Exploring Spatiotemporal Pattern and Agglomeration of Road CO2 Emissions in Guangdong, China

Xingdong Deng [a,b]
Email: dengxingdong@gzpi.com.cn

Wangyang Chen [a,b]
Email: e0403826@u.nus.edu

Qingya Zhou [a,b,*]
Email: qy_zhou@yeah.net

Yuming Zheng [a,b]
Email: 779076427@qq.com

Hongbao Li [a,b]
Email: lihongbao@gzpi.com.cn

Shunyi Liao [a,b]
Email: liaoshunyi@gzpi.com.cn

Filip Biljecki [c,d]
Email: filip@nus.edu.sg

[*] Corresponding Author
[a] Guangzhou Urban Planning and Design Survey Research Institute, Guangzhou 510060, China
[b] Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510030, China
[c] Department of Architecture, National University of Singapore, Singapore
[d] Department of Real Estate, National University of Singapore, Singapore

November 23, 2022

Exploring Spatiotemporal Pattern and Agglomeration of Road CO2 Emissions in Guangdong, China

Abstract

Road transport is a prominent source of carbon emissions. However, fine-grained regional estimations on road carbon dioxide (CO2) emissions are still lacking. This study estimates road CO2 emissions in Guangdong Province, China, at high spatiotemporal resolution, with a bottom-up framework leveraging massive vehicle trajectory data. We unveil the spatiotemporal pattern of regional road CO2 emissions and highlight the contrasts among cities. The Greater Bay Area (GBA) is found to produce 76% of the total emissions, wherein Guangzhou emits the most while Shenzhen has the highest emission intensity. Emission agglomeration is still an under-explored field, which we advance in this paper. We propose Quantile-based Hierarchical DBSCAN (QH-DBSCAN) to explore road CO2 emission agglomeration in GBA. Our method is the first one to identify the specific location and scope of emission hotspots. Emission hotspots exhibit significant concentration on major urban centers. Considering emission characteristics from multiple perspectives, we derive six emission categories, including four emission zones and two emission connectors. The density-based property of our method results in spatially contiguous regions with similar emission patterns. Accordingly, we divide policy zones and propose targeted strategies for road carbon reduction. The study provides new technologies and insights to achieve regional sustainable development.

Keywords: Road transport, Carbon emission, Spatiotemporal distribution, Vehicle trajectory, GBA, Emission agglomeration

1. Introduction

Carbon dioxide (CO2) emissions pose a great threat to the global environment as a culprit of global warming. The pattern of carbon emission and its relationship with different aspects of human development is an enduring topic to discuss (Shindell et al., 2008; Zhang et al., 2014; Zhang and Da, 2015; Huang et al., 2018; Shan et al., 2021). Among the multitudinous sectors of human activities, transport sector accounts for more than 20% of the global carbon emissions. (Yan et al., 2017; Van Fan et al., 2018). The share continues to climb up (IEA, 2019; Mohsin et al., 2019), rendering transport a challenging sector in carbon emission mitigation (Yang et al., 2015; Batur et al., 2019). Among all the transport modes, road transport is the closest to our

Abbreviations: GBA, Guangdong-Hong Kong-Macao Greater Bay Area ; HDFV, Heavy-Duty Freight Vehicle; QH-DBSCAN, Quantile-Based Hierarchical Density-Based Spatial Clustering Of Applications With Noise; O-D, Origin-Destination.

daily lives and contributes to 82% of the total carbon emission in the sector (IEA, 2020). In this regard, estimating road carbon emissions and exploring the patterns are essential for relevant carbon reduction policy making.

To estimate road CO2 emissions, various kinds of data sources have been leveraged, but each faces different deficiencies. One group of studies use collective data, such as statistical yearbooks (Lin and Li, 2020; Zhou et al., 2018; Cai et al., 2018). Such method usually covers a broad geographic scale, but suffers from coarse resolution in both space and time. Some scholars seek to ameliorate the estimation granularity by using survey (Pérez-Martínez et al., 2020; McQueen et al., 2020; Sobrino and Arce, 2021; Patiño-Aroca et al., 2022), sensors (Liu and Zimmerman, 2021), smartphones (Manzoni et al., 2010) or video surveillance data (Li et al., 2019b). Although these methods can obtain fine-grained results, they are only applicable at a limited geographic scale. Balancing the trade-off between both data fineness and wide spatial coverage, vehicle trajectory is considered to be a suitable instrument to estimate road emissions (Kan et al., 2018; Xia et al., 2020; Sun et al., 2015; Zhao et al., 2017). However, existing vehicle trajectory datasets often only contain a subset of vehicle types, e.g. taxis (Zhao et al., 2017), ride-hailing (Sui et al., 2019), and light-duty vehicles (Li et al., 2019a). Emission patterns obtained with these datasets are characterized by low sampling rate and thus may considerably deviate from the actual pattern of the entire fleet, which is more valuable for policy-makers to understand the status quo and design emission mitigation strategies. Solving the problem calls for a trajectory dataset with higher sampling rate and more vehicle types. Correspondingly, a framework for estimating road CO2 emissions with massive vehicle trajectory dataset is also required.

Besides deficiencies in emission estimation, limited attention is paid to emission agglomeration. As an effective strategy to catalyze the development of economy, urbanization and industry (Fujita and Thisse, 1996; Malmberg and Maskell, 1997; Fang and Yu, 2017), agglomeration also brings some negative externalities, of which an important aspect is the convergence of carbon emissions (Yu et al., 2020; Wang et al., 2019). In China, 1% of the land contributes to 70% national CO2 emissions, and the regions with high level of urban agglomerations are also those emitting the most (Wang et al., 2014). In our context, we define the concept of emission agglomeration as the phenomenon that massive emissions are concentrated in certain contiguous territory. To better spatialize and visualize emission agglomeration, we introduce the "emission hotspot", which indicates the spatially continuous region with high emissions per unit area. Although regional emission agglomeration has been verified statistically with city-level or county-level emissions, studies on identifying emission hotspots is extremely lacking. The hotspots of CO2 emissions may be prone to worse environmental problems and could also be the crux for overall emission reduction. Thus, how to determine the location and scope of emission hotspots becomes crucial. However, the literature falls short on relevant methods, which is another gap that this study intends to fill.

In this paper, we conduct a regional study in Guangdong Province, which has the largest population and economy in China, and accordingly produces intensive emissions. Guangdong-Hong Kong-Macao Greater Bay Area (GBA) denotes the most developed region in Guangdong and is primary to the overall development of China (Hui et al., 2020). According to the *Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area* published by Chinese State Council, low-carbon development is among the major objectives of GBA. GBA maintains great economic, population and urban agglomeration (Chen et al., 2017, 2020; Yu, 2019), making it a favorable region to study mesoscale emission agglomeration (Zhou et al., 2022a). Although some insights have been given in spatial (Lin and Li, 2020; Chen et al., 2017) and temporal (Zhou et al., 2018; Chen et al., 2017) patterns of carbon emissions in Guangdong or GBA, the results are either coarse in granularity or focus on total emissions instead of road emissions. Fine-grained regional patterns of road transportation CO2 emission and agglomerations in Guangdong and GBA is still under-explored. Exploring regional patterns of road CO2 emission is the prerequisite GBA to formulate strategies for sustainable development.

Taking into account the research opportunities and deficiencies described so far, we aim to bridge the following gaps and present the following contributions to the field: (1) We implement a bottom-up framework for estimating road CO2 emissions based on a vehicle trajectory dataset with sampling rates surpassing the counterparts in the literature, and develop approaches to handle inherent data deficiencies. (2) We present fine-grained spatiotemporal patterns of regional road CO2 emissions in Guangdong Province where similar analysis is lacking. (3) We propose an approach to identify hierarchical emission hotspots, which uncover the spatiotemporal emission agglomeration patterns in GBA. (4) We categorize our study area into emission zones with

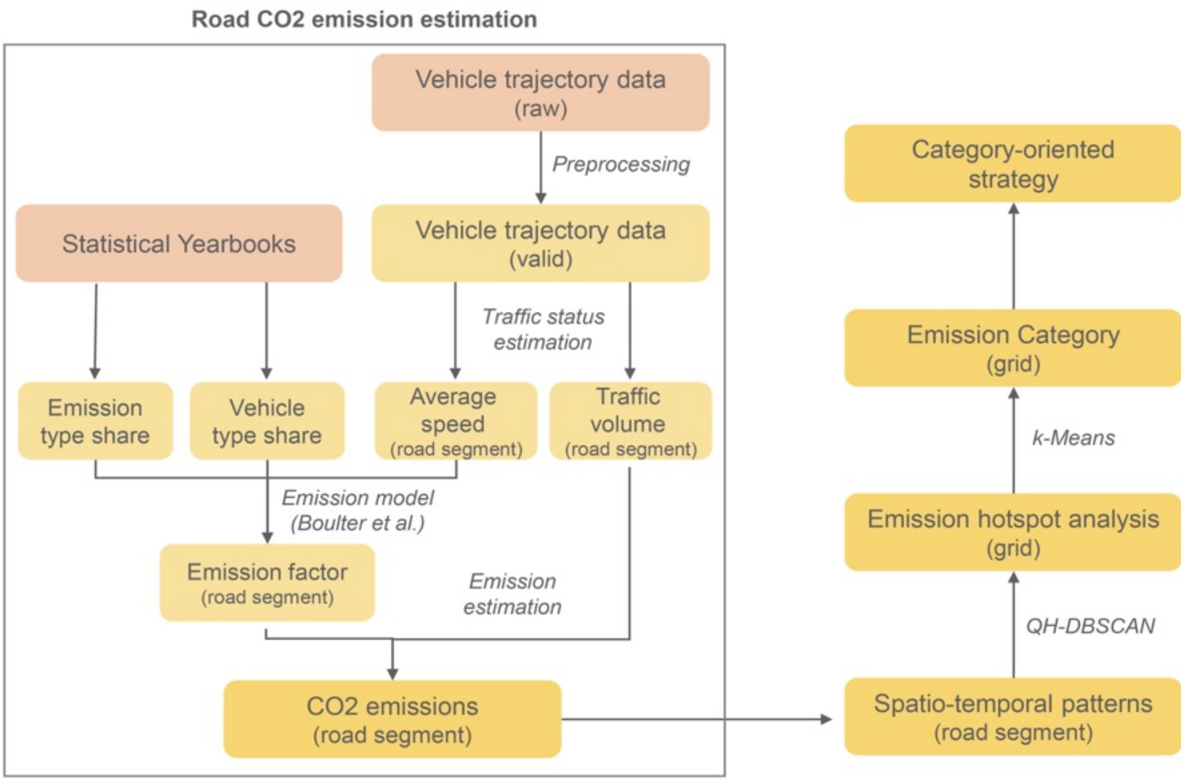


Figure 1: Analytical framework.

distinct emission patterns and propose targeted emission reduction strategies.

The analytical framework of this study is designed as illustrated in Figure 1. Our study estimates CO2 emissions at road segment level at the first stage. To achieve such an advancement, we harness a massive vehicle trajectory dataset that contains 12.9 million records in a day, encompassing all vehicle types. Its sampling rate exceeds most counterparts used hitherto. We firstly introduce the dataset and elucidate the estimation process. Then we analyze the spatial distribution and the hourly variation of road CO2 emissions in Guangdong, highlighting the disparities among cities. Next, a density-based clustering algorithm is proposed to detect hierarchical emission hotspots. Their temporal variations are also explored. Finally, we integrate all facets of emission features and divide our study area into different emission categories. Based on the emission characteristics of each category, targeted carbon reduction strategies are proposed for the sustainable development in GBA and Guangdong.

2. Methodology

2.1. Study area

Our study area (Figure 2) includes all the twenty-one prefectural-Level cities in Guangzhou Province. We highlight the nine GBA cities (Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai,

Figure 2: Study area: Guangdong Province (colored area) and GBA (pink area). Source of the basemap: (c) Esri.

Huizhou, Zhongshan, Jiangmen and Zhaoqing), considering their paramount role in the region. Hong Kong and Macao, which also belong to GBA but are not part of Guangdong Province, are not included in this study due to their absence in the trajectory dataset. Thus, we refer to the nine aforementioned and contiguous cities when we refer to GBA in this paper. The socioeconomic indicators of the corresponding cities are displayed in Supplementary Table S1 according to the Statistical Yearbook of Guangdong Province (2021) (Guangdong Province Statistical Bureau, 2021). GBA accommodates 62% of the population and contributes to 81% of the GDP with only 30% of land in Guangdong. It is four times as densely populated and three times as economically developed as the rest areas in Guangdong, showing significant agglomerations. To put that in context for international readers, these values are comparable with the population of Germany, economy of Brazil, and area of Croatia. Considering that GBA also has denser road networks and heavier traffic (Hui et al., 2020), in this study, we estimate the road CO2 emissions and delineate the patterns in the entire province, but focus on GBA when exploring emission hotspot patterns.

