Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Identification of carbon emission mechanism typologies: A Data-Driven approach integrating GWR and SOM-FCM clustering

Yang Sun [*,a]

Manchester School of Architecture, The University of Manchester, Manchester, M1 7ED, England

Article Info

Article History:

Received 21 July 2026

Accepted 10 Sep 2026

Keywords:

CO2 emissions;
Built environment;
Urban morphology;
Geographically Weighted Regression (GWR);
SOM-FCM clustering;
Machine learning;
Spatial heterogeneity;
Precision spatial governance;
Guangzhou;
China

Abstract

Spatial optimization of the built environment has become an important strategy to achieve carbon neutrality. However, most existing models ignore the spatial non-stationarity of driving forces. Therefore, this study combines GWR with an integrated SOM-FCM method for unraveling the local variations of built environment drivers of carbon emissions in downtown Guangzhou. The findings show that: (1) The geographical patterns demonstrate an obvious "south-west high and north-east low" emission trend with a clear core-periphery pattern in which intense emissions tend to be significantly concentrated in the downtown historical area. (2) Influencing factors have strong spatial variability; especially, road density and functional intensity aggravate emissions in the mature core, whereas mass transit and green space act as effective mitigating factors in the high-density southern areas. (3) The SOM-FCM framework synthesized these complex patterns into four distinct "carbon emission mechanism zones", namely Vitality-Traffic Driven, Morphology Dominant, Location-Sensitive, and Industrial-Logistics, each characterized by unique profiles of built environment determinants. Ultimately, this research contributes a data-driven methodology for transitioning from uniform density regulations to zoning-based precision spatial governance in high-density metropolises.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

Global climate change is now an issue that must be addressed to ensure continued human progress. The International Energy Agency reports that global energy-related carbon dioxide (CO_2) emissions increased to a record level of 37.8 Gt in 2024 (International Energy Agency, 2025), which leads to the enhancement of the greenhouse effect that causes climate change as well as environmental pollution with subsequent impacts on public health and ecosystem security [1]. Cities are disproportionately responsible for this disaster; despite covering less than 3% of Earth's surface area, cities generate about 71% of anthropogenic carbon emissions [2]. As the largest developing country in the world, China has committed to achieving carbon neutrality by 2060 [3]. The assumption underlying this pledge is that the physical environment has a substantial effect on energy use. These dynamics put tremendous pressure on fast-developing urban areas to decarbonize economic activities through the transformation of their physical environment [1,2,4]. Therefore, it is an important task to improve the built environment by using relevant information to explore emission patterns and to conduct targeted spatial management for controlling energy use.

Urban carbon emissions are not randomly distributed but exhibit significant spatial dependence and clustering characteristics [5]. Inspired by spatial econometrics, studies of the relationship between built environment and carbon emission have developed in four directions: First, with

*Corresponding author: sunyang602@outlook.com
DOI: http://dx.doi.org/10.17515/resm2026-1963vk0721rs

regard to the scale of studies, there has been a clear shift towards finer-scale analyses [3]. Although many previous works use low-resolution input, they are prone to losing the detailed information about the environment. Thus, fine-grained units such as street blocks or high-resolution grids are also widely used in order to accurately locate a particular emission hotspot or characterize the detailed spatial heterogeneity [4,6]. Secondly, with respect to the built environment indicator construction, research has extended from simplistic density measures into multi-dimensional indexes. There is a growing consensus that both two-dimensional metrics and three-dimensional factors, such as building height and local climate zones, are important drivers of carbon emission [7-9]. In addition, the entire urban spatial pattern is also significant as some studies indicate that polycentric structures could be better at reducing emissions compared to monocentric ones [1,10]. Third, in terms of methodology, there has been increased focus towards spatial non-stationarity [11]. Impact of built environment on emission is hardly homogeneous through space. However, traditional global regressions such as OLS assume spatial homogeneity and usually lead to biased estimation in a heterogeneous urban context [4,12,13]. To remedy this problem, local regression models such as Geographically Weighted Regression (GWR) are required [14], since they can reflect that different built environment factors exert varying impacts in different areas, indicating that drivers may act in completely opposite directions or with varying magnitudes between different urban growth areas [5,9]. Fourth, in addition to detecting space-variant conditions, recent literature stresses that it is important to aggregate high-dimensional local models into interpretable regions using mixed-effects clustering techniques, in particular, SOM combined with FCM, which has proved effective for preserving topological properties and defining functional areas even in complex scenarios [15,16]. Therefore, this type of data-driven zoning logic provides a transferable theory foundation to identify the distinct zones with different carbon emission mechanisms under high population density scenarios.

However, with all this methodological advancement there are still some important knowledge gaps to address. First, in terms of the scale at which studies were conducted, most research is limited either in extent and/or resolution by administrative units [3], which may introduce the well-known Modifiable Areal Unit Problem (MAUP), leading to a loss of local variation that characterizes urban areas [4]. Second, there is inconsistent evidence on important built environment determinants such as urban density. Such mixed results indicate that the underlying spatial heterogeneity in these mechanisms remains unexplained [17,18]. Third, there is a gap between the identification of mechanisms and actual spatial management. However, few studies aggregate such high-dimensional local fits to meaningful planning units and hence translate spatially varying non-stationarity into actionable, zoning-based approaches [8,9]. Thus, they are unable to capture how various aspects of the built environment differentially drive emission changes at a local level. Hence, empirical work that disentangles these spatially heterogeneous effects on the fine grid scale is desperately needed.

To fill those gaps, this work proposes to build up a data-driven framework by combining GWR with SOM-FCM clustering for decoding the spatial non-stationarity of carbon emission mechanisms in the central urban area of Guangzhou quantitatively. Integrating multi-source open data (e.g., OpenStreetMap road networks, Baidu Map building footprints, and POI data), this study builds a comprehensive indicator set that covers the aspects of road density, building coverage, and land use mix at the level of small (1 km × 1 km) grid cells. In particular, this study addresses three key issues: (1) How is the anthropogenic CO_2 emission distributed within the dense urban core? (2) How do multidimensional built environment indicators drive emissions, and do these effects exhibit spatial heterogeneity at the grid level? (3) How can high-dimensional local estimates be aggregated via the proposed hybrid GWR and SOM-FCM approach to obtain “Carbon Emission Mechanism Typologies”? Finally, by interpreting the intricate spatial heterogeneity of carbon drivers, this study provides theoretical support to low-carbon urban revitalization and an empirical basis to shift from flat control to zoning-based precision spatial governance in mega-cities.

