



OPEN ACCESS

EDITED BY

Bushra Khalid,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Waishan Qiu,
The University of Hong Kong, Hong Kong, SAR
China
Kevin Gurney,
Northern Arizona University, United States

*CORRESPONDENCE

Max Anjos,
✉ maxanhos@campus.ul.pt

RECEIVED 08 July 2024

ACCEPTED 18 December 2024

PUBLISHED 07 January 2025

CITATION

Anjos M and Meier F (2025) Zooming into Berlin: tracking street-scale CO₂ emissions based on high-resolution traffic modeling using machine learning.
Front. Environ. Sci. 12:1461656.
doi: 10.3389/fenvs.2024.1461656

COPYRIGHT

© 2025 Anjos and Meier. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Zooming into Berlin: tracking street-scale CO₂ emissions based on high-resolution traffic modeling using machine learning

Max Anjos* and Fred Meier

Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Berlin, Germany

Artificial Intelligence (AI) tools based on Machine learning (ML) have demonstrated their potential in modeling climate-related phenomena. However, their application to quantifying greenhouse gas emissions in cities remains under-researched. Here, we introduce a ML-based bottom-up framework to predict hourly CO₂ emissions from vehicular traffic at fine spatial resolution (30 × 30 m). Using data-driven algorithms, traffic counts, spatio-temporal features, and meteorological data, our model predicted hourly traffic flow, average speed, and CO₂ emissions for passenger cars (PC) and heavy-duty trucks (HDT) at the street scale in Berlin. Even with limited traffic information, the model effectively generalized to new road segments. For PC, the Relative Mean Difference (RMD) was +16% on average. For HDT, RMD was 19% for traffic flow and 2.6% for average speed. We modeled seven years of hourly CO₂ emissions from 2015 to 2022 and identified major highways as hotspots for PC emissions, with peak values reaching 1.639 kgCO₂ m⁻² d⁻¹. We also analyzed the impact of COVID-19 lockdown and individual policy stringency on traffic CO₂ emissions. During the lockdown period (March 15 to 1 June 2020), weekend emissions dropped substantially by 25% (−18.3 tCO₂ day⁻¹), with stay-at-home requirements, workplace closures, and school closures contributing significantly to this reduction. The continuation of these measures resulted in sustained reductions in traffic flow and CO₂ emissions throughout 2020 and 2022. These results highlight the effectiveness of ML models in quantifying vehicle traffic CO₂ emissions at a high spatial resolution. Our ML-based bottom-up approach offers a useful tool for urban climate research, especially in areas lacking detailed CO₂ emissions data.

KEYWORDS

artificial intelligence, machine learning, carbon accounting, urban climate, COVID-19

1 Introduction

The transportation sector is one of the major contributors to global carbon dioxide (CO₂) emissions from fossil fuels, with road vehicles alone accounting for three-quarters of the emissions in this sector (EEA, 2017; IEA, 2019). This contribution is particularly pronounced in urban areas, where high concentration of vehicles and increased travel distances result in large CO₂ emission levels, making road transportation an important component of city carbon accounting (Gately et al., 2015; Gurney et al., 2012; Huo et al., 2022; Nangini et al., 2019). However, with a few exceptions, cities still face challenges in

accurately quantifying traffic CO₂ emissions at a high spatio-temporal resolution, such as rush hours, daily-seasonal circles, holidays, and shifts in mobility routines at street scale.

At the city scale, understanding where, when, and how much emissions happen is crucial for informed climate actions, detailed greenhouse gas monitoring and inventories (Duren and Miller, 2012; Jungmann et al., 2022; Ku et al., 2022; Roest et al., 2020; Seto et al., 2021; Turnbull et al., 2022). The Hestia project (Gurney et al., 2019) exemplifies this endeavor by focusing on the creation of high-resolution CO₂ emission datasets through street-level modeling in cities across the United States. However, such initiatives are still exceptions, as many cities worldwide struggle to provide high spatiotemporal estimates of traffic CO₂ emissions for their entire road networks due to limited monitoring stations. While many cities have traffic monitoring systems in road segments with significant traffic volumes, these systems often do not cover all roads in the network. This incomplete data coverage makes it difficult for cities to accurately quantify high spatiotemporal CO₂ emissions and assess their carbon footprint, leading to often under-reported urban emission inventories (Gurney et al., 2021).

Recent advancements in Artificial Intelligence (AI) have demonstrated significant potential for improving traffic prediction accuracy (Shaygan et al., 2022). Machine Learning (ML) models, a subset of AI capable of identifying patterns in large, complex datasets (Aurélien Geron, 2022; Chollet, 2018; Kuhn and Johnson, 2013), have effectively captured spatial and temporal correlations in big data (Lv et al., 2014), leading to accurate and timely predictions of traffic indicators such flow, speed, and accident risk (Liu et al., 2018). While ML models show promise in applications related to urban road traffic, the quantification of greenhouse gas emissions at high spatial-temporal resolution remains relatively under-researched.

Here, the core idea of our study is to estimate vehicle traffic information, specifically focusing on extrapolating traffic flow and speed predictions from road segments with available data to those without, considering similarities in both location and time. These variables are essential for calculating CO₂ emissions in urban environments. By utilizing ML algorithms and bottom-up approach, we can predict hourly traffic flow, average speed, and CO₂ emissions at the street scale for the entire road network. Built upon Geographic Information System, road infrastructure data, meteorological conditions, and local traffic measurements, our ML-based bottom-up model identifies truly predictive traffic CO₂ patterns with a grid of 30 m in horizontal resolution and captures temporal changes from hourly to yearly scales.

A key aspect of the proposed ML-based bottom-up is that it can be used particularly in areas where data on traffic CO₂ emissions is scarce, offering new possibilities for city-scale carbon accounting. It also aligns with the existing practices of environment researchers and authorities who consider hourly and zoomed-in emission maps useful to construct detailed and gridded vehicular emission inventories (Huo et al., 2009; Wang et al., 2010). Hence, the primary goal of this paper is to demonstrate the ML-based bottom-up application in estimating the spatiotemporal variability of road traffic CO₂ emissions from passenger cars and heavy-duty- trucks in Berlin, Germany. Specifically, our aims are as follows.

1. Introducing a ML-based bottom-up to estimate hourly traffic flow and average speed at the street scale across an entire city, along with evaluating modeled estimates using local traffic measurements. We also identified key spatial and temporal features that influence the final ML predictions;
2. Mapping spatio-temporal variability of CO₂ emissions by vehicle type at the street scale with an hourly 30-m grid resolution from 2015 to 2022. We compared these estimates with data from the Carbon Monitor Cities dataset, which provides CO₂ emission from the road transportation sector (Huo et al., 2022).
3. Exploring the impact of COVID-19 lockdown and various response measures, such as school closures, travel restrictions, workplace closures, and stay-at-home requirements, on traffic CO₂ emission behavior. We will merge our emissions data with the Oxford pandemic policies database (Hale et al., 2021) to analyze the effectiveness of different governmental measures in reducing emissions in cities.

2 Data and methods

2.1 Target city

Berlin, Germany's capital, boasts a population exceeding 3.7 million inhabitants and spans an area of approximately 892 km² (Statistical Office of Berlin-Brandenburg, 2019). As one of the European Union's largest cities, Berlin showcases diverse transportation modes, including roads, rails, and aviation. These various networks contribute to substantial CO₂ emissions levels, with transportation sector ranking as the second-largest emitter (after buildings), responsible for 20% of the city's total CO₂ emissions and road traffic counting 70% of this share (Senuvk (Senate Department for the Environment, U. M., Consumer Protection and Climate Action), 2019; Hirschl and Harnisch, 2016). In 2019, the city had 330 cars per 1.000 residents, distributed across the road network (SenStadtWohn, 2019).

Aligned with global trends, Berlin is committed to achieving climate neutrality by 2050. To achieve this goal, the city has instituted the Berlin Energy and Climate Protection Program, encompassing numerous strategies for reducing emissions across sectors, including road transportation (Hirschl and Harnisch, 2016). In our study, we employ the proposed ML model to estimate the direct CO₂ emissions from road transportation that physically occurs within the city's boundary, corresponding to scope 1 for emissions accounting and reporting (refer to Chen et al. (2019) for details on scopes 1, 2 and 3).

2.2 Dataset description

The proposed ML-based bottom-up model (ML model) integrates three primary datasets: local traffic measurements, spatial information, and meteorological conditions. Local traffic data were sourced from the Digital Platform City Traffic Berlin/Traffic Detection Berlin (<https://api.viz.berlin.de/daten/verkehrsdetektion>) and Bundesanstalt für Straßenwesen (BAST) (https://www.bast.de/DE/Verkehrstechnik/Fachthemen/v2-verkehrszahlung/zaehl_node.html). The Digital

platform provides hourly vehicle volume and average speed data from 583 lane-specific detectors at counting stations, for passenger cars (PC) and heavy-duty trucks (HDT). The BAST dataset includes hourly vehicle volumes for different vehicle types, including PC and HDT, from 17 counting stations on motorways and non-urban federal roads in Berlin. In this study, the data covers from January 2015 to December 2022.

Spatial information was gathered from OpenStreetMap (OSM) and Berlin Digital Environmental Atlas (Berlin Atlas). OSM, a global collaborative project, provides crowdsourced Geographic Information Voluntary ([OpenStreetMap contributors, 2017](https://www.openstreetmap.org/wiki/Map_features)). OSM features such as road types (motorway, trunk, primary, secondary, and tertiary), leisure, land use, amenity, building types, and more were utilized ([https://wiki.openstreetmap.org/wiki/Map_features](https://www.openstreetmap.org/wiki/Map_features)). Berlin Atlas contributed GIS features, including land use, population density, and daily mean traffic volumes in 2019 ([SenUVK, 2021](https://www.senuvk.de)). The shapefile of Berlin's road network, containing road length, speed limit, and road classification based on Functional Road Class (FRC), was obtained from TomTom's Historical Traffic Stats (<https://www.tomtom.com/products/traffic-stats/>). FRC reflects the road importance based on traffic volume, speed and connectivity, enhancing the model's reproducibility under TomTom's non-commercial usage permission.

To complete the ML model's variables set, hourly meteorological data (air temperature, relative humidity, sunshine, rainfall, wind direction, and wind speed) were acquired from the weather station Berlin-Dahlem (latitude 52.4537, longitude 13.3017) managed by the German Weather Service Climate Data Center ([DWD, 2020](https://www.dwd.de)).

2.3 Data preparation

The ML model was built with the open-source R statistical computing platform ([R Core Team, 2018](https://www.r-project.org/)). In the R environment, the data preparation step ensures the dataset is properly formatted, cleaned, and prepared for analysis. A key task involves defining the dependent variables and independent variables (predictors). In this study, we defined the mean traffic flow per hour and average speed in km per hour from all counting stations at a road segment represented by a line (link) in shapefile format, as dependent variables.

