

Sergio Alejandro Ibarra Espinosa

**Air pollution modeling in São Paulo using  
bottom-up vehicular emissions inventories**

**Modelagem da poluição atmosférica em São Paulo utilizando inventários de emissões  
veiculares bottom-up**

São Paulo

2017



Sergio Alejandro Ibarra Espinosa

**Air pollution modeling in São Paulo using bottom-up  
vehicular emissions inventories**

**Modelagem da poluição atmosférica em São Paulo utilizando inventários de emissões  
veiculares bottom-up**

Thesis submitted to Instituto de Astronomia,  
Geofísica e Ciências Atmosféricas of the  
Universidade de São Paulo in partial fulfill-  
ment of the requirements for the degree of  
Doctor of Sciences. Concentration area: Me-  
teorology.

Universidade de São Paulo  
Instituto de Astronomia, Geofísica e Ciências Atmosféricas  
Departamento de Ciências Atmosféricas

Supervisor: Rita Yuri Ynoue

**Versão Corrigida. O original encontra-se disponível na Unidade.**

São Paulo  
2017

---

Sergio Alejandro Ibarra Espinosa

Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories

Modelagem da poluição atmosférica em São Paulo utilizando inventários de emissões veiculares bottom-up/ Sergio Alejandro Ibarra Espinosa. – São Paulo, 2017-

156 p. : il. (some colorful) ; 30 cm.

Supervisor: Rita Yuri Ynoue

**Versão Corrigida. O original encontra-se disponível na Unidade.**

Thesis (Ph.D.) – Universidade de São Paulo  
Instituto de Astronomia, Geofísica e Ciências Atmosféricas  
Departamento de Ciências Atmosféricas , 2017.

1. Introduction, 2. Traffic data 3. VEIN v0.2.2: an R package for bottom-up Vehicular Emissions Inventories 4. Weather and research forecast model with chemistry (WRF-Chem) 5. Results 6. Conclusions 7. Academic Literature Produced During Ph.D. Bibliography Appendix A: Evaluation of tests of Wilcoxon for scenarios of simulation CDU 02:141:005.7

---



Sergio Alejandro Ibarra Espinosa

# **Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories**

**Modelagem da poluição atmosférica em São Paulo utilizando inventários de emissões  
veiculares bottom-up**

Thesis submitted to Instituto de Astronomia,  
Geofísica e Ciências Atmosféricas of the  
Universidade de São Paulo in partial fulfill-  
ment of the requirements for the degree of  
Doctor of Sciences. Concentration area: Me-  
teorology.

---

**Rita Yuri Ynoue**  
**Versão Corrigida. O original**  
**encontra-se disponível na Unidade.**  
Supervisor

---

**Mauricio Osses**  
Invited 1

---

**Edzer Pebesma**  
Invited 2

---

**Edmilson Freitas**  
Invited 3

---

**Marcelo Felix Alonso**  
Invited 4

São Paulo  
2017



*Con todo mi amor a mi madre QEPD*



# Acknowledgements

Gracias mamá que siempre creíste en mí y gracias por el tiempo que tuvimos juntos. Recuerdo cuando estabas en cama y me dijiste? "Recuerda lo que fuiste a hacer a Brasil" y aquí estoy entregando mi tesis de doctorado.... Un día nos volveremos a encontrar. Mama, gracias por enseñarme lo que es el trabajo, desde la feria a los 5 años, vendiendo cubos con mis hermanas para comprar útiles escolares.... tantas cosas importantes que tú me enseñó mamá... y estoy seguro que todavía tengo muchas cosas para aprender de ti.

Agradezco a mi familia, a mi polola Amanda y a su familia. Agradezco a mi orientadora Rita, Fátima y Edmilson, También agradezco a Mauricio Osses quien me ha estado ayudando, en realidad desde hace muchos tiempo partiendo en el Centro Nacional del Medio Ambiente en Chile. A los trabajadores del IAG y de la USP. A mis compañeros de la universidad. A Brasil y a Chile.

Agradezco al programa de becas CAPES de Brazil por el primer año de beca. También agradezco al programa de Becas Chilede Conicyt, Chile, por la beca de los últimos 3 años de doctorado.

Agradezco a CETESB, CET, SPTRANS, Laboratorio de Geoprocésamiento de la Poli-USP, Maplink y ARTESP por el otorgamiento de los datos cruciales para este tesis.

Agradezco a todas las personas que me han apoyado y también todos los que me han odiado y que han intentado colocarme el pie encima.

Quiero agregar que es importante compartir el conocimiento, guardarlo no sirve a nadie.



*"Sometimes i feel like giving up  
Then i remember i have a lot motherfuckers  
to prove wrong" (anonymous)*





# Abstract

IBARRA, S. **Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories**, 2017. XXXf. Tese (Doutorado) - Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, 2017.

"Modelagem da poluição atmosférica em São Paulo utilizando inventários de emissões veiculares bottom-up"

Neste trabalho, investiga-se o impacto de diferentes cenários de inventários de emissões veiculares na qualidade do ar nas regiões metropolitanas de São Paulo, Baixada Santista, Vale do Paraíba, Sorocaba e Campinas. A construção de inventários de emissões veiculares bottom-up é complexa, tendo que agregar informações diversas, como a composição da frota veicular (com a distribuição de idade, tipos de veículos, tipos de combustível) e os processos emissores (fatores de emissão para partidas a frio, emissões de escapamento e evaporativas). Além disso, para modelos de qualidade do ar, estas emissões ainda devem ser distribuídas no tempo e no espaço. Os cenários foram construídos utilizando diferentes fontes de dados, destacando-se as simulações de tráfego e contagem de veículos da CET e SPTRANS, e registros de deslocamentos de veículos através de GPS para a distribuição espaço-temporal do fluxo veicular. Para o cálculo de emissões foi desenvolvido um software open source chamado VEIN (Vehicular Emissions Inventories, disponível em <https://github.com/ibarraespinoza/vein>). As emissões simuladas para as regiões metropolitanas de São Paulo são maiores do que as emissões estimadas pela CETESB para todos os poluentes. A partir destes cenários, foram realizadas simulações de qualidade do ar com o modelo WRF-Chem. Os resultados variam para os diferentes poluentes. De uma forma geral, a variação diurna dos poluentes é bem simulada, mostrando que as emissões estão consistentes. Apesar dos maiores valores de emissão encontrados neste trabalho, as concentrações simuladas dos poluentes primários foi, em média, menor do que as concentrações observadas. Isto provavelmente é decorrente do fato dos ventos simulados serem mais fortes do que os ventos observados. Este trabalho mostra novos métodos para desenvolver inventários de emissões com diferentes dados fornecendo um novo enfoque para compreender os problemas de qualidade do ar.

**Key-words:** VEIN, Air Pollution, Vehicular Emissions COPERT, WRF-CHEM, SÃO PAULO.



# Abstract

IBARRA, S. **Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories**, 2017. XXXf. Tese (Doutorado) - Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, 2017.

"Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories"

In this work, the impact of different vehicle emission inventory scenarios on air quality in the metropolitan areas of São Paulo, Baixada Santista, Vale do Paraíba, Sorocaba and Campinas is investigated. The construction of bottom-up vehicular emissions inventories is complex, being necessary to aggregate diverse information, such as the composition of the vehicle fleet (with the distribution of age, types of vehicles, types of fuel) and the emitting processes (emission factors for cold starts, exhaust and evaporative emissions). In addition, for air quality models, these emissions must still be distributed in time and space. The scenarios were constructed using different data sources, highlighting traffic simulations and vehicle counting of CET and SPTRANS, and records of vehicular displacements through GPS for the spatial-temporal distribution of vehicular flow. For the calculation of emissions, an open source software called VEIN (Vehicular Emissions Inventories, available at <https://github.com/ibarraespinoza/vein>) was developed. The simulated emissions for the metropolitan regions of São Paulo are larger than the emissions estimated by CETESB for all pollutants. From these scenarios, air quality simulations were performed with the WRF-Chem model. The results vary for different pollutants. In general, the daily variation of the pollutants is well simulated, showing that emissions are consistent. Despite the higher emission values found in this work, the simulated concentrations of the primary pollutants were, on average, lower than the observed concentrations. This is probably due to the fact that the simulated winds are stronger than the observed winds. This work shows new methods to develop emission inventories with different data providing a new approach to understanding air quality problems.

**Key-words:** VEIN, Air Pollution, Vehicular Emissions COPERT, WRF-CHEM, SÃO PAULO.



# Resumen

IBARRA, S. **Air pollution modeling in São Paulo using bottom-up vehicular emissions inventories**, 2017. XXXf. Tese (Doutorado) - Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, 2017.

"Modelación de la contaminación atmosférica en São Paulo utilizando inventarios de emisiones vehiculares bottom-up"

En este trabajo se investiga el impacto de diferentes escenarios de inventarios de emisiones sobre la calidad del aire en las regiones metropolitanas de São Paulo, Baixada Santista, Vale de Paraíba, Sorocaba y Campinas. La construcción de los inventarios de emisiones es compleja, teniendo que agregar diversas informaciones como la composición de la flota vehicular (con la distribución de edad, tipos de vehículos, tipos de combustible) y los procesos emisores (factores de emisión para partidas en frío, emisiones de tubo de escape y evaporativas). Además de esto, para los modelos de calidad del aire, estos modelos deben ser distribuidos en tiempo y espacio. Los escenarios fueron construidos utilizando diferentes datos, destacándose las simulaciones de tráfico y conteo vehicular de la CET y SPTRANS, y los registros de desplazamiento de vehículos a través de GPS para la distribución espacio-temporal del flujo vehicular. Para el cálculo de las emisiones fue desarrollado un software llamado VEIN (Vehicular Emissions Inventories, disponible en <https://github.com/ibarraespinosa/vein>). Las emisiones estimadas para las regiones metropolitanas de São Paulo son mayores que las emisiones estimadas por la CETESB para todos los contaminantes. A partir de estos escenarios, fueron realizadas simulaciones de calidad del aire con el modelo WRF-Chem. Los resultados varían para los diferentes contaminantes. De forma general, la variación diurna de los contaminantes es bien simulada, mostrando que las emisiones están consistentes. A pesar que las emisiones son mayores en este trabajo, las concentraciones simuladas de los contaminantes primarios fue, en media, menor que las concentraciones observadas. Esto es probablemente por el hecho de que los vientos simulados son más fuertes que los vientos observados. Este trabajo muestra métodos nuevos para desarrollar inventarios de emisiones con diferentes datos entregando un nuevo enfoque para comprender los problemas de calidad del aire.