2. Data and Methods

2.1 Study Area

The study area is the central urban area of Guangzhou, the capital of Guangdong Province and a core city of the Guangdong-Hong Kong-Macao Greater Bay Area. It covers the contiguous built-up core of the metropolis, extending from about 113.14°E to 113.61°E and from 23.05°N to 23.43°N. The southwestern part corresponds to the historical urban core (Liwan, Yuexiu, and western Haizhu), characterized by dense road networks, compact building fabric, and highly mixed functions, whereas the northeastern part contains newly developed frontiers and extensive ecological space such as the forested hills on the north side of Tianhe. This pronounced internal contrast between a mature high-density core and a developing periphery makes the area a representative case for examining the spatial non-stationarity of carbon emission mechanisms in high-density Asian metropolises. For the analysis, the study area is partitioned into 1,476 grid cells of 1 km × 1 km, covering approximately 1,476 km^2 and serving as the basic spatial units of all subsequent statistical modeling. The built fabric within this extent is highly uneven, with building coverage averaging 0.17 and reaching almost full coverage in the densest cells, road density averaging 5.7 km/km^2 with a maximum of 18.8 km/km^2, and the distance to the CBD spanning from 0.3 km to 37.5 km, which quantifies the core-periphery gradient described above.

2.2 Data Sources

This study integrates multiple geospatial datasets to build the analysis database. The high-resolution anthropogenic carbon emission dataset comes from the ODIAC database (2024 release), from which the 2020 annual emission layer used in this study was extracted; the database gives a global grid-based estimate of fossil fuel burning. This study used three main data sets to describe the physical form of the city, namely OSM vector roads, Baidu Map for the building footprint vector with detailed morphological information, and Gaode Map for POIs as a representation of the functional density. In addition, urban land use classes are extracted from the EULUC-China 2.0 dataset which employs the fusion of many sources and modalities of geospatial data with deep learning approaches for improved land use classification [13]. To attain a coherent representation across the heterogeneous sources, all data sets were projected to the WGS 1984 UTM Zone 49N coordinate system and then aggregated spatially into 1 km × 1 km grid units, which serve as the basic units of the subsequent OLS and GWR analysis.

Table 1. Definitions, calculation, and units of the variables (variables with |skewness| > 1 were log-transformed as log(1+x), and all variables were z-standardized before regression)

Variable	Description and calculation	Unit
Carbon emissions (dependent)	ODIAC fossil-fuel CO_2 emissions aggregated to each grid cell	t/year
Road Density	Total road length (OpenStreetMap) per unit grid area	km/km^2
Transit Density	Number of public transit stops within the grid cell	$count/km^2$
Building Coverage	Share of grid area covered by building footprints (Baidu Map)	ratio (0-1)
POI Density	Number of points of interest (Gaode Map) within the grid cell	$count/km^2$
Industrial / Residential / Commercial / Green Space	Share of grid area in the corresponding EULUC-China 2.0 land use class	ratio (0-1)
Distance to CBD / Distance to River	Euclidean distance from the grid centroid to the CBD / nearest major river	km

2.3 Research Methods

2.3.1 Spatial Autocorrelation Analysis

In order to find out whether CO_2 emission of the central urban area of Guangzhou has a spatial dependency relationship, this study uses the Global Moran's I statistic, which measures the level of spatial autocorrelation of carbon emissions over the whole region. This index can determine if the distribution is clustered, dispersed, or random [14]. The calculation equation is shown below:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

Where I represents the Global Moran's I index, ranging from -1 to 1. A value closer to 1 indicates a strong positive spatial correlation (high-high or low-low clustering) of CO_2 emissions; a value closer to -1 indicates a negative correlation (dispersed); and a value near 0 implies a random distribution; n is the total number of grid units (1 km × 1 km) in the study area; x_i and x_j represent the anthropogenic CO_2 emission values of grid unit i and grid unit j, respectively.$\bar{x}$ is the average CO_2 emission value of all grid units; $w_{i,j}$ is the spatial weight between grid i and grid j, representing the spatial proximity relationship; S_0 is the aggregate of all spatial weights, calculated as

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{2}$$

To test whether the spatial autocorrelation was significant we calculate the z score:

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}} \tag{3}$$

Where $E[I] = -1/(n-1)$ is the expected value, and $V[I]$ is the variance. A statistically significant positive Z-score ($p < 0.05$) confirms the spatial agglomeration of carbon emissions.

2.3.2 High/Low Clustering

Although Global Moran's I detects that there exists a clustering phenomenon in space but cannot distinguish whether this cluster consists of high values or low values. Therefore, the present paper uses the High/Low Clustering (Getis-Ord General G) statistic [14]. It is a global statistic that quantifies the degree to which high or low values for CO_2 emissions are clustered within the whole study area, and it is calculated by:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j}, \forall j \neq i \tag{4}$$

Where x_i and x_j are the CO_2 emission values for grid unit i and grid unit j; $w_{i,j}$ is the spatial weight between grid i and grid j; n is the total number of grid units in the dataset; ∀ j ≠ i indicates that features i and j cannot be the same feature.

To evaluate the statistical significance of the clustering pattern, the z-score for the General G statistic is calculated as follows:

$$Z_G = \frac{G - E[G]}{\sqrt{V[G]}} \tag{5}$$

Where the expected value $E[G]$ and variance $V[G]$ are calculated as:

$$E[G] = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \forall j \neq i \tag{6}$$

$$V[G] = E[G^2] - E[G]^2 \tag{7}$$

The interpretation of the General G statistic depends on the Z-score (Z_G): High-Value Clustering (High Z_G): If Z_G is positive and statistically significant, it indicates that the spatial pattern of CO_2 emissions is dominated by clusters of high values (i.e., emission hotspots are more intense). Low-Value Clustering (Low Z_G): If Z_G is negative and statistically significant, it indicates that the spatial pattern is dominated by clusters of low values. Random Distribution: If Z_G is close to zero, there is no apparent high or low clustering.

2.3.3 Regression Analysis

To quantitatively reveal the driving mechanisms of multidimensional built environment indicators on CO_2 emissions, this study constructs both global and local regression models.

(1) Multicollinearity Diagnosis

Before building up the regressions, it is necessary to check if there are any dependencies between the independent variables (built environment indicators), i.e., multicollinearity. VIF was used for detecting possible multicollinearity. Generally, a VIF value less than 10 implies no serious multicollinearity. Indicators that meet such requirements in this paper are taken as the final independent variables (x) for the regression analysis.

