

Simulation and Feature Analysis of Road Carbon Emissions in Wuhan Based on Random Forest

Yuan Wang¹, Sixian Qin², Haiting Li³, Xuewei Zhang⁴, Bo Tan⁵

¹ Wuhan Geomatics Institute, Wuhan 430022, China - 573094354@qq.com

² Wuhan Geomatics Institute, Wuhan 430022, China -sixianqin@qq.com

³ Wuhan Geomatics Institute, Wuhan 430022, China - 120059329@qq.com

⁴ Wuhan Geomatics Institute, Wuhan 430022, China - 799037417@qq.com

⁵ Wuhan Geomatics Institute, Wuhan 430022, China - 327840525@qq.com

Keywords: Machine Learning, Random Forest, Road CO₂ Concentration, Traffic Carbon Emissions, GIS Technology

Abstract

Against the backdrop of global climate change, reducing CO₂ emissions from transportation is urgent. This study focuses on Wuhan, integrating multi-source data such as road monitoring, traffic flow, vehicle types, and speeds to build a spatiotemporal carbon emission model using the Random Forest algorithm. Results show the model achieves an R² of 0.74 and RMSE of 26.22 ppm, accurately simulating hourly road CO₂ variations. Emissions exhibit "peak concentration, directional asymmetry, and arterial dependency," with morning peaks averaging 492.46 ppm (peak 804.18 ppm), evening peaks at 488.79 ppm (peak 788.27 ppm), and nighttime lows averaging 477.31 ppm. Enclosed corridors like the East Lake Tunnel show significantly higher CO₂ levels (550 – 810 ppm), while radial arterials connecting urban cores and peripheries account for over half of total emissions. Inner and second ring roads act as emission hotspots, with concentrations 3.9% – 4.5% higher than outer rings. Nighttime emissions drop by 4.7% – 5.3%. Tunnels exhibit the highest average CO₂ (620 ppm), 28.8% above other road types, due to restricted exhaust dispersion and high traffic density. These findings highlight the need for optimized urban planning and traffic management.

1. Introduction

Under the global climate governance framework, the temperature control goals established by the Paris Agreement (UNFCCC, 2016) have accelerated low-carbon transitions worldwide. As the world's largest carbon emitter, China has pledged to achieve a carbon peak by 2030 through its “Dual Carbon” strategy (NDRC, 2022), leveraging low-carbon pilot cities as platforms to explore differentiated emission reduction pathways (He et al., 2022). Wuhan, a representative megacity, faces acute carbon emission challenges in its transportation sector: 2024 data reveals that transportation-related CO₂ emissions (excluding aviation) reached 13.78 million tons, accounting for 15% of the city's total emissions (Zhang et al., 2024), with an annual growth rate (5.8%) far exceeding those of industrial and building sectors. The spatiotemporal heterogeneity of CO₂ concentrations in central urban areas further underscores the urgent need for refined governance and innovative methodologies to enable precise emission control.

Current research on traffic carbon emissions faces three key constraints: (1) Traditional energy inventory methods lack spatial resolution to support road-segment-level emission accounting (Zhang et al., 2023); (2) Existing mobile monitoring technologies primarily focus on atmospheric pollutants, with insufficient spatiotemporal coverage for greenhouse gas measurements. For instance, short-term studies (e.g., an 8-day campaign in Shanghai) successfully characterized NH₃ emissions but failed to address complex road network monitoring demands (Yang, H., 2023); (3) Mainstream emission factor models (e.g., MOVES) generate aggregated estimates for entire road segments or regions, lacking detailed spatiotemporal disaggregation and resulting in significant simulation biases during peak hours. To address these challenges, the Wuhan Low-Carbon Transportation Development Plan (2024) explicitly prioritizes the development of road-level emission inventories, with emerging mobile sensing technologies offering novel solutions. Recent studies demonstrate that high-resolution greenhouse gas concentration data, combined with traffic

parameters (Yang et al., 2022) and machine learning algorithms (e.g., Random Forest models), can effectively resolve spatiotemporal emission patterns (Yifan W., 2022; Chaerin P., 2023). However, their application in dynamic carbon emission simulations remains underexplored.

