and relationships ([Zebari et al., 2020](#)). Feature selection addresses this issue by identifying a subset of features that effectively represent the data while excluding irrelevant and redundant ones ([Pan et al., 2023](#)). Traditional feature selection methods can be categorized into three approaches: filter methods, wrapper methods, and embedding methods ([Khaire and Dhanalakshmi, 2022](#)). In the domain of traffic emissions, these methods have been employed to improve model accuracy by identifying key factors traffic emissions ([Cheng et al., 2023; Zebari et al., 2020](#)). By reducing the dimensionality of the feature space, this approach identifies and retains only the most significant factors affecting emissions predictions ([Pan et al., 2023](#)). However, feature selection methods typically aim to balance multiple objectives: reducing dimensionality, improving predictive performance, managing trade-offs associated with computational efficiency, and enhancing data

interpretability (Khaire and Dhanalakshmi, 2022).

In recent years, XAI has emerged as a pivotal field aimed at improving the transparency and interpretability of machine learning models. Various XAI techniques have been applied in transportation research, including global interpretability methods such as partial dependence plots (PDP) and accumulated local effects (ALE), as well as local methods like local interpretable model-agnostic explanations (LIME) and SHAP. These approaches have been widely utilized to explore travel behavior (Tu et al., 2021; Ding et al., 2022) and urban mobility patterns (Sun et al., 2024; Yang et al., 2024). In the realm of traffic emissions, XAI methods have effectively revealed critical insights. For instance, PDP has been used to assess the impact of location accessibility and population density on travel-related CO₂ emissions (Shao et al., 2023). Similarly, SHAP has also been employed to interpret the nonlinear relationships between built environment variables and taxi CO₂ emissions, highlighting the need for adaptive urban management strategies (Wu et al. 2023); ALE has been applied to examine the heterogeneous effects of urban form and road network features on per capita traffic CO₂ emissions (Zhi et al., 2024).

While XAI has primarily been utilized to elucidate decision-making processes in ML models, it has also proven to be an efficient tool for feature selection (Veerappa et al., 2022). Integrating XAI techniques can enhance model interpretability and optimize feature engineering, especially in high-dimensional datasets. Among the existing XAI methods, SHAP has shown superior performance in feature selection over traditional approaches (van Zyl et al., 2024). Several studies in the transportation field have combined XAI with feature selection to tackle tasks such as traffic safety, traffic classification, and traffic speed prediction. SHAP has been integrated into a deep learning-based intrusion detection system for internet of vehicles networks, improving cybersecurity in autonomous driving applications (Oseni et al., 2022). The technique has also been applied to traffic classification, where feature importance ranking enhances classifier accuracy (Khani et al., 2024). In traffic speed prediction, a combination of recursive feature elimination and SHAP has been used to identify influential road segments, with results validated through LIME and traffic flow simulations (Nesa and Yoon, 2024). Additionally, SHAP has been employed to analyze risk factors influencing injury severity in traffic accidents, helping categorize crash data into high-level causal factors using aggregate SHAP scores (Abdulrashid et al., 2024).

To the best of our knowledge, in the field of traffic emissions, most studies utilizing linear models address feature selection through methods such as correlation coefficients (McCaffery et al., 2021) or variance inflation factors (VIF) (Ko et al., 2019). In contrast, studies based on black-box ML models often rely on the models' inherent importance scores (Wu et al. 2023). A notable gap exists in leveraging XAI-driven feature selection specifically for traffic emissions research. This study aims to bridge this gap by investigating the potential of XAI-based methods for feature selection, thus extending their application to this critical area of environmental research.

2.3. Methodological approaches for traffic emissions estimation

Early research on emissions analysis primarily employed linear models such as ordinary least squares (OLS) regression, structural equation models (SEMs), and geographically weighted regression (GWR). While these models offered straightforward interpretability, they were limited by assumptions of linearity and predictor independence, failing to capture complex, nonlinear relationships (Yang et al., 2024; Tu et al., 2021; Ding et al., 2022). To address these limitations, researchers increasingly turned to ML techniques, particularly tree-based methods like random forests and gradient boosting machines (Ding et al., 2024; Li et al., 2024; He et al., 2024). These approaches effectively uncovered nonlinear relationships in high-dimensional data, as demonstrated by (Wu et al., 2023) in their analyses of built environment factors. However, tree-based methods assume conditional independence among samples, which is often violated in hierarchical or clustered data with interdependent observations across time or space. Ignoring temporal and spatial heterogeneity can lead to inaccurate or irrelevant predictions for specific regions (Hajjem et al., 2014).

To mitigate this challenge, linear mixed-effects models were initially employed to account for variability across different groups or clusters (Schielzeth et al., 2020). Subsequently, nonlinear mixed-effects models were developed, combining the flexibility of machine learning with the robustness of mixed-effects modeling. These models effectively accommodate the complex, hierarchical nature of emission data, capturing both fixed and random effects while addressing nonlinearity and interactions (Sigrist, 2023).

2.4. Research gaps

The above review reveals several limitations in existing studies on truck emissions and their influencing factors. Firstly, while some studies have explored the impact of various factors on HDDT emissions, there is a lack of comprehensive analysis that simultaneously considers the complex interplay of traffic conditions, road infrastructure, socioeconomic attributes, and built environment characteristics. Most studies focus on a subset of these factors, potentially overlooking important interactions and confounding effects. Secondly, XAI-based feature selection techniques have gained traction for enhancing model interpretability. However, in the domain of traffic emissions, there is limited research on achieving a balanced trade-off between predictive accuracy, feature interpretability, and the dimensionality reduction required for high-dimensional datasets. This gap indicates a need for more sophisticated approaches that can handle high-dimensional datasets while maintaining interpretability and computational efficiency. Third, while the spatial and temporal heterogeneity of emission patterns and their influencing factors have been acknowledged (Li et al., 2022; Peng et al., 2024), they are not fully addressed in current modeling approaches. Traditional linear models and even advanced machine learning techniques often fail to capture the hierarchical nature of emission data (Schielzeth et al., 2020; Sigrist, 2023), leading to biased or inaccurate results when applied across different spatial and temporal scales. Additionally, the differential effects of policy

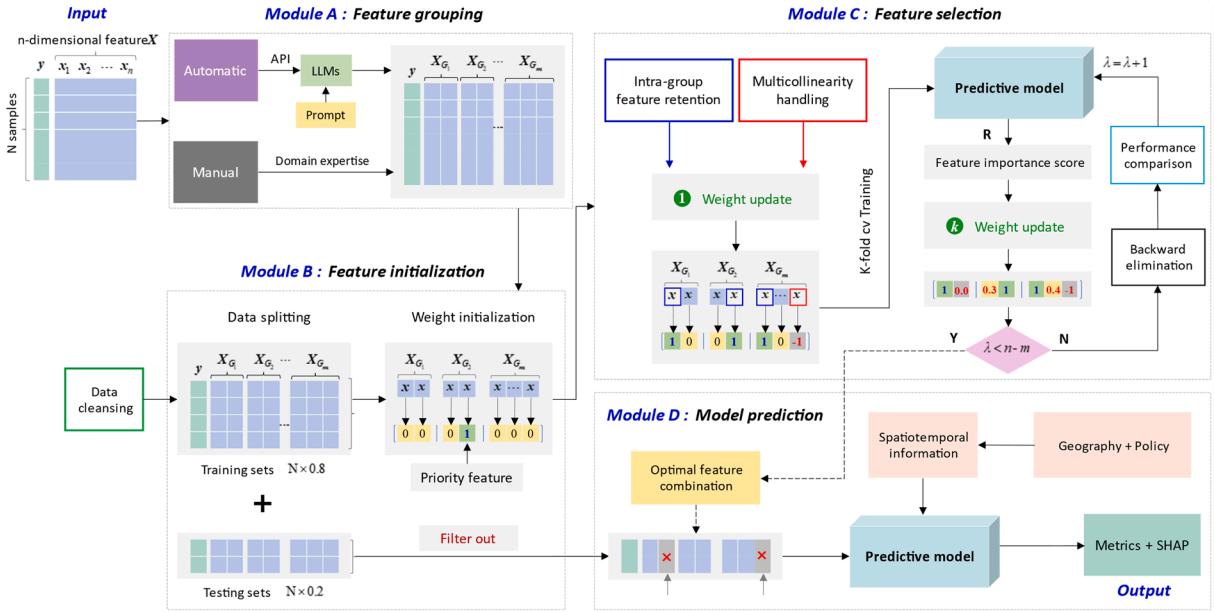


Fig. 2. Overall framework for spatiotemporal analysis of HDDT emissions.

interventions across various spatial configurations and temporal periods are not well understood, limiting the ability to design and implement effective emission reduction strategies.

3. Methodology

3.1. Overview

This study introduces a novel, integrated framework encompassing multiple stages of model development, including data processing, feature selection, model prediction, and result interpretation, as illustrated in Fig. 2. This comprehensive approach is applicable to linear models, machine learning models, and mixed-effects models for processing panel data and has been applied and verified in the experiments.

3.2. Data preprocessing and feature engineering

3.2.1. Feature grouping and initialization

The incorporation of feature grouping serves multiple critical purposes in model development and interpretation. Primarily, it helps identify and manage multicollinearity within and across feature subsets, potentially improving model stability. Additionally, it also reduces the dataset's dimensionality by allowing selective feature retention, minimizing redundancy, and enhancing computational efficiency (Khaire and Dhanalakshmi, 2022). In this study, the framework supports both manual and automatic classification for effective feature grouping, denoted as $\text{GroupType} \in \{\text{"manual"}, \text{"automatic"}\}$. For datasets with a manageable number of features $X = \{x_1, x_2, \dots, x_p\}$, manual type leverages domain expertise to categorize features into groups based on the meaning inherent to the feature. In contrast, for larger datasets with numerous features, the automatic type leveraging Large Language Models (LLMs) such as "Gemini-api" (<https://aistudio.google.com/app/apikey>) is implemented. This method employs semantic analysis on feature names and descriptions, following a two-step process: i) Feature naming conventions: Each feature must have a descriptive name that clearly conveys its meaning, such as "speed". ii) Semantic analysis: The LLM conducts a semantic analysis of feature names and descriptions by crafting appropriate prompts to cluster features based on their semantic similarity and relevance.

Based on the aforementioned steps, the framework systematically partitions the n -dimensional feature space into m distinct feature subsets $\{X_{G_1}, \dots, X_{G_m}\}$. Subsequently, the dataset undergoes a rigorous data cleaning process to ensure quality and reliability. This includes detecting and addressing missing values and anomalies using techniques such as imputation, outlier detection, and correction. Following this, initial weights W_{G_i} are assigned to each feature to guide the feature selection process, striking a balance between flexibility and specificity. Features designated as "priority features," due to their critical role in prediction or their importance in preserving essential analytical characteristics, are assigned an initial weight of $W_{G_i} = 1$, reflecting the researchers' intent to ensure their retention. This prioritization helps safeguard these features during selection, particularly when they exhibit high correlations with other features. Conversely, non-priority features are initialized with a weight of $W_{G_i} = 0$. It is worth noting that the initial weight

of 1 is not equivalent to the weight applied in predictive tasks but is solely intended to emphasize the importance of retaining these features in the feature selection phase.

3.2.2. Feature selection

Two tasks are included in this section: solving the multicollinearity problem and selecting the optimal feature combination. To identify highly correlated features while retaining those of greatest interest to researchers, we first conduct intra-group correlation analysis, prioritizing features with an initial weight of 1 and assigning a weight of -1 to highly correlated features. Second, we assess inter-group multicollinearity using an enhanced Variance Inflation Factor (VIF) calculation method based on backward elimination (BE). This iterative process involves: i) Calculating initial VIF values for all features and setting a threshold (e.g., $VIF > 10$). ii) Identifying the feature with the highest VIF value in each iteration. If it exceeds the threshold, its weight is updated to -1 ; otherwise, the iteration terminates. iii) Recalculating VIF values for remaining features, excluding those with weights of -1 . iv) Repeating steps 2 and 3 until no feature's VIF exceeds the threshold.

To efficiently select the optimal feature combination, we integrate the SHAP with tree-based feature importance, balancing the impact of feature split importance $L_1(X_q)$ and gain importance $L_2(X_q)$ through a ratio δ (detailed in [Section 5.1](#)). Thus, the combined feature importance can be calculated as Eq.(1a):

$$L(X_q) = \delta \times \hat{L}_1(X_q) + (1 - \delta) \times \hat{L}_2(X_q) \quad (1a)$$

$$L_1(X_q) = \frac{1}{N_q} \sum_{s \in \text{splits}(X_q)} \text{cover}(s) \quad (1b)$$

where the $L(X_q)$ is the final importance value of X_q ; $\hat{L}_1(X_q)$ is the normalized value of $L_1(X_q)$, calculated as Eq.(1b); $\hat{L}_2(X_q)$ is the normalized value of $L_2(X_q)$, calculated as described in [Section 3.4](#). N_q is the number of times feature X_q is used for splitting across all trees; $\text{splits}(X_q)$ is the set of nodes where feature X_q is used for splitting; $\text{cover}(s)$ is the number of samples covered by node s .

By adopting the BE strategy for multiple iterations, update the feature weights (each time the weight of the last feature importance ranking is assigned a value of -1) until the iteration number λ is less than $n - m$. Further, the optimal feature combination is determined by evaluating the performance of each updated feature set using Mean Squared Error (MSE) as the evaluation metric. Specifically, each feature combination is applied to the predictive task, and the combination yielding the lowest MSE is identified as the optimal feature set. This process ensures that the selected features maximize predictive performance. To further clarify this procedure, we provide detailed experimental results and analysis in [Section 5](#).

