Fung, Pak Lun; Al-Jaghbeer, Omar; Chen, Jia; Paunu, Ville-Veikko; Vosough, Shaghayegh; Roncoli, Claudio; Järvi, Leena

A geospatial approach for dynamic on-road emission through open-access floating car data

Published in:
Environmental Research Letters

DOI:
10.1088/1748-9326/ad984d

Published: 01/01/2025

Document Version
Publisher's PDF, also known as Version of record

ENVIRONMENTAL RESEARCH LETTERS

PURPOSE-LED PUBLISHING™

LETTER • OPEN ACCESS

A geospatial approach for dynamic on-road emission through open-access floating car data

To cite this article: Pak Lun Fung *et al* 2025 *Environ. Res. Lett.* **20** 014033

IOP Publishing Environ. Res. Lett. 20 (2025) 014033 https://doi.org/10.1088/1748-9326/ad984d

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
19 June 2024

REVISED
5 November 2024

ACCEPTED FOR PUBLICATION
28 November 2024

PUBLISHED
17 December 2024

LETTER

A geospatial approach for dynamic on-road emission through open-access floating car data

Pak Lun Fung[1,2,*], Omar Al-Jaghbeer[1], Jia Chen[3], Ville-Veikko Paunu[4], Shaghayegh Vosough[5], Claudio Roncoli[5,6] and Leena Järvi[1,2]

1 Institute for Atmospheric and Earth System Research (INAR)/Physics, Faculty of Science, University of Helsinki, Helsinki 00014, Finland
2 Helsinki Institute of Sustainability Science (HELSUS), Faculty of Science, University of Helsinki, Helsinki 00014, Finland
3 Department of Electrical and Computer Engineering, Technical University of Munich, Munich 80333, Germany
4 Finnish Environment Institute (Syke), Helsinki 00790, Finland
5 Department of Built Environment, Aalto University, Espoo 02150, Finland
6 Centre for Industrial Management / Traffic and Infrastructure, KU Leuven, Leuven 3001, Belgium
* Author to whom any correspondence should be addressed.

E-mail: pak.fung@helsinki.fi

Keywords: congestion, open-access, traffic emission, GIS, spatial analysis, air pollution, climate change

Abstract

This paper presents a geospatial approach for quantifying street-level on-road emissions of carbon dioxide (CO_2), nitrogen oxides (NO_x), and carbon monoxide (CO). By leveraging an existing open-access database of real-time congestion information derived from floating car data, we tested three methods to map high-resolution dynamic traffic emissions. To demonstrate the robustness and accuracy of the methods, we showcased results for summer workdays and winter weekends in the Helsinki Metropolitan Area (HMA). The three methods employed include (1) a physics-based relation known as the macroscopic fundamental diagram, (2) a data-driven input-adaptive generalized linear model (GLM), and (3) their ensemble (ENS). These methods estimated traffic density with satisfactory accuracy (R^2 = 0.60–0.88, sMAPE = 31%–68%). Utilizing speed-dependent emission factors retrieved from a European database, the results compared favorably against the downscaled national emission inventory, particularly for CO_2 (R^2 = 0.70–0.77). Among the three methods, GLM exhibited the best overall performance in the HMA, while ENS provided a robust upscaling solution. The modeled emissions exhibited dynamic diurnal and spatial behavior, influenced by different functional road classes, fleet compositions and congestion patterns. Congestion-induced emissions were calculated to account for up to 10% of the total vehicular emissions. Furthermore, to anticipate the forthcoming transportation transformation, we calculated emission changes under scenarios with various penetration rates of connected and autonomous vehicles (CAVs) using this geospatial approach. The introduction of CAVs could result in emission reductions of 3%–14% owing to congestion improvements.

1. Introduction

Despite the increasing adoption of alternative fuels and electric vehicles (EV), road transport emissions of greenhouse gases (GHGs) and air pollutants (APs) remain substantial due to growing transport volumes [1]. In Europe, road transport was responsible for 24% of total carbon dioxide (CO_2) emissions in 2020 [1]. Additionally, road transport is a major source of APs, responsible for 37% of total nitrogen oxides (NO_x) emissions in Europe in 2020 [1], and 11% of total carbon monoxide (CO) emissions in the UK in 2021 [2]. The emissions of GHGs and APs from road transport play significant roles in climate and health impacts [3–5].

Traditionally, quantifying traffic emissions has relied on inventory-based, bottom-up approaches in a static manner with top-down generalized linear model measurements as validation [6]. One prevalent method involves deriving emissions from aggregated inventories, such as annual average daily traffic (AADT) measured by *in-situ* traffic counters [7]. However, these inventories are often generalized over long timescales (e.g. annual and seasonal averages) and broad spatial contexts (e.g. city and municipality levels). While AADT assists decision-makers in grasping overall traffic emissions, it overlooks the spatial heterogeneity and temporal dynamics of real-world street-level traffic situations, especially in densely populated cities with complex road networks [8, 9]. Real-world traffic emissions do not linearly correlate with AADT but fluctuate with driving speeds and idling time influenced by traffic congestion [8]. Therefore, understanding actual traffic activity patterns and variations in local on-road emissions resulting from unexpected traffic issues, such as congestion [10] or lockdown measures [11–14], can enhance the reliability of emission estimates. Insights from traffic reports and real-world measurements show traffic congestion would exacerbate vehicular emissions of up to 34% [15] and 4%–55% [16, 17], respectively, compared to optimal driving conditions in highly congested urban areas. Consequently, real-time traffic information becomes a valuable input for refining the spatio-temporal granularity of on-road emissions, which is essential for crafting effective local action plans to combat climate change and improve air quality in cities.

Collection of citywide dynamic traffic information is essential for enhancing traffic management and understanding congestion patterns [18]. In contrast to traditional traffic data collected by stationary traffic counters, floating car data (FCD), obtained directly from moving vehicles [19], has emerged as a crucial traffic data source due to its low cost and extensive spatial coverage [20]. Multiple navigation and map companies provide real-time traffic congestion information (as relative driving speeds in comparison to free-flow speeds) for route optimisation to alleviate congestion [18, 21]. However, these real-time data have received limited attention in terms of their environmental implications, particularly regarding congestion-induced emissions [21].

From real-time FCD, the macroscopic fundamental diagram (MFD) approach has been utilized over the whole road network to estimate traffic flow (q) or traffic density (d), which are critical components to quantify traffic emission [22–26], using driving speed (v) in both congested and uncongested scenarios. The physics-based MFD approach has been criticized for their limited applicability in urban settings and are discouraged for estimating emissions for entire cities, but rather recommended for specific well-defined major roads [27, 28] because MFDs are independent of the traffic demand [29]. In other words, applying MFDs to roads with low utility would lead to high uncertainties. Despite this, some researchers advocate for developing MFDs by calibrating them for the entire city adaptively based on the traffic volumes using simple traffic flow equations [30, 31] to deduce traffic density and the resulting emissions. While some studies have focused on major cities in China using congestion index information from navigation apps like Gaode Map and Baidu Navigation, these applications have been limited to offline use and restricted to specific regions [22–25]. In a similar vein, [26] employed crowd-sourced Google traffic congestion maps and MFDs to calculate traffic emissions in Hong Kong using their local fleet and fuel composition database.

To address the limitations of using MFDs covering various urban road types, data-driven models have been introduced. For example, traffic emissions with different congestion levels were estimated with dynamic data and daily mean car counts in Paris using a sigmoid regression function [32]. Similarly, [11, 14] utilized data from Apple and TomTom as a proxy for vehicle activity and applied linear scaling of city-scale emissions with these activity datasets. Some incorporated land use predictors, such as CORINE land cover classes and local climate zones (LCZs), into a regression framework [33–35]. This method has demonstrated promising outcomes in estimating traffic density and emissions, particularly in regions with ample training data [33, 36]. However, in areas with limited training data, such estimations from data-driven models may yield high uncertainties due to their site-specific nature and dependence of available data [37]. It is important to note that neither MFDs nor data-driven models with land use predictors are flawless in this regard, as each approach has its own strengths and weaknesses when estimating traffic density and emissions.

In this study, we develop a geospatial framework to estimate on-road emissions based on real-time traffic congestion information and FCD obtained from an open-access data provider TomTom. We test the framework by assessing the influence of real-world dynamic traffic activities on emissions (GHGs: CO_2, APs: NO_x and CO) in the Helsinki Metropolitan Area (HMA, figures B1 and B2), Finland using three methods: (1) a physics-based relation MFD, (2) a data-driven mixed-effects generalized linear regression model (GLM) with adaptive land use inputs, and (3) their ensemble model (ENS), on a summer workday and winter weekend, representing two distinctive traffic patterns as demonstration. Using the proposed framework, we analyze the emission patterns caused by congestion, and examine the potential environmental impacts of emerging transportation technologies, such as connected and autonomous vehicles (CAVs). This is the first study of its kind to employ physics-based and data-driven methods to estimate dynamic traffic emissions using various sources of traffic information in the HMA.

2. Material and methods

Figure 1 illustrates the workflow of the proposed framework to quantify dynamic on-road traffic emissions uses real-time FCD, traffic counter data,

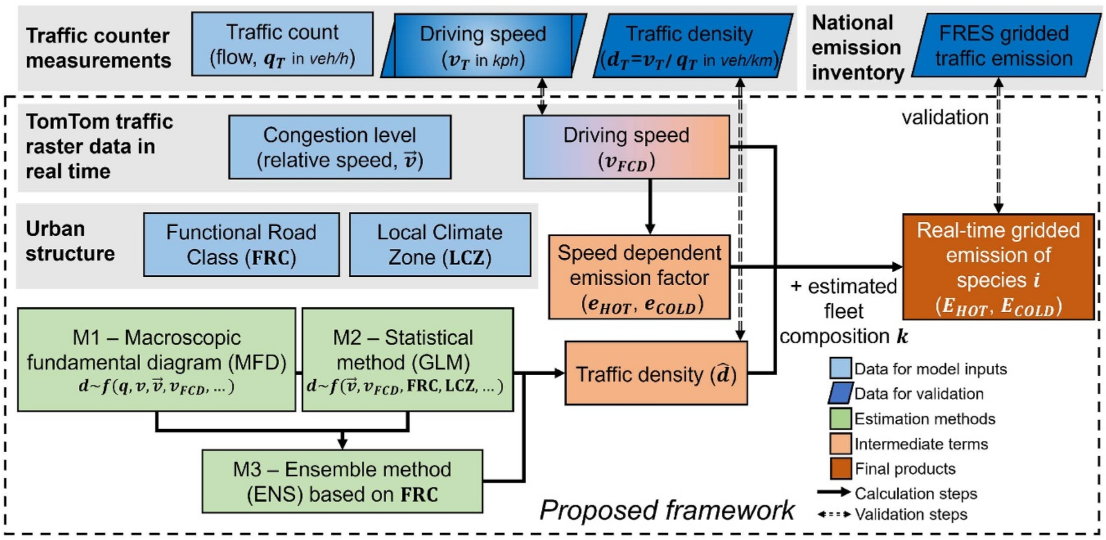

Figure 1. Workflow of the proposed framework. Blue blocks represent input materials, encompassing TomTom traffic floating car data in real time, traffic counter data, national emission inventory, and supporting materials of urban structure. Methods used are macroscopic fundamental diagram MFD, statistical method GLM, and their ensemble model ENS, which are indicated in green. Red blocks are intermediate terms required for the quantification of the final product gridded emission.

national emission inventory, and supporting materials of urban structure as inputs. Physics-based MFD, data-driven GLM, and their ensemble ENS are used to calculate traffic density, which is then used as an input to quantify gridded emissions. Two study periods, summer workdays and winter weekends, were selected. The selected periods represent their corresponding distinctive traffic patterns in the city due to factors including road conditions and social behavior.

2.1. Real-time FCD

We downloaded real-time FCD (driving speed v_{FCD} and congestion level $\vec{v}$ as relative driving speed) in raster format [19] provided by TomTom [15] with verified data quality [38] every hour for the two study periods. Raw data were originally optimized for electronic devices and the retrieval resolution was 500 pixels, which is equivalent to around 4 m at the latitude of the HMA. The data were resampled to 250 × 250 m^2 for comparison.

2.2. Traffic counter data

We collected vehicle counts (q_T) and driving speed (v_T) provided by 85 traffic counters in the HMA from Digitraffic and the City of Helsinki in summer and winter months (both workdays and weekends, figure B1). The traffic data, together with details of SLs, number of lanes, and proportion of light-duty vehicles, were aggregated into hourly format for road characterization and training our data-driven model (table B1). 86% of the traffic counters were situated along major roads or secondary roads while local roads were not well represented. We also estimated the fleet composition from the database maintained by the Finnish Transport and Communications Agency [39, 40] (table B2).

2.3. National emission inventory

In this study, annual exhaust emissions of CO_2, NO_x, and CO calculated based on the Finnish Regional Emission Scenario (FRES) model provided by the Finnish Environment Institute (Syke) at a spatial resolution of 250 × 250 m^2 for 2015 and 2030 were used [41]. The future scenario was based on the official Finnish national climate and energy strategy. The model calculated the emissions on national level based on fuel consumption by vehicle type (from statistics), vehicle fleet (based on estimates of the ages of vehicles in use), EURO emission standards, and corresponding emission factors (from the global GAINS model). The emissions were spatially distributed first to municipalities based on municipal fuel use data, and from there to grid with road network and traffic volume data. Detailed specifications of the inventory can be found in [42]. In this study, the FRES emissions were downscaled to daily resolution using traffic temporal profiles provided by Syke (figure B5) and used as a reference dataset to compare with the 24 h summed emissions using our proposed framework (see section 2.8).

2.4. Supporting datasets of urban structure

We used the standardized functional road classes (FRCs) given by TomTom [15]. The road classification is based on the character of the service they are intended to provide, and the number of vehicles detected through a network of roads. Eight classifications have been made according to their roads' functional importance in the HMA (table B3). For each FRC, we also differentiated them by SLs, resulting in 14 distinct road types.

LCZs (figure B2, table B4), which are classified based on surface structure and cover [43], have been demonstrated to strongly correlate with traffic emissions [34] and be able to explain the discrepancy between top-down and bottom-up traffic emission estimates [40].

