Transportation Research Part D 119 (2023) 103757

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Estimating public transport emissions from General Transit Feed Specification data

João Pedro Bazzo Vieira [a,*], Rafael H.M. Pereira [a], Pedro R. Andrade [b]

[a] Institute for Applied Economic Research, Brasília 70390-025, Brazil
[b] National Institute of Space Research, São José dos Campos, 12227-010, Brazil

ARTICLE INFO

Keywords:
Gtfs2emis
Public transport emissions
Emission models
Urban bus
GTFS
Emission factors

ABSTRACT

This paper introduces the gtfs2emis model, a bottom-up method available as an R package to estimate emissions of public transport systems. The method uses General Transit Feed Specification (GTFS) data, a standard format for public transport data widely adopted worldwide, which makes the method easily applicable to cities with limited data. The model requires a GTFS feed of a given transport system and a table with general characteristics of the vehicle fleet profile. The package can estimate over 16 pollutants and energy consumption based on emission factor models from Europe, the United States, and Brazil. It also includes functions to help users examine how emissions are distributed across space, at different times of the day, and by types of vehicles. This paper presents a reproducible example of the city of São Paulo (Brazil) to demonstrate the gtfs2emis package and to discuss the potential applications and limitations of the proposed model.

1. Introduction

The transport sector has been widely recognized among the leading and growing contributors to global emissions (Caiazzo et al., 2013; Nocera et al., 2018). There is scant evidence on how the pollution generated from transportation activities impacts air quality in cities (Landrigan et al., 2018) and harms people's health (Brook et al., 2010; Currie & Walker, 2009; Fu et al., 2019; Shehab & Pope, 2019; Zhang et al., 2018). In this context, the development of new open data and methods to estimate vehicle emissions has become particularly relevant amid the debates on urban health climate change and vehicle technology (Linero et al., 2020; Böhm et al., 2022; Yeh et al., 2022). As a result, there has been growing attention to public transport emissions, with several studies conducting onboard measurements in order to assess how emission levels are affected by operating conditions (Rosero et al., 2020), after-treatment technologies (López et al., 2009), vehicle age (Huang et al., 2022), road grade (Rosero et al., 2021), passenger load (Frey et al., 2020), and vehicle powertrains (Zhang et al., 2018; Zhao et al., 2021; García et al., 2022; Muñoz et al., 2022). Although these methods are useful to quantify real-world vehicle activity and energy use, there is still a lack of open-source models that assess the emissions of public transport systems at scale for large urban environments.

Here we introduce the gtfs2emis model, a novel bottom-up method to estimate tailpipe emissions for public transport systems. The model leverages public transport data in the General Transit Feed Specification (GTFS) format, an open data standard widely adopted by transport agencies worldwide, making the method easily applicable to different contexts. The method "gtfs2emis" is presented as an

* Corresponding author.
E-mail addresses: joao.bazzo@gmail.com, joao.vieira@ipea.gov.br (J.P.B. Vieira), rafael.pereira@ipea.gov.br (R.H.M. Pereira), pedro.andrade@inpe.br (P.R. Andrade).

https://doi.org/10.1016/j.trd.2023.103757
Received 10 November 2022; Received in revised form 18 April 2023; Accepted 21 April 2023
1361-9209/

R package (R Core Team, 2021) that allows friendly and reproducible use for different applications. Based on data from emission factor models for Europe (EMEP/EEA, 2019a), the United States (EPA, 2020; CARB, 2021), and Brazil (CETESB, 2019), the gtfs2emis package supports the efficient computation of over 16 pollutants at the vehicle level using seamless parallel computing. In this paper, we present the main functions of the package and use a case study of São Paulo (Brazil) to demonstrate how the gtfs2emis package can be used to estimate hot exhaust emissions of CO_2, NO_X, PM_{10}, and CH_4, and analyze the spatial and temporal distribution of pollutants.

In general, studies that investigate emissions for large transport networks often rely on large-scale inventories, based on traffic simulations or GPS data of vehicles (Liu et al., 2013; Shan et al., 2019a; Chan et al., 2022; Ibarra-Espinosa et al., 2020; Böhm et al., 2022). The work of Ibarra-Espinosa et al. (2020) and Böhm et al. (2022), for example, can be used to generate traffic emission inventories for large transportation systems but they focus exclusively on private vehicles and overlook public transportation networks.

In the case of urban bus emissions, the work of S. Chan et al. (2013) uses onboard GPS data to assess the impact of bus technology on greenhouse gas emissions of a busy bus corridor in Montreal (Canada). Using a similar method, Shan et al. (2019b) estimate bus emissions using instantaneous speed information collected from sparse GPS data of 43 buses in Shanghai. Meanwhile, López-Martínez et al. (2017) propose a methodology to estimate fuel consumption and emissions, using the public transport system of Madrid (Spain) as a case study. The authors relied on onboard measurements from a sample of vehicles and bus routes in order to determine the actual operating condition and the emission factors for different vehicle sizes, technologies, and powertrains (López-Martínez et al., 2017).

In relation to real-world emissions assessments, several studies use portable emission measurement systems (PEMS) to measure emission levels of urban buses in a few public transport routes and corridors (Liu et al., 2011; Huang et al., 2013; Zhang et al., 2014). Despite the insightful contributions of these studies, the use of PEMS is often limited to a few vehicle categories, bus routes, and driving conditions because of research costs. As an alternative to using real-world emissions data, which is not often available, several studies have relied on emission factor databases adjusted by vehicle activity to estimate emissions in a finer resolution. One example is found in Waraich et al. (2020), who proposed a modeling framework to estimate public transport emissions at the micro level using the bus transport system of Montreal (Canada) as a case study. Their method is composed of 1) a ridership module simulation (based on a series of regression models that predict hourly boardings and alightings at bus stops), 2) a model to estimate bus occupancy rates between consecutive stops at different times, and 3) emission estimates of running and idling conditions (Waraich et al., 2020).

A limitation of previous methods found in the literature is that they have strong data requirements. The data from onboard GPS devices often present significant computational challenges due to data size, and they are often incomplete or unavailable for several cities, especially in emerging economies. Moreover, these data are not structured in a standardized format, which creates barriers to aggregating and comparing the data from different transport agencies. Meanwhile, due to costly data collection, measuring emissions through PEMs can also be very challenging to scale to city-level analysis.

The gtfs2emis model proposed in this paper advances the literature by proposing a public transport emission model with minimal data requirements that can be easily applicable and scalable to multiple cities through a user-friendly R package. The GTFS specification used in gtfs2emis provides a structured framework for public transport data in a format that is easy to maintain and commonly used by transport agencies across several countries. Because GTFS data sets are meant to reflect the level of services planned for public transport systems, emission estimates based on GTFS can be useful to examine the emission levels expected from transport plans and policies. They can also be useful for benchmark studies to anticipate how alternative interventions to public transport systems could impact hot-exhaust emission levels. Additionally, the proposed method allows a reasonable understanding of the spatiotemporal patterns of how public transport emissions are distributed within cities.

The remainder of this paper is organized as follows. Section 2 describes the methods used in the gtfs2emis model. Section 3 presents the results of the case study of São Paulo, looking at the overall levels of pollutant emissions and their spatial and temporal distribution. Finally, Section 4 presents the final remarks and discusses some environmental insights that can be drawn from the package as well as a potential research agenda.

2. Methods

The gtfs2emis implements a two-step model to estimate emissions from public transport data in an R package (R Core Team, 2021). The first step is the transport model, called in R with the transport_model function. It converts a GTFS data input into a trajectory data table, similar to GPS records with the space–time position of every public transport vehicle. The second step is the emission model, called the emission_model function. This step estimates the pollutants emitted by each vehicle at each road segment and time of the day by combining the output from the transport model with additional data on fleet characteristics provided by the user, and emission factors provided in the gtfs2emis package.

The gtfs2emis package requires two main inputs: GTFS data and a table with fleet characteristics. The steps, data workflow, and functions of gtfs2emis are illustrated in Fig. 1. In the next sections, we describe the transport and emission models in more detail.

2.1. Transport model

The first step of gtfs2emis is to run the function transport_model. The only data input required is a public transport data set in the GTFS format. A GTFS feed is a zipped file that gathers detailed geolocated information on scheduled services, including its stops,

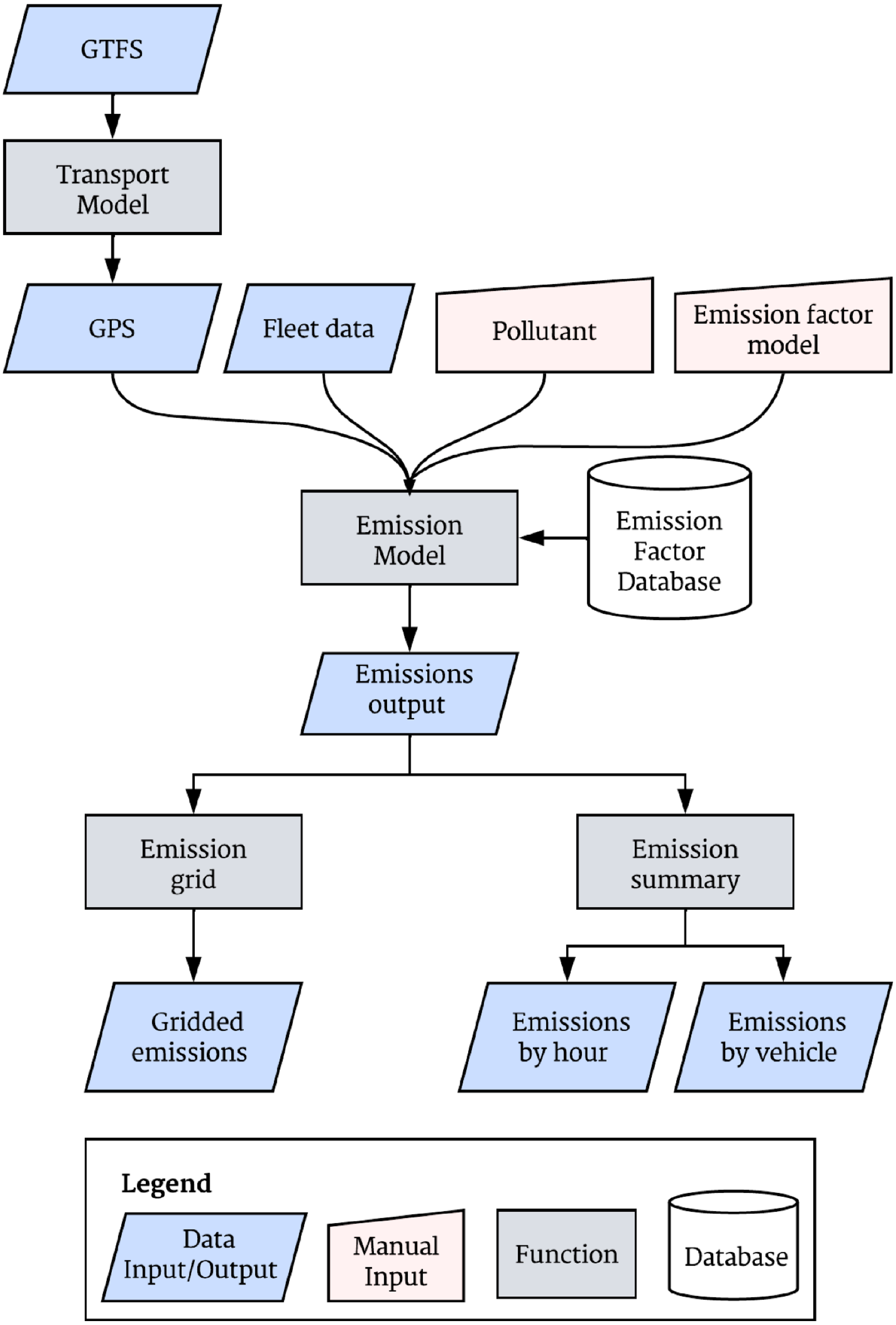

Fig. 1. Data requirements and workflow of gtfs2emis package.

routes, trips, timetables, and calendar, organized in structured text files. Each file is a plain text file with standard columns and formats. Similar to a relational database, with key columns linking the information between routes, trips, timetables, stops, etc. The general schema of GTFS is shown in Fig. 2, with key columns highlighted as the endpoints of the arrows[1].