This study focuses on dynamic simulation of transportation carbon emissions in Wuhan, innovatively integrating vehicle-mounted sensor networks and multi-source data to construct an hourly prediction model. Through continuous monitoring across diverse traffic scenarios (e.g., cross-river tunnels and elevated expressways), we establish the first mobile carbon monitoring dataset for a megacity, addressing critical data gaps. By leveraging Random Forest algorithms to overcome the linear assumption limitations of traditional models, we achieve high-precision dynamic simulations under complex traffic flow conditions. Finally, we unveil spatiotemporal emission differentiation patterns and propose targeted governance strategies. The outcomes not only provide scientific foundations for low-carbon transportation management in Wuhan but also offer a replicable paradigm for similar cities to advance carbon reduction through “data-model” synergy.

2. Materials and Methods

2.1 Study area

Wuhan (29° 58' – 31° 22' N, 113° 41' – 115° 05' E), located in central China along the middle reaches of the Yangtze River, is the capital of Hubei Province, a National Central City, and a core hub of the Yangtze River Economic Belt. As a pivotal national transportation junction and advanced manufacturing base, it covers a total area of 8,569.15 km², comprising 13 administrative districts. By the end of 2023, its permanent population reached 13.774 million, with an urbanization rate of 84.79%. The city experiences a northern subtropical monsoon humid climate, characterized by an annual average temperature of 16.6°C, precipitation of 1,269 mm, relative humidity of 75–80%, and distinct seasonal variations.

Wuhan's road system features a three-dimensional "ring-radial" structure, including 5 ring roads, 21 bridges and 2 tunnels spanning the Yangtze River, and 18 arterial roads, with a road network density of 6.26 km/km². Rapid urbanization has driven explosive growth in motor vehicle ownership, surging from 2.1325 million in 2015 to 4.497 million in 2023, representing an annual growth rate of 9.8%. During weekday morning peak hours in the main urban area, the average traffic speed is approximately 26.2 km/h, with a congestion delay index of 1.92. Critical corridors such as the Jiefang Avenue Second Ring Road and Yangtze River Second Bridge exhibit particularly severe traffic-related pollution emissions. Consequently, Wuhan faces intensifying challenges of urban traffic congestion and air pollution, necessitating urgent research on road-level CO₂ concentration patterns (Wuhan Transportation Development Annual Report, 2024).

2.2 Methodology

This study proposes a technical framework for simulating road carbon emissions in Wuhan based on the Random Forest algorithm, comprising three steps (figure 1): data collection, construction of road CO₂ emission simulation model and spatiotemporal feature analysis. First, a mobile monitoring system was used to collect multi-source data, which were fused and processed to extract features related to road categories, traffic operations, meteorological conditions, and environmental factors. Next, a Random Forest model was employed to build the spatiotemporal carbon emission model, where the dataset was divided for model training and validation. Finally, the spatiotemporal characteristics of urban road carbon emissions were analyzed based on the model, including the impacts of different time periods, administrative districts, and road types on emissions.

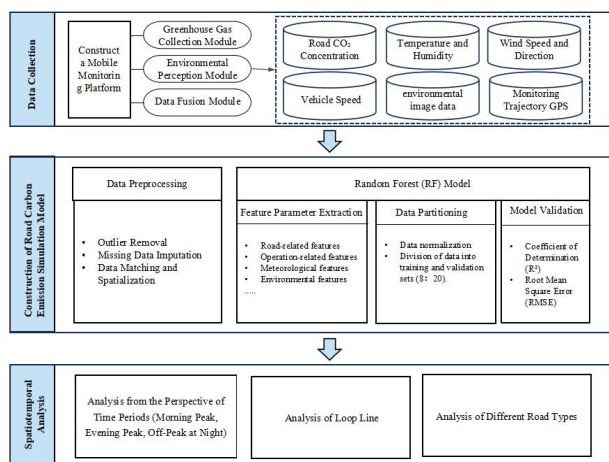


Figure 1. Overall technical framework.

2.2.1 Data collection

Following the technical specifications for mobile greenhouse gas (CO₂ and CH₄) monitoring, this study integrated multi-source sensing devices to construct a high-precision mobile monitoring platform. The system's core modules include: (1) Greenhouse Gas Analysis Module, Configured with a Picarro G2301 (CO₂ measurement accuracy: 0.1 ppm) and Vaisala GMP343 (accuracy: 1–2 ppm) to acquire greenhouse gas data at varying resolutions; (2) Environmental Perception Module, Integrated GPS sensors (recording trajectories and driving status), panoramic cameras (360° image acquisition), and meteorological sensors (synchronized measurements of

temperature, humidity, wind speed, and direction); (3) Data Fusion Module, Enabled real-time multi-source data transmission via customized communication protocols, with high-precision time synchronization technology (error <1 ms) ensuring spatiotemporal consistency. The system is equipped with dual battery packs supporting 10-hour continuous operation.