Table 1: Summary statistics of the trajectory data.

Vehicle type	Raw trajectories	Valid trajectories	Samples	Total travel distance (10^4 km)	Sampling rate (%)
HDFV	527,396	520,924	137,963,458	505.1	100
Non-HDFV	12,371,644	10,623,090	385,648,577	1022.4	40
Total	12,899,040	11,144,014	523,612,035	1527.5	–

2.2. Data description and preprocessing

In this study, we introduce a new vehicle trajectory dataset to estimate road CO2 emissions. The dataset is provided by PalmGo, a company offering driving navigation services in China. The dataset contains 12.9 million navigation trajectories of individual vehicles in a day (Monday 19 October 2020) in Guangdong (Table 1). The dataset covers our study area well and is verified representative to the general traffic condition (Supplementary Note S1). A trajectory sample is shown in Supplementary Table S3. Each trajectory records the location and time of departure and arrival. The route is represented by a series of passed road segments (*SegmentIDs*). Every time the vehicle drives onto a new road segment, the road segment ID, direction, and time of entry would be appended to the record. By matching trajectory records with road network geodata provided by PalmGo, we can reconstruct the trajectories. Vehicles are divided by two types: heavy-duty freight vehicle (HDFV) and the others (non-HDFV). Since all HDFVs are obliged to install Global Positioning System (GPS) devices by law in China, the sampling rate of HDFV fleets is 100%. For non-HDFVs, the overall sampling rate is 40% according to the company. Despite covering the entire fleet with favorable sampling rate, the dataset is subject to one major issue. All the time-related fields of trajectory records are only accurate to minute, which causes difficulty in computing vehicular instantaneous status (speed and acceleration). To exclude abnormal records that may root from device errors, trajectories less than 500 m or with single sample are filtered out. As a result, we obtain 11.1 million valid trajectories.

Meshing is a common approach to regularize and integrate anisotropic spatial attributes (road networks) to isotropic ones (grids). In our case, CO2 emissions would be measured at road segment level as the foundation. To facilitate emission hotspot analysis, the emissions would be further allocated to regular grids. Instead of defining our own grids, we use the 1 km-resolution grids of WorldPop (WorldPop, 2018; Tatem, 2017) to integrate the population information and make our results attachable to other relevant initiatives using the same grid system (Chen et al., 2021).

2.3. Traffic status estimation

To estimate road CO2 emissions with trajectories, the first step is to calculate the traffic status on road. Some studies are able to compute the instantaneous speeds and accelerations of individual vehicles at different positions based on second-level trajectory data (Böhm et al., 2022; Deng et al., 2020). Since our dataset is only accurate to minute, we propose the following strategies to estimate the traffic status with minimal deviation. Instead of distinguishing individual vehicles, we emphasize the average traffic status per road segment and update the status every hour. Besides, we only calculate speeds and avoid computing accelerations which require better fineness in time. Traffic volume per road segment per hour is gleaned by counting the number of passed trajectories. The volume of HDFVs and non-HDFVs is counted separately, but they are summed up to estimate the speed. The average speed on a road segment at an hour is calculated by dividing the aggregated travel distance by the total time cost (Equation 1).

$$s_{i,t} = \frac{|U(i,t)| \cdot l_i}{\sum\limits_{u \in U(i,t)} T_{u,i}} \tag{1}$$

where $s_{i,t}$ denotes the average speed on road segment i at hour t, $U(i,t)$ is the set of vehicle trajectories that pass road segment i at hour t, and $|\cdot|$ represents the size of the set. In our case, $|U(i,t)|$ or $q_{i,t}$ denotes the traffic volume on road segment i at hour t. l_i is the length of road segment i, and $T_{u,i}$ is the passing time of vehicle u through road segment i.

Another challenge is how to estimate actual passing times on road segments ($T_{u,i}$) with minute-level trajectories. Considering the way the trajectory is recorded (Supplementary Table S3), simply regarding time differences between consecutive road segments as passing times becomes infeasible because of two issues. First, if a vehicle passes through more than one road segment at the same minute, the time difference would be zero, resulting in the inability to calculate speed. Second, even with a non-zero time difference, the actual passing time would still be out of calibration. For example, a time difference of 1 minute could correspond to any actual passing time between 1 and 120 seconds. Out of our control, since 93.2% of the road segments are shorter than 500 m, 95.4% of the time differences are no larger than 1 min, rendering it a pervasive problem. To alleviate the problem, we amend the time differences based on following strategies. We introduce the minimal time cost (T^i_{min}) for each road segment, which is calculated with the maximum speed limit. The maximum speed is determined by the road class referring to the speed limit regulation (100 km/h for highways, 80 km/h for provincial and urban expressways, 60 km/h for arterial roads, 40 km/h for ramps and secondary roads, and 20 km/h for branch roads). $T_{u,i}$ smaller than T^i_{min} is replaced by T^i_{min}. Furthermore, we mitigate the uncertainty by accumulating the length and time difference of continuous segments until the added-up time difference exceeds a minimum threshold (τ). Then we calculate the average speed based on the cumulative length and time, and reallocate the cumulative passing time to each road segment proportional to the segment length. After grid searching with τ from 2 min to 7 min, τ is finally set to 5 min as the trade-off between estimation accuracy and variance. Mathematically, the speed error caused by time uncertainty is constrained within a range of $\pm 20\%$ (-1/5 to 1/5).

2.4. Road CO2 emission estimation

Table 2: Weighted average parameters to calculate emission factors for HDFV and non-HDFV. (Parameter $a-k$ are the coefficients of the polynomial fitting function between speeds and emission factors provided by (Boulter et al., 2009).)

Vehicle type	a	b	c	d	e	f	g	k
HDFV	5237.8	1003.4	-20.164	0.1603	1.28e-3	-1.67e-5	3.28e-8	1
Non-HDFV	2735.7	104.2	-0.468	0.0098	-3.43e-5	2.39e-7	-4.09e-10	1

With the segment-level traffic status per hour, we estimate the road CO2 emissions with a speed-based microscopic emission model (Boulter et al., 2009), which fits our data characteristics and emission standard system well, and has been widely applied in various scenarios (Carslaw et al., 2010; Lomas et al., 2010; Hicks et al., 2021). The model determines the emission factor f by speed and parameters varying with vehicle attributes including the vehicle type, fuel type and emission standard (Equation 2). Due to lack of these vehicle attributes, we follow the approach commonly used in the literature (Pla et al., 2021; Zhou et al., 2022b), by assuming that the on-road shares of vehicle types, fuel types, and emission standards are consistent across all road segments, and estimating CO2 emissions with weighted average emission factors. Non-HDFV consists of all the vehicle types except for HDFV, including large passenger vehicle, medium passenger vehicle, small passenger vehicle, mini passenger vehicle, medium freight vehicle, light freight vehicle, and mini freight vehicle. The shares of vehicle types in non-HDFVs are obtained from the vehicle possession structure in 2020 (Guangdong Province Statistical Bureau, 2021). The shares of fuel types and emission standards refer to the national level in 2018 (MEE, 2019). Despite some time inconsistencies, it is the official data closest in time to our trajectories. CO2 emissions from new energy vehicles are ignored in this study, given that new energy vehicles only account for 1.75% of the total vehicles in China in 2020 (MEE, 2021). Details in shares of vehicle types, fuel types, and emission standards could be found in Supplementary Table S4-S6, while the parameters to calculate CO2 emission factors for each combination are available Supplementary Table S7. With all of this information, we obtain the weighted average emission factors for HDFVs and non-HDFVs (Table 2), respectively. Accordingly, road segment-level CO2 emissions are estimated with the following formulas (Equation 3, 4, and 5):

$$f_{i,t} = k \cdot (a + bs_{i,t} + cs_{i,t}{}^2 + ds_{i,t}{}^3 + es_{i,t}{}^4 + fs_{i,t}{}^5 + gs_{i,t}{}^6)/s_{i,t} \tag{2}$$

$$E_{i,t}^{non} = \frac{1}{\beta_{non}} \times (f_{i,t}^{non} \times l_i \times q_{i,t}^{non}) \tag{3}$$

$$E_{i,t}^{hdfv} = \frac{1}{\beta_{hdfv}} \times (f_{i,t}^{hdfv} \times l_i \times q_{i,t}^{hdfv}) \tag{4}$$

$$E_{i,t} = E_{i,t}^{non} + E_{i,t}^{hdfv} \tag{5}$$

where a-f are the weighted average parameters depending on vehicle type (Table 2), $E_{i,t}^{hdfv}$ and $E_{i,t}^{non}$ denote CO2 emissions at segment i and hour t from HDFVs and non-HDFVs respectively, $E_{i,t}$ is the total CO2 emissions, $f_{i,t}^{hdfv}$ and $f_{i,t}^{non}$ are the average CO2 emission factors (g/km), and $q_{i,t}^{hdfv}$ and $q_{i,t}^{non}$ indicate traffic volumes. The sampling rate of HDFVs β_{hdfv} and non-HDFVs is 1.0 and 0.4 respectively.

2.5. *Emission hotspot detection (QH-DBSCAN)*

Considering that emissions are continuously distributed in space, we deem density-based clustering algorithms an appropriate option to capture the hotspots. Density-based spatial clus-

Table 3: Description of emission hotspots at different levels.

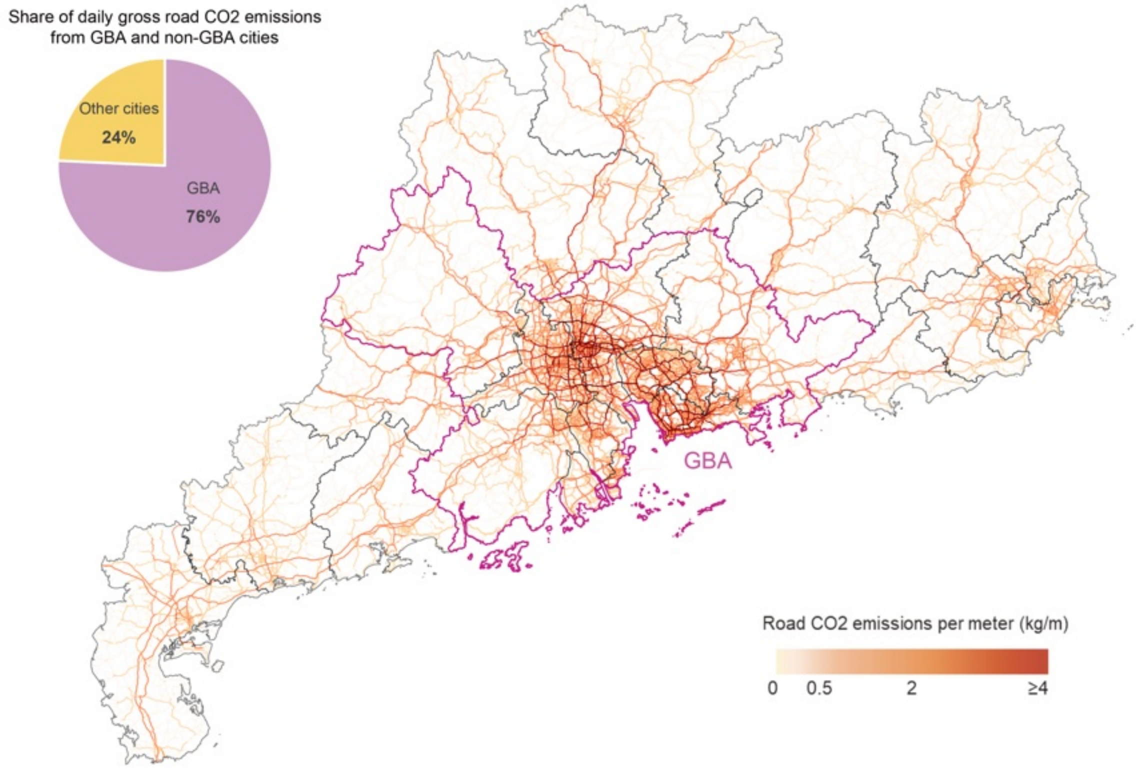
Level	Color	Quantile	Description
High	Red	90%	Clustered grids with emissions higher than the 90% quantile
Middle	Orange	75%	Clustered grids with emissions higher than the 75% quantile
Low	Green	50%	Clustered grids with emissions higher than the 50% quantile

tering of applications with noise (DBSCAN) (Ester et al., 1996) is widely used to detect clusters based on the proximity in the feature space. To decipher the hierarchy of emission hotspots, on the ground of DBSCAN, we propose a method called Quantile-based Hierarchical DBSCAN (QH-DBSCAN) to recognize hierarchical emission hotspots, which refer to spatially clustered high-emission areas. First, road-level CO2 emissions are allocated onto WorldPop grids. We use quantiles to identify high-emission grids. To explore a hierarchy of emission hotspots, we prepare a list of quantiles and repeat emission hotspot detection with each. For each quantile, the corresponding emission hotspots are identified by feeding the grids with emissions over the quantile into a DBSCAN model. We fully appraise the uncertainty hidden behind the method. The list of quantiles is determined through exploratory experiment on the relationship between the quantiles and the covered percentages of total emissions, while the hyper-parameters of DBSCAN are determined by sensitivity analysis. Both experiments would be elaborated in Section 3.3. The selected quantiles with the associated levels of emission hotspots are displayed in Table 3. To clarify, the 90%-quantile grids are equivalent to the top 10% emitting grids, and so on. QH-DBSCAN has two major merits. First, it could detect hierarchical emission hotspots and present the results with intuitive maps. Second, as a density-based clustering method, it leads to results with spatial contiguity, which is more useful for potential policy making.