(2) Ordinary Least Squares (OLS)

First, the OLS model was employed to estimate the global average impact of built environment factors on carbon emissions. The OLS model assumes that the relationship between built environment and CO_2 emission is spatially constant (homogeneous), i.e., it does not vary over the study space [12]. The general equation can be written as:

$$y = \beta_0 + \sum_{k=1}^{p} \beta_k x_k + \varepsilon \tag{8}$$

In this equation, y represents the dependent variable, which is the CO_2 emissions of the grid unit; β_0 is the intercept of the regression model; x_k represents the value of the k-th built environment indicator (e.g., road density, building coverage, POI density); β_k is the regression coefficient for the explanatory variable k, reflecting the global marginal effect of that built environment factor on CO_2 emissions; ε is the random error term.

(3) Geographically Weighted Regression (GWR)

Because OLS ignores spatial variation, it cannot capture localized effect in a dense downtown area. To handle spatial non-stationarity, this paper uses the GWR model. In comparison with OLS, the GWR allows for spatial heterogeneity of the regression coefficients and assumes that nearby observations exert more influence in estimating parameters than distant ones [14]. The formula can be expressed below:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{9}$$

In this equation, y_i is the CO_2 emission value at grid unit i; (u_i, v_i) denotes the geographic coordinates (centroid) of grid unit i; $\beta_0(u_i, v_i)$ is the localized intercept at location i; $\beta_k(u_i, v_i)$ represents the local regression coefficient of the k-th built environment indicator at grid i. This coefficient varies across the grid, visualizing the spatial heterogeneity of how the built environment drives carbon emissions; ε_i is the random error term at location i; The model uses a fixed Gaussian kernel to calculate weights, a choice suited to the regular 1 km × 1 km lattice of analysis units, with the optimal bandwidth determined by the corrected Akaike Information Criterion (AICc); the AICc-optimized bandwidth is approximately 2.9 km. The spatial randomness of the GWR residuals was further examined as a specification check. The Global Moran's I of the residuals is 0.059 ($z = 4.57$, $p < 0.001$), an order of magnitude lower than that of the dependent variable (0.216), indicating that the local model absorbs the bulk of the spatial dependence in the data. The remaining autocorrelation is weak but still statistically significant, which suggests that a small part of the local variation is driven by factors outside the present indicator set.

2.3.4 Spatial Zoning

To extract spatially contiguous areas with similar driving mechanisms out of the high dimensional GWR coefficients, we applied a SOM-based FCM clustering algorithm as follows. First, SOM was used for projecting the multi-dimensional coefficient vector into a topology preserving two-dimensional feature map. SOM is an unsupervised neural network that can reduce the dimensionality of, and the noise in, the data while retaining nonlinear spatial relationships due to its competitive learning mechanism. Then, the FCM algorithm was applied to the SOM weight vectors to deal with the continuous gradients that are typical of urban spatial phenomena [15],[16]. In contrast to hard clustering, FCM assigns membership degrees (u_{ij}) to grid units, minimizing the objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2 \tag{10}$$

where N is the sample size, C is the number of clusters, and $m = 2.0$ is the fuzziness exponent. Finally, the optimal number of clusters ($k = 4$) was rigorously determined via sensitivity analysis (testing k from 2 to 8). As summarized in Table 2, this selection provides the best overall trade-off among the Fuzzy Partition Coefficient (FPC), the Silhouette Score, and the interpretability and spatial contiguity of the resulting zones; smaller k merges functionally distinct mechanism types, whereas $k \geq 5$ destabilizes the partition without adding interpretable structure. In the implementation, the SOM network was configured with a 13×13 grid and trained for 5,000 iterations. The model yielded a Quantization Error of 0.4805 and a Topographic Error of 0.229, confirming robust topological preservation. The final zoning solution ($k = 4$) exhibited a stable FPC of 0.6021, validating the reliability of the partition results.

Table 2. Sensitivity analysis of the number of clusters k (FCM applied to the SOM output; values averaged over five random initializations; the final archived solution at k = 4 has FPC = 0.6021)

Number of clusters k	FPC	Silhouette Score
2	0.709	0.035
3	0.639	0.059
4	0.611	0.042
5	0.567	0.016
6	0.539	0.029
7	0.514	0.006
8	0.509	0.042

3. Results

3.1 Spatial Distribution Characteristics of CO_2

As seen from Fig. 1, the spatial distribution of anthropogenic CO_2 emissions in the central urban region of Guangzhou has considerable spatial difference and clustering features according to the 1 km × 1 km grid size. In general, it can be said that there is a very clear spatial gradient of emission intensity, which decreases from southwest to northeast. High-value zones (red and dark orange grids, corresponding to the upper classes of the Figure 1 legend of up to about 3.8 million tons per year) are mainly concentrated in the western and southern parts of the study area, forming a continuous and dense high-emission cluster corresponding to the spatially developed urban core (including Liwan, Yuexiu, and western Haizhu districts). On the contrary, the low-value zones (yellow grids) are mostly located in the northern and northeastern outskirts such as the northern part of Tianhe District, where the emission intensity is much lower than the average of the whole region. This strong core-periphery pattern implies that carbon emissions are not evenly spread throughout the city.

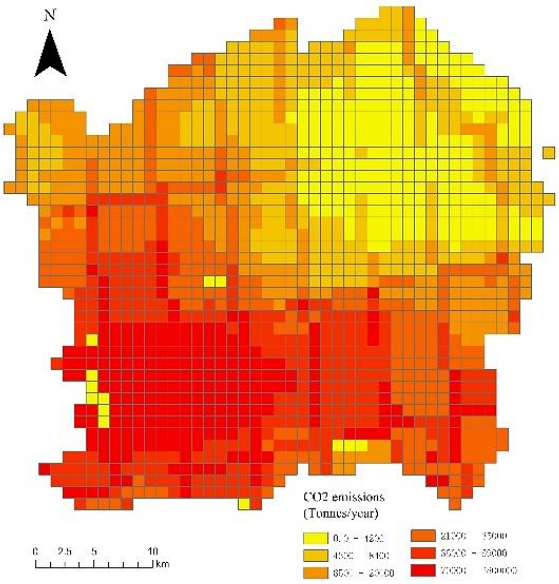


Fig. 1. The spatial distribution of anthropogenic CO_2 emissions in central Guangzhou (Unit: Tonnes/year)

Additional spatial autocorrelation analyses reveal that the emission is not random and there are clusters of emissions at a statistically significant level. As shown in Figure 2a, the Global Moran's I statistic value of human-caused CO_2 emission data is 0.216, with a z-score of 21.68 and $p < 0.001$, which shows a significant positive spatial autocorrelation relationship. To identify particular agglomeration modes, the map of Local Indicators of Spatial Association (LISA) shows a clear local spatial regime (Figure 2b). The high-high cluster (pink grids) is clearly located in the southwest area, including the old core districts of Liwan and Yuexiu, which makes up a continuous dense high-emission area. On the other hand, the low-low groups (light-blue grids) occupy the northeast part of the study area, which corresponds to the underdeveloped regions and ecological protection areas on the north side of Tianhe. Furthermore, scattered low-high outliers (dark blue) and high-low outliers (red) are intermingled on the transition edges of both main groups, capturing the local spatial variability of carbon intensity.