2.5. Physics-based MFDs

MFDs consist of three different relation graphs: flow–density, speed–flow, and speed–density (q–d–v) relationship (1), which physically describes the traffic dynamics under all traffic conditions, including congested-flow regimes, free-flow regimes, and traffic breakpoints. The three graphs were derived and calibrated by plotting field data and giving these data a best fit curve for each road type [28]:

$$d = \frac{q}{v}. \tag{1}$$

We calibrated MFDs hourly using the traffic counter data collected in the HMA as ground truth (q_T and v_T). We employed a quadratic traffic model developed by Greenshields *et al* [31] (2), which has been demonstrated to perform well for free-flow conditions compared to other fundamental diagram equations in reality [26]:

$$q_\mathrm{T} = k_0 \cdot v_\mathrm{T} \cdot \left(1 - \frac{v_\mathrm{T}}{k_1}\right), \tag{2}$$

where coefficients k_0 and k_1 were updated hourly for each road type through simple statistical fittings using least-squares method. We then estimated dynamic traffic density ($\widehat{d}$) by inputting the speed obtained through TomTom (v_{FCD}):

$$\widehat{d} = k_0 \cdot \left(1 - \frac{v_\mathrm{FCD}}{k_1}\right). \tag{3}$$

2.6. Data-driven input-adpative mixed-effects GLM

Using the data from TomTom and the traffic counters, we employed a GLM taking input adaptability and mixed effects of land use predictors into account (4), which demonstrated better robustness than traditional statistical models [36, 37]:

$$g(\mu) = -\frac{1}{\mu} = \beta_0 + \sum_{k=1}^{m} \beta_k x_k + \sum_{j=1}^{n} b_j, \tag{4}$$

where $g(\mu)$ is the reciprocal link function where gamma distribution was chosen for its suitable domain for response variable $\widehat{d}$, for which we used the traffic data from the vehicle counters d_T as ground truth. β_0 denotes the fixed intercept of the equation. The second term of the equation represents the total contribution by the fixed-effects variables x ($\overrightarrow{v}$ and v_{FCD}) with a slope β. SL was initially considered but excluded due to its collinearity with other traffic-related inputs. The categorical inputs for random effects (hour of day, FRC, and LCZ) are indicated by b as intercepts of the corresponding hierarchical subgroups. The number of adaptive variables used (m and n) are dynamic to provide adaptability to address data scarcity issues as traffic counters do not cover all LCZs and road types. Standard procedures including a train-test split of 80:20 and a five-fold cross validation were followed.

2.7. Ensemble method

This ensemble ENS adaptively chose the physics-based MFD for major roads, and the data-driven GLM for non-major roads. It circumvents the limitations the two individual methods have, as suggested in [44]: MFD works better in describing the traffic dynamics of major roads [27]; GLM is site-specific due to the scarce amounts of traffic counters in the city.

2.8. Quantification of gridded emissions

Two commonly used emission models have been considered: HBEFA (the Handbook Emission Factors for Road Transport, [45]) and COPERT (the Computer Programme to calculate Emissions from Road Transport, [46]), representing 'traffic situation' and 'average-speed' model types, respectively. A comparison in a European city [47] has found that COPERT was able to provide more realistic emission estimates (lower errors despite slightly lower correlation coefficients), although the emission factors of HBEFA have been presumably considered to be more representative of real-world traffic emissions.

We employed Tier 3 formulas documented in COPERT version 5.7.1 [46] to compute traffic emissions by inputting dynamic traffic density. This includes the quantification of hot emissions (E_HOT) and cold emissions (E_COLD). In general, E_HOT were calculated from hot emission factor ($\mathrm{e_{HOT}}$), which is an integral function of mean speed distribution curves $f_k(v)$ over the emission curves $\mathrm{e}(v)$ for each emittant of interest i and vehicle class k in our modeled fleet composition (figure B3):

$$E_{\mathrm{HOT};i,k} = \widehat{d}_k \times v_{FCD;k} \times A \times \mathrm{e}_{\mathrm{HOT};i,k}, \tag{5}$$

$$\mathrm{e_{HOT};i,k} = \int [\mathrm{e}(v) \times f_k(v)]\,\mathrm{d}v. \tag{6}$$

E_COLD of NO_x and CO were calculated and added to their total emissions using a cold/hot emission quotient ($\frac{\mathrm{e^{COLD}}}{\mathrm{e^{HOT}}}$, assigned to be >1):

$$E_{\mathrm{COLD};i,k} = \beta_{\mathrm{TEMP};i,k} \times \widehat{d}_k \times v_{\mathrm{FCD};k} \times A \times \mathrm{e}_{\mathrm{HOT};i,k} \times \left(\frac{\mathrm{e^{COLD}}}{\mathrm{e^{HOT}}}\Big|_{i,k} - 1\right), \tag{7}$$

where β_TEMP parameter depends upon monthly ambient temperature, and the pattern of vehicle use, i.e. the average trip length (17 km per trip estimated for Finland [46]). In (5) and (7), $\widehat{d}$ and v_{FCD} are, respectively, the dynamic traffic density estimated by our three methods and the driving speed obtained from FCD in grid area A, an added term to the original Tier 3 formulas to cope with this geospatial problem. The gridded emission products were then aggregated to our two temporal cases for evaluation against the downscaled emission inventory FRES as described in section 2.3.

2.9. Scenario settings for CAV penetration

Previous studies conducted in Dublin, Ireland, with comparable population size and road networks as the HMA [48, 49] revealed the introduction of CAVs would result in a change of driving behavior, including driving speeds and the overall congestion levels, compared to traditional human-driving vehicles. We applied the changes in congestion levels from three types of road network (major road category, minor road category, and downtown region) to our framework at CAV penetration rates of 25%, 50%, 75%, and 100% (table B5). The trend of electric vehicle usage was not taken into account.

3. Results and discussion

3.1. Evaluation of driving speed and traffic density

Although the traffic dynamics of the HMA have strong temporal patterns in terms of their MFDs, congestion level, and traffic volumes (figures B4 and B5), the driving speeds measured by traffic counters and obtained through FCD have high correlation and low errors in both summer workday and winter weekend cases (figure B6, table B6, $r = 0.74$–0.77, sMAPE = 11%–12%). On the contrary, the traffic density calculated using measurements by traffic counters has a varying correlation with that calculated using the three methods (figure B7, $R^2 = 0.60$–0.88, sMAPE = 31%–68%). This result goes in line with previous studies which demonstrated a direct relationship between TomTom congestion levels and the local daily vehicle counts [14]. When combining the two mobility metrics, our data collected for the HMA had much higher accuracy in driving speeds but slightly better performance in traffic density than [12] who found the deviations between novel mobility data and governmental traffic flow data larger than 60%.

The moderate correlation for the hourly calibrated MFD is attributed to the fact that it provides a generalized q–d–v relationship only based on 14 road types. It assumes that traffic density has the same relationship with driving speed for the same road type, which might not be the case in reality (figure B4, table B7). For the data-driven GLM, the high correlation (table B8, $R^2 = 0.84$–0.88) is attributed to the specialization of the training data. The scope of the comparison has to be limited to the locations where traffic counters were under operation in the studied periods. Besides, the input-adaptive nature and the inclusion of mixed-effects predictors also help boost the model adaptability and accuracy [37]. Despite ENS results in a slightly lower R^2 (0.62–0.72) compared to GLM, it could complement the individual methods MFD and GLM, creating synergy of both physical meanings of traffic flow and statistical evidence reflected from real-world data. For example, ENS manages to improve the estimation at local roads (better distributed along the 1:1 line in figure B7) compared to single benchmark model MFD. This agrees with [44] where ENSs comprising fundamental diagrams and statistical algorithms have shown significant model flexibility to tackle issues of high prediction errors and the lack of physical interpretability arising respectively in the single benchmark models.

3.2. Evaluation of the gridded traffic emission product

Under our framework with speed-dependent emission factors using a projected fleet composition (figure B3), the total CO_2, NO_x, and CO emissions in the HMA are, respectively, 13 300, 36.36, and 17.28 t on a summer workday, and 7810, 15.44, and 9.40 t on a winter weekend. Diurnally, the emission patterns roughly follow the typical traffic volume, with bimodal peaks on workdays and a single afternoon peak on weekends. In both periods, the emissions partitioned by different FRCs roughly correspond to their proportion in the road network in the HMA [9] (figure B8). Besides, fleet composition has a major effect, with passenger cars dominating the GHG emissions and heavy-duty vehicles being the primary source of the AP emissions [26] (figure B9).

The two CO_2 emission products (national emission inventory FRES and our framework using the three proposed methods) show moderate correlation and discrepancies with low variability across the HMA in both studied periods (figure B10, table B6, $R^2 = 0.70$–0.77, sMAPE = 35%–36%) regardless of the method used. Besides, large discrepancies of up to 50% are found when NO_x emission fluxes exceed 4×10^{-6} g m^{-2} s^{-1}, resulting in low overall R^2 values (figure B11, $R^2 = 0.08$–0.18). Although all three methods demonstrate moderate R^2 values (0.53–0.66) for CO, they exhibit very high sMAPE values (70%–79%), indicating the largest discrepancies among the three emittants (figure B12). These discrepancies are also observed spatially, for example in CO_2 emission maps (figures 2 and B13). While both national emission inventory FRES (a reference dataset) and calculated CO_2 emission map using ENS show emission hotspots mostly located along highways in the HMA, their discrepancies are scattered in the studied region regardless of FRCs including minor roads [9, 35] for both workday and weekend cases (figure B14). Besides, at a street canyon with close proximity of traffic in the HMA, we found that our emission products match some of the rush-hour peaks observed in the concentrations profiles measured with reference instruments (figure B15).

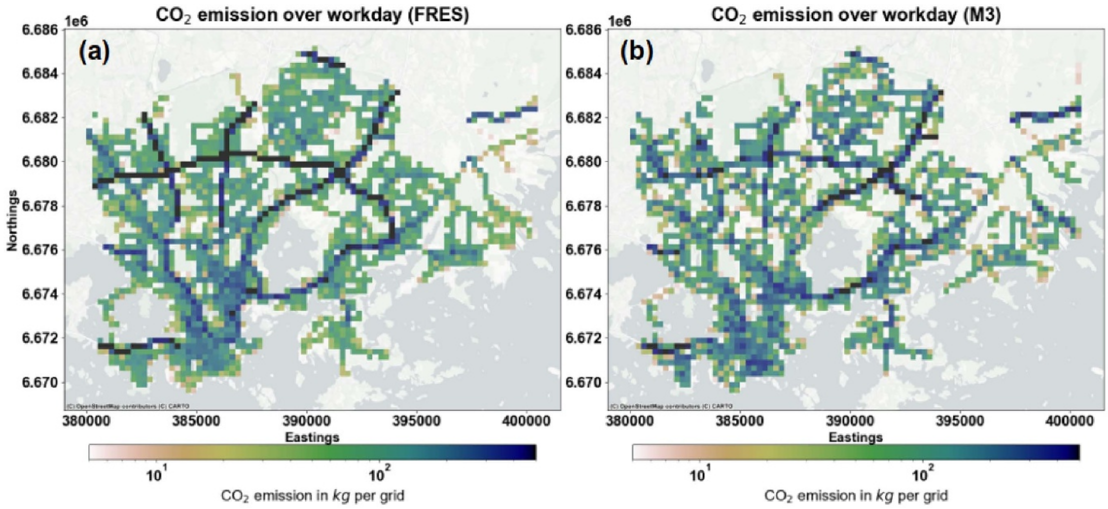

Figure 2. Color-coded emission map (in kg per grid) of daily exhaust CO_2 on a summer workday. (a) Illustrates the spatial distribution of daily emissions using the national emission inventory FRES model (after adjustment from annual to daily emission). (b) Illustrates the one using our proposed ensemble ENS model based on real-time hourly data (summation of hourly profile to one day). $N = 3769$.

The observed discrepancies are common for different emission products [50, 51]. In our case, the discrepancies between FRES and this framework are likely due to the adoption of different emission calculation methods and inputs, including the use of COPERT's Tier 3 emission formulas [47], and the projection of fleet composition [52], respectively. Our framework calculates dynamic emissions for the current year while the national inventory FRES model represents annual emissions in retrospective years, which might not have accounted for the advancements in vehicle emission control technology. Although comparing these datasets may not be perfect due to the difference in the target years and temporal resolutions, FRES remains the best available inventory for spatial evaluation purposes in the HMA. Similar issues can also be found in other parts of Europe, where updated, detailed national emission inventories are not available, let alone in less developed regions worldwide. Despite these challenges, the physics-based method MFD demonstrates promising robustness in estimating traffic emissions. It is worth noting that although GLM excels in estimating traffic density mentioned in the previous section, it gives high discrepancies in traffic emission, in general, compared to FRES. The ensemble method ENS, however, shows performance consistency in both traffic density and gridded emission product with all evaluation metrics combined across both study periods.

3.3. Impact of congestion

To investigate the additional emission induced by congestion, we found that traffic congestion occurrence was distributed mostly in the downtown region in the HMA (figure B16), dominated by residential and commercial low-rise buildings with local roads, where the frequency of inhabitants' movement is the most intensive [53]. Local and other minor roads are designed for relatively small traffic flow with more curves and narrower lanes, which might not be able to cope with unforeseen incidents, for example, traffic accidents and traffic light malfunctioning. Occasional congestion instances with smaller intensity are also recorded along highways with more and wider lanes because of their high traffic flow in terms of volume [39].

The daily emissions of CO_2, NO_x, and CO induced by congestion (emission difference between real-world driving schedule and ideal free-flow driving scenario) were, respectively, 35 900, 203, and 43.6 kg on a summer workday and 18 700, 40.4, and 24.5 kg on a winter weekend, which constitute up to 5%–10% of the total on-road emissions in the HMA (figures 3(a), (b) and B17). The largest emission enhancements were noticed during workday peak hours (7–8 a.m. and 3–4 p.m.). The congestion-induced emissions on the winter weekend were smoothly distributed along the day with a plateau peak between 12–5 p.m.. The similarity of the diurnal cycles of congestion instances (lines) and congestion-induced emissions (bars) agrees with [10] who suggested that congestion plays a role in traffic emission enhancement. This reinforces the need for real-time congestion information to capture the traffic dynamics in an urban road network.

3.4. Scenario analysis on CAV penetration

During peak hours on summer workdays, when congestion is the most severe, the introduction of 25% CAVs would reduce the emission of NO_x by 3.2% (figure 4). On the contrary, it could increase the emissions of CO_2 and CO by 3.4% and 0.3%, respectively. This finding aligns with previous literature that low

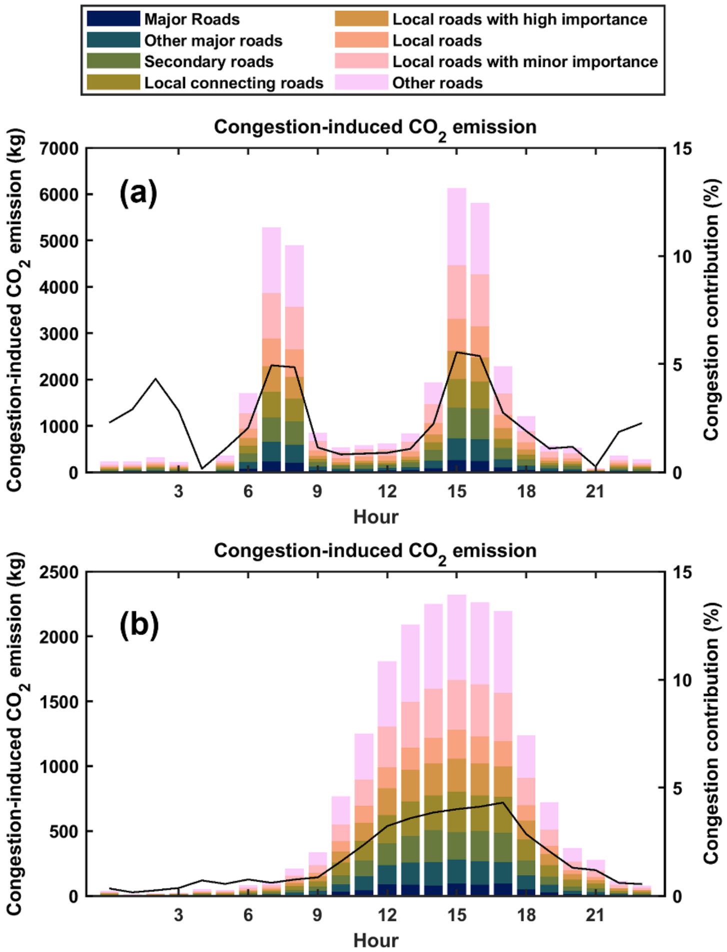

Figure 3. Hourly congestion-induced emissions as bars partitioned color-coded by functional road classes in the Helsinki Metropolitan Area (HMA). (a) Illustrates the situation on a summer workday while (b) displays the winter weekend case. The solid lines indicate the contribution of congestion to total on-road emissions.