We geographically linked traffic count points with road network segments, aggregating traffic flow and average speed and independent variables, such as OSM features, population density and daily mean traffic volumes using the `st_join` and `st_nearest_feature` functions from R `sf` package ([Pebesma, 2018](https://r-spatial.org/r/packages/sf/)). We then separated all road segments into two categories: "sampled," referring to those covered by traffic count points, and "non-sampled," indicating those without such coverage (see [Supplementary Figure S1](#) in the Supplementary Information). Subsequently, we applied a set of data pre-processing techniques, called feature engineering ([Kuhn and Johnson, 2019](https://www.kuhnandjohnson.com/)) to build good predictors. It involves handling missing data in both numeric and categorical variables, as well as performing data transformations to extract useful information (new predictors) and select potential ones. For example, temporal predictors such as time of day, weekdays, weekends, and indicators for holidays were derived from the

date-time column (e.g., 2021-01-01 01:00:00) using the `step_timeseries_signature` function from `recipes` and `timetk` R packages ([Kuhn and Wickham, 2023](https://www.kuhnandjohnson.com/); [Dancho and Vaughan, 2023](https://www.danchoandvaughan.com/)).

The ML model consisted of 35 spatio-temporal predictors that can represent the traffic flow and average speed estimates at the street scale. See [Supplementary Table S1](#) for more details on dependent variables and each predictor. Thus, the idea behind the ML model is that the algorithm learns from the independent variables of sampled roads (measured) to predict the dependent variables for non-sampled (unmeasured) roads.

2.4 Model development

To ensure the robustness and generalisability of the ML model ([Chollet, 2018](https://www.cholletandjohnson.com/); [Kuhn and Johnson, 2013](https://www.kuhnandjohnson.com/)), we divided our dataset into three sets: training, validation, and test. Initially, we randomly attributed 80% (440) of our traffic count stations and respective sampled road segments to the training set, 10% (148) to the validation, and 10% (147) to the test set. We made sure this fraction split across different road types to ensure a representative sample, as shown in [Supplementary Figure S2](#). To assess the model's performance across seasons, days of the week, and rush hours, we chronologically split the traffic data for 4 months (February, July, and November) in 2018 so that earliest 80% of each month data was assigned to training, the next 10% for validation and the latest 10% for testing.

We used the Random Forest (RF), a popular ensemble learning algorithm known for its ability to combine a large number of decision trees for classification or regression tasks ([Breiman, 2001](https://www.breiman.com/)). RF has been widely used in the fields of traffic demand and air-pollutant research ([Liu and Wu, 2017](https://www.liuandwu.com/); [Wen et al., 2022](https://www.wenetal.com/)). The RF algorithm was implemented as a supervised regression task using the R `ranger` package ([Wright and Ziegler, 2017](https://www.wrightandziegler.com/)).

We trained the RF iteratively on the testing set, making adjustments to hyperparameters as needed ([Kuhn and Johnson, 2019](https://www.kuhnandjohnson.com/)). Once validated, the model was tested on the unseen test set in order to evaluate its performance in real-world situations. Default settings of the RF used to predict traffic flow and average speed at the street scale are shown in [Supplementary Table S2](#). To simplify the synthesis of our model's performance, we applied the following metrics: r (correlation coefficient - Pearson), root mean square error (RMSE), mean absolute error (MAE), and relative mean difference (RMD).

To measure how much each feature contributes to the traffic flow and average speed predictions, we employed the permutation method ([Wright and Ziegler, 2017](https://www.wrightandziegler.com/)). This technique involves randomly shuffling the values of a specific feature and measuring the resulting change in model performance. The permutation method is particularly useful for understanding the importance of features in black-box models, such as RF, which can be difficult to interpret directly. However, it may be computationally expensive, especially for large datasets and complex models, such as RF.

2.5 Deploy model

After training the ML model for each PC and HDT categories individually, we deployed it to predict hourly traffic flow and average

speed for each road segment. Using these predictions, we calculated the corresponding hourly CO₂ emissions (E_{hi}) in g km⁻² h⁻¹ for the PC and HDT, following Equation 1 (Stagakis et al., 2023):

$$E_{i,l,h} = \sum_{i=1}^n \frac{q_{i,l,h} * L_l * EF_{v,i}}{As} \quad (1)$$

where $q_{i,l,h}$ is the modeled traffic flow of vehicle category i on road link l at hour h , L is the link-road length in km⁻¹, As is the area of the road segment in m⁻², and $EF_{v,i}$ is the emission factor for of vehicle category i at average speed v in g km⁻¹ (see Supplementary Table S3 for more details on EF functions). The speed-dependence EF for PC and HDT were derived from the European Road Transport Emission Inventory Model - COPERT, which offers EF that are expressed as functions of the mean traveling speed over a complete driving cycle, taking into account specific vehicle types, vehicle fleet layers differentiated by size, technology, and emission standards (Ntziachristos et al., 2009). This approach enables accurate estimation of emissions under various real driving emissions. We chose the COPERT model due to its methodology aligning with both the EMEP/EEA air pollutant emission inventory guidebook 2023 for road emission calculations in the European context (EEA, 2023), as well as the guidelines provided by the Intergovernmental Panel on Climate Change (IPCC), which utilizes CO₂ EF related to fuel consumption factors and distance traveled to estimate fuel usage (IPCC, 2000).

To obtain a detailed E_{Ti} map, we aggregated the estimated emissions to 30 × 30 m grid cells, a scale that is useful for informed climate actions in urban areas (Christen, 2014). To do this, we resampled the data to a 30-meter resolution land cover raster. We then intersected the road segment links (line sources) with the grid. CO₂ emissions within each grid cell were calculated by multiplying the fraction of intersected line values by the original length, ensuring accurate estimates without underestimation.

To gain a preliminary understanding of the performance of our CO₂ emission estimates, we compared them with data from the Carbon Monitor Cities (CM-Cities) dataset, which provides near-real-time daily CO₂ emissions estimates by various sectors across cities globally (<https://cities.carbonmonitor.org/>; Huo et al., 2022). For the ground transportation sector, the CM-Cities uses TomTom daily transport congestion and Emissions Database for Global Atmospheric Research (EDGAR) data as inputs, and provides daily CO₂ emission estimates covering the city-scale. Hence, daily CO₂ emissions from Berlin's ground transport sector in 2019 were matched with our ML model's daily CO₂ emissions for the same period. Both estimates were aligned as total city-scale emissions in the unit of thousands of metric tons per day for the entire Berlin area.

2.6 COVID-19 data and analysis

To assess the magnitude and timing of COVID-19 lockdown effects on traffic-related CO₂ emissions, we analyzed changes in 7-day running mean and weekday mean values during two distinct periods: the lockdown and a baseline period. The lockdown period was identified based on significant reductions in the mobility patterns of Berlin residents, as reported by Schatke et al. (2022).

For this analysis, we focused on the spring lockdown, spanning from March 15 to 1 June 2020. The baseline period, serving as a pre-pandemic reference, consisted of the corresponding calendar days in 2019, which were unaffected by COVID-19 restrictions.

To further assess the relationship between COVID-19 lockdown and various response measures on traffic CO₂ emissions in Berlin, we used data from the Oxford COVID-19 Government Response Tracker (OxCGRT) dataset (Hale et al., 2021). The OxCGRT dataset includes 21 indicators for target policies, categorized into five groups: containment and closure, economic, health system, vaccination, and miscellaneous. Given their direct influence on mobility patterns, this study particularly focused on the impact of containment and closure policies: school closures, workplace closures, restrictions on gatherings, stay-at-home requirements, internal movement restrictions, and international travel controls. Each policy index ordinarily ranges from 0 to 4, with higher values reflecting more stringent measures.

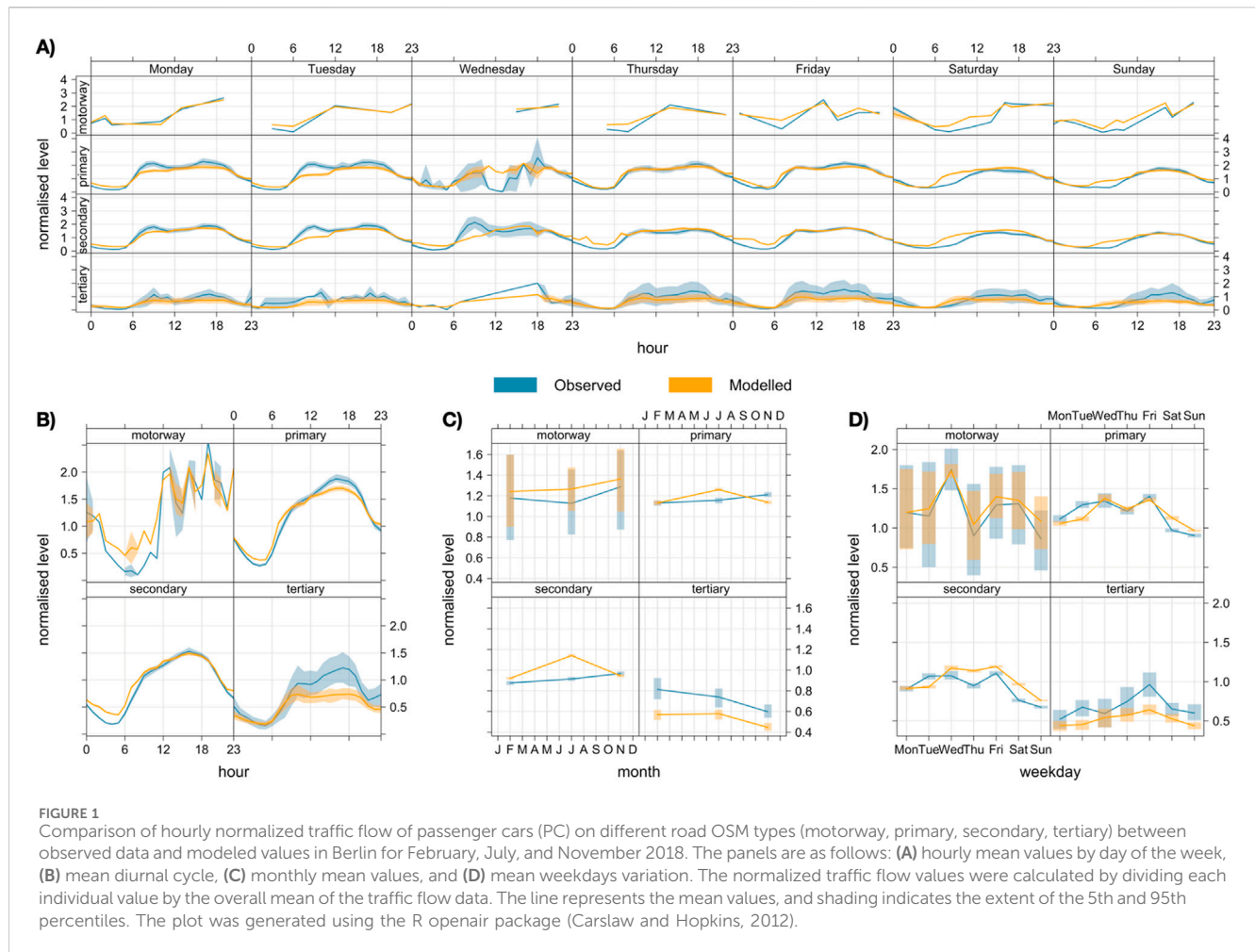
To analyze the statistical association between OxCGRT containment and closure policies on daily CO₂ emissions during the lockdown period (March 15 to 1 June 2020), we employed a combination of Spearman correlation and Partial Least Squares (PLS) regression model. Spearman correlation was used to assess the strength and direction of the relationship between individual policy measures and CO₂ emissions. In this study, PLS identified the most influential policy measures driving changes in daily CO₂ emissions. PLS combines features of principal component analysis and multiple regression, making it particularly suitable for datasets with multicollinearity among predictors (e.g., government policies), as it simultaneously reduces dimensionality and explains variance in the dependent variable (e.g., CO₂ emissions) (Geladi and Kowalski, 1986). The use of PLS was supported by multicollinearity diagnostics, which revealed moderate (>5) and high (>10) Variance Inflation Factor (VIF) values among the response measures (see Supplementary Table S4).