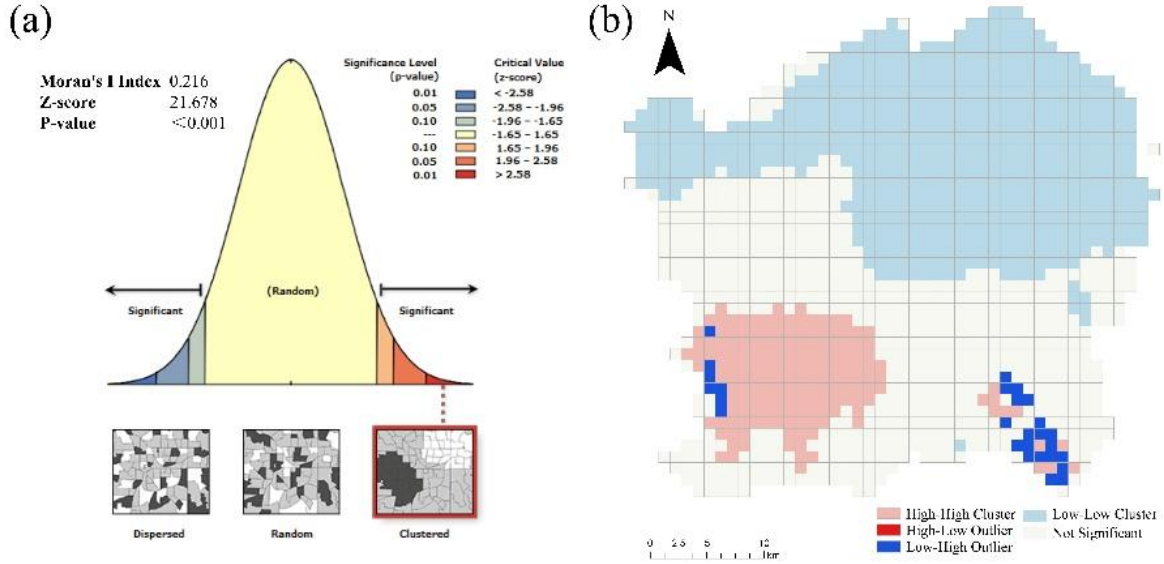


Fig. 2. Spatial autocorrelation analysis of carbon emissions. (a) Global Moran's I statistic and its significance test report based on the z-score distribution. (b) Spatial distribution of local indicators of spatial association clusters

3.2 Driving Effects of Built Environment on CO_2 Emissions

3.2.1 Global Regression Results (OLS)

Table 3 presents the results of the global Ordinary Least Squares (OLS) regression analysis. The model yielded an R^2 of 0.459, indicating that the selected multidimensional built environment indicators explain approximately 45.9% of the variance in spatial carbon emissions. Regarding specific drivers, Road Density ($\beta = 0.137, p < 0.001$) and Building Coverage ($\beta = 0.090, p < 0.05$) exhibited significant positive correlations with carbon emission intensity. Similarly, land use structure indicators, including Industrial ($\beta = 0.114, p < 0.001$), Residential ($\beta = 0.108, p < 0.001$), Commercial ($\beta = 0.052, p < 0.05$), and Green Space ($\beta = 0.135, p < 0.001$), all showed statistically significant positive associations. In terms of locational factors, both Distance to CBD ($\beta = -0.246, p < 0.001$) and Distance to River ($\beta = -0.185, p < 0.001$) demonstrated significant negative coefficients, suggesting that emission intensity tends to decrease as the distance from the city center and water bodies increases. Conversely, Transit Density ($\beta = -0.048, p = 0.177$) and POI Density ($\beta = 0.090, p = 0.075$) did not reach statistical significance at the 0.05 level in the global model.

Table 3. Results of the Global Ordinary Least Squares (OLS) regression

Variable	Coef.	Std. Error	t-Statistic	p-value
Intercept	-0.000	0.019	-0.000	1.000
Road Density	0.137	0.032	4.327	0.000
Transit Density	-0.048	0.035	-1.350	0.177
Building Cov.	0.090	0.037	2.470	0.014
POI Density	0.090	0.051	1.780	0.075
Industrial	0.114	0.024	4.831	0.000
Residential	0.108	0.027	4.048	0.000
Commercial	0.052	0.022	2.344	0.019
Green Space	0.135	0.023	5.778	0.000
Dist to CBD	-0.246	0.040	-6.123	0.000
Dist to River	-0.185	0.037	-4.971	0.000
R2	0.459	-	-	-
AICc	3305.232	-	-	-

3.2.2 Geographically Weighted Regression (GWR) Results

Compared with the global OLS model, the GWR model has better explanatory power and goodness-of-fit; the AICc decreases from 3305.232 to 3128.049, and the R2 increases from 0.459 to 0.632. It means that the GWR model is able to consider the spatial non-stationarity of carbon emissions. The summary statistics of the parameter estimates are reported in Tables 4 and 5. Table 4 summarizes the distribution of the local coefficients across all 1,476 grid units, whereas Table 5 reports, for each variable, the mean of the local coefficients computed only over the grid units where the estimate is statistically significant ($|t| > 1.645$), together with the shares of significantly positive and negative grids. The local coefficients of every variable vary considerably among grids, which confirms strong spatial non-stationarity. More specifically, Road Density (mean significant local coefficient of 0.388, positive and significant in 12.7% of grids) and POI Density (0.259, positive and significant in 11.2% of grids) are major driving factors, showing a positive relationship with carbon emission in most of the grids where they are significant. On the contrary, Transit Density (mean significant local coefficient of -0.148) and Green Space (-0.127) mainly play an inhibiting role. Distance to CBD has the strongest negative decay effect (mean significant local coefficient of -1.039), and 32.2% of the grid units show a statistically significant negative correlation. In particular, Distance to River shows a complicated mixed influence (significantly positive in 20.4% and significantly negative in 25.3% of grids), whereas Building Coverage, Residential, and Commercial land proportions are significant in only a small share of grids, indicating that their local impact is spatially limited.