3.3. Prediction algorithm

3.3.1. ST-MELightGB model components

The ST-MELightGB model, which integrates mixed-effects models with tree boosting. The following form of a mixed-effects model is assumed as Eq. (2):

$$y_i = f(X_i) + Z_i b_i + \epsilon_i, \\ b_i \sim N(0, D), \quad \epsilon_i \sim N(0, R_i), \quad i = 1, \dots, n, \quad (2)$$

where the total number of samples is $N = \sum_{i=1}^n n_i$; $y_i = [y_{i1}, \dots, y_{in_i}]^T$ is the $n_i \times 1$ vector of responses for the n_i observations in cluster i ; $X_i = [x_{i1}, \dots, x_{in_i}]^T$ is the $n_i \times p$ matrix of fixed-effects covariates; $Z_i = [z_{i1}, \dots, z_{in_i}]^T$ is the $n_i \times q$ matrix of random-effects covariates, which can be structured along multiple dimensions, including the temporal dimension (accounting for time-specific variations), the spatial dimension (capturing location-specific effects), and the policy restriction dimension (incorporating the heterogeneous impacts of HDDTs policy restrictions across both space and time). The flexibility in defining Z allows for the establishment of different clustering schemes based on these dimensions, either individually or in combination; $b_i = [b_{i1}, \dots, b_{iq}]^T$ is the $q \times 1$ unknown vector of random effects for cluster i , assumed to follow a multivariate normal distribution $b_i \sim N(0, D)$; $\epsilon_i = [\epsilon_{i1}, \dots, \epsilon_{in_i}]^T$ is the $n_i \times 1$ vector of errors, $\epsilon_i \sim N(0, R_i)$. D and R_i are the covariance matrices of b_i and ϵ_i respectively, that is, $D = \text{Cov}(b_i)$ and $R_i = \text{Cov}(\epsilon_i)$.

In linear mixed effects models (LMEM) and linear Gaussian process models in spatial statistics, it is assumed that $y_i = X_i \beta + Z_i b_i + \epsilon_i$, where β is a vector of coefficients. The standard LMEM further assumes that b_i and ϵ_i are independent and normally distributed, with observations between clusters also being independent. In this study, the ST-MELightGB extends the concepts of mixed-effects random forests (MERF) ([Hajjem et al., 2014](#)), where the tree-based structure used to estimate the fixed effects component $f(X_i)$ of the model is replaced by a gradient boosting machine, specifically LightGBM model ([Ke et al., 2017](#)). Thus, $f(X_i)$ will be given by the following as Eq. (3):

$$f(X_i) = \sum_{k=1}^K \eta \cdot T_k(X_i) \quad (3)$$

where $T_k(X_i)$ is the output of the k -th decision tree; K is the total number of trees in the model; η is the learning rate. The training process for LightGBM model will follow several steps:

Step 1. The initial model prediction is set as a constant value that minimizes the overall loss as Eq. (4):

$$f_0(X) = \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (4)$$

where $L(y_i, \gamma)$ is the loss function, typically squared error loss $L(y_i, \gamma) = (y_i - \gamma)^2$.

Step 2. For the k -th iteration, the residual for each data point is:

$$r_i^{(k)} = -\frac{\partial L(y_i, f_{k-1}(X_i))}{\partial f_{k-1}(X_i)} \quad (5)$$

For squared error loss, the residual simplifies to:

$$r_i^{(k)} = y_i - f_{k-1}(X_i) \quad (6)$$

Step 3. A decision tree $T_k(X)$ is fitted to predict the residuals $r_i^{(k)}$:

$$T_k(X) = \operatorname{argmin}_T \sum_{i=1}^n (r_i^{(k)} - T(X_i))^2 \quad (7)$$

Step 4. The new prediction for the model is updated by adding the output of the k -th tree, scaled by the learning rate η :

$$f_k(X) = f_{k-1}(X) + \eta \cdot T_k(X) \quad (8)$$

Step 5. After K iterations, the final fixed effects estimate $\hat{f}(X_i)$ is given by:

$$\hat{f}(X_i) = \sum_{k=1}^K \eta \cdot T_k(X_i) \quad (9)$$

3.3.2. ST-MELightGB model update strategy

The Expectation-Maximization (EM) algorithm (Moon, 1996) is an iterative technique used to find maximum likelihood estimates of model parameters, particularly in the presence of hidden or latent variables. It achieves this by alternating between two key steps: the Expectation (E-step), which calculates the expected value of the hidden variables given the current parameter estimates, and the Maximization (M-step), which updates the parameter estimates to maximize the likelihood of the observed data, considering the newly calculated expected values. In this study, the EM algorithm is employed to effectively estimate both the fixed effects, representing global trends, and the random effects, capturing local variations. The iterative update procedure of the EM algorithm for ST-MELightGB is formally detailed in Algorithm 1. The algorithm begins with initializing parameters, including the random effects (\hat{b}_i^0), residual variance ($\hat{\sigma}^{2(0)}$), and covariance matrix ($\hat{D}^{(0)}$). The iterative process comprises several key steps. Initially, the response variable ($\hat{y}_i^{(r)}$) is adjusted for each data point by removing the previously estimated random effects, and the LightGBM model is trained using this adjusted response variable and the covariates (X_i), yielding an updated fixed effects estimate ($\hat{f}(X_i)$). Next, the random effects are updated by computing the covariance matrix ($\hat{V}_i^{(r-1)}$) for each data point and using it to refine the random effects $\hat{b}_i^{(r)}$ with the fixed effects estimate. Following this, the residual variance ($\hat{\sigma}^{2(r)}$) is updated by averaging the sum of squared residuals and the trace of the inverse covariance matrix based on the residuals ($\epsilon_i^{(r)}$). The random effects covariance matrix ($\hat{D}^{(r)}$) is then updated by averaging the random effects estimates. Throughout the iterations, the generalized log-likelihood (GLL) is evaluated by summing the contributions from each data point, including residuals and determinant terms for covariance matrices. Convergence is assessed by comparing the change in GLL with the threshold (ϵ), terminating the algorithm if the change is below this threshold, and updating GLL_prev for the subsequent iteration.

Algorithm 1: ST-MELightGB: Iterative mixed-effects tree-boosting optimization with spatiotemporal heterogeneity

Input: Set initial residual variance $\hat{\sigma}^{2(0)}$, random effects \hat{b}_i^0 and random effects covariance matrix $\hat{D}^{(0)}$. Initial log-likelihood $GLL_prev = Inf$

Output: Predictor function $\hat{f}(X_i)$, updated parameters \hat{b}_i , $\hat{\sigma}^2$ and \hat{D} .

```

1: Initialize  $\hat{\sigma}^{2(0)} = 1$ ,  $\hat{b}_i^0 = 0$ ,  $\hat{D}^{(0)} = I_q$  and  $r = 0$ ,  $R =$  maximum number of iterations.

2: while  $r < R$  :
   r = r + 1
   # step 1: Adjust response variable and estimate fixed effects
   4: for  $i=1$  to  $n$  do:
      5:    $\hat{y}_i^{*(r)} = y_i - Z_i \hat{b}_i^{(r-1)}$ 
   6: end for
   Train LightGBM model with  $\hat{y}_i^{*(r)}$  as response and  $X_i$  as covariates. Obtain updated fixed effects
   7: estimate  $\hat{f}(X_i)$ .
   # step 2: Update random effects
   8: for  $i=1$  to  $n$  do:
      9:   for  $i=1$  to  $n$  do:
         10:     $\hat{V}_i^{(r-1)} = Z_i \hat{D}^{(r-1)} Z_i^T + \hat{\sigma}^{2(r-1)} I_n$  # Compute covariance matrix for random effects
         11:     $\hat{b}_i^{(r)} = \hat{D}^{(r-1)} Z_i^T (\hat{V}_i^{(r-1)})^{-1} (y_i - \hat{f}(X_i))$  # Update random effects
      12: end for
      13: # step 3: Update residual variance
      14: for  $i=1$  to  $n$  do:
         15:    $\epsilon_i^{(r)} = y_i - \hat{f}(X_i) - Z_i \hat{b}_i^{(r)}$ 
      16: end for
      17: Update  $\hat{\sigma}^{2(r)} = \frac{1}{N} \sum_{i=1}^n (\epsilon_i^{(r)})^2 + \hat{\sigma}^{2(r-1)} \text{trace}((\hat{V}_i^{(r-1)})^{-1})$ 
      18: # step 4: Update random effects covariance matrix
      19:  $\hat{D}^{(r)} = \frac{1}{n} \sum_{i=1}^n \hat{b}_i^{(r)} (\hat{b}_i^{(r)})^T + \hat{D}^{(r-1)} - \hat{D}^{(r-1)} Z_i^T (\hat{V}_i^{(r-1)})^{-1} Z_i \hat{D}^{(r-1)}$ 
      20: # Evaluate generalized log-likelihood
      21: for  $i=1$  to  $n$  do:
         22:    $\epsilon_i = y_i - \hat{f}(X_i) - Z_i \hat{b}_i^{(r)}$ 
         23:    $GLL(f, b | y) = \sum_{i=1}^n \epsilon_i^T R_i^{-1} \epsilon_i + \log |D| + \log |R_i|$ 
      24: end for
      25: if  $|GLL - GLL\_prev| < \varepsilon$  :
         26:   Break
      27: end if
      28:  $GLL\_prev = GLL$ 
29: end while

```

3.4. Tree-based models explainability

SHAP is a game theoretic approach to explain the output of any machine learning model (Lundberg and Lee, 2017). It is based on Shapley values, a concept from cooperative game theory, which provides a method for fairly distributing the “payout” among the features. Shapley values (ϕ_i) quantify the contribution of each feature to the prediction by considering all possible feature combinations. For a model f with input features x and a specific prediction $f(x)$, the Shapley value for feature i is calculated as Eq. (10):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (10)$$

where N is the set of all features. It assigns an importance value to each feature that represents the effect on the model prediction of including that feature. S is a subset of N that does not include feature i , $f(S)$ is the prediction with features in subset S , and $|S|$ is the size of subset S . The term $(|S|!(|N| - |S| - 1)!)/(|N|!)$ is a weighting factor ensuring all feature combinations are considered.

While SHAP values offer a theoretically grounded approach to model interpretation, their direct computation involves an exponential complexity with respect to the number of features, rendering the approach impractical for large datasets or models with numerous features. The TreeExplainer, developed by (Lundberg et al., 2020), leverages the unique architecture of tree-based models to dramatically speed up Shapley value computations through the employment of efficient polynomial-time algorithms. The Shapley value for a feature i in a tree-based model is given by Eq. (11):

$$\phi_i = \sum_{t=1}^T \sum_{j=1}^{L_t} \frac{n_j^t}{N} \left(v_{ij}^t - v_{i(j-1)}^t \right) \quad (11)$$

where T is the number of trees, L_t is the number of leaves in tree t , n_j^t is the number of samples reaching node j in tree t , v_{ij}^t is the value at node j for feature i , and $v_{i(j-1)}^t$ is the value at the previous node for feature i . For each tree, the contribution of feature i is computed by summing the contributions along all paths in the tree where feature i is used for splitting. Within each tree, the contribution of feature i at a node is calculated by comparing the model’s prediction before and after the node’s split on feature i . The final Shapley value for feature i is obtained by averaging its contributions across all trees in the ensemble, weighted by the number of samples reaching each node.

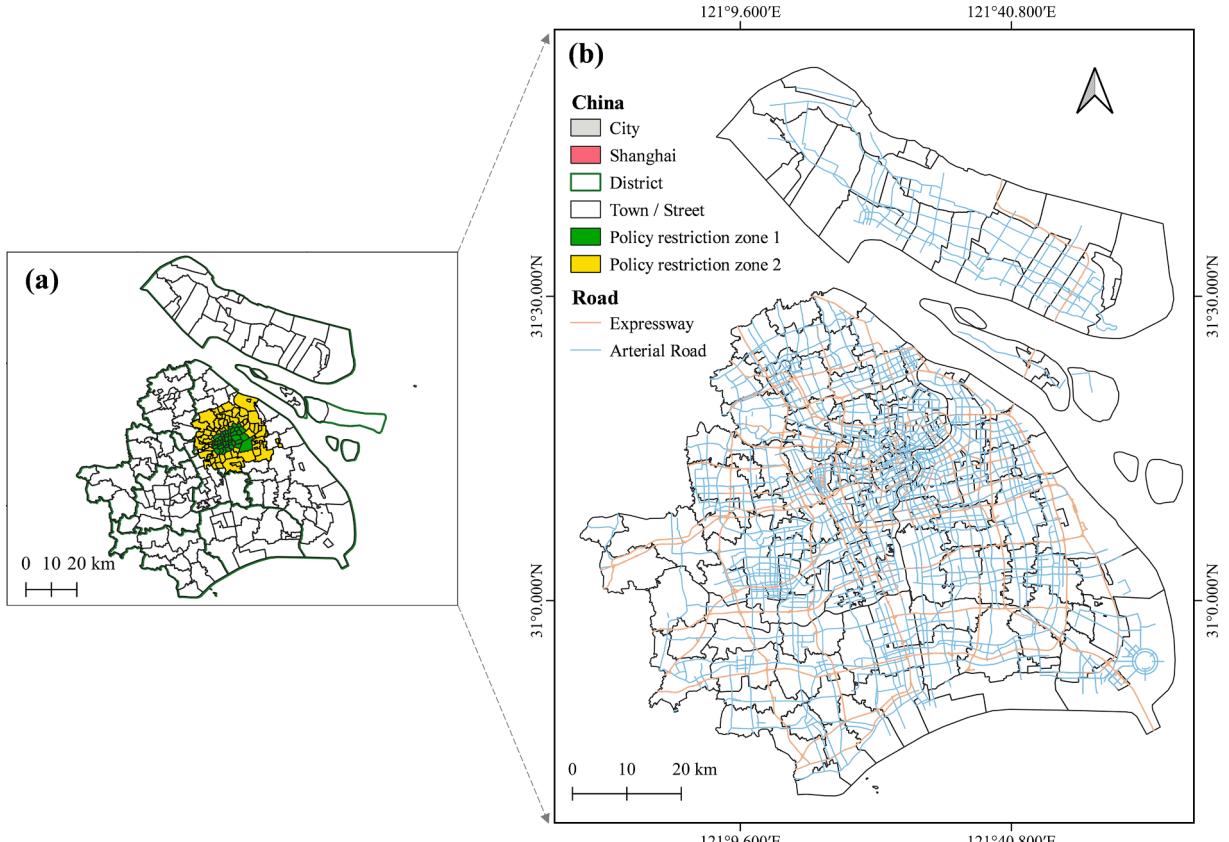


Fig. 3. Study area.