**Key-words:** VEIN, Air Pollution, Vehicular Emissions COPERT, WRF-CHEM, SÃO PAULO.



# List of Figures

Figure 1 – Area of study: Metropolitan Areas (delimited by black lines), Topography (gray shades) and road network (colored lines) . . . . .	35
Figure 2 – CET Traffic simulation of LDV from 08:00 to 09:00 LT. . . . .	38
Figure 3 – CET Traffic simulation of HGV from 08:00 to 09:00 LT. . . . .	39
Figure 4 – SPTrans simulation of buses from 08:00 to 09:00 LT. . . . .	39
Figure 5 – Traffic counts for (a) LDV and (b) T2A, between 07:00 and 08:00 LT, 2012. . . . .	42
Figure 6 – GPS . . . . .	47
Figure 7 – GPS . . . . .	48
Figure 8 – Speed ( $km \cdot h^{-1}$ ) on Monday at 00:00 LT. . . . .	50
Figure 9 – Temporal distribution of average max speed of Taxi ( $km \cdot h^{-1}$ ) by type of street for Monday 2014-10-06. . . . .	51
Figure 10 – Representation of the VEIN model. Boxes and circles represent functions and data, respectively. . . . .	62
Figure 11 – Traffic flow simulation for LDV ( $veh \cdot h^{-1}$ ) at 08:00-09:00 LT for MASP. . . . .	64
Figure 12 – Traffic speeds (colored lines; $km \cdot h^{-1}$ ) for LDV fleet at 08:00 LT (left panel) and 23:00 LT (right panel) in MASP. . . . .	65
Figure 13 – Traffic average speeds for LDV fleet by type of street (colored lines) at 08:00-09:00 LT in MASP. . . . .	66
Figure 14 – (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 manuscript. . . . .	70
Figure 15 – (a) CO Emissions ( $t \cdot h^{-1}$ ) per hour of the day and day of the week (colored and shaped lines) for LDV from MASP, (b) CO Emissions ( $t \cdot y^{-1}$ ) according to the age of use of the LDV from MASP. . . . .	71
Figure 16 – CO Emissions (colored lines; $g \cdot h^{-1}$ ) for LDV at Friday 19:00 LT over MASP. . . . .	72
Figure 17 – CO emissions ( $kg \cdot m^{-2} \cdot s^{-1}$ ) in MASP for (a) Monday at 00:00:00 LT, estimated with VEIN for a 1 km x 1 km grid spacing, and (b) the emissions of road transport for the same area from EDGAR. . . . .	75
Figure 18 – Fuel sales for 2014 for metropolitan areas Campinas, MASP, Santos, SJDC and Sorocaba. . . . .	77
Figure 19 – Normalized fleet of PC, LCV, HGV, Bus and MC by age of use. . . . .	80
Figure 20 – $CO$ , $NO_x$ , $HC$ and $PM$ Emission factors ( $g \cdot km^{-1}$ ) for PC, LCV, Medium Trucks, UB and MC (CETESB, 2015). . . . .	81
Figure 21 – CO emission factors ( $g \cdot km^{-1}$ ) for PC as a function of speed ( $km \cdot h^{-1}$ ). . . . .	83

Figure 22 – CO emission factors ( $g \cdot km^{-1}$ ) for PC as a function of speed ( $km \cdot h^{-1}$ ) and age of use. . . . .	84
Figure 23 – Evaporative emission factors for a) Diurnal, b) Hot-Soak and c) Running Losses $g \cdot km^{-1}$ by age of use for average monthly temperature between 20 and 35 Celsius based on (CETESB, 2015). . . . .	84
Figure 24 – Fuel stations on MASP, obtained from CETESB. . . . .	85
Figure 25 – Domains of 27, 9 and 3 km of grid spacing. . . . .	92
Figure 26 – Hourly accumulated precipitation at Mirante de Santana station, October 2014, (INMET, 2016) . . . . .	94
Figure 27 – GOES 13 satellite IR image for October 8th, at 15:00 LT. . . . .	95
Figure 28 – Box-plots for traffic counts. . . . .	98
Figure 29 – Box-plots for traffic counts of vehicles ( $veh \cdot h^{-1}$ ) for (a) type of streets and (b) number of lanes. . . . .	100
Figure 30 – Traffic count interpolation for (a) LDV, (b) UB, (c) T2A and (d) MC . .	106
Figure 31 – Traffic flow of Cars generated from GPS data ( $veh \cdot h^{-1}$ per lane). . .	109
Figure 32 – Traffic flow of Cars generated from GPS data with spatial correction using speed as proxy ( $veh \cdot h^{-1}$ ). . . . .	110
Figure 33 – Traffic flows derived from GPS recordings for Cars ( $veh \cdot h^{-1}$ ) on Monday 2014-10-06 08:00 - 09:00 LT. . . . .	111
Figure 34 – Traffic flows derived from GPS recordings for Trucks ( $veh \cdot h^{-1}$ ) on Monday 2014-10-06 08:00-09:00 LT. . . . .	112
Figure 35 – Hourly speeds in MASP with data from GPS, simulations and interpolations ( $km \cdot h^{-1}$ ). . . . .	114
Figure 36 – Signature (vehicular emissions by age of use and type of vehicle) in all metropolitan regions with GPS traffic data for different pollutants ( $t \cdot y^{-1}$ ). . . . .	116
Figure 37 – Emission maps of $CO$ , $NO_x$ , $PM$ and $NMHC$ due traffic activity from GPS, Interpolation and Simulations ( $g \cdot h^{-1}$ ) for 08:00-09:00LT Monday 2014. . . . .	118
Figure 38 – Emission maps of $NO$ using LAPAt top-down approach ( $g \cdot h^{-1}$ ) for 08:00-09:00LT. . . . .	119
Figure 39 – Hourly emissions of $CO$ , $NO_x$ , $NMHC$ and $PM$ due traffic activity from GPS, Interpolation and Simulations in MASP ( $g \cdot h^{-1}$ ). . . . .	120
Figure 40 – Meteorological and Air Quality Stations from CETESB used in this study. . . . .	121
Figure 41 – Hourly mean of observed and simulated temperature (Celsius) by Metropolitan Area. . . . .	122
Figure 42 – Hourly mean and standard deviation of observed and simulated wind speed ( $m \cdot s^{-1}$ ) by Metropolitan Area. . . . .	123



Figure 43 – Hourly mean and standard deviation of <i>CO</i> (ppm) by region and scenario. . . . .	124
Figure 44 – Hourly mean and standard deviation of <i>NO</i> ( $ug \cdot m^{-3}$ ) by region. . .	126
Figure 45 – Hourly mean and standard deviation of <i>NO2</i> ( $ug \cdot m^{-3}$ ) by region. . .	127
Figure 46 – Hourly mean and standard deviation of <i>SO2</i> ( $ug \cdot m^{-3}$ ) by region. . .	128
Figure 47 – Hourly mean and standard deviation of <i>O3</i> ( $ug \cdot m^{-3}$ ) by region. . . .	130



# List of Tables

Table 1 – Classification of vehicles . . . . .	34
Table 2 – Speeds Monday 2014-10-06 00:00 by type of street $km \cdot h^{-1}$ . . . . .	49
Table 3 – Average traffic and speeds by type of street and number of lanes at 08:00-09:00 in São Paulo, 2012 . . . . .	51
Table 4 – Summary of the VEIN classes, functions and internal data. . . . .	63
Table 5 – Proposed equivalence of emission standards used in São Paulo study. . . . .	67
Table 6 – Vehicular composition used in this study. . . . .	79
Table 7 – Speciation for NMHC by type of fuel and process $mol \cdot 100g^{-1}$ . MIR values from (MARTINS, 2006) and speciation values from (RAFEE et al., 2017) . . . . .	88
Table 8 – Parameterizations used in WRF-Chem . . . . .	93
Table 9 – Air quality and meteorology stations from CETESB used in this study . . . . .	96
Table 10 – Summary of statistics for traffic counts by type of street. . . . .	99
Table 11 – Regressions coefficients for LDV. . . . .	102
Table 12 – Regressions coefficients for UB. . . . .	103
Table 13 – Regression coefficients for T2A. . . . .	103
Table 14 – Regression coefficients for MC with Binomial Negative and Quasi-Poisson distributions. . . . .	104
Table 15 – Calibration factors for traffic by type of fuel in each region and scenario . . . . .	113
Table 16 – Emissions by each regions ( $t \cdot y^{-1}$ ) 2014 . . . . .	115
Table 17 – Evaluation of atmospheric simulation in each region . . . . .	131
Table 18 – Comparison with vehicular emissions inventories in different cities ( $t \cdot y^{-1}$ ) . . . . .	134
Table 19 – Test Wilcoxon for $CO$ , $NO$ and $NO_2$ . . . . .	155
Table 20 – Test Wilcoxon for $SO_2$ and $O_3$ . . . . .	156