Table 4. Summary statistics of GWR local parameter estimates over all 1,476 grid units

Variable	Mean	STD	Min	Median	Max
Intercept	-0.312	1.425	-4.017	-0.509	3.134
Road Density	0.072	0.145	-0.146	0.038	0.699
Transit Density	-0.021	0.127	-0.389	-0.005	0.441
Building Cov.	0.357	1.230	-0.601	0.057	6.861
POI Density	0.064	0.202	-0.528	0.041	0.700
Industrial	0.003	0.218	-0.376	0.005	1.448
Residential	-0.019	0.166	-0.457	-0.001	0.845
Commercial	-0.017	0.192	-0.202	-0.126	0.384
Green Space	-0.047	0.163	-0.417	-0.017	0.771
Dist to CBD	-0.466	1.192	-6.173	-0.294	3.495
Dist to River	0.016	1.437	-3.694	-0.153	6.752
R2	0.632 (OLS: 0.459)				
AICc	3128.049 (OLS: 3305.232)				

Table 5. Significance patterns of GWR local estimates (means computed over statistically significant grid units, |t| > 1.645)

Variable	Mean Sig. Coef.	Positive Sig. (%)	Negative Sig. (%)	Dominant Impact
Intercept	-0.312	21.May	26.Oca	Mixed
Road Density	0.388	12.Tem	0.0	Promoting (+)
Transit Density	-0.148	3.Nis	12.Mar	Inhibiting (-)
Building Cov.	0.048	4.Mar	3.0	Insignificant
POI Density	0.259	11.Şub	3.Haz	Promoting (+)
Industrial	0.228	6.0	5.Oca	Promoting (+)
Residential	0.074	3.0	3.Haz	Insignificant
Commercial	-0.017	1.Ağu	3.Tem	Insignificant
Green Space	-0.127	2.Tem	8.May	Inhibiting (-)
Dist to CBD	-1.039	8.Tem	32.2	Decay (-)
Dist to River	0.141	20.Nis	25.Mar	Mixed

3.2.3 Spatial Distribution of Driving Effects

The visualization of GWR local coefficients (Figure 3) further reveals significant spatial heterogeneity in the driving mechanisms of built environment on carbon emissions. The promoting factors, i.e., Road Density (Fig. 3c) and POI Density (Fig. 3f), also show an obvious "Southwest-High, Northeast-Low" spatial pattern, with their strongest positive coefficients represented by the densest deep red clusters located in the southwestern districts (Liwan and Yuexiu), suggesting that roads and functionally concentrated agglomerations within the more developed city center are under the greatest pressure to grow their emissions. Similarly, Industrial land use (Fig. 3g) shows a localized high-intensity promoting effect in the southwestern riverside areas. Conversely, the inhibitory factors demonstrate different spatial regimes; Public Transport (Fig. 3a) shows large negative coefficients (blue clusters) mainly concentrated in the southwest and the south, suggesting that the emission reduction potentials are higher for these high-density areas. Regarding locational variables, Distance to CBD (Fig. 3d) shows an evident negative gradient on the central northern region validating the distance-decay effect, while Distance to river (Fig. 3e) presents a mixed complex pattern of intermingled positive and negative effects within the northern part. Moreover, there is a strong spatial agreement among the driver factors Residential land (Fig. 3h), Commercial land (Fig. 3i), and Building Coverage (Fig. 3j). These three morphological indicators have an obvious "West-High, East-Low" spatial gradient.

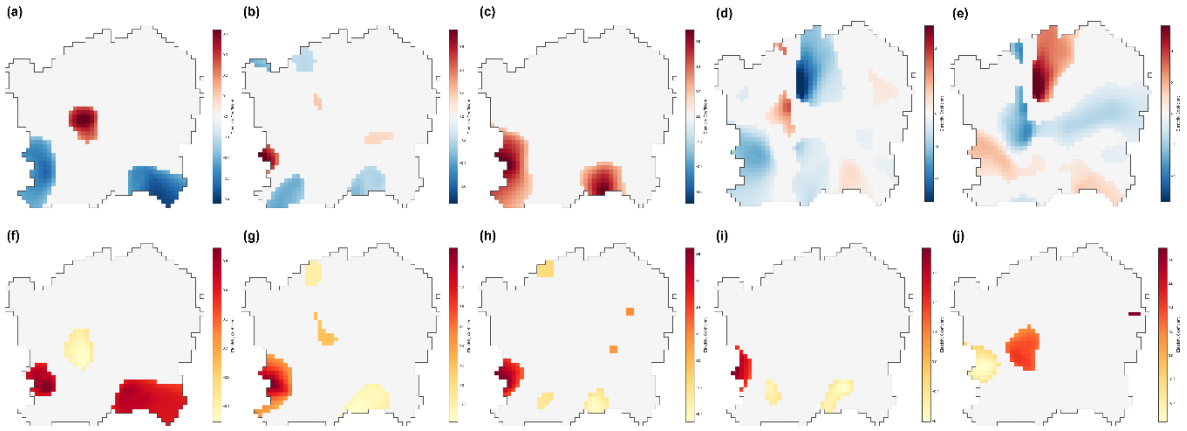


Fig. 3. Spatial distribution of local regression coefficients for built environment indicators based on the GWR model. The maps illustrate the spatial heterogeneity of driving effects for ten key variables. (a) Public Transit Density; (b) Green Space Ratio; (c) Road Network Density; (d) Distance to CBD; (e) Distance to River; (f) POI Density; (g) Industrial Land Prop.; (h) Residential Land Prop.; (i) Commercial Land Prop.; (j) Building Coverage

3.3 Spatial Zoning of Carbon Emission Driving Mechanisms

3.3.1 Spatial Distribution Characteristics of Mechanism Zones

Using the SOM-FCM model, this study integrated the complicated multi-dimensional regression coefficients to obtain four contiguous mechanism regions in space (Fig. 4). As shown, the study area has an obvious "mosaic" pattern instead of a simple concentric gradient, which proves the driving forces of carbon emissions have significant heterogeneity in this high-density central urban area. Zone 1 is mainly distributed in the western and southwestern areas of Guangzhou, which include the old city area (e.g., Liwan and the western part of Haizhu); Zone 2 dominates the eastern area, covering the growing development frontier (e.g., portions of Tianhe and Huangpu); Zone 3 concentrates largely on the central-north strip, while extending partly to the south-central strip. Zone 4 is represented by interstitial patches between the center and the edge, especially at the northeastern and central-western borders, indicating a mix of driving patterns typical of mixed-use areas.