4. Study area and data sources

4.1. Study area

Shanghai, located in the eastern part of China along the Pacific coast and at the eastern edge of the Eurasian continent, sits at the geographical center of China's northern and southern coastlines. This research focuses on the 16 municipal districts and 215 sub-districts (towns) of Shanghai, encompassing an area of 6,340.5 square kilometers, as illustrated in Fig. 3. This area is characterized by a high concentration of economic activity, a dense population exceeding 24.87 million residents (as of October 2023), and a sophisticated transportation network. To mitigate NOx emissions from HDDTs, Shanghai has implemented a series of spatiotemporal policies with differentiated time controls based on vehicle type, registration place, and emission standards. The restricted areas are mainly divided into two zones: the first is within the inner ring road (the green area in Fig. 3(a) – Zone 1), and the second is from the middle ring road to the outer ring road (the yellow area in Fig. 3(a) – Zone 2). For Zone 1, which is closer to the city center, the policy prohibits freight vehicles with local license plates from passing between 7:00 AM and 8:00 PM on weekdays, except for special vehicles. For Zone 2, the policy states: i) Freight vehicles with license plates from other provinces and cities are prohibited from passing between 7:00 AM and 8:00 PM every day; ii) Diesel trucks that meet the China 3 emission standard are prohibited from passing between 6:00 AM and 1:00 AM the next day.

In this study, we further decompose the above policy regulations on heavy trucks in Shanghai into three dimensions: temporal, spatial, and spatiotemporal policy restriction dimensions, as shown in Fig. 4. The data corresponds to December 2022, a period when Shanghai was transitioning back to normalcy after the strict Covid-19 lockdown. Specifically, the traffic control policies effective on December 1, 2022, were selected to ensure alignment with the HDDT emissions data, reflecting the conditions and restrictions during the same period. The temporal dimension includes four periods: T1, T2, T3, and T4. T1 is from 1:00 AM to 6:00 AM, with no policy restrictions; T2 is divided into two parts, from 6:00 AM to 7:00 AM and from 8:00 PM to 1:00 AM the next day, focusing on restrictions for diesel trucks meeting China 3 emission standards; T3 and T4 are periods with restrictions for all heavy trucks, with T3 covering the morning peak (7:00 AM to 10:00 AM) and the evening peak (5:00 PM to 8:00 PM). The spatial dimension includes three zones: Z1, Z2, and Z3, where Z3 is the white area in Fig. 3(a) (the area remaining after removing Z1 and Z2). The policy restriction unit (PRU) includes PRU1, PRU2, PRU3, and PRU4, categorizing similar policy regulations across spatial and temporal dimensions. This division allows for a more detailed analysis of policy effects.

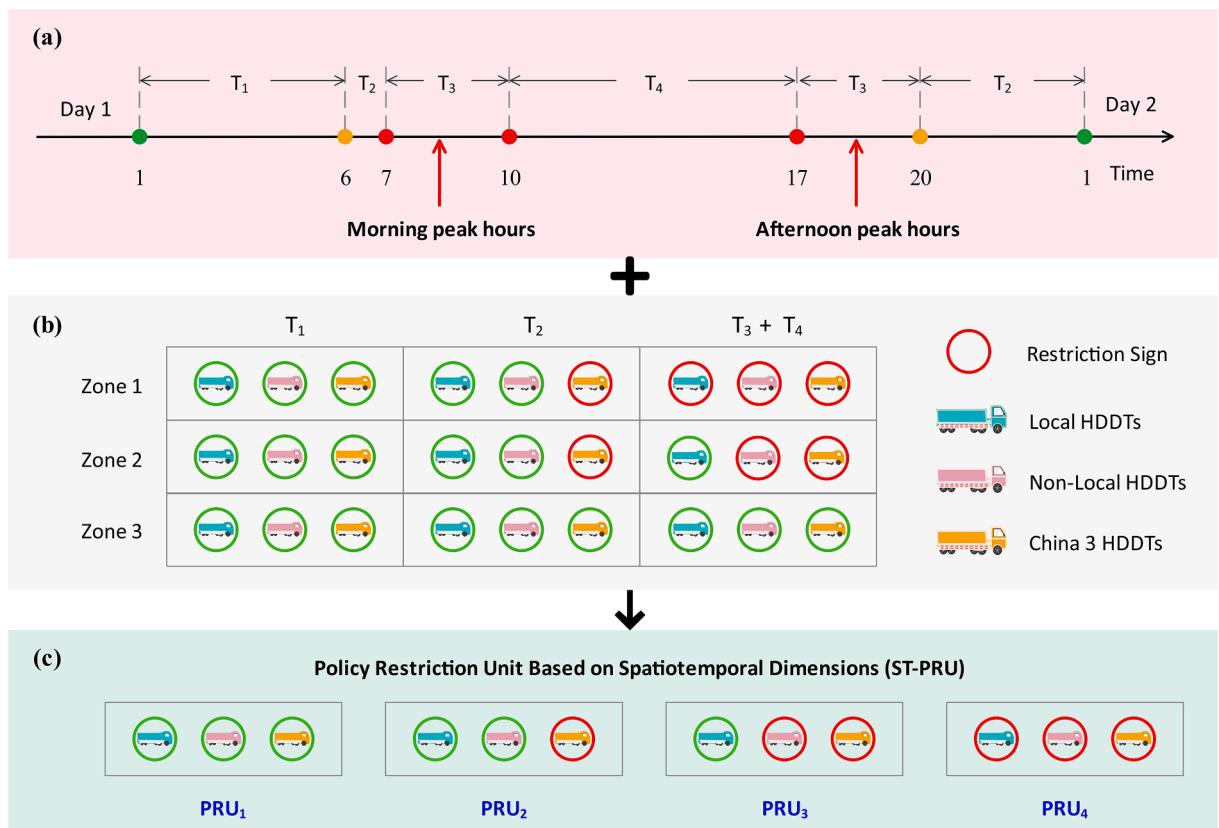


Fig. 4. Spatiotemporal decomposition of Shanghai's heavy truck policy regulations.

4.2. Data collection and dataset building

4.2.1. Trajectory data of HDDTs

This study utilizes GPS data for 87,930 heavy vehicles, provided by the Shanghai Municipal Transportation Bureau on December 1, 2022. The dataset, sampled at intervals ranging from 2 to 60 s, comprises 8.54 GB of information, including over 80 million records. Each record contains vehicle ID, geolocation (longitude and latitude), timestamp, emission standard classification, and vehicle type. Following data preprocessing to eliminate duplicate, missing, and anomalous records, the final dataset is refined to 7.97 GB, encompassing over 75 million valid records. Table 1 presents a representative sample.

4.2.2. Built environment data

In this study, we select five Ds as built environment factors: density, design, land use diversity, distribution of industries, and distance to the urban center. Density indicators include the density of public transportation facilities and land use intensity. Public transportation density reflects the frequency and number of transportation hubs, such as airports, seaports, river ports, and railway stations per unit area. Land Use Density represents the intensity of land use in the study area, with different land use types leading to distinct NOx emission patterns. We collect the proportions of three major land use types: industrial, public administration/service, and commercial land. Design involves factors related to the proportion of road networks and the number of road intersections. According to OpenStreetMap's road classification fields, the road network analyzed includes elevated roads and expressways, suburban arterial roads, urban arterial roads, and urban collector roads. Road intersections are defined as the points where two or more roads of the aforementioned types intersect. The road network and road intersection density are measured by road length and the number of intersections per unit area, respectively. These data are obtained through the API of the AutoNavi Map Web Service (<https://lbs.amap.com/>). Land use diversity refers to the combination and distribution of different land use types (e.g., commercial, industrial, public facilities) within a specific area (Zhu et al., 2024). It quantifies the diversity and uniformity of land use types using an entropy index. Higher entropy indicates greater land use diversity and more even distribution. Distance to the Urban Center describes the spatial proximity of a location to the urban core area. In this study, it measures the distance between each sub-district (town) and the center of Shanghai. The AutoNavi Map Web API route planning service was used to obtain the actual travel distance from the starting point to the endpoint. Distribution of Industries covers the primary, secondary, and tertiary industries in economics. This study collects the distribution of all industries in Shanghai and calculated the number of industries per unit area.

4.2.3. Socio-demographic data

The population density data are utilized from China's seventh national census (2020), including the number of household members, the number of employed members, gender, age and educational level. Here, we map the latitude and longitude information of residential locations to each sub-district (town) using ArcGIS 10.6 software. PGDP, a key indicator of regional economic intensity, is also considered. Regions with higher PGDP tend to have more developed economic systems, characterized by increased business activities, industrial production, and consumer behavior, ultimately leading to higher demands for logistics and transportation. The 2023 Shanghai PGDP data used in this study were sourced from the official releases of the Shanghai Statistics Bureau (<https://tjj.sh.gov.cn/>).

4.2.4. Traffic flow data

This study employs three key traffic flow parameters: average speed, speed variability, and congestion level, to examine their influence on HDDT emissions. All traffic parameters were directly derived from HDDT GPS trajectory data, ensuring temporal consistency and precise spatiotemporal alignment between truck activity and traffic conditions. Average speed represents the mean velocity of vehicles along road segments during specific time intervals and was computed by aggregating speed data from GPS trajectories within each of sub-districts, with geospatial integration using tools such as ArcGIS. Speed variability, calculated as the standard deviation of vehicle speeds within the same intervals, serves as a proxy for traffic dynamics, where higher variability indicates frequent speed fluctuations. Traffic congestion levels, categorized into five classes (free flow, basically free flow, light congestion, moderate congestion, and severe congestion), were identified using an unsupervised clustering algorithm applied to HDDT speed data. This data-driven method, validated by previous studies (Shi et al., 2023), captures congestion patterns based on HDDT-specific speed distributions rather than relying on fixed thresholds, ensuring greater adaptability to varying sub-district conditions.

It is worth noting that the traffic data used in this study corresponds to December 2022, a period following Shanghai's strict COVID-19 lockdown (March 18–May 30, 2022), during which urban transportation was gradually recovering. By December, most citywide

Table 1
Trajectory records of HDDTs.

Vehicle ID	Longitude	Latitude	Time	Type	Standard
Jin46ae69ad87bd72a278923ea5c1003e45	121.33002	31.350355	2022/12/1 1:31	China 5	tractor
Hu00385d1ccae128a14f48df1b6541efb0	121.581225	31.364961	2022/12/1 1:30	China 5	tractor
Hu00385d1ccae128a14f48df1b6541efb0	121.581201	31.364546	2022/12/1 1:30	China 5	tractor
Hub7a0115f6a0ffa079035559a3a225f48	121.001071	31.100056	2022/12/1 0:16	China 6	tractor
Hu67b7afc27ce6fc180f1bc48ef7e8cd4	121.814815	31.107302	2022/12/1 0:17	China 3	Van-type
Hu622f948ae16d083e2806902cb353290a	121.299101	31.187295	2022/12/1 0:00	China 5	bulk lorry
.....

restrictions had been lifted, and freight transportation had largely resumed. While overall traffic conditions were approaching pre-pandemic levels, residual effects such as shifts in travel demand, increased reliance on freight logistics, and modifications in commuter behavior may have introduced localized variations in traffic dynamics. By directly deriving traffic flow parameters from HDDT GPS trajectory data, this study provides high-resolution, real-time insights into localized traffic patterns, potentially capturing the impacts of pandemic-induced changes in mobility and freight demand.

4.2.5. NO_x emissions data for HDDTs

Accurate modeling of NO_x emissions from HDDTs requires high spatiotemporal resolution data, comprehensive road traffic information, and robust emission models. GPS data have proven effective in monitoring and estimating transportation-related emissions (Cheng et al., 2022). The specifications and load status of HDDTs, alongside road traffic conditions and the traffic environment, significantly influence the quantitative parameters of emission models (McCaffery et al., 2021). The tonnage levels and emission standards of HDDTs were obtained through a static vehicle attribute information interface. Emission standards are categorized as China 3, China 4, China 5, and China 6, corresponding to Euro III, Euro IV, Euro V, and Euro VI, respectively. The latest version of the COPERT V model (version 5.7.3-Jan 2024; <https://copert.emisia.com/>) was employed to simulate basic vehicle emission factors. The process integrates emission factors, vehicle-specific data, and traffic conditions to estimate emissions at various scales, ranging from individual vehicles to the entire network.

(a) *Localization of emission factors:* The emission factors for different types of HDDTs were derived from the COPERT V model. However, due to variations in fuel types, modeling years, and emission standards across regions, it was necessary to adjust these emission factors using local fuel data to ensure more accurate representation of emissions. The following form of adjustment is assumed as Eq. (12):

$$EF_{p,m,n,v,i}^{\text{corr}} = \frac{F_{\text{corr},\text{NO}_x,\text{Fuel}}}{F_{\text{corr},\text{NO}_x,\text{Base}}} \times \frac{\left(\alpha \times v^2 + \beta \times v + \gamma + \frac{\delta}{v} \right)}{\varepsilon \times v^2 + \theta \times v + \varphi} \times (1 - RF) \quad (12)$$

where $EF_{p,m,n,v,i}^{\text{corr}}$ represents the corrected emission factor for an HDDT with an emission standard n and tonnage level m at a traffic speed v on road segment i at time interval t , in g/km; v denotes the average speed of HDDTs during time interval; The basic emission factor parameters $\alpha, \beta, \gamma, \delta, \varepsilon, \theta, \varphi$ and the reduction factor (RF), were obtained directly from the COPERT V model. $F_{\text{corr},\text{NO}_x,\text{Fuel}}$ and $F_{\text{corr},\text{NO}_x,\text{Base}}$ are the fuel correction functions.