3 Results

3.1 Assessing the ML model predictions

To evaluate our ML model's performance, we deployed the trained RF algorithm to an independent dataset. The model demonstrated satisfactory results in predicting road traffic volumes and average speed for the main OSM road types in Berlin. For PC, the modeled hourly traffic flow values ($N = 29,300$) exhibit consistency with observed values ($r = 0.73$), with an RMSE of 155 veh h⁻¹ and an MAE of 111 veh h⁻¹. The RMD, calculated as [(modeled_value - observed_value)/mean(modeled_value, observed_value)], was +16% on average, indicating an overestimation. For average speed, the model performed similarly, with RMSE of 10 km h⁻¹, MAE of 6.8 km h⁻¹, and RMD of -3.6%. For HDT, the model's performance was comparable to that for PC, with similar RMD values for traffic flow (19%) and average speed (2.6%). Please refer to Supplementary Table S5 for a detailed breakdown of all metrics.

Figure 1 illustrates the model's performance in representing temporal traffic flow patterns, including mean diurnal cycles, rush hours, day of the week, and monthly variations, across different road



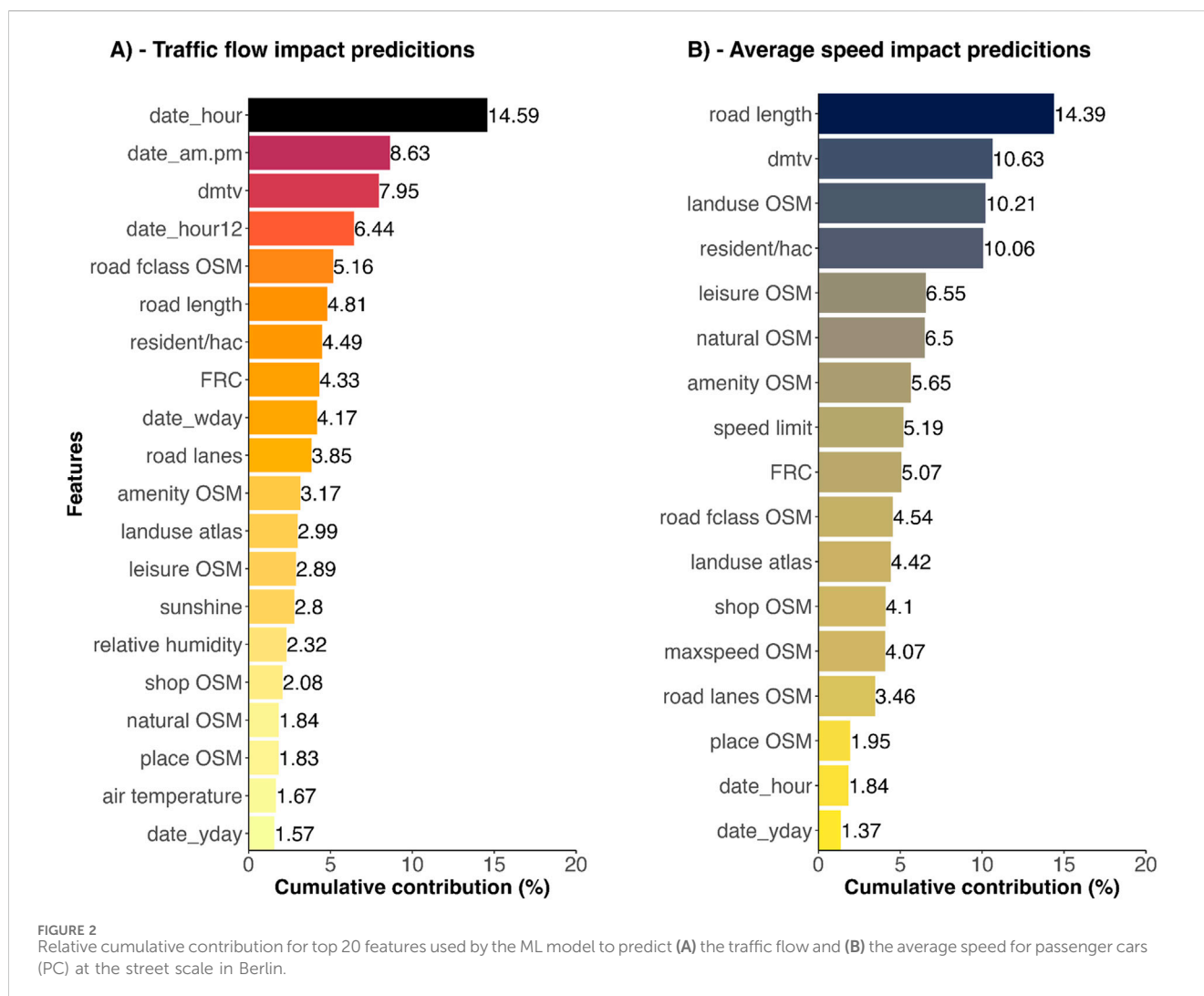
types for PC (see [Supplementary Figure S3](#) for HDT). The model exhibits some biases, such as underestimating traffic flow during afternoon hours on tertiary roads and overestimating traffic flow during morning hours on motorway and secondary roads ([Figure 1B](#)). These biases persist on certain days of the week ([Figures 1A, D](#)). These errors may be attributed to the limited sample size of road segments used for training (see [Supplementary Figure S2](#)), which hinders the model's ability to capture traffic variability, particularly during peak hours (including Wednesdays), as well as unmodeled external factors. Regarding monthly variations ([Figure 1C](#)), the models performed well in estimating the average seasonality of traffic flow across different road types.

[Figure 2](#) shows the key predictors that impact traffic flow and average speed predictions at the street scale. Temporal factors played an important role in the traffic flow predictions. Date_hour (hour of the day) the most influential feature (14.6%), highlighting the temporal aspect in shaping traffic flow estimates. Other relevant temporal features include date_am.pm (daytime/nighttime) with 8.6%, and date_hour12 (12 h of day period) with 6.4%. Dtmv (daily mean traffic volume) also influenced traffic flow prediction with 7.9%. Other spatial features also contribute to traffic flow predictions, although with less impact than temporal features. These include road types (fclass OSM and FRC), road length, and population density (resident/hac), each contributing around 5.0%.

For average speed predictions ([Figure 2B](#)), road length is the most important factor (13.2%). Dtmv and landuse OSM are also influential, along with population density (all around 10%). These results align with prior research ([Medina-Salgado et al., 2022](#)) that highlights the effectiveness of ML models that incorporate both spatial and temporal features in accurately predicting traffic estimates.

3.2 Spatiotemporal CO₂ emission patterns

The most significant outcome of the ML model is its ability to provide detailed, high-spatial data on vehicle traffic CO₂ emissions, as mapped in [Figure 3](#). The model was executed for a specific winter workday, estimating CO₂ emissions on each street segment in the city, totalling 27 thousand modeled links. As anticipated, the model estimated large emissions for PC along major highways across the city, reaching a peak value of 1.639 kgCO₂ m⁻² d⁻¹ on motorway roads. Panels B and C on the map present the time series of daily CO₂ emissions for PC from 2015 to 2022 for two specific streets: AVUS and Straße des 17. Juni. They highlight the different emission patterns observed on these streets over the years. AVUS, a major motorway-highway, shows higher emissions than Straße des 17. Juni, a primary road located in a central urban area.



3.3 Temporal CO₂ emission variation

Figure 4 illustrates the temporal variability of road traffic CO₂ emissions from PC and HDT for the entire city-level. The total emissions were calculated by summing up the emissions from all modeled road segments within Berlin's administrative boundaries (covering an area of 892 km²) in 2019. The upper plot (Figure 4A) depicts the diurnal patterns of hourly mean emissions for all days of the week. Significant peaks occur during rush hours, averaging around 140 tCO₂ h⁻¹ for PC and 70 tCO₂ h⁻¹ for HDT at 12:00 and 18:00, accompanied by a reduction in emissions on weekends. The lower-right plot (Figure 4D) shows a contrast between weekdays and weekends, with a reduction of 20 tCO₂ week⁻¹ for both PC and HDT on Saturdays and Sundays. Furthermore, important variations in the monthly mean emissions are evident in the bottom-center plot (Figure 4C), particularly for HDT. March and November, registered a 40 tCO₂ increase compared to other months.

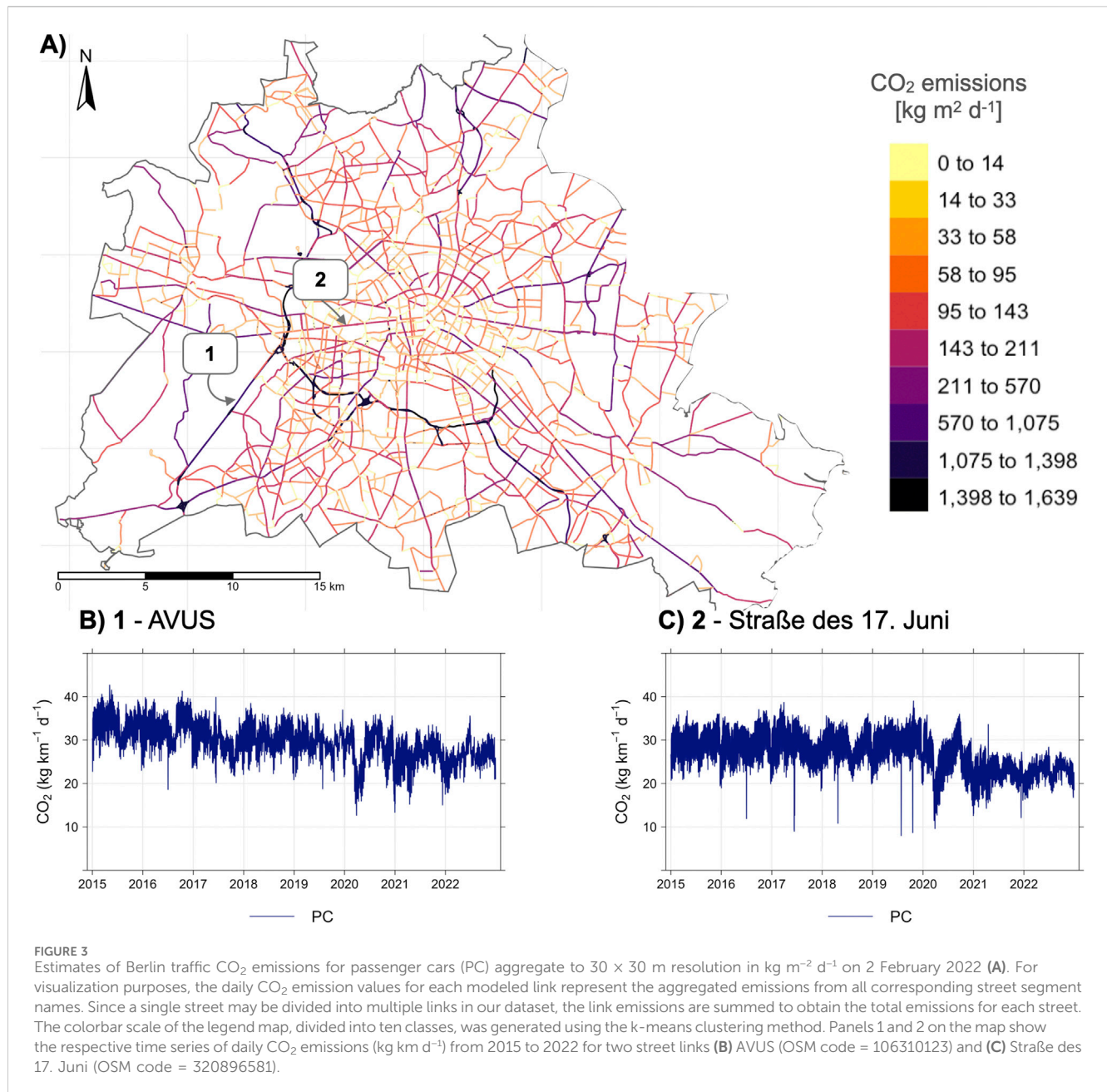
Collectively, these plots clearly demonstrate the pronounced temporal fluctuations in traffic CO₂ emission behavior, primarily influenced by changes in traffic flow. These findings agree with numerous studies that have shown similar temporal variation in