Based on the "Five Rings and Eighteen Radiations" road network characteristics of Wuhan City and the spatiotemporal distribution patterns of traffic flow, this study designed nine mobile monitoring routes covering all road network types, with a total length of 1,676 kilometers (figure 2). These routes prioritized monitoring expressways, cross-river tunnels, and industrial clusters (e.g., areas around steel plants). The route planning integrated considerations of land use diversity (residential, commercial, industrial, and green spaces) and traffic tidal patterns. Monitoring was conducted in two batches (table 1), spanning two complete weeks (including weekdays and non-weekdays) and 24 hours per day. Ultimately, the study accumulated 5,139 kilometers of mobile monitoring mileage, collecting 8,198 entries of minute-by-minute CO₂ concentration data and 13,872 environmental images.

Batch	Date	Period Type	Areas	km	Data (Records)
1	2024-08-03 to 2024-08-09	Morning/Evening Peak: 7:00-9:30, 17:00-19:30 Midday: 11:00-15:00 Night: 23:00-5:00	1. Wuchang District, Jiangan District, and 14 other administrative areas 2. Inner to Fourth Ring Roads and Cross-River Tunnels	2,626	4,830
2	2024-09-07 to 2024-09-13	Morning/Evening Peak: 7:00-9:30, 17:00-19:30 Midday: 12:00-14:00 Night: 0:00-2:00	1. Jiangnan District, Qingshan District, and 12 other administrative areas 2. Inner to Fourth Ring Roads and Industrial Area Road Networks	2,513	3,368

Table 1. Mobile monitoring data collection information

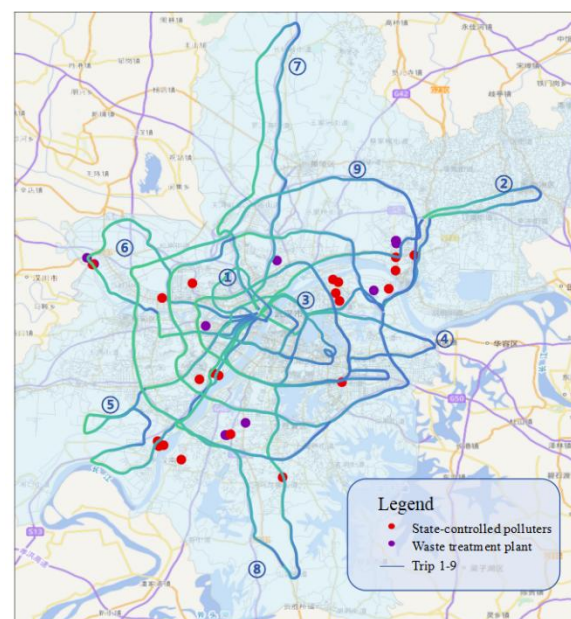


Figure 2. Mobile monitoring routes

2.2.2 Multi-Source data processing

During the actual data collection process, errors and inaccuracies inevitably occur due to sensor malfunctions, measurement deviations, or environmental interference. To meet the data quality requirements of the model, all collected data undergo rigorous preprocessing, including outlier removal, missing data imputation, temporal alignment, and spatialization. The specific steps are as follows:

(1) Outlier Detection: For each data type, the Z-score of each data point is calculated based on the dataset's mean (μ) and standard deviation (σ). Data points exceeding a predefined Z-score threshold (e.g., $|Z| > 3$) are flagged as outliers and removed.

(2) Missing Data Imputation: When vehicles pass through areas with GPS signal loss (e.g., tunnels or under bridges), linear interpolation or spline interpolation is applied using adjacent valid data points to estimate missing coordinates.

(3) Temporal Alignment and Spatialization: Greenhouse gas concentration data, meteorological parameters, and panoramic images are synchronized with GPS trajectories using timestamp matching, ensuring temporal consistency across multi-source datasets. Subsequently, geospatial coordinates are assigned to each data point, enabling spatial analysis and visualization.