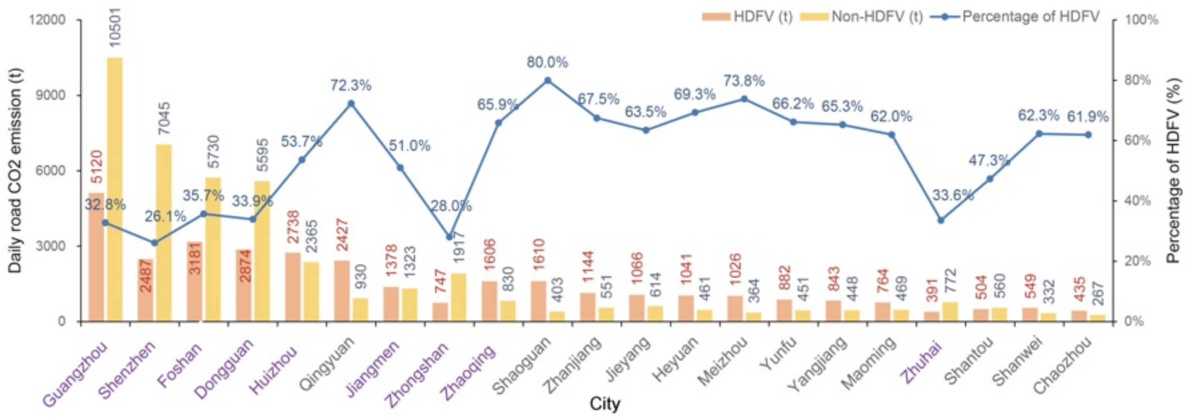
3. Results

3.1. Spatial pattern of road CO2 emissions in Guangdong.

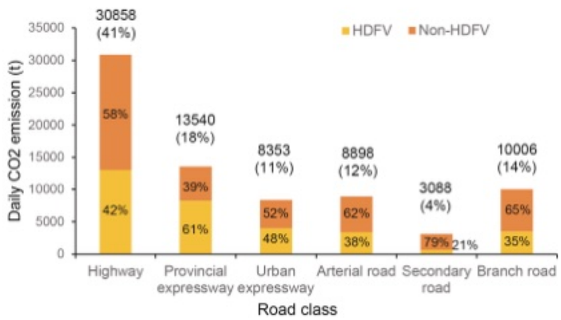
Based on the vehicle trajectory dataset and the proposed emission estimation method, segment-level CO2 emissions per hour are obtained. The spatial distribution of daily gross road CO2 emission intensity is illustrated in Figure 3(a). Emission intensity denotes CO2 emissions per meter for each road segment. Road CO2 emissions in Guangdong Province show significant spatial agglomeration. GBA accounts for 76% of the total emissions, while non-GBA cities



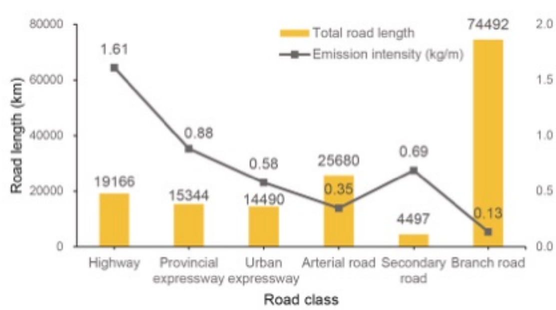
(a)



(b)



(c)



(d)

Figure 3: Daily gross road CO2 emission pattern in Guangdong Province. **(a)** Daily gross CO2 emission intensity per road segment in Guangdong. **(b)** Daily gross road CO2 emissions per city and vehicle type (GBA cities in purple). **(c)** Daily gross road CO2 emissions per road class and vehicle type. **(d)** Emission intensity and total road length per road class.

generate considerably less emissions on a broader land. Comparing the daily gross emissions among the cities (Fig 3(b)), the capital city, Guangzhou, appears to be the topmost emitter, producing 15621t CO2 and contributing to 21% of the total. It is followed by the other major city, Shenzhen, with 9532t of road CO2 emissions. Although the total emission is lower than Guangzhou, Shenzhen is the most densely emitting city, with an average emission intensity (1.01 kg/m) twice as much as the provincial average (0.49 kg/m). Foshan and Dongguan are the other two emission-intensive cities with a total emission close to the one of Shenzhen. Their emission intensities are between Shenzhen and Guangzhou. These four cities are the major industrial cities in Guangdong. For the rest cities, both the total emissions and emission intensity drop dramatically. Zhuhai is the only GBA city at the bottom of the emission ranking. Qingyuan produces the most road CO2 among the non-GBA cities.

GBA cities appear to have higher percentages of road CO2 emissions from non-HDFVs. Comparing the shares of emissions from different vehicle types across the cities, we find that GBA cities tend to have less proportions of HDFV emissions (generally around 30%), while those of most non-GBA cities are over 60%. GBA cities possess 71.7% of the non-HDFVs in Guangdong (Guangdong Province Statistical Bureau, 2021), but they generate 86.0% of the provincial non-HDFV emissions by our estimation. To discover whether non-HDFVs in GBA cities tend to emit more in average, we calculate and display the road CO2 emissions per non-HDFV possession across the cities in Supplementary Table S8. The results are affirmative. The average emissions per non-HDFV possession for GBA cities are 2.24 kg, while those for non-GBA cities are only 0.92 kg. GBA cities occupy the top 8 places in this indicator. In comparison, regarding emissions per HDFV possession, the mean of non-GBA cities is 16.02 kg, which is a bit larger than that of GBA cities (11.14 kg). The difference is not as significant as that of non-HDFV. Our findings suggest that non-HDFVs are used in a more emission-intensive way in GBA than outside. It may be due to longer travel distances, more frequent vehicle travel, and severer congestion in GBA. However, we also cannot rule out the possibility that drivers in non-GBA cities use navigation less due to a less complicated traffic system and smaller active area, which leads to undersampling of trajectories in these regions.

The distribution of CO2 emissions by road class is also heterogeneous (Figure 3(c)). Highway generates the most emissions (41% of the total), with 42% from HDFVs. The emission intensity on highways is also the highest with 1.61 kg/m in average (Figure 3(d)). Provincial expressways as the other major regional connector, emit the second most CO2 (18% of the total). It has the highest share of HDFV emissions (61%), indicating its vital role in regional freight transportation. The rest road classes mainly serves local transportation. Although their emission intensities are generally more moderate, they cover 77% of the total road length and accordingly contribute to 50% of the total emissions. In particular, the branch roads exhibit substantially minor emission intensity (0.13 kg/m), but the massive length makes them generate 14% of the total emissions.

3.2. Temporal pattern of road CO2 emissions

The hourly variations of road CO2 emissions during the day are revealed in Figure 4. Given that the total emissions of GBA and non-GBA cities differ considerably, we further compare their temporal fluctuation on emissions (Figure 4(a), 4(b)). In general, the city-level road CO2 emissions share a "three-stage" pattern, including an ascending stage (4:00-10:00), a plateau stage (10:00-19:00) and a descending stage (19:00-4:00 the next day). During the plateau stage, the emissions experience a slight decrease at noon (11:00-13:00). An obvious trough is observed at dawn (around 4:00). To quantify the emission variance of a city in a day, we define the variation rate as the maximum hourly emissions divided by the minimum. Variation rates of GBA cities (ranging from 2.7 for Zhaoqing to 9.6 for Zhongshan) surpass non-GBA cities (ranging from 1.8 for Shaoguan to 6.8 for Shantou). Guangzhou is the top emitter at all hours, with a variation rate of 4.7. The other three emission-intensive cities, Shenzhen, Foshan and Dongguan, possess larger variation rates, which are 6.7, 6.8, and 5.3 respectively.

To provide more insights in GBA, we compare the temporal patterns by vehicle types. The hourly variations of road CO2 emissions from HDFVs and non-HDFVs for GBA cities are demonstrated in Figure 4(c) and 4(d), respectively. Non-HDFV emissions follow a very similar "three-stage" pattern with the total emissions. In comparison, HDFV emissions fluctuate less during the day. Guangzhou is the city with the most stable and intensive HDFV emissions during the day, manifesting its dominant position in freight transportation in GBA. There are two obvious valleys of HDFV emissions at 6:00-9:00 and 17:00-20:00 in Guangdong, Shenzhen, Dongguan and Foshan. The declines may be attributed to HDFV traffic control during the peak time. In these four major cities, HDFVs show some degree of off-peak travel, and the pattern

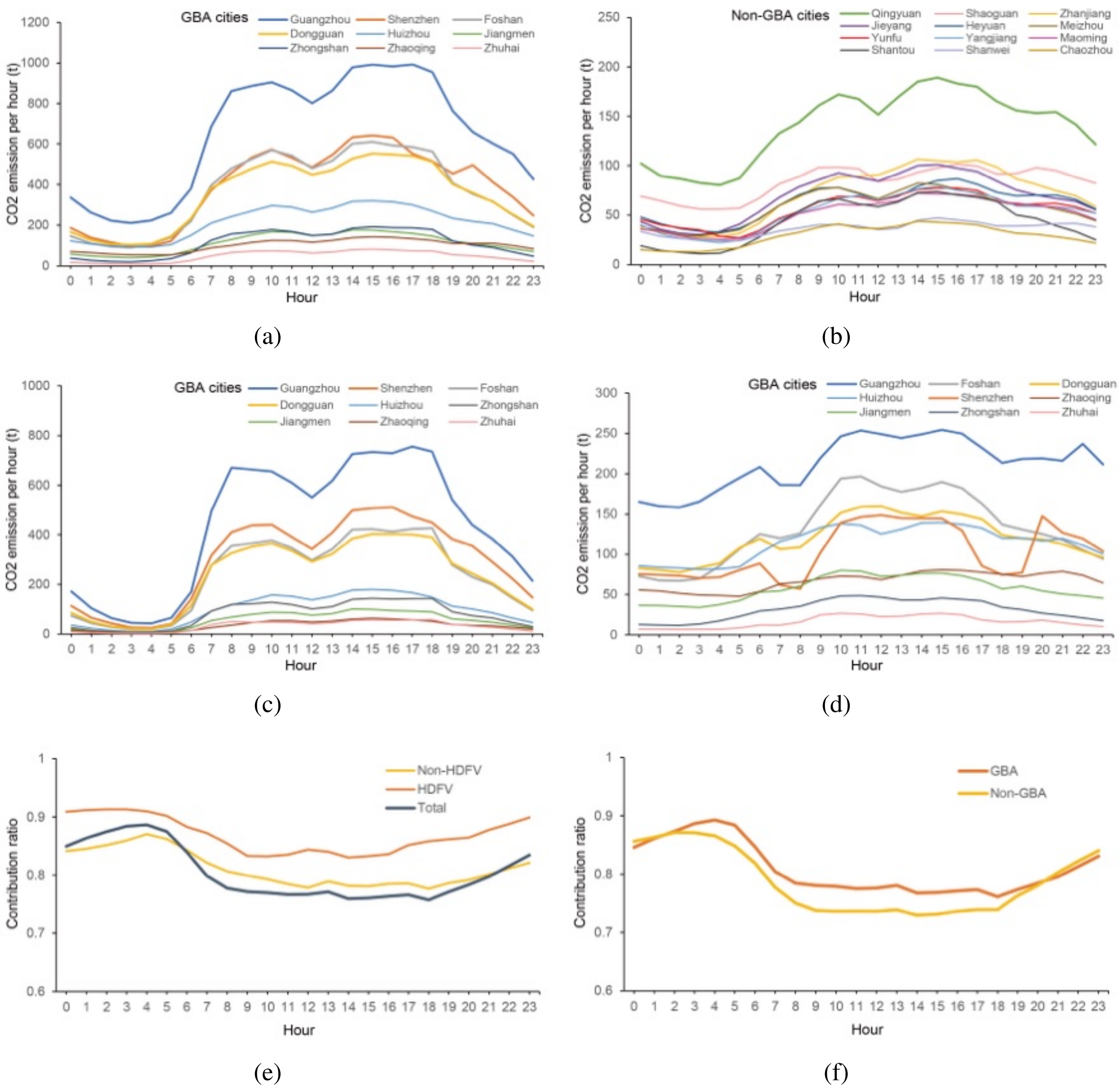


Figure 4: Temporal patterns of road CO2 emissions in Guangdong. **(a)** Hourly variations of road CO2 emissions for GBA cities. **(b)** Hourly variations of road CO2 emissions for non-GBA cities. **(c)** Hourly variations of non-HDFV CO2 emissions for GBA cities. **(d)** Hourly variation of HDFV CO2 emissions for GBA cities. **(e)** Hourly variations of contribution ratio for GBA and non-GBA cities. **(f)** Hourly variations of contribution ratio per vehicle type in Guangdong.