both modeled and measured traffic CO₂ emissions in urban settings (Buckley et al., 2016; Gurney et al., 2012; Mitchell et al., 2018; Park et al., 2022; Ueyama and Ando, 2016).

Figure 5 presents the daily total CO₂ emissions, including PC and HDT, in Berlin between 2015 and 2022. Emissions were calculated by averaging daily CO₂ estimates from 23,000 modeled road links across the city. On average, daily emissions were 73 tCO₂ d⁻¹, ranging from a minimum of 19 tCO₂ d⁻¹ to a maximum of 120 tCO₂ d⁻¹. Note that, in March 2020, emissions dropped sharply to approximately 60 tCO₂ d⁻¹ due to COVID-19 restrictions and the associated decline in traffic.

The reduction in CO₂ emissions persisted into 2021 and 2022, reflecting an decrease of 18% in traffic flow by 2021, as shown in Figure 6A. The continuation of COVID-19 measures, indicated by the daily Stringency Index (SI) in Figure 6B, highlights factors such as remote work and shifts in travel behavior that contributed to sustained reductions in traffic and CO₂ emissions. Section 3.5 provides a detailed analysis of these impacts.

Furthermore, the reductions in road traffic emissions align with annual CO₂ flux observations from *in situ* tower measurements at two Berlin neighborhood sites, TUCC and ROTH (Nicolini et al., 2022; Fenner et al., 2024). Further details on TUCC and ROTH are



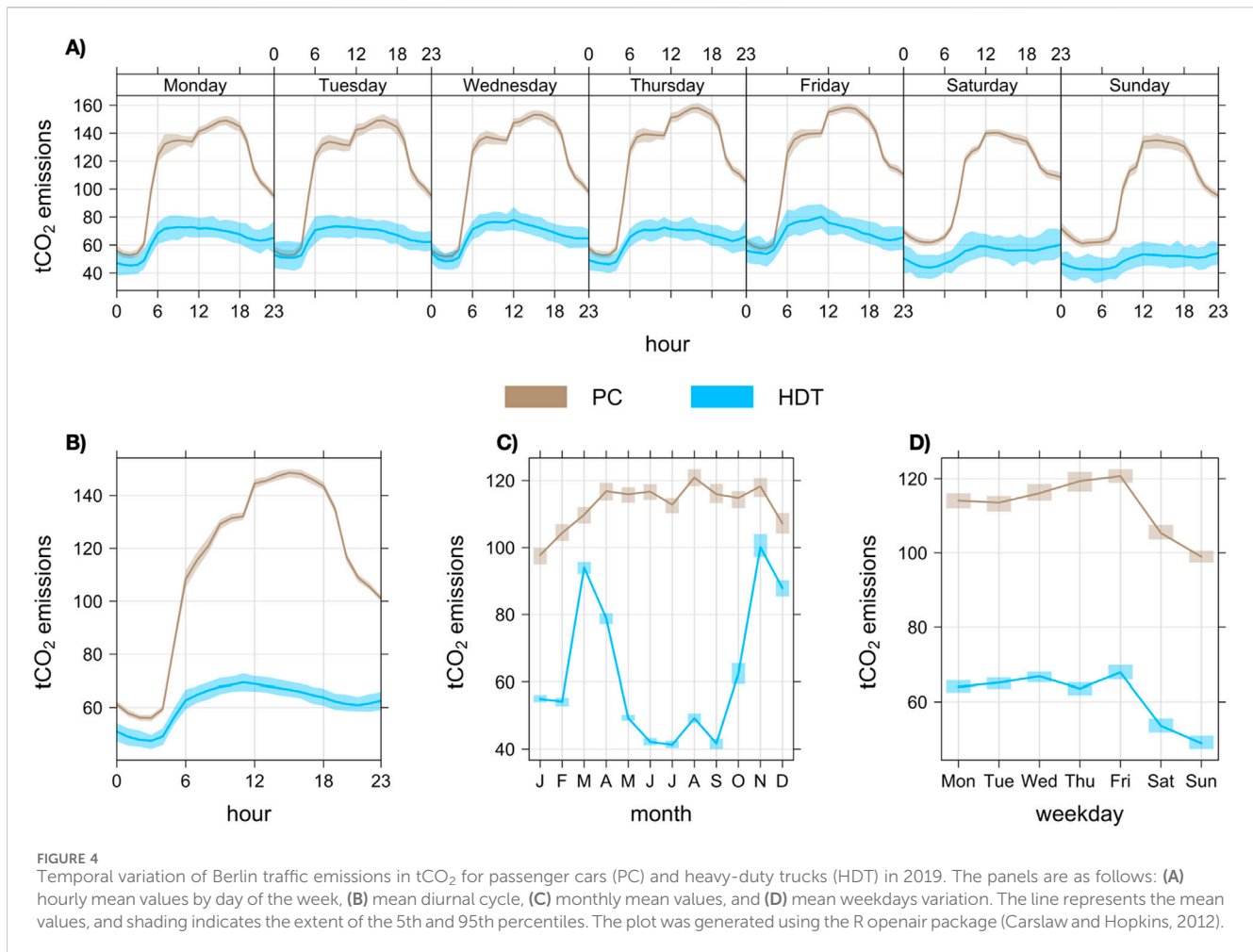
provided in [Supplementary Figure S4](#). Between 2019 and 2021, CO₂ fluxes at TUCC, primarily originating from road traffic, decreased by 25%, from 11.21 to 8.40 kgCO₂ m⁻² y⁻¹. Similarly, ROTH showed an 18% reduction, from 10.21 to 8.37 kgCO₂ m⁻² y⁻¹ by 2022 ([Figure 6C](#)). These decreases in CO₂ fluxes reflect changes in local anthropogenic activities (such as building energy consumption), vegetation dynamics, and weather conditions, beyond the impact of traffic flow.

3.4 Comparison with CM-cities daily CO₂ emissions

[Figure 7](#) compares the daily CO₂ emissions predicted by our ML model with those estimated by CM-cities in Berlin for 2019.

Both models exhibit similar daily patterns of CO₂ emissions variation, as evidenced by the high Pearson correlation coefficient ($R^2 = 0.76$) ([Supplementary Figure S5](#)). The relative mean difference (RMD), calculated as $[(ML_model - CM_cities) / \text{mean}(ML_model, CM_cities)]$, between the two models is 14%, indicating that our ML model exceeded the CO₂ emissions of CM-cities by an average of 814.7 tons per day. This RMD is within a reasonable range, as suggested by the interquartile range of 7%–22.4%.

The difference between these estimates can be attributed to the data sources and methods used for CO₂ emission calculations. For example, CM-Cities relies on TomTom's traffic data as a proxy for trace CO₂ emissions from traffic, which includes millions of anonymous consumer-driven GPS-based measurements representing the traffic flow and average speed across road



segments. It is important to note that some studies have shown that when novel mobility data like TomTom's data are compared to local traffic data, significant discrepancies can arise, with some studies reporting errors in emission estimates exceeding 60% (Gensheimer et al., 2020; Gensheimer et al., 2021). The authors reported that these errors are often due to different baselines, the omission of seasonal variations, and, primarily, the individual representations of the datasets, which may partially explain the differences found in CM-Cities' estimates relative to our ML model.