The transport_model R function converts a GTFS data feed into a GPS-like table with space–time positions and speeds of public transport vehicles. It does so by interpolating the space–time position of each vehicle in each trip, considering the network distance and vehicle speed between stops registered in the inputted GTFS. The space–time positions of vehicles are interpolated along the road

[1] More information about the GTFS standard format can be found at ≪https://developers.google.com/transit/gtfs≫.

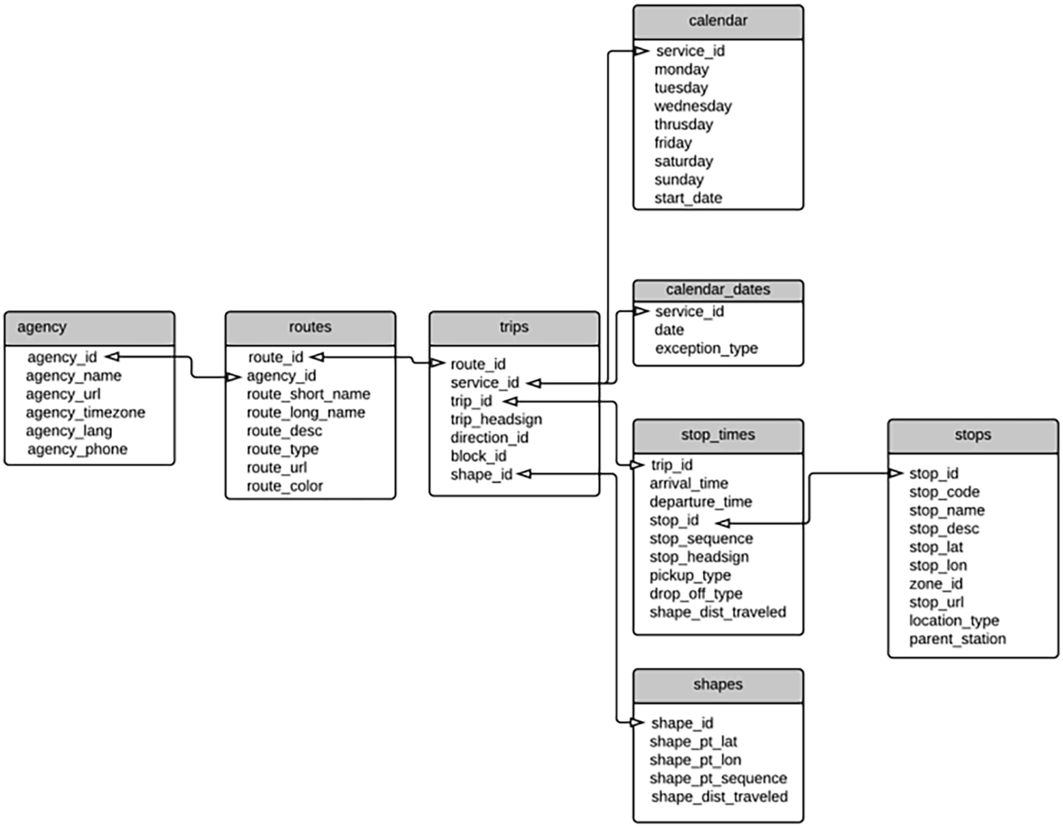

Fig. 2. Schema of the General Transit Feed Specification (GTFS) data format.

network given a maximum spatial resolution set by the user to guarantee a maximum distance between two consecutive data points in each trip. The model assumes a constant average speed between consecutive stops, although the speed of vehicles can vary between segments of the same route and across routes according to the GTFS input data. The step-by-step computational process of the transport model is presented in more detail by Pereira et al. (2022).

By using GTFS data, the gtfs2emis model becomes easily applicable to various cities across multiple countries due to the way the GTFS data standard has become popular among transport agencies worldwide. GTFS feeds are now widely adopted by thousands of transport agencies worldwide and can be freely downloaded from public repositories such as TransitLand (2023) and TransitFeeds (2023). The method can be applied to data from any context as long as it is organized following the data structure and headers of the GTFS specification. A caveat of using GTFS data, though, is that they are based on information on scheduled services. Consequently, emission estimates based on this type of dataset do not consider the variability of travel-time estimates due to unplanned factors such as traffic incidents or nonrecurrent congestion levels. Recurrent congestion, nonetheless, is normally taken into account by transport authorities in the planning service schedules[2].

The output of the transport_model R function is a trajectory data table with the space–time position and speed of each trip segment for every vehicle of the public transport system. A simplified example is shown in Table 1. Each row is a road segment that connects two consecutive public transport stops, where “from_timestamp” represents the time that each vehicle departs from a stop (“from_stop_id”), while “to_timestamp” conveys the arrival time to the next stop (“to_stop_id”). The R function transport_model computes the distance of each stop sequence, mean speed, as well as cumulative distance considering the network distance along the shape of public transport routes.

2.2. Emission model

The second step of gtfs2emis is to run the R function emission_model. This operation estimates the pollutants emitted by each public transport vehicle at each road segment and the time of the day. To do so, it combines the space–time positions and speeds generated by the transport model with some fleet characteristics provided by the user and emission factors data available in the gtfs2emis package.

[2] For studies that examine how scheduled services in GTFS format may deviate from GPS data, see the works of Wessel et al (2017), Braga et al (2020), and Liu et al (2022).

Table 1
Simplified example of the R function transport_model output of one trip.

stop_sequence	from_stop_id	to_stop_id	from_timestamp	to_timestamp	speed	dist	cumdist
1	410,003,313	410,003,312	5:00:00	5:01:55	10.5	336.5	336.5
2	410,003,312	410,003,314	5:01:55	5:03:50	20.7	660.3	996.9
3	410,003,314	410,003,317	5:03:50	5:05:45	7	224.4	1221.3
…	…	…	…	…	…	…	…
21	110,001,116	1,113,267	5:36:56	5:38:51	14	447.7	8253.1
22	1,113,267	130,001,842	5:38:51	5:39:17	17	121.2	8374.2
23	130,001,842	130,001,849	5:39:17	5:40:06	17	231.2	8605.4

Obs: “speed” is given in km/h; “dist” is the distance of the segment in meters; “cumdist” is the cumulative distance in meters.

The estimates of hot exhaust emissions for each vehicle are given by equation (1).

$$EH_{i,j,k,l} = L_i \cdot EF(v)_{i,j,k,l} \tag{1}$$

Where:
$EH_{i,j,k,l}$ is the emission for the street link i, vehicle category j, fuel k, and age l;
L_i is the length of the street link i (km);
$EH(v)_{i,j,k,l}$ is the emission factor (g/km) for the vehicle of category j, fuel k, age l traveling at the average speed v.

The emission_model function requires users to input (i) the output of the transport model step, (ii) a data table with the characteristics of the public transport fleet, and (iii) a selection of which pollutants and emission factor model should be considered. The data on the fleet profile generally includes the following vehicle characteristics: age, type of vehicle, fuel, and technology.

The gtfs2emis package currently includes different emissions factor sources, covering European countries, the United States, and Brazil. This pre-processing included reading different file sources, standardizing variable names, creating dictionaries for each emission factor provider, as well as organizing the data to be easily analyzed in the R environment. Different emission factor models require information on different fleet characteristics. Table 2 describes the required information for the available emission factor models available in the package.

Users can input fleet data in two ways. The simplest option is a table with the average composition of the fleet that indicates the proportion of vehicles with each combination of characteristics (e.g. vehicle category, fuel, model year, technology). This is a general type of data commonly held by most transport agencies, which helps to make the gtfs2emis model easily applicable in data-poor contexts. Alternatively, users may input a detailed table with the characteristics of each individual vehicle and indicate to which public transport route each vehicle is assigned.

A key element of gtfs2emis is selecting the emission factor model that should be used to estimate public transport emissions. Emission factors are empirical functional relations between pollutant emissions and the activity that causes them (Franco et al., 2013). Various emission factor models are developed by environmental agencies based on local data using dynamometer testing and PEMS (CETESB, 2019; EMEP/EEA, 2019a; CARB, 2021).

The gtfs2emis package currently includes four emission factor models from Europe, the United States, and Brazil. More information on these models is presented below. Each model supports a different set of pollutants. For CETESB, EMEP/EEA, and MOVES/EPA emissions databases, users can also estimate energy consumption rates. Overall, the package allows users to estimate over 16 pollutants (Table 3).

The localized scaling of emission factors allows users to apply emission factor models originally generated for other contexts and vehicle types to a particular context of interest. In gtfs2emis, for instance, we presented an adjusted scale to apply European emission factor models to estimate emissions for the Brazilian context, which allows accounting for vehicle speeds, which would not be possible in the current CETESB emission factor model. However, the provided MOVES emission factor refers to the unlocalized rates, which do not incorporate local vehicle operating modes but instead use the default average speed driving cycles. The use of localized MOVES EF is appropriate when real bus trajectory data are available, especially for regions outside the U.S., as the differences between localized and default emission rates can be more expressive, as noted by Liu et al. (2013), Perugu (2019) and Shan et al. (2019b).

Table 2
Hot-exhaust emission factor models, their respective vehicle categories, and fleet variables.

Region	Emission factor model	Bus categories	Variables of vehicle and route link characteristics
Brazil	CETESB (2019)	Micro, Standard, Articulated	Age, Fuel, EURO stage
Europe	EMEP/EEA (2019)	Micro, Standard, Articulated	Fuel, EURO stage, technology, load, slope
United States	MOVES/EPA (2021)	Urban Buses	Age, Fuel
United States	EMFAC/CARB (2021)	Urban Buses	Age, Fuel

*The scripts to pre-process raw emission factors data to the package format are available in the gtfs2emis documentation.

Table 3
List of pollutants included in gtfs2emis package by emission factor model.

Region	EF Model	Pollutants
Brazil	CETESB	CH_4, CO, CO_2, ETOH, FC (Fuel Consumption), FS (Fuel Sales), gCO_2/KWH, gD/kWh (grams of diesel per kWh), HC, KML (total traveled distance per year), N_2O, NH_3, NMHC, NO, NO_2, NO_X, PM_{10}, and RCHO (Aldehyde)
Europe	EMEP / EEA	CH_4, CO, CO_2, EC (Energy Consumption, in MJ/km), FC, N_2O, NH_3, NO_X, PM_{10}, SPN_{23} (#/kWh), and VOC
United States	EMFAC / CARB	CH_4, CO, CO_2, N_2O, NO_X, PM_{10}, $PM_{2.5}$, ROG (Reactive Organic Gases), SO_X, and TOG (Total Organic Gases)
United States	MOVES / EPA	CH_4, CO, CO_2, EC (Energy Consumption, in kJ/km), HONO (Nitrous Acid), N_2O, NH_3, NH_4, NO, NO_2, NO_3, NO_X (NO + HONO + NO_2), PM_{10}, $PM_{2.5}$, SO_2, THC, TOG, and VOC

2.2.1. *Europe*

European exhaust emission factors come from the air pollutant inventory guidebook published by the European Monitoring and Evaluation Programme at the European Environment Agency (EMEP/EEA, 2019a), developed using the COPERT 5.4 Software. In this model, urban buses are classified into three categories, four fuel types, and six euro stages (Table 4).