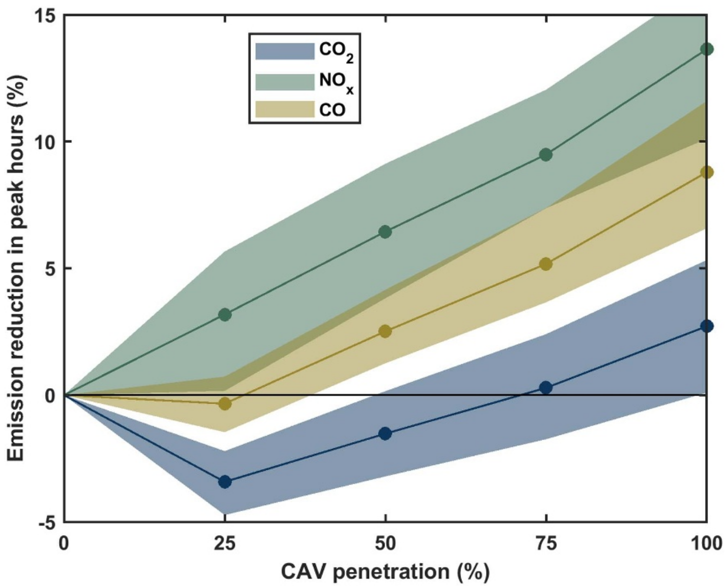

Figure 4. Emission reduction (in %, negative values represents emission increase as compared to no introduction of CAVs) under scenarios of different penetration rates (25%, 50%, 75%, and 100%) of CAVs to the road network during peak hours in the Helsinki Metropolitan Area (HMA). Three emittant species (blue: CO_2, green: NO_x, brown: CO) were illustrated with error ribbons representing their interquatile ranges (20th–80th percentile).

penetration levels of CAVs could increase emissions due to inefficient behavior of non-connected human-driving vehicles [54] and/or escalation of driving acceleration of CAVs [55]. The emission reduction increases with penetration rates of CAVs, as supported by a review paper of the environmental impacts, including exhaust emissions, on CAVs [56], owing to congestion improvements [55]. Through interpolation between the set penetration rates, the traffic emissions of CO and CO_2 are expected to be offset only when more than 28% and 70% of CAVs are introduced to the city road network. In the case of full substitution of human-driving vehicles, emission reductions would rise to 3% for CO_2, 14% for NO_x, and 9% for CO on average.

3.5. Limitations and applications

A main limitation of our framework is the lack of ground truth data of high spatio-temporal resolution to evaluate the three models. While the driving speed from TomTom and the traffic density calculated using the three methods are compared with *in-situ* traffic counter data, the uneven distribution of these counters undermines the representativeness of the evaluation across various road types in the city. The emissions obtained from our proposed method are additionally assessed using the national emission inventory FRES, which is not designed for evaluating dynamic traffic emissions of high temporal resolution with the emerging transportation transformation. Nevertheless, it remains the most reliable reference dataset available for this purpose within the region of concern.

Another limitation is the heavy reliance of the processed FCD provided by the navigation company. Despite the open-access data collected are somewhat documented, the algorithms essentially operate as black boxes, leaving uncertainties about how congestion information is precisely deduced. Consequently, if any discrepancies arise with reality, researchers are left with no option but to trust the data provider. Moreover, the availability of openly accessible data is contingent upon the company's policies, which are not guaranteed to remain consistent. Currently, alternatives of similar traffic products include, but not limited to, Google Maps and MapBox [57].

Furthermore, there are limitations in estimating fleet composition and the corresponding emission factors. Detailed information on fleet distribution at high temporal resolution has been lacking. In this study, we considered the proportion of light-duty and heavy-duty vehicles along roads with traffic counter installations. However, projecting the exact fleet composition requires making bold assumptions. Neither did we account for upstream emissions, which occur during the production, processing, and delivery stages and depend on the types of fuel or energy consumed.

Despite the limitations, the framework for traffic density and emission estimation has demonstrated strong performance. As long as local traffic measurements are available, the framework can be calibrated for the computation of dynamic traffic emissions by inputting real-time traffic raster data from TomTom and urban structure datasets (standardized FRCs and LCZs), which are readily available across a wide spatial coverage. Besides, this framework also unlocks possibilities for conducting scenario analyses on social anomalies, such as strikes or travel restrictions, which cannot be predicted using typical aggregated bottom-up approaches, assuming consistent traffic patterns for workdays and weekends. Moreover, this framework enables simple scenario analysis for varying congestion levels as a function of the penetration of emerging transportation transformations with different driving behavior as in this study.

4. Conclusion

This study presents a geospatial framework for estimating dynamic on-road traffic emissions by leveraging open-access real-time FCD. By employing an ensemble of physics-based and data-driven algorithms ENS, the method demonstrates robustness and accuracy in mapping high-resolution traffic density and on-road emissions (particularly for CO_2, $R^2 = 0.70$–0.77), even in locations lacking *in-situ* measurements. Through comparisons with national emission inventories of exhaust emittants, the study sheds light on the performance variations among different emission models due to the difference in road types and the adoption of different emission calculation methods and inputs. The calibrated framework provides finer spatio-temporal resolution data to support existing urban climate and air quality models for tackling the planetary goals of carbon neutrality and pollution control. It also facilitates scenario analyses for varying congestion levels, essential for understanding the impacts of emerging transportation transformations, which can be used to inform urban planning strategies aimed at mitigating traffic emissions. Future research could incorporate more case cities of traffic patterns with varying similarity to investigate the framework scalability by evaluating the resulting traffic density and emissions with various bottom-up and top-down data sources.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors. The underlying codes and findings generated and analyzed for the geospatial framework of this study are partly available in Zenodo repository [58], which will be available from 1 January 2025.

Acknowledgments

The work is supported by Artificial Intelligence for Urban Low-Emission Autonomous Traffic (AIforLEssAuto) under Green and Digital transition call from the Research Council of Finland (Project Nos.: 347197, 347198, 347200), Research Council of Finland ACCC Flagship (Project No.: 337549) and the European Union's Horizon 2020 research and innovation programme under Grant Nos. 101036245 and 101037319. The authors would like to acknowledge the City of Helsinki for providing traffic information, and Jarkko V Niemi from the Helsinki Region Environmental Services (HSY) for providing measurement data in Mäkelänkatu supersite.

Author Contribution

P L F: Conceptualization, Methodology, Formal Analysis, Investigation, Validation, Writing—Original Draft. O A-J: Investigation. J C: Validation. V-V P, S V and C R: Resources. L J: Resources, Supervision, Funding acquisition. All co-authors have reviewed and edited the manuscript before submission. No potential competing interest was reported by the authors.

Appendix A. Supplementary text

This supporting information document includes the description of the study area, together with the distribution of LCZs, and road network characterization using measurements by traffic counters and national emission inventory using FRES model. We provide an in-depth discussion of traditional MFDs for road characterization. We describe the evaluation metrics used for comparing various models. Additionally, extended results, including evaluation of driving speed and traffic density with traffic counter data, evaluation of gridded emission with national emission inventory of greenhouse gases (GHGs: CO_2) and air pollutants (APs: NO_x and CO), and description of congestion variation, are also elaborated.

A.1. Study area

Helsinki, together with the neighboring municipalities Espoo, Vantaa, and Kauniainen, forms the HMA. The HMA is located at the coast of the Gulf of Finland with a humid continental climate with cool summers and cold winters. The HMA covers an area of 213.8 km^2 with a population of over 1.5 million. In this study, we used a bounding box within the latitudes of 60°8′2′′ N and 60°18′ N and the longitudes of 24°48′ E and 25°12′ E in WGS 84 reference ellipsoid system. This bounding box covers part of the HMA (figures B1 and B2).

A.2. LCZs

In LCZ classification, urban areas are divided into 10 built classes (LCZ 1–10) and seven land cover classes (LCZ A–G) based on the surface properties (e.g. building and tree height & density, pervious vs. impervious). The classification (table B4), is subjected to variable or ephemeral land cover properties that change significantly with synoptic weather patterns, agricultural practices, and/or seasonal cycles. Each zone is local in scale, meaning it represents horizontal distances of hundreds of meters to several kilometers. The classification was generated worldwide by [43] using random forest techniques. We extracted the data with the bounding box of the HMA at a resolution of 250 × 250 m^2. Inside the bounding box, 14 LCZs are found (figure B2). Built LCZs cover 60% of the bounding box. The most commonly built LCZ is LCZ 5 (open midrise, 24%), followed by LCZ 6 (open lowrise, 15%) and LCZ 8 (large lowrise, 14%). The most commonly found land cover LCZ is LCZ A (dense trees), which covers 24% of the total land area, followed by LCZ D (low plants, 7%) and LCZ B (scattered trees, 6%). The spatial distribution of LCZ is shown in figure B2.

A.3. Road networks

The HMA has a well-developed and well-maintained road network covering a spectrum of road types from four-lane highways to narrow alleys with one-lane traffic. The road network is dominated by seven access roads leading to the city center of Helsinki and these access roads are connected by Ring I (Kehä I). The fleet composition the HMA has been estimated as some of the most found car types operated: passenger cars using petrol (63.0%), passenger cars using diesel fuels (23.5%), light-duty vans using diesel (9.8%), heavy-duty trucks using diesel (2.7%), and light-duty vans and heavy-duty trucks using petrol, both less than 1%, from the database given by [39]. The share of other vehicle types including fully EV, plug-in hybrid electric vehicles, and natural gas vehicles have been on the rise since the last decade; however, these new energy vehicles still account for a minor share in the fleet composition in the HMA. More details on the selected car types in this study can be found in table B2.

The road network in the HMA was classified using the Functional Road Class (FRC) framework developed by TomTom [15]. Eight classes can be found inside our bounding box in the HMA (figure B1): major roads (10%), other major roads (7%), secondary roads (6%), local connecting roads (10%), local roads with high importance (13%), local roads (16%), local roads with low importance (10%) and other roads (28%). The proportion of light-duty vehicles on the specific FRC is also calculated, which ranges from 0.89 to 0.97. This information is important to estimate the local fleet composition for later stage in the framework. Roads are grouped into classes according to the character of the service they are intended to provide. FRC defines the nature of this channeling process by defining the role that any particular road or street should play in serving the flow of trips through a road network. The heavy travel movements are directly served by major channels, and the lesser trips are channeled into somewhat indirect paths. FRC is designed to categorize segments based on their functional importance within the transportation network. More details of the classes can be found in table B3. The use of FRC instead of local road classification is that TomTom has classified roads using this standardized framework in their operating countries. The worldwide operation covers most parts of Europe, America, and Oceania, and extends to some Asian and African countries. This extensive coverage of the road class standardization could help improve the scalability of the proposed model.

FRCs were further categorized by their SLs. They form a total of 14 road types in total. All road types have their corresponding parameterization of free-flow speed v_m, critical speed v_c, and capacity volume q_c, which are derived by traditional MFDs (see more in the following section) using the corresponding traffic counter data.

A.4. Road characterization using MFDs

We adopted [28]'s method to calibrate traditional MFDs based on traffic flow data. Inspired by the special properties of flow–speed (q–v) diagram as well as the insights from previous studies, we fit pre-defined distributions to traffic flow q and driving speed v. Speed and volume distributions of a road encode important information about the shape of q–v diagram, which makes it possible to infer MFD-related parameters, for example, the critical volume (q_c), the critical speed (v_c) and the free-flow speed (v_m). While the free-flow speed of a road r can be easily estimated as $v_m = \max\{v \in V_r \text{fitted in best dist.}, SL_r\}$, the estimation of critical speed v_c is less straightforward. We also used K-means to cluster the speed data into classes of free-flow regime C_{free} and congested flow regime $C_{congest}$. Finally, the critical speed v_c is estimated as $v_c = 0.5 \times (\max\{v \in C_{\mathrm{congest}}\} + \min\{v \in C_{\mathrm{free}}\})$. The critical volume q_c was estimated as the 98th percentile of the fitted distribution Q_r for each road.

The exact traffic volume q and its corresponding traffic density d are estimated using the q–v diagram by introducing a road utilization function λ, which approximates on what level the road is utilizing its total volume capacity, and explicitly encodes the shape of the q–v diagram thereby providing extra information for volume estimation (A.1)

$$q \approx q_{\mathrm{c}} \cdot \lambda(v) = \begin{cases} q_{\mathrm{c}} \cdot \sqrt{\frac{v}{v_{\mathrm{c}}}} & , 0 < v \leqslant v_{\mathrm{c}} \\ q_{\mathrm{c}} \cdot \max\left\{\frac{v_m - v}{v_m - v_{\mathrm{c}}}, 0\right\} & , v_{\mathrm{c}} < v \leqslant v_m. \end{cases} \tag{A.1}$$

However, due to the incapability to capture the hourly dynamics of traffic volume and density, this traditional MFD has not been implemented into our proposed geospatial framework. Compared to the dynamic MFDs applied in this study and Hong Kong [26] where they investigated the relationships between hourly-averaged single-lane traffic volume and speed across different road types using Greenshield's equation [31], traditional MFDs (black line) might fail to capture the temporal variability (for example 8 a.m. and 4 p.m. in figure B4) of the traffic activity on certain roads. The two MFDs only share some similar patterns for other major roads with SL of 60 kph. The results using traditional MFDs, on the other hand, are very important to characterize the different road types with various FRCs and SLs in the HMA. The deduced traffic variables (v_m, v_c, and q_c) are presented in table B1. Together with the traffic profiles based on congestion index inferred by FCD and *in-situ* traffic counter data figure B5, an overview of traffic conditions can be portrayed. The FCD we collected from TomTom is based on four sources: data from global system for mobile communications networks, real-time GPS data from users of TomTom navigation devices, third-party data such as government traffic control centers, and historic information collected from navigation devices [15].