(b) *Individual vehicle emissions calculation:* The NO_x emissions from an HDDT with emission standard m and tonnage level n on-road segment i at time interval t , denoted as $E_{\text{NO}_x,m,n,i,t}$, is calculated in kilograms (kg):

$$E_{\text{NO}_x,m,n,i,t} = EF_{p,m,n,v,i}^{\text{corr}} \times L_{i,t} \times 10^{-3} \quad (13)$$

where L_i represents the distance traveled by the vehicle on segment i during time interval t ;

(c) *Segment-Level Emissions Estimation:* Emissions for each road segment were calculated by aggregating the emissions of all vehicles traveling on that segment:

$$E_{\text{NO}_x,i,t} = \sum_{N,t} E_{\text{NO}_x,m,n,i,t} \quad (14)$$

where $E_{\text{NO}_x,i,t}$ represents the total emissions on segment i during time interval t ; N is the total number of vehicles passing through that segment i during time interval t .

(d) *Network-Level Emissions Estimation:* This study employed geospatial analysis tools, such as spatial joins and intersection operations in ArcGIS, to align emission data with predefined geographic units (sub-districts). Once the NO_x emissions for each road segment were mapped, the emissions of all segments within the study area were summed to calculate the total NO_x emissions at the network level, as expressed in Eq. (15):

$$E_{\text{NO}_x,t} = \sum_i E_{\text{NO}_x,i,t} \quad (15)$$

where $E_{\text{NO}_x,t}$ represents the total NO_x emissions for the unit area during time interval t . To understand the spatial distribution of emissions, the emission density $E_{z,\text{density}}$ of unit area z (sub-district) (kg/km²/h) is assumed as Eq. (16):

$$E_{z,\text{density}} = E_{\text{NO}_x,t}/A_z \quad (16)$$

where A_z is the area of sub-district z . This calculation provides a measure of the emission intensity, which is essential for evaluating the environmental impact at different spatial scales.

4.3. Variables description

We visually illustrate the spatial distribution of key explanatory variables across the study area in Fig. 5, providing insights into potential spatial patterns and variations. Additionally, we conducted a descriptive analysis of the collected data. Table 2 presents descriptive statistics for the explanatory variables, including the mean, standard deviation, minimum, and maximum values for each

variable. These statistics offer a quantitative overview of the built environment, demographic characteristics, and traffic flow characteristics within the study area. Similarly, Table 3 provides descriptive statistics for the emission density, including the mean, standard deviation, minimum, and maximum observed values.

5. Results and discussion

5.1. Experimental setting

This study evaluates the performance of our proposed ST-MELightGB algorithm compared to classical statistical models, including multiple linear regression (MLR), LMEM, and generalized additive mixed models (GAMM). Additionally, we compare against ensemble machine learning models, namely random forest (RF), extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016), LightGBM (Ke et al., 2017), and MERF (Hajjem et al., 2014). Furthermore, to compare the combined advantages of tree-boosting and mixed-effects models and to evaluate the impact of spatiotemporal heterogeneity on the results, we tested the following models: the T-MELightGB model, which uses only temporal information as the random grouping variable; the S-MELightGB model, which uses only spatial information as the random grouping variable; and the ST-MEXGB model, which combines spatiotemporal information and mixed-effects models with XGBoost. All computations are performed on a laptop equipped with a 2.9 GHz quad-core processor and 16 GB of random-access memory. To ensure optimal performance, we employ 5-fold cross-validation on the training data for each run, selecting the combination of tuning parameters from a comprehensive grid that minimizes the average mean square error on the validation data. For each boosting algorithm, we consider the following candidate tuning parameters: learning_rate $\in \{0.1, 0.05, 0.01\}$, max_depth $\in \{-1, 1, 5, 10\}$, the number of boosting rounds or trees to be built n_estimators $\in \{50, \dots, 1000\}$, minimum number of samples required to form a leaf min_data_in_leaf $\in \{10, \dots, 100\}$, and the number of boosting iterations $\in \{1, \dots, 1000\}$.

Additionally, all the aforementioned models are evaluated based on the screening results derived from the feature selection framework proposed in this study. The parameters are set as follows: GroupType = "automatic", handle_multicollinearity = "True" to address the screening of highly correlated features. Feature scaling is performed using the MinMaxScaler method. To balance the impact of feature split importance (inherent to the model) and gain importance (derived from SHAP), we explore a range of ratios δ

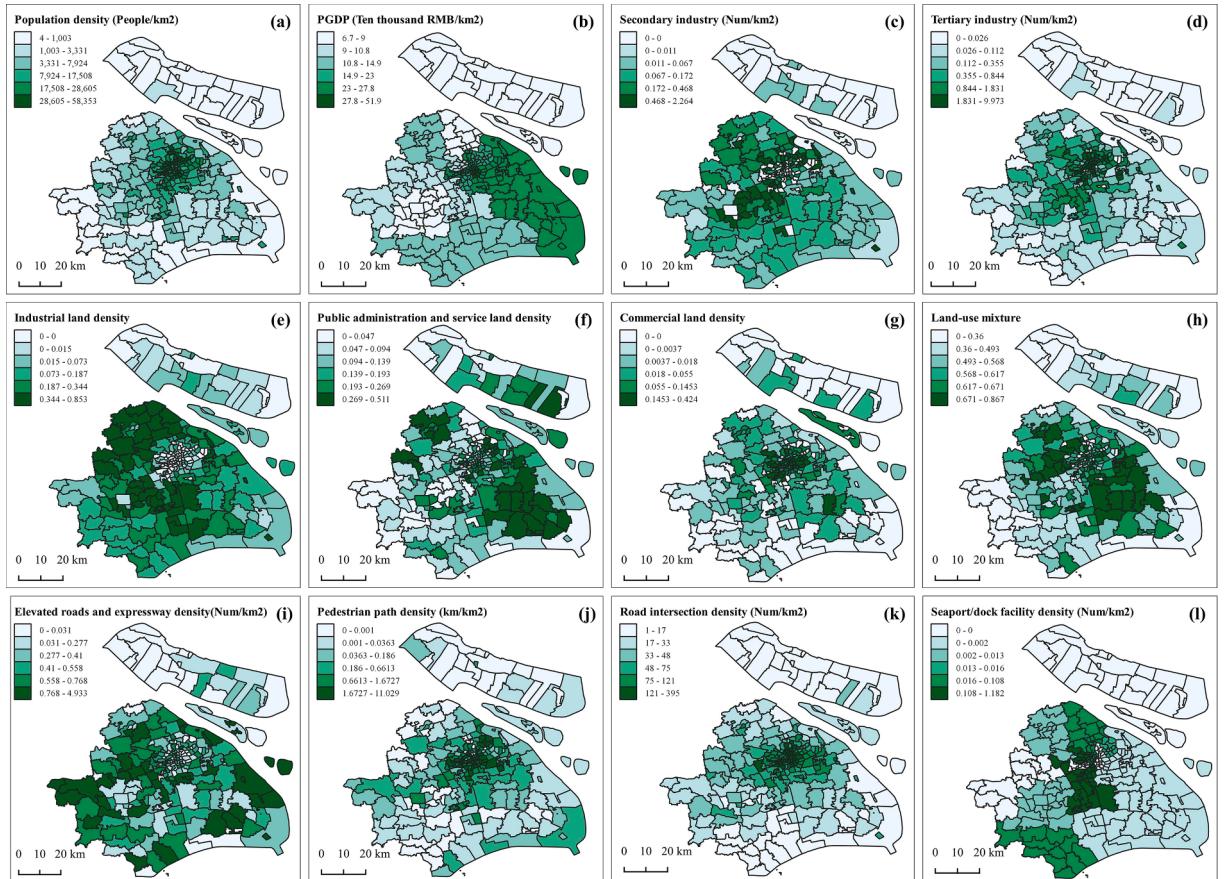


Fig. 5. Distributions of some explanatory variables.

Table 2

Descriptive statistics of independent variables.

Variable	Description	Mean	St.D	Min	Max
Built environment					
Industrial land density	Size of industrial land area /area size (%)	0.166	0.185	0	0.853
Public administration and service land density	Size of public administration land area /area size (%)	0.161	0.112	0	0.511
Land-use mixture	Entropy index of the land use mix	0.541	0.160	0	0.867
Commercial land density	Size of commercial land area /area size (%)	0.059	0.087	0	0.424
Distance to city center	Shortest distance from city center (km)	31.995	23.396	2.070	115.19
Seaport/dock facility density	Number of port terminals and related facilities per unit area (Num/km ²)	0.113	0.286	0	1.182
Suburban arterial road density	Length of suburban arterial roads per unit area within urban area (km/km ²)	1.209	0.770	0.038	4.184
Elevated roads and expressway density	Length of elevated roads and expressways per unit area within urban area (km/km ²)	0.513	0.556	0	4.933
Suburban and rural road density	Length of suburban and rural roads per unit area in suburban and rural regions (km/km ²)	0.101	0.390	0	4.385
Internal road density	Length of internal roads within defined area (km/km ²)	1.701	2.370	0	17.088
Pedestrian path density	Length of pedestrian paths per unit area (km/km ²)	0.750	1.355	0	11.029
Elevated bridges and expressway intersection density	Number of intersections involving elevated bridges and expressways per unit area (num/km ²)	11.056	16.484	0	126.22
Road intersection density	Total number of road intersections per unit area (num/km ²)	69.584	70.068	1.344	394.91
Primary industry	Number of Primary industry /area size (num/km ²)	0.025	0.101	0	0.997
Secondary industry	Number of Secondary industry /area size (num/km ²)	0.237	0.392	0	2.264
Tertiary industry	Number of Tertiary industry /area size (num/km ²)	0.915	1.410	0	9.973
Demographic characteristics					
Population density	Population/area size (person per km ²)	12,823	13,281	4	58,353
PGPD	Gross domestic product/population (10 ⁴ RMB per capita)	18.159	10.478	6.655	51.933
Traffic flow characteristics					
Speed	Average speed of different streets over aggregation period (km/h)	36.171	13.271	7.659	75.011
Speed stddev	Measure of dispersion in speed data over time interval	17.381	5.676	0.625	43.579
Congestion level	Traffic congestion levels	3	1.675	1	5

Table 3Descriptive statistics of response variable (unit: kg/km²/h).

Variable	Description	Mean	St.D	Min	Max
Total	NOx emissions per unit area per hour in each sub-district (town)	0.148	0.218	0	2.375
DayPeriod					
T1	NOx emissions from 1:00 AM to 6:00 AM	0.075	0.135	0	1.504
T2	NOx emissions from 6:00 AM to 7:00 AM and from 8:00 PM to 1:00 AM the next day	0.124	0.181	0	1.805
T3	NOx emissions during the morning peak (7:00 AM to 10:00 AM) and evening peak (5:00 PM to 8:00 PM)	0.160	0.222	0	2.042
T4	NOx emissions from 10:00 AM to 5:00 PM	0.207	0.267	0	2.375
Zone					
Z1	NOx emissions in the central urban area of Shanghai	0.056	0.051	0	0.338
Z2	NOx emissions in the area between the inner and outer rings of Shanghai	0.129	0.197	0	1.741
Z3	NOx emissions in the area outside of Shanghai's outer ring road	0.189	0.251	0	2.375
Policy restriction					
PRU1	NOx emissions with no spatiotemporal policy restrictions	0.168	0.239	0	2.375
PRU2	NOx emissions under policy restriction specific to China 3 standard	0.095	0.131	0	1.359
PRU3	NOx emissions under policy restriction specific to local/non-local HDDTs	0.164	0.226	0	1.741
PRU4	NOx emissions where spatiotemporal policy restrictions are the strictest	0.060	0.052	0	0.296

from 0 to 1, with increments of 0.1. Fig. 6 illustrates the relationship between the number of eliminated features and the MSE loss across varying ratios. When $\delta = 0.2$, a minimum loss is attained, corresponding to an optimal dimensionality reduction.

5.2. Model prediction performance

In this study, we compare the predictive performance using root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2) metrics. These metrics are evaluated both across the total test sample (overall performance) and within specific dimensions (local performance). Table 4 presents the prediction results for the ST-MELightGB model and the baseline methods.

Fig. 7 displays the performance improvement of the ST-MELightGB model over other baseline models across different dimensions. For overall results, ST-MELightGB model demonstrates a substantial improvement exceeding 55 % over the classical linear statistical

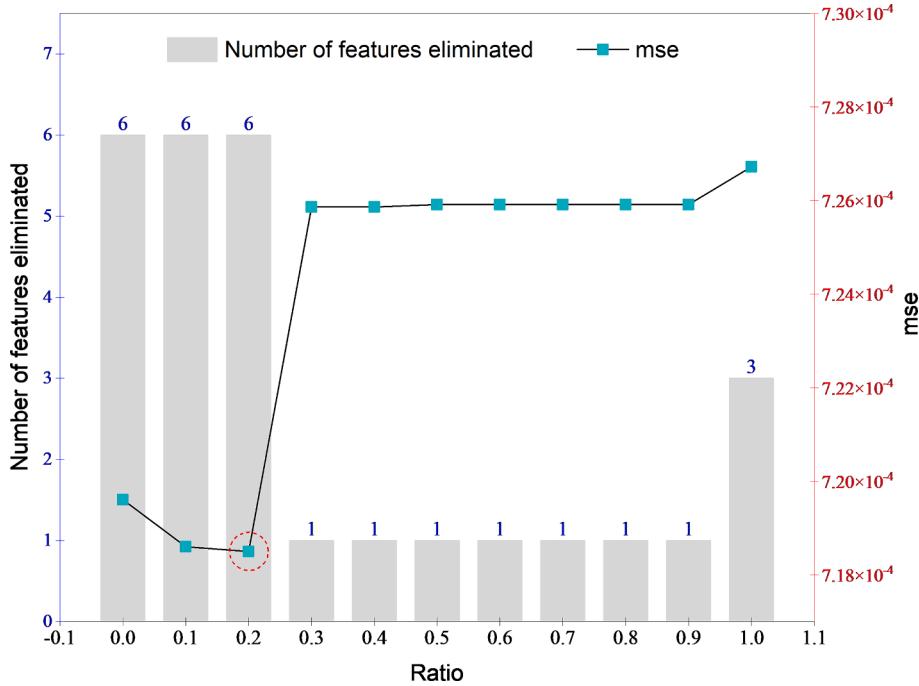


Fig. 6. Balancing feature importances: Effect of ratio on dimensionality reduction and performance.

models (MLR and LMEM), highlighting the limitations of linear approaches in capturing the nonlinear relationships inherent in this study. Compared to classical machine learning methods (RF, XGBoost, and LightGBM), the ST-MELightGB model exhibits a notable enhancement of approximately 15 %–20 %. Furthermore, it surpasses methods that combine mixed effects and machine learning (MERF, ST-MEXGB, T-MELightGB, and S-MELightGB) by 6 %–12 %. These overall results underscore the effectiveness of integrating spatiotemporal information and mixed-effects models within a machine learning framework for capturing the complex relationships between NO_x and explanatory variables. For local prediction tasks, this performance advantage is consistently observed across various temporal (T1–T4), spatial (Z1–Z4), and policy restriction units (PRU1–PRU4).

