



VEIN v0.2.2: an R package for bottom–up vehicular emissions inventories

Sergio Ibarra-Espinosa¹, Rita Ynoue¹, Shane O’Sullivan², Edzer Pebesma³, María de Fátima Andrade¹, and Mauricio Osses⁴

¹Department of Atmospheric Sciences, Universidade de São Paulo, Rua do Matão 1226, São Paulo, SP, Brazil

²Department of Pathology, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo 455, São Paulo, SP, Brazil

³Institute for Geoinformatics, Westfälische Wilhelms-Universität Münster, Heisenbergstrasse 2, 48149 Münster, Germany

⁴Department of Mechanical Engineering, Universidad Técnica Federico Santa María, Vicuña Mackenna 3939, Santiago, Chile

Correspondence: Sergio Ibarra-Espinosa (sergio.ibarra@usp.br)

Received: 15 August 2017 – Discussion started: 25 September 2017

Revised: 3 May 2018 – Accepted: 4 May 2018 – Published: 14 June 2018

Abstract. Emission inventories are the quantification of pollutants from different sources. They provide important information not only for climate and weather studies but also for urban planning and environmental health protection. We developed an open-source model (called Vehicular Emissions Inventory – VEIN v0.2.2) that provides high-resolution vehicular emissions inventories for different fields of studies. We focused on vehicular sources at street and hourly levels due to the current lack of information about these sources, mainly in developing countries.

The type of emissions covered by VEIN are exhaust (hot and cold) and evaporative considering the deterioration of the factors. VEIN also performs speciation and incorporates functions to generate and spatially allocate emissions databases. It allows users to load their own emission factors, but it also provides emission factors from the road transport model (Copert), the United States Environmental Protection Agency (EPA) and Brazilian databases. The VEIN model reads, distributes by age of use and extrapolates hourly traffic data, and it estimates emissions hourly and spatially. Based on our knowledge, VEIN is the first bottom–up vehicle emissions software that allows input to the WRF-Chem model. Therefore, the VEIN model provides an important, easy and fast way of elaborating or analyzing vehicular emissions inventories under different scenarios. The VEIN results can be used as an input for atmospheric models, health studies, air quality standardizations and decision making.

1 Introduction

Emissions inventory is a quantification of pollutants discharged into the atmosphere by different sources (Pulles and Heslinga, 2010). This quantification is vital for regulatory and scientific purposes because it allows us to monitor the state of the Earth’s atmosphere and climate. It also allows us to create air quality standards, which will protect ecosystems and human health. For instance, the Intergovernmental Panel on Climate Change (IPCC) includes a dedicated task force, separated from the other three working groups, only for the purpose of greenhouse gas emissions inventory issues (Paus-tian et al., 2006).

In this instance, there are several emissions inventories that use different input data and approaches for different scales. One of the most frequently used inventories is the Emission Database for Global Atmospheric Research (EDGAR; Olivier et al., 1996), which provides estimates for the total emissions worldwide. This inventory uses national statistics that do not provide detailed characterizations of high-resolution applications. These detailed characterizations are needed for urban studies. There are also continental emissions inventories, such as the European Monitoring and Evaluation Programme (EMEP), which compile emissions from the parties of the Convention on Long-range and Trans-boundary Air Pollution (CLRTAP) (EEA, 2013). Moreover, there is the Regional Emissions inventory in Asia (REAS), which covers China, Japan and other countries (Streets et al., 2003). Also, there are nationwide inventories such as the National Emissions Inventory (NEI) in the United States, which

is released every 3 years and is based on data from state, local and tribal agencies and compiled by the Environmental Protection Agency (U.S. EPA, 2018). However, there are many countries and cities that do not include estimates of emissions for environmental and climate planning.

Vehicular emissions are becoming increasingly important in urban centers (Andrade et al., 2017) and measurements have shown that compounds emitted from exhausts can be highly reactive in the atmosphere, contributing to critical episodes of photochemical smog (Nogueira et al., 2015). However, obtaining this type of emissions database can be complicated, since the sources are moving and the emissions process is complex with many variables. This can be a challenge, especially in developing countries, due to the lack of information about the vehicle type, technology, age, motor size, fuel, speeds, accelerations, street type, environmental temperature and humidity, among other aspects. Besides that, there are other aspects involving the emissions inventory. The most common aspects are the accuracy and complexity relating to the exact contribution of the different pollutant sources and the fact that, in most cases, emissions inventories are usually seen as a scapegoat when simulations do not match observation (Pulles and Heslinga, 2010).

Vehicular emissions inventories are classified according to top-down and bottom-up approaches. Top-down approaches are based on statistics of vehicle composition, representative speeds and country balances, while bottom-up approaches are based on traffic counts, vehicle composition and speed recording (Ntziachristos and Samaras, 2016). An example of a bottom-up emissions model is the SPARE-Truck (Spatial Regression and Output Optimization Truck) (Perugu et al., 2017), which produces truck activity data at the level of road sections between junctions or interchanges (hereafter referred to as “link”) and then estimates emissions using the U.S. EPA Motor Vehicle Emission Simulator (MOVES) (Koupal et al., 2003). The accuracy of the emissions inventory will reflect the representation of the pollutants in the atmosphere. It is not always related to the complexity of the model. For instance, a meta-analysis of several studies on vehicular emissions (Smit et al., 2010) concluded that there is no evidence that the more complex models perform better than the less complex ones and that emissions estimation techniques must be chosen according the available traffic data. An emissions inventory must be comprehensive, including all the important sources and aspects regarding the emissions.

These complexities were addressed by the Department of Atmospheric Sciences at the University of São Paulo (USP) when modeling the atmospheric chemistry over Brazilian cities using a top-down vehicular emissions inventory with an online atmospheric model (Andrade et al., 2015). The Metropolitan Area of São Paulo (MASP), is the most populated megacity in Latin America (IBGE, 2014) and its most important source of pollution comes from the 11 million vehicles that circulate within the region (CETESB, 2013; DE-

NATRAN, 2015; Andrade et al., 2017). Furthermore, half of all emissions of carbon monoxide (CO), volatile organic compounds (VOCs) and nitrogen oxides (NO_x) in the MASP are from vehicles that are more than 10 years old (Andrade et al., 2017). Despite the inventories developed by (Andrade et al., 2015) being useful, they still suffer from top-down limitations as they use surrogates to produce spatial and temporal distributions, hence limiting the representation of the emissions. Therefore, we decided to develop our own bottom-up vehicular emissions inventory model. With this model, we aim to generate scientific estimates and provide useful information to decision makers and urban/environmental planners.

The main goal of this project was to develop a high spatial- and temporal-resolution vehicular emissions inventory model, which was named the Vehicular Emissions INventory (VEIN) model. The VEIN model follows a bottom-up approach and includes methods from Ntziachristos and Samaras (2016) and other authors. It is designed for working with and extrapolating outputs of four-stage travel demands; however, the model is flexible and could be used with a top-down approach. It allows the classification of vehicles into several categories, different options of emission factors and specification of pollutants, and input traffic from traffic simulations or other network-based sources. The model includes functions to represent spatial objects and is capable of producing gridded emissions outputs. It is open source, user-friendly and available to run in any computational platform (Mac, Windows, Linux, etc.).

2 VEIN model: methodology to estimate vehicular emissions

Temporal and spatial disaggregated emissions are estimated following a general approach of multiplication between activities and emission factors (Pulles and Heslinga, 2010), as shown in Eq. (1).

$$\text{Emission}_{\text{pollutant}} = \sum_{\text{activity}} (\text{AR}_{\text{activity}} \cdot \text{EF}_{\text{pollutant, activity}}), \quad (1)$$

where $\text{Emission}_{\text{pollutant}}$ for any type of pollutant depends on the activity rate AR and the emission factors EF, which is the mass of pollutants generated according to the level of activity. In the context of vehicular emissions, $\text{AR}_{\text{activity}}$ represents the number of vehicles times the distance (km) that they travel. $\text{EF}_{\text{pollutant, activity}}$ is the emission factor (g km^{-1}) for pollutants of the vehicles.

For a bottom-up estimation of vehicular emissions, a large number of parameters influence activity (traffic flow, vehicle composition, speed recording, length of road) and emission factors (speed or acceleration dependent) (Ntziachristos and Samaras, 2016). In this instance, the following sections provide the theory behind the VEIN model regarding traffic data

arrangement, selection of emission factors, emissions estimation, spatial allocation and inputs for atmospheric models.

2.1 Traffic data

Traffic data required for the VEIN model must be represented as an hourly amount of vehicles per street.

This traffic data can be provided by traffic simulations, interpolations or by other sources. In the first step, VEIN reads spatial morning rush hour traffic data from each street of a desired area or city. After reading them, VEIN arranges and organizes the data by vehicle composition, according to Eq. (2).

$$F_{i,j,k}^* = Q_i \cdot VC_{i,j} \cdot Age_{j,k}, \quad (2)$$

where $F_{i,j,k}^*$ is the vehicular flow at street link i for vehicle type j by age of use k . j defines the vehicular composition according to their type of use, type of fuel, size of engine and gross weight, based on the definitions of (Corvalán et al., 2002). Q_i is the traffic flow at street link i . $VC_{i,j}$ is the fraction of vehicles varying according to the type of vehicles j in the composition for street link i . $Age_{j,k}$ is the age distribution by vehicular composition j and age of use k . This equation shows that VC splits the total vehicular flow Q to identify the vehicular fraction, which varies according to the type of fuel, size of motor and gross weight. For example, if Q is light-duty vehicles (LDVs) and we know that 5% of the Q are passenger cars (PCs), with an engine of less than 1400 cc, VC is 0.05. This characterization of the fleet depends on the amount and quality of the available information. VEIN then multiplies the traffic with age to obtain the amount of each type of vehicle by age of use.

Traffic data must be temporally extrapolated because they are usually available for the morning rush hour. Traffic data can be estimated from short period traffic count datasets, then expanded to represent longer timespan, such as annual average daily traffic (AADT; Wang and Kockelman, 2009; Lam and Xu, 2000). The next step is to extrapolate the vehicular flow at street link i , vehicle type j and age of use k , to obtain the vehicular flow for the hour of the week l ($F_{i,j,k,l}$; see Eq. 3).

$$F_{i,j,k,l} = F_{i,j,k}^* \cdot \mathbf{TF}_{j,l}, \quad (3)$$

where $F_{i,j,k,l}$ is the traffic flow for each link i and type of vehicle j of the vehicular composition, with the age of use k for the hour l . $\mathbf{TF}_{j,l}$ are the temporal factors varying according to each hour of l and type of vehicle j . \mathbf{TF} is defined as a matrix with 24 lines, and the number of columns represents each day from Monday to Sunday. For instance, if the user has the output of a travel-demand model for the morning rush hour and also hourly traffic counts, the information of the hourly traffic count must be arranged to create a matrix with 24 lines and with the number of columns to each day. Then, in order to output the travel-demand model for other hours, \mathbf{TF} matrices must be normalized to the hour that

represents the traffic data. This means that \mathbf{TF} values for the morning peak hour must be 1 and the respective proportion must be assigned to the other hours, which is obtained by simply dividing the traffic of each hour of the matrix by the traffic of the morning rush hour. \mathbf{TF} values can be obtained from automatic traffic count stations or other hourly traffic data to obtain the temporal profile.

The average speed of traffic flow is very important, and it must be determined for each link and hour. Once the vehicular flow is identified for each hour, the average speed is then identified for each hour. This was accomplished by employing curves from the Bureau of Public Roads (BPR; Bureau of Public Roads, 1964), as shown in Eq. (4). The process involves calculating speed by dividing the length of road by the time. The time is calculated using the total traffic expanded to each street link i and hour l .

$$T_{i,l} = To_i \cdot \left(1 + \alpha \cdot \left(\frac{Q_{i,l}}{C_i} \right)^\beta \right) \quad (4)$$

In Eq. (4), T_i is the travel time per street link i at each hour of the week l . To_i is the travel time under free-flow conditions where maximum speed was used. $Q_{i,l}$ is the traffic flow at peak hour for each street link i and hour of the week l . C_i is the capacity of vehicles on street link i . The parameters α and β are adjustments with default values of 0.15 and 4, respectively (Bureau of Public Roads, 1964). It has been described that speed decreases as the flow increases until it reaches the capacity of the link, and as the traffic continues to increase, the speed decreases and travel time increases (Suh et al., 1990). Capacity is the maximum number of vehicles during a specific period under prevailing roadway, traffic and control conditions (Manual, 2000). The values of BPR parameters are obtained by regressions of speed and volume traffic recordings and the default values. The default BPR parameters α 0.15 and β 4 are representative of traffic and circulation characteristics of the United States. As a consequence, other authors have investigated the determination of capacity based on local data. For instance, Suh et al. (1990) used Korean data to reestimate α and β by level of service (LOS) A, B, C, D and E, from A (less congested) to E (more congested), finding values of $\alpha = 2.72$ and $\beta = 6$ for LOS D. In addition, Manzo et al. (2014) estimated the distribution of α and β with Danish data, and Kucharski and Drabicki (2017) found values of $\alpha = 0.52$ and $\beta = 3.47$. Consequently, the user can use different values in the VEIN model instead of suggested values from the Bureau of Public Roads (1964). This information is important because it allows the user to calculate the time to travel during link i at any hour l and, therefore, to obtain the speed on that length of the road.