The EMEP-EEA speed-dependent emission factors for diesel urban buses have been taken from the Handbook Emission Factors for Road Transport (Notter et al., 2019), for Euro I to Euro VI emission standards. The database considers emission factors with average speeds higher than a given minimum speed, which can be at least 12 km/h depending on the pollutant. Distinct parameters of emission factors were considered for Euro V, according to the control technology, which can be Exhaust Gas Recirculation (EGR) or Selective Catalytic Reduction (SCR). According to EMEP/EEA (2019a), it is estimated that 75% of Euro V heavy-duty vehicles are equipped with SCR. For the category of Compressed Natural Gas (CNG) buses, it has an additional emission standard known as Enhanced Environmental Vehicles (EEV), since it may have different combustion and after-treatment technology, being associated with lower PM and NO_X emission rates compared with diesel buses (EMEP/EEA, 2019a). Only older CNG buses are classified as EURO I, II, or III.

EMEP/EEA database can account for variations in emission factors based on passenger load and street slopes. The current version of gtfs2emis allows users to set the mean load factor for all trips. The package also allows users to input a raster file with digital elevation model data, which is then automatically used to calculate the average terrain slope between consecutive stops. Detailed aspects of hot exhaust emissions rates for Europe are described in the EMEP/EEA Guidebook (EMEP/EEA, 2019b).

2.2.2. *United States*

The gtfs2emis package currently includes two emission factor models for the United States. The first one is the EMFAC (EMission FACtor) model, developed by the California Air Resources Board (CARB, 2021). In this model, the running exhaust emission factors vary according to vehicles' speed following speed bins from 5 to 90 mph (8.04 to 144.84 km/h), with increments of 5 mph (8.04 km/h). The gtfs2emis package currently includes only emission factors representative of the annual season and the California statewide area. However, the EMFAC model contains data for other seasons (summer and winter), and geographic areas (Air Basin, Air District, Metropolitan Planning Organization, Country, and Sub-Area), which could be included in future versions of the package.

The second emission factor model for the United States included in gtfs2emis comes from the Vehicle Emission Simulator (MOVES), developed by the Environmental Protection Agency (EPA, 2020). This is the official model for transportation conformity analysis in the US outside the state of California. The gtfs2emis package stores default hot-exhaust emission factors from the MOVES database according to pollutant, fuel type, model year, speed range, and vehicle age. Both emission factor models for the US include data for urban buses without distinction of vehicle size. The available data are classified according to fuel type (CNG, Diesel, Gasoline), model year, and reference year (Table 5). A detailed discussion on heavy-duty emission rates is found in EPA (2020).

2.2.3. *Brazil*

Finally, the gtfs2emis package includes an emission factor model for Brazil, developed by the Environmental Company of São Paulo (CETESB, 2019). In the package, we include the database pre-processed by Ibarra-Espinosa (2022), who provided corrections to ensure data quality and completeness, such as dealing with missing data, and anomalous values for a few vehicle types, model years, and pollutants. The list of available emission factors by vehicle type is summarized in Table 6.

While the CETESB model does not account for the effect of vehicle speed, gtfs2emis allows one to use adjusted factors to consider the effect of average speed. This correction takes into account the European EMEP/EEA model as a reference based on the expression:

$$EF^{scaled}_{i,j,k,l}(V) = EF^{local}_{i,j,k,l} \cdot \frac{EF_{i,j,k,l}(V_i)}{EF_{i,j,k,l}(V_{dc})} \tag{2}$$

Where:

$EF^{scaled}_{i,j,k,l}(V)$ is the speed-adjusted emission factors for each street link i, bus type j, fuel k, age l at traveling at speed V;

$EF^{local}_{i,j,k,l}$ is the local emission factor;

$EF_{i,j,k,l}(V_i)$ is the EEA emission factor at the speed of V_i;

$EF_{i,j,k,l}(V_{dc})$ is the EEA emission factor at the mean speed of urban driving conditions (19 km/h) adopted by EMEP/EEA (2019a).

Table 4
Summary of all vehicle classes covered by the EMEP/EEA Tier 2 methodology (EMEP/EEA, 2019b).

Bus category	Fuel	Euro Stage
Standard (15–18 t)	CNG	Euro I, Euro II, Euro III, EEV
	Diesel, Biodiesel	Conventional, Euro I - Euro VI
	Hybrid Diesel	Euro VI
Midi (<15 t), Articulated (>18 t)	Diesel	Conventional, Euro I - Euro VI

Table 5
Summary of all vehicle classes included in gtfs2emis of MOVES and EMFAC.

Source	Fuel	Model year	Reference year[1]
EMFAC	Diesel	1982–2020	2010–2020
	CNG	1988–2020	
	Gasoline	1990–2020	
MOVES	Diesel, CNG, Gasoline	1985–2022	2015–2022

Note: [1]The year in which the emissions inventory is performed.

Table 6
Summary of all vehicle classes included in gtfs2emis of CETESB.

Vehicle category	Fuel	Model year
Standard bus	Diesel	1990–2019
Micro bus		
Articulated bus		

2.2.4. *Comparison of emission factor models*

Estimates of public transport emissions can vary significantly because of fleet characteristics and driving conditions. Previous research has shown that vehicle age influences emission factor estimates, as degradation of emissions and engine technology directly affect emission efficiency (Frey et al., 2020). Nonetheless, much of the results from emission estimates are affected by the different methods of emission factor models. Using the emission factor data included in gtfs2emis, Fig. 3 shows how different emission factor models generate estimates of CO_2, NO_X, and PM_{10} for a standard/conventional urban bus at different speeds.

The output of the R function emission_model is a list with several vectors and data frames with emission estimates, emission factors, vehicles' travel speed and distance, and associated information on vehicle characteristics (fuel, age, tech, euro, vehicle type).

Finally, gtfs2emis has two functions to help users analyze the data outputs. For convenience, the emis_grid function can be used to examine the spatial distribution emission estimates while the emis_summary allows users to easily aggregate emission estimates by type of vehicle and time of the day.

3. Results of the case study

To demonstrate the package, in this section we present a reproducible example using the gtfs2emis v.0.1.0 that estimates the urban bus emissions of the public transport system of São Paulo (Brazil). Also, we analyze the effects of hot-exhaust emissions per capita using

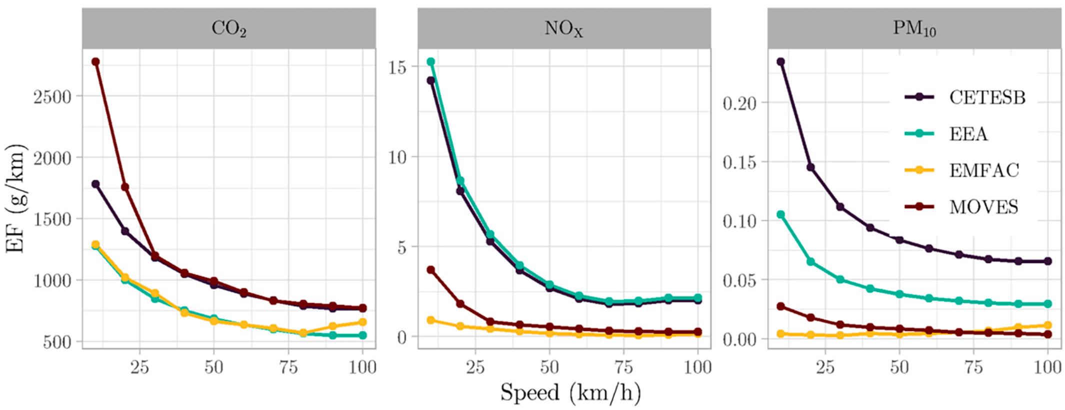

Fig. 3. CO_2, NO_X, and PM_{10} emissions factor for a standard urban bus. Obs. Emission estimates were calculated considering a standard urban bus with Euro III standard for Europe, and the 2011 model year for Brazil and the United States.

data on urban bus occupancy from Arbex and Cunha (2020), which contains hourly data of passengers on board of each bus route between stops on a typical business day for all the bus routes in São Paulo. Approximately 54% of all motorized trips in São Paulo in 2017 were conducted in the public transport system, of which more than half (54%) were done in buses (Metrô, 2019). The case study illustrates how the gtfs2emis model can be used to examine the spatial and temporal distribution of public transport emissions on a typical business day. The data and code to replicate this analysis are shared openly by Vieira et al. (2022).

3.1. Input data

The public transport data in GTFS format for the city of São Paulo was retrieved from the São Paulo Transportation Agency (SPTRANS). It includes all scheduled services for the month of June 2019. If we only consider public transport services by bus, the transport system has 1,323 routes and 193,765 trips on a typical business day. Fig. 4 shows the bus routes of São Paulo, where route links with lower transparency (darker shades of red) indicate higher densities of bus traffic.

The data on the characteristics of the bus fleet of São Paulo were also obtained from SPTRANS. The relevant information to compute emissions were vehicles' age, fuel, and type. The data set does not identify which vehicles are assigned to which routes. In cases like this, we perform a proportional assignment, assuming that all routes have the same distribution of vehicles — which is ultimately based on fleet type frequency. This limitation could underestimate the emissions of particular routes or neighborhoods if they are systematically served by older vehicles, for example. This limitation could be minimized if there were additional information on which vehicles are allocated to specific routes or regions of the city.

The bus fleet of São Paulo has 14,057 urban buses, of which 98.6% are powered by diesel. The remaining 1.4% (202 vehicles) were electric buses. For the purpose of this study, these electric vehicles were removed from the analysis because they do not account for tailpipe emissions. Moreover, the diesel bus fleet consists of Midi buses (2700 buses or 19.5% of the total fleet), Standard buses (8754 buses or 63.2%), and Articulated buses (2401 buses or 17.3%). Fig. 5 shows the composition of the bus fleet of São Paulo by age and category.