According to the measurements by traffic counters operated in the HMA, we analyzed the road characteristics of different FRCs and SLs based on their vehicle counts and driving speed in the summer workdays and winter weekends. The overall characteristics behave in such a way that v_m, v_c, and q_c in all measured roads are higher in the summer than winter. This is in alignment with previous studies which suggested that road condition is a critical determinant governing the shape of MFD (e.g. [27, 28]). Driving conditions can become difficult under adverse weather conditions, such as heavy rain, fog, ice, snow, or slush [59], especially in the wintertime in high-latitude cities. This could lead to consequent anticipatory driving, reduced speed, increased following distance and extra vigilance, which in turn alter the inter-relationship among driving speed, traffic flow, and traffic density.

However, traffic counters in the HMA have been restricted mostly along major roads, other major roads, secondary roads, and local connecting roads where elevated vehicle volumes have been observed. In addition to the limited data, MFD has been argued to be valid only in major roads as a broad category because any travel time would be possible for any level of demand [27]. Therefore, this traditional method might not be applicable to smaller roads that constitute more than half of the total road network in the HMA. This reinforces the idea of developing a framework to mitigate the high spatial and temporal variation in traffic patterns by selecting dynamic MFDs (using traffic flow equations as the physics-based model) for major roads and another data-driven model, such as GLM with land use predictors for non-major roads.

A.5. Urban air quality and greenhouse gas measurements

In addition to the FRES model, we collected hourly concentration measurements of our target emittants (CO_2 in ppm, NO_x in µg m^{-3}, and CO in µg m^{-3}) from Mäkelänkatu supersite (60°11′ N, 24°57′ E, 24 m a.s.l.), operated by the Helsinki Region Environmental Services Authority (HSY) for the summer workday and winter weekend cases. The supersite is situated at 3 km from the city center at a street canyon in the immediate vicinity to one of the main roads leading to downtown Helsinki, with six lanes and two tramlines. The annual mean traffic volume in 2018 per workday was 28 100 vehicles, 11% of which were recorded as the heavy-duty vehicles. The traffic loads are especially high during rush hours at 8 a.m. and 5 p.m.. The street canyon of 42 m width is surrounded by rows of 17 m high buildings, which weaken the dispersion process of the direct traffic emissions. All the inlets for the measuring devices are positioned approximately at a height of 4 m from the ground level [40].

A.6. Evaluation metrics

The correlation coefficient (r) is a statistical measure indicating the strength and direction of a relationship (A.2). It ranges from –1 to 1, with positive values denoting a positive correlation and negative values representing a negative correlation. A value of 0 implies no correlation at all. In practical applications, r is utilized to assess correlations between different parameters from two separate data sources. For example, it can be employed to compare driving speeds obtained from TomTom with those recorded by traffic counters.

The coefficient of determination, denoted by R^2, measures the goodness of fit of a model (A.3). R^2 is often interpreted as the proportion of response variation 'explained' by the regressors in the model. Thus, $R^2 = 1$ indicates that the fitted model explains all variability, while $R^2 = 0$ indicates no 'linear' relationship. R^2 was used when we intended to show how well our proposed models explain the variability of the traffic density, and also the corresponding gridded emission in the HMA.

Symmetric mean absolute percentage error (sMAPE) is an accuracy measure based on percentage (or relative) errors (A.4). sMAPE was used when the ground truth is unknown. As neither the national emission inventory nor the real-time emission estimated by TomTom is the ground truth, sMAPE is, therefore, applied to avoid leaning towards either one. The percentage error ranges between 0% and 100%. Both over- and under-estimates have the same positive sign

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \tag{A.2}$$

where x_i and y_i are the two parameters of interest at data point i, respectively, and $\bar{x}$ and $\bar{y}$ the corresponding mean of the datasets

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \tag{A.3}$$

where y_i and $\hat{y}_i$ are the output parameter and the estimated output parameter at data point i, respectively

$$\mathrm{sMAPE} = \frac{100}{n} \sum \frac{|\hat{y}_i - x_i|}{|\hat{y}_i| + |x_i|}, \tag{A.4}$$

where x_i and $\hat{y}_i$ are the input parameter and the estimated output parameter at data point i, respectively. n is the total number of data points.

A.7. Extended results for driving speed and traffic density evaluation

Scatter plots of the driving speed of TomTom raster against traffic counters are plotted for summer workdays and winter weekends in figure B6. The correlations in both cases are high (r = 0.74–0.77) with around 11% of sMAPE. This indicates that the hourly raster data retrieved from TomTom has a high consistency with the *in-situ* measurements collected in certain road types in the HMA. The small discrepancy might be due to the difference in temporal sampling method [38]. The traffic counter data were averaged hourly while data from TomTom were retrieved once per hour representing traffic conditions at that very moment. Driving speed in the winter case is slower due to adverse weather conditions leading to anticipatory driving [59].

Furthermore, we used three methods to quantify traffic density. Table B7 illustrates the coefficients of k_0 and k_1 using the dynamic MFD (Greenshield's equation, [31]) at some hours (8 a.m. and 4 p.m. to represent rush hours) along different road types (varying FRCs and SL). The flow–density–speed (q–d–v) relationship of other major roads with SL of 60 kph using this dynamic MFD is visualized in figure B4 as an example. The individual R^2 at certain hours along certain road types vary a lot ($R^2 = 0.05$–0.94). For cases with high R^2, for example, other major roads with SLs of 60 kph, the reasons could be due to the traffic dynamics during that hour are towards its full road capacity, so that the q–d–v relationship of that road type can be well portrayed. On the contrary, the R^2 in other cases would be due to the lack of road utilization. If there are a few cars utilizing the road, there may exist many possibilities that govern the relationship of q–d–v in that road segment. Combining all individual road types, the R^2 would be fairly good between 0.60–0.69, which is described in the main text.

For the second method GLM, we explained a bit further the choice of adaptive inputs in table B8. As driving speed and congestion level (fixed effect) and hour (random effect) are always available, the difference between the four sub-models is whether or not to include the other two random effects FRCs and/or LCZs. Apparently, if we consider all four input variables, the sub-model performance would be the best ($R^2 = 0.82$–0.86, sMAPE = 33%–43%) as compared to sub-models with fewer input variables ($R^2 = 0.73$–0.85, sMAPE = 36%–55%). The overall model, surprisingly, shows even better performance ($R^2 = 0.84$–0.88, sMAPE = 33%–43%) in both studied times. This implies that the input-adaptive framework indeed helps boost the flexibility and accuracy as shown in previous studies (e.g. [37]). The use of GLM is to employ a simple model with widely accessible data that can be effectively scaled to other test locations.

To evaluate the traffic density calculated using the proposed three methods, we present scatter plots of calculated versus measured traffic density from the traffic counters in the HMA (figure B7). Comparing the three methods, GLM appears to have the best performance while MFD fails to estimate traffic density in local roads. ENS, as the ensemble of both, ranks between them. This noticeable difference is attributed to the difference in model architecture. As GLM is entirely a data-driven model, it would work relatively well if we test it at the same site where we collect our training data. The hourly calibrated MFD follows some physical equations that might not work perfectly in every condition.

A.8. Extended results for emission evaluation

Before quantifying the gridded emission, one important component is the speed-dependent emission factor. We visualize the emission factor curves for the six selected vehicle types, and the total fleet (figure B3). While speed-dependent emission factor equations for individual vehicle class have been established with real-world measurements, the overall emission factor characteristics depend on the unique fleet composition of each city. The curves describing the relationship between emission factors and driving speed, to some extent, have a concave shape for all three emittant species. At driving speeds below 40 kph and above 100 kph, the emission factors are higher than driving at medium speeds. For CO, the elevated emission factors even show exponential increments after 80 kph. The non-linearity properties of both the fundamental diagrams and emission factors for various emittant species imply that the relationships among traffic congestion, driving speed, traffic volume, traffic density, and the resulting emission are more than a one-to-one straightforward function.

Calculated emissions were plotted against the national emission inventory FRES in raster format. Each point represents a grid of 250 × 250 m^2 inside the bounding box. Boxplots showing their binned median and interquatiles can be found in figure B10 for CO_2, figure B11 for NO_x and figure B12 for CO. The calculated CO_2 has an overall estimation for the inventory as shown by their R^2s. The total CO_2, NO_x, and CO emissions using ENS in the HMA are, respectively, 1330, 3.65, and 1.74 t on a summer workday and 747, 1.42, and 0.89 t on a winter weekend. The variances for the boxes in figure B10 within the range of small CO_2 values (e.g. below 1×10^{-3} g m^{-2} s^{-1}) are much smaller than that of larger values. MFD apparently outperforms the other two methods ($R^2 = 0.76$–0.77, sMAPE = 35%). The case for NO_x is very different from CO_2 in such a way that the whole framework fails to give a good estimate for large inventory emissions regardless of which method is used. The correlation for small emissions appears to be good until emissions reach 4×10^{-6} g m^{-2} s^{-1} where the proposed models introduce larger discrepancies (figure B11). This also leads to a low overall R^2 ($R^2 = 0.08$–0.18). The overall errors are as high as 50%. CO, interestingly, shows another distinct pattern from the other two emittants (figure B12). The three models all show moderate R^2 (0.53–0.66), but very high sMAPE (70%–79%). This high sMAPE representing a large discrepancy is plausibly owing to the technological advancement of emission control for vehicles. Considering the data used for projecting the national emission inventory was collected for the year 2015, the rapid change in transportation transformation has been significant. Meanwhile, the designed framework uses the most up-to-date emission factors taking into account of the new models of vehicles. The comparison of these two datasets might not be perfect as they were not implemented using the same baseline; however, the national emission inventory is the best available material so far for evaluation purpose. The insufficiency of reference datasets is not an individual issue. Updated national emission inventory is not available in every European country, let alone in less developed regions in the world.

As for their spatial evaluation, figure B13 shows two CO_2 emission maps for winter weekends: one for the FRES model and another is the calculated emission using our proposed framework ENS. Similar to the one presented in the main text for summer workdays, the emission hotspots are located similarly, mostly along main roads. Figure B14 shows the spatial discrepancies of the emissions modeled by the proposed framework versus the FRES model (after adjustment from annual to daily emission). The overall discrepancies are not large, but considerable grid variances can be observed for CO_2 emissions, in alignment with the boxplots in figure B10.

In figure B15, we also compared our emission products calculated using ENS (orange dotted lines) with hourly concentration measurements (blue solid lines) of CO_2, NO_x, and CO at a traffic site in a street canyon in the HMA for both temporal cases. During summer workdays, the morning peaks of CO_2 and NO_x traffic emissions match the concentrations measured at the traffic site. However, due to the progression of the meteorological conditions and enhanced mixing throughout the day, the measured concentrations failed to match the afternoon peaks of the corresponding emissions calculated using our model. For CO, due to external factors, such as more diverse emission sources, very low ambient concentrations, and the meteorological conditions mentioned above, no diurnal patterns could be observed for the measured concentrations. For winter weekends, as the emissions of all the three emittants were even lower, it was more likely that the abovementioned external factors suppressed the development of a noticeable diurnal cycle for concentrations.

A.9. Extended results for congestion variation

To investigate to additional emission induced by congestion, we first illustrate the congestion hotspots in the HMA in figure B16 for the two studied periods (summer workdays and winter weekends). Apparently, congestion instances are more intense on summer workdays, especially in the south of the HMA where commercial activities actively take place intensive [53]. Highways do not show frequent congestion despite they usually have large traffic flow because they have been designed to support large traffic flux by building more lanes. Therefore the road capacity for highways is high enough to keep vehicles moving at a (near-)optimal speed. Therefore, most of the congestion-induced emissions do not take place along major roads, but in smaller roads, such as local roads and other roads in (figure B8). This situation can be observed for all emittants, especially NO_x, for which major roads constitute a trivial amount of emissions induced by congestion. Similar partitioning was carried out for fleet composition (figure B9). Passenger cars (using both diesel and petrol) account for most of the congestion-induced emissions of CO_2 and CO, due to the substantial amounts of vehicles. As for NO_x, heavy-duty trucks fueled with petrol apparently become more significant in contributing to the total emission. Despite the relatively small quantity of heavy-duty vehicles, they account for more than 80% of NO_x emission.

The daily emissions (CO_2, NO_x and CO) induced by congestion (emission difference between real-world driving schedule and ideal free-flow driving scenario) are, respectively, 35 900, 203, and 43.6 kg on a summer workday and 18 700, 40.4, and 24.5 kg on a winter weekend, which constitute up to 10% of the total on-road emission in the HMA (figure B17).

Appendix B. List of supplementary figures and tables

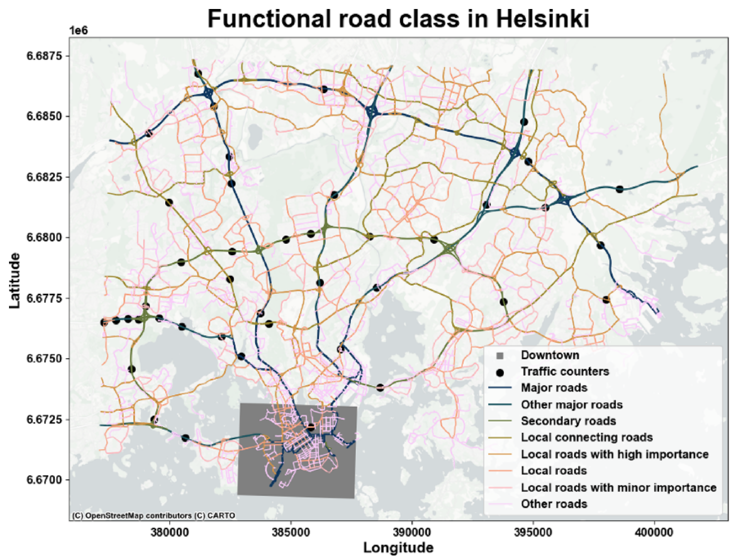

Figure B1. Road network in the Helsinki Metropolitan Area (HMA) colored by road classes using Functional Road Class (FRC). Black markers represent the locations of the traffic counters. Grey region indicates a downtown bounding box for the scenario setting.

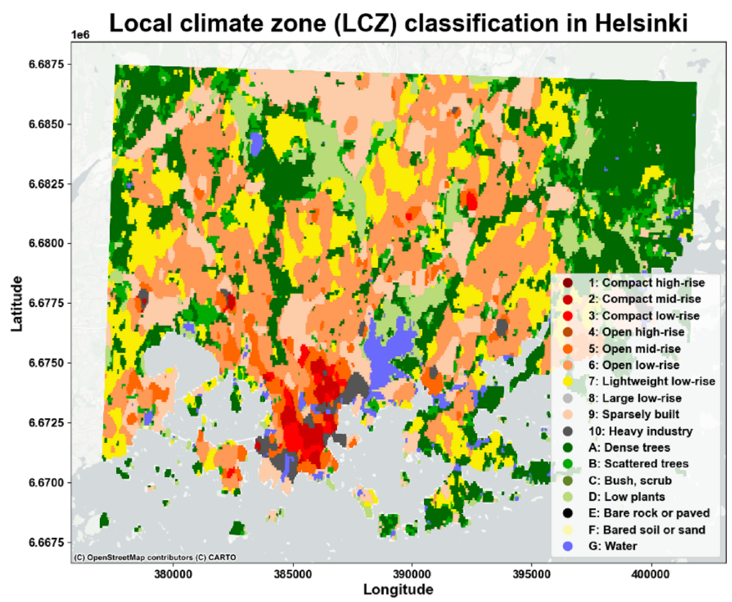

Figure B2. Color-coded local climate zone (LCZ) distribution in the Helsinki Metropolitan Area (HMA).