However, the information of the capacity is not always available. When information exists at least for the peak and free-flow speeds, it is possible to apply a simple average of both speeds to obtain the average speeds and then distribute these speeds at different hours.

2.2 Selection of the emission factors

The emission factors describe the relationship intensity of activity and emissions for a given technology (Pulles and Heslinga, 2010). In the case of our model, an emission factor is the mass of pollutant emitted by the vehicular type, technology and years of use by traveled distance, as mass of pollutant/distance. VEIN takes into account emission factors for hot and cold exhaust, evaporative, deterioration and wear emissions. VEIN allows three types of hot exhaust emission factors:

1. Speed functions from the COmputer Programme to calculate Emissions from Road Transport (Copert; Ntziachristos and Samaras, 2016), which are stored internally in the model. This approach can be used if there is no local emission factors and if there is information about vehicular speed recordings, simulations or knowledge of the representative speeds.
2. Emission factors from local sources. The values must be mass per kilometer (g km^{-1}) per specific type of vehicle, including fuel type, size and weight by age of use. For instance, Brazilian emission factors in the official emissions inventory report are averaged emissions of annual emission certification tests by type of vehicle, fuel type, size and weight (CETESB, 2013), and they are not speed functions. Nevertheless, this capability can be utilized with different types of emission factors such as tunnel studies or dynamometer measurements which are not speed functions. Appendix B shows an example of VEIN being used with different emission factors.
3. Scaled local emission factors with Copert in order to incorporate speed variation for local factors, as shown in Eq. (5). This produces a specific speed dependence on emission factor by age of use for the vehicle.

Copert emission factors are based on emission measurements made in Europe. This means that there is an inherent fuel composition associated with these factors. When the user intends to use these factors, they must be corrected with the local fuel properties as shown by Ntziachristos and Samaras (2016). However, if the user uses emission factors based on local measurements, this correction is not necessary.

$$EF_{\text{scaled}}(V_{i,l})_{j,k,m} = EF(V_{i,l})_{j,k,m} \cdot \frac{EF_{\text{local}}_{j,k,m}}{EF(V_{dc_{i,l}})_{j,k,m}}, \quad (5)$$

where $EF_{\text{scaled}}(V_{i,l})_{j,k,m}$ is the scaled emission factor and $EF(V_{i,l})_{j,k,m}$ is the Copert emission factor for each street link i , vehicle from composition k , hour l and pollutant m . $EF_{\text{local}}_{j,k,m}$ represents the constant emission factor (not speed functions). $EF(V_{dc_{i,l}})_{j,k,m}$ are Copert emission factors with average speed value of the respective driving cycle for the vehicular category j . The São Paulo emission factors data include recordings of the Federal Test Procedure

(FTP-75) driving cycle for LDV with an average speed of 34.12 km h^{-1} , as shown in the report of driving cycles (Barlow et al., 2009).

By default, VEIN includes a deterioration factor from Copert (Ntziachristos and Samaras, 2016). This deterioration factor depends on accumulated mileage and technology associated with Euro standards on vehicles with three-way catalysts. However, it is possible to include other sources, such as from Corvalán and Vargas (2003).

2.3 Emissions estimation

VEIN estimates type of emissions including hot exhaust (EH; Eq. 6), cold-start exhaust (EC; Eq. 7), evaporative (EV; Eq. 8), deterioration factors and speciation. The total vehicular emission is the sum of all types of emissions.

2.3.1 Hot exhaust emission

The VEIN process of emissions estimation is performed per street link, vehicle type, hour of week and pollutant. Equation (6) shows the hot exhaust estimation:

$$EH_{i,j,k,l,m} = F_{i,j,k,l} \cdot L_i \cdot EF(V_{i,l})_{j,k,m} \cdot DF_{j,k}. \quad (6)$$

In Eq. (6), $EH_{i,j,k,l,m}$ is the emissions for each street link i , vehicle category from composition k , hour l and pollutant m , where $F_{i,j,k,l}$ is the vehicular flow calculated in Eq. (1). L_i is the length of the street link i . $EF(V_{i,l})_{j,k,m}$ is the emission factor of each pollutant m . $DF_{j,k}$ is the deterioration factor for vehicle of type j and age of use k . Deterioration factors are emission degradation functions of gasoline vehicles using catalysts due to accumulated mileage based on European measurements (Ntziachristos and Samaras, 2016). Nevertheless, the users could use their own set of deterioration factors such as that proposed by Corvalán et al. (2002).

2.3.2 Cold-start emissions

Cold-start emissions are produced during engine startup, when the engine and/or catalytic converter system has not reached its normal operational temperature. Consequently, these emissions occur when a vehicle starts its journey at the parking location or mostly in residential streets. Several studies have shown the significant impact of these types of emissions (Chen et al., 2011; Weilenmann et al., 2009). VEIN also considers cold-start emissions – under this condition, emissions will be higher, and if the atmospheric temperature decreases, cold-start emissions will increase regardless of whether the catalyst has reached its optimum temperature for functioning (Boulter, 1997). For example, Ludykar et al. (1999) measured vehicular emissions with the European driving cycle (EDC) at 22, -7 and -20 °C, finding that at lower temperatures, CO increased 5 times, NO_x 1 time, hydrocarbons (HC) 14 times and particulate matter (PM) 46 times, but specific hydrocarbons increased more, such

as Toluene, which increased 21 times, and PAH, which increased 33 times.

The VEIN model caters to these emissions by using the approach outlined in Copert (Ntziachristos and Samaras, 2016), as shown in Eq. (7).

$$EC_{i,j,k,l,m} = \beta_j \cdot F_{i,j,k,l} \cdot L_i \cdot EF(V_{i,l})_{j,k,m} \cdot DF_{j,k} \cdot (EF_{\text{cold}}(ta_n, V_{i,l})_{j,k,m} - 1) \quad (7)$$

This approach adds two terms to Eq. (6). The first term $EF_{\text{cold}}(ta_n, V_{i,l})_{j,k,m} - 1$ is the emission factors for cold-start conditions at each street link i , vehicle category from composition k , hour l , pollutant m and monthly average temperature n . Ntziachristos and Samaras (2016) suggest using monthly average temperature. As the information about the parking locations is limited, in Eq. (7) we propose assigning the emissions by links. In cases when better information is available, these emissions should be assigned into parking locations.

The second term β_j is defined as the fraction of mileage driven with a cold engine/catalyst (Ntziachristos and Samaras, 2016). The VEIN model incorporates a dataset of cold starts recorded during the implementation of the International Vehicle Emissions (IVE) model (Davis et al., 2005) in São Paulo (Lents et al., 2004), which provides the hourly mileage driven with cold-start conditions. The data of start patterns should be generated by each user with local data. For instance González et al. (2017) generated a start pattern using surveys to estimate vehicular emissions with IVE. However, the user eventually could use the start pattern available in VEIN only in the absence of other data. Alternatively, the user could follow the Ntziachristos and Samaras (2016) for estimating β , which consists in $\beta = 0.6474 - 0.02545 * \text{ltrip} - (0.00974 - 0.000385 * \text{ltrip}) * ta$, where ltrip is the average length of the trip and ta is the average monthly temperature.

2.3.3 Evaporative emissions

Evaporative emissions are important sources of hydrocarbons, and these emissions are produced by vaporization of fuel due to variations in ambient temperatures (Mellios and Ntziachristos, 2016; Andrade et al., 2017). There are mainly three types of evaporative emissions: diurnal emissions, due to increases in atmospheric temperature, which lead to thermal expansion of vapor fuel inside the tank; running losses, when the fuel evaporates inside the tank due to normal operation of the vehicle; and hot-soak emissions, which occur when the hot engine is turned off. These methods implemented in VEIN were sourced from the evaporative emissions methods of Copert (Mellios and Ntziachristos, 2016). This approach is shown in Eqs. (8), (9) and (10).

$$EV_{j,k} = \sum_s D_s \cdot \sum_j F_j \cdot (HS_{j,k} + de_{j,k} + RL_{j,k}), \quad (8)$$

where EV_j is the volatile organic compound (VOC) evaporative emissions due to each type of vehicle j . D_s is the

“seasonal days” or number of days when the mean monthly temperature is within a determined range: $[-5^\circ, 10^\circ\text{C}]$, $[0^\circ, 15^\circ\text{C}]$, $[10^\circ, 25^\circ\text{C}]$ and $[20^\circ, 35^\circ\text{C}]$. $F_{j,k}$ is the number of vehicles with to the same type j and age of use k . $HS_{j,k}$, $de_{j,k}$ and $RL_{j,k}$ are average hot/warm-soak, diurnal and running loss evaporative emissions (g day^{-1}), respectively, according to the vehicle type j and age of use k . $HS_{j,k}$ and $RL_{j,k}$ are obtained using equations also sourced from

$$HS_{j,k} = x_{j,k} \cdot (c \cdot (p \cdot e_{\text{shc}} + (1-p) \cdot e_{\text{swc}}) + (1-c) \cdot e_{\text{shfi}}), \quad (9)$$

where x is the number of trips per day for the vehicular type j and age of use k . c is the fraction of vehicles with fuel return systems. p is the fraction of trips finished with a hot engine, for example, an engine that has reached its normal operating temperature and the catalyst has reached its light-off temperature (Ntziachristos and Samaras, 2016). The light-off temperature is the temperature at the point when catalytic reactions occur inside a catalytic converter. e_{shc} and e_{swc} are average hot-soak and warm-soak emission factors for gasoline vehicles with carburettor or fuel return systems (g parking^{-1}). e_{shfi} is the average hot-soak emission factors for gasoline vehicles equipped with fuel injection and non-return fuel systems (g parking^{-1}).

$$RL_{j,k} = x_{j,k} \cdot (c \cdot (p \cdot e_{\text{rhc}} + (1-p) \cdot e_{\text{rwc}}) + (1-c) \cdot e_{\text{rffi}}) \quad (10)$$

x and p have the same meanings as in Eq. (9). e_{rhc} and e_{rwc} are average hot and warm running loss emission factors for gasoline vehicles with carburettor or fuel return systems (g trip^{-1}), and e_{rffi} is the average hot running loss emission factors for gasoline vehicles equipped with fuel injection and non-return fuel systems (g trip^{-1}). It is recommended to estimate the number of trips per day (Mellios and Ntziachristos, 2016), x , as the division between the mileage and 365 times the length of trip: $x = \frac{\text{mileage}_j}{(365 \cdot \text{ltrip})}$. However, the mileage of a vehicle is not constant over the years. Therefore, VEIN incorporates a dataset of equations to estimate mileage of different types of vehicles by age of use (Bruni and Bales, 2013). The methods presented in this part correspond to the evaporative emission factors Tier 2 (Mellios and Ntziachristos, 2016). However, we expect to add Tier 3 methods in future versions of VEIN, including hourly distribution parking patterns. Alternatively, for convenience, the user could transform the units to produce emission factors in grams per kilometer for convenience. For instance, in Eqs. (9) and (10), x is the number of trips per day, knowing that the mean distance per trip would allow x to be calculated as the distance traveled per day.

2.4 Speciation of emissions in chemical subcomponents

Particulate matter and hydrocarbons are a mixture of several chemical compounds that play an important role in atmospheric chemistry (Seinfeld and Pandis, 2016). VEIN includes speciation profiles for hydrocarbons and particulate

matter from Ntziachristos and Samaras (2016) and Ibarra (2017). These profiles are percentages of the emissions by vehicle type, fuel, emission standard and other characteristics. Speciations are also included for particulate matter in black carbon and organic matter; particulate matter fractions for tire, brake and road wear; non-methanic hydrocarbons (NMHC); and nitrogen oxides. In this case, the NMHC speciations from Ibarra (2017) cover emissions from vehicles consuming gasohol (gasoline with 25 % of ethanol), ethanol and diesel from exhaust, evaporative and liquid fuel released to the atmosphere. This speciation is based on measurements made in Brazil during the studies of Rafee (2015), Ynoue (2004), Albuquerque (2005), Oliveira (2007), Miranda (2001), Vara-Vela et al. (2016) and Andrade et al. (2015). Hence, it is advisable that users from other nations/locations use suitable data for their local conditions.