Each bus type in São Paulo was associated with an equivalent category in the EMEP/EEA database, following a correspondence table proposed by Miller & Posada (2019) — who analyzed the Brazilian regulation for heavy-duty vehicles and its equivalence with

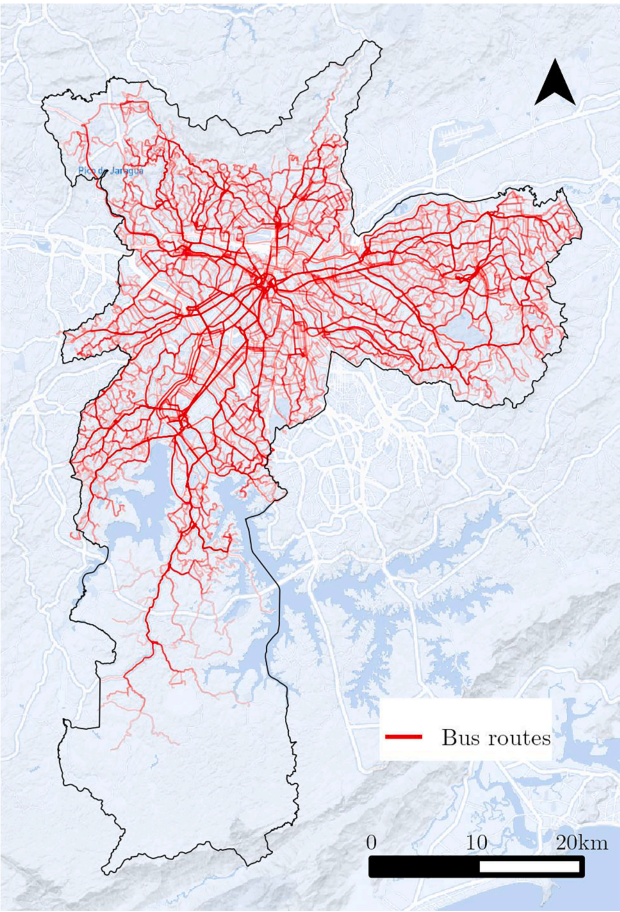

Fig. 4. Bus routes of São Paulo, June 2019.

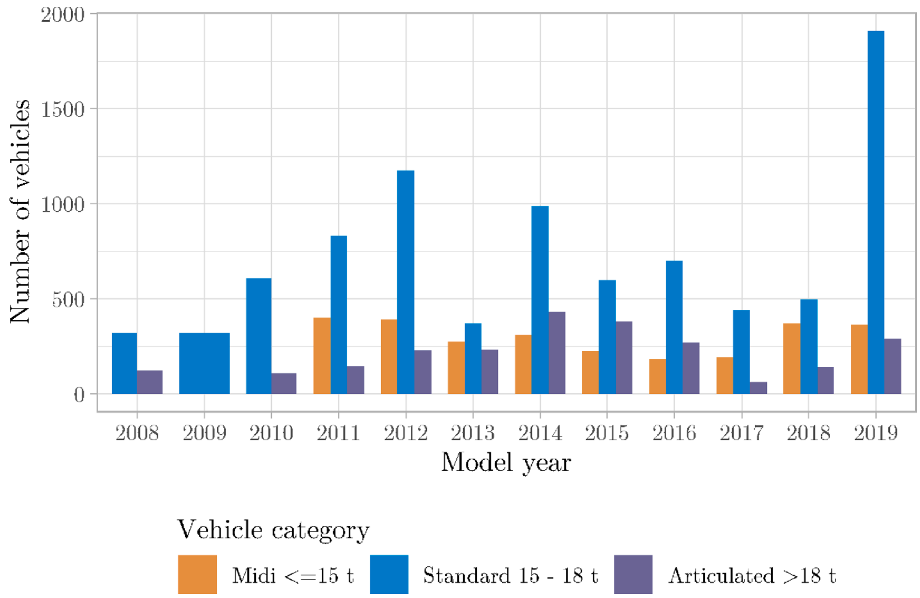

Fig. 5. Urban bus fleet of São Paulo by type of vehicle and manufacturing year.

Table 7
Relationship between urban buses category.

SPTRANS category	Passenger capacity	EMEP/EEA category
Micro	21	Midi bus (<15 t)
Mini	40	
Midi	60	Bus Standard (15–18 t)
Basic	68	
Padron	88	
Articulated	112	Articulated bus (>18 t)
Biarticulated	170	

Euro Standards. According to this study, Brazilian urban buses from 2004 to 2011 are equivalent to Euro III technology, while vehicles from 2012 to 2022 are equivalent to Euro IV. Regarding the vehicle size, we adopted the correspondence table between SPTRANS' and EMEP/EEA's fleets proposed by IEMA (2017), as shown in Table 7.

3.2. Transport model

The first step of the gtfs2emis model is to run the transport model component. Before running the model, we used the gtfstools package (Herszenhut et al., 2022) to filter the GTFS feed and keep only bus trips on a typical business day. In this case study, no changes in the total number of trips were found between weekdays, so we adopted Wednesday as a baseline. The output data of the R function transport_model is a spatial data set that represents every interval of GPS data points between stops as a line segment for each trip.

Table 8
Main statistics of the GTFS used in the case study.

Variable	Value	Units
25th percentile of speed	10.3	km/h
Mean speed	14.3	km/h
75th percentile of speed	20.3	km/h
Average trips per route	146.5	trips
Average bus stops per route	15.5	stops
Total bus stops	97,134	stops
Total VKT	2,89 $\times$ 10^6	km
Total traveled time	2,19 $\times$ 10^5	hours
Total routes	1323	routes

Table 8 summarizes some key statistics of the bus public transport system of São Paulo and the output of the transport model. According to the GTFS data, the bus system of São Paulo totals approximately 2.89 million vehicle kilometers traveled (VKT) on a typical business day, with an average speed of 14.2 km/h. This result based on planned services in the GTFS data differs from the estimates generated by IEMA (2019) using recorded GPS data for June 2019, which led to approximately 2.1 million VKT, and 16.6 km/h. These differences of 27.3% for VKT and −16.9% for average speed can be attributed to changes in service levels between the planned and provided services, service disruptions, etc.

3.3. Emission estimates

We estimate hot exhaust emissions of our public transport system with the emission_model function, using the EEA emission factor model to account for the effect of vehicle speeds on emission levels. The results are consolidated into a single data structure.

According to our estimates, the public transport bus system of São Paulo emitted approximately 3.34 Gt of CO_2, 35.11 t of NO_X, 344.76 kg of PM_{10}, and 73.4 kg of CH_4, on a typical business day in June 2019. Adjusting the results by traveled distance produces average emission factors of 1157.50, 12.15, 0.12, and 0.03 g/km for CO_2, NO_X, PM_{10}, and CH_4, respectively. In the next sub-section, these results are compared with a similar study conducted in São Paulo for urban buses.

Fig. 6 shows the NO_X emissions distributed by the time of the day. As expected, the morning and afternoon peak periods present substantially high emission levels. This is largely due to how these periods generally have both higher service levels and lower speeds. These rush hour periods are also the times with the highest levels of air pollution exposure because of a combination of higher emission levels and a larger number of commuters (Sanderson et al., 2005).

The gtfs2emis package also allows us to examine how public transport hot-exhaust emissions vary in space. Fig. 7 shows the spatial distribution of CO_2 and NO_X bus hot-exhaust emissions in São Paulo. It also presents the downtown area of the city, where important public transport routes start and end. The spatial patterns are similar as both pollutants are directly proportional to total VKT. The maps indicate in dark colors the areas with higher emission concentrations. Such areas can help inform further research on air pollution exposure, as it usually represents busy roads with a high frequency of public transport services and consequently high VKT, often due to the location of bus terminals or overlaps of multiple bus routes. These results are in line with previous research that show public transport terminals as hot spots of pollution concentration and exposure (Yang et al., 2015; Nogueira et al., 2019; Shan et al., 2019a).

The CO_2e (carbon dioxide equivalent) and NO_X running exhaust emission rates from buses represent a substantial proportion of the total emissions generated during the entire life cycle of the vehicle. Specifically, in the case of CNG and diesel transit buses, Pump-To-Wheel CO_2e emissions contribute to over 80 % of life-cycle CO_2 emissions (Xu et al., 2015). Similarly, for a diesel transit bus, tailpipe NO_X emissions account for approximately 78% of life-cycle NO_X emissions (Xu et al., 2015; Cuéllar-Álvarez et al., 2023). However, hot-exhaust particulate matter represents only a small share of the total diesel urban bus life cycle, where resuspension, wear, and well-to-pump emissions processes play a significant role (Xu et al., 2015; Cuéllar-Álvarez et al., 2023). In addition, the increase in gross vehicle weight contributes to more non-exhaust PM_{10}, especially for newer technologies, electric and hybrid powertrains (Beddows & Harrison, 2021; Piscitello et al., 2021). In this sense, to develop informed policies and decision-making, it is essential to have precise and improved bus emission factors not only for the urban bus life cycle but also for the non-exhaust process.

Fig. 8 shows the spatial distribution of NO_X for three periods, between 0:00 and 2:59 am; from 6:00 am to 08:59 am, and from 9:00 pm to 11:59 pm. The spatial aggregation uses the H3 geospatial indexing system, which partitions the world into hexagonal cells at various spatial resolutions ranked from 1 to 15. We adopted the hexagonal grid of resolution #9, where each cell has an area of 0.11 km^2. Our estimates indicate a total of 35.1 t of NO_X emitted in the morning peak, which corresponds to 18.4% of all daily NO_X hot-exhaust emissions. This pollutant is particularly relevant in the context of air pollution from heavy-duty vehicles, as most vehicles use diesel.

For contexts where detailed data on transit ridership is available, it is possible to combine these data with the results from gtfs2emis to examine how emission levels per passenger vary across space and time. Fig. 9 shows the CO_2 hot-exhaust emissions distributed by time of the day, along with additional information on total passengers, overall vehicle capacity[3], total passengers, average vehicle occupancy, emissions per capita, and emissions per km per capita. As expected, the morning and afternoon peak periods present substantially higher emission levels and total numbers of passengers. Emission levels tend to decline over the evenings, presenting a sharp drop during nighttime when bus frequencies are extremely low. The level of occupancy is highest at the morning peak while remaining below 10% for midnight trips. These results in lower emissions per capita in the morning and afternoon peaks, and 2–3 times more emissions for midnight trips. In a similar study, Waraich et al. (2020) found rates of 87–90 g of CO_2 per capita for urban buses in Montreal, while our results indicate less than 27 g of CO_2 per person. Such differences can occur due to public transport characteristics (levels of vehicle occupancy, frequency of services between bus networks) and variations in vehicle technology.

From these results, at least two different strategies could be considered to reduce emissions emissions per capita. One would be adopting discount fares for off-peak periods to increase transit ridership. For instance, according to emission factor data for São Paulo, increasing vehicle occupancy from 5 to 50% could substantially reduce hot-exhaust CO_2 emission per capita considering that this increase in passenger load would only cause a 20% increase in total hot-exhaust CO_2 emission. Another strategy would be deploying smaller vehicles in periods with lower occupancy rates. If only micro-bus operated overnight trips (between 0 and 4 a.m.), the CO_2 hot-

[3] Estimated as the mean vehicle capacity of São Paulo's urban bus fleet, which has different vehicle maximum capacities depending on the category (microbus, standard, articulated, and biarticulated).

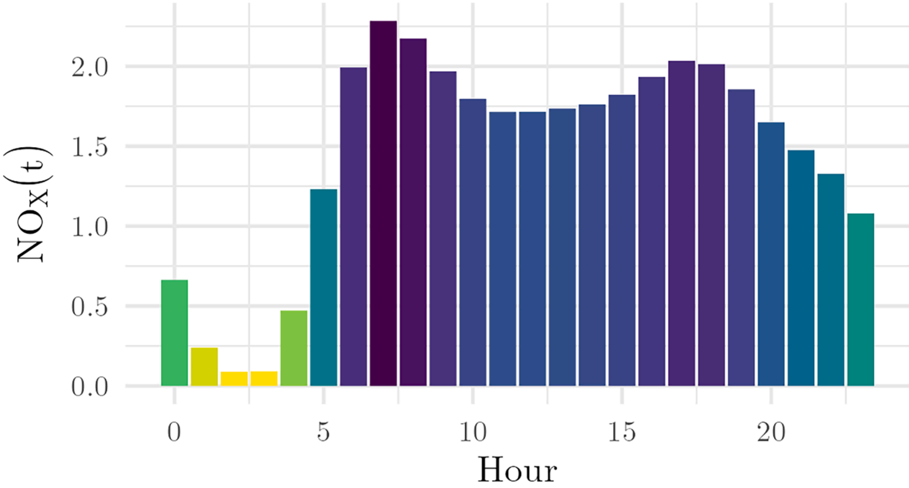

Fig. 6. NO_X hot-exhaust emissions aggregated temporally.

exhaust emissions per capita in this period would be reduced by 19.6 % (from 33.4 to 26.8 g of CO_2 per person), compared to the current scenario where all vehicles are equally distributed in time.