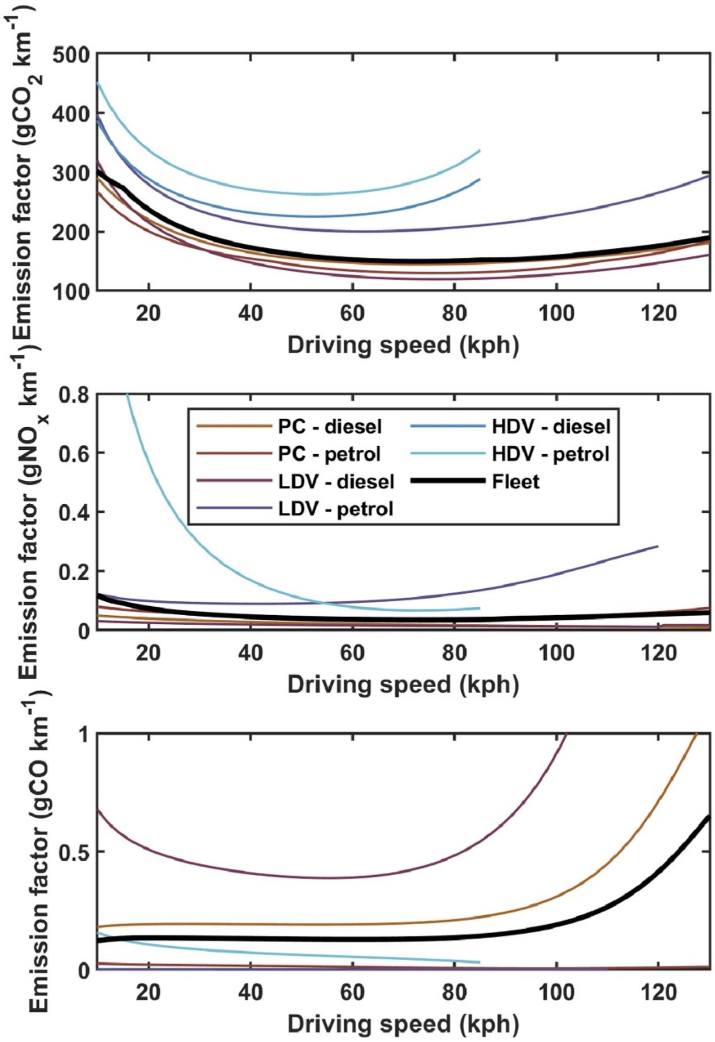

Figure B3. Speed-dependent emission factors (in g km^{-1}) for the six selected vehicle types (colored lines) and the fleet composition (black bold line) in the Helsinki Metropolitan Area (HMA). The first row represents the case for CO_2, followed by NO_x and CO in rows 2 and 3, respectively. For vehicles types, PC, LCV, and HDT stand for passenger cars, light commercial vehicles, and heavy-duty trucks, respectively. $N = 3769$. All p-values are below 0.05.

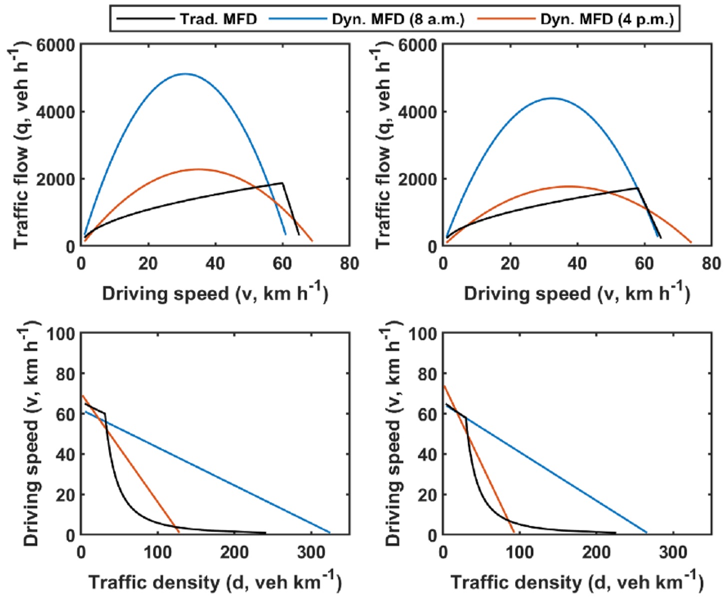

Figure B4. Comparison of flow–density–speed (q–d–v) relationship using traditional MFDs [28] against dynamic MFDs ([31], at 8 a.m. and 4 p.m. as examples) for other major roads with speed limit of 60 kph. The relationships of q–v and v–d are shown in row 1 and 2, respectively. The first column illustrates the interrelationship for the summer workday case while the second column represents the winter weekend. The MFD has been fitted using real-world traffic counter data.

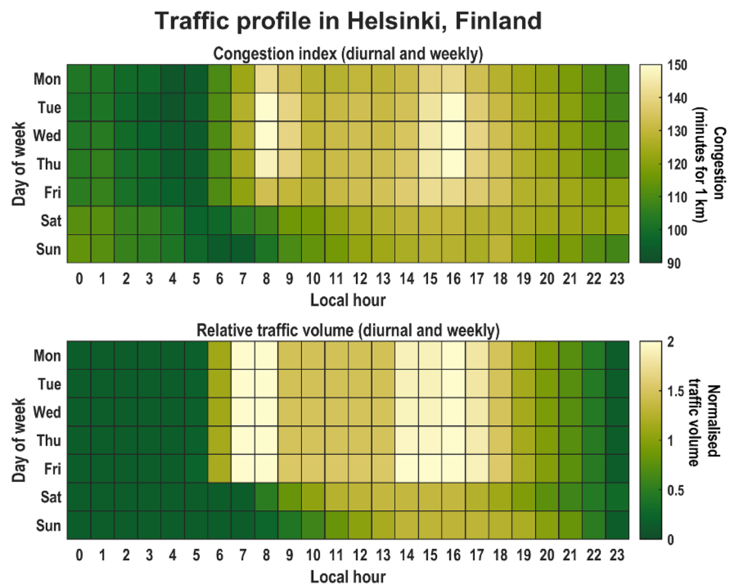

Figure B5. Color-coded diurnal and weekly traffic profile in The Helsinki Metropolitan Area (HMA) in terms of TomTom aggregated congestion index (upper panel) and normalized traffic volume used in our reference dataset FRES national emission inventory (lower panel).

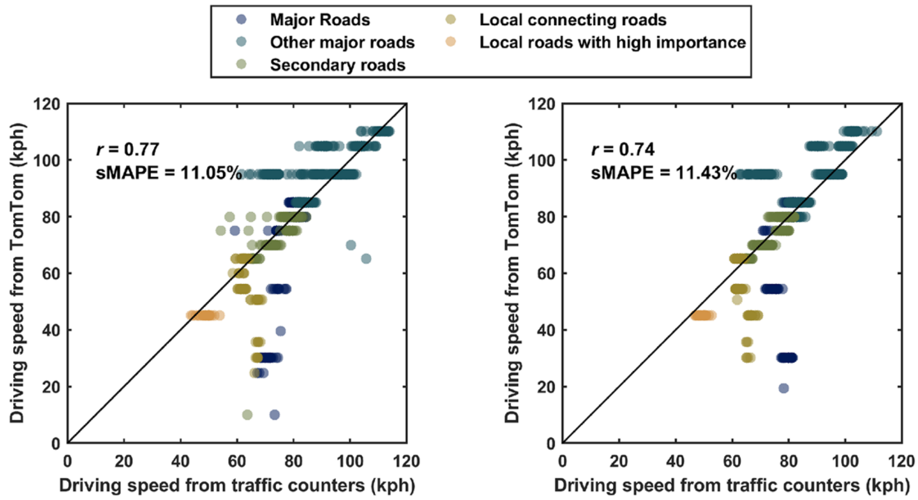

Figure B6. Comparison of driving speed retrieved from TomTom against traffic counters colored by Functional Road Classes (FRCs). The left panel shows the summer workday case. The right panel shows the winter weekend case. r and sMAPE are shown for each case. $N = 1512$. All p-values are below 0.05.

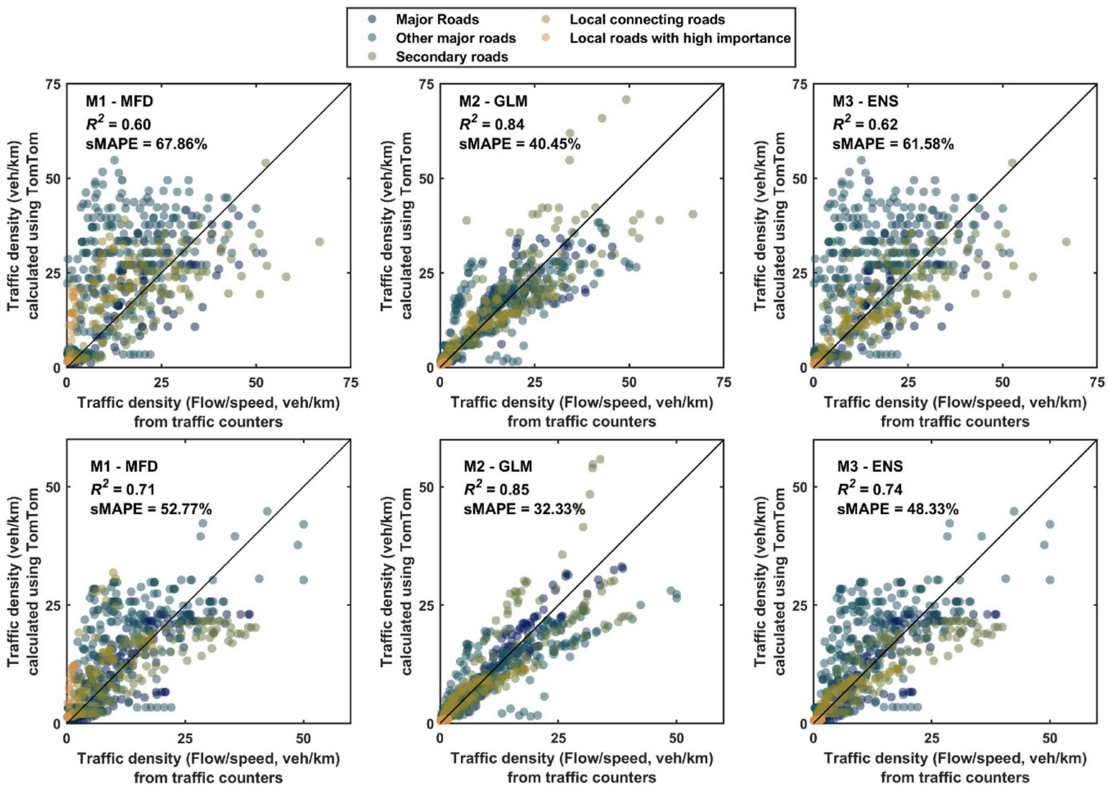

Figure B7. Comparison of traffic density using proposed models (M1: macroscopic fundamental diagram MFD in column 1, M2: generalized linear model GLM in column 2, M3: ensemble method ENS in column 3) against that calculated from traffic counters colored by Functional Road Classes (FRCs). The upper row shows the summer workday case. The lower row shows the winter weekend case. R^2 and sMAPE are shown for each case. $N = 1512$. All p-values are below 0.05.

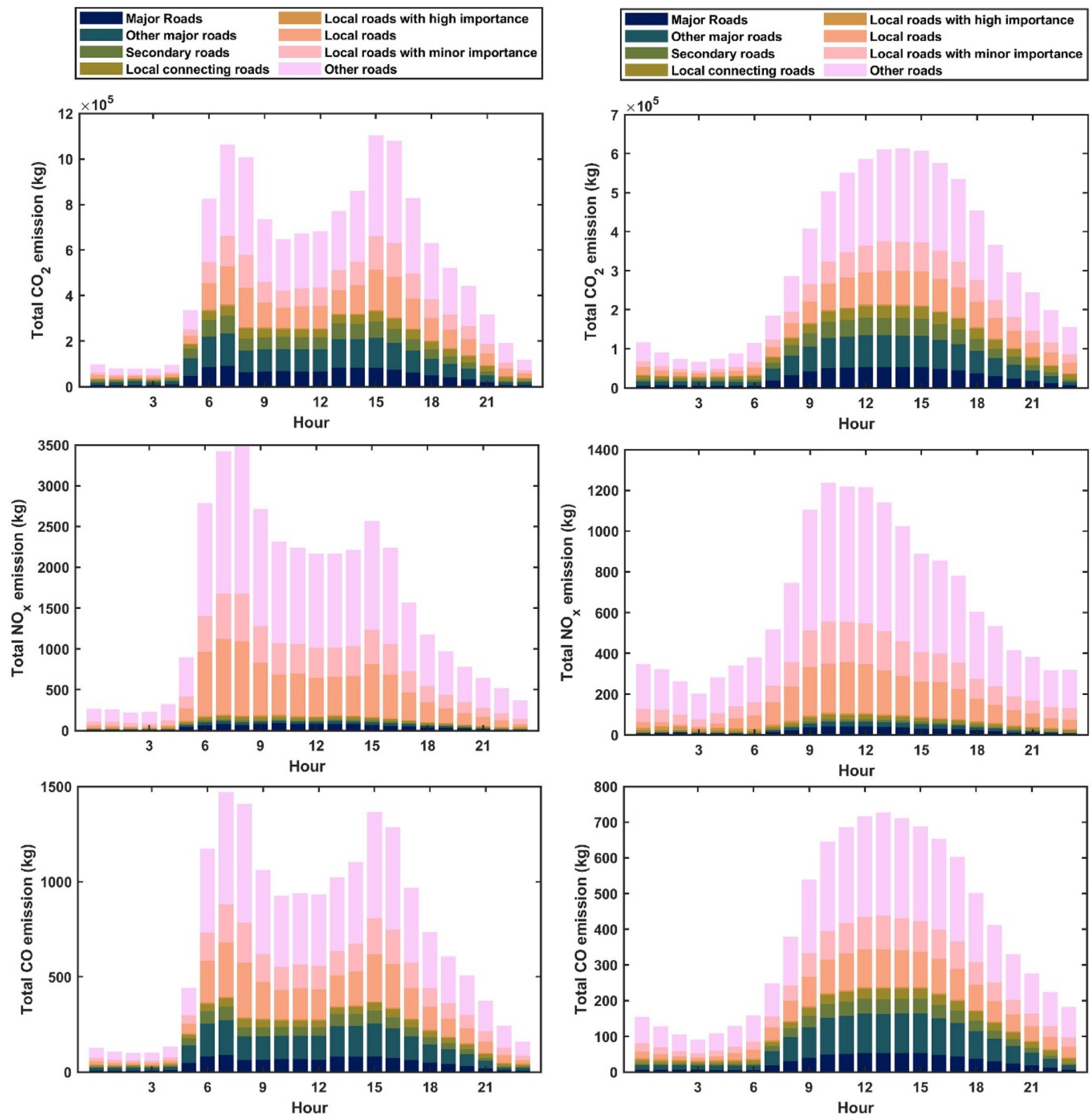

Figure B8. Emission partitioned by road classes. The left column shows the summer workday case. The right column shows the winter weekend case. Row 1 illustrates CO_2 emission, followed by NO_x and CO in rows 2 and 3, respectively.