2.5 Spatial allocation and databases

VEIN provides functions to generate grids and spatially allocate emissions into grids. This is helpful for the visualization and generation of inputs for atmospheric models and as a tool for urban planning. In addition, VEIN includes functions to produce a database of hourly emissions for vehicular composition by age of use. Section 4.4 provides details and examples about the emissions grids and databases.

3 VEIN model design

The VEIN model was constructed using the free open-source R software (R Core Team, 2017). R is a programming language and environment for statistical computing and graphics (R Core Team, 2017). It was developed primarily for analyzing data. However, since its capabilities have grown over time, R has become a flexible language with many different areas of application. It includes elements of programming languages such as Lisp and the syntax of S, as described by Leiner et al. (1997).

The VEIN R package depends on the package “sp” (Bivand et al., 2013), as it uses several of its classes. VEIN imports some functions from the package “rgeos” (Bivand and Rundel, 2016), which is an interface for the Geometry open-source (GEOS) library (<https://trac.osgeo.org/geos/>, last access: 31 May 2018). It also imports functions from “rgdal” (Bivand et al., 2016), which provides bindings to the Geospatial Data Abstraction Library (GDAL; <http://www.gdal.org/>, last access: 31 May 2018). Therefore, these R packages must be installed prior to using the VEIN package.

VEIN started between 2014 and 2016 as a collection of several R scripts, initially named R-EMISSIONS (REMI; Ibarra-Espinosa and Ynoue, 2017), which later evolved into an R package. It was developed in R due to the free open-source advantages and because R allows easier reproducibility. VEIN is open to scrutiny from its community of users,

thus allowing opportunities for user feedback and improvements. This facilitates widespread use of the model and identifying any software bugs/errors, with the potential for adding new capabilities. VEIN has its own functions, but it also incorporates other data and functions such as emission factors and mileage.

VEIN can be installed from the Comprehensive R Archive Network (CRAN; <https://CRAN.R-project.org/package=vein>, last access: 31 May 2018) or from the website Github (<https://github.com/atmoschem/vein>). In order to use the VEIN library and run the demo, it is necessary to run the following scripts in R:

```
install.packages("vein") #or
library(devtools)
install_github("atmoschem/vein")
library(vein)
demo(VEIN)
```

The diagram process for estimating emissions is shown in Fig. 1. The circles in this figure refer to the data and the boxes refer to the functions inside the model. The VEIN model diagram starts at the circle for traffic, which represents the morning rush hour traffic data for each street link. Then the “age” functions (“age_ldv”, “age_hdv” or “my_ldv”) determine the vehicular composition by age of use as shown in Eq. (2). The data “profile” allows us to temporally extrapolate traffic data to the other hours, and this allows us to estimate the average vehicular speed at any hour and link using the function “netspeed”. Emission factor selections start by adding the deterioration effect with the function “emis_det” to local, speed-dependent emission factors from Ntziachristos and Samaras (2016), denoted as “speed_ef”, or scaled emission factors, denoted as *scaled_ef* in Fig. 1. Besides including speed-dependent emission factors from Ntziachristos and Samaras (2016), VEIN also includes local emission factors from CETESB (2016). Once the input data are ready, the function “emis” estimates hourly emissions for each hour of the day and day of the week. The function “emis_post” produces an emissions database by vehicle category or by street, denoted as “df” and “street” in Fig. 1, respectively. These emissions are then speciated with the function “speciate”. At this time, the user can create a grid with the function “make_grid”, which creates a rectangular spatial grid with format `SpatialPolygonsDataFrame` used to allocate emissions spatially with the function “emis_grid”. The function “emis_wrf” reads the emissions grids and creates a data frame ready to be used with the NCL script AS4WRF (Vara-Vela et al., 2016), which creates an input for the WRF-Chem model (Grell et al., 2005). The user can also use the model export (Ibarra-Espinosa et al., 2018) to create inputs for the model WRF-Chem or BRAMS (Freitas et al., 2005) using only R (R Core Team, 2017).

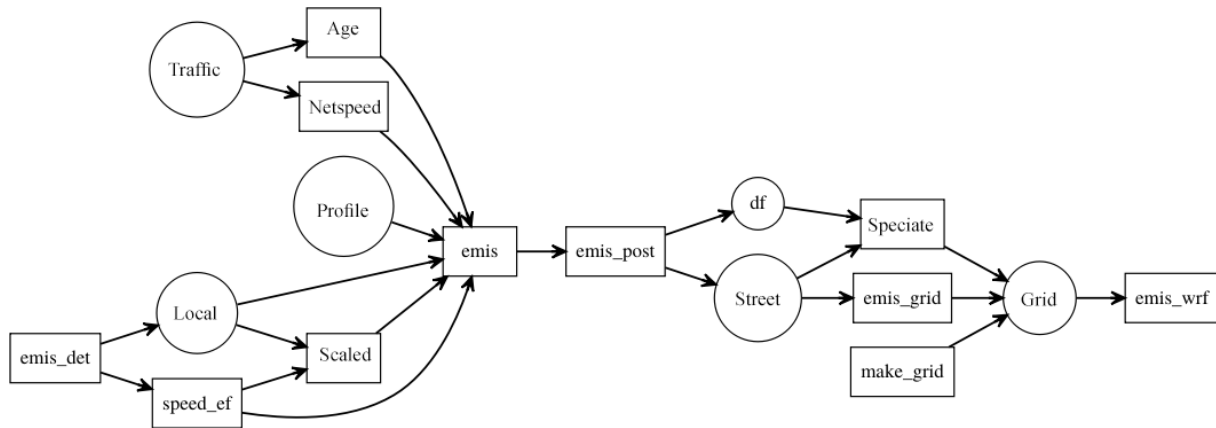


Figure 1. Representation of the VEIN model. Boxes and circles represent functions and data, respectively.

3.1 Functions and classes

VEIN uses objects of the class “Spatial” (Pebesma and Bivand, 2005), including `SpatialLinesDataFrame`. To read geospatial data, there are several packages, such as `rgdal` (Bivand et al., 2016) or “`maptools`” (Bivand and Lewin-Koh, 2015). The main requirement is that the network must be a `SpatialLinesDataFrame`, a class of `sp` (Pebesma and Bivand, 2005).

We included several functions to arrange traffic data, select or scale emission factors, as well as estimate and process emissions in VEIN, as shown in Table 1. These functions implement the equations shown in Sect. 2.

VEIN incorporates eight classes (see Table 1), which are objects with specific characteristics: methods and units. The methods are `print`, `summary` and `plot`. They are functions that return a specific result depending on each class. Another important characteristic of each class is that they include explicit units, in an effort to reduce human error and improve usability. For this task, VEIN imports some functions of the package “`units`” (Pebesma et al., 2016), which is an interface in the C library “`udunits`” from University Corporation for Atmospheric Research (UCAR). Therefore, this library must be installed on the system prior to using VEIN. Only the `EmissionFactorsList` and `EmissionsArray` do not show their units explicitly due to limitations with the `units` package. The classes outlined in Table 1 are also constructor functions, which means that they can create VEIN classes and add the respective units. VEIN incorporate constructor functions to create classes such as `Vehicles` or `Emissions`. These functions are incorporated inside other VEIN functions in order for the output of VEIN to have a class. When the constructor functions are applied to a numeric element, the constructor simply adds the units and the resulting object has class units. For example, applying the function `EmissionsArray` to a numeric vector will add the units grams per hour to the numeric vector.

4 Estimating MASP vehicular emissions using VEIN model

This section presents the application of the most important functions of the VEIN model. These functions obtain an estimate of CO emissions from LDV fleets in MASP for 2015 (for a typical non-holiday week).

4.1 Traffic data for MASP

Hourly traffic is a requirement for these data. These data can only be represented as 1 h of data, which can then be extrapolated with the VEIN functions or as a list of hourly traffic data for all the covered hours. The present application includes a four-stage travel-demand model for MASP CET (2014) for a morning rush hour loaded into R as a `SpatialLinesDataFrame`. It includes peak and free-flow speeds, along with capacity (maximum amount of vehicles that can circulate in a road per hour) and traffic flow from LDV and HDVs (heavy-duty vehicles). Figure 2 shows the traffic simulation of LDV at 08:00–09:00 local time (LT), where urban motorways concentrate the highest amount of vehicles. The total volume of LDV is 24 708 767 veh h^{-1} and the number of streets are 34 733, with a mean of 711 $\text{veh h}^{-1} \text{street}^{-1}$. It is important to note that the VEIN model provides an extraction of the traffic simulation for the western part of São Paulo. The traffic simulation for MASP has a size of 61.6 Mb and the extraction for the western part of São Paulo 3.1 Mb. We provided the extraction and not the whole traffic simulation in VEIN to make it faster. This section provides codes to run VEIN so that the reader can follow them with the data provided in the model.

After loading traffic data, the traffic flow was expanded to each hour of the week with the function “`temp_fact`”, as shown in the following scripts. It is also necessary to extrapolate hourly vehicle speeds. Therefore, we created the function `netspeed`, which applies the function BPR (Bureau of Public Roads, 1964) curves, according to Eq. (4). To use

Table 1. Summary of the VEIN classes, functions and internal data.

Function	Description	Reference
age_hdv	Distribution of HDV by age of use	Ministério do Meio Ambiente (2011)
age_ldv	Distribution of LDV by age of use	Ministério do Meio Ambiente (2011)
age_moto	Distribution of motorcycle by age of use	Ministério do Meio Ambiente (2011)
ef_evap	Evaporative emission factors	Mellios and Ntziachristos (2016)
ef_hdv_scaled	List of scaled emission factors for HDV	Ntziachristos and Samaras (2016)
ef_hdv_speed	HDV emission factors	Ntziachristos and Samaras (2016)
ef_ldv_cold	LDV cold-start emission factors	Ntziachristos and Samaras (2016)
ef_ldv_cold_list	List of LDV cold-start emission factors	Ntziachristos and Samaras (2016)
ef_ldv_scaled	List of scaled emission factors for LDV	Ntziachristos and Samaras (2016)
ef_ldv_speed	LDV emission factors	Ntziachristos and Samaras (2016)
ef_wear	Tire and brake wear, road abrasion	Ntziachristos and Boulter (2009)
EmissionFactors	Creates class <code>EmissionFactors</code> (g km^{-1})	
EmissionFactorsList	Creates class <code>EmissionFactorsList</code> (g km^{-1})	
Emissions	Creates class <code>Emissions</code> (g h^{-1})	
EmissionsArray	Creates class <code>EmissionsArray</code> (g h^{-1})	
EmissionsList	Creates class <code>EmissionsList</code> (g h^{-1})	
emis	Estimation to hour and day of the week	
emis_cold	Cold-start estimation	Ntziachristos and Samaras (2016)
emis_det	Deterioration factors	Ntziachristos and Samaras (2016)
emis_evap	Evaporative estimation	Mellios and Ntziachristos (2016)
emis_grid	Allocation on rectangular grid	
emis_paved	Resuspension of paved roads	U.S. EPA (2016)
emis_post	Post-processing of emissions	
emis_wear	Estimation of wear emissions	Ntziachristos and Boulter (2009)
emis_wrf	Creating data frame to NCL AS4WRF	Vara-Vela et al. (2016)
Evaporative	Creates class <code>Evaporative</code> ($\text{g}\cdot\text{d}^{-1}$)	
fe2015	Data of CETESB emission factors	CETESB (2016)
fkm	Data of mileage functions by vehicle	Bruni and Bales (2013)
hot_soak	Hot-soak evaporative	Mellios and Ntziachristos (2016)
make_grid	Rectangular grid	
my_age	Distribution of vehicles by age of use	
net	Data of traffic simulation of west São Paulo	CET (2014)
netspeed	Estimate average speed	
pc_profile	Data of temporal factors	ARTESP (2012)
pc_cold	Data of vehicle start pattern	Lents et al. (2004)
running_losses	Evaporative estimation	Mellios and Ntziachristos (2016)
speciate	Split by species	Ntziachristos and Samaras (2016), Ibarra (2017)
Speed	Creates class <code>Speed</code> (km h^{-1})	
temp_fact	Expand hourly traffic	
Vehicles	Creates class <code>Vehicles</code> (1 h^{-1})	
vkm	Determination of vehicle kilometers	