# List of abbreviations and acronyms

<i>CH<sub>4</sub></i>	Methane
<i>O<sub>3</sub></i>	Ozone
<i>PM</i>	Particulate Matter
<i>SO<sub>x</sub></i>	Sulphur Oxides
<i>PM<sub>10</sub></i>	Particulate matter with aerodynamic diameter of 10 $\mu m$
<i>PM<sub>2.5</sub></i>	Particulate matter with aerodynamic diameter of 2.5 $\mu m$
<i>NO<sub>2</sub></i>	Dioxide of nitrogen
<i>NO</i>	Monoxide of nitrogen
<i>SO<sub>2</sub></i>	Dioxide of sulfur
<i>SO<sub>2</sub></i>	Dioxide of sulfur
<i>BEN</i>	Benzene
<i>TOL</i>	Toluene
<i>O<sub>3</sub></i>	Ozone
<i>TEMP</i>	Temperature at 2 m
<i>WS</i>	Wind speed
<i>WD</i>	Wind direction
<i>RAD</i>	Solar radiation
<i>ERT</i>	Total Reduced Sulfur
<i>P</i>	Atmospheric pressure
CLRTAP	Convention of 1979 on Long-Range Transboundary Air Pollution
COPERT	Computer programme to calculate emissions from road Transport
EDGAR	Emissions Database for Global Atmospheric Research
EPA	U.S. Environmental Protection Agency

IPCC	Intergovernmental Panel on Climate Change
NAAQS	National Ambient Air Quality Standards
MASP	Metropolitan Region of São Paulo
MRBS	Metropolitan Region of Baixada Santista
MRC	Metropolitan Region of Campinas
MRRP	Metropolitan Region of Ribeirão Preto
MRS	Metropolitan Region of Sorocaba
MRVPLN	Metropolitan do Vale do Paraíba e Litoral Norte
VEIN	Vehicular Emissions Inventories
VOC	Volatile Organic Compounds
WRF-Chem	Weather Research and Forecasting model coupled to Chemistry
LDV	Light Duty Vehicles covering Passenger Cars and Light Commercial Vehicles
PC	Passenger Cars
LCV	Light Commercial Vehicles
HDV	Heavy Duty Vehicles covering Urban Buses (UB) and Heavy Good Vehicles (HGV)
UB	Urban Buses (UB)
HGV	Heavy Good Vehicles or Trucks
MC	Motorcycles
E25	Gasoline blended with 25-27 % ethanol known as Gasohol
E10	Ethanol 100 %
B5	Diesel blended with 5 % bio-diesel
CET	Traffic Company of Engineering
SPTRANS	Secretary of transport and mobility of São Paulo
GW	Gross Weight

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>29</b>
<b>1.1</b>	<b>Objectives</b>	<b>31</b>
1.1.1	Specific Objectives	32
<b>2</b>	<b>TRAFFIC DATA</b>	<b>33</b>
<b>2.1</b>	<b>Types of vehicles</b>	<b>33</b>
<b>2.2</b>	<b>Road network</b>	<b>34</b>
2.2.1	Study area	35
2.2.2	Road network	36
<b>2.3</b>	<b>Traffic activity</b>	<b>36</b>
2.3.1	Traffic simulation from CET	37
2.3.2	Traffic simulation from SPTRANS	38
2.3.3	Traffic counts	38
2.3.3.1	Short Period Traffic Count (SPTC)	40
2.3.3.2	Regression approach	41
2.3.3.3	Evaluation	44
2.3.4	GPS tracking data	45
2.3.4.1	GPS data processing	45
<b>3</b>	<b>VEIN V0.2.2: AN R PACKAGE FOR BOTTOM-UP VEHICULAR EMISSIONS INVENTORIES</b>	<b>53</b>
<b>3.1</b>	<b>VEIN model: methodology to estimate vehicular emissions</b>	<b>54</b>
3.1.1	Traffic data	55
3.1.2	Selection of the emissions factors	56
3.1.3	Emissions estimation	57
3.1.4	Speciation of emissions in chemical sub-components	59
3.1.5	Spatial allocation and data-bases	60
<b>3.2</b>	<b>VEIN model design</b>	<b>60</b>
3.2.1	Functions and classes	61
<b>3.3</b>	<b>Estimating MASP vehicular emissions using VEIN model</b>	<b>62</b>
3.3.1	Traffic data for MASP	63
3.3.2	Emission Factors	67
3.3.3	Emission estimation	69
3.3.4	Post estimation	71
3.3.5	Speciation	73
3.3.6	Input of atmospheric models	74

<b>3.4</b>	<b>VEIN model applied to different traffic activity data</b>	<b>76</b>
3.4.1	Conciliating bottom-up and top-down emissions estimation	76
3.4.2	Running VEIN model for the different activity data scenarios	77
<b>3.5</b>	<b>Partial Conclusions</b>	<b>88</b>
<b>4</b>	<b>WEATHER AND RESEARCH FORECAST MODEL WITH CHEMISTRY (WRF-CHEM)</b>	<b>91</b>
<b>4.1</b>	<b>Model configuration</b>	<b>91</b>
<b>4.2</b>	<b>Meteorological conditions</b>	<b>93</b>
<b>4.3</b>	<b>Evaluation</b>	<b>94</b>
<b>5</b>	<b>RESULTS</b>	<b>97</b>
<b>5.1</b>	<b>Traffic data</b>	<b>98</b>
5.1.1	Traffic flow from interpolation	98
5.1.1.1	Discussion	105
5.1.2	Traffic flow from GPS	108
5.1.3	Traffic flow from simulations	109
5.1.4	Emissions scenarios	110
<b>5.2</b>	<b>Emissions inventories</b>	<b>113</b>
5.2.1	Traffic flow calibration with fuel consumption	113
5.2.2	Emissions in each region by type of vehicle	113
5.2.3	Emissions by age of use of vehicle	114
5.2.4	Emission maps	116
5.2.5	Emissions by hour of the week	117
<b>5.3</b>	<b>Atmospheric simulation</b>	<b>117</b>
5.3.1	Meteorological simulation	119
5.3.2	Simulation of CO, NO, NO <sub>2</sub> , SO <sub>2</sub> and O <sub>3</sub>	120
5.3.3	Evaluation of the WRF-Chem simulation in each region	129
<b>5.4</b>	<b>Discussion and uncertainty</b>	<b>129</b>
<b>6</b>	<b>CONCLUSIONS</b>	<b>137</b>
<b>6.1</b>	<b>Suggestions for future works</b>	<b>138</b>
<b>7</b>	<b>ACADEMIC LITERATURE PRODUCED DURING THE PH.D.</b>	<b>139</b>
<b>7.1</b>	<b>Publications with peer review process</b>	<b>139</b>
<b>7.2</b>	<b>Submitted publications under review process</b>	<b>139</b>
<b>7.3</b>	<b>Software</b>	<b>139</b>
<b>7.4</b>	<b>Conferences</b>	<b>140</b>
<b>7.5</b>	<b>Other events</b>	<b>140</b>
	<b>BIBLIOGRAPHY</b>	<b>143</b>



<b>APPENDIX A – EVALUATION OF TESTS OF WILCOXON FOR SCE- NARIOS OF SIMULATION . . . . .</b>	<b>155</b>
---	------------



# 1 Introduction

Air pollution is a global threat to the health of population, causing 3% of cardiovascular mortality and 5% of several types of cancer (COHEN et al., 2005). The levels of air pollution grew rapidly with the industrial revolution in 19th century but it is during the 20th century that the pollution reached the category of global threat for human health by exposition and for the deleterious effect on ecosystemic services. One consequence of this issue is the Antarctic ozone hole (MOLINA, 1996), but perhaps, the main threat of population to the biosphere is the climate change. There is consensus in the scientific community that climate change is due to human activity (COOK et al., 2016). Anthropogenic emissions of greenhouse gases are causing global warming producing a climate change with significant impacts on ecosystems (STOCKER, 2014). Therefore, in order to glimpse strategies to diminish air pollution, fighting back against global warming and improving population health, it is crucial to better understand the physical and chemical mechanism of generation of the air pollutants, the emissions.

One important tool for air quality management is the emissions inventory. Emissions inventory is the compilation of all the pollutants released by activities for a defined area and time-lapse (PULLES; HESLINGA, 2010).

Inventories can have two main applications: Scientific or Policy. When focused on policy making, emissions inventories are used to try to identify which sources contribute most to selected pollutants. The Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories, for example, provide guidance to the process of developing inventories for these gases in order to reduce their emissions. Another example is the Convention on Long-Range Transboundary Air Pollution (CLRTAP<sup>1</sup>) signed by 51 member States who must report their emissions inventory annually. In a smaller scale, in the USA, areas of the country that violate air quality standards<sup>2</sup> must develop a plan to reduce the levels of air pollution. The plan involves an emissions inventory to identify key sources that are responsible for most of the pollution. An example of scientific emissions inventory is the Emissions Database for Global Atmospheric Research (EDGAR) (OLIVIER et al., 1994), which is a worldwide emissions inventory of greenhouse gases and ozone-depleting substances with a spatial resolution of 1 degree. This model has been updated improving the spatial and temporal coverage, and also, to include more pollutants. Policy emissions inventories seek agreement and scientific inventories question the methods and develop new ones.

---

<sup>1</sup> <http://www.unece.org/env/lrtap/welcome.html>

<sup>2</sup> <https://www.epa.gov/green-book>

The elaboration of emissions inventories is a complex task where the **approach** used plays an important role. There are basically two approaches, **bottom-up** and **top-down**. Top-Down approach uses national statistics and average emissions ratios, meaning that they are faster to be produced but present lacks of detail, as well as lower temporal and spatial definition. Bottom-up emissions inventories require a larger and more detailed amount of information, with a higher temporal and spatial resolution.

Atmospheric pollutants are released by natural and anthropogenic sources. Biogenic emissions have no human intervention, for example, volcanic ashes, lightning, desert dust or oceanic aerosols. For anthropogenic activities, US-EPA identifies 8 main sources<sup>3</sup>: Agriculture, Dust, Fires, Fuel Combustion, Industrial Processes, Miscellaneous, Mobile and Solvent.