To enhance road traffic state analysis, this study processes environmental images captured by panoramic cameras. Specifically, the camera simultaneously acquires images in five directions (front, rear, left, right, and upward) with 40%–60% overlap between adjacent frames. Using image stitching technology, these images are reconstructed into seamless 360° panoramas (figure 3). Furthermore, the YOLOv8 (You Only Look Once Version 8) deep learning model is employed to detect vehicle types (e.g., cars, trucks) and count traffic flow in real time during mobile monitoring (figure 4).



Figure 3. Panoramic environmental image stitching results



Figure 4. Vehicle information recognition results

2.2.3 Road carbon emission simulation model

Random Forest is an ensemble learning algorithm that improves prediction accuracy and robustness by constructing multiple decision trees and averaging their results. The Random Forest model has several advantages, such as handling high-dimensional data, capturing non-linear relationships, strong robustness, and preventing overfitting, making it suitable for simulating road traffic CO₂ concentration data. The input variables primarily consist of mobile monitoring CO₂ concentration data, which also serve as the target values for model training. The feature variables comprehensively consider factors related to the built environment, socio-economic activities, and transportation, with a total of 19 feature variables constructed. These are divided into four types: road characteristics, operational characteristics, meteorological characteristics, and environmental characteristics (table 2).

Types	Feature Vectors
Road Characteristics	Road length, Road class, Number of lanes, Road network density
Operational Characteristics	Congestion index, Traffic volume, Average speed
Meteorological Characteristics	Shortwave radiation intensity, 2-meter air temperature, 2-meter specific humidity, 10-meter meridional and zonal wind speed, Surface pressure, Hourly precipitation
Environmental Characteristics	Population density, Building density, Land use type, POI density, Environmental Air Quality Index (AQI)

Table 2. Road carbon emission feature factor table

Road Characteristics form the core component of constructing carbon emission feature vectors, providing a detailed description of the urban road network structure and attributes. These are sourced from Wuhan's territorial space monitoring road network data. Operational Characteristics reflect traffic operation features such as road flow, speed, and operating conditions, derived from Baidu Maps and statistical data provided by transportation management authorities. Meteorological Characteristics are key driving factors for variations in road CO₂ concentrations, with meteorological parameter data primarily sourced from the China Meteorological Administration's Land Data Assimilation System (CLDAS-V2.0), published by the National Meteorological Information Center. Environmental Characteristics capture the impact of regional natural resources and socio-economic activities on CO₂ concentrations, mainly sourced from population distribution grid data, land use data, and hourly Air Quality Index (AQI) data.

To evaluate the predictive capability of the model, the dataset is typically divided into two parts: the Training Set and the Validation Set. The training set is used to train the model, allowing it to enhance its predictive ability through learning. The validation set is used for model tuning and evaluation. During the training process, model parameters, structure adjustments, and overfitting control are optimized based on performance metrics from the validation set. After completing the training and adjustment phases, the validation set is used to assess the model's final performance, providing an accurate representation of its real-world capabilities.

Based on these requirements, the dataset is proposed to be split into an 8:2 ratio for the training and validation sets.

After completing the model training, the test set data is used to evaluate model performance by calculating evaluation metrics such as the coefficient of determination (R^2) and root mean square error (RMSE). The coefficient of determination (R^2) measures the model's ability to explain the variance in the data, with values ranging from 0 to 1. The closer R^2 is to 1, the better the model fits the data. The root mean square error (RMSE) is the square root of the average squared difference between predicted and actual values. A smaller RMSE indicates better predictive performance. Below are the formulas for calculating R^2 and RMSE:

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (p_i - \bar{p})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2} \quad (2)$$

where p_i = observed CO₂ concentration
 o_i = predicted CO₂ concentration
 $\bar{p}$ = the arithmetic mean of the observed CO₂ concentrations
 n = the number of matched data pairs

To evaluate the model's accuracy, this study assessed the model using the validation set data (figure 5). By comparing the model's simulated results with the measured data, it was found that the coefficient of determination (R^2) was 0.74, and the root mean square error (RMSE) was 26.22 ppm. The validation results indicate that the model has a good fitting effect and can simulate hourly carbon concentration variations on Wuhan's roads with high accuracy. This demonstrates the reliability and practicality of the model in predicting urban road carbon concentrations.