is the most significant in Shenzhen. However, there are still apparent overlaps of high emissions from HDFVs and non-HDFVs during the period between the morning and evening peaks, rendering a space for relevant policy making.

The study also outlines the temporal pattern of road CO2 emission agglomeration. We use the contribution ratio of the top 10% emitting road segments to the total emissions to quantify the level of emission hotspots in a region. In Guangdong, the top 10% emitting roads produce 77.6% of the total emissions, presenting significant emission agglomeration. The phenomenon is also reported in other city-level road emission studies (Böhm et al., 2022; Chen et al., 2022). The contribution ratio of GBA cities (78.6%) exceeds the provincial average, while non-GBA cities (75.9%) is below. The hourly variation of the contribution ratio of GBA and non-GBA cities is plotted in Figure 4(e). For both groups of cities, the contribution ratio culminates near 90% around the valley hour of emission (4:00-5:00). It indicates that, although the total emission shrinks at dawn, emissions are actually more concentrated on few roads. In terms of vehicle types (Figure 4(f)), HDFV emissions are more concentrated than non-HDFV emissions at all time. Nevertheless, the emission agglomeration of non-HDFV emissions is also nonnegligible.

3.3. Emission hotspot analysis

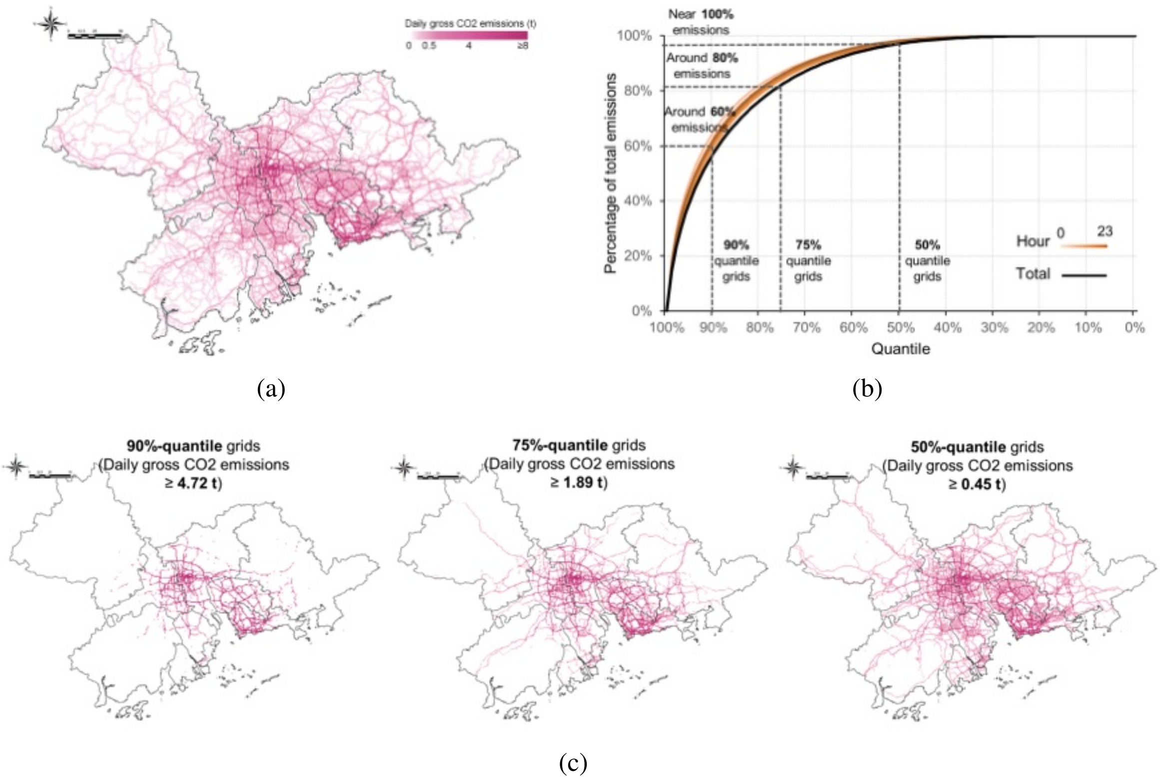


Figure 5: Process of quantile selection for emission hotspot and analysis. **(a)** Spatial distribution of grid-level daily gross road CO2 emissions in GBA. **(b)** Curves on relationships between quantiles and the covered percentages of total emissions. **(c)** Spatial distributions of grids associated with the three selected quantiles.

In the preceding sections, we have presented the spatiotemporal pattern of road CO2 emissions in Guangdong. The findings suggest that GBA contributes to 76% of the total emissions with only 30% of the land. GBA is also found to exhibit more significant emission agglomeration gauged by contribution ratios. Thus, in the following analysis, we focus on GBA to conduct emission hotspot analysis. To regularize the road-level emissions spatially, we lattice the GBA area into 70,750 1km-resolution (0.01°) grids and aggregate the emissions within each grid, resulting in the grid-level map of daily gross emissions in GBA (Figure 5(a)).

Before implementing QH-DBSCAN method to identify emission hotspots, we need to determine the list of quantiles. The relationship between quantiles and the percentages of the covered daily gross CO2 emissions is illustrated in Figure 5(b) (the black curve). As the quantile decreases,the percentage increases dramatically at first. The 90%-quantile girds have covered around 60% of the total emissions. After that, the slope becomes flatter gradually and near all the emissions are covered with the minimum amount of grids at the 50% quantile. Emissions at different hours also follow a similar trend. Given the observation, we select 90% and 50% quantiles as the upper and lower limits respectively. To introduce some continuity and hierarchy, the 75% quantile is added, which is associated with a percentage in the middle (80%). As a result, we confirm three quantiles, namely 90%, 75% and 50%. The spatial distribution of the grids fulfilling each quantile is shown in Figure 5(c). The 90%-quantile grids primarily distribute at the downtown of Guangzhou and Shenzhen and along the regional connectors among the four major cities. More areas with backbone roads are included for the 75%-quantile, and most built areas in GBA are involved for the 50%-quantile.

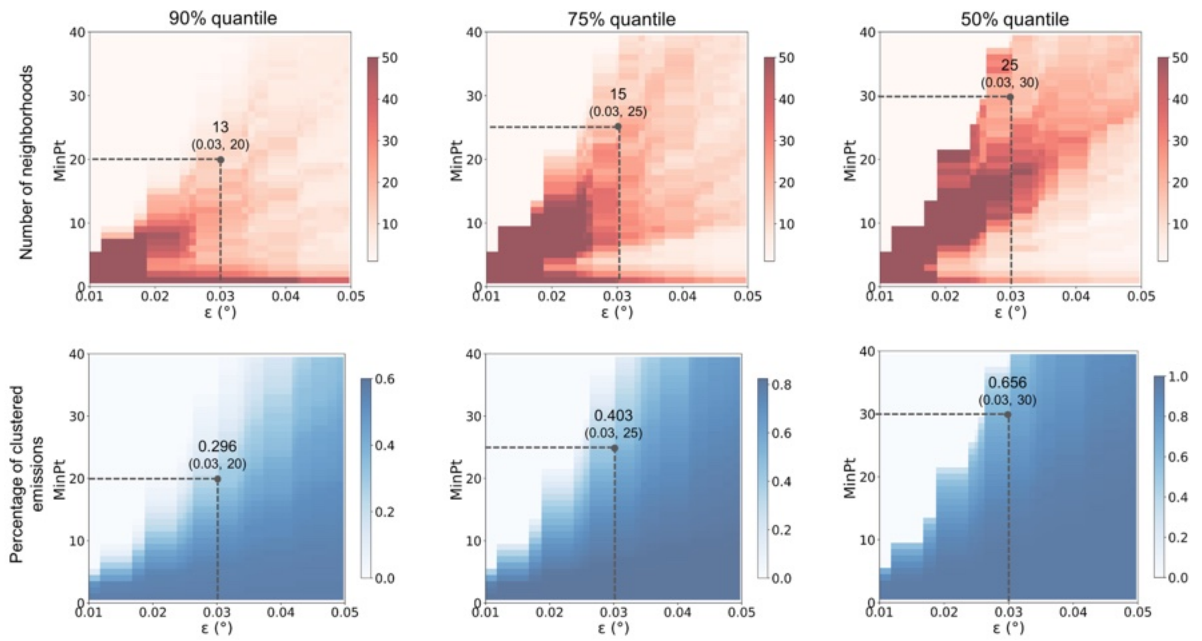


Figure 6: The process of determining hyper-parameters of DBSCAN: the maximum neighborhood-searching distance(ε) and the minimum sample in a neighborhood ($MinPt$).

Have the quantiles determined, we perform DBSCAN with grids with emissions over each quantile respectively. To minimize the uncertainty caused by parameter selection, we design a sensitivity analysis on the two main hyper-parameters of DBSCAN, the maximum neighborhood-searching distance(ε) and the minimum sample in a neighborhood ($MinPt$). The number of neighborhoods and the percentage of clustered emissions, are the two metrics that we use to evaluate the rationality of clustering results. Based on daily gross emissions, we visualize how these two metrics change with different combinations of parameters (Figure 6). In general, the number of neighborhoods decreases the fastest when either ε or $MinPt$ increments individually. When the hyper-parameters grow synchronously, the rate of decrease slows down, and the smaller the quantile, the more obvious the decrease is. For the percentage of clustered emissions, the trend is simpler. Larger ε induces more clustered emissions, while $MinPt$ is the opposite. We note that extreme values of either metric can lead to unwanted clustering results. A too large number of neighborhoods could result in small and dispersive emission hotspots, while a too small one makes emission hotspots over-merged. Similarly, the percentage of clustered emissions reflects the severity of the conditions for identifying emission hotspots, which cannot be too loose nor too tight. Therefore, our strategy is to select a parameter couple that yields moderate values of both metrics at the same time. ε determines how stringent we define the proximity between grids, so it should be constrained by the size of the grids. Given that our grids are in 0.01° resolution, the available range of ε is from 0.01° to 0.05°. Finally, striking a balance among all the factors aforementioned, the hyper-parameters are settled for each quantile: 90% quantile ($\varepsilon = 0.03°$, $MinPt = 20$), 75% quantile ($\varepsilon = 0.03°$, $MinPt = 25$) and 50% quantile ($\varepsilon = 0.03°$, $MinPt = 30$).