The importance of the source depends on the air pollution problem in the study area. In general, emissions inventories studies are made in zones where air pollutant concentrations exceed air quality standards. Therefore, the most important sources of a study area are the ones that are causing the exceeding of air quality standards. Hence, the knowledge of the atmospheric chemistry and meteorology sciences is crucial for the understanding of air pollution problem, which is inherently inter-disciplinary. For instance, residential wood combustion is the main source for  $PM_{2.5}$  concentrations at Temuco, a city in south of Chile (SANHUEZA et al., 2009). Photochemical smog is a persistent problem in mega-cities - cities with more than 10 million of habitants - where, on the other hand, vehicles are the main air pollutants source, specially for  $CO$ ,  $NO_x$  and VOC (MOLINA; MOLINA, 2004a).

In the state of São Paulo, Brazil, the pollutants  $O_3$  and  $PM_{10}$  frequently present violation of their air quality standards. The official emissions inventory was developed by the Environmental Agency of São Paulo (CETESB) following a top-down approach. The 2013 CETESB Air Quality Report (CETESB, 2013a) shows that vehicles in the Metropolitan Area of São Paulo were responsible for 96% of the Carbon Monoxide (CO), 81% of the Hydrocarbons (HC), 80% of the Oxides of Nitrogen ( $NO_x$ ), 31% of the Inhalable Particulate Matter ( $PM_{10}$ ) and 47% of the Oxides of Sulfur ( $SO_x$ ) emissions of 2012. Hence, vehicles are the most important source in São Paulo.

Another important tool for policy decision is air quality models. They have been used not only to study the impact of different emissions scenarios for policy making but also to understand the dynamics of air pollutants in various parts of the world. MASP, for example, was the area studied with different air quality models: CIT photochemical model (Ulke Andrade, 2001; Vivanco Andrade, 2006), CMAQ (Albuquerque et al., 2007), SPM-BRAMS (ANDRADE et al., 2015) and WRF-Chem (VARA-VELA et al., 2016). There were many interesting results from these air quality simulations. It was shown

<sup>3</sup> Emission Data Source: <https://www.epa.gov/air-emissions-inventories/air-emissions-sources>

that vehicles were responsible for 40% of  $PM_{2,5}$  (ANDRADE et al., 2012) and for the formation of 20-30% of secondary aerosols (VARA-VELA et al., 2016). MASP emissions also cause high  $O_3$  concentrations which is transported to the nearby metropolitan regions of Campinas, Santos, São Jose dos Campos and Sorocaba (BOIAN; ANDRADE, 2012). This work also questioned the values of the official emissions inventory, making adjustments to HC (an increase of 20%) and  $NO_X$  (divided by half) emissions, following the suggestion given in VIVANCO; ANDRADE (2006).

Although air quality models are easily available, emissions - that must be used as input for these models - are not. One of the biggest problem is lack of information regarding vehicular emissions. All the top-down emissions inventories cited above inform the amount of pollutant released in one year for a relative large area, for example, for a whole country (National Emissions Inventories) or a Metropolitan Area (CETESB Emissions Inventory). Air quality models require gridded hourly emissions data. Therefore, spatial disaggregation was performed based on land use (VIVANCO; ANDRADE, 2006; ULKE; ANDRADE, 2001), night satellite images (ANDRADE et al., 2015), international databases (VARA-VELA, 2013), street density (EICKER et al., 2008), density of a combination of street length and category of street density (ANDRADE et al., 2015) (VARA-VELA et al., 2016), population density (WILSON et al., 2006) and correlations of emissions, traffic counts and land use maps (SAIDE et al., 2009). For temporal distribution, usually an average profile was assumed (ANDRADE et al., 2015). These approaches, although cumbersome, have the advantages of being relatively fast, cheap and not so complicated, but the main disadvantage is the lack of representativity of the dynamic of traffic conditions, and misrepresentation of the emissions for particular regions (NTZIACHRISTOS; SAMARAS, 2016).

Therefore, there is a need for an emissions inventory model for scientific applications, with high level of detail and high temporal and spatial resolution. This present doctoral thesis pursues to provide a scientific and bottom-up vehicular emissions inventory model in order to clarify air quality questions. The study area comprehends the metropolitan areas of São Paulo, Santos, São José Dos Campos, Campinas and Sorocaba.

## 1.1 Objectives

The main object of this study was to develop a bottom-up vehicular emissions inventory model applicable for São Paulo metropolitan areas and evaluate its emissions impact on air pollution.

### 1.1.1 Specific Objectives

In order to achieve the main objective, the following actions were taken:

- Review, identify and elaborate activity traffic data.
- Identify and elaborate vehicular emission factors for Brazilian reality.
- Construct a vehicular emissions model that integrates different sources of traffic data, emission factors and produces gridded emissions for air quality models.
- Elaborate a high resolution emissions inventory for the main metropolitan regions of São Paulo.
- Simulate air pollutant concentrations with an air quality model and evaluate the impact of the emissions

This thesis is organized as:

- Chapter 2: Traffic data. In this chapter, all the traffic data (vehicular types, road network, traffic simulations, traffic counts, GPS tracking data) are presented. Since traffic counts must be interpolated in order to be useful for air quality modeling purposes, the interpolation methodology is also described.
- Chapter 3: VEIN v0.2.2: an R package for bottom-up Vehicular Emissions Inventories. In this chapter, the development of the Vehicular Emissions Inventories Model (VEIN) is described. Initially, the usage of the R-package is explained using only Light Duty Vehicles and CO emissions for MASP. Then, VEIN model is applied to the different traffic activity data presented in chapter 2. A more detailed description is given for the classification of types of vehicles, calculation of emission factors, speciation of NMHC, and traffic calibration considering fuel consumption.
- Chapter 4: Weather and Research Forecast Model with Chemistry (WRF-Chem). In this chapter, the configuration of the WRF-Chem model, the meteorological conditions of the period of simulation and the evaluation indexes are presented.
- Chapter 5: Results. In this chapter, all the results from the previous chapters are shown and discussed.
- Chapter 6: Conclusions. In this chapter, the main conclusions of the work are given.

## 2 Traffic data

A key element to develop bottom-up vehicular emissions inventories is the availability of traffic data. Some studies use macro (CORVALÁN; OSSES; URRUTIA, 2002) and micro (PANIS; BROEKX; LIU, 2006) traffic simulations to estimate vehicular emissions. Others use traffic counts to allocate or produce traffic flow to later estimate emissions (KINNEE et al., 2004). New technologies such as GPS incorporated into smart-phones, vehicles with assurance policy with GPS and other fleet management services are generating new data available to the public. Companies such as Google (<https://www.google.com.br/maps>) and Maplink (<http://transito.maplink.global/>) in Brazil use this data to provide real-time traffic conditions at any time. Accessing this data to generate inputs for vehicular emissions could enhance the inventories and provide better input for air quality simulation and health studies.

In this chapter it will be described the data and the procedures developed to generate the traffic flow input for the bottom-up vehicular emissions inventory model (VEIN model) (IBARRA-ESPINOSA et al., 2017).

### 2.1 Types of vehicles

*Vehicular composition* is the characterization of the fleet into its components and it is one the first steps when developing a vehicular emissions inventory. CORVALÁN; OSSES; URRUTIA (2002) defined vehicular composition characterizing the fleet by the type of use, type of fuel and gross weight. In this work the following categories, partly based on the classification presented in the vehicular emissions inventory for São Paulo (CETESB, 2015), were adopted: Light Duty Vehicles (LDV = Passenger Cars (PC) + Light Commercial Vehicles (LCV)); Heavy Duty Vehicles (HDV composed by Heavy Good Vehicles (HGV) or trucks + Buses); and also Motorcycles (MC). Type of fuel consist of: Gasohol as a mixture of gasoline and 25 % of ethanol- E25; hydrated ethanol 100% - E100; and diesel with 5 % of biodiesel - B5. In 2003, car manufacturers started producing vehicles capable of running with E25, E100 or any mixture in between with the development of the flex engine (GIROLDO et al., 2005). FE25 is a vehicle with flex engine using E25 and FE100, when consuming E100. Table (1) shows the vehicle classification for this study. This table was defined to match vehicular categories from CETESB and the European Emission Guidelines (NTZIACHRISTOS; SAMARAS, 2016). The table also included size of engines, gross weight, and the emission standards for each type of vehicle (Ministerio de Medio Ambiente, 2011; CONAMA, 1986; MACEDO et al., 2011). These informations as well as the emission factors used in this work will

Table 1 – Classification of vehicles

Vehicle	Fuel	Size/Gross weight	Standard
Passenger Car (PC)	E25	<1400cc	PRE L1
	FE25		L1
	FE100	1400-2000cc	L2
	E100	>2000cc	L3
			L4
		L5	
Light Commercial Vehicles (LCV)	E25	<=3.5t	PRE L1
	FE25		L1
	FE100		L2
	E100		L3
			B5
		L5	
Light Truck (LT)	B5	<7.5t	PRE P1
Semi Light Truck (LT)		12-14t	P1
Medium Truck (MT)		14-20t	P2
Semi Heavy Truck (SHT)		20-26t	P3
Heavy Trucks		>32t	P4
			P5
			P7
Small Urban Bus (SUB)	B5	<=15t	PRE P1
Urban Bus (UB)			P1
Urban Bus Articulated (UBA)		15-18t	P2
Motorway Bus (Coach)		>=18t	P3
			P4
		P5	
			P7
Motorcycle (MC)	E25	<=150cc	PRE M1
	FE25	150-500cc	M1
	FE100	>=500cc	M2
			M3

be treated in more detail in Chapter 3.

## 2.2 Road network

The bottom-up vehicular emissions inventory model developed during this work, requires traffic data as an hourly amount of vehicles per street. The following sections describe the study area and the map of streets adopted for the emissions model.



### 2.2.1 Study area

The study area is delimited by the coordinates -47.85848 to -45.48376 degrees of longitude and -24.488 to -22.46691 degrees of latitude. The area is located at Southeast Brazil and covers the metropolitan areas of São Paulo (MASP), Santos, São José dos Campos (SJDC), Sorocaba and Campinas. The area of study is shown in Fig. (1). The political delimitations of each metropolitan area are smaller than the regions presented, with exception of MASP. In other words, it was added more municipalities into each metropolitan area because data was available for the whole area, and not only each geopolitically defined metropolitan area.