3.5 Assessment of impact of COVID-19 on traffic emissions

3.5.1 Overall impact of the lockdown on traffic emissions

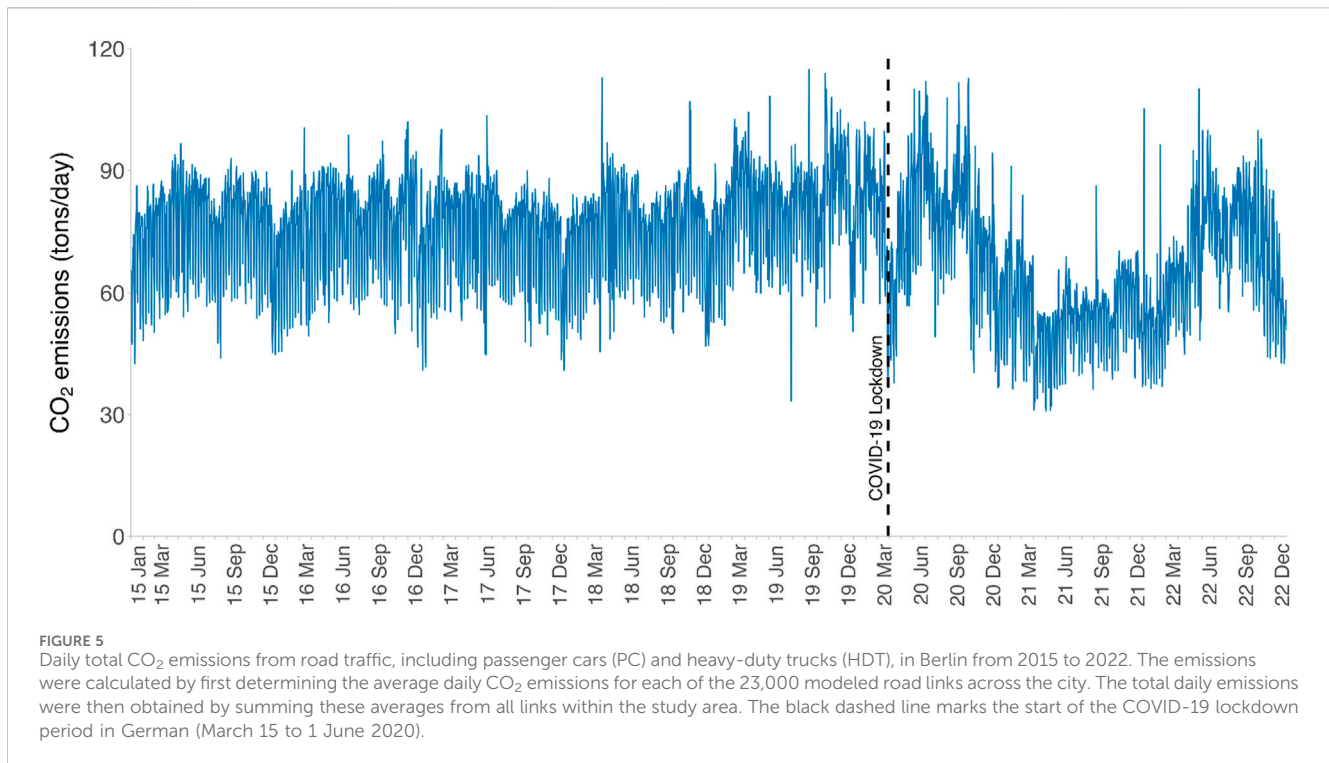
During the COVID-19 lockdown, Berlin's daily CO₂ emissions from road traffic in 2022 decreased by an estimated 14.2 tCO₂ day⁻¹ compared to 2019, representing a relative reduction of 16.6% (Figure 8A). Across all weekdays, average CO₂ emissions fell by 17.3% (95% CI: 12.1%–22.5%), corresponding to a reduction of approximately 12.6 tCO₂ day⁻¹

(CI: 9.0–16.4 tCO₂) from the 2019 baseline of 73 tCO₂ day⁻¹ (Figure 8B). Weekday emissions dropped by 12% (CI: 8.5%–15.2%), equivalent to an average decrease of 8.8 tCO₂ day⁻¹ (CI: 6.5–11.1 tCO₂), while weekends recorded the largest reductions, declining by 25% (CI: 20.3%–30.1%), which corresponds to a decrease of 18.3 tCO₂ day⁻¹ (CI: 14.8–21.8 tCO₂). The pronounced reductions on weekends are partially explained by shifts in public behavior and restricted leisure and labor activities during the lockdown.

Our findings align with results observed across both global and local scales. Liu et al. (2022) reported a significant 17% reduction in global CO₂ emissions during the peak weekly decline of the COVID-19 lockdown in 2020 compared to the same period in 2019. At the city scale, Schatke et al. (2022) found similar results, noting an average NO₂ concentration reduction of –21.9% across various sites (air pollution stations) in Berlin, during the lockdown period (March 15 to 1 June 2020).

3.5.2 Effects of specific policies on traffic emissions

To assess the impact of specific containment and closure policies on traffic CO₂ emissions during the defined lockdown, we analyzed correlations and performed PLS regression between daily CO₂ emissions and six individual OxCGRT policy measures



(Supplementary Figure S6). The correlation analysis revealed moderate negative correlations for stay-at-home requirements ($r = -0.39$), workplace closing ($r = -0.31$) and school closing ($r = -0.34$), indicating their immediate effect in reducing mobility and associated CO₂ emissions. This suggests that limiting the movement of people by enforcing “shelter-in-place” orders substantially reduced vehicle traffic emissions at street scale during the pandemic. In contrast, international travel controls ($r = -0.28$) and restrictions on gatherings ($r = -0.21$) showed weaker correlations with CO₂ emissions, indicating that these measures alone had less direct influence on traffic emission reductions.

In a multivariate analysis, the PLS regression further highlighted workplace closures as having the most substantial negative impact on emissions (coefficient = -24.39) (Figure 9). This indicates that each unit increase in the workplace closure index (e.g., from no restrictions to partial or severe restrictions) was associated with a reduction of $24.39 \text{ tCO}_2 \text{ day}^{-1}$, largely due to the decrease in commuting as remote work policies were implemented. Stay-at-home requirements ($-14.37 \text{ tCO}_2 \text{ day}^{-1}$) and restrictions on gatherings ($-12.33 \text{ tCO}_2 \text{ day}^{-1}$) also contributed to emission reduction, due to the decline in leisure and social travel. School closures had a moderate negative impact ($-7.21 \text{ tCO}_2 \text{ day}^{-1}$), dropping traffic flow linked to educational activities.

It is worth mentioning that restrictions on internal movement showed a positive coefficient ($12.45 \text{ tCO}_2 \text{ day}^{-1}$), suggesting potential compensatory traffic patterns, such as localized travel within restricted areas (Figure 9). International travel controls had the smallest impact ($-3.05 \text{ tCO}_2 \text{ day}^{-1}$), likely due to the lower contribution of international trips to urban traffic.

4 Discussion

Our ML model provides accurate hourly street-scale CO₂ emissions, overcoming the difficulties posed by limited observational data. This particularly addresses the issue of unavailable traffic flow and average speed data in various roadways, so that the proposed ML-model can enable cities to compute their CO₂ emissions at a manageable scale. Our findings underscore the significance of accounting for the high spatio-temporal variability of vehicle traffic CO₂ emissions when formulating and implementing carbon management strategies in urban areas. As outlined in the 3.2 and 3.3 sections, our results can be useful for assessment of policies targeting CO₂ emission reduction, particularly in high-emitting roads during peak emission periods.

To facilitate reproducibility and effective communication of our ML model’s outcomes, we have developed an Emission Geographic Information Platform accessible at <https://bymaxanjos.github.io/CO2-traffic-emissions/>. This platform offers an interactive interface for visualizing traffic emissions through zoomable CO₂ maps categorized by district (Figure 10). It also provides summary statistics for a comprehensive understanding of the data, covering a diverse audience, including users, stakeholders, the research community, and the general public.

Furthermore, our ML model has the potential to enhance the generation of comprehensive and detailed greenhouse gas (GHG) inventories for urban areas. Traditional GHG inventories, often based on annual averages, oversimplify the spatio-temporal variability of emissions, leading to discrepancies when compared to high-resolution modeling results (Chen et al., 2020; Gurney et al., 2021). Our ML model offers a promising way for improving urban GHG inventories by providing detailed

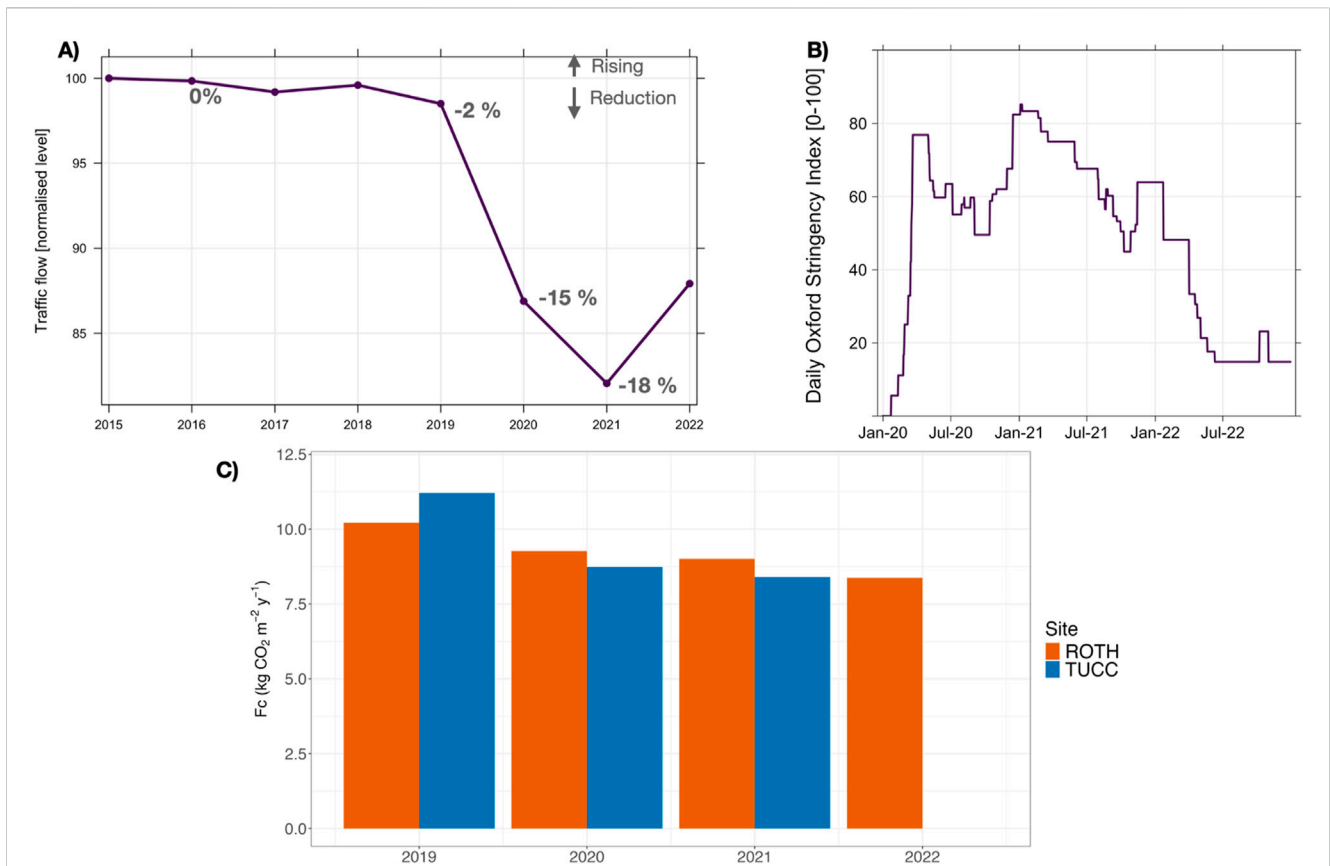


FIGURE 6 (A) Annual normalized variation in traffic flow in Berlin from 2015 to 2022, based on monthly totals from 583 counting stations. Traffic flow is normalized to a 2015 baseline (100), with values below 100 indicating a relative decrease (B) Germany’s daily Stringency Index (SI) values from the Oxford COVID-19 Government Response Tracker (OxCGRT) for 2020–2022, where SI ranges from 0 (no measures) to 100 (maximum stringency) (C) Annual CO₂ fluxes (kg m⁻² y⁻¹) measured at two Berlin tower sites: TU Campus Charlottenburg (TUCC, 52.45723°N, 13.31583°E) and Steglitz Rothenburgstrasse (ROTH, 52.51228°N, 13.32786°E). No data available for TUCC in 2022 due to technical problems.