Fig. 10 illustrates the spatial distribution of CO_2 hot-exhaust emissions and passengers in São Paulo's transportation network, focusing on areas of high emissions and potential targets for emissions reduction. The major transit corridors exhibit the highest levels of emissions and passenger volume. However, we do not find particularly high emissions levels per passenger along these corridors. While we do see high emissions per capita near the city, it is in a few peripheral routes that we find the highest emissions per capita.

To examine the influence of fleet age on total emissions in the case study, Fig. 11 shows the marginal emission factors aggregated by euro stage, vehicle category, and pollutant. The Euro III vehicles, produced before 2012, strongly influence overall PM_{10} and NO_X hot-exhaust emissions. This gives us a general estimate of the potential environmental benefits that could be achieved with policies to renew the public transport fleet, as discussed in depth in the literature (Morales Betancourt et al., 2019; Nogueira et al., 2019; Frey et al., 2020; Morales Betancourt et al., 2022). If all buses with Euro III stage in the public transport system of São Paulo were to be replaced by new buses in the same category but with Euro V standards, the total hot-exhaust emissions of NO_X would be cut by 4.4%, respectively. A more radical policy of electrifying all Euro III vehicles, for example, would drastically cut hot exhaust emissions, avoiding 24.4% of total CO_2 (815.9 t), 46.5 % of PM_{10} (160.4 kg), and 24.3% of NO_X (8.5 t) in the atmosphere. Although this type of policy could achieve a substantial reduction in hot-exhaust emissions, other emissions processes such as cold-start, non-exhaust, and life-cycle emissions should be conducted to fully assess the benefits of EV adoption, as noted by Chan et al. (2013); Zhao, Walker, and Surawski (2021); and Cuéllar-Álvarez et al. (2023).

3.4. Results benchmark

How accurate are the results of the gtfs2emis model applied to this case study of São Paulo? Unfortunately, there is no blueprint against which to compare our model results. For the sake of a benchmark, though, we compare our estimates against the results from a recent research report published by the Institute for Energy and the Environment (IEMA, 2019). In this report, the authors estimated PM_{10} and NO_X hot-exhaust emissions for the public transport of São Paulo using historical GPS data of buses for June 2019. The authors also used the emission factor model developed by the European Environment Agency (EMEP/EEA) and a fleet data set provided by SPTRANS, using very similar input data as we used in our model. In other words, the main difference between the methods used in our case study and IEMA's report is the transport input data.

As mentioned in Section 3.2, the differences between the GTFS and GPS input data make our estimates of VKT and average vehicle speed to be 27% larger and 17% smaller than the results from IEMA (2019), respectively. Consequently, our estimates of PM_{10} were 26% higher than the value estimated by IEMA (Table 9). The difference in VKT between both methods explains the most divergence in PM_{10} estimates because emission factors are almost entirely affected by traveled distances.

The estimates of NOx emissions, on the other hand, were 42.5% higher by our method compared to the results from IEMA (2019). This large difference occurs because NO_X emission factors are much more sensitive to vehicle speed, particularly low driving speeds below 25 km/h (shown previously in Fig. 3). The small difference of 2.4 km/h in the average speed between the GTFS and GPS data inputs explains approximately 20% of the difference in total NO_X emissions found between the gtfs2emis and IEMA's models. According to Dixit et al. (2017), bus engines tend to emit more NOx emissions under lower speed and/or load conditions, which leads to lower exhaust temperatures, under which emission control systems are less effective. Compared to GTFS feeds, data from onboard GPS devices can better represent the public transport services provided to the population. Emissions estimates are therefore generally more reliable when they are based on GPS records. However, our results show how emission estimates based on GTFS data using the gtfs2emis model converge with estimates generated from GPS data. These findings indicate how the gtfs2emis model can be a valuable

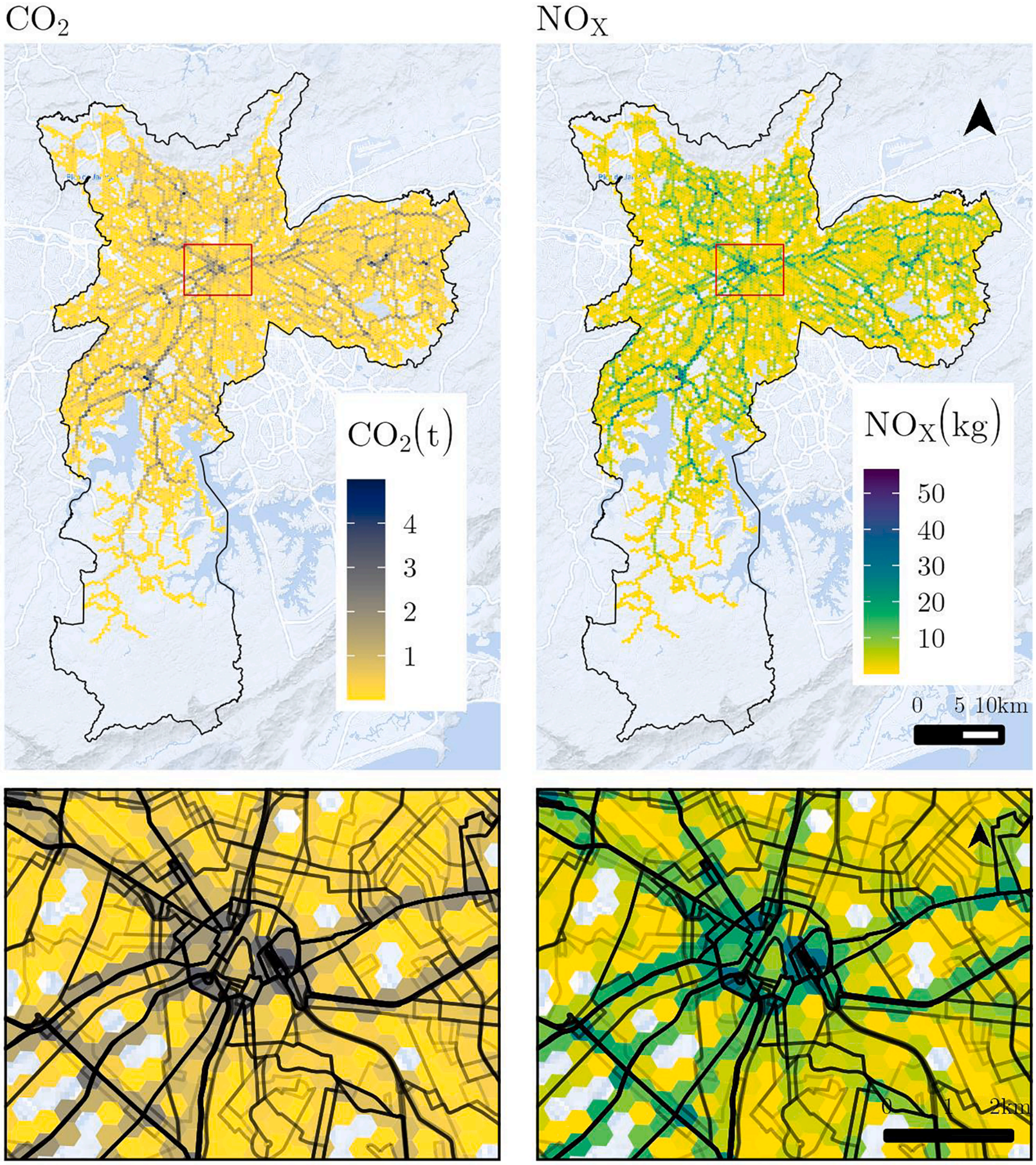

Fig. 7. Total CO_2 and NO_X emissions of buses in São Paulo's public transport system on a typical business day.

alternative to estimate public transport emissions in data-poor contexts where GPS records are unavailable. Our results also stress the importance of generating high-fidelity GTFS data that represents service levels and timetables as accurately as possible.

4. Conclusions

This study introduced the gtfs2emis model, a bottom-up method to estimate public transport emissions. By leveraging GTFS data, a standard format for public transport data widely adopted worldwide, the proposed model can be used in different contexts with few data requirements. The model is freely available through the gtfs2emis package in R, which makes the model easily scalable and applicable to estimate the emissions of various pollutants from single routes to entire public transport systems at high spatial and temporal resolutions. In addition, given the different data structures of the emission factors from CETESB, EMEP/EEA, MOVES, and EMFAC, the package also contributes by providing an easy framework to incorporate urban bus hot-exhaust emission factors into computational emission models.

The paper also presented a case study, where we estimated CO_2, NO_X, CH_4, and PM_{10} emissions of the public transport bus system of São Paulo to illustrate an application of gtfs2emis. We have found that the bus services of São Paulo emitted approximately 3344.9 t of CO_2, 73.4 kg of CH_4, 35.1 t of NO_X, and 344.8 kg of PM_{10} on a typical business in June 2019. The emissions estimates vary substantially

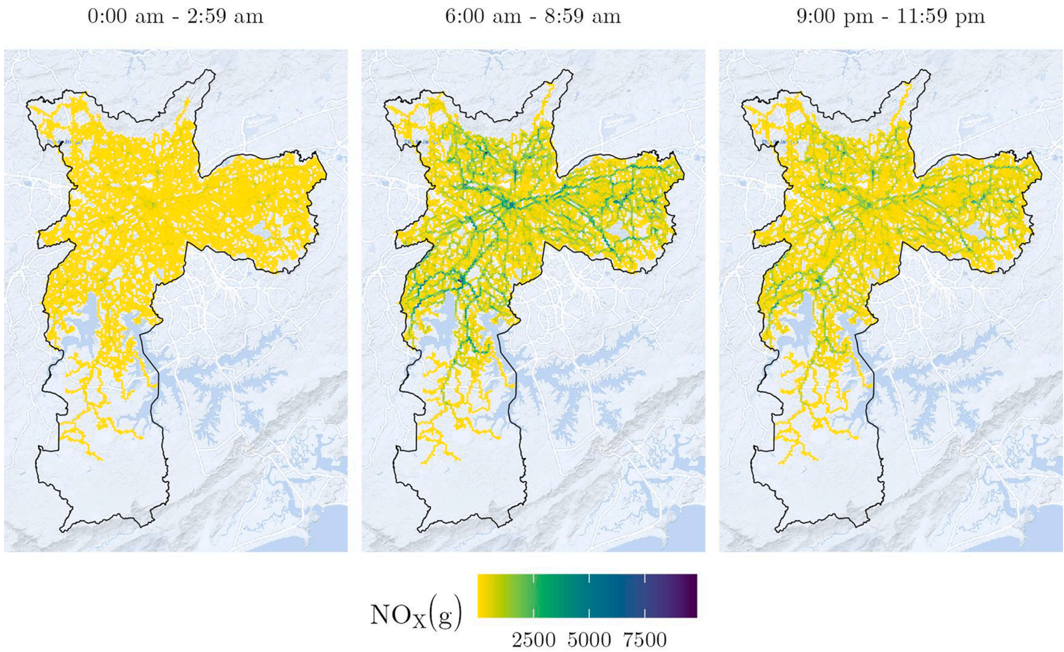

Fig. 8. Spatial and temporal distribution of NO_X emissions.