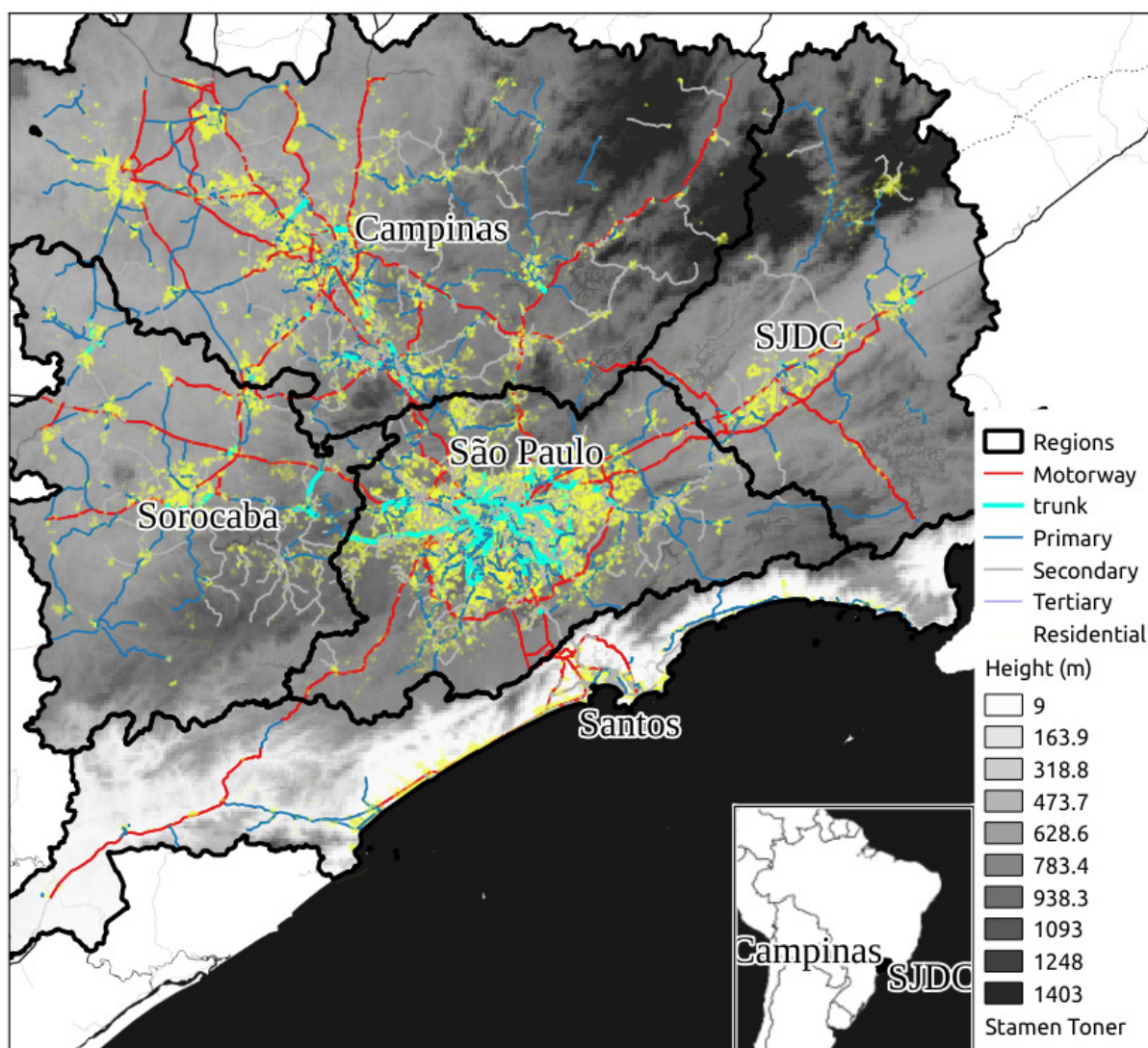


Figure 1 – Area of study: Metropolitan Areas (delimited by black lines), Topography (gray shades) and road network (colored lines)

## 2.2.2 Road network

OpenStreetMap (OpenStreetMap contributors, 2017) is an open mapping project. Data is freely available and covers many parts of the world where geographical information is either nonexistent or very expensive. Fig. (1) also shows the types of street from OpenStreetMap. There are many road categorization covering not only streets for vehicles displacement, but also rails, pedestrian roads, among other. In the case of streets, there is a discussion<sup>1</sup> at OpenStreetMap regarding the classification of streets with Highway Functional Classification System, proposing this equivalence: Interstates & other freeways equivalent to motorway, Expressways equivalent to trunk or primary, Other Principal Arterials equivalent to secondary, Minor Arterial Streets equivalent to tertiary and local streets equivalent to residential.

The data was downloaded with the QGIS software (QGIS Development Team, 2017) and it filtered only streets field identified as 'highway' with the following attributes: *motorway*, *motorway\_link*, *trunk*, *trunk\_link*, *primary*, *primary\_link*, *secondary*, *secondary\_link*, *tertiary*, *tertiary\_link* and *residential*. In Brazil, the characteristics of the type of road are: Motorway is a paved express highway, with speed  $\geq 80 \text{ km} \cdot \text{h}^{-1}$ , without crossings or obstructions, being a major street, normally with 2 or more running lanes. Trunk is similar to motorway with the exception that it includes obstructions such as traffic lights, being national or urban expressways with speed limit 70 - 80  $\text{km} \cdot \text{h}^{-1}$ . Primary roads are paved with speed  $\geq 60 \text{ km} \cdot \text{h}^{-1}$  connecting many neighborhoods and designed as urban arterial roads. Secondary roads collect traffic, speed  $\geq 40 \text{ km} \cdot \text{h}^{-1}$  and circulation between neighborhoods. Tertiary roads are similar to Secondary, connecting fewer neighborhoods. The last category is Residential which is a local road with few vehicles located in a residential areas. Streets terminated in '\_link' are junctions between these streets and other categories. The road network of Fig. (1) shows how the metropolitan areas are connected by motorways and trunk roads. It is also possible to see the other roads especially the numerous residential ones (yellow lines in the map).

## 2.3 Traffic activity

Traffic data is a very important source of information for urban planning. Urban planners use this information to make decisions about projections, costs, quality of life and air pollution. However, this information is not always available to researchers, consultants and civil servants. In addition, this data can be particularly difficult to obtain in developing countries.

Traffic data can be developed generally in two ways, either by a multi-step demand estimation or a direct demand estimation (LOWRY, 2014). Multi-step demand

<sup>1</sup> [https://wiki.openstreetmap.org/wiki/Talk:Highway\\_Functional\\_Classification\\_System](https://wiki.openstreetmap.org/wiki/Talk:Highway_Functional_Classification_System)

estimation refers to a 4-stage traffic simulation. This data is generally available for morning rush hour, however, it requires a burdensome and costly amount of input data (ORTUZAR; WILLUMSEN et al., 1994). Direct demand estimation consists of collecting and extrapolating traffic data based on characteristics of roads and the adjacent areas.

For this work, traffic activity data were obtained from different sources. Traffic simulations were obtained from Traffic Engineering Company (CET <http://www.cetsp.com.br/>) and from SPtrans (<http://www.sptrans.com.br/>), and since they could be directly inserted in the VEIN model, they are just described in the next subsections. Another source was vehicle counts from CET and GPS data and the procedures made to transform this information into traffic simulations are detailed in subsections 2.3.3 and 2.3.4, respectively.

### 2.3.1 Traffic simulation from CET

The Traffic Engineering Company of São Paulo (CET) is the agency that performs the transport planning in São Paulo city. They have a specific sector dedicated to 4-stage traffic simulations using data from an Origin-Destination-Survey (ODS) (METRO, 2017) which started in the decade of 1950 in São Paulo. A classic reference of a 4-stage modeling transport is (ORTUZAR; WILLUMSEN, 2002). The 4 stages of the traffic modeling includes characterization of trip attractions and productions by zone in some regions, distribution of these trips, preferred mode of transport for traveling and finally the allocation of the trips at each mode, in this case into the road network. The ODS is made every 10 years by Metro (<http://www.metro.sp.gov.br/>), which is the underground company, and they perform a smaller update of ODS after 5 years. The information gathered in the ODS is massive with the participation of thousands of commuters. It helps to identify characteristics of the trips inside MASP. CET uses the information from ODS and performs the traffic simulation. In this case, it is a macro or strategic traffic simulation which represents the equilibrium between offer and supply of transportation at maximum load of the road network, that is, at the rush hour, which is from 08:00 to 09:00 Local Time (LT).

The CET traffic simulation provides initial and final id of node, volume of LDV ( $veh \cdot h^{-1}$ ), HGV (trucks,  $veh \cdot h^{-1}$ ), number of lanes, capacity ( $veh \cdot h^{-1}$ ), free flow speed (FFS,  $km \cdot h^{-1}$ ), Peak Speed (PS,  $km \cdot h^{-1}$ ), type of street and travel-time for each link.