With all the prerequisites set, we obtain the hierarchical emission hotspots of daily gross road CO2 emissions in GBA (Figure 7(a)). Metrics of different levels of emission hotspots are presented in Figure 7(b). Most hotspots locate in the core area of GBA (the zoomed-in area). Statistically, the 90%-quantile hotspots emit 29.6% of the total emissions with only 5% of land in GBA . The 75%-quantile hotspots cover twice the area of the 90%-quantile hotspots, contributing to 40.3% emissions. 50%-quantile hotspots aggregate 65.7% of emissions on 23% of land. For each quantile, we plot the location and scope of every single emission hotspot with the ranking of total emissions (Figure 7(c)). The 90%-quantile hotspots include two primary components, namely the Guangzhou-Foshan metropolis and Shenzhen center, exhibiting huge emission volume that far exceeds other hotspots scattering at the major industry and transportation nodes in Guangzhou, Shenzhen, Foshan and Dongguan. The results suggest that road CO2 emissions in Guangzhou and Foshan have formed a unified high-level emission hotspot, reflect-

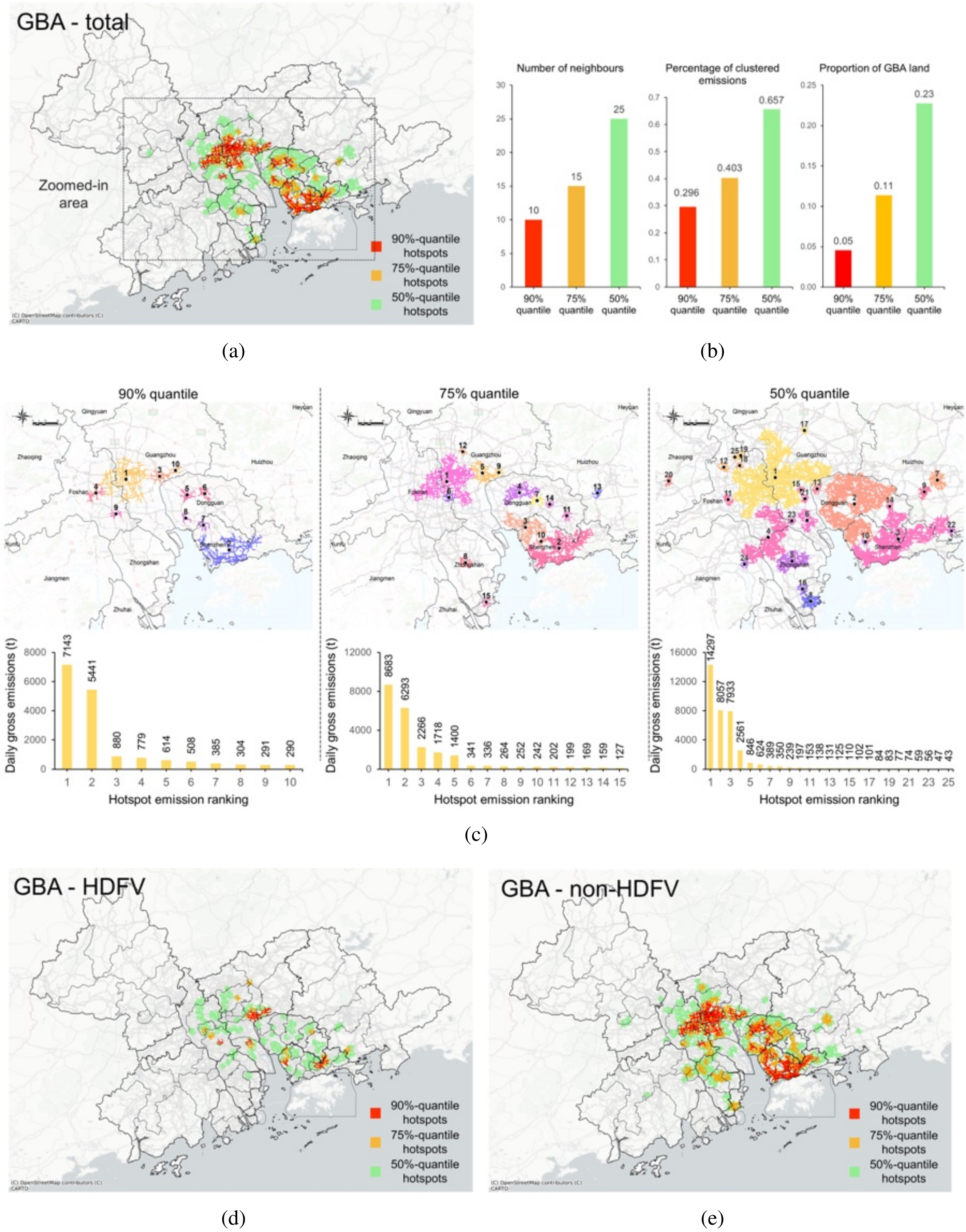


Figure 7: Maps and statistics of hierarchical emission hotspots in GBA. **(a)** Hierarchical emission hotspots of daily gross road CO2 emissions in GBA. **(b)** Metrics of different levels of emission hotspots. **(c)** Spatial distribution and emission rankings of different levels of emission hotspots. **(d)** Hierarchical emission hotspots of daily CO2 emissions from HDFVs in GBA. **(e)** Hierarchical emission hotspots of daily CO2 emissions from non-HDFVs in GBA. Source of the basemap: (c) OpenStreetMap contributors.

ing the promotion of integrated urbanization of the two cities (Zhang et al., 2021b). For the 75%-quantile hotspots, the two primary hotspots retain and absorb broader area around. Some adjacent 90%-quantile hotspots are merged into 75%-quantile hotspots, and glean second-tier emissions. 90%-quantile hotspot 5 and 6 are integrated into 75%-quantile hotspot 4 in Dongguan downtown. 90%-quantile hotspot 7 and 8 become 75%-quantile hotspot 3 at Shenzhen-Dongguan junction area. Meanwhile, some new hotspots pop up in central area of other small cities, including Zhongshan (75%-quantile hotspot 8), Huizhou (75%-quantile hotspot 13) and Zhuhai (75%-quantile 15). The top 50%-quantile hotspots (1 - 4) are characterized by huge footprints. Since the 50%-quantile hotspots are obtained with nearly all the grids with emissions, they reflect the spatially contiguous road CO2 emitting zones in GBA. One observation is that all the top hotspots cross city boundaries. It indicates that regional integration development in GBA is promoting huge spatially continuous road emission zones. Cross-border emission hotspots have been constituted between Guangzhou and Foshan, Dongguan and Shenzhen, and Foshan, Zhongshan and Jiangmen, while the contiguous discharge in the east and south directions of Guangzhou has not yet formed. Even so, the current top 50%-hotspots have gathered the majority of road CO2 emissions. Although a large amount of small hotspots exists around the top hotspots, they only contribute to a tiny proportion.

The disparity of agglomeration patterns between HDFV and non-HDFV emissions is highlighted (Figure 7(d), 7(e)). HDFV CO2 emissions are significantly less clustered than the counterpart. Even the 50%-quantile hotspots only account for 32.8% of the total HDFV emissions. The 90%-quantile HDFV hotspots mainly refer to Huangpu District in east Guangzhou, Longgang District in north Shenzhen, and Baoan District in west Shenzhen. These areas are the principal industrial zones in GBA. Comparatively, emission agglomeration of non-HDFV is far more considerable. The hierarchical pattern resembles that of the total emissions. Non-HDFV emissions are even more concentrated than the total emissions, with the three levels of hotspots covering 42.5%, 61.6% and 78.9% of the total non-HDFV emissions respectively.

With the hyper parameters consistent, the temporal variation of hierarchical emission hotspots in GBA is analyzed (Supplementary Figure S2). The patterns from 7:00 to 0:00 the next day are analogous and are basically consistent with that of the daily gross emissions. From 1:00, all levels of emission hotspots start to shrink in scale visibly and become the smallest at 4:00.

After that, emission hotspots resume spreading and finally return the full state at around 7:00. As a scale metric of emission hotspots, the percentage of clustered emissions records the same process of variation in a quantitative way (Supplementary Figure S2(b)). These curves have a similar trend with the daily gross emissions (Figure 4(a)). However, they carry different information. QH-DBSCAN reflects the relative level of emission agglomeration regardless of the absolute emissions. Thus, their remarkable consistency indicates that, in GBA, the rise of absolute emissions would promote the level of emission agglomeration.

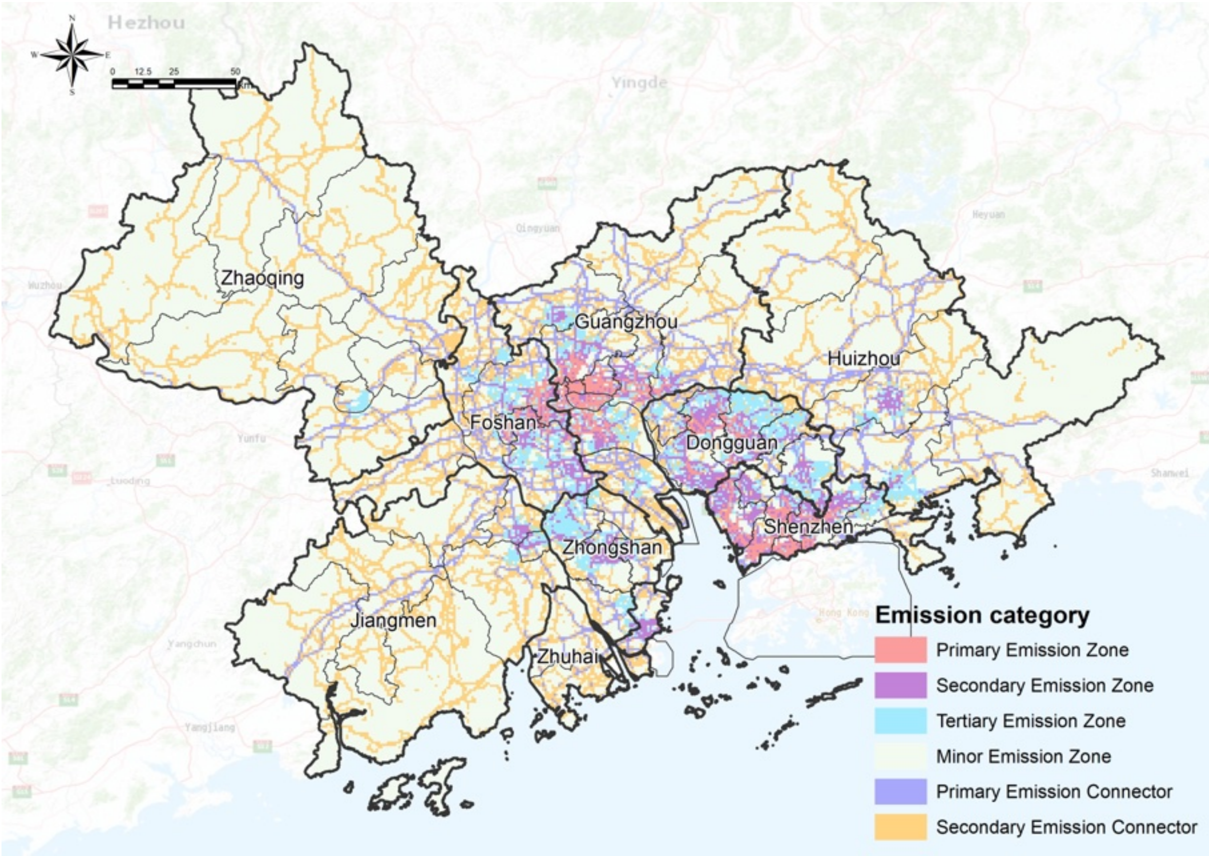
3.4. Category of emission patterns

The preceding analysis has unfolded the road CO2 emission patterns in GBA from the perspective of space, time, vehicle type and agglomeration. To gain an all-round insight from those sides, we attempt to categorize the grid-level emission patterns. The attributes used for the categorization consist of two portions. The first part indicates the overall emission quantities, including daily gross emissions, daily gross HDFV emissions, daily gross non-HDFV emissions, emission intensity and the percentage of HDFV emissions. The second one contains the cases where the grids belong to different levels of emission hotspots at different times. We use hotspot level rather than the absolute emissions to guide the generation of spatially continuous emission categories, which is conductive to policy making. Since emission hotspots do not vary drastically in consecutive hours, we use typical hours to represent the entire day. Focusing on representative hours alleviates the multicollinearity among attributes and is instrumental to interpret the results. Specifically, we choose five hours, that is the emission valley hour (4:00), the two emission peak hours (8:00 and 18:00), and two transitional hours (12:00 and 0:00). The hierarchical emission hotspots at each selected hour are converted into three dummy variables, indicating whether the grid belongs to each level of hotspot. As a result, the process gives rise to 15 dummies. In summary, the attribute set contains 20 explanatory variables. We employ K-Means clustering algorithm (Hartigan and Wong, 1979) to categorize the emission patterns. All the non-dummy attributes are normalized before entering the model. We experiment among Z-score, min-max scaling and inverse hyperbolic sine function, and find that the last one fits the best because it is capable of dealing with zero-inflated attribute sets as ours. To determine the optimal hyper-parameter k (number of clusters), we conduct assessments with multiple methods including elbow method (Bholowalia and Kumar, 2014), Silhouette score (Rousseeuw, 1987), and gap statistics (Tibshirani et al., 2001). After testing with 100 random seeds, the optimal k is set to be 6. Consequently, we obtain the spatial distribution of the six emission categories in GBA (Figure 8(a)). There are two main geographical forms for the emission categories. Among the six categories, four are in zone form, whilst two are in linear form. According to the geographical form and the emission intensity, we define the zone categories as Primary, Secondary, Tertiary and Minor Emission Zone, and name the linear categories as Primary and Secondary Emission Connector. The differences in emission metrics among the categories are demonstrated in Table 4.