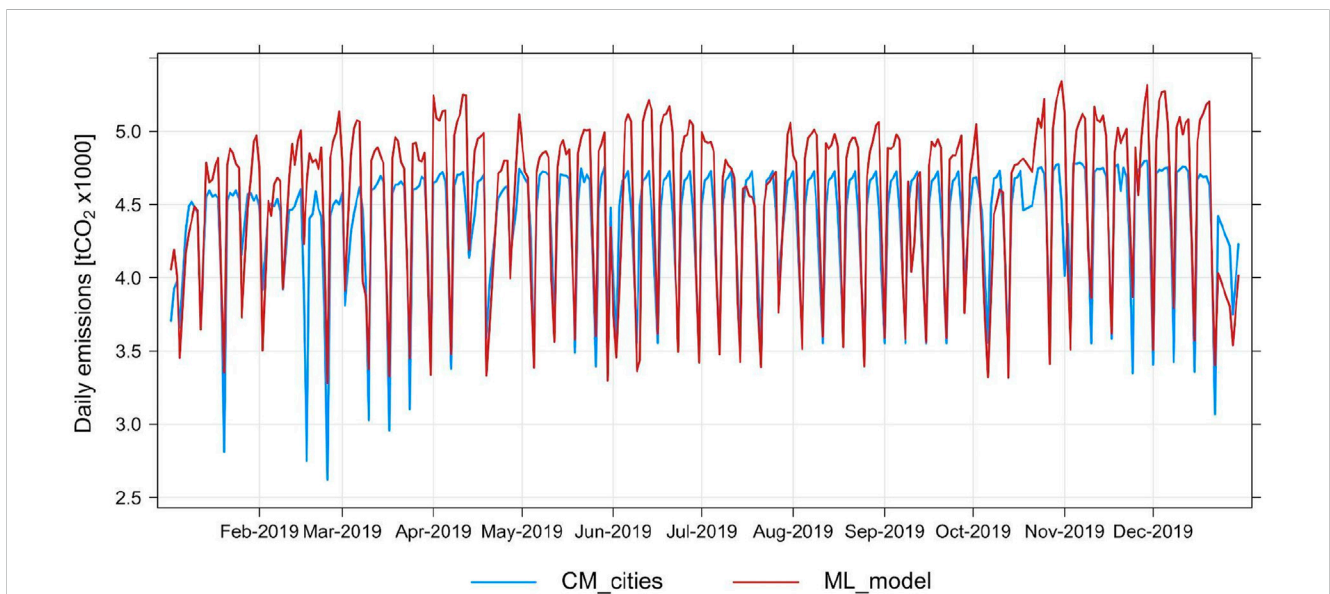


FIGURE 7 Time series of daily traffic emissions including passenger cars (PC) and heavy-duty trucks (HDT) in tCO₂ (x1,000) for ML model and CM-Cities in Berlin from January 2019 to December 2019.

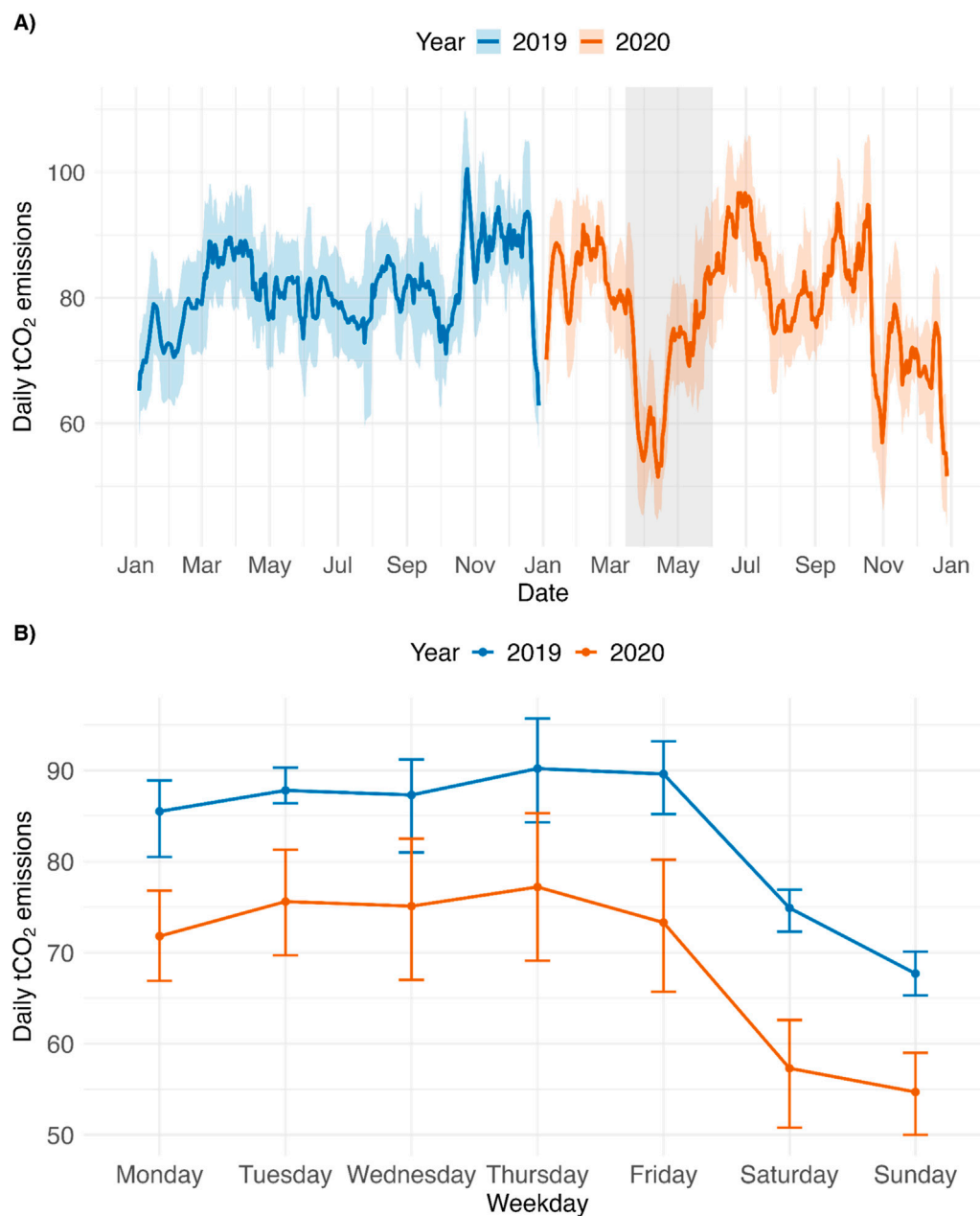


FIGURE 8 Berlin's daily CO₂ traffic emissions, including heavy-duty trucks (HDT) and passenger cars (PC), presented as **(A)** a 7-day running mean and **(B)** weekday averages tCO₂ day⁻¹. The lockdown period, spanning March 15 to 1 June 2020, is shaded gray, while the baseline period represents emissions during March 15 to 1 June 2019, unaffected by COVID-19 restrictions. In **(A)**, the blue and orange shaded areas indicate uncertainty ranges and 95% confidence intervals for the 2019 and 2020 estimates, respectively. In **(B)**, error bars represent the 95% confidence intervals.

and accurate estimates of traffic CO₂ emissions, capturing hourly fluctuations at the street scale. This level of detail is crucial for identifying and mitigating traffic emissions hotspots within the city, aiding in the analysis and monitoring of traffic flow, velocity, and emissions patterns.

While our study primarily focuses on CO₂, the ML-based bottom-up model can also be easily adapted to estimate other GHG species and traffic-related pollutants in urban areas. The integration of high spatio-temporal resolution in emissions

modeling can help to identify areas vulnerable to air pollution, contributing to the development of localized health policies and interventions. In addition, the flexibility and CO₂ emission disaggregation can support the recognition and monitoring of potential geographical zones in the city suitable for traffic regulation initiatives, such as Low Emissions Zones (Chukwunonye Ezeah et al., 2015).

This study provides a general quantification of urban traffic CO₂ emissions only differentiating traffic flow and average speed by PC and

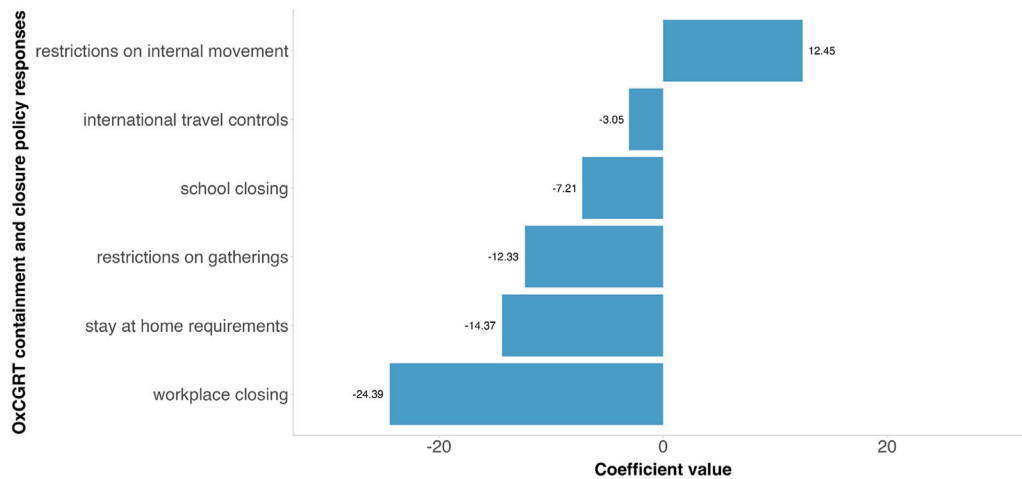


FIGURE 9 PLS regression between daily traffic CO₂ emissions (tCO₂ day⁻¹), including passenger cars (PC) and heavy-duty trucks (HDT), and the OxCGRt containment and closure policy responses to COVID-19 in Berlin during the lockdown (March 15 to 1 June 2020). Negative coefficients indicate stricter policies are associated with reduction in CO₂ emissions, while positive coefficients suggest increases.