Fig. 8 presents the predictive performance results across the three categories of models for all defined dimensions. The results show that linear models, specifically MLR and LMEM, exhibit the least robust predictive capabilities, with R² scores persistently dipping below 0.8 throughout all examined dimensions. Besides, their performance for central zones with stringent policy restrictions (Z1 and PRU4) and the early morning period (T1) is particularly poor, as evidenced by negative R² scores, which signify predictions lacking reliability. In contrast, the GAMM model, which incorporates smooth functions to capture nonlinear relationships, shows improved performance compared to the linear models. This is evidenced by R² values exceeding 0.8 in multiple dimensions, with a notable increase in the PRU2 task from 0.353 (LMEM) to 0.761. However, despite this improvement, the local prediction accuracy of statistical models remains suboptimal, particularly for dimensions such as T1, Z1, and PRU4, where complex spatial–temporal variations are present. Secondly, classic machine learning models (RF, XGBoost, and LightGBM) exhibit similar predictive performance across dimensions, but their results for Z1 and PRU4 remain significantly weaker (R² < 0.55) compared to other tasks. This suggests that although machine learning models capture nonlinear relationships more effectively than linear models, they may still struggle to fully account for the intricate spatial dependencies and policy-induced variations in these regions. In contrast, the MERF, ST-MEXGB, and ST-MELightGB models outperform baseline methods across all dimensions, with ST-MELightGB demonstrating the best predictive performance for Z1 and PRU4 (R² > 0.7). The superior performance of these mixed-effects models highlights their ability to incorporate spatial and temporal dependencies, making them more effective in handling the localized variations in Z1 and the unique traffic patterns in T1. Evidently, these findings underscore the heightened challenge in accurately modeling the localized effects within Z1 and PRU4 relative to other dimensions.

5.3. Relative importance of the explanatory variables

Fig. 9 illustrates the relative contributions and importance rankings of various factors influencing NO_x emission, derived from the average absolute SHAP values across overall test samples. Land use factors emerge as the most significant (36.35 %), closely followed by road design factors (35.91 %). Traffic characteristics rank third (13.68 %), with industrial structure (8.90 %), demographics (3.64 %), and urban planning (1.54 %) having less influence. Notably, urban land use and road system design collectively account for 72.26 % of the total impact, highlighting the critical role of integrated land planning and transportation infrastructure in mitigating NO_x emissions. Within land use factors, industrial land density dominates (31.28 %), while commercial land density (1.50 %) and public

Table 4

The predication results of the ST-MELightGB model and other baseline methods (* denote $R^2 < 0.5$).

Model	Overall performance			Performance under different policy restriction units											
				PRU1			PRU2			PRU3			PRU4		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
MLR	0.072	0.123	0.669	0.077	0.129	0.683	0.062	0.125	*	0.064	0.121	0.737	0.062	0.077	*
sLMEM	0.073	0.123	0.667	0.078	0.130	0.680	0.062	0.125	*	0.064	0.121	0.738	0.062	0.076	*
GAMM	0.056	0.086	0.840	0.062	0.092	0.840	0.045	0.076	0.761	0.053	0.088	0.861	0.032	0.040	*
RF	0.041	0.068	0.899	0.043	0.073	0.899	0.036	0.054	0.880	0.047	0.075	0.901	0.027	0.036	*
XGBoost	0.036	0.064	0.910	0.037	0.071	0.904	0.031	0.044	0.921	0.041	0.065	0.921	0.025	0.033	0.511
LightGBM	0.040	0.068	0.899	0.042	0.073	0.898	0.032	0.047	0.908	0.047	0.076	0.896	0.026	0.033	0.498
MERF	0.034	0.059	0.925	0.035	0.061	0.928	0.034	0.052	0.889	0.040	0.068	0.917	0.021	0.026	0.695
Our model															
ST-MEXGB	0.031	0.059	0.925	0.032	0.064	0.923	0.027	0.043	0.923	0.038	0.064	0.928	0.021	0.027	0.684
T-MELightGB	0.033	0.058	0.927	0.034	0.062	0.928	0.029	0.042	0.928	0.042	0.069	0.915	0.020	0.025	0.709
S-MELightGB	0.035	0.062	0.916	0.037	0.067	0.915	0.031	0.046	0.912	0.041	0.069	0.915	0.023	0.030	0.609
ST-MELightGB	0.031	0.054	0.935	0.033	0.059	0.934	0.024	0.033	0.954	0.039	0.063	0.928	0.019	0.024	0.742

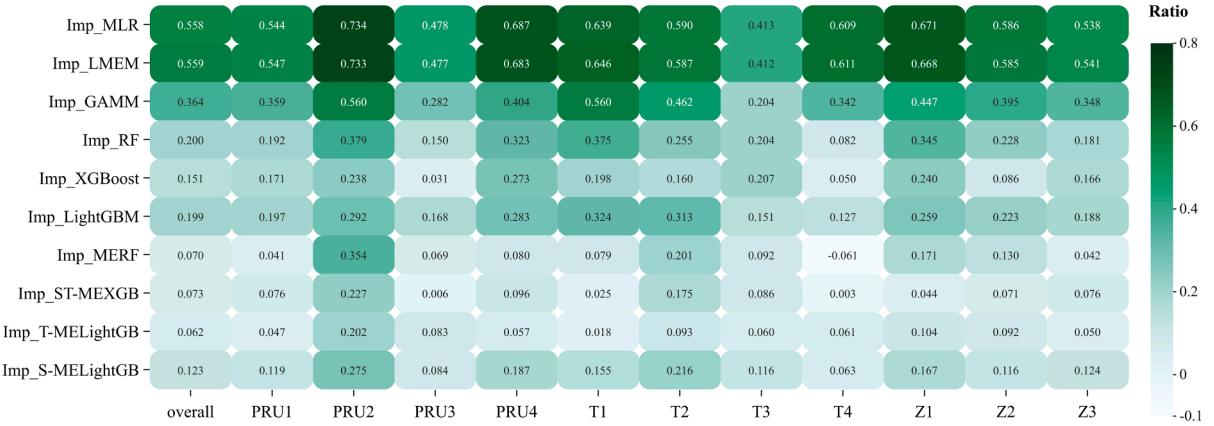


Fig. 7. Performance improvement (RMSE) of the ST-MELightGB model over the baseline methods.

administration and service land density (1.43 %) contribute significantly less. This contrasts with passenger car emission studies, where commercial land density typically prevails (Wu et al., 2023). Road design factors are primarily influenced by elevated roads and expressway density (13.11 %) and their intersections (11.14 %), which indicating that the spatial distribution of NOx emission hotspots for HDDTs is predominantly concentrated along select inter-city highway corridors, aligning with the results reported by (Cheng et al., 2023). Interestingly, pedestrian path density (4.69 %) surpasses road intersection density (3.84 %), suburban arterial road density (1.16 %), and rural road density (0.72 %), suggesting a unique relationship between urban walkability and HDDTs emissions, possibly due to the increased frequency of acceleration and deceleration in pedestrian-oriented areas (Zavala-Reyes et al., 2019). Among traffic characteristics, vehicle speed (9.69 %) contributes most to NOx emissions, followed by speed standard deviation (3.29 %) and congestion levels (0.70 %). The industrial structure analysis reveals that the tertiary sector (5.94 %) exerts a greater influence on HDDTs emissions compared to the secondary (2.54 %) and primary (0.42 %) sectors. In the demographic sphere, the pronounced impact of PGDP (2.33 %) relative to population density (1.31 %) on HDDTs emissions suggests a stronger correlation with economic vitality than urban population concentration.

Fig. 10 illustrates the average contribution (SHAP value) of partial explanatory variables to NOx emissions across various dimensions. In the temporal dimension (Fig. 10(a)-(b)), dynamic traffic factors speed and speed standard deviation (speed stddev) exhibit distinct patterns. Speed shows a higher contribution during T3 (morning and evening peak hours) compared to other periods, while speed stddev does not, likely due to its limited variation during peak hours. In the spatial dimension (Fig. 10(c)-(d)), elevated roads and expressway density contributes more significantly to NOx emissions in Z2 (between Shanghai's inner and outer rings) than in other areas. Conversely, industrial land density has higher contributions in Z3 (outside of Shanghai's outer ring road), suggesting a strong influence of industrial activities on emissions in this region. Regarding policy restrictions (Fig. 10(e)-(h)), speed and elevated roads and expressway density contribute more to PRU3 (policy restriction specific to local/non-local HDDTs) than to other policy conditions. In contrast, industrial land density and elevated bridges and expressway intersection density contribute more to PRU1 (no spatiotemporal policy restrictions), with the latter showing a particularly high contribution.

To further examine the impact of various factors on NOx emissions under different policy conditions, the contribution of each explanatory variable was quantified across PRU1–PRU4. The results indicate that industrial land density, elevated bridges and expressway intersection density, elevated roads and expressway density, and speed consistently emerge as the most influential variables across all policy scenarios. However, as policy restrictions become more stringent, the relative importance of these key factors increases significantly. Specifically, the contribution of industrial land density rises from 29.68 % in PRU1 to 34.73 % in PRU4, while elevated roads and expressway density increases from 11.54 % in PRU1 to 14.84 % in PRU4. Similarly, the contribution of speed grows from 8.19 % in PRU1 to 12.68 % in PRU4. This pattern suggests that under stricter regulatory frameworks, restrictions on non-compliant vehicles, such as those with older emission standards or non-local registrations, lead to a redistribution of freight traffic. This redistribution results in increased reliance on major transportation corridors, particularly elevated expressways, thereby concentrating NOx emissions in areas with high industrial density and extensive expressway networks. Consequently, these factors exert a greater influence on emission estimations as policy stringency escalates.

5.4. Nonlinear effects of the key explanatory variables

5.4.1. Nonlinear association between land use and NOx emissions of HDDTs

Fig. 11 illustrates the nonlinear impacts of industrial land density (Fig. 11(a)-(e)), public administration and service land density (Fig. 11(f)-(j)), commercial land density (Fig. 11(k)-(o)), and land-use mixture (Fig. 11(p)-(t)) on NOx emissions across the total test samples and four policy restriction units. For industrial land density (Fig. 10(a)-(e)), a critical threshold is observed at a density of 0.2 in the overall test sample (Fig. 11(a)). Below this threshold, SHAP values are predominantly negative, indicating that lower densities are associated with reduced NOx emissions. Above this threshold, there is a marked shift to positive SHAP values, indicating that

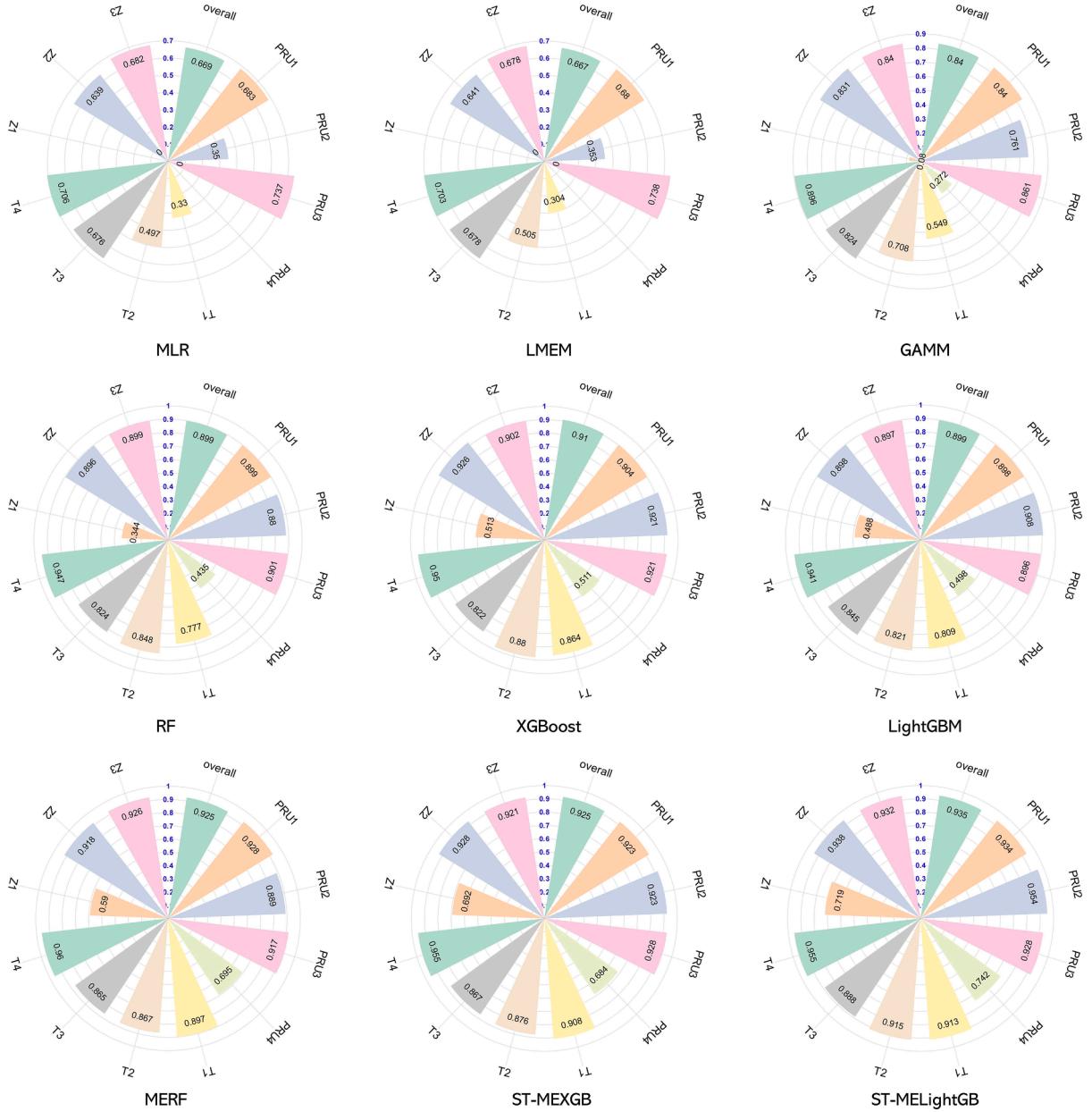


Fig. 8. Multi-dimensional performance assessment (R^2) of three model categories.

higher industrial land density values contribute to increasing the predicted NOx emissions. Furthermore, the SHAP values exhibit a nonlinear increasing trend as industrial land density rises, demonstrating a positive correlation between increasing industrial land density and its impact on NOx emissions. This pattern is largely consistent across PRU1, PRU2, and PRU3 (Fig. 11(b)-(d)). In contrast, PRU4, which is subject to the most stringent restrictions, exhibits consistently negative SHAP values across its narrow range of industrial land density (0–0.1). This observation aligns with the spatial distribution depicted in Fig. 5(e), where the majority of areas within PRU4 are characterized by industrial land density values below 0.1.