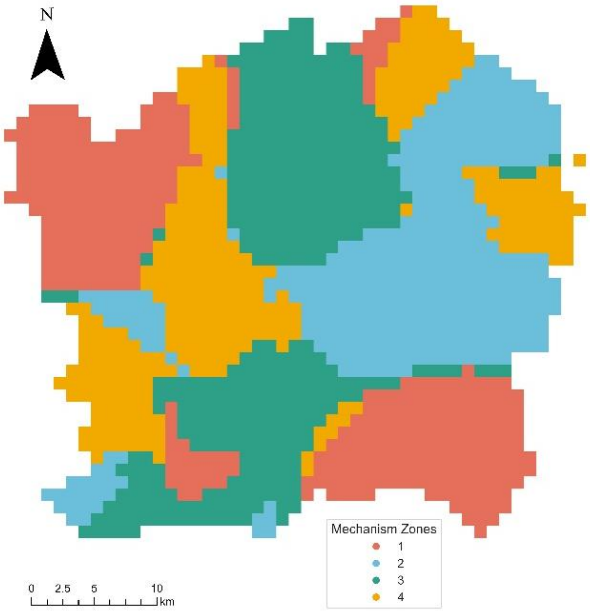


Fig. 4. Spatial distribution of identified carbon emission mechanism zones

3.3.2 Differentiated Driving Mechanisms Across Zones

The relative importance of driving factors per zone is shown in Fig. 5, indicating different mechanism typologies. The values quoted below are zone-averaged local GWR coefficients, whereas the bars in Fig. 5 display within-zone standardized relative effects rescaled for cross-variable comparison. Zone 1 (Vitality-Traffic Driven), which lies in the mature core area, is characterized by the strongest influence of POI Density (0.201) and Road Density (0.107), while Green Space has a large negative coefficient (-0.219), indicating that the lack of carbon sinks aggravates emissions. Meanwhile, Zone 2 (Morphology Dominant) covers the eastward expansion areas, where Building Coverage (1.074) dominates with a much larger weight than all the other factors, suggesting that emissions there are mainly determined by the physical amount of construction. Zone 3 (Location-Sensitive), which spreads from the north to the center, is highly sensitive to macro-location factors, with the highest positive weight for Distance to River (1.198) and a large negative weight for Distance to CBD (-1.480). Lastly, Zone 4 (Industrial-Logistics) is dispersed in outlying interstitial areas and is captured by the greatest positive Industrial land use effect (0.165) along with a significant Road Density influence (0.112), separate from the business logic of the hub region.

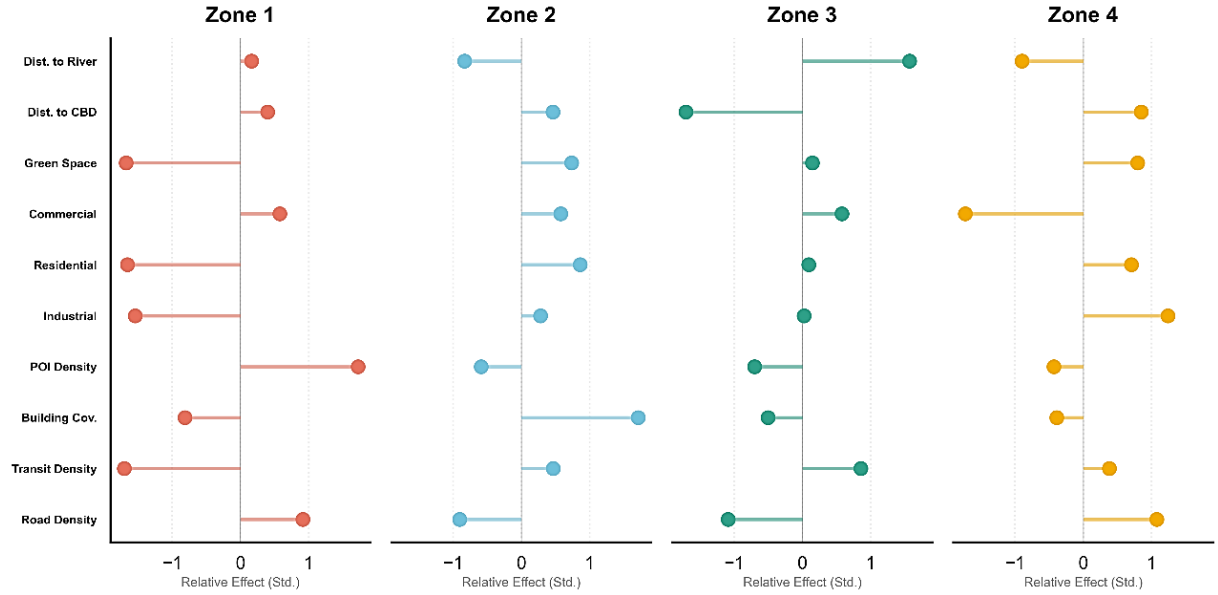


Fig. 5. The distinct profiles of carbon emission driving mechanisms across the four identified zones (bars show standardized relative effects; the corresponding zone-averaged local coefficients are reported in the text)

4. Discussion

4.1 Spatial Heterogeneity

With the goal for global emission reductions moving toward a net zero target, changing the spatial form of the urban fabric is becoming one approach towards reducing building energy demands. However, global regression models are common in the literature of conventional planning approaches; this implies an implicit “one-size-fits-all” assumption about the relation of built environment to carbon emissions, which ignores the complicated local realities in fast-growing areas. Using a fine-grained GWR model, this study further dissects such spatially varying driving forces of multidimensional built environments (e.g., road network and land use intensity) on CO_2 emission, which shows that the parameters at each location are much more predictive than a globally averaged set of parameters. Specifically, in the core area of Guangzhou, an archetypal mega-city with dense population and typical characteristics of Asian cities, this study finds strong regional differences across the driving forces for CO_2 emissions, casting doubt upon homogeneous regulations and underscoring that a shift toward differentiated, place-based spatial governance is needed in order to address the internal imbalances of carbon intensity.