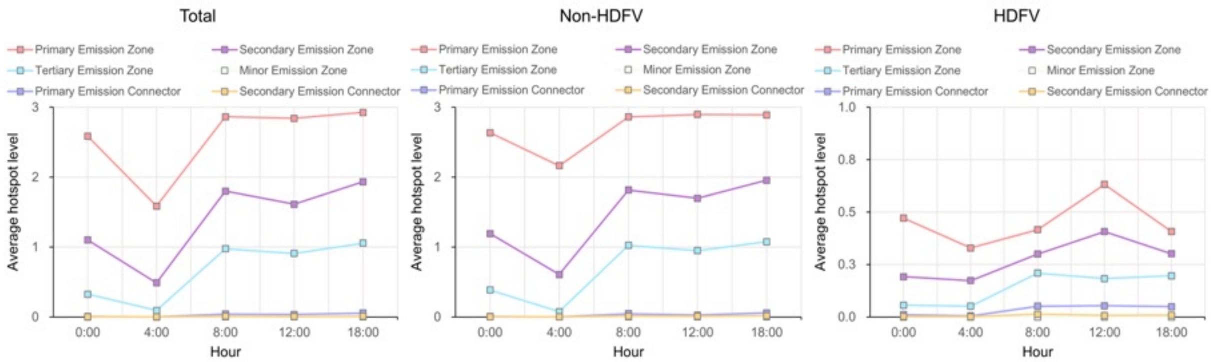
Table 4: Differences in road CO2 emissions metrics among emission categories.

Category	Population (million)	Daily gross emissions (t)	Per capita emissions (kg/person)	Emission intensity (kg/m)	Non-HDFV emissions (t)	HDFV emissions (t)	HDFV percentage (%)
Primary Emission Zone	16.46	17,002	1.03	1.63	13,850	3152	18.5
Secondary Emission Zone	16.29	12,808	0.79	0.80	8897	3911	30.5
Tertiary Emission Zone	12.46	5333	0.43	0.34	3385	1948	36.5
Minor Emission Zone	11.22	344	0.03	0.04	283	61	17.7
Primary Emission Connector	5.73	15,890	2.77	1.03	7704	8186	51.5
Secondary Emission Connector	9.88	4864	0.49	0.24	1761	3103	63.8

The Primary, Secondary and Tertiary Emission Zone overlap with the urban area in GBA. The Primary Emission Zone includes the Guangzhou-Foshan metropolis and the main cities of Shenzhen and Dongguan, which are the most developed area in GBA, accommodating 16.46 million residents according to the WorldPop (WorldPop, 2018). It not only generates the largest total emissions (17002 t), but also possesses the topmost emission intensity (1.63 kg/m) and per capita emission (1.03 kg/person). In Primary Emission Zone, non-HDFVs are responsible for the majority of emissions, with only 18.5% of the total emissions from HDFVs. The Secondary Emission Zone points to the satellite towns around Guangzhou, Shenzhen, Dongguan and Foshan, and also the major cities of Zhongshan, Huizhou, Jiangmen and Zhuhai. The total population is close to the Primary Emission Zone, but less road CO2 emissions are produced (12808 t) because of the lower per capita emissions (0.79 kg/person) and emission intensity (0.80 kg/m). The Tertiary Emission Zone contains more outer suburbs around the Primary and Secondary Emission Zone. With 12.46 million residents in the zone, the zone products 5333t road CO2 emissions per day. The per capita emissions (0.43 kg/person) and emission intensity (0.34 kg/m) decline further. In the meantime, HDFVs account for a higher proportion of the total emissions (36.5%), which nearly doubles that of the Primary Emission Zone. The findings unveil that population size may not be the decisive factor of road CO2 emissions. In fact,



(a)



(b)

Figure 8: Emission Category in GBA. **(a)** Spatial distribution of road CO2 emission categories in GBA. **(b)** Temporal variations of average emission hotspot level. Source of the basemap: (c) Esri.

some zones emit surplus emissions because of a higher unit average emission instead of excessive population. Besides, the emission addition is mainly triggered by non-HDFVs, implying a possibly more emission-intensive mode of mobility in those high-emission zones. The broad rural and natural area in GBA is mostly demarcated into the Minor Emission Zone. Road CO2 emissions in this zone are comparatively negligible.

By contrast with zones, the rest two emission categories are characterized by a spatially linear distribution along the road, especially the regional connector (highway and provincial expressway). The Primary Emission Connector mainly consists of highways that radiate from the emission-intensive zones. These connectors produce 15890 t road CO2 a day, which is second only to the volume of Primary Emission Zone. This emission category involves the least residents and has the highest per capita emissions (2.77 kg/person). Most emissions are related to inter-city long-distance travel or freight transportation. HDFVs contribute to 51.5% of the total emissions, which is considerably larger than those of the emission zones. The Secondary Emission Connector refers to the provincial expressways that primarily serve intra-city medium-distance mobility. The daily gross emissions and the emission intensity are much lower than the Primary Emission Connector. However, the HDFV percentage (63.8%) is the highest among all the categories. On emission connectors, HDFVs appear to be the dominant emitter, with 55.4% of the daily gross HDFV emissions concentrating on Primary and Secondary Emission Connector.

To further interpret the differences in the emission hotspot level among emission categories across the time, we conduct the following analysis. For grids included in multiple levels of emission hotspots, the highest level is taken. For each grid, a numerical score is used to represent the highest hotspot level (90%-quantile hotspot— 3, 75%-quantile hotspot— 2, 50%-quantile hotspot — 1, no hotspot — 0). For each category, its average emission hotspot level is reflected by the mean score across the grids. In this way, we obtain the temporal variations of the average hotspot level in Figure 8(b). Regarding the total emissions, only the Primary, Secondary and Tertiary Emission Zone exhibit significant emission agglomeration. Grids in the Primary Emission Zone mostly belong to 90%-quantile hotspots from 8:00 to 18:00, while one level of downgrade occurs at 4:00. Similar trend is also observed for the Secondary and Tertiary Emission Zone with lower overall levels. The average hotspot level of Non-HDFV emissions has analogous characteristics with the total emissions. In contrast, HDFV emission agglomeration is less considerable across all the categories, with average hotspot levels lower than 1 during the day.

4. Discussion

4.1. Validation of emission estimation

Table 5: Comparison of estimation results between this study and the literature.

Study area	Road type	Study	Year	Data Source	Approach	CO2 Emissions (Mt/year)
Guangdong Province	All	Jia et al. (2018)	2011	Statistical Yearbooks	Bottom-up	84.5
		Jia et al. (2018)	2012	Statistical Yearbooks	Bottom-up	94.1
		Guo et al. (2014)	2012	Statistical Yearbooks	Top-down	56.82
		Jia et al. (2018)	2013	Statistical Yearbooks	Bottom-up	101.9
		Jia et al. (2018)	2014	Statistical Yearbooks	Bottom-up	109.4
		Jia et al. (2018)	2015	Statistical Yearbooks	Bottom-up	112.5
		Crippa et al. (2021)	2018	Energy Balances	Top-down	90.9
		Guan et al. (2021)	2019	Statistical Yearbooks	Top-down	61.17
		Xu et al. (2021)	2019	Statistical Yearbooks	Bottom-up	50.1
		Gao et al. (2022)	2020	China's continuous emissions monitoring systems (CEMS)	Top-down	70.15
		Liu et al. (2022)	2021	EDGAR transport emissions	Top-down	80.3
		This study	2020	Vehicle trajectories	Bottom-up	**27.28**
Guangdong Province	Highway	Li et al. (2022)	2021	Highway toll data	Bottom-up	5.59
		This study	2020	Vehicle trajectories	Bottom-up	**11.26**
Futian and Nanshan, Shenzhen	All	Zhou et al. (2022b)	2017	Vehicle trajectories	Bottom-up	0.03
		This study	2020	Vehicle trajectories	Bottom-up	**0.84**

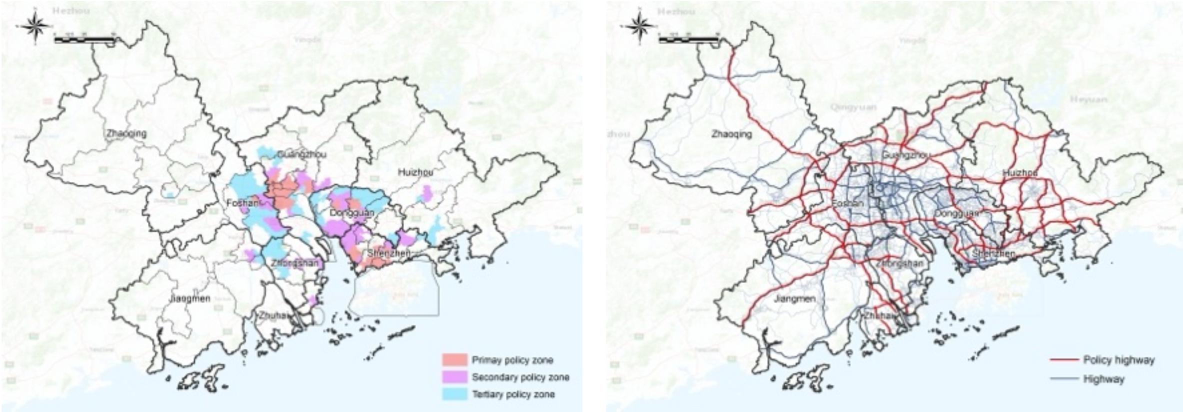
Emission estimation is always vulnerable to the method and data source used. To validate our results and discover the pros and cons, we compare our results with other studies using different datasets and methods (Table 5). We make necessary conversions on our results to guarantee the consistency of study scenario (year, study area, road type and emission type) with previous studies and ensure the comparability. It turns out that road CO2 emission estimates vary significantly across studies. Top-down and bottom-up approach can lead to very different results using similar data source. Besides, data source also has a profound impact. Estimates based on trajectory data are commonly lower than those based on collective data. As the first study that uses vehicle trajectories to estimate daily gross road CO2 emissions in Guangdong, our estimates are considerably lower than those obtained from collective data such as China Statistical Yearbook and Energy Balances. However, it cannot be assumed that the sampling rate of our dataset is overrated, because our study yields higher emission estimates than other studies using similar dataset such as highway toll data (Li et al., 2022) or vehicle trajectories (Zhou et al., 2022b) at the same time. Li et al. (2022) harness full-sample O-D pairs between highway tolls and route simulations rather than real trajectories to estimate highway CO_2 emissions. Their lower estimates may root from overlooking accidents and detours that often happen in real life. Zhou et al. (2022b) use real trajectories in 2017 with limited samples. Considering the three-year time difference and the fact that our dataset contains over 16 times as many samples as theirs in the same area, the difference in results is generally acceptable. Even though most inputs are fixed, uncertainty of estimates can also be introduced by different empirical schemes adopted to determine the emission factor, vehicle kilometer traveled, fuel consumption, etc.. Therefore, in general, the estimates of this study lie in an appropriate range, and they compensate the lack of road carbon emission estimation in Guangdong using individual level activity data.

4.2. Policy implications

Based on the findings above, we intend to recommend some targeted measures to mitigate road CO_2 emissions in our study area, especially in GBA.