Regarding the relationship between public administration and service land density and NOx emissions, analysis of the total sample (Fig. 11(f)) reveals that within the density range of 0 to 0.35, the majority of SHAP values are negative and close to zero, indicating a slight negative contribution to predicted NOx emissions. In the 0 to 0.15 range, SHAP values exhibit a decreasing trend, reaching a minimum at approximately 0.15 density. Upon surpassing the 0.15 threshold, SHAP values exhibit a positive correlation with public administration and service land density, transitioning from negative to positive at 0.35 density. This indicates a shift from a mitigating effect to an augmenting effect on predicted NOx emissions as density increases. This pattern of a shift from negative to positive SHAP values with increasing land density is largely consistent across PRU1, PRU2, and PRU3 (Fig. 11(g), (h), (i)), aligning with the

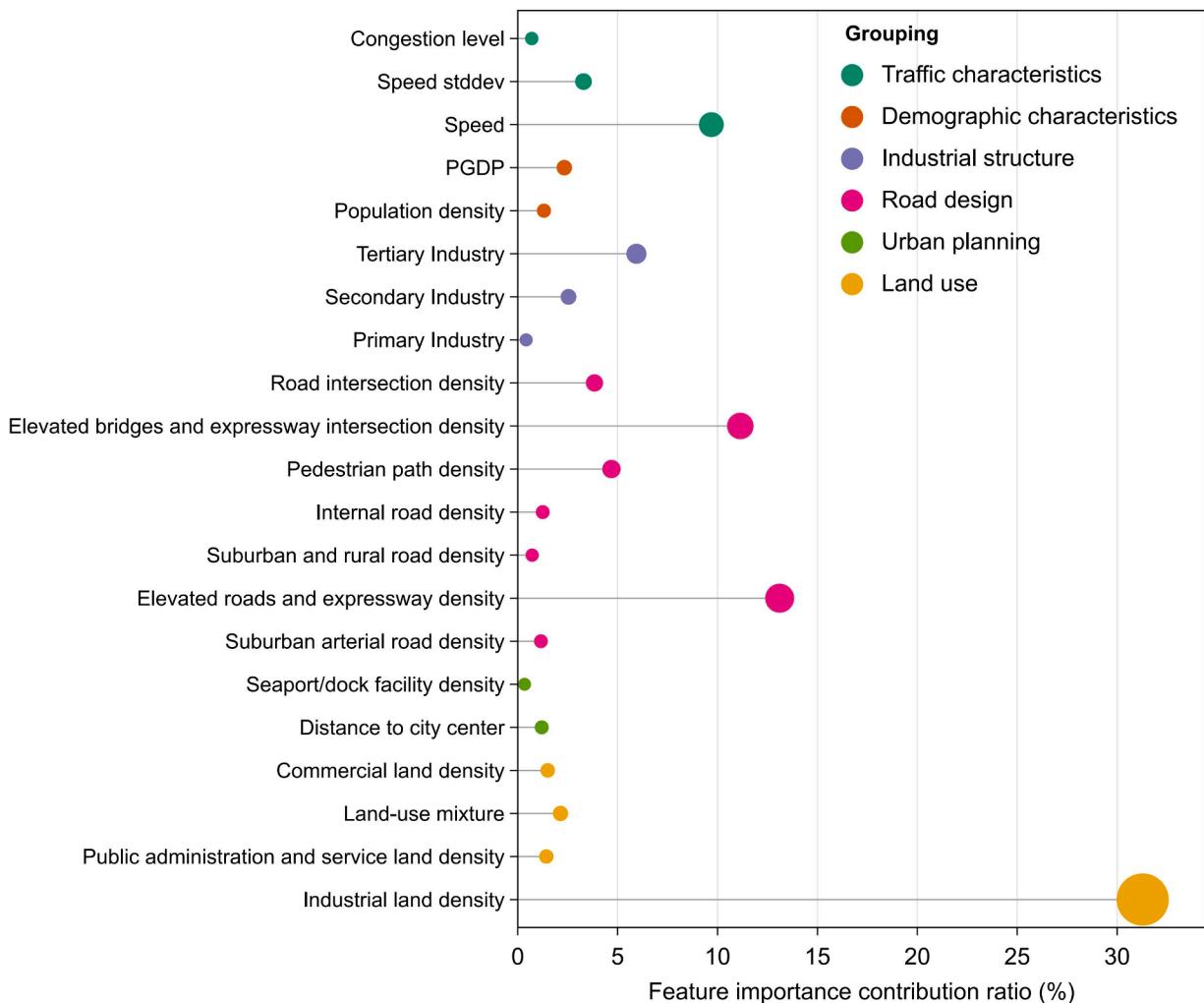


Fig. 9. Contribution ratios of factor groups influencing NOx emissions based on SHAP values.

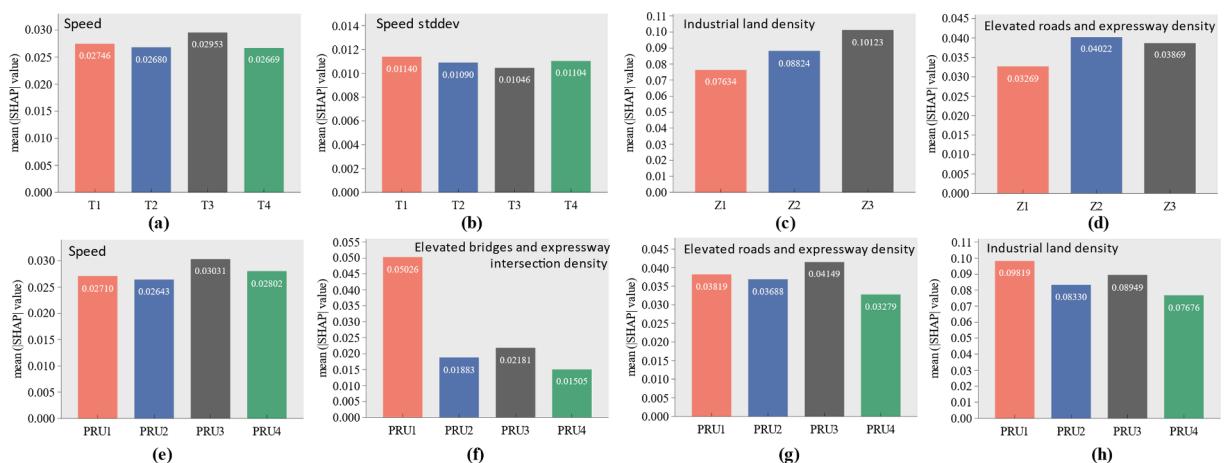


Fig. 10. The mean contribution (SHAP value) of partial explanatory variables to NOx emissions across different dimensions.

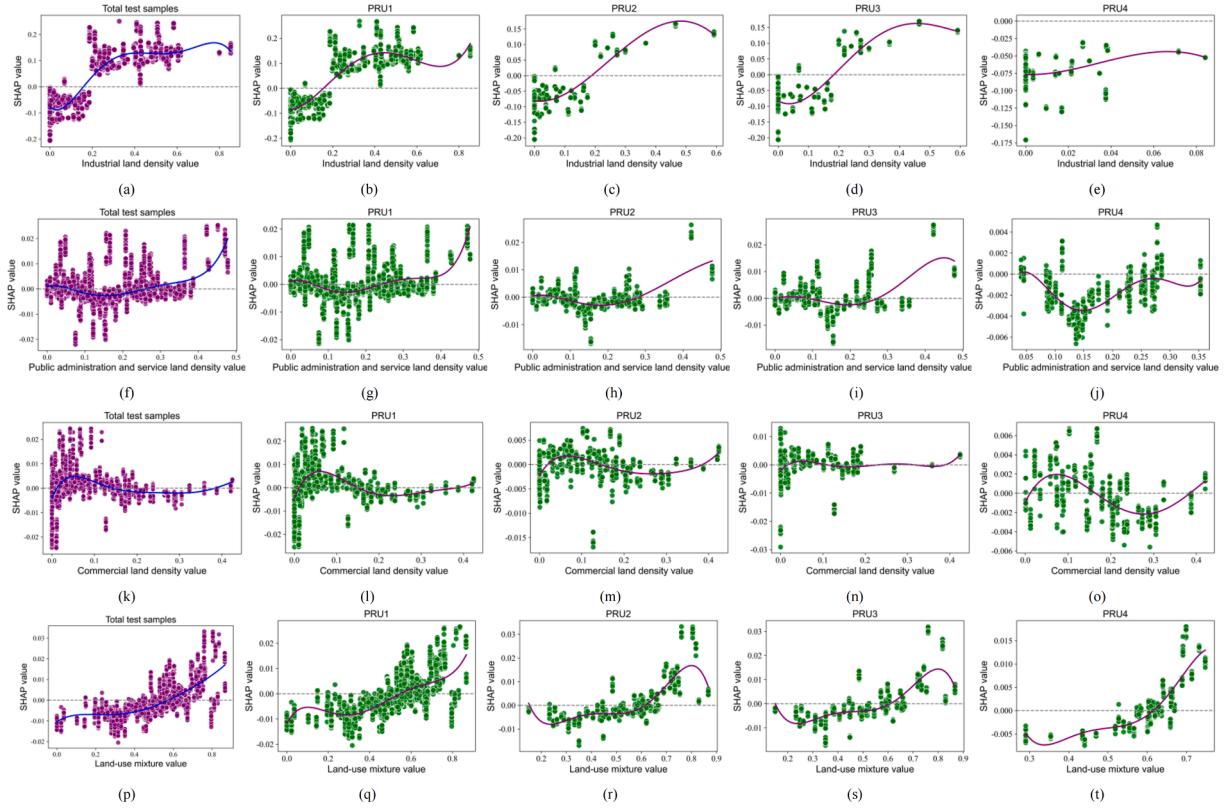


Fig. 11. Nonlinear relationships between land use density and NOx emissions across different policy restriction units (SHAP value > 0 : increasing the predicted NOx emission from the base value; SHAP value < 0 : decreasing the predicted NOx emission from the base value).

observations from the overall sample. In contrast, PRU4 (Fig. 11(j)), under the most stringent policy constraints, the SHAP values persistently retain a negative disposition across all density gradients, consistently signaling a diminution in predicted emissions relative to the baseline, underscoring the effectiveness of stringent policies in suppressing emissions.

The relationship between commercial land density and NOx emissions reveals distinct patterns. In the total sample and PRU1-3 (Fig. 11(k)-(n)), SHAP values are positive when commercial land density ranges between 0 and 0.12, with a local maximum observed within the 0–0.05 range. As density increases to 0.3, SHAP values demonstrate a gradual decline, transitioning to predominantly negative values. This trend suggests that lower commercial densities initially contribute to increased predicted NOx emissions relative to the base value, while higher densities tend to mitigate emissions. In PRU4 (Fig. 11(o)), SHAP values exhibit high variability across different commercial land densities, yet the generally smaller absolute values indicate a diminished impact on NOx emissions due to stringent policy conditions.

For the land-use mixture, the total test sample and PRU1 (Fig. 11(p)) display a threshold effect at a mixture ratio of 0.58. Below this threshold, the SHAP values are negative, indicating that lower mixture ratios contribute to decreasing the predicted NOx emissions. However, above this 0.58 threshold, the SHAP values shift to positive and steadily increase as the mixture ratio rises further, suggesting a positive contribution to NOx emissions at higher mixture levels. Interestingly, for PRU2, PRU3, and PRU4, this critical threshold shifts to a higher mixture ratio of 0.65. This implies that the policy interventions in these more restricted zones may extend the range over which mixed land-use patterns contribute to lower NOx emissions, compared to the overall sample and PRU1.

To summarize the above results in the land use patterns, we can reveal that when the following conditions occur, it is conducive to reducing NOx emissions from HDDTs:

(a) **Threshold effect:** The industrial land density is below a critical threshold (< 0.2); The public administration and service land density is maintained within a moderate range (0–0.35), with optimal effects in the lower range (0–0.15); The commercial land density initially (0–0.12) shows positive SHAP values, which transition to negative values at higher densities, suggesting a potential mitigating effect on NOx emissions in denser commercial areas; The land-use mixture ratio is below a certain threshold (< 0.58 in general scenarios, or < 0.65 in policy-restricted environments);

(b) **Policy restriction effectiveness:** In the absence of policy restrictions, increasing industrial land density, public administration and service land density, and land-use mixture significantly increases the NOx emission contribution of HDDTs. However, under stricter policy restrictions, such as PRU4, these land use factors contribute significantly to reducing NOx emissions. Moreover, the implementation of policy restrictions (PRU2-PRU4) modulates the emission-increasing trends observed in unrestricted conditions. This

is particularly evident in the land-use mixture model, where policy interventions extend the threshold for NOx emission reduction from a mixture ratio of 0.58 to 0.65.