The clear “Southwest-High, Northeast-Low” spatial pattern found here suggests that Guangzhou’s emission pattern is strongly conditioned by its past growth characteristics. In contrast with the results of [5], where emission structures were defined by peripheral sprawl, this study finds that emissions within this high-density setting are driven by the accumulated rigidity of the built environment. The tight clustering of these emitting hot spots inside the developed and dense downtown area (Liwan and Yuexiu districts) forms a continuous “Carbon Lock-in” region. This result is consistent with carbon lock-in theory [11], which suggests that the spatiotemporal direction of history generates a path dependence. In this core, it is the inflexible dense form of its urban shape which poses an obstacle to decarbonization and thus proves that even with policies promoting polycentrism as a strategy for growth, the environmental pressure is still mainly dominated by the history of the monocentric solution.

Regarding global driving factors, our OLS results highlight Road Density ($\beta = 0.137$) and Building Coverage ($\beta = 0.090$) as the most robust promoters of carbon emissions. Ju et al. [7] identified "spatial congestion" as a critical determinant of transport emissions; our study provides empirical validation of this mechanism within a high-density urban context. The significant positive correlation suggests that road networks in the city center function less as facilitators of mobility and more as proxies for traffic friction and induced demand. Similarly, while [19] proposed a "two-sided" effect of building density, suggesting potential reductions through efficiency, our findings reveal a predominantly negative environmental outcome. This aligns with the perspective of [17], which argued that in affluent city centers, the "scale effect" of high consumption and intense activity often overshadows the "technological effect" of shared infrastructure. Consequently, in Guangzhou’s central districts, physical compactness manifests as a stressor that exacerbates per capita emissions rather than a mitigation tool.

The GWR local coefficient maps also provide insights into how various indicators have distinct effects, either positive or negative, depending on location. For instance, the fact that “Distance to CBD” has strong negative coefficients in the center-north area (Fig. 3d, blue clusters) confirms the persistence of a classic monocentric decay model. This stands in contrast to the view that polycentric structures effectively lower emissions, implying that Guangzhou’s functional centrality remains highly concentrated despite planning interventions [10]. The "Distance to River" indicator exhibits a complex mixed effect. In some northern areas there is a negative sign, indicating that being closer to the river means more emissions, which is probably due to the past planning of the industrial and logistics belt along the riverside [10]. In some parts, the positive sign indicates that emission increases when moving away from the river, suggesting that these particular waterfronts serve as a low-carbon ecological buffer zone or a high-quality living area. Public Transport has a strong negative (inhibiting) impact mainly in the dense southern areas, providing a spatial refinement of [20] concerning the emission-reducing potential of transit efficiency; our spatial evidence suggests that such a reduction is conditional. Public transit acts as a "relief valve" only where population density is sufficient to render it a viable substitute for private vehicles. In lower-density areas, its potential to offset emissions remains unrealized. A further point deserving explanation is the sign reversal of Green Space between the global and local models. In the global OLS model, Green Space carries a positive and significant coefficient, whereas the zone-level and local GWR estimates indicate a predominantly negative, carbon-inhibiting role. This apparent anomaly arises because green space in central Guangzhou spatially co-locates with high-density, high-consumption residential districts, so the global coefficient partly absorbs the emission load of the surrounding urban fabric; once spatial non-stationarity is modeled explicitly, the mitigating role of vegetated land emerges locally. Methodologically, this is a typical case in which a global estimate conflates opposing local processes, underscoring the necessity of the local modeling strategy adopted here.

4.2 Policy Implications

In practice, these distinct mechanism regimes force us to move away from homogeneous density laws and propose instead an approach that we call “Precision Zoning Policy”. Firstly, for Zone 1 (Vitality-Traffic Driven) in the mature core, where emissions are strictly coupled with high functional intensity, policies should shift from growth management to "stock management." This

means carefully managing new high intensity construction not to exacerbate congestion, and applying "acupuncture-style" greening (e.g., vertical gardens, pocket parks). These micro-scale regeneration interventions not only compensate for the identified ecological deficits but also mitigate the urban heat island effect, thus indirectly reducing the cooling energy use of buildings. Secondly, at the growth fronts of Zone 2 (Morphology Dominant), i.e., where the building volume plays a dominant role, a “preventative low-carbon planning” approach should be adopted, i.e., planners should impose floor area ratio (FAR) ceilings, compulsory green building criteria, and related requirements, so as to decarbonize the physical expansion before the high-carbon mode of development is spatially locked in. Thirdly, for the outer Zones 3 and 4, policies must be shifted toward a “functional restructuring” in order to reduce travel demand. That is, for the industry-dominant Zone 4 (where commercial land shows a negative influence), it is suggested that more commercial facilities be introduced to create a harmony of production and living, minimizing the need to travel far from home; at the same time, within the location-sensitive Zone 3, investment in “last-mile” transport will be key to making mass transit a viable replacement for, rather than just an addition to, private cars, maximizing the reuse of existing infrastructure.

4.3 Limitations and Future Directions

While acknowledging the theoretical and practical value of this research, it is also necessary to point out its limitations to guide future work. First, due to the time dimension, this analysis is only based on the cross-sectional data of 2020. Considering the maturity of Guangzhou's urban form, although its spatial structure is relatively stable, the lack of longitudinal comparison still restricts the in-depth exploration of the dynamic evolution of the carbon-built environment relationship. In particular, 2020 was marked by the COVID-19 pandemic, during which mobility restrictions suppressed transport activity and the associated emissions; the cross-sectional relationships estimated here should therefore be read as reflecting a partly atypical year, and the pandemic-induced deviation from normal emission patterns is itself a caveat on the generality of the findings. It is necessary to introduce panel data covering both pandemic and non-pandemic years in subsequent research to reveal the long-term impact of the driving mechanism. Second, in terms of data accuracy, the CO_2 emissions are taken from the Open-Data Inventory for Anthropogenic Carbon dioxide (ODIAC), which focuses on fossil fuel burning. Although it is an extensively used and cited source at larger scales, it is a top-down estimate and therefore does not necessarily reflect bottom-up details such as non-energy-related emissions or particular biogenic sources within overall city-level emission budgets. Third, in terms of methodological depth, although this paper managed to decode the spatial heterogeneity of driving factors through GWR (where and how the impacts change), it did not quantify threshold effects. The locally linear regression model does not capture the “turning points” of density in cities. Future work can incorporate artificial intelligence methods and segmented regression models, which will allow us to estimate those turning points with greater accuracy for better governance.