Several universal strategies can be widely applied. We outline the necessity of periodically supervising fine-grained regional road carbon emissions with intelligent systems and all sorts of big data approaches (Wang et al., 2022a). Our study suggests that special attention should be given to the core area of Guangzhou-Foshan metropolis and Shenzhen where emissions are the most intensive and clustered. The continuously updated database lays the foundation for downstream applications, such as emission pattern analysis and emission reduction policy making. This study enriches the technical toolkits for both the supervision side and the application side, by proposing methods to handle trajectory data with deficiency in time granularity, identify emission hotspots, and categorize emission zones. Besides, our research confirms that raising vehicle emission standards and promoting new energy vehicles is instrumental to reduce overall road carbon emissions (Zhang et al., 2022). HDFVs are found to be associated with much more intensive emissions per vehicle possession than non-HDFVs (Supplementary Table S8). Thus, HDFVs, especially those with higher accumulated mileage (Wang et al., 2022b), should be prioritized for emission standard upgrade or electrification.

Furthermore, considering that resources are limited and extensive policies may cause negative externalities, emission reduction policies should be adaptive to local emission patterns (Cai et al., 2018). To achieve that, we introduce policy zones where targeted measures are proposed,



(a) (b)

Figure 9: Maps for policy implications. **(a)** Spatial distribution of emission policy zones **(b)** Spatial distribution of policy highways.

based on the spatial distribution of emission categories. Our methodology generates spatially continuous regions with similar emission patterns, which is inductive to defining policy zones. To further facilitate policy implementation, we use street (the smallest administrative division in China) as the unit to divide policy zones. The emission category of a street is determined by the emission category of the grids covering the largest area within the street. In line with the three major emission zones (Figure 8(a), Table 4), we define the Primary, Secondary and Tertiary Policy Zone (Figure 9(a)). The spatial distribution of emission zones and policy zones are similar. Primary Emission Zone is characterized by the highest emission intensity but low proportions from HDFVs. The measures should accordingly focus on mitigating emissions from daily travel. Considering that areas within Primary Emission Zone are equipped with comparatively more advanced public transportation resources in GBA, we suggest encouraging sustainability-oriented travel mode choice from two directions. First, more people should be guided to use the public transits by improving the competitiveness of the system. For example, the transportation department could optimize the bus lines and subway operations within this zone to achieve more volume and better accessibility. Second, government could toll extra emission taxes on personal vehicles driving in the Primary Policy Zone. Under our estimation regarding road CO2 emission reduction, converting residents in the Primary Policy Zone to public transits is five times more efficient than converting the same number of random people in GBA. The Secondary Policy Zone indicates the streets with most of their area categorized as Secondary Emission Zone whose emission intensity halved but proportions from HDFVs doubled compared with the Primary Emission Zone. Although some level of off-peak HDFV delivery has been observed (Figure 4(c), 4(d)), there are still apparent overlaps of high emissions from HDFVs and non-HDFVs during the period between the morning and evening peaks. Therefore, we recommend enacting incentive policies such as midnight toll discount to guide HDFVs to reschedule their itinerary and operate during low-emission hours (22:00 - 6:00). Promoting off-peak delivery could alleviate the emission surplus caused by traffic congestion at peak hours (Chen et al., 2022). Last, Tertiary Policy Zone mainly indicates the marginal streets around the major urban areas. Although both the local emission volume and intensity are much less, the nonnegligible population size and the geographic adjacency with the major urban areas may mean heavy external traffics. We would recommend more detailed future studies to uncover the pattern of emissions from traffics between these streets and the major urban areas. Some measures might be necessary to reduce emission surplus from long-distance vehicle travels.

The emission category, Primary Emission Connector, mainly consists of the inter-city highways with intensive emissions and a high proportion from HDFVs. We highlight these highways as the “policy highway”, which should be prioritized for carbon reduction policies (Figure 9(b)). We advocate promoting electrified HDFVs to run the transportation routes on these highways. Some policy and infrastructure facilities could be provided on these highways for electrified HDFVs, such as more convenient and affordable charging services and lower passing tolls. Also, we note that many of the policy highways are those radiating from Guangzhou and Shenzhen. Thus, another strategy could be strengthening the construction and integration of regional freight railway infrastructure around these two cities, to divert corresponding transportation needs from roads to railways.

4.3. Limitations and future studies

The findings of this study have to be seen in light of the following limitations. The vehicle trajectory data used in our study is at minute-level granularity, which may still induce uncertainty in traffic status and CO2 emission estimation, although some corrections are implemented. Besides, due to limited data availability, this analysis is conducted with data in a single day. Panel data should be helpful to examine the emission fluctuation in a longer time span. Since the vehicle type, fuel type and emission standard of each trajectory are unavailable, these factors are considered to be constant across all the road segments. We suggest using more advanced data sources and local emission model to validate our results. Besides, our results identify a considerably higher proportion of HDFV CO2 emissions in non-GBA cities, which might be partly attributed to the possibility that drivers in these cities count less on navigation during daily travel. It means the possible spatial heterogeneity in the sampling rates, which may influence trajectory-based emission estimation but remains under-explored in the literature.

This study obtains emission category zones in GBA and accordingly divides policy zones with targeted carbon reduction strategies. We are aware that policy making is a rigorous and complicated process where multiple aspects should be considered. Due to space limitations, this study focuses more on emission pattern analysis, and thus does not involve other perspectives such as demography, land use, finance, and so on. A more detailed depiction on the demographic structure in each emission category zone may help to understand the status quo better and formulate more solid strategies (Deng et al., 2021; Zhang et al., 2021a). Other indicators regarding urban form (Shi et al., 2020; Wang et al., 2017; Chen et al., 2021; Biljecki and Chow, 2022) and built environment (Cao and Yang, 2017) could also be taken into consideration. Furthermore, distinguishing the emission patterns of different vehicle usages and trip purposes (Zhao et al., 2017) is also instrumental to deepen the understanding of road emissions. Our follow-up research would be intended to remedy these deficiencies.

5. Conclusions

Based on massive vehicle trajectories, this study demonstrates the spatiotemporal pattern of road CO2 emissions in Guangdong. Emissions are estimated per road segment and per hour, with both broad spatial coverage and fine granularity. Overall, GBA produces 76% of the total emissions with only 30% of land in Guangdong. Guangzhou is the primary emitter in Guangdong, producing 15621 t CO2 and contributing to 21% of the provincial emissions. Shenzhen has the highest average emission intensity (1.01 kg/m). Most GBA cities rank high in road CO2 emissions except for Zhuhai. Compared with GBA cities, non-GBA cities have twice the percentage of HDFV emissions. Regarding the road class, highway appears to generate the most emissions, accounting for 41% of the total. Road classes serving local transportation also contribute to 50% of the total emissions. Temporally, we discover a "three-stage" pattern for all the cities, while the trend of non-GBA cities is flatter. We use the contribution ratio of the top 10% emitting road segments to quantify the level of emission agglomeration. Top 10% emitting roads discharge 77.6% of the total emissions in Guangzhou. The contribution ratio culminates at the emission valley hour. Besides, GBA cities exhibit more significant contribution ratio.

We propose QH-DBSCAN to detect hierarchical emission hotspots with emissions projected onto 1 km grids in GBA, identifying the location and scope of emission hotspots. Then we demonstrate a comprehensive sensitivity analysis to examine the influence of hyper-parameter selection for the method. The results show that the 90%-quantile hotspots emit 29.6% of the daily gross road CO2 emissions with only 5% of land in GBA. The associated regions refer to the Guangzhou-Foshan metropolis and Shenzhen center. The 50%-quantile emission hotspots cover most urban areas in GBA, presenting cross-border integration development between adjacent cities represented by Guangzhou and Foshan, and Shenzhen and Dongguan. HDFV CO2 emissions show less significant agglomeration than the non-HDFV counterpart. The temporal variation of emission hotspots is highly synchronized with the daily gross emissions. Different interactions among the three levels of emission hotspots are found in high-emission and low-emission periods.

We aggregate all aspects of emission features and derive six emission categories, including four emission zones and two emission connectors. The Primary Emission Zone refers to the most developed urban area in GBA, which produces the largest total emissions (30% of GBA total). It has the topmost emission intensity but the lowest percentage of HDFV emissions in emission zones. The Secondary and Tertiary Emission Zone emit less intensively but have a higher percentage of HDFV emissions. The Primary Emission Connector mainly contains the inter-city highways radiating from the emission-intensive zones, with the highest emission intensity and over 50% of emissions contributed by HDFVs. On the ground of the differences of emission patterns among emission categories, we propose policy zones and recommend targeted strategies for carbon reduction.

In summary, the study demonstrate a bottom-up road CO2 estimation method based on a massive vehicle trajectory dataset. The results unveil the spatiotemporal pattern of road CO2 emission in Guangdong and emission agglomeration in GBA. This study expands the toolkit of regional emission studies and offers insightful findings that support the carbon reduction policy making and sustainable development in Guangdong and especially GBA.

Funding

This research was supported by Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning (No.2020B121202019) and The Science and Technology Foundation of Guangzhou Urban Planning & Design Survey Research Institute (RDI2220205141).

CRediT authorship contribution statement

Xingdong Deng: Conceptualization, Methodology, Funding acquisition, Writing - Original draft preparation. **Wangyang Chen:** Writing- Original draft preparation, Writing - Review & Editing, Software. **Qingya Zhou:** Writing - Review & Editing, Investigation, Validation, Supervision. **Yuming Zheng:** Writing- Original draft preparation, Formal analysis, Visualization. **Hongbao Li:** Writing - Review & Editing, Supervision. **Shunyi Liao:** Resources, Project administration. **Filip Biljecki:** Conceptualization, Writing - Review & Editing.

Declarations of interest

None

Acknowledgments

We sincerely appreciate the kindly help from the editors and reviewers to improve this study. We also acknowledge the data sources and the open-source packages used in this study.

References

Batur, İ., Bayram, I.S., Koc, M., 2019. Impact assessment of supply-side and demand-side policies on energy consumption and co2 emissions from urban passenger transportation: The case of istanbul. Journal of Cleaner Production 219, 391–410.

Bholowalia, P., Kumar, A., 2014. Ebk-means: A clustering technique based on elbow method and k-means in wsn. International Journal of Computer Applications 105.

Biljecki, F., Chow, Y.S., 2022. Global Building Morphology Indicators. Computers, Environment and Urban Systems 95, 101809.

Böhm, M., Nanni, M., Pappalardo, L., 2022. Gross polluters and vehicle emissions reduction. Nature Sustainability , 1–9.

Boulter, P., Barlow, T., McCrae, I., Latham, S., 2009. Emission factors 2009: Final summary report. TRL Published Project Report .

Cai, B., Guo, H., Cao, L., Guan, D., Bai, H., 2018. Local strategies for china's carbon mitigation: An investigation of chinese city-level co2 emissions. Journal of Cleaner Production 178, 890–902.

Cao, X., Yang, W., 2017. Examining the effects of the built environment and residential self-selection on commuting trips and the related co2 emissions: An empirical study in guangzhou, china. Transportation Research Part D: Transport and Environment 52, 480–494.

Carslaw, D.C., Goodman, P.S., Lai, F.C., Carsten, O.M., 2010. Comprehensive analysis of the carbon impacts of vehicle intelligent speed control. Atmospheric Environment 44, 2674–2680.

Chen, L., Xu, L., Yang, Z., 2017. Accounting carbon emission changes under regional industrial transfer in an urban agglomeration in china's pearl river delta. Journal of Cleaner Production 167, 110–119.

Chen, W., Wu, A.N., Biljecki, F., 2021. Classification of urban morphology with deep learning: Application on urban vitality. Computers, Environment and Urban Systems 90, 101706.

Chen, X., Jiang, L., Xia, Y., Wang, L., Ye, J., Hou, T., Zhang, Y., Li, M., Li, Z., Song, Z., et al., 2022. Quantifying on-road vehicle emissions during traffic congestion using updated emission factors of light-duty gasoline vehicles and real-world traffic monitoring big data. Science of the Total Environment 847, 157581.

Chen, Z., Li, Y., Wang, P., 2020. Transportation accessibility and regional growth in the greater bay area of china. Transportation Research Part D: Transport and Environment 86, 102453.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Lo Vullo, E., Solazzo, E., Monforti-Ferrario, F., Olivier, J., Vignati, E.E., 2021. v6. 0 global greenhouse gas emissions. European Commission, Joint Research Centre (JRC): Vienna, Austria .

Deng, F., Lv, Z., Qi, L., Wang, X., Shi, M., Liu, H., 2020. A big data approach to improving the vehicle emission inventory in china. Nature communications 11, 1–12.

Deng, X., Liu, Y., Gao, F., Liao, S., Zhou, F., Cai, G., 2021. Spatial distribution and mechanism of urban occupation mixture in guangzhou: An optimized geodetector-based index to compare individual and interactive effects. ISPRS International Journal of Geo-Information 10, 659.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise., in: kdd, pp. 226–231.