5.4.2. Nonlinear association between road design and NOx emissions of HDDTs

Fig. 12 illustrates the nonlinear impacts of elevated roads and expressway density (**Fig. 12(a)-(e)**), suburban and rural road density (**Fig. 12(f)-(j)**), road intersection density (**Fig. 12(k)-(o)**), and elevated bridges and expressway intersection density (**Fig. 12(p)-(t)**) on NOx emissions across the total test samples and four policy restriction units. For elevated roads and expressway density (**Fig. 12(a)-(e)**), a positive correlation with NOx emissions is observed across all dimensions when the density ranges between 0 and 1.4 km/km². Beyond 1.4 km/km², the relationship stabilizes (**Fig. 12(a), (b)**, suggesting a saturation point where additional high-capacity infrastructure does not further increase NOx emissions. Besides, in areas where elevated road and expressway density is above 0.68 km/km², implement targeted traffic management measures to mitigate the increase in NOx emissions.

The relationship between suburban and rural road density and NOx emissions exhibits complex nonlinearity, particularly in the overall sample (**Fig. 12(f)**) and PRU1 (**Fig. 12(g)**). Negative SHAP values at low densities (<0.9 km/km²) suggest that sparse rural road networks may contribute to reduced emissions, possibly due to lower traffic volumes. The positive SHAP values observed in the 0.9–3.4 km/km² range, peaking at 2.5 km/km², indicate an optimal density beyond which emissions begin to decrease. Notably, in PRU4 (**Fig. 12(j)**), where spatiotemporal restrictions are the strictest, SHAP values are predominantly positive and increase monotonically with suburban and rural road density, indicating a strong positive correlation with NOx emissions. This pattern suggests that even with stringent spatiotemporal policy restrictions, increased road density in suburban and rural areas leads to higher NOx emissions.

In the total samples (**Fig. 12(k)**), PRU1(**Fig. 12(l)**), PRU2(**Fig. 12(m)**), and PRU4(**Fig. 12(o)**), the relationship between road intersection density and NOx emissions exhibits a U-shaped pattern. SHAP values decrease and then increase with increasing intersection density. Positive SHAP values are observed for densities up to 80 num/km², while SHAP values are negative beyond 80 num/km², reaching a minimum at approximately 200 num/km². The U-shaped relationship implies that both very low and very high intersection densities are associated with lower NOx emissions, and moderate densities (80–200 num/km²) may require targeted interventions to reduce emissions.

The relationship between elevated bridges and expressway intersection density and NOx emissions exhibits clear nonlinearity (**Fig. 12(p)-(t)**). A positive correlation is observed for densities up to 10 num/km², beyond which SHAP values stabilize while remaining positive. This suggests that while increasing the density of elevated infrastructure initially contributes to higher emissions,

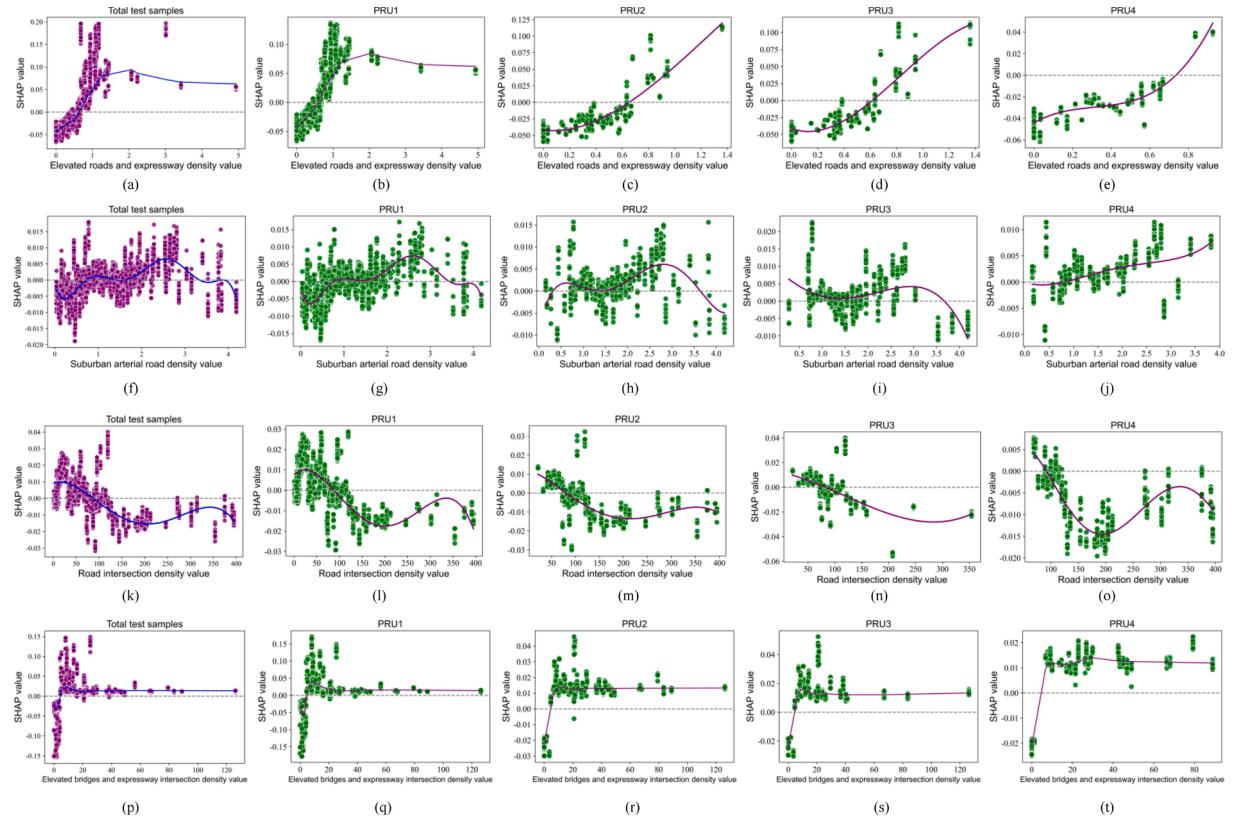


Fig. 12. Nonlinear relationships between road design and NOx emissions across different policy restriction units.

there is a threshold beyond which further increases have minimal additional impact.

To summarize the above results regarding road infrastructure characteristics, we can reveal that the following conditions are conducive to reducing NOx emissions from HDDTs:

(a) **Threshold effect:** Elevated roads and expressway density between 0.68 and 1.4 km/km² contributes to increased emissions, necessitating targeted management measures when the density exceeds 0.68 km/km²; Suburban and rural roads decrease NOx emissions when density is below 0.9 km/km² or above 3.4 km/km², but increase emissions in the medium density range (0.9–3.4 km/km²), with peak contribution at 2.5 km/km²; Road intersection density is most effective in reducing NOx emissions at a medium density of about 200 num/km²; Elevated bridges and expressway intersection density below 10 num/km² is conducive to emission reduction, while exceeding this threshold positively contributes to increased NOx emissions.

(b) **Policy restriction effectiveness:** In the absence of policy restrictions, the four facility types exhibit distinct emission contribution patterns. Elevated roads and expressway density and its corresponding intersection density show a positive correlation with emissions up to a certain threshold, after which the relationship plateaus. Road intersection density initially demonstrates a negative correlation with emissions, followed by a positive trend. Conversely, “Suburban and rural roads density” displays an initial positive correlation, transitioning to a negative one. Under stricter policy restrictions, particularly in PRU4 (the most stringent spatiotemporal regulation), divergent emission reduction effects emerge. While elevated roads and expressway density, as well as road intersection density, contribute to NOx emission reduction within specific ranges (Fig. 12(e), (o)), suburban and rural road density, along with elevated bridge and expressway intersection density, appears to exacerbate emissions (Fig. 12(j), (t)). This phenomenon may be attributed to the redistribution of HDDT traffic under PRU4, where restrictions in core urban areas force non-compliant vehicles onto suburban and rural roads. Unlike expressways, these roads often feature more frequent intersections, stop-and-go traffic, and greater variability in driving speeds, all of which can lead to higher NOx emissions. Thus, while PRU4 effectively reduces emissions in highly regulated areas, it may contribute to localized emission increases in alternative routes. Based on these findings, we propose targeted mitigation strategies. Under PRU1, PRU2, and PRU3 conditions, promoting HDDT use of secondary roads in areas with high elevated road and expressway density can help distribute traffic more evenly and alleviate congestion-induced emissions.

5.4.3. Nonlinear association between demographic and industrial with NOx emissions of HDDTs

Fig. 13 shows the nonlinear impacts of secondary industry (Fig. 13(a)-(e)), population density (Fig. 13(f)-(j)) on NOx emissions for the overall test sample and four policy restriction units. In the analysis of secondary industry density, for densities between 0 and 0.4 num/km², SHAP values are negative but increasing, indicating a diminishing negative impact on NOx emissions. As density exceeds 0.4 num/km², SHAP values turn positive, suggesting that higher concentrations of secondary industries contribute to increased NOx emissions. The overall test sample (Fig. 13(a)) and PRU1 (Fig. 13(b)) display similar trends, with SHAP values peaking at approximately 0.7 num/km² within the 0–1.5 num/km² range. This pattern implies that moderate levels of industrial concentration may have the most significant positive impact on NOx emissions, possibly due to agglomeration effects and economies of scale in pollution control.

For population density, the overall sample (Fig. 13(f)) and PRU1 (Fig. 13(g)) results indicate that NOx emissions are positively correlated with population density when it is below 8000 person/km². SHAP values decrease between 8000 and 31,000 person/km² and then increase again. However, in PRU2, PRU3, and PRU4, the nonlinear relationship is less pronounced. Notably, in PRU2 (Fig. 13(h)) and PRU4 (Fig. 13(j)), SHAP values are negative when population density is below 41,000 person/km², after which they increase and become positive, demonstrating the moderating effect of policy restrictions.

To summarize the above results regarding secondary industry and population density characteristics, we can reveal that the following conditions are conducive to reducing NOx emissions from HDDTs:

(a) **Threshold effect:** The secondary industry density is below a critical threshold (< 0.4 num/km²); In the absence of policy restrictions, the population density is most likely to contribute to emissions when it is around 10,000 person/km², indicating that policy restrictions should be specifically implemented in this range.

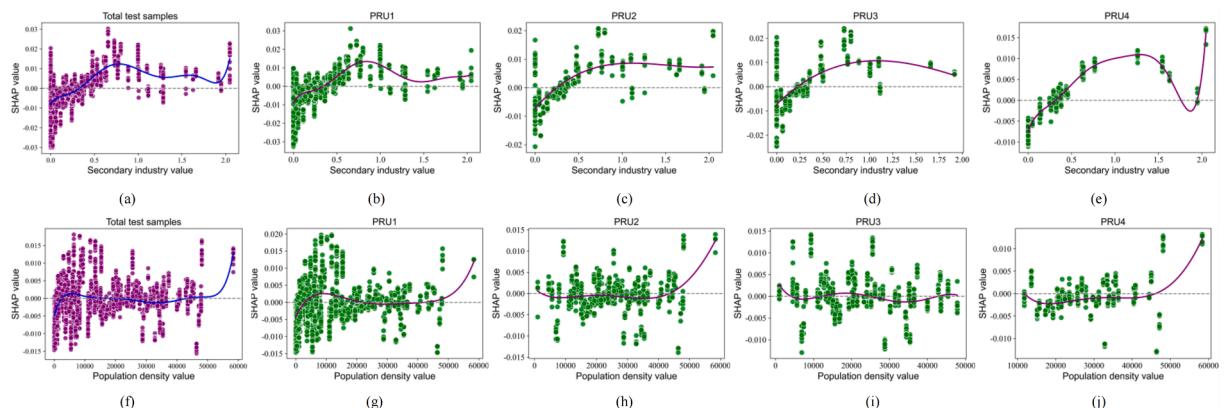


Fig. 13. Nonlinear relationships between demographic and industrial with NOx emissions.

(b) Policy restriction effectiveness: Under policy restrictions, particularly the strictest regulations (PRU4), population density demonstrates a significant effect in reducing NOx emissions. Conversely, even with the strictest restrictions, SHAP values remain positive for secondary industry density above 0.4 num/km², underscoring the role of industrial concentration in contributing to NOx emissions. Based on these findings, we propose the following targeted recommendations: 1) Maintain and potentially expand strict policy measures in densely populated areas, as they prove effective in reducing NOx emissions from HDDTs; 2) Implement strict zoning policies to limit secondary industry density to below 0.4 num/km² in urban and suburban areas. For existing high-density industrial areas, develop long-term plans for gradual relocation or restructuring of industries to reduce concentration.

5.4.4. Nonlinear associations of speed and distance to city center with NOx emissions from HDDTs

Fig. 14 demonstrates the nonlinear impacts of speed (Fig. 14(a)-(i)) and distance to city center (Fig. 14(j)) on NOx emissions from HDDTs. The overall test sample (Fig. 14(a)) and PRU1 (Fig. 14(b)) demonstrate similar nonlinear relationships between speed and NOx emissions. For speed below 42 km/h, SHAP values are positive and decreasing, indicating a positive but diminishing effect on NOx emissions. Above 42 km/h, SHAP values become negative, with a local minimum occurring between 60–70 km/h, suggesting this speed range is optimal for minimizing NOx emissions. In contrast, PRUs with stricter policy restrictions (PRU2, PRU3, and PRU4) show more pronounced positive effects of speed on NOx emissions. The critical point shifts to 44 km/h and 46 km/h for PRU2 and PRU3, respectively. Notably, in PRU4, the most strictly regulated area, speed consistently contributes to higher NOx emissions, as evidenced by consistently positive SHAP values (Fig. 14(e)). Since speed is a time-varying factor, we also analyzed its relationship with NOx emissions across four time periods (T1-T4). Fig. 14(f) and (g) show that the nonlinear trends during T1 (1:00–6:00 AM) and T2 (6:00–7:00 AM and 8:00 PM–1:00 AM) are similar to those observed in the overall sample. Notably, during T3 (morning peak 7:00–10:00 AM and evening peak 5:00–8:00 PM), more samples exhibit positive SHAP values, indicating a stronger positive impact of speed on NOx emissions due to higher traffic volumes and potential congestion during peak hours.