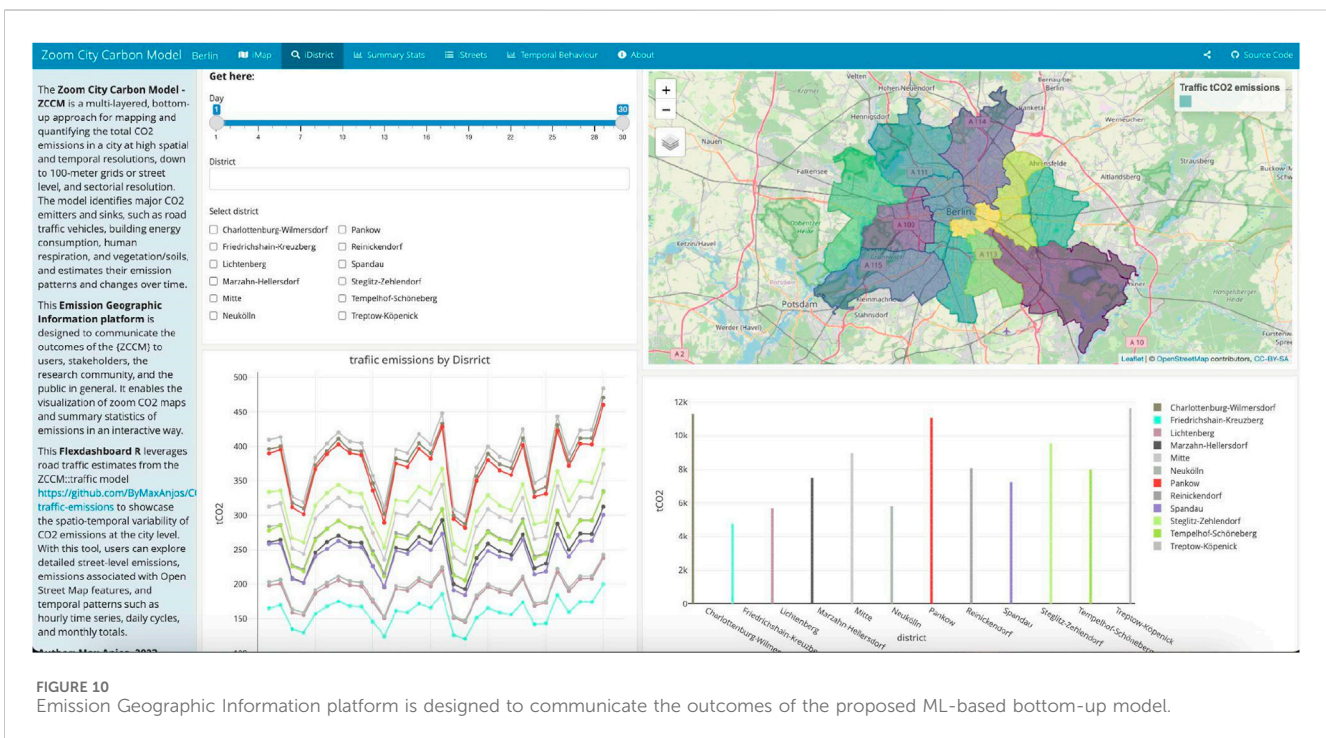


FIGURE 10 Emission Geographic Information platform is designed to communicate the outcomes of the proposed ML-based bottom-up model.

HDT (including only trucks). Future studies can evaluate the distinct impacts of other vehicle engine types on local emissions, considering the growing prevalence of electric vehicles globally (IEA, 2019). Applying the spatio-temporal modeling approach would be useful for assessing the influence of electric vehicles on reducing emissions.

Our evaluation metrics, with an RMSE of 155 veh h⁻¹ for traffic flow and 10 km h⁻¹ for average speed, are comparable to, and in some cases surpass, those of other state-of-the-art models reported in the literature. For instance, a review by Medina-Salgado et al.

(2022), which analyzed computational techniques from 61 studies on urban traffic flow prediction, reported a maximum RMSE of 240.98 veh h⁻¹. This highlights the robustness of our ML model in predicting traffic flow, as it effectively estimates data from measured road segments to unmeasured ones in Berlin, maintaining a reasonable error margin within the bounds established by prior studies.

It is important to acknowledge that the use of data from 583 traffic count stations for ML model development may lead to

under (over)estimation of results in specific streets. This because certain road types, such as residential, were excluded from the analysis due to measurement data unavailability. To further enhance the model's predictive power and provide more understanding of traffic patterns and associated emissions, feature research could include additional variables such as road network structure, proximity to Central Business District, and public transit accessibility.

To assess the performance of our CO₂ emission estimates, we compared them with the Carbon Monitor Cities dataset, which provides confidence in our findings. However, further validation using *in situ* CO₂ measurement data is recommended. One effective approach for validation is the Eddy Covariance (EC) method, a well-established technique for measuring greenhouse gas, water, and energy fluxes in urban areas (Valesco and Roth, 2010). By combining our traffic CO₂ emissions estimates with EC fluxes and footprint analysis, an advanced assessment of the model performance can be carried out.

We found that the stay-at-home was the strictest measure, associated with an approximate reduction of 45% in traffic CO₂ emissions on weekdays during the lockdown period. This result highlights the important influence that “shelter-in-place” orders can have on reducing traffic emissions at street scale during COVID-19 pandemic events. Traffic emissions at street scale during COVID-19 pandemic events. This is consistent with other studies (Bekbulat et al., 2021; Le Quéré et al., 2020; Liu et al., 2021; Liu et al., 2020; Turner et al., 2020) that have also emphasized the effectiveness of “shelter-in-place” orders in reducing traffic emissions at global and urban scales. It is important to consider, however, that while staying at home during lockdowns reduces traffic emissions, it may lead to an increased building CO₂ emissions related to domestic heating and cooking, as noted by Nicolini et al. (2022) in European areas.

5 Conclusions

This study highlights the potential of AI tools, such as ML models, in quantifying CO₂ emissions in urban environments. By incorporating ML techniques into the bottom-up approach, we accurately estimated CO₂ emissions from vehicular traffic in Berlin at a high spatio-temporal resolution, with hourly emissions on a 30-meter grid at the street scale. Our approach, leveraging traffic counts, spatial features, and meteorological data, provides detailed CO₂ emissions patterns over a 7-year period (2015–2022).

The comparison shows that the ML model and CM-Cities are able to capture daily variations in traffic emissions. Nevertheless, our model has the advantage of estimating CO₂ emissions on a single road segment and hourly basis, which may be relevant for informed climatic action within a city, although it requires more computational effort and diverse inputs compared to CM-Cities. User objectives, research questions, and available datasets determine the choice between approaches.

The results highlighted significant CO₂ emissions concentrated along major highways and demonstrated the substantial impact of COVID-19 lockdown measures on reducing traffic-related

emissions. The OxCGRT containment and closure policies, including stay-at-home orders, workplace closures, school closures, were associated with reductions in road traffic CO₂ emissions during the lockdown period (March 15 to 1 June 2020). These measures continued to influence emissions throughout 2020 and 2022.

This research underscores the utility of ML-based bottom-up models in areas with limited emissions data, providing researchers and policymakers with potential tools for monitoring of traffic-related emissions and their impact on air quality and public health. Our findings contribute to the growing field of ML applications in emissions modeling and spatio-temporal analysis of greenhouse gas variations, paving the way for more data-driven, sustainable urban planning and policy-making.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. All the codes used in this study have been uploaded on the same public GitHub repository (<https://github.com/ByMaxAnjos/CO2-traffic-emissions>).

Author contributions

MA: Conceptualization, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. FM: Funding acquisition, Investigation, Methodology, Supervision, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, and by the Alexander Von Humboldt Foundation. We acknowledge support by the German Research Foundation and the Open Access Publication Fund of Technische Universität Berlin for APC payment.

Acknowledgments

We would like to thank Marcos Alves for his thoughtful comments on Machine Learning techniques and Gabriel Leitoles for his contribution to the early stages of the approach presented in this study. We would also like to express our appreciation to all peer reviewers who provided feedback on this work, and we welcome further suggestions and comments as we continue to refine our approach.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fenvs.2024.1461656/full#supplementary-material>