BPR, a data frame is required with total traffic at all hours and the morning rush parameters capacity, peak speed, free-flow speed and length of the road and with the BPR parameters alpha and beta. The argument “scheme” produces a 24 h speed data frame, based only on peak and free-flow speed with a profile of free-flow speeds in the early mornings, peak speeds, and morning and evening rush hours and the average at hours in between. If the time lapse for the emissions estimation is longer than a week, the user could simply replicate the hours until it reaches the desired hours.

```
data(net)
data(pc_profile)
pcw <- temp_fact(net$ldv+net$hdv,
pc_profile)
speed <- netspeed(pcw, net$ps, net$ffs,
net$capacity, net$lkm, alpha = 1)
```

To illustrate this, the resulting speeds are observed in Fig. 3, which shows two different speed maps: one for 08:00 LT (panel a) and the other for 23:00 LT (panel b). This

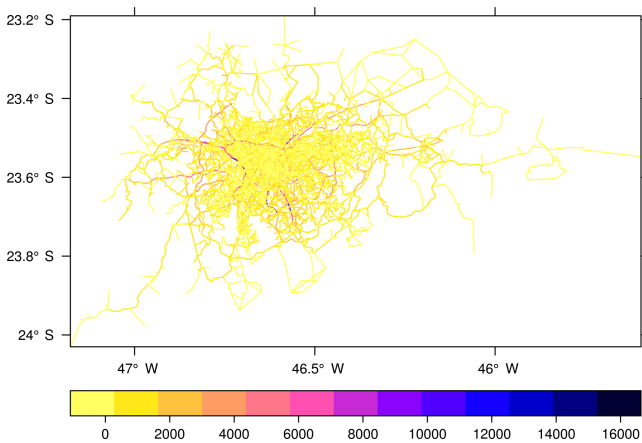


Figure 2. Traffic flow simulation for LDV (veh h^{-1}) at 08:00–09:00 LT for MASP.

figure shows that the highest speeds are found in most of the streets further away from the MASP center at both times (08:00 and 23:00 LT). The major difference between the two panels (a and b), is that late at night, the flow is faster near the center of MASP. This seems reasonable since the vehicular flow tends to diminish during the night. The average speeds also show a pattern related to the type of street as shown in Fig. 4. The type of street comes from São Paulo traffic simulation; the names of the different types were translated into English and are as follows. Motorway: roads with speed limits above 80 km h^{-1} without physical intersections. Arterial: roads with a speed limit of 60 km h^{-1} with intersections such as traffic lights. Collector: roads with a speed limit of 40 km h^{-1} that collect and distribute traffic between arterial roads. Local: roads with a speed limit of 30 km h^{-1} that access restricted zones. Figure 4 shows that lower speeds are found during the morning (07:00–10:00 LT) and evening (17:00–20:00 LT) rush hours. This is important in terms of air pollution because at lower speeds vehicles emit more pollutants (Ntziachristos and Samaras, 2016). By contrast, maximum average speeds for each type of road are obtained during night hours and on Sundays at all hours.

After calculating the São Paulo traffic flow average speeds for each hour of the week and each street link, the age distribution of the fleet was obtained by type of vehicle. The “age*” functions (age_ldv, age_hdv and “age_moto”) distribute the traffic data by the vehicle’s age of use. These functions return a data frame with the number of rows matching the number of streets, columns representing the amount of vehicles by age of use and a message indicating the average age of the fleet. The age* functions are related to Eq. (2), where they split the vehicular flow at street link Q_i by type of vehicle j and the vehicle’s age of use k . These functions are based on the survival equations presented in the Brazilian Emissions Inventory Report by the Ministério do Meio Ambiente (2011) and are parameterized for the VEIN model.

They allow the use of different coefficients to obtain different age distributions allowing the representation of different realities. Furthermore, the function “my_age” distributes the traffic from an existing dataset, e.g., yearly vehicle licensing.

The following code shows three uses of age* functions. The first, my_age, uses yearly traffic data from the report of the São Paulo emissions inventory (CETESB, 2016), expressed as CETESB_PC based on vehicle sales. The second, age_ldv, uses default parameters, and the third, age_ldv, uses $b = -0.14$.

```
CETESB_PC <- c(33491,22340,24818,31808,
46458,28574, 24856,28972,37818,49050,
87923,133833, 138441,142682,171029,151048,
115228,98664, 126444,101027,84771,55864,
36306,21079, 20138,17439,7854,2215,656,
1262,476,512, 1181,4991,3711, 5653, 7039,
5839, 4257, 3824,3068)

pc1 <- my_age(x = net$ldv,
y = CETESB_PC, name = "PC")
pc2 <- age_ldv(x = net$ldv,
name = "PC", agemax = 41)
pc3 <- age_ldv(x = net$ldv,
name = "PC", b = - 0.14, agemax = 41)
```

Figure 5a shows three age distributions, each one has $24\,708\,767 \text{ veh h}^{-1}$, and each has a different average age. Each curve represents the São Paulo LDV fleet according to its age of use, with an estimated average age of 11.09 years (red line), 15.53 years (blue line) and 15.17 years (green line). This figure shows that in MASP there are more newer vehicles than older ones. age* functions also include a logical option named “bystreet”, with a default value equal to FALSE. When this value is TRUE, age* expects that the coefficients a and b for the age* functions are numeric vectors, with length matching the number of streets. This allows different age distributions within the same road network, and this is particularly useful for areas with less information about the vehicles’ age of use.

4.2 Emission factors

Once we obtain the traffic flow for the desired type of vehicles (in our LDV example), for each hour of the day, for all (desired) days of the week, for each age distribution and for each street link, then we can proceed to the emissions calculation itself.

The VEIN package includes a database entitled “fe2015” with emission factors for PC and light trucks by age of use from the São Paulo official vehicular emissions inventory (CETESB, 2016). This inventory was compiled using a top–down approach, and the pollutants estimated were CH_4 , CO , CO_2 , HC , N_2O , NMHC , NO_x and PM . These data in-

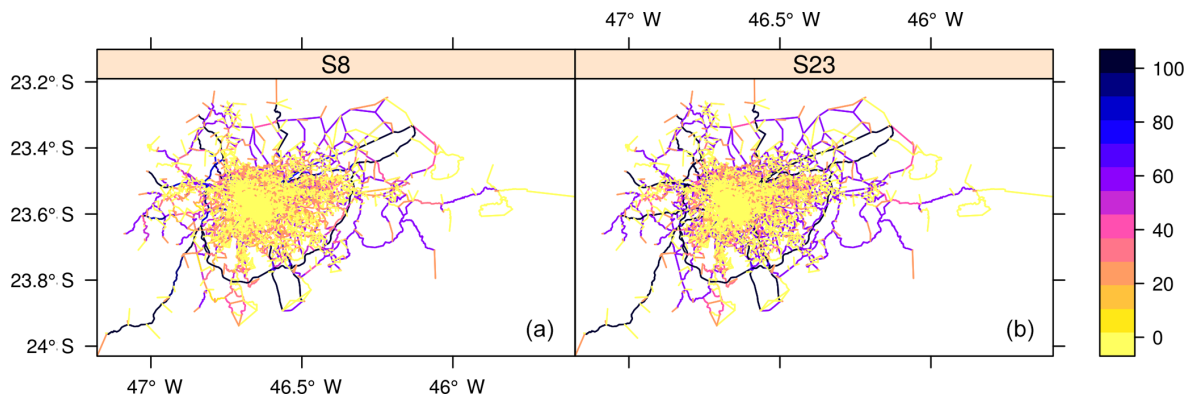


Figure 3. Traffic speeds (colored lines; km h^{-1}) for LDV fleet at 08:00 LT (a) and 23:00 LT (b) in MASP.

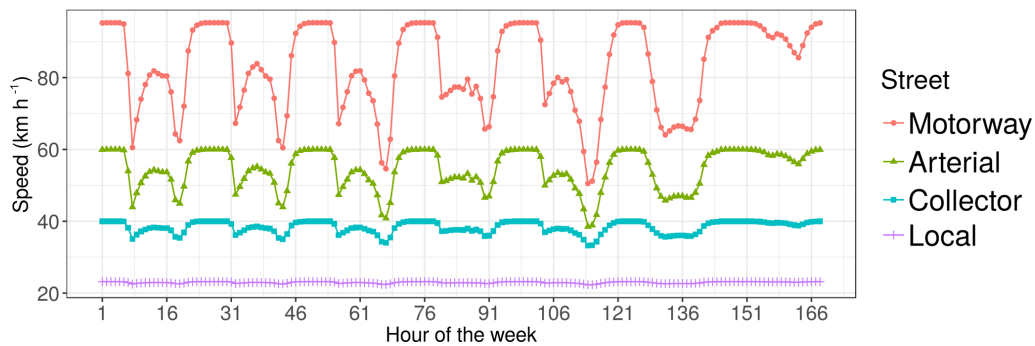


Figure 4. Traffic average speeds for LDV fleet by type of street (colored lines) at 08:00–09:00 LT in MASP.

clude national and equivalent Euro emission standards by year and age. The equivalence among Brazilian CONAMA (1986), MMA (2011), MMA (2015) and the Euro Directive 70/220/EEC (1991) was added to this database in order to choose the corresponding matching vehicle and emissions standard. The equivalence can be seen in Table 2.

fe2015 emission factors do not include the deterioration effect due to the accumulated age of vehicles, and it must be included. This is done with the deterioration factor function *emis_det*, which has the arguments pollutant, size of engine, Euro standard and mileage in kilometers. VEIN includes a Brazilian database of mileage functions named “fkm”, which is a list of functions with each element of the list corresponding to vehicle type. These functions depend on the vehicle’s age of use, and they originate from the odometer readings of more than $1.6 \cdot 10^6$ vehicles (Bruni and Bales, 2013). *emis_det* includes deterioration factors from Ntziachristos and Samaras (2016), which are based on measurements of European vehicles consuming gasoline; however, most of Brazilian automotive fuel sold has a mixture of bio-fuels such as ethanol from sugarcane. To our knowledge, there are no published deterioration factors for Brazilian conditions. Hence, we believe that including deterioration factors from Ntziachristos and Samaras (2016) is a valid option in the absence of better data. Another aspect that it is impor-

Table 2. Proposed equivalence of emission standards used in São Paulo study.

Vehicle	Brazilian standard	Euro standard	Year
LDV	L – 1	Pre Euro	1988–1991
	L – 2	Euro 1	1992–1996
	L – 3	Euro 2	1997–2004
	L – 4	Euro 3	2005–2011
	L – 5	Euro 4	2012–2013
	L – 6	Euro 5	2014
HDV	P – 1	Pre Euro	1990–1992
	P – 2	Pre Euro	1993
	P – 3	Euro 1	1994–1997
	P – 4	Euro 2	1998–2003
	P – 5	Euro 3	2004–2011
	P – 6	Euro 4	–
	P – 7	Euro 5	2012
Motorcycle	M – 1	Euro 1	2003–2005
	M – 2	Euro 2	2006–2008
	M – 3	Euro 3	2009

Based on <https://www.transportpolicy.net/standard/brazil-light-duty-emissions/>, last access: 31 May 2018.

tant to mention is that, when improved fuels are consumed in an older fleet, for instance fuel designed for Euro III technology in a pre-Euro fleet, this will diminish the emissions in older fleet. This fuel effect it is being included in the development repository of VEIN.

Emission factors for PCs, light commercial vehicles (LCVs) and motorcycles are called with the function “ef_ldv_speed”. In the case of trucks and buses, the function “ef_hdv_speed” is used. The arguments are filters for an internal database of emission factors which include several parameters such as fuel, Euro standard, volume of engine and load, among others. These functions also include a multiplication factor with a default value of 1. Exact spelling is required when using the arguments. If the argument names are entered incorrectly, VEIN will not return the emission factor functions.