For distance to city center (Fig. 14(j)), within the immediate urban core (0–18 km), NOx emissions show a positive correlation with distance. This trend reverses in the suburban and *peri-urban* areas (18–105 km), where a negative correlation is observed. Besides, positive SHAP values, indicating an increase in predicted NOx emissions from the base value, are confined to the 8–35 km range. All other distances are associated with negative SHAP values, suggesting a reduction in predicted NOx emissions relative to the base value.

6. Policy implications and conclusions

6.1. Policy implications

(1) Industrial land density is the primary contributor (31.28 %) to NOx emissions among land use factors, while commercial land density has a significantly lower impact (1.50 %). This pattern contrasts with passenger car emission studies, where commercial land density typically plays a more prominent role (Wu et al., 2023a). The distinct characteristics of freight emissions compared to passenger vehicles necessitate a differentiated approach to vehicle restriction policies. While bus-oriented policies focus on high-density commercial areas, HDDT regulations should prioritize industrial zones and areas with high mixed land use density. Cities could benefit from establishing dedicated freight corridors optimized for emission reduction, potentially linking industrial zones with key logistics hubs. These freight-specific green corridors would address the unique emission patterns of HDDTs and contribute to more efficient and environmentally friendly urban logistics.

(2) In the road design factors, NOx emissions are most influenced by elevated roads and expressway density (13.11 %) and their intersections (11.14 %). Pedestrian path density (4.69 %) also makes a significant contribution compared to other road types, likely due to increased acceleration and deceleration in pedestrian-centric areas (Zavala-Reyes et al., 2019). Based on the analysis of the results, we propose the following recommendations: 1) Dynamic traffic management and route optimization: Implement dynamic traffic management systems in areas where elevated road and expressway density surpass critical thresholds. Optimize traffic signal timing in regions with high road intersection density to maximize emission reduction benefits. Encourage HDDTs to use secondary roads in areas with high elevated road and expressway density, especially under policy restriction conditions, to distribute traffic more evenly and reduce congestion on primary routes. 2) Promote alternative delivery methods and electrification initiatives: In pedestrian-centric areas with high path density, promote alternative last-mile delivery methods, such as cargo bikes or small electric vehicles, to reduce HDDT traffic.

(3) For industrial structure and demographic factors, the tertiary sector (5.94 %) has the most significant impact on NOx emissions, followed by the secondary (2.54 %) and primary (0.42 %) sectors. In terms of demographics, PGDP (2.33 %) shows a stronger correlation with HDDT emissions than population density (1.31 %), suggesting that economic vitality influences emissions more than urban population concentration. Secondary industry density exhibits a threshold effect at 0.4 num/km², beyond which NOx emissions increase substantially, highlighting the crucial role of industrial concentration in emission patterns. Population density's impact on emissions varies across policy restriction units, with densities around 10,000 person/km² contributing the most to emissions in unrestricted areas. While policy restrictions effectively moderate emissions, especially in high-density urban areas, their impact on industrial-related emissions is less pronounced, indicating a need for targeted industrial policies. Based on these insights, we propose the following recommendations: 1) Maintain and expand strict policy measures: Enforce stringent policies in densely populated areas to reduce emissions; 2) Implement strict zoning policies: Limit secondary industry density to below 0.4 num/km² in urban and suburban areas in future urban planning. For existing high-density industrial areas, develop long-term plans for the gradual relocation or restructuring of industries to reduce concentration.

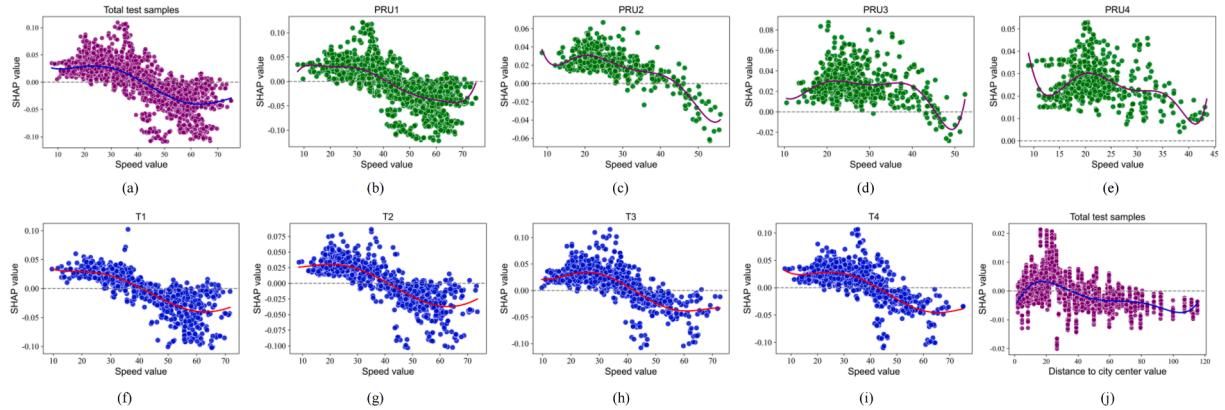


Fig. 14. Nonlinear relationships between speed and distance to city center with NOx emissions.

6.2. Conclusion and prospects

We developed a novel interpretable machine learning framework to explore the nonlinear relationship between influencing factors and truck emissions. In this framework, we designed the ST-MELightGB model to capture time-specific variations, location-specific effects, and the spatial-temporal heterogeneity of HDDT policy impacts. Compared to traditional linear models and machine learning algorithms, the ST-MELightGB algorithm demonstrated significantly better predictive performance due to its comprehensive consideration of data heterogeneity. Especially in local prediction tasks, we found that linear models and machine learning models face greater difficulties in modeling in the central zones with stringent policy restrictions, but the ST-MELightGB model achieved substantial improvements in these challenging regions. Finally, we employed SHAP analysis to identify the critical characteristics of distinct policy restriction zones and to quantify their nonlinear, localized spatiotemporal impacts on HDDT emissions. The analysis revealed that land use and road design are primary contributors to NOx emissions, while industrial structure has a moderate influence, and demographics and urban planning have minimal impact. Specifically, industrial land density and expressway-related features consistently play crucial roles, but their relative importance varies across spatial zones and policy scenarios.

However, this study also has limitations that future research could address. While our analysis focuses solely on Shanghai, expanding the scope to encompass urban agglomerations could offer a more nuanced understanding of HDDTs' emission behaviors in complex, interconnected urban environments. This approach would provide insights into how urban form, transportation networks, and emissions interact across different scales and contexts. Additionally, the effects of the Covid-19 pandemic on traffic flow and emissions, which were not explicitly analyzed here, warrant further exploration, and future studies could incorporate data from both the pandemic and post-pandemic periods for a comparative analysis. Furthermore, although our current study has addressed several factors related to Freight Transportation Centers (FTC), the scope of this research is limited by data availability, which constrained the ability to fully analyze the combined effects of various FTC-related factors on HDDT emissions. Therefore, further research is needed to comprehensively explore the spatial distribution, density, and operational characteristics of FTCs in different regions, and to better understand how these factors collectively influence emissions.

CRediT authorship contribution statement

Tongtong Shi: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Meiting Tu:** Writing – review & editing, Methodology, Funding acquisition, Data curation, Conceptualization. **Ye Li:** Writing – review & editing, Supervision, Resources. **Haobing Liu:** Formal analysis, Funding acquisition, Supervision, Writing – review & editing. **Dominique Gruyer:** Writing – review & editing, Supervision.

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.

Acknowledgments

This research is supported by National Key R&D Program of China (grant number. 2023YFB3906900), and the National Natural Science Foundation of China [grant number 52302441].

- Shi, K., Yu, B., Zhou, Y., Chen, Y., Yang, C., Chen, Z., Wu, J., 2019. Spatiotemporal variations of CO₂ emissions and their impact factors in China: A comparative analysis between the provincial and prefectural levels. *Appl. Energy* 233–234, 170–181. <https://doi.org/10.1016/j.apenergy.2018.10.050>.
- Shao, Q., Zhang, W., Cao, X.J., Yang, J., 2023. Built environment interventions for emission mitigation: A machine learning analysis of travel-related CO₂ in a developing city. *Journal of Transport Geography* 110, 103632. <https://doi.org/10.1016/j.jtrangeo.2023.103632>.
- Shi, T., Wang, P., Qi, X., Yang, J., He, R., Yang, J., Han, Y., 2023. CPT-DF: Congestion Prediction on Toll-Gates Using Deep Learning and Fuzzy Evaluation for Freeway Network in China. *J. Adv. Transp.* 2023, 2941035. <https://doi.org/10.1155/2023/2941035>.
- Sigrist, F., 2023. Latent Gaussian Model Boosting. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 1894–1905. <https://doi.org/10.1109/TPAMI.2022.3168152>.
- Stohl, A., Aamaas, B., Amann, M., Baker, L.H., Bellouin, N., Berntsen, T.K., 2015. Evaluating the climate and air quality impacts of short-lived pollutants. *Atmos. Chem. Phys.* 15, 10529–10566. <https://doi.org/10.5194/acp-15-10529-2015>.
- Sun, S., Yang, S., Guo, L., Cao, X., (Jason), 2024. Individual and synergistic associations among transit allowance, land use, and transit commuting. *Transp. Res. Part D: Transp. Environ.* 128, 104100. <https://doi.org/10.1016/j.trd.2024.104100>.
- Tang, G., Wang, S., Du, B., Cui, L., Huang, Y., Xiao, W., 2022. Study on pollutant emission characteristics of different types of diesel vehicles during actual road cold start. *Sci. Total Environ.* 823, 153598. <https://doi.org/10.1016/j.scitotenv.2022.153598>.
- Tu, M., Li, W., Orfila, O., Li, Y., Guyer, D., 2021. Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. *Transp. Res. Part D: Transp. Environ.* 93, 102776. <https://doi.org/10.1016/j.trd.2021.102776>.
- van Zyl, C., Ye, X., Naidoo, R., 2024. Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP. *Appl. Energy* 353, 122079. <https://doi.org/10.1016/j.apenergy.2023.122079>.
- Veerappa, M., Anneken, M., Burkart, N., Huber, M.F., 2022. Validation of XAI explanations for multivariate time series classification in the maritime domain. *Journal of Computational Science* 58, 101539. <https://doi.org/10.1016/j.jocs.2021.101539>.
- Wu, J., Jia, P., Feng, T., Li, H., Kuang, H., 2023. Spatiotemporal analysis of built environment restrained traffic carbon emissions and policy implications. *Transp. Res. Part D: Transp. Environ.* 121, 103839. <https://doi.org/10.1016/j.trd.2023.103839>.
- Wu, X., Tao, T., Cao, J., Fan, Y., Ramaswami, A., 2019. Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. *Transp. Res. Part D: Transp. Environ.* 75, 1–12. <https://doi.org/10.1016/j.trd.2019.08.018>.
- Yang, L., Yang, H., Yu, B., Lu, Y., Cui, J., Lin, D., 2024. Exploring non-linear and synergistic effects of green spaces on active travel using crowdsourced data and interpretable machine learning. *Travel Behav. Soc.* 34, 100673. <https://doi.org/10.1016/j.tbs.2023.100673>.
- Yang, W., Cao, X., 2018. Examining the effects of the neighborhood built environment on CO₂ emissions from different residential trip purposes: A case study in Guangzhou, China. *Cities* 81, 24–34. <https://doi.org/10.1016/j.cities.2018.03.009>.
- Yang, Z., Chen, X., Deng, J., Li, T., Yuan, Q., 2023. Footprints of goods movements: Spatial heterogeneity of heavy-duty truck activities and its influencing factors in the urban context. *J. Transp. Geogr.* 113, 103737. <https://doi.org/10.1016/j.jtrangeo.2023.103737>.
- Zavala-Reyes, J.C., Jeanjean, A.P.R., Leigh, R.J., Hernández-Paniagua, I.Y., Rosas-Pérez, I., Jazcilevich, A., 2019. Studying human exposure to vehicular emissions using computational fluid dynamics and an urban mobility simulator: The effect of sidewalk residence time, vehicular technologies and a traffic-calming device. *Sci. Total Environ.* 687, 720–731. <https://doi.org/10.1016/j.scitotenv.2019.05.422>.
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J., 2020. A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends* 1, 56–70. <https://doi.org/10.38094/jastt1224>.
- Zhang, S., Fu, M., Zhang, H., Yin, H., Ding, Y., 2025. Emission control status and future perspectives of diesel trucks in China. *J. Environ. Sci.* 148, 702–713. <https://doi.org/10.1016/j.jes.2023.06.010>.
- Zhao, P., Li, Z., Xiao, Z., Jiang, S., He, Z., Zhang, M., 2023. Spatiotemporal characteristics and driving factors of CO₂ emissions from road freight transportation. *Transp. Res. Part D: Transp. Environ.* 125, 103983. <https://doi.org/10.1016/j.trd.2023.103983>.
- Zhi, D., Zhao, H., Chen, Y., et al., 2024. Quantifying the heterogeneous impacts of the urban built environment on traffic carbon emissions: New insights from machine learning techniques. *Urban Climate* 53, 101765. <https://doi.org/10.1016/j.uclim.2023.101765>.
- Zhu, H., Peng, J., Dai, Q., Yang, H., 2024. Exploring the long-term threshold effects of density and diversity on metro ridership. *Transp. Res. Part D: Transp. Environ.* 128, 104101. <https://doi.org/10.1016/j.trd.2024.104101>.