References

- Aurélien Géron (2022). *Hands-on machine learning with scikit-learn, keras, and TensorFlow*. Sebastopol, CA, United States: O'Reilly Media, Inc.
- Bekbulat, B., Apte, J. S., Millet, D. B., Robinson, A. L., Wells, K. C., Presto, A. A., et al. (2021). Changes in criteria air pollution levels in the US before, during, and after Covid-19 stay-at-home orders: evidence from regulatory monitors. *Sci. Total Environ.* 769, 144693. doi:10.1016/j.scitotenv.2020.144693
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi:10.1023/A:1010933404324
- Buckley, S. M., Mitchell, M. J., McHale, P. J., and Millard, G. D. (2016). Variations in carbon dioxide fluxes within a city landscape: identifying a vehicular influence. *Urban Ecosyst.* 19 (4), 1479–1498. doi:10.1007/s1252-013-0341-0
- Chen, G., Shan, Y., Hu, Y., Tong, K., Wiedmann, T., Ramaswami, A., et al. (2019). Review on city-level carbon accounting. *Environ. Sci. and Technol.* 53 (10), 5545–5558. doi:10.1021/acs.est.8b07071
- Chen, J., Zhao, F., Zeng, N., and Oda, T. (2020). Comparing a global high-resolution downscaled fossil fuel CO₂ emission dataset to local inventory-based estimates over 14 global cities. *Carbon Balance Manag.* 15 (1), 9. doi:10.1186/s13021-020-00146-3
- Chollet, F. (2018). in *Deep learning with R*. Editor J. J. Allaire (Shelter Island, NY, United States: Manning Publications).
- Christen, A. (2014). Atmospheric measurement techniques to quantify greenhouse gas emissions from cities. *Urban Clim.* 10, 241–260. doi:10.1016/j.uclim.2014.04.006
- Dancho, M., and Vaughan, D. (2024). *Timetk: A tool kit for working with time series. R package version 2.9.0*. Available at: <https://business-science.github.io/timetk/>.
- Duren, R. M., and Miller, C. E. (2012). Measuring the carbon emissions of megacities. *Nat. Clim. Change* 2 (8), 560–562. doi:10.1038/nclimate1629
- DWD (2020). Open data area of the climate data center. *Dtsch. Wetterd.* Available at: https://opendata.dwd.de/climate_environment/CDC/.
- EEA (2017). *Annual European Union greenhouse gas inventory 1990–2015 and inventory report 2017*. Copenhagen, Denmark: European Environment Agency. Available at: <https://www.eea.europa.eu/publications/european-union-greenhouse-gas-inventory-2017>.
- EEA (2023). *EMEP/EEA air pollutant emission inventory guidebook 2023 (1.A.3.b.i-iv Road transport)*. Copenhagen, Denmark: European Environment Agency. Available at: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2023/part-b-sectorial-guidance-chapters/1-energy/1-a-combustion/1-a-3-b-i/view>.
- Ezeah, C., Finney, K., and Nnajide, C. (2015). A critical review of the effectiveness of Low emission zones (lez) as A strategy for the management of air quality in major European cities. *J. Multidiscip. Eng. Sci. Technol.* 2 (7).
- Fenner, D., Christen, A., Grimmond, S., Meier, F., Morrison, W., Zeeman, M., et al. (2024). Urbisphere-berlin campaign: investigating multiscale urban impacts on the atmospheric boundary layer. *Bull. Amer. Meteor. Soc.* 105, E1929–E1961. doi:10.1175/BAMS-D-23-0030.1
- Gately, C. K., Hutyra, L. R., and Sue Wing, I. (2015). Cities, traffic, and CO₂: a multidecadal assessment of trends, drivers, and scaling relationships. *Proc. Natl. Acad. Sci.* 112 (16), 4999–5004. doi:10.1073/pnas.1421723112
- Geladi, P., and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Anal. Chim. Acta* 185, 1–17. doi:10.1016/0003-2670(86)80028-9
- Gensheimer, J., Turner, A., Shekhar, A., Wenzel, A., and Chen, J. (2020). What are different measures of mobility changes telling us about emissions during the COVID-19 pandemic? doi:10.1002/essoar.10504783.1
- Gensheimer, J., Turner, A. J., Shekhar, A., Wenzel, A., Keutsch, F. N., and Chen, J. (2021). Error assessment of traffic emission estimates using novel mobility datasets. doi:10.5194/egusphere-egu21-5419
- Gurney, K. R., Liang, J., Roest, G., Song, Y., Mueller, K., and Lauvaux, T. (2021). Under-reporting of greenhouse gas emissions in U.S. cities. *Nat. Commun.* 12 (1), 553. doi:10.1038/s41467-020-20871-0
- Gurney, K. R., Patarasuk, R., Liang, J., Song, Y., O'Keefe, D., Rao, P., et al. (2019). The Hestia fossil fuel CO₂ emissions data product for the Los Angeles megacity (Hestia-LA). *Earth Syst. Sci. Data* 11 (3), 1309–1335. doi:10.5194/essd-11-1309-2019
- Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M. (2012). Quantification of fossil fuel CO₂ emissions on the building/street scale for a large U.S. City. *Environ. Sci. and Technol.* 46 (21), 12194–12202. doi:10.1021/es3011282
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* 5 (4), 529–538. doi:10.1038/s41562-021-01079-8
- Hirschl, B., and Harnisch, R. (2016). *Climate-neutral Berlin 2050: recommendations for a Berlin energy and climate protection programme (BEK)*. Berlin, Germany: Senate Department for Urban Development and the Environment.
- Huo, D., Huang, X., Dou, X., Ciaia, P., Li, Y., Deng, Z., et al. (2022). Carbon Monitor Cities near-real-time daily estimates of CO₂ emissions from 1500 cities worldwide. *Sci. Data* 9 (1), 533. doi:10.1038/s41597-022-01657-z
- Huo, H., Zhang, Q., He, K., Wang, Q., Yao, Z., and Streets, D. G. (2009). High-resolution vehicular emission inventory using a link-based method: a case study of light-duty vehicles in Beijing. *Environ. Sci. and Technol.* 43 (7), 2394–2399. doi:10.1021/es802757a
- IEA (2019). *CO₂ emissions from fuel combustion 2019*. Paris, France: OECD. doi:10.1787/2a701673-en
- IPCC (2000). *IPCC good practice guidance and uncertainty management in national greenhouse gas inventories*. Geneva, Switzerland: Intergovernmental Panel on Climate Change. Available at: https://www.ipcc-nggip.iges.or.jp/public/gp/english/2_Energy.pdf.
- Jungmann, M., Vardag, S. N., Kutzner, F., Keppler, F., Schmidt, M., Aeschbach, N., et al. (2022). Zooming-in for climate action—hyperlocal greenhouse gas data for mitigation action? *Clim. Action* 1 (1), 8. Article 1. doi:10.1007/s44168-022-00007-4
- Ku, A. Y., Greig, C., and Larson, E. (2022). Traffic ahead: navigating the road to carbon neutrality. *Energy Res. and Soc. Sci.* 91, 102686. doi:10.1016/j.erss.2022.102686
- Kuhn, M., and Johnson, K. (2013). *Applied predictive modeling*. Springer.
- Kuhn, M., and Johnson, K. (2019). *Feature engineering and selection*. Chapman and Hall/CRC. doi:10.1201/9781315108230
- Kuhn, M., and Wickham, H. (2023). *Recipes: Preprocessing and feature engineering steps for modeling. R package version 1.1.0*. Available at: <https://github.com/tidymodels/recipes>.
- Le Quérec, C., Jackson, R. B., Jones, M. W., Smith, A. J. P., Abernethy, S., Andrew, R. M., et al. (2020). Temporary reduction in daily global CO₂ emissions during the COVID-19 forced confinement. *Nat. Clim. Change* 10 (7), 647–653. doi:10.1038/s41558-020-0797-x
- Liu, D., Sun, W., Zeng, N., Han, P., Yao, B., Liu, Z., et al. (2021). Observed decreases in on-road CO₂ concentrations in Beijing during COVID-19 restrictions. *Atmos. Chem. Phys.* 21 (6), 4599–4614. doi:10.5194/acp-21-4599-2021
- Liu, Y., and Wu, H. (2017). "Prediction of road traffic congestion based on random forest," in *2017 10th international symposium on computational intelligence and design (ISCID)*, 361–364. doi:10.1109/ISCID.2017.216
- Liu, Z., Ciaia, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., et al. (2020). Near-real-time monitoring of global CO₂ emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* 11 (1), 5172. doi:10.1038/s41467-020-18922-7

- Liu, Z., Deng, Z., Zhu, B., Ciais, P., Davis, S. J., Tan, J., et al. (2022). Global patterns of daily CO₂ emissions reductions in the first year of COVID-19. *Nat. Geosci.* 15, 615–620. doi:10.1038/s41561-022-00965-8
- Liu, Z., Li, Z., Wu, K., and Li, M. (2018). Urban traffic prediction from mobility data using deep learning. *IEEE Netw.* 32 (4), 40–46. doi:10.1109/MNET.2018.1700411
- Lv, Y., Duan, Y., Kang, W., Li, Z., and Wang, F. Y. (2014). Traffic flow prediction with big data: a deep learning approach. *IEEE Trans. Intelligent Transp. Syst.* 16 (2), 1–9. doi:10.1109/TITS.2014.2345663
- Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., and Sierra, J. E. (2022). Urban traffic flow prediction techniques: a review. *Sustain. Comput. Inf. Syst.* 35, 100739. doi:10.1016/j.suscom.2022.100739
- Mitchell, L. E., Lin, J. C., Bowling, D. R., Pataki, D. E., Strong, C., Schauer, A. J., et al. (2018). Long-term urban carbon dioxide observations reveal spatial and temporal dynamics related to urban characteristics and growth. *Proc. Natl. Acad. Sci.* 115 (12), 2912–2917. doi:10.1073/pnas.1702393115
- Nangini, C., Peregón, A., Ciais, P., Weddige, U., Vogel, F., Wang, J., et al. (2019). A global dataset of CO₂ emissions and ancillary data related to emissions for 343 cities. *Sci. Data* 6 (1), 180280. doi:10.1038/sdata.2018.280
- Nicolini, G., Antoniella, G., Carotenuto, F., Christen, A., Ciais, P., Feigenwinter, C., et al. (2022). Direct observations of CO₂ emission reductions due to COVID-19 lockdown across European urban districts. *Sci. Total Environ.* 830, 154662. doi:10.1016/j.scitotenv.2022.154662
- Ntziachristos, L., Gkatzoflias, D., Kouridis, C., and Samaras, Z. (2009). “COPERT: a European road transport emission inventory model,” in *Information technologies in environmental engineering*. Editors I. N. Athanasiadis, A. E. Rizzoli, P. A. Mitkas, and J. M. Gómez (Springer Berlin Heidelberg), 491–504. doi:10.1007/978-3-540-88351-7_37
- OpenStreetMap contributors (2017). *Planet dump*. Available at: <https://planet.osm.org>.
- Park, C., Jeong, S., Park, M.-S., Park, H., Yun, J., Lee, S.-S., et al. (2022). Spatiotemporal variations in urban CO₂ flux with land-use types in Seoul. *Carbon Balance Manag.* 17 (1), 3. doi:10.1186/s13021-022-00206-w
- Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *R J.* 10 (1), 439. doi:10.32614/RJ-2018-009
- R Core Team (2018). *A language and environment for statistical computing [computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Roest, G. S., Gurney, K. R., Miller, S. M., and Liang, J. (2020). Informing urban climate planning with high resolution data: the Hestia fossil fuel CO₂ emissions for Baltimore, Maryland. *Carbon Balance Manag.* 15 (1), 22. doi:10.1186/s13021-020-00157-0
- Schatke, M., Meier, F., Schröder, B., and Weber, S. (2022). Impact of the 2020 COVID-19 lockdown on NO₂ and PM₁₀ concentrations in Berlin, Germany. *Atmos. Environ.* 290, 119372. doi:10.1016/j.atmosenv.2022.119372
- SenStadtWohn (2019). *Traffic volumes ADT 2019*. Berlin, Germany: Senate Department for Urban Development and Housing. Available at: <https://www.berlin.de/umweltatlas/en/traffic-noise/traffic-volumes/2019/summary/>.
- SenUVK (2021). “Geoportal Berlin/Grünanlagenbestand Berlin (einschließlich der öffentlichen Spielplätze),” in *FIS-Broker*. Available at: <https://fbinter.stadt-berlin.de/fb/index.jsp>.
- Senuvk (Senate Department for the Environment, U. M., Consumer Protection and Climate Action) (2018). Climate protection in Berlin. Available at: <https://www.berlin.de/sen/uvk/en/climate-action/publications/>.
- Seto, K. C., Churkina, G., Hsu, A., Keller, M., Newman, P. W. G., Qin, B., et al. (2021). From Low-to net-zero carbon cities: the next global agenda. *Annu. Rev. Environ. Resour.* 46 (1), 377–415. doi:10.1146/annurev-environ-050120-113117
- Shaygan, M., Meese, C., Li, W., Zhao, X. G., and Nejad, M. (2022). Traffic prediction using artificial intelligence: review of recent advances and emerging opportunities. *Transp. Res. Part C Emerg. Technol.* 145, 103921. doi:10.1016/j.trc.2022.103921
- Stagakis, S., Feigenwinter, C., Vogt, R., and Kalberer, M. (2023). A high-resolution monitoring approach of urban CO₂ fluxes. Part 1 - bottom-up model development. *Sci. Total Environ.* 858 (2023), 160216. doi:10.1016/j.scitotenv.2022.160216
- Statistical Office of Berlin-Brandenburg (2019). *Inhabitants of the state of Berlin on 31 december 2018*. Potsdam, Germany: Potsdam.
- Turnbull, J., DeCola, P., Mueller, K., Vogel, F., Karion, A., Coto, I. L., et al. (2022). IG3IS urban greenhouse gas emission observation and monitoring best research practices. *World Meteorol. Organ. Integr. Greenh. Gas. Inf. Syst.*
- Ueyama, M., and Ando, T. (2016). Diurnal, weekly, seasonal, and spatial variabilities in carbon dioxide flux in different urban landscapes in Sakai, Japan. *Atmos. Chem. Phys.* 16 (22), 14727–14740. doi:10.5194/acp-16-14727-2016
- Velasco, E., and Roth, M. (2010). Cities as net sources of CO₂: review of atmospheric CO₂ exchange in urban environments measured by Eddy covariance technique. *E. Geogr. Compass* 4, 1238–1259. doi:10.1111/j.1749-8198.2010.00384.x
- Wang, H., Fu, L., and Chen, J. (2010). Developing a high-resolution vehicular emission inventory by integrating an emission model and a traffic model: Part 2—a case study in Beijing. *J. Air and Waste Manag. Assoc.* 60 (12), 1471–1475. doi:10.3155/1047-3289.60.12.1471
- Wen, Y., Wu, R., Zhou, Z., Zhang, S., Yang, S., Wallington, T. J., et al. (2022). A data-driven method of traffic emissions mapping with land use random forest models. *Appl. Energy* 305, 117916. doi:10.1016/j.apenergy.2021.117916
- Wright, M. N., and Ziegler, A. (2017). ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* 77, 1–17. doi:10.18637/jss.v077.i01