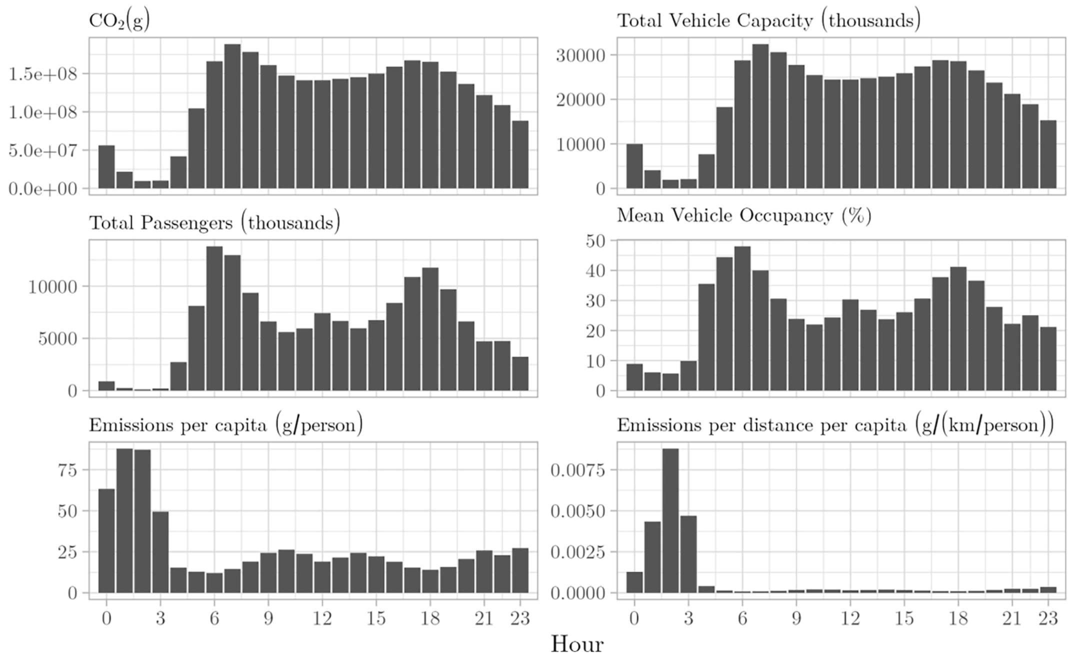

Fig. 9. CO_2 hot-exhaust emission indicators categorized by time of day.

across space and time of the day, with larger levels along roads with larger VKT due to higher service frequency and overlaps of multiple routes. Finally, when analyzing data on vehicle occupancy and emissions per capita, the study highlights the need for targeted emissions reduction strategies in the high-traffic corridors of São Paulo's transportation network, while also emphasizing the importance of considering emissions per capita as a key metric for evaluating the environmental impact of transportation systems.

Comparing our results against emissions estimates published by IEMA (2019) using a similar method but based on GPS records of buses, we estimated approximately 26% – 27% more VKT and PM_{10}. Our results for NO_X were 42% higher based on GTFS data compared to estimates based on GPS records, which largely occurred because of how the GTFS of São Paulo underestimated average

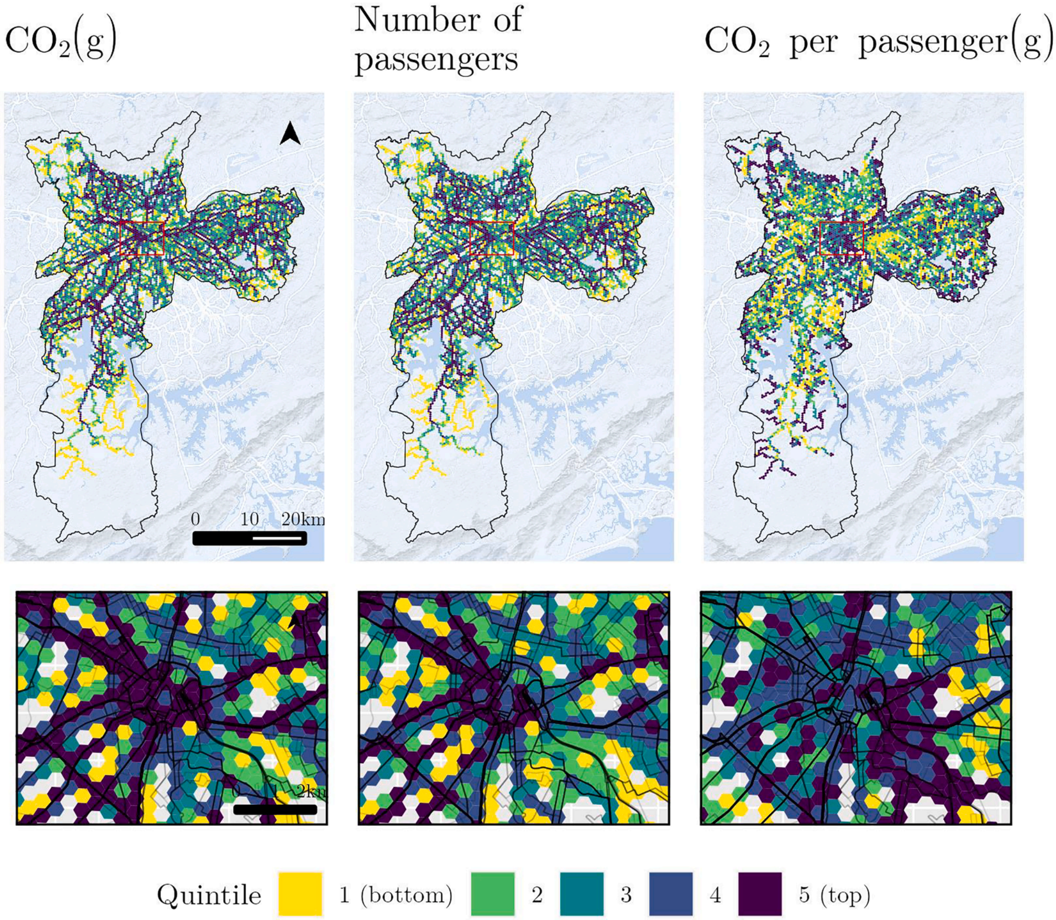

Fig. 10. CO_2 hot-exhaust emissions, total passengers, and emissions per capita.

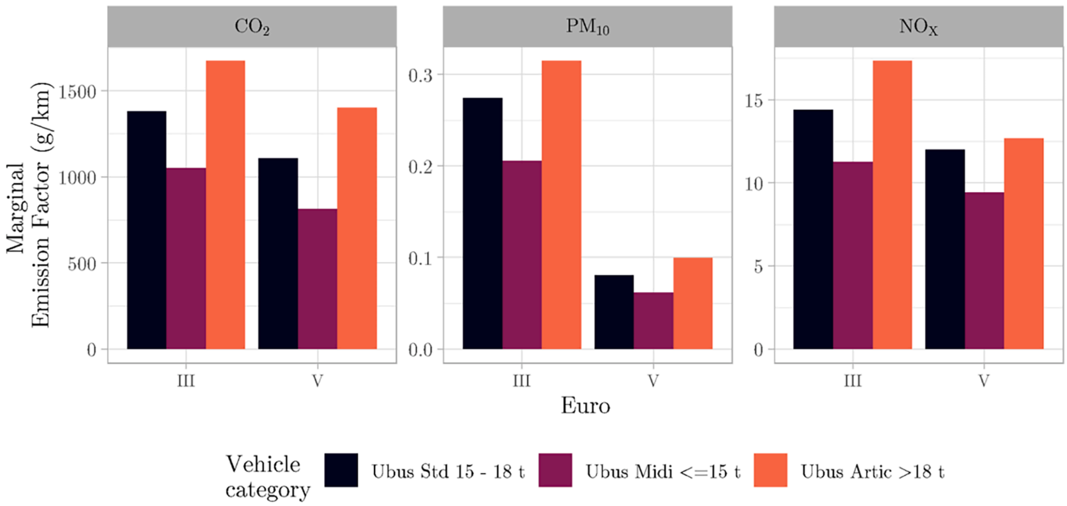

Fig. 11. Hot-exhaust emission factors of CO_2, PM_{10}, and NO_X, according to euro standard and vehicle category.

vehicle speeds by 17% and because NO_X emission factors are highly sensitive to speed values. Despite these differences in overall emissions, our findings suggest that the gtfs2emis model can be a valuable alternative to estimate public transport emissions in data-poor contexts where GPS records are unavailable, especially in developing countries. In this sense, GTFS data can be easier and less

Table 9
Summary of transport model variables and hot-exhaust emissions from gtfs2emis and IEMA (2019).

Source	Transport model		Emissions		Emission Factors	
	VKT (10^6 km)	Mean Speed (km/h)	PM_{10} (kg)	NO_X (t)	PM_{10} (g/km)	NO_X (g/km)
gtfs2emis	2.89	14.2	344.76	35.15	0.119	12.2
IEMA (2019)	2.1	16.6	255	20.2	0.121	9.6
Difference (%)	27.3	−16.9	26	42.5	−1.8	20.9

Note: The fleet data in gtfs2emis and IEMA (2019) were considered equivalent, as the differences in vehicle distribution by category and technology were lower than 1%.

costly to produce and update. Our results also call for the importance of emerging research methods to generate high-fidelity GTFS data based on historical GPS records (Wessel et al., 2017; Elliott & Lumley, 2020; Liu & Miller, 2022).

The current version gtfs2emis has a few caveats. Because the model considers average speeds between bus stops, it does not account for variations in vehicle performance under different driving, acceleration, and deceleration conditions due to signal priorities, stop/start driving, queue jumper lanes, intersections, and bus corridors. The emission factor models currently included in the gtfs2emis package do not allow estimating emissions from idle vehicles, which usually occurs at street intersections and stops to embark/disembark passengers. However, as more extensions of gtfs2emis are implemented, a larger range of investigations of public transport emissions can be performed.

Looking particularly at the emission factor databases, the package does not yet include emissions factors for other vehicles besides urban buses, such as metro, train, or light-rail vehicles. Given the important presence of diesel-powered railways in the public transport systems of many cities worldwide, future updates of gtfs2emis should include emissions factors for rail-based transport modes. Moreover, gtfs2emis currently focuses on hot-exhaust emissions and overlooks other important emission sources such as cold-start, evaporative losses, resuspension, tire, and brake wear. The growing discussion of hybrid and electric vehicles calls for the integration of life-cycle analysis into the overall discussions of electrification policies and could be incorporated into gtfs2emis in the future. Next versions of the gtfs2emis package could also be expanded to include emission factor models to cover more regions of the world, including countries in Asia, Africa, Latin America, and Oceania.

In addition, more research is also needed to validate the results from the gtfs2emis against other methods and local measurements, as suggested by Smit et al. (2010) to improve the emission model. The method still could be improved by incorporating the quantification of variability and uncertainty for the key variables, as discussed by Frey & Zheng (2011). For instance, systematic errors inherent to emission factors models could be used to incorporate uncertainty measures into the emission inventory. The model from EMEP/EEA (2019b), for example, includes a detailed discussion on the precision of emission factors, according to the pollutant, vehicle type, and fuel, which could be incorporated in future developments of gtfs2emis. The overall quantification of errors can be done, for example, by defining uncertainties in the activity factors (total number of trips, vehicle speeds, and composition of vehicle technologies), and using bootstrap simulation as a means to calculate confidence intervals for the emissions inventory.