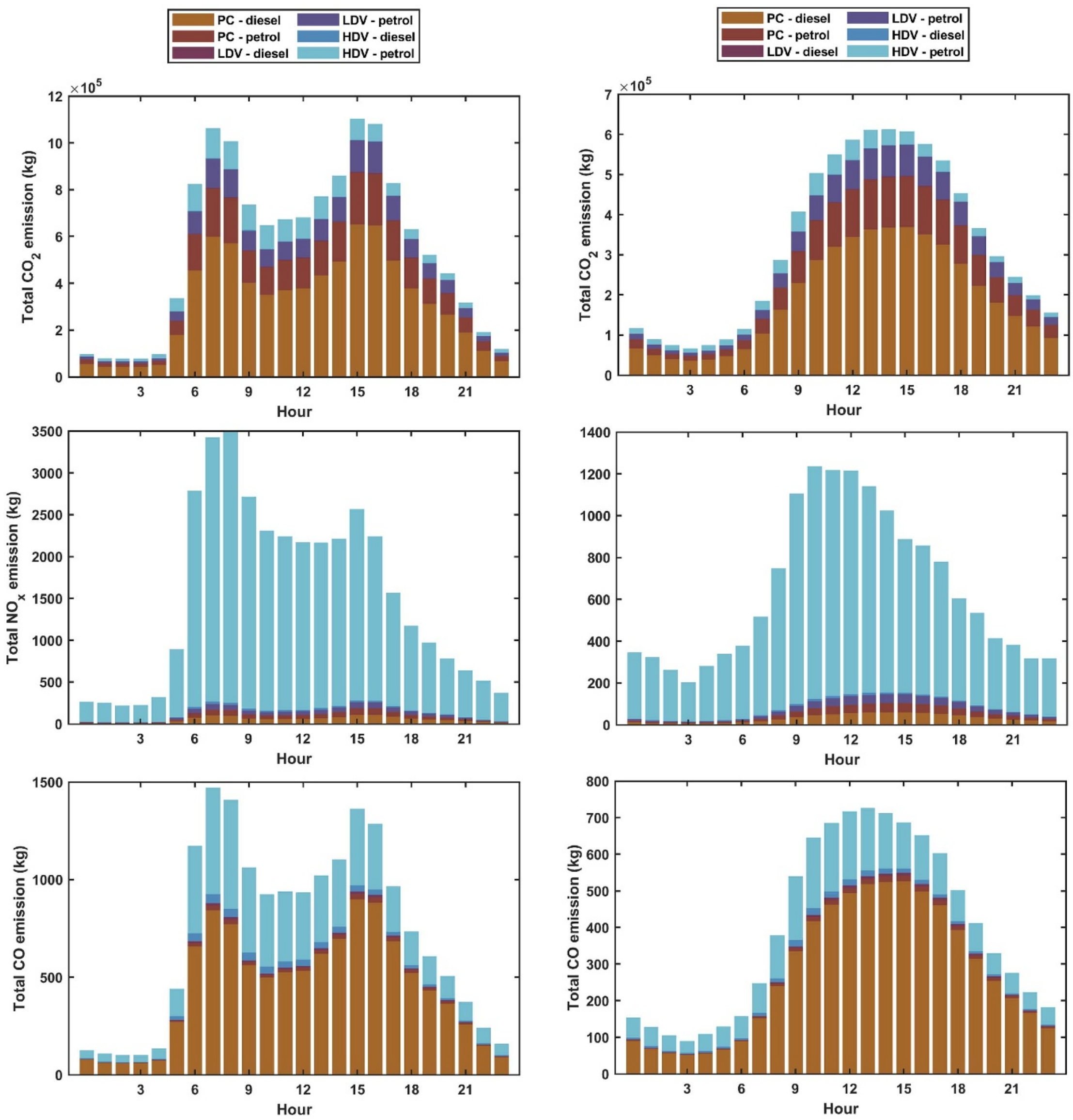

Figure B9. Emission partitioned by fleet composition. The left column shows the summer workday case. The right column shows the winter weekend case. Row 1 illustrates CO_2 emission, followed by NO_x and CO in rows 2 and 3, respectively.

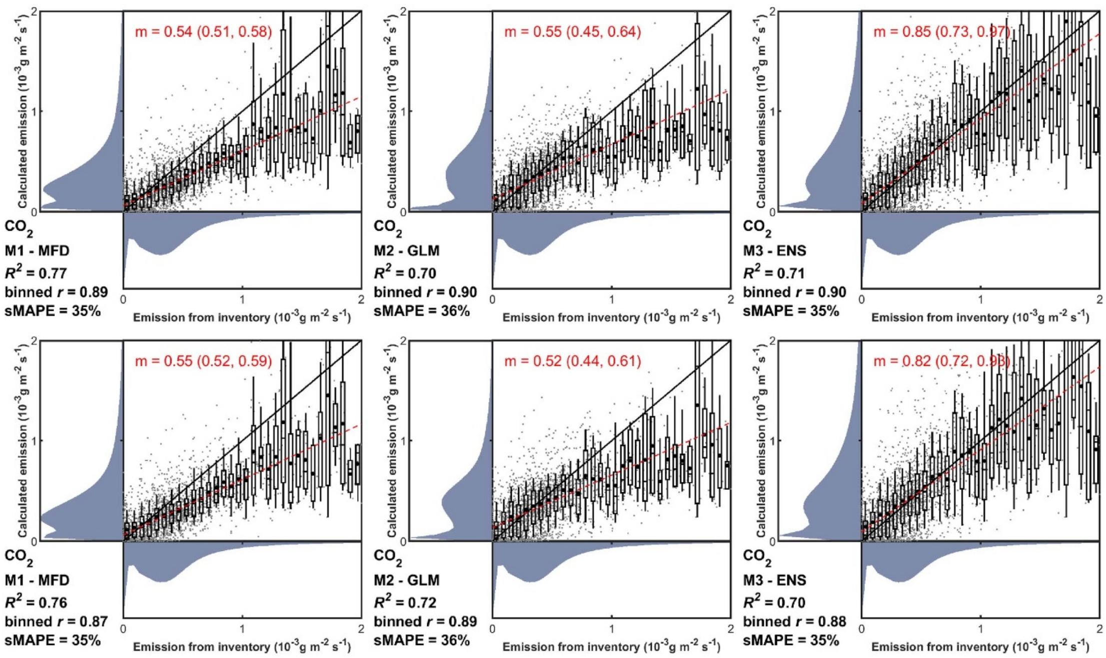

Figure B10. Comparison of CO_2 emission using proposed models (M1: macroscopic fundamental diagram MFD in column 1, M2: generalized linear model GLM in column 2, M3: ensemble method ENS in column 3) against emission inventory. The upper row shows the summer workday case. The lower row shows the winter weekend case. Model evaluation metrics including R^2, binned r, sMAPE, and slope m are shown for each case. $N = 3769$. All p-values are below 0.05.

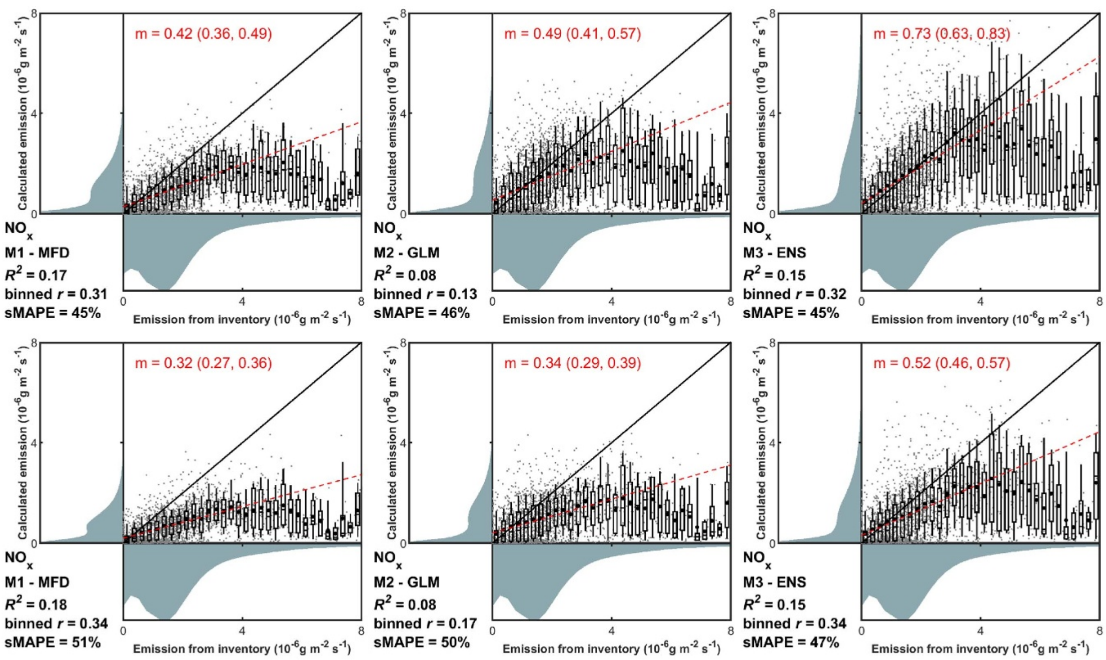

Figure B11. Comparison of NO_x emission using proposed models (M1: macroscopic fundamental diagram MFD in column 1, M2: generalized linear model GLM in column 2, M3: ensemble method ENS in column 3) against emission inventory. The upper row shows the summer workday case. The lower row shows the winter weekend case. Model evaluation metrics including R^2, binned r, sMAPE, and slope m are shown for each case. $N = 3769$. All p-values are below 0.05.

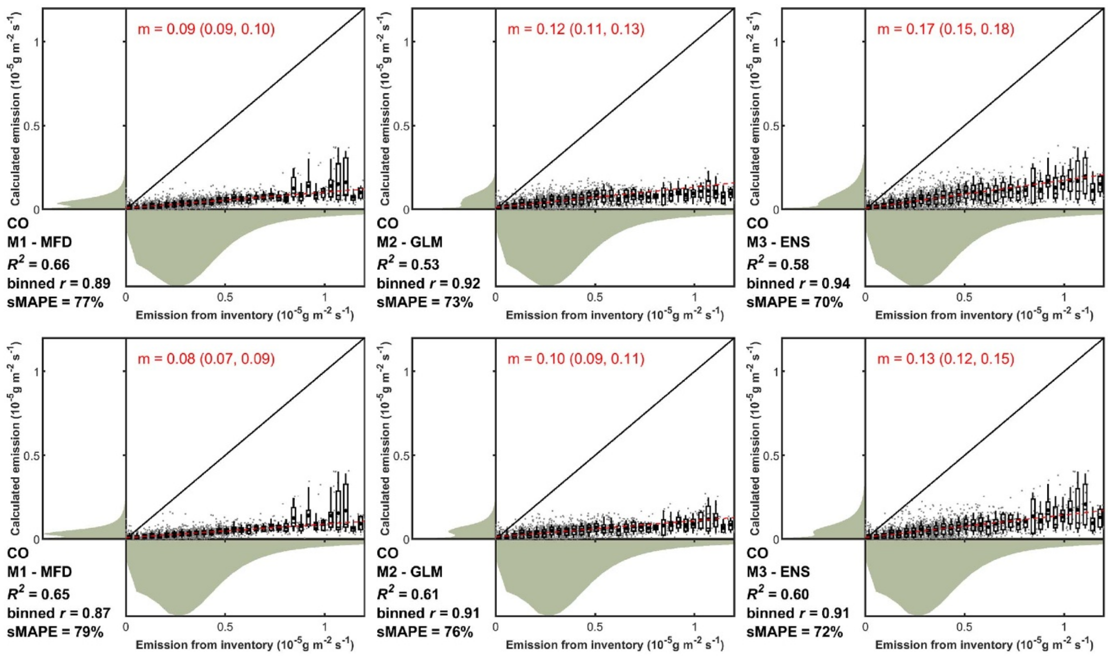

Figure B12. Comparison of CO emission using proposed models (M1: macroscopic fundamental diagram MFD in column 1, M2: generalized linear model GLM in column 2, M3: ensemble method ENS in column 3) against emission inventory. The upper row shows the summer workday case. The lower row shows the winter weekend case. Model evaluation metrics including R^2, binned r, sMAPE and slope m are shown for each case. $N = 3769$. All p-values are below 0.05.

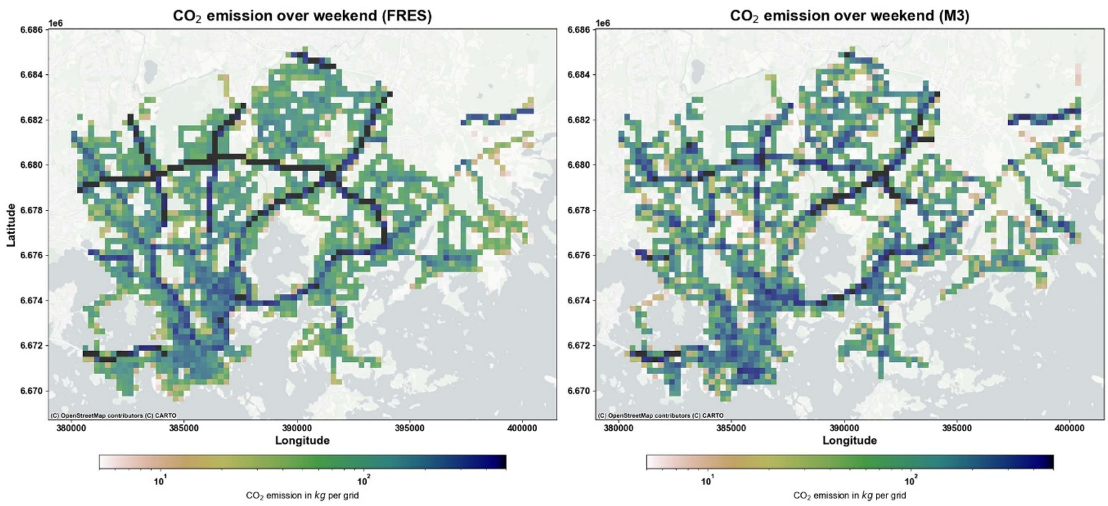

Figure B13. Color-coded emission map (in kg per grid) of daily exhaust CO_2 on a winter weekend. The left panel illustrates the spatial distribution of daily emission using the national emission inventory Finnish Regional Emission Scenario model (after adjustment from annual to daily emission). The right panel illustrates the one using our proposed ensemble ENS model based on real-time hourly data (summation of hourly profile to one day). $N = 3769$.