The traffic simulations for LDV and HGV are shown on Fig. (2) and (3) respectively. LDV traffic is more concentrated in urban motorways located near the center of the city. The color scale shows that there are streets where LDV traffic overpasses more than 15000  $veh \cdot h^{-1}$ . At this simulated rush hour, the total volume of LDV is 24708767, distributed through 34733 streets, resulting on an average of 711  $veh \cdot h^{-1} \cdot street^{-1}$ . HGV is concentrated in motorways around the city where the volume can overpass

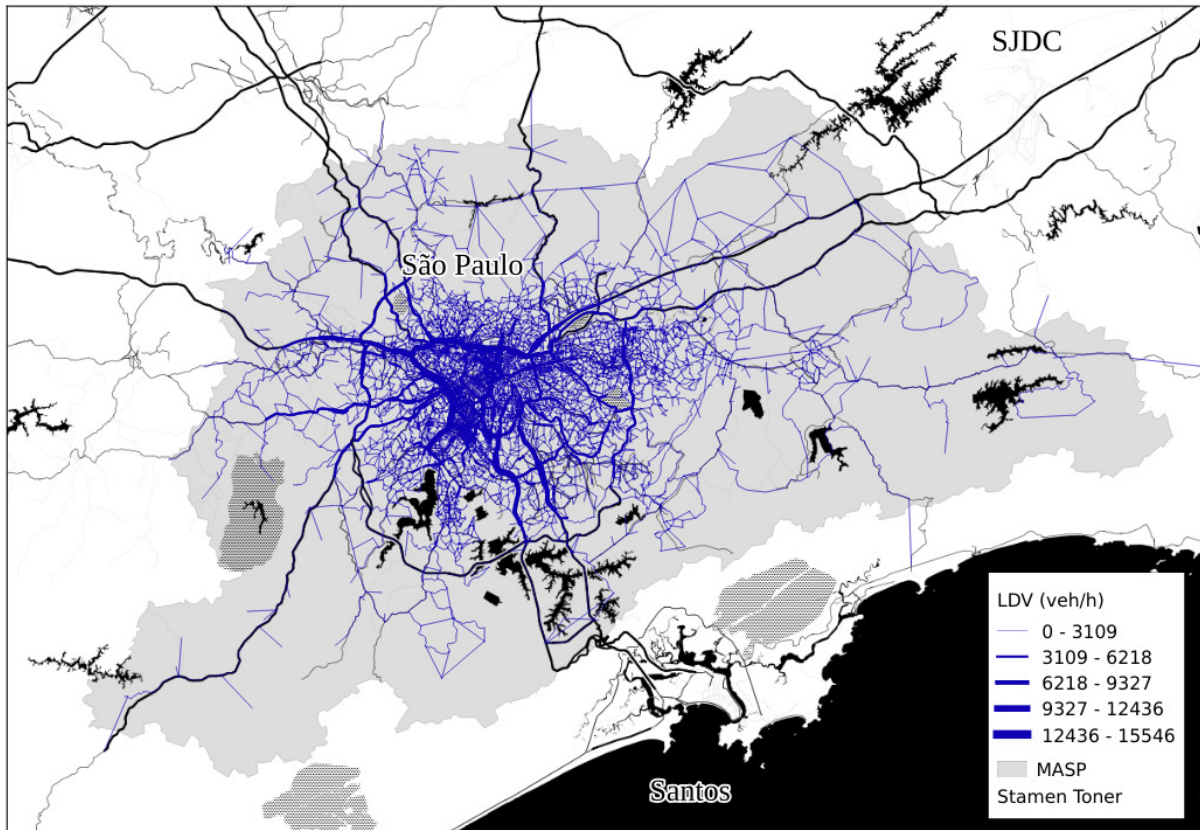


Figure 2 – CET Traffic simulation of LDV from 08:00 to 09:00 LT.

$7000 \text{ veh} \cdot \text{h}^{-1}$ . The total amount of HGV is 2809297 and the average number of HGV per street is 80.

### 2.3.2 Traffic simulation from SPTRANS

The Secretary of Transport and Mobility of São Paulo (SPTrans) is the agency responsible for public transportation by buses, and also uses ODS (METRO, 2017) to make traffic simulation for the rush hour. SPTrans traffic simulation is also a spatial network with the following information: initial and final id of node, LDV ( $\text{veh} \cdot \text{h}^{-1}$ ), capacity ( $\text{veh} \cdot \text{h}^{-1}$ ), volume of passengers ( $\text{passengers} \cdot \text{h}^{-1}$ ), volume of buses ( $\text{veh} \cdot \text{h}^{-1}$ ) and Peak Speed (PS,  $\text{km} \cdot \text{h}^{-1}$ ). it was used only volume of buses, since the other information is already covered by CET simulation. Fig. (4) shows that buses are also more concentrated near the center of the city. The volume of buses can overpass  $5000 \text{ veh} \cdot \text{h}^{-1}$ , the total amount of buses is 763097 and the average number of buses per street is 23.

### 2.3.3 Traffic counts

Traffic data can be collected with automatic traffic detectors or manually. When the traffic count is manual, it is usually only for most congested routes over the city at



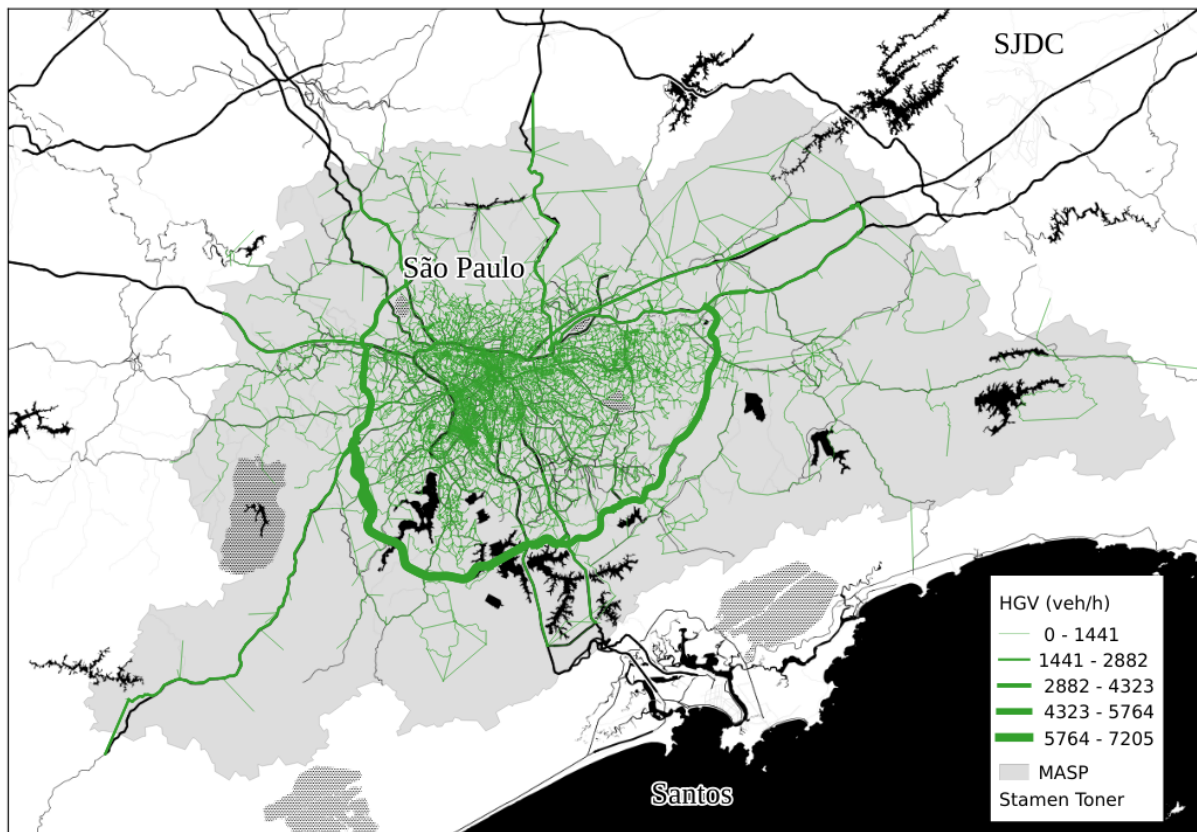


Figure 3 – CET Traffic simulation of HGV from 08:00 to 09:00 LT.

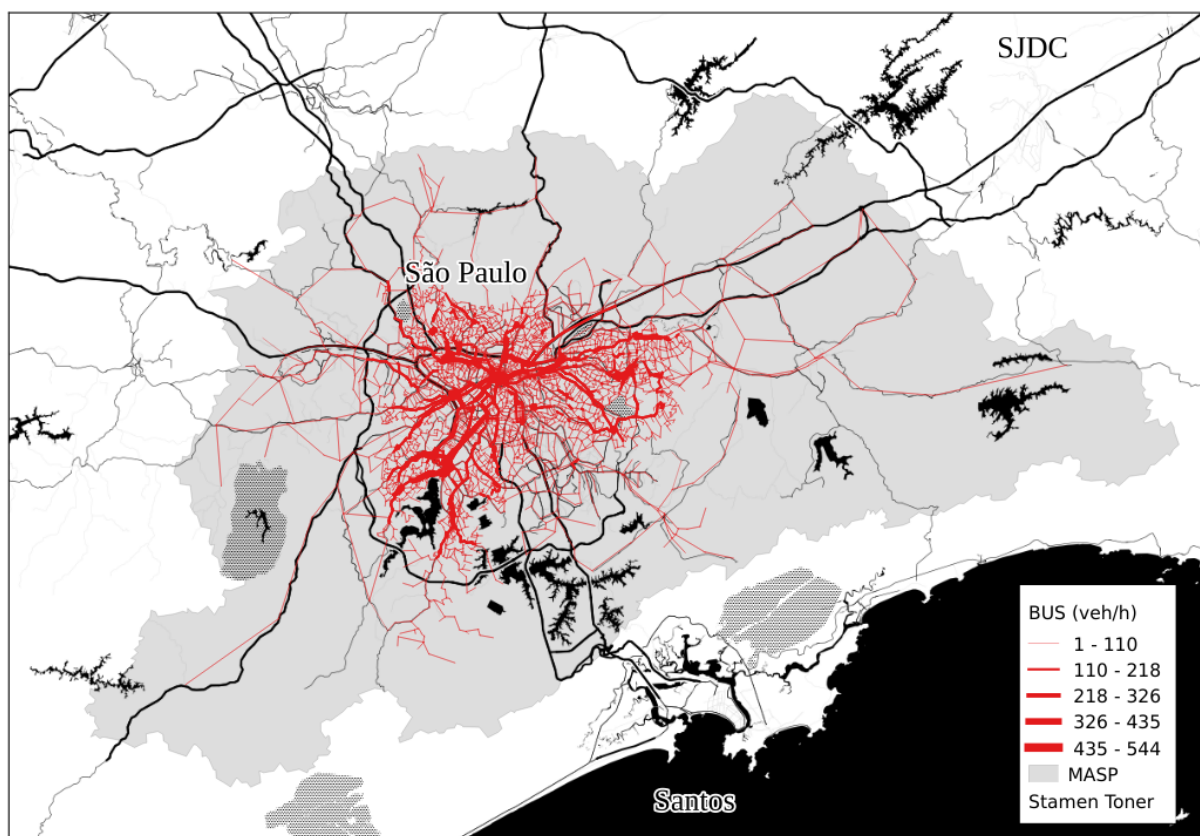


Figure 4 – SPTrans simulation of buses from 08:00 to 09:00 LT.