From a policy-making perspective, the gtfs2emis model could be used to help support policies for a low-carbon transition. The model can be used by transport and environmental agencies, for example, to simulate how much and where CO_2, NO_X, or PM_{10} emissions are reduced by different electrification policy scenarios. Similarly, the gtfs2emis model could help quantify the environmental benefits of fleet renewal investments helping local authorities to prioritize the allocation of new cleaner buses to public transport routes identified with higher pollution levels. Given the capability of gtfs2emis to estimate emission levels and high spatial and temporal resolutions, it could be used in conjunction with other complementary tools that measure other types of emissions to help authorities identify critical zones, such as hospitals and residential areas facing high levels of pollution. This kind of knowledge could inform, for example, in which areas and times of the day authorities could enact driving restrictions to allow only battery-electric or cleaner vehicles to reduce high levels of air pollution exposure.

We believe the proposed gtfs2emis model could also help advance research agendas on the environmental sustainability and health implications of public transport systems. One example of application is using fuel-based emission factors to evaluate emissions performance, as discussed by Singer & Harley (1996). This may be especially relevant to conducting an international comparison between public transport emissions across multiple cities, as this metric shows the efficiency of pollutants by fuel consumption and compares emissions from different vehicles and driving conditions (Kean et al., 2000). In addition, this might open new research avenues to examine how emissions levels per passenger can be affected by different factors such as vehicles' speed and low occupancy vehicles, as investigated by Yu et al (2016) and Waraich et al (2020).

From an environmental justice perspective, the gtfs2emis model could be combined with air quality data to examine which population groups are more or less exposed to the emissions from public transport systems, as discussed in Gardner-Frolick et al (2022) for different applications. This type of research can also be used to better understand the health effects of public transport services and corridors on local communities. While these research questions could be conducted based on GTFS data with the gtfs2emis model, future research could use GPS data to update the speed and departure times of the information of GTFS feed to account for real traffic conditions and improve the accuracy of emission estimates, as done in previous studies (Wessel et al., 2017; Braga et al., 2020; Elliott & Lumley, 2020). Finally, the easy applicability of the gtfs2emis package in R could help further promote open science and reproducible research projects in transport and environmental modeling to pursue more sustainable cities.

Code and data availability. The package gtfs2emis v.0.1.0 as well as the data and scripts used to reproduce this study's result are available in Vieira et al. (2022). The developing version and recent updates can be found at ≪https://github.com/ipeaGIT/gtfs2emis≫. The required packages to run the examples are {aopdata, data.table, gtfstools, gtfs2emis, sf}. Results were modeled in an Intel (R) Xeon(R) Gold 5118 CPU @2.30 GHz, 250 GB memory, but the analysis could be replicated in lower computer specifications with minor changes to the code.

CRediT authorship contribution statement

João Pedro Bazzo Vieira: Conceptualization, Data curation, Investigation, Methodology, Software, Visualization. **Rafael H.M. Pereira:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision. **Pedro R. Andrade:** Conceptualization, Investigation, Methodology, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Institute for Applied Economic Research and the National Institute of Space Research for the support.

References

Arbex, R., Cunha, C.B., 2020. Estimating the influence of crowding and travel time variability on accessibility to jobs in a large public transport network using smart card big data. J. Transp. Geogr. 85, 102671 https://doi.org/10.1016/j.jtrangeo.2020.102671.

Beddows, D.C.S., Harrison, R.M., 2021. PM10 and PM2.5 emission factors for non-exhaust particles from road vehicles: dependence upon vehicle mass and implications for battery electric vehicles. Atmos. Environ. 244, 117886 https://doi.org/10.1016/j.atmosenv.2020.117886.

Böhm, M., Nanni, M., Pappalardo, L., 2022. Gross polluters and vehicle emissions reduction. Nat. Sustainability 5 (8), Art. 8. https://doi.org/10.1038/s41893-022-00903-x.

Braga, C.K.V., Loureiro, C.F.G., Pereira, R.H.M., 2020. Analisando a variabilidade de estimativas de acessibilidade por transporte público a partir de dados de GPS. TRANSPORTES 28 (5), 5. https://doi.org/10.14295/transportes.v28i5.2175.

Brook, R.D., Rajagopalan, S., Pope, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A., Peters, A., Siscovick, D., Smith, S.C., Whitsel, L., Kaufman, J.D., 2010. Particulate matter air pollution and cardiovascular disease. Circulation 121 (21), 2331–2378. https://doi.org/10.1161/CIR.0b013e3181dbece1.

Caiazzo, F., Ashok, A., Waitz, I.A., Yim, S.H.L., Barrett, S.R.H., 2013. Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005. Atmos. Environ. 79, 198–208. https://doi.org/10.1016/j.atmosenv.2013.05.081.

CARB, 2021. EMFAC2021 Volume III (p. 271). California Air Resources Board. Available from: <https://ww2.arb.ca.gov/sites/default/files/2021-08/emfac2021_technical_documentation_april2021.pdf>.

CETESB, 2019. *Emissões veiculares no estado de São Paulo* (Dados Internacionais de Catalogação ISSN 0103-4103; p. 137). CETESB. Available from: <https://cetesb.sp.gov.br/veicular/wp-content/uploads/sites/6/2020/11/Relatorio-Emissoes-Veiculares-no-Estado-de-Sao-Paulo-2019.pdf>.

Chan, E.C., Leitão, J., Kerschbaumer, A., Butler, T.M., 2022. Yeti 1.0: a generalized framework for constructing bottom-up emission inventory from traffic sources. Geosci. Model Dev. Discuss. 1–34. https://doi.org/10.5194/gmd-2022-147.

Chan, S., Miranda-Moreno, L.F., Alam, A., Hatzopoulou, M., 2013. Assessing the impact of bus technology on greenhouse gas emissions along a major corridor: a lifecycle analysis. Transp. Res. Part D: Transp. Environ. 20, 7–11. https://doi.org/10.1016/j.trd.2013.01.004.

Cuéllar-Álvarez, Y., Guevara-Luna, M.A., Belalcázar-Cerón, L.C., Clappier, A., 2023. Well-to-Wheels emission inventory for the passenger vehicles of Bogotá, Colombia. Int. J. Environ. Sci. Technol. https://doi.org/10.1007/s13762-023-04805-z.

Currie, J., Walker, W.R., 2009. Traffic Congestion and Infant Health: Evidence from E-ZPass (Nº w15413). National Bureau of Economic Research. doi: 10.3386/w15413.

Dixit, P., Miller, J.W., Cocker, D.R., Oshinuga, A., Jiang, Y., Durbin, T.D., Johnson, K.C., 2017. Differences between emissions measured in urban driving and certification testing of heavy-duty diesel engines. Atmos. Environ. 166, 276–285. https://doi.org/10.1016/j.atmosenv.2017.06.037.

Elliott, T., Lumley, T., 2020. Modelling the travel time of transit vehicles in real-time through a GTFS-based road network using GPS vehicle locations. Aust. N. Z. J. Stat. 62 (2), 153–167. https://doi.org/10.1111/anzs.12294.

EMEP/EEA, 2019a. EMEP/EEA air pollutant emission inventory guidebook [Publication]. Available from: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2019>.

EMEP/EEA, 2019b. EMEP/EEA air pollutant emission inventory guidebook [Publication]. Available from: <https://www.eea.europa.eu/publications/emep-eea-guidebook-2019>.

EPA, 2020. Exhaust Emission Rates for Heavy-Duty Onroad Vehicles in MOVES3 (Technical Report EPA-420-R-20-018; p. 308). U.S. Environmental Protection Agency. Available from: <https://www.epa.gov/moves/moves-onroad-technical-reports#moves3>.

Franco, V., Kousoulidou, M., Muntean, M., Ntziachristos, L., Hausberger, S., Dilara, P., 2013. Road vehicle emission factors development: a review. Atmos. Environ. 70, 84–97. https://doi.org/10.1016/j.atmosenv.2013.01.006.

Frey, H.C., Zheng, J., 2011. Quantification of variability and uncertainty in air pollutant emission inventories: method and case study for utility NOx emissions. J. Air Waste Manag. Assoc. https://doi.org/10.1080/10473289.2002.10470837.

Frey, H.C., Zheng, X., Hu, J., 2020. Variability in measured real-world operational energy use and emission rates of a plug-in hybrid electric vehicle. Energies 13 (5), 1140. https://doi.org/10.3390/en13051140.

Fu, P., Guo, X., Cheung, F.M.H., Yung, K.K.L., 2019. The association between PM2.5 exposure and neurological disorders: a systematic review and meta-analysis. Sci. Total Environ. 655, 1240–1248. https://doi.org/10.1016/j.scitotenv.2018.11.218.

García, A., Monsalve-Serrano, J., Lago Sari, R., Tripathi, S., 2022. Life cycle CO_2 footprint reduction comparison of hybrid and electric buses for bus transit networks. Appl. Energy 308, 118354. https://doi.org/10.1016/j.apenergy.2021.118354.

Gardner-Frolick, R., Boyd, D., Giang, A., 2022. Selecting data analytic and modeling methods to support air pollution and environmental justice investigations: a critical review and guidance framework. Environ. Sci. Tech. 56 (5), 2843–2860. https://doi.org/10.1021/acs.est.1c01739.

Herszenhut, D., Pereira, R., Andrade, P., Bazzo, J., 2022. gtfstools: General Transit Feed Specification (GTFS) Editing and Analysing Tools (1.1.0). doi: 10.5281/zenodo.6577028.

Huang, Y., Lee, C.K.C., Yam, Y.-S., Mok, W.-C., Zhou, J.L., Zhuang, Y., Surawski, N.C., Organ, B., Chan, E.F.C., 2022. Rapid detection of high-emitting vehicles by on-road remote sensing technology improves urban air quality. Sci. Adv. 8 (5), eabl7575. https://doi.org/10.1126/sciadv.abl7575.

Huang, C., Lou, D., Hu, Z., Feng, Q., Chen, Y., Chen, C., Tan, P., Yao, D., 2013. A PEMS study of the emissions of gaseous pollutants and ultrafine particles from gasoline- and diesel-fueled vehicles. Atmos. Environ. 77, 703–710. https://doi.org/10.1016/j.atmosenv.2013.05.059.

Ibarra-Espinosa, S., Ynoue, R.Y., Ropkins, K., Zhang, X., de Freitas, E.D., 2020. High spatial and temporal resolution vehicular emissions in south-east Brazil with traffic data from real-time GPS and travel demand models. Atmos. Environ. 222, 117136 https://doi.org/10.1016/j.atmosenv.2019.117136.

Ibarra-Espinosa, S., 2022. vein: Vehicular Emissions Inventories (0.9.5) [R]. Available from: <https://cran.r-project.org/web/packages/vein/vein.pdf>.

IEMA, 2017. Inventário de Emissões Atmosféricas do Transporte Rodoviário de Passageiros no Município de São Paulo. Emissiões atmosféricas - Gráficos. Available from: <http://emissoes.energiaeambiente.org.br/graficos>.

IEMA, 2019. Monitor de Ônibus SP. Instituto de Energia e Meio Ambiente. Available from: <http://energiaeambiente.org.br/onibus-sp>.