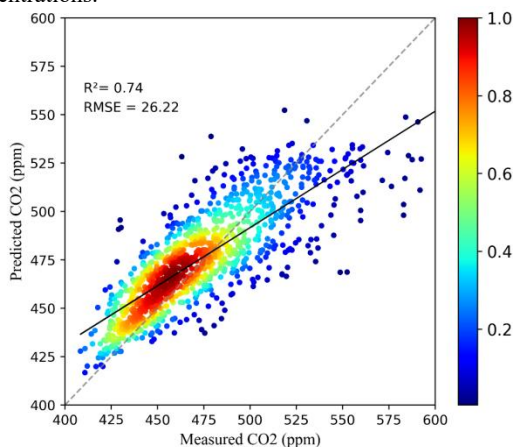


Figure 5. Model accuracy evaluation

3. Results

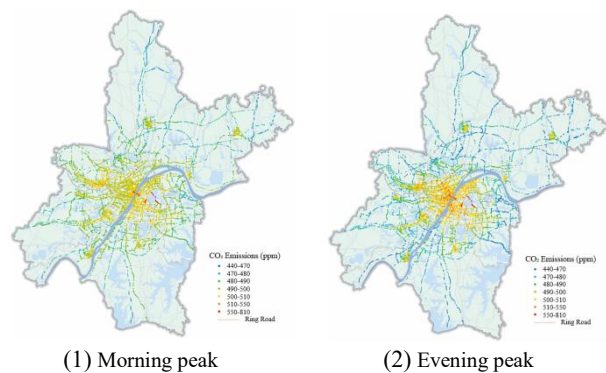
3.1 Carbon emission from a time-of-day perspective

Based on the simulation results, this study selected three representative periods: morning peak (7:00 – 9:30), evening peak (17:00 – 19:30), and nighttime off-peak (0:00 – 5:00)—to analyze the spatiotemporal patterns of road carbon emissions in Wuhan (figures 6). The results indicate that the tidal nature of traffic activities coupled with the polycentric urban spatial structure drives carbon emissions characterized by "peak aggregation, directional asymmetry, and arterial road

dependency", with the following specific patterns of differentiation:

Carbon emissions exhibit pronounced temporal variations. During the morning peak, the highest carbon emissions of the day occur, with an average CO₂ concentration of 492.46 ppm and a peak concentration reaching 804.18 ppm, which is 3.2% higher than during the nighttime off-peak period. This elevated emission level can be attributed to the tidal flow of inbound commuting from new urban areas (e.g., Jiangxia to Optics Valley), resulting in high saturation on arterial roads, compounded by energy consumption from early industrial shifts and high emissions from vehicle cold starts. During the evening peak, carbon emission intensity is slightly lower, with an average CO₂ concentration of 488.79 ppm and a peak of 788.27 ppm (2.4% higher than the nighttime off-peak). This is primarily driven by outward commuting flows (e.g., centrifugal traffic from peripheral new towns such as Houhu and Nanhu), exhaust emissions from commercial activities (exacerbated by congestion around commercial zones), and ongoing emissions from some late-shift industrial operations. However, due to staggered production schedules and traffic diversion policies (e.g., truck restrictions from 7:00 – 20:00 banning entry into the Third Ring Road), peak pressure is somewhat alleviated. During the nighttime off-peak, the average CO₂ concentration decreases to 477.31 ppm, but the maximum value still reaches 736.25 ppm. Despite a significant reduction in traffic volume and industrial activities operating at low capacity, baseline emission sources (e.g., ventilation systems in cross-river tunnels, 24-hour logistics transportation, and continuous operation of power plants) maintain concentrations at a medium level, revealing the hidden contribution of urban basic functions to nighttime carbon emissions.

Spatially, carbon emissions exhibit a "corridor polarization" phenomenon. Enclosed sections, such as the East Lake Tunnel, Shuiguo Lake Tunnel, and Yangtze River Tunnel, show significantly higher CO₂ concentrations (550 – 810 ppm) compared to other areas. Carbon emissions during the morning and evening peaks are concentrated on radial arterial roads connecting the central urban areas with peripheral new towns (e.g., Wuhan Avenue, Luoyu Road). Influenced by the "multi-center cluster" urban form, which extends commuting distances, these roads account for more than half of the city's total road emissions. This reflects the path dependency on cross-river traffic and arterial roads under the "job-housing separation" spatial structure, driven by long-distance commuting.