morning and evening rush hours. When the traffic is counted over 72 hours or less, it is known as Short Period Traffic Counts (SPTC) and this data is usually employed to calculate Annual Average Daily Traffic (AADT). AADT plays a key role in this study and it is a very effective indicator for urban planning. The methods used to obtain AADT are distinguished between ordinary least square, neural networks and Kriging regression. For instance, ZHAO; CHUNG (2001) tested several multiple stepwise regressions incorporating land-use characteristics. They considered several variables: number of lanes, classification of road type, employment numbers and access to expressway. Their results had correlations between 0.66 and 0.82. LAM; XU (2000) compared neural networks and regressions for a dataset of 13 locations obtaining better results with a neural network approach. Kriging methods have also been used in AADT interpolation. EOM et al. (2006) predicted AADT for non-motorway streets. WANG; KOCKELMAN (2009) also predicted AADT but applied it where no traffic detectors were present. SELBY; KOCKELMAN (2013) compared Kriging and geographically weighted regressions in the prediction of AADT. Road characteristics and land use can show little variation in road networks of small communities. LOWRY (2014) presented a new method for predicting AADT based on the concept of origin-destination centrality. The idea is to obtain predictor variables directly from the road network. He identified origin and destination zones and added multiplication factors based on the land-use characteristics.

In general, most studies are interested in predicting total vehicular count and not the type of vehicle. Light duty vehicles have different rush hours than heavy duty vehicles. Moreover, regressions and interpolations use factors that expand traffic to obtain AADT and no results are shown for hourly interpolation. Furthermore, traffic regressions are assuming normal distribution, and literature shows that when working with count data, this assumption is not appropriate (ZEILEIS; KLEIBER; JACKMAN, 2008). Even more, it has been shown that Poisson distribution should be used with count data only when variance is equal to mean, but this assumption is highly restrictive, especially in environmental sciences (HOEF; BOVENG, 2007). This section presents stepwise regressions with quasi-Poisson and negative binomial distributions to interpolate hourly traffic data by type of vehicle. This involved the use of a traffic counts dataset for the Metropolitan Area of São Paulo (MASP). The dataset included LDV, Trucks, Buses (UB and private urban buses, RB) and MC. Road network data from OpenStreetMap was also used.

### 2.3.3.1 Short Period Traffic Count (SPTC)

CET performs SPTC for São Paulo City. The reports are available on-line as pdf documents (CET, 2013). This study used the report from year 2012 as a base year. CET identifies 31 main traffic routes and then traffic is counted manually in several points of each route, in each way. Traffic is counted in street segments and each segment is

explicitly identified in the CET reports. This allowed geo-referencing each point over a OpenStreetMap road network with the software QGIS (QGIS Development Team, 2017). The traffic is counted hourly during morning and evening rush hours between 07:00 - 10:00 and 17:00 - 18:00. All the mentioned times are in the Local Time (LT). This is performed in order to represent usual conditions of workdays. Most of the points are counted once per year, but the point with highest volume of traffic in each route is counted twice per year and only the average value is shown in the CET reports.

The traffic count dataset was aggregated to have total vehicles in each point, in both ways. Traffic counts distinguish the following type of vehicles: Light Duty Vehicles (LDV) which can be Passenger Cars (PC) or Light Commercial Vehicles (LCV), Trucks with two axes (T2A or LT in Table (1)), Trucks with three axes (T3A or MT), Trucks with four axes (T4A or SHT), Urban buses (UB), Rented buses (RB) and Motorcycles (MC). The points are shown in Figure 5, where it is shown higher amount of LDV towards the center of the city. This Figure also shows the OpenStreetMap road network. In the case of T2A, initially there were more vehicles outside the city showing spatial bias towards southeast. This issue was corrected by adding traffic counts from toll stations located near São Paulo City, as seen on the left of Figure 5(b). The toll database comes from the Transport Agency of São Paulo State (Artesp, <http://www.artesp.sp.gov.br/>) and consists of 6 traffic points located at motorways near São Paulo City. The toll station points are located on the motorways SP280-KM180 (Castelo Branco), SP021-KM14 (Castelo Exter), SP021-KM24 (Raposo Exter), SP021-KM24700 (Raposo Inter), SP021-KM25 (Regis) and SP021-KM14 (Castelo Inter). A decision was made to append this Artesp dataset only for trucks and not for other types of vehicles. This decision was made because light vehicles and urban buses circulate mainly in the urban center, while load transport travel to other cities using mainly routes such as motorways. The resulting database consisted of 107 counting points from CET for LDV, UB, RB, and MC; and 107 from CET plus 6 from Artesp for T2A, T3A and T4A, resulting in 113 traffic counts for trucks. Concerning the road network, there were 48 traffic counting points over Trunk roads, 43 over Primary roads, 14 in Secondary roads, 2 in Tertiary and 6 in Motorways (only for T2A).

### 2.3.3.2 Regression approach

After analyzing the traffic count data, the next step is to use a regression for each type of vehicle and each hour. As previously mentioned, similar studies have focused on estimating AADT, however, this study is novel since it proposes an hourly approach. Before making any assumption about traffic count distribution, it was tested the normality of each type of vehicle at each hour. It was used the test of SHAPIRO; WILK (1965) to check normality with the software R (R Core Team, 2017) and the command `shapiro.test` (ROYSTON, 1982).

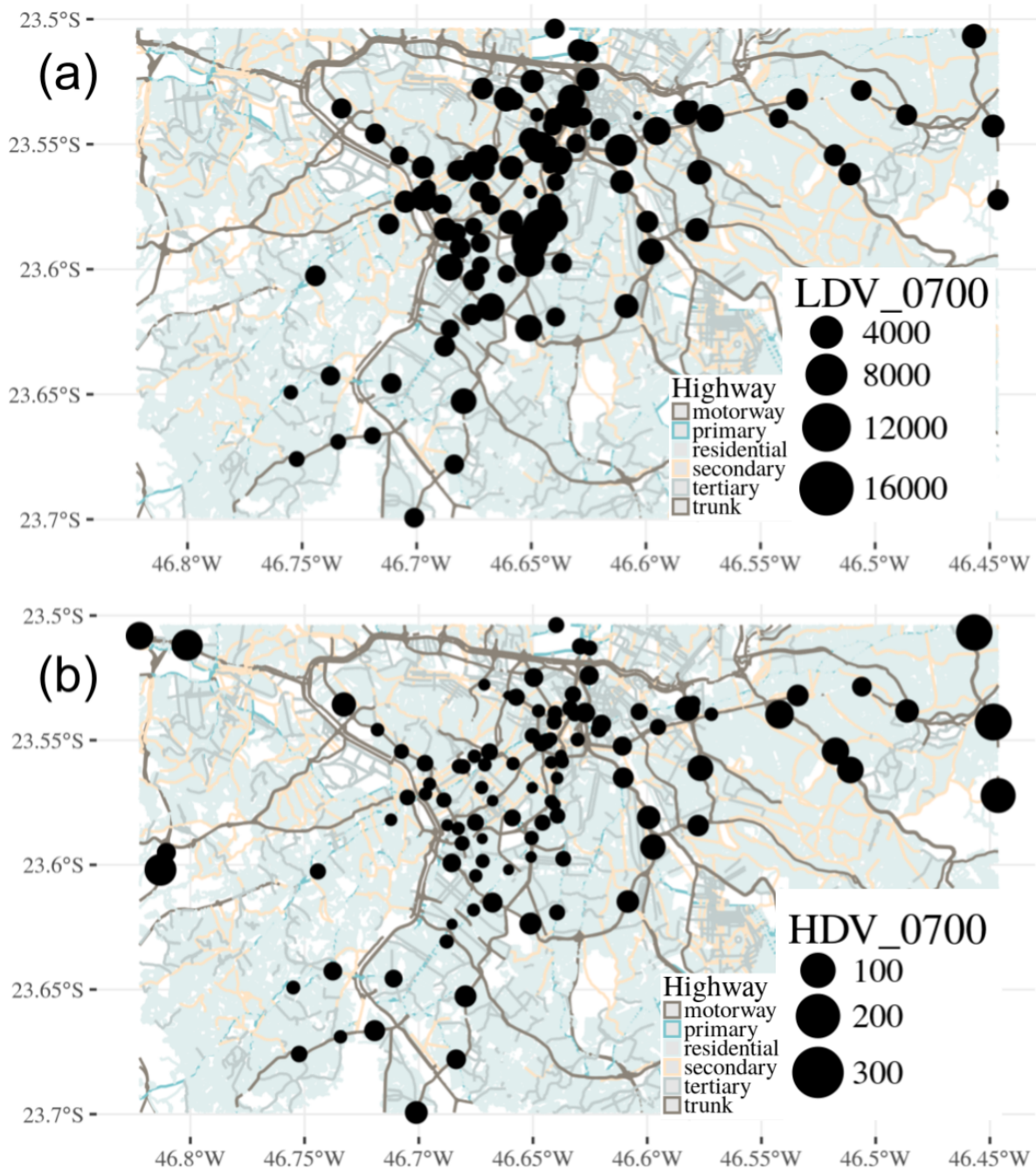


Figure 5 – Traffic counts for (a) LDV and (b) T2A, between 07:00 and 08:00 LT, 2012.