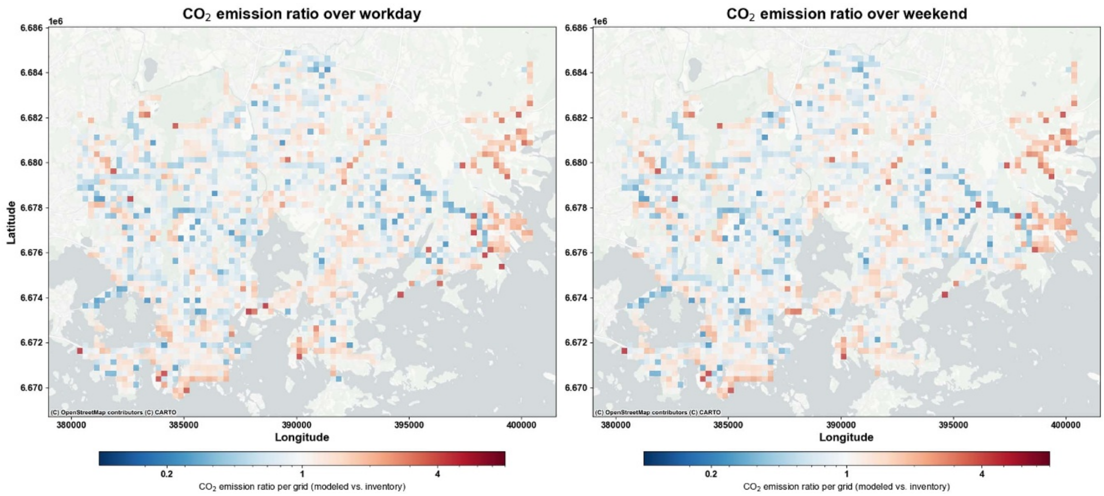

Figure B14. Color-coded emission ratio map of daily exhaust CO_2 modeled by the proposed framework versus the national emission inventory Finnish Regional Emission Scenario model (after adjustment from annual to daily emission). Red colors indicate higher values of modeled emissions while blue colors show vice verse. The left panel illustrates the case on a summer workday while the right panel represents a winter weekend. $N = 3769$.

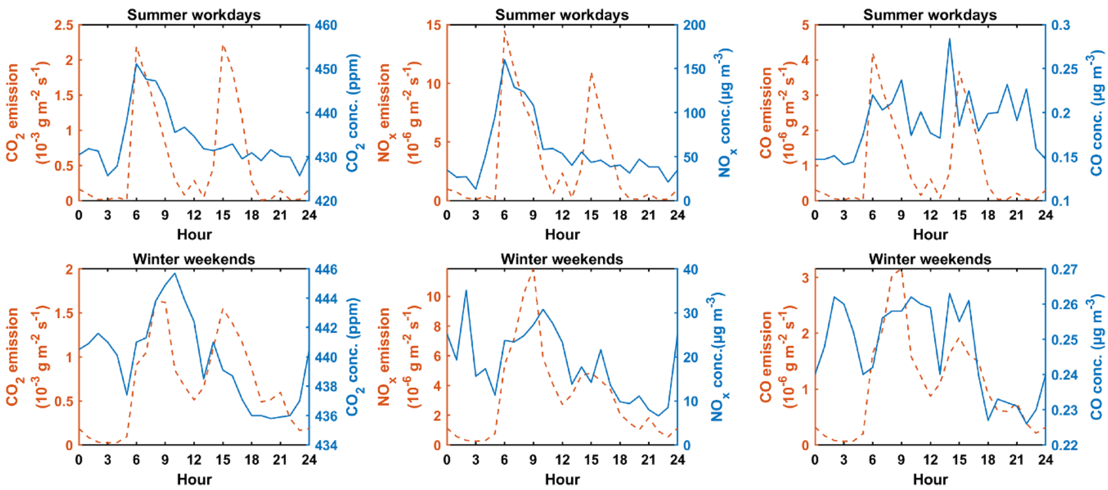

Figure B15. A comparison of our emission products using the ensemble method ENS (orange dotted lines) with hourly concentration measurements (blue solid lines) of CO_2 (first column), NO_x (second column), and CO (third column) at a traffic street canyon in the Helsinki Metropolitan Area (HMA) for both temporal cases (upper panel: summer workdays; lower panel: winter weekends).

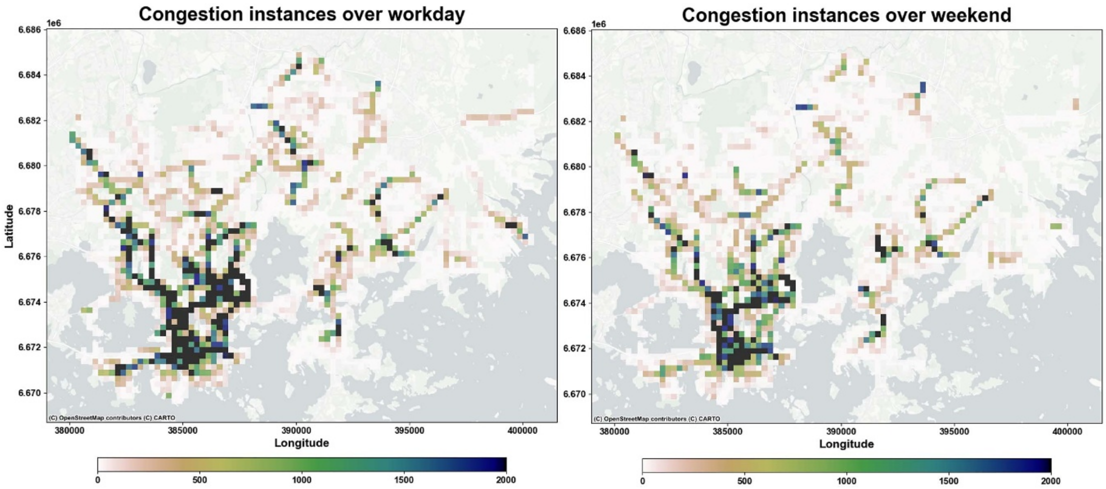

Figure B16. Congestion distribution in the Helsinki Metropolitan Area (HMA) on summer workdays (left panel) and winter weekends (right panel).

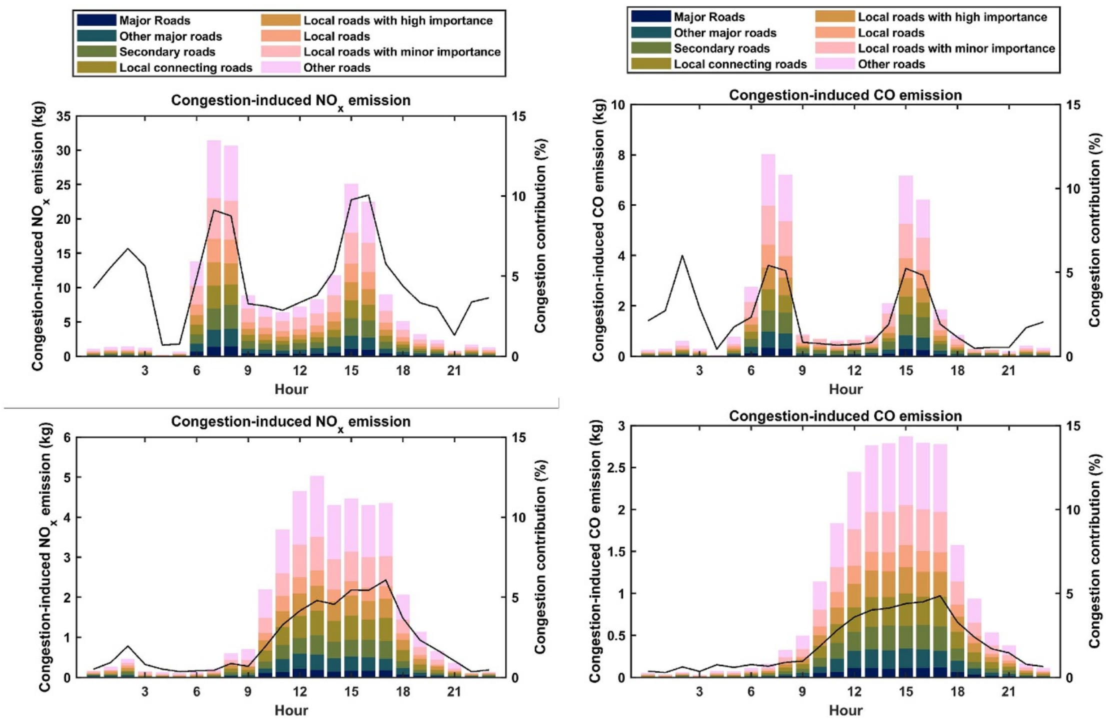

Figure B17. Hourly congestion-induced emissions as bars partitioned color-coded by functional road classes in the Helsinki Metropolitan Area (HMA). The upper panel illustrates the situation on a summer workday while the lower panel displays the winter weekend case. The left column demonstrates the NO_x emissions while the right column shows the CO emissions. Solid lines indicate the contribution of congestion to total on-road emissions.

Table B1. Parameterization (free-flow speed: v_m in kph, critical speed: v_c in kph and capacity volume: q_c in veh/h) for fundamental diagrams of various functional road classes (FRCs) with varying speed limits (SLs, in kph) in summer and winter in the Helsinki Metropolitan Area (HMA). The number of corresponding traffic counters for each road type is denoted by n. The proportion of light-duty vehicles, including passenger cars and light commercial vehicles, is shown under the columns ph% and nph% during peak hours and non-peak hours, respectively.

			Summer workdays						Winter weekends					
	%	SL	v_m	v_c	q_c	n	ph%	nph%	v_m	v_c	q_c	n	ph%	nph%
Major roads	10	60	78	64	1740	2	91	89	75	60	1594	2	93	91
Other major roads	7	80	89	77	3403	17	96	93	93	76	3064	17	96	94
		60	66	60	1868	2			66	58	1719	2		
		80	94	81	3445	8			90	75	3152	8		
		100	110	91	3886	17			106	87	4101	22		
		120	120	102	3424	5			—	—	—	—	—	—
Secondary roads	6	60	74	72	2208	4	96	94	71	60	2111	4	97	95
		70	77	76	3476	4			75	66	3031	4		
		80	85	84	4942	14			82	73	4408	14		
Local cnct. roads	10	60	72	64	1674	6	96	94	70	61	1604	6	96	94
Local roads with high imp.	13	—	53	40	1387	4	92	89	52	42	1229	4	92	89
Local roads	16	—	29	22	379	2	92	94	29	22	248	2	94	96
Local roads with low imp.	10	—	—	—	—	0	—	—	—	—	—	0	—	—
Other roads	28	—	—	—	—	0	—	—	—	—	—	0	—	—

Table B2. The specifications of six selected vehicle types in Helsinki fleet composition (FC1–FC6). For vehicles types, PC, LCV, and HDT stand for passenger cars, light commercial vehicles, and heavy-duty trucks, respectively. The corresponding percentage has been rounded to its nearest integer.

	FC1	FC2	FC3	FC4	FC5	FC6
Vehicle type	PC	PC	LCV	LCV	HDT	HDT
Fuel used	Petrol	Diesel	Petrol	Diesel	Petrol	Diesel
Mass (kg)	1334	1334	1752	1752	6839	6839
Percentage (%)	63	24	<1	10	<1	3

Table B3. Functional Road Class (FRC) defined by TomTom in short and long descriptions [15]. The proportion (%) of FRC in the Helsinki Metropolitan Area (HMA) is calculated. The corresponding percentage has been rounded to its nearest integer.

FRC	%	Short Description	Long Description
1	10	Major Roads; Motorways; Freeways	All roads that are officially assigned as motorways. All roads of high importance, but not officially assigned as motorways, that are part of a connection used for international and national traffic and transport.
2	7	Other Major Roads	All roads used to travel between different neighboring regions of a country.
3	6	Secondary Roads	All roads used to travel between different parts of the same region.
4	10	Local Connecting Roads	All roads making all settlements accessible or making parts (north, south, east, west, and central) of a settlement accessible.
5	13	Local Roads of High Importance	All local roads that are the main connections in a settlement. These are the roads where important through traffic is possible e.g. arterial roads within suburban areas, industrial areas or residential areas, a rural road, which has the sole function of connecting to a national park or important tourist attraction.
6	16	Local Roads	All roads used to travel within a part of a settlement or roads of minor connecting importance in a rural area.
7	10	Local Roads of Minor Importance	All roads that only have a destination function, e.g. dead-end roads, roads inside a living area, alleys: narrow roads between buildings, in a park or garden.
8	28	Other Roads	All other roads that are less important for a navigation system: a road that is too small to be driven by a passenger car, bicycle paths or footpaths that are especially designed as such stairs, pedestrian tunnel, pedestrian bridge, and alleys that are too small to be driven by a passenger car

Table B4. Urban (aka built types, 1–10) and natural (aka land cover types, A–G) Local Climate Zone (LCZ) definitions in short and long descriptions [43]. The proportion (%) of LCZ in the Helsinki Metropolitan Area (HMA) excluding open sea area is calculated. The corresponding percentage has been rounded to its nearest integer.

LCZ	%	Short Description	Long Description
1	1	Compact highrise	Dense mix of tall buildings to tens of stories. Few or no trees. Land cover mostly paved. Concrete, steel, stone, and glass construction materials.
2	1	Compact midrise	Dense mix of midrise buildings (3–9 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.
3	<1	Compact lowrise	Dense mix of lowrise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Stone, brick, tile, and concrete construction materials.
4	4	Open highrise	Open arrangement of tall buildings to tens of stories. Abundance of previous land cover (low plants, trees). Concrete, steel, stone, and glass construction materials.
5	24	Open midrise	Open arrangement of lowrise buildings (3–9 stories). Abundance of pervious land cover (low plants, scattered trees). Concrete, steel, stone, and glass construction materials.
6	15	Open lowrise	Open arrangement of lowrise buildings (1–3 stories). Abundance of pervious land cover (low plants, scattered trees). Wood, brick, stone, tile, and concrete construction materials.
7	0	Lightweight lowrise	Dense mix of single-story buildings. Few or no trees. Land cover mostly hard-packed. Lightweight construction materials (e.g. wood, thatch, corrugated metal).

(Continued.)

Table B4. (Continued.)