The following code shows how to read the emission factors of the VEIN databases fe2015, “pc_profile” and fkm, in order to incorporate the deterioration effect into the CETESB (2016) emission factors. The age of LDV shown in Fig. 5a has a length of 41 years. This means that it needs 41 emission factors, one per each age of use. We used a 41-year distribution because we want to have a realistic representation of the vehicles in circulation by age of use, despite the oldest vehicles having a small number. However, the user could use a different age distribution, such as 50 years. This allow us to obtain emissions by the age of use of the vehicles; hence, VEIN can be used as a fast tool for scrapping policies and other applications. It calls the function emis_det, which requires the accumulated mileage, obtained from the list of mileage equations fkm. Fig. 5c) shows the emission factors from CETESB with and without deterioration by age of use. We are using deterioration factors from Ntziachristos and Samaras (2016) that affect only vehicles with a catalytic system. The base year of this emissions estimation is 2015, and vehicles with catalytic systems started in 1992 in Brazil (23 years before 2015). Therefore, the vehicles that entered the market before 1992 do not include deterioration. The emission factors dataset fe2015 includes emission factors for vehicles with only 36 years of use, but the vehicular distribution calculated in the last script has 41 years of use. Therefore, we repeated the oldest emission factors to have 41 emission factors. Here we are assuming that the emission factors of vehicles with 36 years of use are the same as for the vehicles with 41 years of use. The last line of the following script calculates the deteriorated emission factors of passenger cars by age of use.

```
data(fe2015)
data(fkm)
pckm <- fkm[[1]](1:24)
pckma <- cumsum(pckm)
cod1 <- emis_det(po = "CO", cc = 1000,
eu = "III", km = pckma[1:11])
cod2 <- emis_det(po = "CO", cc = 1000,
```

```
eu = "I", km = pckma[12:24])
col <- fe2015[fe2015$Pollutant == "CO", ]

col[37:41, ] <- col[36, ]
cod <- c(col$PC_G[1:24] * c(cod1, cod2),
col$PC_G[25:nrow(col)])
```

Once the deterioration effect was added to the Brazilian emission factors (CETESB, 2016), they were scaled to account for speed with the function “ef_ldv_scaled”. This function is used to multiply emission factors from ef_ldv_speed with a constant.

The new emission factor (dependent on speed) has the same value as the local emission factor, which is evaluated at the reference speed of the measurement 34.12 km h^{-1} for FTP-75. The default speed value is 34.12 km h^{-1} , but this value must change correspondingly to the speed of the driving conditions. To use this function, it is necessary to scale emission standards of local emission factors with Euro standards. In the following code, “Euro_LDV” is a vector indicating Euro standard by age of use.

```
lef <- ef_ldv_scaled(col, cod, v = "PC",
cc = "<=1400", f = "G", p = "CO",
eu = col$Euro_LDV)
```

4.3 Emission estimation

After inputting the database of vehicles and their respective emission factors, VEIN is ready to use the emis function. The VEIN package uses several emis functions according to the type of emission being estimated. The emis function assembles data and outputs from other VEIN functions and estimates the emissions for the number of hours and days in the week. This function reads the morning rush hour traffic data by age of vehicle use and extrapolates it with the profile data frame, as previously explained. It reads the emission factors stored in a list with length matching the age distribution of the vehicle category and then reads the list of speeds. This function returns the emissions at each street in an array with four dimensions: (1) number of streets; (2) maximum age of age distribution; (3) hours (usually 24); and (4) days (usually 7). For convenience, there are defined default values for this function: hour – 24; day – 7; array – TRUE. The values can be changed as necessary.

For example, the estimation of the traffic simulations shown in Fig. 2 has 34 733 streets, a fleet with a 41-year age distribution, 24 h of the day, and 7 days of the week. Therefore, it will produce an emissions array with the dimensions 34 733, 41, 24 and 7. The vehicle fleet used to produce the age distribution is shown in green in Fig. 5a, and it is 41 years in length.

```
E_CO <- emis(veh = pcl, lkm = net$lkm,
ef = lef, speed = speed,
profile = pc_profile)
```

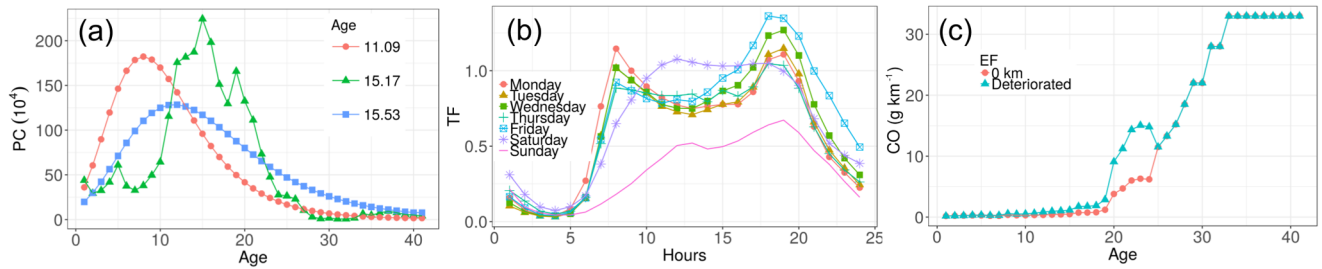


Figure 5. (a) Distribution of LDV composition by age of use, (b) temporal factor for expanding morning rush hour traffic data and (c) CO emission factors used in the estimation presented in this paper.

This emissions array output for 34 733 streets and a vehicle fleet with a 41-year age distribution, 24 h and 7 days of the week, has the size of 1.8 Gb. Hence, it is recommended to use the function `emis_post` and then delete the original emissions array. The arguments include the emissions array, type of vehicle, size or weight, fuel, pollutant, and the boolean argument “by”. The `emis_post` function was created to preserve the most important information in the emissions array, to use less memory size and to be compatible with the packages `sp` (Pebesma and Bivand, 2005) and “`ggplot2`” (Wickham, 2009). VEIN outputs could also be used with the package “`openair`” (Carslaw and Ropkins, 2012). `emis_post` returns a data frame, but the argument `by` determines the shape of the data frame. When `by` has the value “veh”, it returns a data frame with an aggregation of the emissions array by each vehicle’s age of use with the following columns: vehicle name, emission (in grams), vehicle type, size, fuel, pollutant, age, hour and day. This output allows the user to visualize hourly emissions at each day of the week, as shown in Fig. 6. Higher emissions are found in the morning and evening rush hours from Monday to Friday. Saturday has peak higher emissions at noon and Sunday has the lowest emissions.

VEIN enables the user to identify which type of vehicle emits more by age of use. This is particularly useful for environmental authorities, who aim to reduce local traffic emissions and restrict the circulation of highly emitting vehicles. Figure 6 shows the CO emissions of gasoline fueled LDV by the vehicle’s age of use. The average age of these vehicles is 15.17 years, as shown by the green curve in Fig. 5a. The total number of vehicles is 24 708 767 veh h⁻¹ (08:00–09:00 LT on a Monday). The total CO emissions is 233 095 t yr⁻¹, considering a year of 52 weeks, but the emissions are concentrated for the LDV between 20 and 23 years of use. The vehicles in this age interval represent 14.76 % of the fleet, emitting 63.79 % (148 712 t yr⁻¹) of the total emissions. In other words, 15 % of the fleet emit more than 60 % of the CO. Between 1992 and 1996, the emissions standard was Proconve L2, equivalence with Euro 1 (see Table 2), and the catalytic system was also introduced in Brazil. Therefore, the high emissions are due to vehicles with a deteriorated catalytic converter. This information is useful for reducing air pollu-

tion, thus supporting the aims of environmental planners and local authorities.

4.4 Post estimation

The spatial dimensions of the emissions estimation is an important feature of VEIN because it allows the representation of the streets with spatial vectors. This is accomplished by using the function `emis_post` with the argument `by` equal to “`streets_narrow`” or “`streets_wide`”. Both options return a data frame with different characteristics, which can be converted into spatial vectors. When `by` is equal to `streets_narrow`, it returns a data frame with four columns: ID, indicating the number of rows; emissions; hour and day of the week. The number of rows in the data frame is the original number of selected streets multiplied by the hours and days of the week. For example, when there are 34 733 streets, 24 h and 7 days of the week, it returns a data frame with 5 835 144 rows having a size of 133.6 Mb. This option is useful to visualize the temporal behavior of specific streets with `ggplot2` (Wickham, 2009) or “`ggmap`” (Kahle and Wickham, 2013), for instance.

In most cases, users will be particularly interested when the argument `by` is equal to `streets_wide`. This produces a data frame with the number of rows matching the number of streets for the domain and the number of columns as hours. Figure 7 shows the CO emissions for LDV at each street on a Friday at 19:00 LT. The following code shows how to produce hourly emissions by street and then add these emissions back into the `SpatialLinesDataFrame` `net`. This is possible because the number of rows in `E_CO_STREETS` is equal and it matches the number of rows in `net`.

```
E_CO_STREETS <- emis_post(arr = E_CO,
  pollutant = "CO", by = "streets_wide")
net@data <- cbind(net@data, E_CO_STREETS)
```

The emissions shown in Fig. 7 are concentrated in two streets: a motorway and a trunk street in the northern part of the emissions map. This image was generated with the function `spplot` in the package `sp`.

It also shows a rectangular grid, which can be used for allocating the emissions. The allocation of emissions into

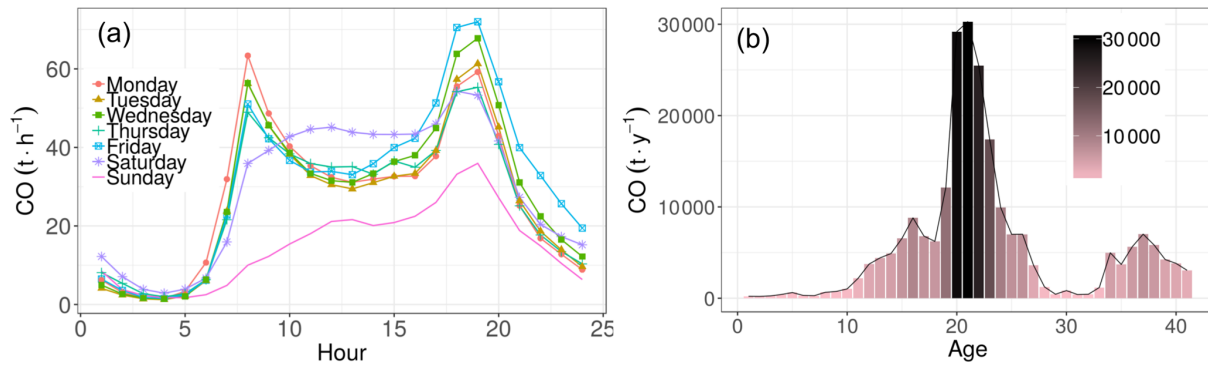


Figure 6. (a) CO emissions (g h^{-1}) per hour of the day and day of the week (colored and shaped lines) for LDV from MASP; (b) CO emissions (t yr^{-1}) according to the age of use of the LDV from MASP.

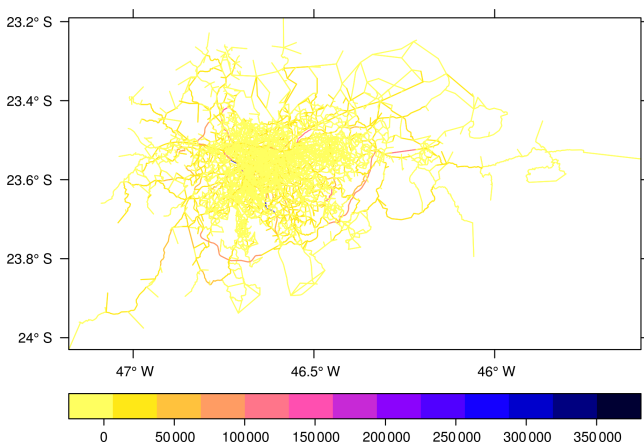


Figure 7. CO emissions (colored lines; g h^{-1}) for LDV on Friday 19:00 LT over MASP.

the grid is very important for visualization and for inputs to air quality models. We included a simple function to create a rectangular grid in VEIN. The function was named, `make_grid`, which has the arguments, “width”, “height” and a boolean argument “polygon” for determining the type of output. When the argument `polygon` is `TRUE`, it returns a `SpatialPolygonsDataFrame`, and when it is `FALSE`, it returns a `SpatialGridDataFrame`. The units of width and height depend on the coordinate reference systems of the data.

The allocation of emissions in each grid cell is produced by a spatial interception between the emissions at each street and the polygon grid. Firstly, the `SpatialLinesDataFrame` object with emission must contain a column with the length of the street. The length is calculated with the function “`gLength`” in the package `rgeos` (Bivand and Rundel, 2016). Secondly, it is performed at the intersection between the `SpatialLinesDataFrame` of emissions and the grid `SpatialPolygonsDataFrame`. The intersection is performed by importing the function “in-

tersect” in the package “`raster`” (Hijmans, 2016). The grid must have a column with the ID for each cell. Thirdly, it calculates, in another column, the length of the street in the resulting `SpatialPolygonsDataFrame`. Then it multiplies the emissions with the proportion of the new and old length of the street. This allows proportional emissions in each grid cell. Fourthly, it aggregates the emissions by the ID of the grid and adds these emissions by grid ID into the grid. The results are given in an emissions grid with the format `SpatialPolygonsDataFrame`. These calculations can be performed automatically by the function `emis_grid`.