(1) Morning peak

(2) Evening peak

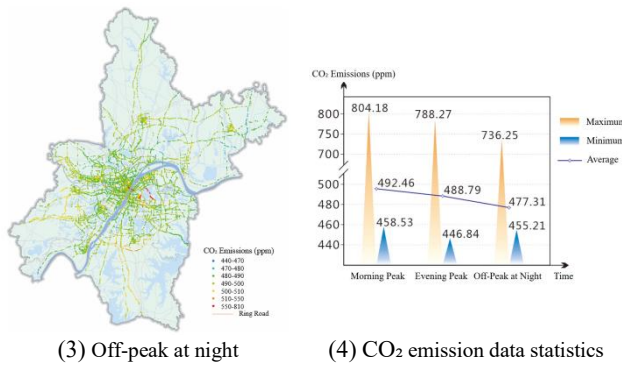


Figure 6. Distribution of CO₂ emissions in typical time periods

3.2 Carbon emission analysis based on ring roads

Wuhan's transportation network primarily consists of concentric ring roads and radial arteries. Monitoring results (figure 7) reveal significant spatiotemporal variations in CO₂ emission concentrations across different ring road zones, with fluctuation patterns strongly correlated to traffic intensity and urban spatial functions. The key findings are:

(1) The Inner Ring and Second Ring roads form the core emission zones, with morning peak concentrations reaching 500.99 ppm and 507.42 ppm respectively, further increasing to 508.36 ppm and 513.0 ppm during evening peaks. This confirms the centripetal commuting pattern dominated by job-housing separation in central urban areas.

(2) Network-wide concentrations decreased by 4.7%-5.3% from daytime peaks (Inner Ring: 488.57 ppm, Second Ring: 487.6 ppm, Third Ring: 489.89 ppm, Fourth Ring: 486.99 ppm), indicating lower urban carbon emission intensity.

(3) The Inner and Second Ring roads maintained 3.9%-4.5% higher daily mean concentrations than outer rings (Third/Fourth), demonstrating outward-decreasing emission intensity. The central urban area forms a stable high-emission pattern due to high-density urban functions and inelastic commuting demand.

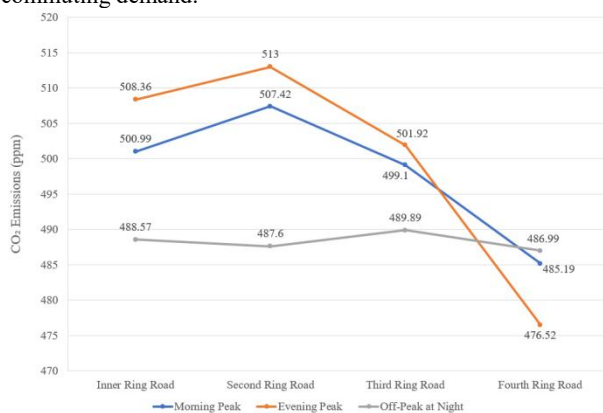


Figure 7. Average CO₂ emissions of ring roads

3.3 Carbon emission analysis based on road types

Analysis of CO₂ concentration data across different areas and road types in Wuhan (figure 8) reveals hotspot emissions in tunnels and a dependency on arterial roads. Specifically, the average CO₂ concentration in tunnel environments is 620 ppm, exceeding the average of other road types (481.33 ± 49.52 ppm)

by 28.8%. This is closely related to restricted exhaust dispersion in enclosed tunnel spaces, ventilation system efficiency, and significantly increased traffic density during peak hours. Expressways and ring roads (488.74 ppm), along with elevated bridges (480.62 ppm), exhibit the next highest concentrations. Notably, branch roads show the lowest concentration levels (475.35 ppm), likely due to the carbon sink effect of surrounding green belts and lower traffic volumes.

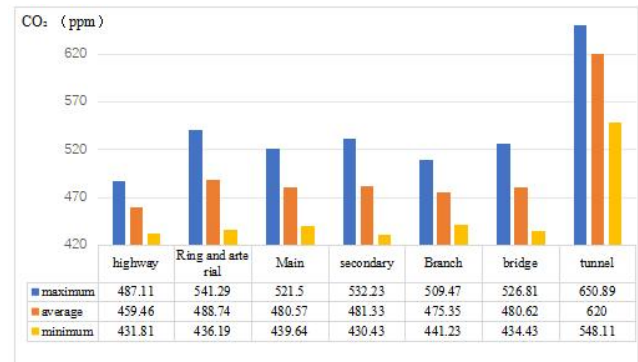


Figure 8. Carbon emissions from different road types

4. Discussion and Conclusions

This study takes Wuhan as a case study, integrating multi-source data such as road-based mobile monitoring, traffic flow, vehicle type proportions, and speed, to construct a spatiotemporal carbon emission model for roads using the Random Forest algorithm. The model achieves a coefficient of determination (R^2) of 0.74 and a root mean square error (RMSE) of 26.22 ppm, enabling high-accuracy simulation of hourly CO₂ concentration variations on Wuhan's roads. Based on this, the study analyzes the spatiotemporal characteristics of carbon emissions across different time periods, ring roads, and road types, revealing their distribution patterns and differences.