5. Conclusions

This study developed and verified an innovative data-driven analysis framework that coupled geographically weighted regression (GWR) and a hybrid machine learning framework (SOM-FCM) for the complex carbon emission mechanisms in high-density metropolitan areas. Empirical results show that the carbon emission pattern in Guangzhou presents a significant gradient of "high in the southwest and low in the northeast", which profoundly reveals the "lock-in effect" of historical urban forms on modern emission paths. By introducing machine learning models to decode high-dimensional spatial heterogeneity, this study successfully identified four mechanism types with significant differences, namely Vitality-Traffic Driven, Morphology Dominant, Location-Sensitive, and Industrial-Logistics, demonstrating the superior performance of this coupled framework in handling both spatially non-stationary relationships (via GWR) and non-linear topological structures (via SOM-FCM) compared to traditional global linear models. This discovery not only provides a scientific basis for the paradigm shift from "generalized density control" to "precise zoning governance", but also constructs an operational and automated methodological blueprint of decarbonization governance for global high-density cities, thereby underscoring the value of data-driven techniques in optimizing urban spatial planning and assisting decision support.

References

[1] Deng Z, Li L, Du Y. Can polycentric urban morphology improve transportation carbon emission efficiency? Evidence from 285 Chinese cities, 2005-2020. *Transportation*. 2024 Nov; https://doi.org/10.1007/s11116-024-10551-z

[2] Ding G et al. The influence of urban form compactness on CO2 emissions and its threshold effect: evidence from cities in China. *J Environ Manage*. 2022 Nov;322:116032. https://doi.org/10.1016/j.jenvman.2022.116032

[3] Cai M et al. The need for urban form data in spatial modeling of urban carbon emissions in China: a critical review. *J Cleaner Prod*. 2021 Oct;319:128792. https://doi.org/10.1016/j.jclepro.2021.128792

[4] Feng T, Zhou B. Impact of urban spatial structure elements on carbon emissions efficiency in growing megacities: the case of Chengdu. *Sci Rep*. 2023 Jun;13(1):36575. https://doi.org/10.1038/s41598-023-36575-6

[5] Wei S, Xue Y, Zhang M. Study on spatiotemporal pattern evolution and regional heterogeneity of carbon emissions at the county scale of major cities, Inner Mongolia Autonomous Region. *Sustainability*. 2025 Oct;17(20):9222. https://doi.org/10.3390/su17209222

[6] Zheng Y et al. Spatial modelling of street-level carbon emissions with multi-source open data: a case study of Guangzhou. *Urban Clim*. 2024 May;55:101974. https://doi.org/10.1016/j.uclim.2024.101974

[7] Ju J et al. Constructing generalised local carbon emission zones (LCEZs): a 2D/3D urban morphology clustering framework for low-carbon planning. *Build Environ*. 2025 Nov;285:113599. https://doi.org/10.1016/j.buildenv.2025.113599

[8] Qin Y et al. Assessment of the impact of urban block morphological factors on carbon emissions introducing the different context of local climate zones. *Sustainable Cities Soc*. 2025 Feb;119:106073. https://doi.org/10.1016/j.scs.2024.106073

[9] Zhang X, Meng Q, Xu J, Li K. Urbanization impacts on anthropogenic carbon dioxide emissions and the roles of urban morphologies: insights from urban socioeconomic clusters and local climate zones in Shanghai, China. *Sustainable Cities Soc*. 2025 Jul;129:106494. https://doi.org/10.1016/j.scs.2025.106494

[10] Zhang B, Xin Q, Chen S, Yang Z, Wang Z. Urban spatial structure and commuting-related carbon emissions in China: do monocentric cities emit more? *Energy Policy*. 2024 Mar;186:113990. https://doi.org/10.1016/j.enpol.2024.113990

[11] Li C, Guan C, Zhang B, Ma R, Chen X. Unpacking the law of spatial directionality on urban expansion morphology and carbon emissions. *Sustainable Cities Soc*. 2025 Nov;134:106935. https://doi.org/10.1016/j.scs.2025.106935

[12] Li L, Sun S, Zhong L, Han J, Qian X. Novel spatiotemporal nonlinear regression approach for unveiling the impact of urban spatial morphology on carbon emissions. *Sustainable Cities Soc*. 2025 May;125:106381. https://doi.org/10.1016/j.scs.2025.106381

[13] Li Z et al. Enhanced mapping of essential urban land use categories in China (EULUC-China 2.0): integrating multimodal deep learning with multisource geospatial data. *Sci Bull*. 2025 Sep;70(18):3029–41. https://doi.org/10.1016/j.scib.2025.07.006

[14] Zhao P, Tian BS, Yang Q, Zhang S. Influencing factors and their spatial-temporal heterogeneity of urban transport carbon emissions in China. *Energies*. 2024 Feb;17(3):756. https://doi.org/10.3390/en17030756

[15] Wu WK et al. Unlocking the ecological blueprint: strategic zoning and policy guidelines for ecosystem services in Guangdong Province, China. *J Urban Plann Dev*. 2025 Jun;151(2):4025003. https://doi.org/10.1061/JUPDDM.UPENG-5315

[16] Wu W, Yang H, Wei L, Fang Z, Liu X. Spatiotemporal dynamics of ecosystem services and ecological zoning in rapidly urbanizing eastern Guangdong. *Front Earth Sci*. 2025 Sep;13:1593217. https://doi.org/10.3389/feart.2025.1593217

[17] Heinonen J, Kyro R, Junnila S. Dense downtown living more carbon intense due to higher consumption: a case study of Helsinki. *Environ Res Lett*. 2011 Sep;6(3):034034. https://doi.org/10.1088/1748-9326/6/3/034034

[18] Liu Y, Huang L, Kaloudis A, Store-Valen M. Does urbanization lead to less energy use on road transport? Evidence from municipalities in Norway. *Transp Res Part D Transp Environ*. 2017 Dec;57:363–77. https://doi.org/10.1016/j.trd.2017.09.021

[19] Wang J et al. Evaluation of the impact of urban morphology on commercial building carbon emissions at the block scale-a study of commercial buildings in Beijing. *J Cleaner Prod*. 2023 Jul;408:137191. https://doi.org/10.1016/j.jclepro.2023.137191

[20] Lyu S, Huang Y, Sun T. Urban sprawl, public transportation efficiency and carbon emissions. *J Cleaner Prod*. 2025 Jan;489:144652. https://doi.org/10.1016/j.jclepro.2025.144652