LCZ	%	Short Description	Long Description
8	14	Large lowrise	Open arrangement of large lowrise buildings (1–3 stories). Few or no trees. Land cover mostly paved. Steel, concrete, metal, and stone construction materials.
9	<1	Sparsely built	Sparse arrangement of small or medium-sized buildings in a natural setting. Abundance of pervious land cover (low plants, scattered trees)
10	1	Heavy industry	Lowrise and midrise industrial structures (towers, tanks, stacks). Few or no trees. Land cover mostly paved or hard-packed. Metal, steel, and concrete construction materials.
A	24	Dense trees	Heavily wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
B	6	Scattered trees	Lightly wooded landscape of deciduous and/or evergreen trees. Land cover mostly pervious (low plants). Zone function is natural forest, tree cultivation, or urban park.
C	0	Bush, scrub	Open arrangement of bushes, shrubs, and short, woody trees. Land cover mostly pervious (bare soil or sand). Zone function is natural scrubland or agriculture.
D	7	Low plants	Featureless landscape of grass or herbaceous plants/crops. Few or no trees. Zone function is natural grassland, agriculture, or urban park.
E	0	Bare rock or paved	Featureless landscape of rock or paved cover. Few or no trees or plants. Zone feature is natural desert (rock) or urban transportation.
F	<1	Bare soil or sand	Featureless landscape of soil or sand cover. Few or no trees or plants. Zone function is natural desert or agriculture.
G	2	Water	Large, open water bodies such as seas and lakes, or small bodies such as rivers, reservoirs, and lagoons.

Table B5. Reduction in congestion with varying penetration rates of connected and autonomous vehicles (CAVs) using traffic flow simulated for Dublin, Ireland [48, 49].

CAV penetration rate	Main roads	Minor roads	Downtown
25%	60.92%	6.88%	9.48%
50%	62.46%	13.87%	19.47%
75%	63.08%	20.51%	30.95%
100%	63.69%	30.95%	55.74%

Table B6. Evaluation results of driving speed, together with traffic density and gridded emissions using three methods (M1: macroscopic fundamental diagram MFD, M2: generalized linear model GLM, M3: ensemble method ENS) for two studied time (summer workday and winter weekend) against data from fixed traffic counters (all p-value below 0.05). $N = 1512$ for speed and traffic density. $N = 3769$ for the gridded emissions.

		Summer		Winter	
		r	sMAPE	r	sMAPE
Speed (v_{FCD} vs v_T)		0.77	12%	0.74	11%
		Summer		Winter	
		R^2	sMAPE	R^2	sMAPE
M1 – MFD	Density ($\widehat{d}$ vs d_T)	0.60	68%	0.69	53%
	CO_2	0.77	35%	0.76	35%
	NO_x	0.17	45%	0.18	51%
	CO	0.66	77%	0.65	79%
M2 – GLM	Density ($\widehat{d}$ vs d_T)	0.84	40%	0.88	31%
	CO_2	0.70	36%	0.72	36%
	NO_x	0.08	46%	0.08	50%
	CO	0.53	73%	0.61	76%
M3 – ENS	Density ($\widehat{d}$ vs d_T)	0.62	62%	0.72	48%
	CO_2	0.71	35%	0.70	35%
	NO_x	0.15	45%	0.15	47%
	CO	0.58	70%	0.60	72%

Table B7. Coefficients for MFD in different hours (8 a.m. and 4 p.m. for rush hours), k_0 and k_1 for the two studied periods. R^2 as model evaluation is shown. All p-values are below 0.05.

		Summer workdays						Winter weekends					
		8 a.m.			4 p.m.			8 a.m.			4 p.m.		
	SL	k_0	k_1	R^2	k_0	k_1	R^2	k_0	k_1	R^2	k_0	k_1	R^2
Major roads	60	130	77	0.81	45	110	0.12	190	73	0.89	67	89	0.43
	80	130	96	0.34	120	100	0.26	55	99	0.17	64	120	0.13
Other major roads	60	270	65	0.94	94	75	0.48	330	62	0.98	130	70	0.61
	80	130	100	0.35	75	130	0.09	53	100	0.15	50	160	0.03
	100	140	110	0.57	190	110	0.69	50	120	0.19	130	110	0.54
	120	85	130	0.42	130	120	0.86	—	—	—	—	—	—
Secondary roads	60	85	80	0.44	84	93	0.40	56	74	0.36	100	80	0.60
	70	130	86	0.29	210	81	0.41	170	74	0.42	120	89	0.12
	80	180	95	0.29	140	110	0.22	100	92	0.19	99	120	0.05
Local cnct. roads	60	120	72	0.53	110	75	0.52	75	68	0.41	84	75	0.37
Local roads with high imp.	—	130	51	0.95	36	130	0.17	110	48	0.94	47	72	0.17
Local roads	—	22	30	0.43	20	48	0.21	8.7	37	0.13	18	57	0.42

Table B8. The input-adaptive mixed-effects linear regression model (GLM) using adaptive inputs with their corresponding accuracy in terms of R^2 and sMAPE. The overall performance is also included. Note that the speed limit has once been considered, but due to its high collinearity with congestion, it was discarded in all runs shown in this table. All p-values are below 0.05.

	Summer workdays				Winter weekends			
	1	2	3	4	1	2	3	4
Driving speed (v_{FCD})	x	x	x	x	x	x	x	x
Congestion level ($\overrightarrow{v}$)	x	x	x	x	x	x	x	x
FRC	x	x			x	x		
Hour	x	x	x	x	x	x	x	x
LCZ	x		x		x		x	
R^2	0.82	0.80	0.76	0.73	0.86	0.85	0.79	0.80
sMAPE	43%	45%	48%	55%	33%	36%	39%	39%
Overall R^2	0.84				0.88			
Overall sMAPE	40%				31%			

ORCID iDs

Pak Lun Fung https://orcid.org/0000-0003-3493-1383
Omar Al-Jaghbeer https://orcid.org/0000-0002-1077-5526
Jia Chen https://orcid.org/0000-0002-6350-6610
Ville-Veikko Paunu https://orcid.org/0000-0002-3466-4169
Shaghayegh Vosough https://orcid.org/0000-0003-0197-8214
Claudio Roncoli https://orcid.org/0000-0002-9381-3021
Leena Järvi https://orcid.org/0000-0002-5224-3448

References

[1] European Environment Agency 2022 Air quality in Europe 2022 (available at: www.eea.europa.eu/publications/air-quality-in-europe-2022) (Accessed 10 June 2024)
[2] National Atmospheric Emission Inventory 2023 Pollution information (available at: https://naei.beis.gov.uk/overview/pollutants?pollutant_id=4) (Accessed 10 June 2024)
[3] Monjardino J, Dias L, Fortes P, Tente H, Ferreira F and Seixas J 2021 *Atmosphere* **12** 324
[4] Zhang Q *et al* 2023 *Environ. Sci. Ecotechnol.* **16** 100280
[5] Rodriguez-Rey D, Guevara M, Linares M P, Casanovas J, Armengol J M, Benavides J, Soret A, Jorba O, Tena C and García-Pando C P 2022 *Sci. Total Environ.* **807** 150743
[6] IPCC 2019 *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories* ed E C Buendia, K Tanabe, A Kranjc, J Baasansuren, M Fukuda, S Ngarize, A Osako, Y Pyrozhenko, P Shermanau and S Federici (IPCC) ch 6 (available at: www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/1_Volume1/19R_V1_Ch06_QA_QC.pdf)
[7] Madziel M 2023 *Energies* **16** 3941
[8] Yang D, Zhang Y, Wu R, Song Y and Xu H 2022 A review of on-road vehicle emission inventory *E3S Web Conf.* **360** 01027
[9] Alvarado-Molina M, Curto A, Wheeler A J, Tham R, Cerin E, Nieuwenhuijsen M, Vermeulen R and Donaire-Gonzalez D 2023 *Environ. Pollut.* **338** 122657
[10] Dietrich F, Chen J, Voggenreiter B, Aigner P, Nachtigall N and Reger B 2021 *Atmos. Meas. Tech.* **14** 1111–26
[11] Le Quéré C *et al* 2020 *Nat. Clim. Change* **10** 647–53
[12] Gensheimer J, Chen J, Turner A J, Shekhar A, Wenzel A and Keutsch F N 2021 *J. Geophys. Res.: Atmos.* **126** e2021JD034664
[13] Oda T, Haga C, Hosomi K, Matsui T and Bun R 2021 *Environ. Res. Lett.* **16** 084058
[14] Liu Z *et al* 2020 *Sci. Data* **7** 392
[15] TomTom Navigation 2024 *Introduction Traffic API TomTom Developer Portal* (available at: https://developer.tomtom.com/traffic-api/documentation/product-information/introduction/) (Accessed 25 September 2024)
[16] Kellner F 2016 *Logist. Res.* **9** 21
[17] Rosero F, Fonseca N, López J M and Casanova J 2021 *Appl. Energy* **282** 116195
[18] Wei X, Ren Y, Shen L and Shu T 2022 *Environ. Impact Assess. Rev.* **95** 106808
[19] Shang W L, Song X, Chen Y, Yang X, Liang L, Deveci M, Cao M, Xiang Q and Yu Q 2024 *Urban Clim.* **53** 101794
[20] Altintasi O, Tuydes-Yaman H and Tuncay K 2017 *Transp. Res. Proc.* **22** 382–39
[21] Shaygan M, Meese C, Li W, Zhao X G and Nejad M 2022 *Transp. Res.* C **145** 103921
[22] Yang D, Zhang S, Niu T, Wang Y, Xu H, Zhang K M and Wu Y 2019 *Atmos. Chem. Phys.* **19** 8831–43
[23] Wu L, Chang M, Wang X, Hang J, Zhang J, Wu L and Shao M 2020 *Geosci. Model Dev.* **13** 23–40
[24] Wen Y, Zhang S, Zhang J, Bao S, Wu X, Yang D and Wu Y 2020 *Appl. Energy* **260** 114357
[25] Meng X, Zhang K, Pang K and Xiang X 2020 *Sci. Total Environ.* **709** 136227
[26] Li X, Gu D, Hohenberger T L, Fung Y H, Fung J C, Lau A K and Liang Z 2023 *Atmos. Environ.* **313** 120059
[27] Brockfeld E, Sohr A, Wagner P and Gartner N H 2008 Fundamental diagram on urban roads: myth or truth? *75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symp.* (TRB Transportation Research Board) p 158 (available at: https://trid.trb.org/View/1112875)
[28] Zhan X, Zheng Y, Yi X and Ukkusuri S V 2016 *IEEE Trans. Knowl. Data Eng.* **29** 272–85
[29] Geroliminis N and Daganzo C F 2008 *Transp. Res.* B **42** 759–70
[30] Underwood R T 1961 Speed, volume and density relationships *Quality and Theory of Traffic Flow* (Bureau of Highway Traffic, Yale University) pp 141–88 (available at: https://trid.trb.org/View/115231)
[31] Greenshields B D, Bibbins J, Channing W and Miller H 1935 A study of traffic capacity *Highway Research Board Proc. (Washington, DC)* vol 14 pp 448–77 (available at: https://onlinepubs.trb.org/Onlinepubs/hrbproceedings/14/14P1-023.pdf)
[32] Ke P *et al* 2023 *Sci. Data* **10** 374
[33] Wen Y, Wu R, Zhou Z, Zhang S, Yang S, Wallington T J, Shen W, Tan Q, Deng Y and Wu Y 2022 *Appl. Energy* **305** 117916
[34] Al-Jaghbeer O, Fung P L, Paunu V V and Järvi L 2024 *Urban Clim.* **58** 102171

[35] Apronti D, Ksaibati K, Gerow K and Hepner J J 2016 *J. Traffic Transp. Eng.* **3** 493–506
[36] Liu S V, Chen F L and Xue J 2017 *Int. J. Environ. Res. Public Health* **14** 1581
[37] Fung P L *et al* 2022 *Atmos. Chem. Phys.* **22** 1861–82
[38] Cohn N 2009 *Transp. Res. Rec.* **2129** 129–35
[39] Finnish Transport and Communications Agency 2023 Open data for vehicles contains registration, approval and technical information for all vehicles in use from the register maintained by Traficom (available at: https://tieto.traficom.fi/en) (Accessed 10 June 2024)
[40] Fung P L, Al-Jaghbeer O, Pirjola L, Aaltonen H and Järvi L 2023 *Sci. Total Environ.* **901** 165827
[41] Karvosenoja N 2008 Emission scenario model for regional air pollution *PhD Thesis* Finnish Environment Institute (available at: http://hdl.handle.net/10138/39332)
[42] Paunu V V *et al* 2024 *Earth Syst. Sci. Data* **16** 1453–74
[43] Demuzere M, Kittner J, Martilli A, Mills G, Moede C, Stewart I D, van Vliet J and Bechtel B 2022 *Earth Syst. Sci. Data* **14** 3835–73
[44] Pan Y A, Guo J, Chen Y, Cheng Q, Li W and Liu Y 2024 *Expert Syst. Appl.* **238** 122219
[45] Matzer C, Weller K, Dippold M, Lipp S, Röck M, Rexeis M and Hausberger S 2019 Update of emission factors for HBEFA Version 4.1 *Final Report* I-05/19/CM EM-I-16/26/679 (TU Graz) (available at: https://cdn.prod.website-files.com/6207922a2acc01004530a67e/625e8d14b70af84fba1ef10c_HBEFA41_Report_TUG_09092019.pdf)
[46] Ntziachristos L and Samaras Z 2023 EMEP/EEA air pollutant emission inventory guidebook 2023 *Technical Report* (EMEP/EEA) (available at: www.eea.europa.eu/publications/emep-eea-guidebook-2023/part-b-sectoral-guidance-chapters/1-energy/1-a-combustion/1-a-3-b-i)
[47] Borge R, de Miguel I, de la Paz D, Lumbreras J, Pérez J and Rodríguez E 2012 *Atmos. Environ.* **62** 461–71
[48] Gueriau M and Dusparic I 2020 Quantifying the impact of connected and autonomous vehicles on traffic efficiency and safety in mixed traffic *23rd IEEE Int. Conf. on Intelligent Transportation Systems*
[49] Li Y, Li H, Martinez-Pastor B and Wang S 2021 Can CAV reduce non-recurrent urban road congestion? (arXiv:2110.08507)
[50] Deng F, Lv Z, Qi L, Wang X, Shi M and Liu H 2020 *Nat. Commun.* **11** 2801
[51] Hakkim H, Kumar A, Annadate S, Sinha B and Sinha V 2021 *Atmos. Environ.* X **11** 100118
[52] Cai Y, Zeng X, Li W, He S, Feng Z and Tan Z 2024 *Sustainability* **16** 3504
[53] Bergroth C, Järv O, Tenkanen H, Manninen M and Toivonen T 2022 *Sci. Data* **9** 39
[54] McConky K and Rungta V 2019 *Transp. Res.* C **100** 289–305
[55] Biramo Z and Mekonnen A 2022 *Environ. Syst. Res.* **11** 28
[56] Silva O, Cordera R, González-González E and Nogués S 2022 *Sci. Total Environ.* **830** 154615
[57] Robert M 2024 *Google Traffic and MapBox* (available at: https://dime-worldbank.github.io/googletraffic/articles/googletraffic-vignette.html) (Accessed 25 September 2024)
[58] Fung P L, Al-Jaghbeer O, Chen J, Paunu V-V, Vosough S, Roncoli C and Järvi L 2025 *Zenodo* (https://doi.org/10.5281/zenodo.13861952)
[59] Finnish Road Safety Council 2024 Difficult driving conditions (available at: www.liikenneturva.fi/en/road-safety/difficult-driving-conditions/) (Accessed 10 June 2024)