Temporal Differentiation Characteristics: Carbon emissions exhibit a "double-peak concentration" pattern. The average CO₂ concentrations during morning and evening peaks are 3.2% and 2.4% higher, respectively, than during the nighttime off-peak period. This is primarily driven by commuting tidal flows, industrial energy consumption, and high emissions from cold starts. Although the average concentration decreases during the nighttime off-peak, baseline emission sources (e.g., tunnel ventilation, logistics transportation) maintain medium-to-high fluctuations, highlighting the continuous impact of urban basic functions.

Spatial Polarization Effect: Tunnels and arterial roads form emission hotspots. Enclosed sections like the East Lake Tunnel show CO₂ concentrations ranging from 550 to 810 ppm. Radial arterial roads, main roads, and cross-river tunnels carry higher carbon emissions, reflecting the path dependency of cross-river traffic due to "job-housing separation" and the multi-center urban structure.

Ring Road Gradient Decline: The daily average CO₂ concentrations of the inner and second ring roads are 3.9% – 4.5% higher than those of outer ring roads. The central urban areas, with dense urban functions and rigid commuting demands (daily traffic volume 2.3 times higher than in outer areas), form a high-emission steady-state pattern.

This study innovatively integrates multi-source sensor data with machine learning models, overcoming the spatial

resolution limitations of traditional emission factor models. It establishes, for the first time, a dynamic road-level carbon emission simulation framework for a megacity. The findings indicate that traffic tidal patterns, spatial structural constraints, and urban functional timing are the three core drivers of carbon emission differentiation, providing scientific support for implementing differentiated control measures in Wuhan. Therefore, future carbon reduction strategies for Wuhan's transportation should consider regional, road grade, and temporal differences, promoting inter-regional collaborative emission reduction. These strategies should integrate road network planning, traffic organization, public transport prioritization, non-motorized system development, and vehicle technology upgrades to achieve coordinated control of citywide transportation carbon emissions. This will drive the low-carbon transformation and sustainable development of Wuhan's transportation system.

References

- UNFCCC, 2016: The Paris Agreement. United Nations Framework Convention on Climate Change.
- Ministry of Ecology and Environment of the People's Republic of China, 2022: Progress report on China's implementation of nationally determined contributions.
- He, J., Li, Z., Zhang, X., et al., 2022: Towards carbon neutrality: A study on China's long-term low-carbon transition pathways and strategies. *Environmental Science & Ecotechnology* 1, 2-10.
- Zhang, X. D., Wu, N. N., Gao, W. C., Du, J. K., 2024: Research on the low-emission zone policy for motor vehicles in Wuhan City. In *Green Intelligence and Efficiency Improvement: Proceedings of the 2024 Annual Conference on Urban Transportation Planning in China*, 2207-2223.
- Zhang, L. P., Huang, P., Zhu, J. Q., et al., 2023: Research progress on carbon emissions in the transportation industry. *Energy Research and Management* 15(04), 76-84.
- Yang, H., 2023: Road-Level NH₃ Emissions from On-Road Vehicles in Shanghai: A Mobile Observation Study. *Environmental Pollution* 324, 121397.
- Li, F., Zhu, Z., Wu, X., 2024: Construction of mobile monitoring network for vehicles in Shanghai and its application in air quality management. *Environmental Monitoring in China* 40(S1), 23-29.
- Yifan, W., Ruoxi, W., Zihang, Z., et al., 2022: A data-driven method of traffic emissions mapping with land use random forest models. *Applied Energy* 305.
- Chaerin, P., Sujong, J., Chongmin, K., et al., 2023: Machine learning based estimation of urban on-road CO₂ concentration in Seoul. *Environmental Research* 231(P3), 116256-116256.
- Wuhan Municipal Institute of Planning and Research (Wuhan Transportation Development Strategy Research Institute). 2024. “Wuhan Transportation Development Annual Report.”