The function `make_grid` is suitable in mid-size or small cities when the resolution is approximately 1 km. When dealing with larger cities and higher resolution, it is recommended to use other tools because `make_grid` would take up too much time. This difficulty will be overcome in a future version of VEIN with dependencies on the package `sf` (Pebesma, 2016). In the following code, we show the use of the function `make_grid` only for example purposes. It is recommended that the function `emis_post` be used with the argument “`by = streets_wide`”, in order to return a data frame with hourly emissions for each street. This output can be used with the functions `emis_grid` to create an emissions grid map, as shown in Fig. 8. This is helpful when the user plans to use the data to construct inputs for air quality models. The following code applies with the demo inside the VEIN model.

```
g <- make_grid(spobj = net, width =
0.00976,height = 0.00976, polygon = T)
E_CO_g <- emis_grid(spobj = net, g = g,
sr = "+init=epsg:31983", type =
"lines")
```

Figure 8a shows the resulting emissions of CO in a grid with class `SpatialPolygonsDataFrame` built for a grid with a resolution of 1 km representing the base year 2015. This emissions grid was built using the package `sf` (process shown in Appendix A). Figure 8b shows the CO emissions grid of road transport from EDGAR for the same

area and base year 2010 (EJ-JRC/PBL, 2016), which is the latest available year.

4.5 Speciation

Atmospheric simulations of ozone require knowledge about the VOC compounds and particulate matter speciation, which are necessary for solving the different chemical mechanisms. For example, a São Paulo study of ozone concentrations that used models BRAMS/SPM (Freitas et al., 2005) and WRF/Chem (Grell et al., 2005) involved detailed VOC speciation (Andrade et al., 2015). It is important to mention that there is evidence to prove that reducing black carbon emissions would help lower the global radiative forcing and improve population health (Bond et al., 2013). Hence, the speciation of emissions is important and VEIN provides this information. The VEIN function `speciate` splits VOC and PM into their constituents. The arguments of these functions are emissions estimation, type of speciation, type of vehicle, fuel and Euro standard. There are four types of PM speciation: “bcom” because it splits PM in black carbon and organic matter (Ntziachristos and Samaras, 2016), “tire”, “brake” and “road” (Ntziachristos and Boulter, 2009). However, there is currently only one type of VOC speciation for MASP, the “iag” (Ibarra, 2017). The name of the speciation `iag` comes from the initials of the Institute of Astronomy, Geophysics and Atmospheric Sciences of the University of São Paulo. The speciation `iag` is based on measurements made by students of this institute. The speciation `iag` splits the VOC emissions for the Carbon Bond Mechanism Z (CMB-Z; Zaveri and Peters, 1999). If the user intends to use other mechanisms, then the user needs to know how to speciate the VOCs and PM based on the user’s own data. This means that the user must know the percentages needed to split the pollutants and use them in the argument k of any VEIN function for the emission factor of the respective type of vehicle and then use them to estimate the emissions of that fraction of vehicles. For example, if the user knows that 5 % of COV emissions for LDV consuming diesel are xylenes, then the user must use the function `ef_ldv_speed` or `ef_ldv_scaled` (or its own local emission factors) with the argument $k = 5/100$. The argument k is simply a factor added to the resulting emission function. Finally, the user must aggregate the emissions by pollutant.

4.6 Input of atmospheric models

Meteorological factors influence the chemical process of pollutants in the atmosphere. Therefore, their transport and behavior in the atmosphere must be predicted by a model that includes the meteorological components (“in-line” coupling of meteorology and chemistry), such as the Weather Research and Forecasting Chemistry model (WRF-Chem; Grell et al., 2005). This model has been widely used around the world since its inception (2005 to 2006).

WRF-Chem requires gridded emission fluxes as input data. There are tools to assimilate top-down emissions inventories, such as EDGAR (Olivier et al., 1996) and REanalysis of the TROpospheric chemical composition (RETRO; Schultz, 2007), using the software PREP-Chem (Freitas et al., 2011). These tools are very important to the modeling community; however, their spatial resolutions are very limited. VEIN includes functions to generate WRF-Chem inputs from the emissions grid with any desired resolution in the following way. VEIN estimates emissions of different pollutants at each street and also produces emissions grids needed to do the regional modeling. This is performed through the spatial intersection between emissions at streets and a polygon grid with the required resolution. The resulting grid has total emissions in each grid cell proportional to the length of the streets inside each cell.

Figure 8 shows a comparison for VEIN (Fig. 8a) and EDGAR (Fig. 8b), using an emissions inventory for the CO in MASP. One may note that CO is spatially well represented for VEIN by comparison with EDGAR. Furthermore, VEIN offers much more detail about the emission of this pollutant, which occurs mainly on urban motorways due to the high volume of traffic on these roads. The total CO emissions using VEIN are 1.73×10^{-6} ($\text{kg m}^{-2} \text{s}^{-1}$), considering the first second of a typical Monday at 00:00 LT, and EDGAR gives emissions of 8.46×10^{-8} ($\text{kg m}^{-2} \text{s}^{-1}$). Therefore, VEIN estimates are 20.50 times higher than EDGAR. This difference could be higher if compared with the morning rush hour of VEIN. However, it is important to mention that the estimate with VEIN for this paper is illustrative, and that more detailed emissions inventories should be made when comparing it to others. For example, the inventory for this paper includes estimates only for LDVs assuming that all are PCs. It does not include other types of vehicles as the total amount of vehicles were not calibrated with fuel consumption. Ntziachristos and Samaras (2016) recommends comparing bottom-up estimates with fuel consumption in order to calibrate inputs of emissions inventory (traffic data in this case). These differences highlight the need for development, intercomparison and uncertainty evaluation of emission estimates. These results are very useful for many scientific and standardization purposes such as health effects in air pollution studies, urban planning and strategies to cut greenhouse gas emissions.

The VEIN model provides functions to transform the emissions grids into inputs for the model Assimilating Anthropogenic Emissions (AS4WRF) (Vara-Vela et al., 2016). AAS4WRF consists of an NCL (Boulder, 2017) script that reads the `wrfinput` file and an emissions text file produced by VEIN to create a WRF-Chem input file. The VEIN model provides the function `emis_wrf` to automatically create a data frame in the correct format with the columns longitude, latitude, ID of grid cell, pollutants, local time and GMT time in the format POSIXct. The arguments of `emis_wrf` are “`sdf`”, which is a list of `SpatialPolygonsDataFrames`; each are given per pollutant. “`nr`” indicates how many times the

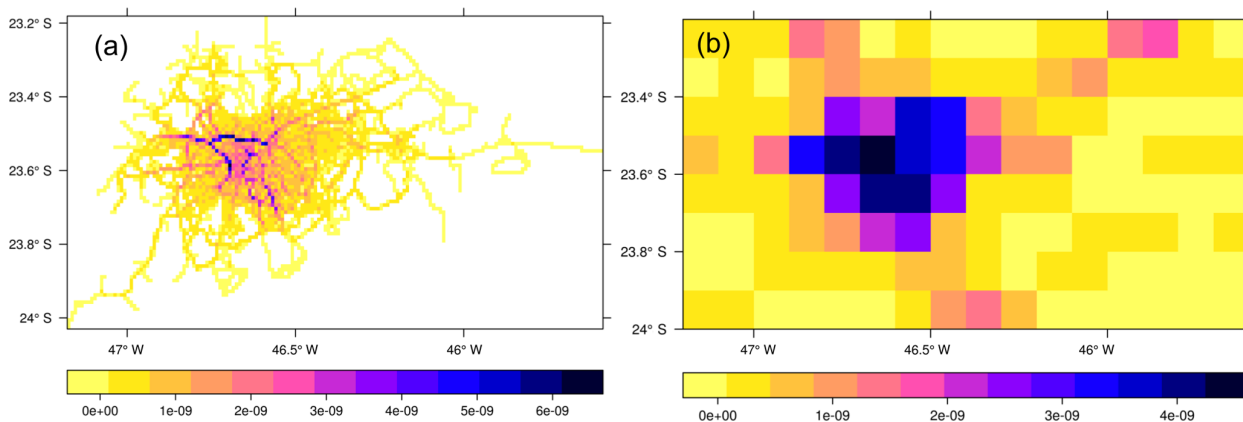


Figure 8. CO emissions ($\text{kg m}^{-2} \text{s}^{-1}$) in MASP for (a) Monday at 00:00:00 LT, estimated with VEIN; (b) the emissions of road transport for the same area from EDGAR.

hours of estimations will be repeated. “dmyhm” indicates the day, month, year, hour and minute of the first hour of emissions in local time. “tz” is the time zone, and “utc” indicates the difference in hours between local and GMT time. “islist” indicates whether the first argument, df, is a list (islist = TRUE) or a data frame (islist = FALSE). The output of VEIN emissions are in grams per hour, and they need to be converted to mol.

```
ldf <- list("co" = E_CO_g)
df_wrf <- emis_wrf(ldf, nr = 1,
dmyhm = "04-08-2014 00:00", tz =
"America/Sao_Paulo", utc = -3,
islist = TRUE)
```

The authors created another model called eixport (Ibarra-Espinosa et al., 2018) for producing emission inputs to atmospheric models. One of the characteristics is creating WRF-Chem input files in NetCDF format inside R (see <https://CRAN.R-project.org/package=eixport>, last access: 31 May 2018). Other models considered in eixport are BRAMS-SPM (Freitas et al., 2005) and R-LINE (Snyder et al., 2013).

5 Discussion and conclusions

In this paper we introduce the development of the Vehicular Emission INventory (VEIN model v0.2.2), an open-source model, to produce high-resolution spatial and hourly emission estimation. VEIN is a tool suited to application in complex environmental science studies, including regional atmospheric modeling. It generates inputs for air quality models in order to forecast air pollutant concentrations or for studies of greenhouse gas emissions from vehicular sources. It can be used to study the relationship between emissions and health effects. VEIN can be used as a tool for urban planning in order to estimate vehicular emissions due to interventions at road networks in most cities. It was written in an R pack-

age that includes several methods for estimating vehicular emissions in a harmonized way.

VEIN provides functions to easily produce inputs of regional air quality models such as WRF-Chem (Grell et al., 2005) and BRAMS/SPM (Freitas et al., 2005). In Fig. 8, the comparison between VEIN and EDGAR (EJ-JRC/PBL, 2016) of CO emissions shows that emissions are heavily concentrated in few streets with a high volume of traffic. EDGAR emissions do not provide this level of detail, and they are lower than VEIN estimates. Furthermore, the highest spatial resolution of EDGAR is 0.1° (approximately 12 km) and it is possible to have a better resolution with VEIN. Based on these factors, it can be concluded that EDGAR is suitable for modeling air pollution in larger domains without considering meteorological mesoscale interactions, including feedbacks. However, with progressing computational advances, it would be possible to perform air pollution modeling for larger domains with a higher resolution and level of detail. VEIN can produce these necessary inputs with a bottom-up approach.

VEIN currently experiences some limitations. The first limitation is the availability of activity data. VEIN needs at least 1 h of traffic data for each street considered in the estimation. These data can be based on traffic simulation or traffic counts; however, most cities do not count this type of data in developing countries. In this case, new data should be generated with traffic counts and interpolations. In addition, the widespread use of applications for smartphones such as Waze (R) or Uber (R), among others, produce traffic data that eventually could be used as activity data for estimating vehicular emissions. For example, currently Google Traffic (<https://developers.google.com/maps/coverage>, last access: 31 May 2018) covers several countries and these data could provide valuable information to estimate vehicular emissions in cities with non-traffic simulation or traffic counts. It is expected that new features will be added in future versions of VEIN. One very promising feature will be the migration of the spa-

tial dependencies into the new package, `sf` (spatial features) (Pebesma, 2016). This package provides S3 classes for handling spatial data faster than its predecessor, the package `sp` (Pebesma and Bivand, 2005).

The emission factors are another aspect of VEIN that can be enhanced in future versions. They could be sourced from several emissions studies, such as tunnel studies (Pérez-Martínez et al., 2014; Martins et al., 2006), or others based on traffic situations whereby emissions are sourced from driving cycles (ARTEMIS for example, André, 2004) or other experimental campaigns (Corvalán and Vargas, 2003). The International Vehicular Emissions (IVE) is a top-down vehicular emission model that has been used in different countries to estimate vehicular emissions (González et al., 2017; Wang et al., 2008). It could be possible to derive emission factors from IVE and estimate their corresponding emissions in VEIN, in order to use the capabilities of VEIN.

VEIN's purpose is to serve as a tool for air quality research and environmental management. Since air quality models need detailed emissions species, VEIN was created with the function `speciate`. VEIN will add several new speciations to these functions, such as those in the EMEP/EEA guidelines (Ntziachristos and Samaras, 2016). In the case of Brazil, there are several studies of tropospheric ozone, which use the speciation of VOC emissions as input (Vara-Vela et al., 2016; Abou Rafee et al., 2017).