Fang, C., Yu, D., 2017. Urban agglomeration: An evolving concept of an emerging phenomenon. Landscape and urban planning 162, 126–136.

Fujita, M., Thisse, J.F., 1996. Economics of agglomeration. Journal of the Japanese and international economies 10, 339–378.

Gao, Y., Zhang, L., Huang, A., Kou, W., Bo, X., Cai, B., Qu, J., 2022. Unveiling the spatial and sectoral characteristics of a high-resolution emission inventory of co2 and air pollutants in china. Science of The Total Environment 847, 157623.

Guan, Y., Shan, Y., Huang, Q., Chen, H., Wang, D., Hubacek, K., 2021. Assessment to china's recent emission pattern shifts. Earth's Future 9, e2021EF002241.

Guangdong Province Statistical Bureau, 2021. Guangdong Statistical Yearbook: 2021. China Statistics Publishers.

Guo, B., Geng, Y., Franke, B., Hao, H., Liu, Y., Chiu, A., 2014. Uncovering china's transport co2 emission patterns at the regional level. Energy Policy 74, 134–146.

Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics) 28, 100–108.

Hicks, W., Beevers, S., Tremper, A.H., Stewart, G., Priestman, M., Kelly, F.J., Lanoisellé, M., Lowry, D., Green, D.C., 2021. Quantification of non-exhaust particulate matter traffic emissions and the impact of covid-19 lockdown at london marylebone road. Atmosphere 12, 190.

Huang, L., Krigsvoll, G., Johansen, F., Liu, Y., Zhang, X., 2018. Carbon emission of global construction sector. Renewable and Sustainable Energy Reviews 81, 1906–1916.

Hui, E.C., Li, X., Chen, T., Lang, W., 2020. Deciphering the spatial structure of china's megacity region: A new bay area—the guangdong-hong kong-macao greater bay area in the making. Cities 105, 102168.

IEA, G.E., 2019. Co2 emissions from fuel combustion. International Energy Agency, Paris .

IEA, G.E., 2020. Co2 emissions from fuel combustion. International Energy Agency, Paris .

Jia, T., Li, Q., Shi, W., 2018. Estimation and analysis of emissions from on-road vehicles in mainland china for the period 2011–2015. Atmospheric Environment 191, 500–512.

Kan, Z., Tang, L., Kwan, M.P., Ren, C., Liu, D., Pei, T., Liu, Y., Deng, M., Li, Q., 2018. Fine-grained analysis on fuel-consumption and emission from vehicles trace. Journal of cleaner production 203, 340–352.

Li, T., Wu, J., Dang, A., Liao, L., Xu, M., 2019a. Emission pattern mining based on taxi trajectory data in beijing. Journal of cleaner production 206, 688–700.

Li, Y., Wu, Q., Zhang, Y., Huang, G., Jin, S., Fang, S., 2022. Mapping highway mobile carbon source emissions using traffic flow big data: A case study of guangdong province, china. Frontiers in Energy Research , 496.

Li, Y., Zheng, J., Dong, S., Wen, X., Jin, X., Zhang, L., Peng, X., 2019b. Temporal variations of local traffic co2 emissions and its relationship with co2 flux in beijing, china. Transportation Research Part D: Transport and Environment 67, 1–15.

Lin, B., Li, Z., 2020. Spatial analysis of mainland cities' carbon emissions of and around guangdong-hong kong-macao greater bay area. Sustainable Cities and Society 61, 102299.

Liu, B., Zimmerman, N., 2021. Fleet-based vehicle emission factors using low-cost sensors: Case study in parking garages. Transportation Research Part D: Transport and Environment 91, 102635.

Liu, Z., Deng, Z., Zhu, B., Ciais, P., Davis, S.J., Tan, J., Andrew, R.M., Boucher, O., Arous, S.B., Canadell, J.G., et al., 2022. Global patterns of daily co2 emissions reductions in the first year of covid-19. Nature Geoscience 15, 615–620.

Lomas, K.J., Bell, M., Firth, S., Gaston, K.J., Goodman, P., Leake, J.R., Namdeo, A., Rylatt, M., Allinson, D., Davies, Z.G., et al., 2010. 4m: Measurement, modelling, mapping and management–the carbon footprint of uk cities .

Malmberg, A., Maskell, P., 1997. Towards an explanation of regional specialization and industry agglomeration. European planning studies 5, 25–41.

Manzoni, V., Maniloff, D., Kloeckl, K., Ratti, C., 2010. Transportation mode identification and real-time co2 emission estimation using smartphones. SENSEable City Lab, Massachusetts Institute of Technology, nd .

McQueen, M., MacArthur, J., Cherry, C., 2020. The e-bike potential: Estimating regional e-bike impacts on greenhouse gas emissions. Transportation Research Part D: Transport and Environment 87, 102482.

MEE, 2019. China Mobile Source Environmental Management Annual Report (2019). Ministry of Ecology and Environment of the People's Republic of China.

MEE, 2021. China Mobile Source Environmental Management Annual Report (2021). Ministry of Ecology and Environment of the People's Republic of China.

Mohsin, M., Abbas, Q., Zhang, J., Ikram, M., Iqbal, N., 2019. Integrated effect of energy consumption, economic development, and population growth on co 2 based environmental degradation: a case of transport sector. Environmental Science and Pollution Research 26, 32824–32835.

Patiño-Aroca, M., Parra, A., Borge, R., 2022. On-road vehicle emission inventory and its spatial and temporal distribution in the city of guayaquil, ecuador. Science of The Total Environment 848, 157664.

Pérez-Martínez, P., Miranda, R., Andrade, M., 2020. Freight road transport analysis in the metro são paulo: Logistical activities and co2 emissions. Transportation Research Part A: Policy and Practice 137, 16–33.

Pla, M.A.M., Lorenzo-Sáez, E., Luzuriaga, J.E., Prats, S.M., Moreno-Pérez, J.A., Urchueguía, J.F., Oliver-Villanueva, J.V., Lemus, L.G., 2021. From traffic data to ghg emissions: A novel bottom-up methodology and its application to valencia city. Sustainable Cities and Society 66, 102643.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65.

Shan, Y., Fang, S., Cai, B., Zhou, Y., Li, D., Feng, K., Hubacek, K., 2021. Chinese cities exhibit varying degrees of decoupling of economic growth and co2 emissions between 2005 and 2015. One Earth 4, 124–134.

Shi, K., Xu, T., Li, Y., Chen, Z., Gong, W., Wu, J., Yu, B., 2020. Effects of urban forms on co2 emissions in china from a multi-perspective analysis. Journal of environmental management 262, 110300.

Shindell, D.T., Levy, H., Schwarzkopf, M.D., Horowitz, L.W., Lamarque, J.F., Faluvegi, G., 2008. Multimodel projections of climate change from short-lived emissions due to human activities. Journal of Geophysical Research: Atmospheres 113.

Sobrino, N., Arce, R., 2021. Understanding per-trip commuting co2 emissions: A case study of the technical university of madrid. Transportation Research Part D: Transport and Environment 96, 102895.

Sui, Y., Zhang, H., Song, X., Shao, F., Yu, X., Shibasaki, R., Sun, R., Yuan, M., Wang, C., Li, S., et al., 2019. Gps data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. Journal of Cleaner Production 227, 495–505.

Sun, Z., Hao, P., Ban, X.J., Yang, D., 2015. Trajectory-based vehicle energy/emissions estimation for signalized arterials using mobile sensing data. Transportation Research Part D: Transport and Environment 34, 27–40.

Tatem, A.J., 2017. Worldpop, open data for spatial demography. Scientific data 4, 1–4.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, 411–423.

Van Fan, Y., Perry, S., Klemeš, J.J., Lee, C.T., 2018. A review on air emissions assessment: Transportation. Journal of cleaner production 194, 673–684.

Wang, F., Wang, G., Liu, J., Chen, H., 2019. How does urbanization affect carbon emission intensity under a hierarchical nesting structure? empirical research on the china yangtze river delta urban agglomeration. Environmental Science and Pollution Research 26, 31770–31785.

Wang, J., Cai, B., Zhang, L., Cao, D., Liu, L., Zhou, Y., Zhang, Z., Xue, W., 2014. High resolution carbon dioxide emission gridded data for china derived from point sources. Environmental science & technology 48, 7085–7093.

Wang, L., Chen, X., Xia, Y., Jiang, L., Ye, J., Hou, T., Wang, L., Zhang, Y., Li, M., Li, Z., et al., 2022a. Operational data-driven intelligent modelling and visualization system for real-world, on-road vehicle emissions—a case study in hangzhou city, china. Sustainability 14, 5434.

Wang, M., Madden, M., Liu, X., 2017. Exploring the relationship between urban forms and co 2 emissions in 104 chinese cities. Journal of Urban Planning and Development 143, 04017014.

Wang, Y., Xing, Z., Zhang, H., Wang, Y., Du, K., 2022b. On-road mileage-based emission factors of gaseous pollutants from bi-fuel taxi fleets in china: The influence of fuel type, vehicle speed, and accumulated mileage. Science of The Total Environment 819, 151999.

WorldPop, 2018. Global 1km population. doi:10.5258/SOTON/WP00647.

Xia, C., Xiang, M., Fang, K., Li, Y., Ye, Y., Shi, Z., Liu, J., 2020. Spatial-temporal distribution of carbon emissions by daily travel and its response to urban form: A case study of hangzhou, china. Journal of Cleaner Production 257, 120797.

Xu, Y., Liu, Z., Xue, W., Yan, G., Shi, X., Zhao, D., Zhang, Y., Lei, Y., Wang, J., 2021. Identification of on-road vehicle co2 emission pattern in china: A study based on a high-resolution emission inventory. Resources, Conservation and Recycling 175, 105891.

Yan, D., Lei, Y., Li, L., Song, W., 2017. Carbon emission efficiency and spatial clustering analyses in china's thermal power industry: Evidence from the provincial level. Journal of Cleaner Production 156, 518–527.

Yang, W., Li, T., Cao, X., 2015. Examining the impacts of socio-economic factors, urban form and transportation development on co2 emissions from transportation in china: a panel data analysis of china's provinces. Habitat International 49, 212–220.

Yu, H., 2019. The guangdong-hong kong-macau greater bay area in the making: development plan and challenges. Cambridge Review of International Affairs , 1–29.

Yu, X., Wu, Z., Zheng, H., Li, M., Tan, T., 2020. How urban agglomeration improve the emission efficiency? a spatial econometric analysis of the yangtze river delta urban agglomeration in china. Journal of environmental management 260, 110061.

Zhang, J., Jia, R., Yang, H., Dong, K., 2022. Does electric vehicle promotion in the public sector contribute to urban transport carbon emissions reduction? Transport Policy 125, 151–163.

Zhang, X., Gao, F., Liao, S., Zhou, F., Cai, G., Li, S., 2021a. Portraying citizens' occupations and assessing urban occupation mixture with mobile phone data: A novel spatiotemporal analytical framework. ISPRS International Journal of Geo-Information 10, 392.

Zhang, X., Shen, J., Gao, X., 2021b. Towards a comprehensive understanding of intercity cooperation in china's city-regionalization: A comparative study of shenzhen-hong kong and guangzhou-foshan city groups. Land Use Policy 103, 105339.

Zhang, Y.J., Da, Y.B., 2015. The decomposition of energy-related carbon emission and its decoupling with economic growth in china. Renewable and Sustainable Energy Reviews 41, 1255–1266.

Zhang, Y.J., Liu, Z., Zhang, H., Tan, T.D., 2014. The impact of economic growth, industrial structure and urbanization on carbon emission intensity in china. Natural hazards 73, 579–595.

Zhao, P., Kwan, M.P., Qin, K., 2017. Uncovering the spatiotemporal patterns of co2 emissions by taxis based on individuals' daily travel. Journal of Transport Geography 62, 122–135.

Zhou, X., Bai, L., Bai, J., Tian, Y., Li, W., 2022a. Scenario prediction and critical factors of co2 emissions in the pearl river delta: A regional imbalanced development perspective. Urban Climate 44, 101226.

Zhou, X., Wang, H., Huang, Z., Bao, Y., Zhou, G., Liu, Y., 2022b. Identifying spatiotemporal characteristics and driving factors for road traffic co2 emissions. Science of The Total Environment 834, 155270.

Zhou, Y., Shan, Y., Liu, G., Guan, D., 2018. Emissions and low-carbon development in guangdong-hong kong-macao greater bay area cities and their surroundings. Applied energy 228, 1683–1692.