Kean, A.J., Sawyer, R.F., Harley, R.A., 2000. A fuel-based assessment of off-road diesel engine emissions. J. Air Waste Manag. Assoc. 50 (11), 1929–1939. https://doi.org/10.1080/10473289.2000.10464233.

Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N.(Nil), Baldé, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Zhong, M., 2018. The Lancet Commission on pollution and health. Lancet 391 (10119), 462–512. https://doi.org/10.1016/S0140-6736(17)32345-0.

Linero, H., Yeh, S., Kishimoto, P., Cazzola, P., Fulton, L., McCollum, D., Miller, J., Kyle, P., 2020. The international transport energy modeling (iTEM) open data & harmonized transport database. Zenodo. https://doi.org/10.5281/zenodo.4287423.

Liu, H., Chen, X., Wang, Y., Han, S., 2013. Vehicle emission and near-road air quality modeling for Shanghai, China: based on global positioning system data from taxis and revised MOVES emission inventory. Transp. Res. Rec. 2340 (1), 38–48. https://doi.org/10.3141/2340-05.

Liu, Z., Ge, Y., Johnson, K.C., Shah, A.N., Tan, J., Wang, C., Yu, L., 2011. Real-world operation conditions and on-road emissions of Beijing diesel buses measured by using portable emission measurement system and electric low-pressure impactor. Sci. Total Environ. 409 (8), 1476–1480. https://doi.org/10.1016/j.scitotenv.2010.12.042.

Liu, L., Miller, H.J., 2022. Measuring the impacts of dockless micro-mobility services on public transit accessibility. Comput. Environ. Urban Syst. 98, 101885 https://doi.org/10.1016/j.compenvurbsys.2022.101885.

Liu, L., Porr, A., Miller, H.J., 2022. Realizable accessibility: Evaluating the reliability of public transit accessibility using high-resolution real-time data. J. Geogr. Syst. https://doi.org/10.1007/s10109-022-00382-w.

López, J.M., Jiménez, F., Aparicio, F., Flores, N., 2009. On-road emissions from urban buses with SCR+Urea and EGR+DPF systems using diesel and biodiesel. Transp. Res. Part D: Transp. Environ. 14 (1), 1–5. https://doi.org/10.1016/j.trd.2008.07.004.

López-Martínez, J.M., Jiménez, F., Páez-Ayuso, F.J., Flores-Holgado, M.N., Arenas, A.N., Arenas-Ramirez, B., Aparicio-Izquierdo, F., 2017. Modelling the fuel consumption and pollutant emissions of the urban bus fleet of the city of Madrid. Transp. Res. Part D: Transp. Environ. 52, 112–127. https://doi.org/10.1016/j.trd.2017.02.016.

Metrô. (2019). Pesquisa Origem Destino 2017, p. 152 [Relatório Síntese]. Governo do Estado de São Paulo. Available from: <https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino/resource/b3d93105-f91e-43c6-b4c0-8d9c617a27fc>.

Miller, J., Posada, F., 2019. Norma PROCONVE P-8 de emissões no Brasil. International Council on Clean Transportation. Available from: <https://theicct.org/sites/default/files/publications/P8_emission_Brazil_policyupdate_20190227.pdf>.

Morales Betancourt, R., Galvis, B., Rincón-Riveros, J.M., Rincón-Caro, M.A., Rodriguez-Valencia, A., Sarmiento, O.L., 2019. Personal exposure to air pollutants in a Bus Rapid Transit System: impact of fleet age and emission standard. Atmos. Environ. 202, 117–127. https://doi.org/10.1016/j.atmosenv.2019.01.026.

Morales Betancourt, R., Galvis, B., Mendez-Molano, D., Rincón-Riveros, J.M., Contreras, Y., Montejo, T.A., Rojas-Neisa, D.R., Casas, O., 2022. Toward cleaner transport alternatives: reduction in exposure to air pollutants in a mass public transport. Environ. Sci. Tech. https://doi.org/10.1021/acs.est.1c07004.

Muñoz, P., Franceschini, E.A., Levitan, D., Rodriguez, C.R., Humana, T., Correa Perelmuter, G., 2022. Comparative analysis of cost, emissions and fuel consumption of diesel, natural gas, electric and hydrogen urban buses. Energ. Conver. Manage. 257, 115412 https://doi.org/10.1016/j.enconman.2022.115412.

Nocera, S., Ruiz-Alarcón-Quintero, C., Cavallaro, F., 2018. Assessing carbon emissions from road transport through traffic flow estimators. Transp. Res. Part C: Emerging Technol. 95, 125–148. https://doi.org/10.1016/j.trc.2018.07.020.

Nogueira, T., Dominutti, P.A., Vieira-Filho, M., Fornaro, A., Andrade, M. de F., 2019. Evaluating atmospheric pollutants from urban buses under real-world conditions: implications of the main public transport mode in São Paulo, Brazil, Atmosphere 10 (3), 3. doi: 10.3390/atmos10030108.

Notter, B., Keller, M., Althaus, H.-J., Cox, B., Knörr, W., Heidt, C., Biemann, K., Räder, D., Marie, J., 2019. HBEFA 4.1 Development Report (N° 7309c; p. 95). Bundesamt für Umwelt BAFU. Available from: <https://www.hbefa.net/e/documents/HBEFA41_Development_Report.pdf>.

Pereira, R.H.M., Andrade, P.R., Vieira, J.P.B., 2022. Exploring the time geography of public transport networks with the gtfs2gps package. J. Geogr. Syst. https://doi.org/10.1007/s10109-022-00400-x.

Perugu, H., 2019. Emission modelling of light-duty vehicles in India using the revamped VSP-based MOVES model: the case study of Hyderabad. Transp. Res. Part D: Transp. Environ. 68, 150–163. https://doi.org/10.1016/j.trd.2018.01.031.

Piscitello, A., Bianco, C., Casasso, A., Sethi, R., 2021. Non-exhaust traffic emissions: sources, characterization, and mitigation measures. Sci. Total Environ. 766, 144440 https://doi.org/10.1016/j.scitotenv.2020.144440.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.

Rosero, F., Fonseca, N., López, J.-M., Casanova, J., 2020. Real-world fuel efficiency and emissions from an urban diesel bus engine under transient operating conditions. Appl. Energy 261, 114442. https://doi.org/10.1016/j.apenergy.2019.114442.

Rosero, F., Fonseca, N., López, J.-M., Casanova, J., 2021. Effects of passenger load, road grade, and congestion level on real-world fuel consumption and emissions from compressed natural gas and diesel urban buses. Appl. Energy 282, 116195. https://doi.org/10.1016/j.apenergy.2020.116195.

Sanderson, E., Briggs, D., Jantunen, M., Forsberg, B., Svartengren, M., Šrám, R., Gulliver, J., Janssen, N., 2005. Human exposure to transport-related air pollution. Krzyzanowski, Em M., Kuna-Dibbert, B., Schneider, J. (Orgs.), Health Effects of Transport-related Air Pollution, first ed. World Health Organization, p. 205. Available from: <https://www.euro.who.int/_data/assets/pdf_file/0006/74715/E86650.pdf>.

Shan, X., Chen, X., Jia, W., Ye, J., 2019. Evaluating urban bus emission characteristics based on localized MOVES using sparse GPS data in Shanghai, China. Sustainability 11 (10), 10, doi: 10.3390/su11102936.

Shan, X., Hao, P., Chen, X., Boriboonsomsin, K., Wu, G., Barth, M.J., 2019b. Vehicle energy/emissions estimation based on vehicle trajectory reconstruction using sparse mobile sensor data. IEEE Trans. Intell. Transp. Syst. 20 (2), 716–726. https://doi.org/10.1109/TITS.2018.2826571.

Shehab, M.A., Pope, F.D., 2019. Effects of short-term exposure to particulate matter air pollution on cognitive performance. Sci. Rep. 9 (1), 1. https://doi.org/10.1038/s41598-019-44561-0.

Singer, B.C., Harley, R.A., 1996. A fuel-based motor vehicle emission inventory. J. Air Waste Manage. Association (1995) 46 (6) 581–593, doi: 10.1080/10473289.1996.10467492.

Smit, R., Ntziachristos, L., Boulter, P., 2010. Validation of road vehicle and traffic emission models - a review and meta-analysis. Atmos. Environ. 44 (25), 2943–2953. https://doi.org/10.1016/j.atmosenv.2010.05.022.

TransitFeeds, 2023. OpenMobilityData—Public transit feeds from around the world. Available from: <https://transitfeeds.com/>.

TransitLand, 2023. Transitland. Available from: <https://www.transit.land/>.

Waraich, A.S., Anowar, S., Tenaglia, T., Sider, T., Alam, A., Minaei, N.S., Hatzopoulou, M., Eluru, N., 2020. Disaggregate level simulation of bus transit emissions in a large urban region. Int. J. Sustain. Transp. 14 (7), 544–553. https://doi.org/10.1080/15568318.2019.1579009.

Wessel, N., Allen, J., Farber, S., 2017. Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. J. Transp. Geogr. 62, 92–97. https://doi.org/10.1016/j.jtrangeo.2017.04.012.

Xu, Y., Gbologah, F.E., Lee, D.-Y., Liu, H., Rodgers, M.O., Guensler, R.L., 2015. Assessment of alternative fuel and powertrain transit bus options using real-world operations data: Life-cycle fuel and emissions modeling. Appl. Energy 154, 143–159. https://doi.org/10.1016/j.apenergy.2015.04.112.

Yang, F., Kaul, D., Wong, K.C., Westerdahl, D., Sun, L., Ho, K., Tian, L., Brimblecombe, P., Ning, Z., 2015. Heterogeneity of passenger exposure to air pollutants in public transport microenvironments. Atmos. Environ. 109, 42–51. https://doi.org/10.1016/j.atmosenv.2015.03.009.

Yeh, S., Gil, J., Kyle, P., Kishimoto, P., Cazzola, P., Craglia, M., Edelenbosch, O., Fragkos, P., Fulton, L., Liao, Y., Martinez, L., McCollum, D.L., Miller, J., Pereira, R.H. M., Teter, J., 2022. Improving future travel demand projections: a pathway with an open science interdisciplinary approach. Progress Energy 4 (4), 043002. https://doi.org/10.1088/2516-1083/ac86b5.

Yu, Q., Li, T., Li, H., 2016. Improving urban bus emission and fuel consumption modeling by incorporating passenger load factor for real world driving. Appl. Energy 161, 101–111. https://doi.org/10.1016/j.apenergy.2015.09.096.

Zhang, X., Chen, X., Zhang, X., 2018. The impact of exposure to air pollution on cognitive performance. Proc. Natl. Acad. Sci. 115 (37), 9193–9197. https://doi.org/10.1073/pnas.1809474115.

Zhang, S., Wu, Y., Liu, H., Huang, R., Yang, L., Li, Z., Fu, L., Hao, J., 2014. Real-world fuel consumption and CO2 emissions of urban public buses in Beijing. Appl. Energy 113, 1645–1655. https://doi.org/10.1016/j.apenergy.2013.09.017.

Vieira, J.P.B., Pereira, R.H.M., Andrade, P.R., 2022. ipeaGIT/gtfs2emis_paper: V.1.0.0 release. Zenodo. doi: <10.5281/zenodo.7309183>.

Zhao, E., Walker, P.D., Surawski, N.C., 2021. Emissions life cycle assessment of diesel, hybrid and electric buses. Proc. Inst. Mech. Eng., Part D: J. Automobile Eng., 09544070211034318. doi: 10.1177/09544070211034318.