Code and data availability. VEIN is available at the Comprehensive R Archive Network (CRAN) <https://CRAN.R-project.org/package=vein> (last access: 31 May 2018) and github <https://github.com/atmoschem/vein> (last access: 31 May 2018); Changelog is available at <https://github.com/atmoschem/vein/blob/master/NEWS.md> (last access: 31 May 2018). The software version Vehicular Emissions INventories (VEIN) v0.2.2 was developed using the R programming language v3.4.1. For further information about the VEIN package, contact the developer Sergio Ibarra-Espinosa at sergio.ibarra@usp.br. Datasets included emission factors `fe2015` and mileage functions `fkm` from the Environmental Agency of São Paulo (CETESB). In addition, emission and deterioration factors are provided in the form of lookup tables from Copert (Ntziachristos and Samaras, 2016). These include hot emission factors for light-duty vehicles (LDVs), heavy-duty vehicles (HDVs), cold start, evaporative factors, cold starts and resuspension and perform speciations. Emission factor functions are included for U.S. EPA/AP42 (U.S. EPA, 2002) estimations with default parameters. There are also two datasets to perform the example functions, a data frame to extrapolate morning peak hour vehicle traffic data to each hour of the week (`pc_profile`), and a vehicle start pattern ("`pc_cold`") to be used for cold-start estimation. Moreover, VEIN includes a traffic simulation for the western region of São Paulo City named "net". Software required: VEIN imports functions of the packages `sp`, `sf`, `rgeos`, `rgdal`, `raster` and `units`, which require libraries GDAL ($\geq 2.0.0$), GEOS ($\geq 3.4.0$) and UDUNITS-2.

Appendix A: Intercept grid and street emissions

Development versions of VEIN already include support for sf including the procedure shown in this Appendix.

```
library(vein)
library(sf)
# Converting objects into "sf" and
# projecting to UTM zone 23S - EPSG:31983
g_sf <- st_as_sf(spTransform(g,
CRSObj = "+init=epsg:31983"))
net_sf <- st_as_sf(spTransform(net,
CRSObj = "+init=epsg:31983"))
# Calculating initial length of net_sf
net_sf$LKM <- st_length(st_cast(net_sf[
st_dimension(net_sf) == 1,]))
# Intersecting net and grid
net_sf_g <- st_intersection(net_sf, g_sf)

# Calculating new length of net_sf
net_sf_g$LKM2 <- st_length(st_cast(
net_sf_g[st_dimension(net_sf_g) ==1,]))
# Checking dimensions
dim(net_sf_g)
# Converting from sf to data.frame
net_sf_gg <- as.data.frame(net_sf_g)
# Eliminating column 'geometry'
net_sf_gg <- net_sf_gg[, -171]
# Obtaining proportional emissions
# of 168 hours of the week
net_sf_gg[, 1:168] <- net_sf_gg[, 1:168] *
as.numeric(net_sf_gg
$LKM2/net_sf_gg$LKM)
# Agregating emissions by id of grid
dfm <- aggregate(cbind(net_sf_gg[, 1:168]),

by = list(net_sf_gg$id),

sum, na.rm = TRUE)
# naming columns of data-frame dfm
names(dfm) <- c("id", paste0("h", 1:168))

# Creating data-frame for id of grid
gx <- data.frame(id = g$id)
# merging data-frames
gx <- merge(gx, dfm, by="id", all.x = T)
# Generating spatial grid and
# converting to "Spatial"
E_CO_g <- as(st_sf(gx,
geometry = g_sf$geometry), "Spatial")
# Converting to WGS84 lat lon
E_CO_g <- spTransform(E_CO_g,
"+init=epsg:4326")
```

Appendix B: Examples using VEIN with different emission factors

B1 Activity data

```
data(net)
data(pc_profile)
data(fe2015)
pc1 <- age_ldv(x = net$ldv, name = "PC")
pcw <- temp_fact(net$ldv+net$hdv, pc_profile)
```

```
speed <- netspeed(pcw, net$ps, net$ffs,
net$capacity, net$lkm, alpha = 1)
```

B2 Emission factors

B2.1 Copert emission factors (without correction of deterioration and fuel effects)

```
euro <- c(rep("V", 5), rep("IV", 5),
rep("III", 5), rep("II", 5),
rep("I", 5), rep("PRE", 15))
lef1 <- lapply(1:40, function(i) {
ef_ldv_speed(v = "PC", t = "4S",
cc = "<=1400", f = "G",
eu = euro[i], p = "CO",
show.equation = FALSE) })
```

B2.2 Local emission factors of passenger cars using gasoline, constant by age of use

```
col <- fe2015[fe2015$Pollutant=="CO", ]
lef2 <- EmissionFactorsList(col$PC_G)
```

B2.3 Scaled emission factors

```
lef3 <- ef_ldv_scaled(dfcol = col$PC_G,
v = "PC", t = "4S", cc = "<=1400",
f = "G", p = "CO", eu=col$Euro_LDV)
```

B3 Estimations

```
E_COv1 <- emis(veh = pc1, lkm = net$lkm,
ef = lef1, speed = speed,
profile = pc_profile)
E_COv2 <- emis(veh = pc1, lkm = net$lkm,
ef = lef2, speed = speed,
profile = pc_profile)
E_COv3 <- emis(veh = pc1, lkm = net$lkm,
ef = lef3, speed = speed,
profile = pc_profile)
```

Author contributions. SIE designed and implemented the entire model. All authors contributed to the final manuscript.

Competing interests. The authors declare they have no actual or potential competing interest.

Acknowledgements. We thank the Environmental Agency of São Paulo State (CETESB), the programs CAPES and CNPQ (Ministry of Education, Brazil), and CONICYT (Ministry of Education, Chile). This work was also partially supported by the FONDAP/CONICYT 15110009 project. These entities provided data and support that were invaluable to this study. We also thank Prof. Dr. Zissis Samaras and Prof. Dr. Leonidas Ntziachristos from Aristotle University, since VEIN relies heavily on all their research about vehicular emissions and the model Copert.

Edited by: Augustin Colette

Reviewed by: two anonymous referees

References

- Abou Rafee, S. A., Martins, L. D., Kawashima, A. B., Almeida, D. S., Morais, M. V. B., Souza, R. V. A., Oliveira, M. B. L., Souza, R. A. F., Medeiros, A. S. S., Urbina, V., Freitas, E. D., Martin, S. T., and Martins, J. A.: Contributions of mobile, stationary and biogenic sources to air pollution in the Amazon rainforest: a numerical study with the WRF-Chem model, *Atmos. Chem. Phys.*, 17, 7977–7995, <https://doi.org/10.5194/acp-17-7977-2017>, 2017.
- Albuquerque, T.: Distribuição de tamanho, composição química e identificação das fontes de aerossol atmosférico de São Paulo: Um estudo de caso para o inverno de 2003, Ph.D. thesis, São Paulo: University of São Paulo, 2005.
- Andrade, M. D. F., Ynoue, R. Y., Freitas, E. D., Todesco, E., Vara Vela, A., Ibarra, S., Martins, L. D., Martins, J. A., and Carvalho, V. S. B.: Air quality forecasting system for Southeastern Brazil, *Front. Environ. Sci.*, 3, 1–12, <https://doi.org/10.3389/fenvs.2015.00009>, 2015.
- Andrade, M. D. F., Kumar, P., de Freitas, E. D., Ynoue, R. Y., Martins, J., Martins, L. D., Nogueira, T., Perez-Martinez, P., de Miranda, R. M., Albuquerque, T., Gonçalves, F. L. T., Oyama, B., and Zhang, Y.: Air quality in the megacity of São Paulo: Evolution over the last 30 years and future perspectives, *Atmos. Environ.*, 159, 66–82, <https://doi.org/10.1016/j.atmosenv.2017.03.051>, 2017.
- André, M.: The ARTEMIS European driving cycles for measuring car pollutant emissions, *Sci. Total. Environ.*, 334, 73–84, 2004.
- ARTESP: Toll station count data, available at: <http://www.artesp.sp.gov.br/> (last access: 31 May 2018), 2012.
- Barlow, T., Latham, S., McCrae, I., and Boulter, P.: A reference book of driving cycles for use in the measurement of road vehicle emissions, TRL Published Project Report, 2009.
- Bivand, R. and Lewin-Koh, N.: *mapproj*: Tools for Reading and Handling Spatial Objects, available at: <http://CRAN.R-project.org/package=mapproj> (last access: 31 May 2018), r package version 0.8-34, 2015.
- Bivand, R. and Rundel, C.: *rgeos*: Interface to Geometry Engine – Open Source (GEOS), available at: <https://CRAN.R-project.org/package=rgeos> (last access: 31 May 2018), r package version 0.3-21, 2016.
- Bivand, R., Pebesma, E., and Gomez-Rubio, V.: *Applied spatial data analysis with R*, Springer, NY, available at: <http://www.asdar-book.org/> (last access: 31 May 2018), 2013.
- Bivand, R., Keitt, T., and Rowlingson, B.: *rgdal*: Bindings for the Geospatial Data Abstraction Library, available at: <https://CRAN.R-project.org/package=rgdal> (last access: 31 May 2018), r package version 1.1-10, 2016.
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, *J. Geophys. Res.-Atmos.*, 118, 5380–5552, <https://doi.org/10.1002/jgrd.50171>, 2013.
- Boulder, C. U.: The NCAR Command Language (NCL) (version 6.4.0), <https://doi.org/10.5065/D6WD3XH5>, 2017.
- Boulter, P.: Environmental traffic management: A review of factors affecting cold start emissions, Tech. rep., Transport Research Laboratory, Old Wokingham Road, Crowthorne, Berkshire, RG45 6AU, UK, 1997.
- Bruni, A. D. C. and Bales, M. P.: Curvas de intensidade de uso por tipo de veículo automotor da frota da cidade de São Paulo, Tech. rep., CETESB, available at: <http://cetesb.sp.gov.br/veicular/wp-content/uploads/sites/6/2013/12/curvas-intensidade-uso-veiculos-automotores-cidade-sao-paulo.pdf> (last access: 31 May 2018), 2013.
- Bureau of Public Roads: Traffic and assignment manual, Tech. rep., Dept. of Commerce, Urban Planning Division, Washington D.C., 1964.
- Carlsaw, D. C. and Ropkins, K.: *Openair* – an R package for air quality data analysis, *Environ. Modell. Softw.*, 27, 52–61, 2012.
- CET: Simulação de tráfego para Região Metropolitana de São Paulo, available at: <http://www.cetsp.com.br> (last access: 31 May 2018), 2014.
- CETESB: Emissões Veiculares no Estado de São Paulo, available at: <http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/> (last access: 31 May 2018), 2013.
- CETESB: Emissões Veiculares no Estado de São Paulo, available at: <http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/> (last access: 31 May 2018), 2016.
- Chen, R.-H., Chiang, L.-B., Chen, C.-N., and Lin, T.-H.: Cold-start emissions of an SI engine using ethanol–gasoline blended fuel, *Appl. Therm. Eng.*, 31, 1463–1467, 2011.
- CONAMA: Programa de controle de emissões veiculares (Proconve), available at: <http://www.mma.gov.br/port/conama/res/res86/res1886.html> (last access: 21 January 2014), 1986.
- Corvalán, R. M. and Vargas, D.: Experimental analysis of emission deterioration factors for light duty catalytic vehicles Case study: Santiago, Chile, *Trans. Res. D*, 8, 315–322, 2003.
- Corvalán, R. M., Osses, M., and Urrutia, C. M.: Hot emission model for mobile sources: application to the metropolitan region of the

- city of Santiago, Chile, *J. Air Waste Manage. Assoc.*, 52, 167–174, 2002.
- Davis, N., Lents, J., Osses, M., Nikkila, N., and Barth, M.: Part 3: Developing countries: development and application of an international vehicle emissions model, *Transportation Research Record, J. Trans. Res. Board*, 8, 155–165, 2005.
- DENATRAN: National Department of Traffic Road, available at: <http://www.denatran.gov.br/frota2015.htm> (last access: 4 February 2016), 2015.
- Directive70/220/EEC: Measures to be taken against the emission of gaseous pollutants from diesel engines for use in vehicles, available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31991L0441&from=EN> (last access: 31 May 2018), 1991.
- EEA: Emission Inventory Report 1990–2011 Under the UNECE Convention on LRTAP, European Environment Agency, Luxembourg, 2013.
- EJ-JRC/PBL: Emission Database for Global Atmospheric Research (EDGAR), release EDGAR v4.3.1_v2 (1970–2010), available at: <http://edgar.jrc.ec.europa.eu> (last access: 31 May 2018), 2016.
- U.S. EPA: Comprehensive analysis of biodiesel impacts on exhaust emissions, Tech. rep., Draft Technical Report, EPA420-02-001, 2002.
- Freitas, E., Martins, L., Silva Dias, P., and Andrade, M. D. F.: A simple photochemical module implemented in RAMS for tropospheric ozone concentration forecast in the metropolitan area of São Paulo, Brazil: Coupling and validation, *Atmos. Environ.*, 39, 6352–6361, 2005.
- Freitas, S. R., Longo, K. M., Alonso, M. F., Pirre, M., Marecal, V., Grell, G., Stockler, R., Mello, R. F., and Sánchez Gácita, M.: PREP-CHEM-SRC – 1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models, *Geosci. Model Dev.*, 4, 419–433, <https://doi.org/10.5194/gmd-4-419-2011>, 2011.
- González, C., Gómez, C., Rojas, N., Acevedo, H., and Aristizábal, B.: Relative impact of on-road vehicular and point-source industrial emissions of air pollutants in a medium-sized Andean city, *Atmos. Environ.*, 152, 279–289, 2017.
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled “online” chemistry within the WRF model, *Atmos. Environ.*, 39, 6957–6975, 2005.
- Hijmans, R. J.: raster: Geographic Data Analysis and Modeling, available at: <https://CRAN.R-project.org/package=raster> (last access: 31 May 2018), r package version 2.5-8, 2016.
- Ibarra, S.: Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories, Ph.D. thesis, São Paulo: University of São Paulo, available at: http://www.iag.usp.br/pos/sites/default/files/t_serjio_a_i_espinosa_corrigida.pdf (last access: 31 May 2018), 2017.
- Ibarra-Espinosa, S.: vein: Vehicular Emissions Inventories, R package version 0.2.2-25, available at: <https://CRAN.R-project.org/package=vein><https://github.com/atmoschem/vein> (last access: 31 May 2018), 2017.
- Ibarra-Espinosa, S. and Pebesma, E.: `ibarraespinosa/vein: VEIN v0.2.2 (Version v0.2.2)`, Zenodo, available at: <https://doi.org/10.5281/zenodo.1039236> (last access: 31 May 2018), 2017.
- Ibarra-Espinosa, S. and Ynoue, R.: REMI model: Bottom-up emissions inventories for cities with lack of data, *J. Earth Sci. Geotech. Eng.*, 1, 277–288, 2017.
- Ibarra-Espinosa, S., Schuch, D., and Dias de Freitas, E.: `eixport`: An R package to export emissions to atmospheric models, *J. Open Source Softw.*, 3, 607, <https://doi.org/10.21105/joss.00607>, 2018.
- IBGE: Instituto Brasileiro de Geografia e Estatística – Población estimada 2014, available at: <http://cod.ibge.gov.br/232IH>, last access: 12 December, 2014.
- Kahle, D. and Wickham, H.: `ggmap`: Spatial Visualization with `ggplot2`, *The R J.*, 5, 144–161, available at: <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf> (last access: 31 May 2018), 2013.
- Koupal, J., Cumberworth, M., Michaels, H., Beardsley, M., and Brzezinski, D.: Design and Implementation of MOVES: EPA’s New Generation Mobile Source Emission Model, *Ann Arbor*, 1001, 48 105, 2003.
- Kucharski, R. and Drabicki, A.: Estimating Macroscopic Volume Delay Functions with the Traffic Density Derived from Measured Speeds and Flows, *J. Adv. Transp.*, 2017, 4629792, <https://doi.org/10.1155/2017/4629792>, 2017.
- Lam, W. H. and Xu, J.: Estimation of AADT from short period counts in Hong Kong – A comparison between neural network method and regression analysis, *J. Adv. Transp.*, 34, 249–268, 2000.
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., and Wolff, S. S.: The past and future history of the Internet, *Commun. ACM*, 40, 102–108, 1997.
- Lents, J., Davis, N., Nikkila, N., and Osses, M.: São Paulo Vehicle Activity Study, Tech. rep., International Sustainable System Research Center, 605 South Palm Street, Suite C, La Habra, CA 90631, USA, 2004.
- Ludykar, D., Westerholm, R., and Almen, J.: Cold start emissions at +22, -7 and -20 C ambient temperatures from a three-way catalyst (TWC) car: regulated and unregulated exhaust components, *Sci. Total Environ.*, 235, 65–69, 1999.
- Manual, H. C.: Highway capacity manual, Washington, DC, p. 11, 2000.
- Manzo, S., Nielsen, O., and Prato, C.: Effects of Uncertainty in Speed-Flow Curve Parameters on a Large-Scale Model: Case Study of the Danish National Model, *Transportation Research Record: J. Transport. Res. Board*, 2429, 30–37, 2014.
- Martins, L. D., Andrade, M. F., Freitas, E. D., Pretto, A., Gatti, L. V., Albuquerque, É. L., Tomaz, E., Guardani, M. L., Martins, M. H., and Junior, O. M.: Emission factors for gas-powered vehicles traveling through road tunnels in São Paulo, Brazil, *Environ. Sci. Technol.*, 40, 6722–6729, <https://doi.org/10.1021/es052441u>, 2006.
- Mellios, G. and Ntziachristos, L.: EMEP/EEA emission inventory guidebook; Gasoline evaporation from vehicles, Tech. rep., European Environment Agency, Copenhagen, 2016.
- Ministério do Meio Ambiente: Primeiro Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários, available at: http://www.mma.gov.br/estruturas/163/_publicacao/163_publicacao27072011055200.pdf (last access: 31 May 2018), 2011.

- Miranda, R.: Caracterização físico química e propriedades ópticas de aerossol urbano da região metropolitana de São Paulo, Ph.D. thesis, São Paulo: University of São Paulo, 2001.
- MMA: Air Pollution Control Program by Motor Vehicles; PROCONVE/PROMOT, available at: <http://www.ibama.gov.br/phocadownload/category/4?download=4793%3Amanual-ingles> (last access: 8 September 2015), 2011.
- MMA: PROCONVE: PROGRAMA DE CONTROLE DE POLUIÇÃO DO AR POR VEÍCULOS AUTOMOTORES, available at: http://www.mma.gov.br/estruturas/163/_arquivos/proconve_163.pdf, last access: 8 September, 2015.
- Nogueira, T., de Souza, K. F., Fornaro, A., de Fatima Andrade, M., and de Carvalho, L. R. F.: On-road emissions of carbonyls from vehicles powered by biofuel blends in traffic tunnels in the Metropolitan Area of São Paulo, Brazil, *Atmos. Environ.*, 108, 88–97, 2015.
- Ntziachristos, L. and Boulter, P.: EMEP/EEA emission inventory guidebook; Road Transport: Automobile tyre and break wear and road abrasion, European Environment Agency, Copenhagen, 2009.
- Ntziachristos, L. and Samaras, Z.: EMEP/EEA emission inventory guidebook; Road Transport: Passenger cars, Light commercial trucks, Heavy-duty vehicles including buses and Motorcycles, European Environment Agency, Copenhagen, 2016.
- Oliveira, M.: Evolução das distribuições de tamanho em massa e número do aerossol atmosférico em São Paulo, Ph.D. thesis, São Paulo: University of São Paulo, 2007.
- Olivier, J. G. J., Bouwman, A. F., van der Maas, C. W. M., Berdowski, J. J. M., Veldt, C., Bloos, J. P. J., Visschedijk, A. J. H., Zandveld, P. Y. J., and Haverlag, J. L.: Description of EDGAR Version 2.0: A set of global emission inventories of greenhouse gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1 degree \times 1 degree grid, National Institute of Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu RIVM), No. RIVM report: 771060 002, 1996.
- Paustian, K., Ravindranath, N., and van Amstel, A.: IPCC Guidelines for National Greenhouse Gas Inventories, Tech. rep., 2006.
- Pebesma, E.: sf: Simple Features for R, available at: <https://github.com/edzer/sf/> (last access: 31 May 2018), r package version 0.4-3, 2016.
- Pebesma, E. and Bivand, R.: Classes and methods for spatial data in R, available at: <http://cran.r-project.org/doc/Rnews/> (last access: 31 May 2018), 2005.
- Pebesma, E., Mailund, T., and Hiebert, J.: Measurement Units in R, *The R J.*, 8, 486–494, available at: <https://journal.r-project.org/archive/2016-2/pebesma-mailund-hiebert.pdf> (last access: 31 May 2018), 2016.
- Pérez-Martínez, P., Miranda, R., Nogueira, T., Guardani, M., Fornaro, A., Ynoue, R., and Andrade, M.: Emission factors of air pollutants from vehicles measured inside road tunnels in São Paulo: case study comparison, *Int. J. Environ. Sci. Technol.*, 11, 2155–2168, 2014.
- Perugu, H., Wei, H., and Yao, Z.: Developing high-resolution urban scale heavy-duty truck emission inventory using the data-driven truck activity model output, *Atmos. Environ.*, 155, 210–230, 2017.
- Pulles, T. and Heslinga, D.: The art of emission inventorying, TNO, Utrecht, 2010.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: <http://www.R-project.org/> (last access: 31 May 2018), 2017.
- Rafee, S.: Estudo numerico do impacto das emissões veiculares e fixas da cidade de Manaus nas concentrações de poluentes atmosféricos da região amazônica, Master's thesis, Londrina: Universidade Tecnológica Federal do Paraná, 2015.
- Schultz, M. G.: REanalysis of the Tropospheric chemical composition over the past 40 years, Max-Planck Institute for Meteorology, 48, ISSN: 1614-1199, Series: Earth System Science, Hamburg, Germany, 2007.
- Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, 3rd Edn., John Wiley & Sons, New Jersey, US, ISBN: 978118947401 2016.
- Smit, R., Ntziachristos, L., and Boulter, P.: Validation of road vehicle and traffic emission models – A review and meta-analysis, *Atmos. Environ.*, 44, 2943–2953, 2010.
- Snyder, M. G., Venkatram, A., Heist, D. K., Perry, S. G., Petersen, W. B., and Isakov, V.: RLINE: A line source dispersion model for near-surface releases, *Atmos. Environ.*, 77, 748–756, 2013.
- Streets, D. G., Bond, T., Carmichael, G., Fernandes, S., Fu, Q., He, D., Klimont, Z., Nelson, S., Tsai, N., Wang, M. Q., Woo, J. H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, *J. Geophys. Res.-Atmos.*, 108, D21, <https://doi.org/10.1029/2002JD003093>, 2003.
- Suh, S., Park, C.-H., and Kim, T. J.: A highway capacity function in Korea: measurement and calibration, *Transport. Res. A*, 24, 177–186, 1990.
- U.S. EPA: National Emissions Inventory (NEI), available at: <https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei>, last access: 31 May 2018.
- U.S. EPA: AP42 Compilation of emission factors; Resuspension emissions from paved roads, Environment Protection Agency, USA, 2016.
- Vara-Vela, A., Andrade, M. F., Kumar, P., Ynoue, R. Y., and Muñoz, A. G.: Impact of vehicular emissions on the formation of fine particles in the São Paulo Metropolitan Area: a numerical study with the WRF-Chem model, *Atmos. Chem. Phys.*, 16, 777–797, <https://doi.org/10.5194/acp-16-777-2016>, 2016.
- Wang, H., Chen, C., Huang, C., and Fu, L.: On-road vehicle emission inventory and its uncertainty analysis for Shanghai, China, *Sci. Total Environ.*, 398, 60–67, 2008.
- Wang, X. and Kockelman, K.: Forecasting network data: Spatial interpolation of traffic counts from Texas data, *Trans. Res. Record: J. Transp. Res. Board*, 1, 100–108, 2009.
- Weilenmann, M., Favez, J.-Y., and Alvarez, R.: Cold-start emissions of modern passenger cars at different low ambient temperatures and their evolution over vehicle legislation categories, *Atmos. Environ.*, 43, 2419–2429, <https://doi.org/10.1016/j.atmosenv.2009.02.005>, 2009.
- Wickham, H.: ggplot2: elegant graphics for data analysis, Springer New York, available at: <https://www.springer.com/us/book/9780387981413> (last access: 31 May 2018), 2009.
- Ynoue, R.: Modelagem numérica da formação, crescimento e transporte das partículas inorgânicas secundárias constituintes do aerossol urbano na região metropolitana de São Paulo, Ph.D. thesis, São Paulo: University of São Paulo, 2004.

Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications, *J. Geophys. Res.-Atmos.*, 104, 30387–30415, 1999.