It is important to consider some checking when working with count dependent variables. Poisson regression is the first candidate to fit count data (GROUP., 2017). However, there are some assumptions that must be considered: variance cannot be higher than mean as it produces over-dispersion, and variables should not contain zeros (ZEILEIS; KLEIBER; JACKMAN, 2008).

In this study, the dataset of traffic counts was tested to check for over-dispersion. Table 10 showed that the standard deviation values are identified near to the mean values for each type of vehicle, meaning that the variance is far greater than the mean. Therefore, the traffic count dataset presents over-dispersion, and the Poisson regression was not used.

After analyzing the data, it was decided to fit and compare a negative binomial and a quasi-poisson regressions. This was performed using also the software R (R Core Team, 2017), with the library *MASS* for negative binomial regressions (VENABLES; RIPLEY, 2002) and the library *stats* for quasi-poisson regressions. ZEILEIS; KLEIBER; JACKMAN (2008) provide a good review of count models.

Regarding the independent variables, it was used OpenStreetMap (OSM) because it is open and free source of spatial information that allows to include several attributes related to traffic counts. The traffic counts were associated with the following type of roads: Trunk, Primary, Secondary and Tertiary. In order to use the regressions to estimate traffic in a city, all variables should be assigned to each link. The MASP road network does not count with number of lanes to each road. Therefore, it was necessary to calculate the average number of lanes per type of road. The mean values are: motorway = 3.2, motorway\_link = 1.9, trunk = 2.1, trunk\_link = 1.9, primary = 2.6, primary\_link = 1.6, secondary = 2.2, secondary\_link = 1.6, tertiary = 2.1, tertiary\_link = 1.5 and residential = 1.

Traffic counts shown in Figure 5a suggest a spatial relationship in which there are more LDV vehicles towards the center of the city. When comparing with the Figure 5b, T2A shows less vehicles towards the center of the city. Considering these observations, it is possible to establish a hypothesis that distance from center  $dfc$  is a predictor for traffic counts. This variable could also be expressed as geographical coordinates  $lat$  and  $lon$  or as the distance from a point representing the center of the city. The point of center was considered to be the location known as "Pátio do Colégio", which is the point where São Paulo City was founded in 1554. This point is located in the middle of the historic center, near to many government and private buildings, bus and subways stations and commerce. The coordinates are latitude = -23.548056 and longitude = -46.6325, WGS84.

Traffic count reports from CET identify the road segment in which traffic was counted. The road segment is limited by streets also identified in the report. This means

that traffic was counted by taking into consideration the traffic that circulated over a distance. It is reasonable to think that when distance for counting vehicles is longer, more vehicles are counted. Therefore, this variable was also added to the model.

Negative binomial and quasi-poisson distributions were tested and compared to count data. Finally, the models were configured in the following way:

$$\log(u) = \beta_0 + \beta_1 \cdot ts + \beta_2 \cdot lan + \beta_3 \cdot dfc + \beta_4 \cdot drs \quad (2.1)$$

where  $u$  is the traffic count,  $ts$  is the type of street,  $lan$  is the number of lanes,  $dfc$  is the distance from center, and  $drs$  is the length of the road segment. Since both the geographical coordinates and distance to the center are related, this could be a source of collinearity. Therefore, this study evaluated separately each model with the variable  $dfc$ , and another with geographical coordinate of longitude  $lon$  and latitude  $lat$ . For the case of values representing type of street  $ts$ , the study assigned the same value to the type of street with name '\_link'. For example, trunk and trunk\_link have the number 1. This could lead to an overestimation of vehicles in these types of streets. However, the amount of these link streets is relatively small so high overestimations are not expected.

The models selected were those with statistical significance  $\Pr(> |z|) < 0.05$ , i.e., small probability that the model coefficients were zero. The study also considered akaike information criterion (AIC) during the stepwise process (VENABLES; RIPLEY, 2002), as well as correlations with observed and predicted traffic counts.

### 2.3.3.3 Evaluation

The predictor capacity of the models was evaluated in terms of forecasting traffic over the road network for the Metropolitan Area of São Paulo (MASP), which is a megacity with more than 20 million habitants. Forecasting traffic is a complex process involving many variables that sometimes are difficult to obtain, mainly in developing countries, and the process involves a propagation of errors. Therefore, predicting traffic over this road network results in an approximation of the traffic that is actually in circulation.

To verify if the regression equations provided good results the fuel consumption resultant from this traffic interpolation was estimated and compared with annual fuel sales for the area where they were interpolated. This approach has been proposed by (NTZIACHRISTOS; SAMARAS, 2016) to compare vehicular emissions inventories produced by a bottom-up and top-down approach and is incorporated in the VEIN model, which is explained in the next chapter.

In order to run VEIN, it was necessary to match the vehicular categories from traffic counts and vehicular fleet, as described in (CETESB, 2013a). Therefore, LDV from traffic counts were divided into PC and LCV. T2A were also divided into five types

of trucks, as described in the CETESB report. In addition, the type of fuel and age distribution were also considered.

The MASP area has a zonal and meridional lengths of 120 km and 80 km, approximately. The assumption is that the automotive fuel sold in this area is consumed in the same area. The fuel sales for MASP were obtained from the Annuary of Statistics of Energy for all municipalities of São Paulo State (PAULO", 2013) for year 2012. The fuel sold for year 2012 was 4780017  $m^3$  of gasohol, 2198794  $m^3$  of ethanol and 3336678  $m^3$  of diesel. The density of each type of fuel is 0.754  $t \cdot m^3$  for gasohol, 0.809  $t \cdot m^3$  for ethanol and 0.852  $t \cdot m^3$  for diesel.

### 2.3.4 GPS tracking data

Nowadays many smart-phones and vehicles includes Global Position Systems (GPS) devices for many uses. This provides a massive source of data that can be used in environmental studies. GPS vehicle tracking data is the continuous registry of geographical position of vehicles including the time at each registry. GPS data was obtained from Maplink (<http://transito.maplink.global/>), one of the companies that provides data to GoogleMap. The GPS data was stored online and the data for the study area was extracted using the Google online tool Big Query (<https://cloud.google.com/bigquery/>). The data is composed of GPS positions of more than 103 million vehicles for the study area and includes the following information: "Vehicle" with an anonymous id for each vehicle; "Type" including Cars, Taxi, Trucks and Undefined; "Collect\_time" time with the format Year-month-day Hour-Minute-Seconds (eg: "2014-10-05 00:00:00") in UTC; "Lat" and "Long" indicating the geographical coordinates in WGS 84. The first registry was recorded at "2014-10-05 00:00:00" and the last one at "2014-10-11 00:00:00" covering 6 days, from Saturday October 4th at 21:00 of 2014 to Friday 10th at 21:00 in Local Time (LT), totalizing 145 hours.

#### 2.3.4.1 GPS data processing

The Laboratory of Geo-processing of USP provided 50 comma separated value (CSV) text files of raw GPS data obtained from Maplink. A quality control had to be developed to avoid errors. Each file contains information for the 145 hours. It was calculated the distance between two consecutive readings of each vehicle to calculate its speed and acceleration. The quality control excluded vehicles with speeds higher than 110  $km \cdot h^{-1}$  and acceleration higher than  $|129600| km \cdot h^{-2}$ . Also, the same vehicle must be tracked at least 5 times in each CSV file. The total number of observations was 104069448. As the data was too numerous only a slice of it is shown in Fig. (6), for LDV (223470 observations) and Fig. (7), for HGV (46207 observations). When comparing

both figures it can be seen that HGV is more spread than LDV. This reflects the behavior of HGV delivering goods and LDV concentrated inside each urban settlement.

Generating traffic flow for emissions inventories from massive GPS records is, to my knowledge, a novel approach and as such it requires caution. One important aspect is the frequency of the observations. Ideally, the frequency of two consecutive observations from the same vehicle would be one second. However, the average frequency of the records is 9155 s for Cars and 38345 s for Trucks. This means that, on average, there is 2.54 hours between two consecutive records of Cars and 10 hours for Trucks. Filtering the data to include only consecutive observations with higher frequency would imply in much less observations. Therefore, the resulting data must be an equilibrium between high frequency of consecutive observations and a minimum set of observations by each street. It was explored several frequencies: a maximum frequency of 5 minutes resulted on an average of 2.21 minutes by consecutive readings and 50 streets; a maximum frequency of 10 minutes resulted on an average of 4.27 minutes and approximately 2000 streets with records; finally, a maximum frequency of 29 minutes resulted on an average of 11.57 minutes and approximately 5000-6000 streets with observations per hour. Despite the relatively lower number of streets, it was decided to use an average frequency of 11 minutes. The percentiles for the Cars data were then P25: 240 seconds, P50: 600 s, P75: 1095 s and P100: 1799 s. For Trucks, the P25 was 3391 seconds; P50, 15609 s and P75, 58272.5 s. Most of the recordings were found in streets different than *residential*. One possible reason is that vehicles goes quickly from a residential street to another type and therefore, most of GPS records were registered in streets different from residential, which was therefore excluded from the filtered data.

The package **data.table** (DOWLE; SRINIVASAN, 2017) was used for the analyses since it performs fast analyses with few computational cost in R. It was calculated the average speed by type of vehicle, but since Taxis, Cars, Trucks and Undefined share a common space on the street network, it was used the information of all the vehicles, allowing to calculate speeds with a bigger number of observations. After that, it was excluded Undefined from the analyses because there was no way to know which type of vehicle it was. It was also excluded Taxis because there were too many Taxi records, which does not mean that there are actually more of these type of vehicles in circulation but could be due to proliferation of related applications for cellphones with GPS included. In other words, there are more Taxis records because nowadays most of Taxis counts with GPS records, however, there are more Cars on the road without GPS devices.

Traffic data consists in points and they must be assigned into the road network. the traffic data was first converted to spatial features with geometry points and then projected to the system of coordinates 31983 SIRGAS 2000 / UTM zone 23 South using

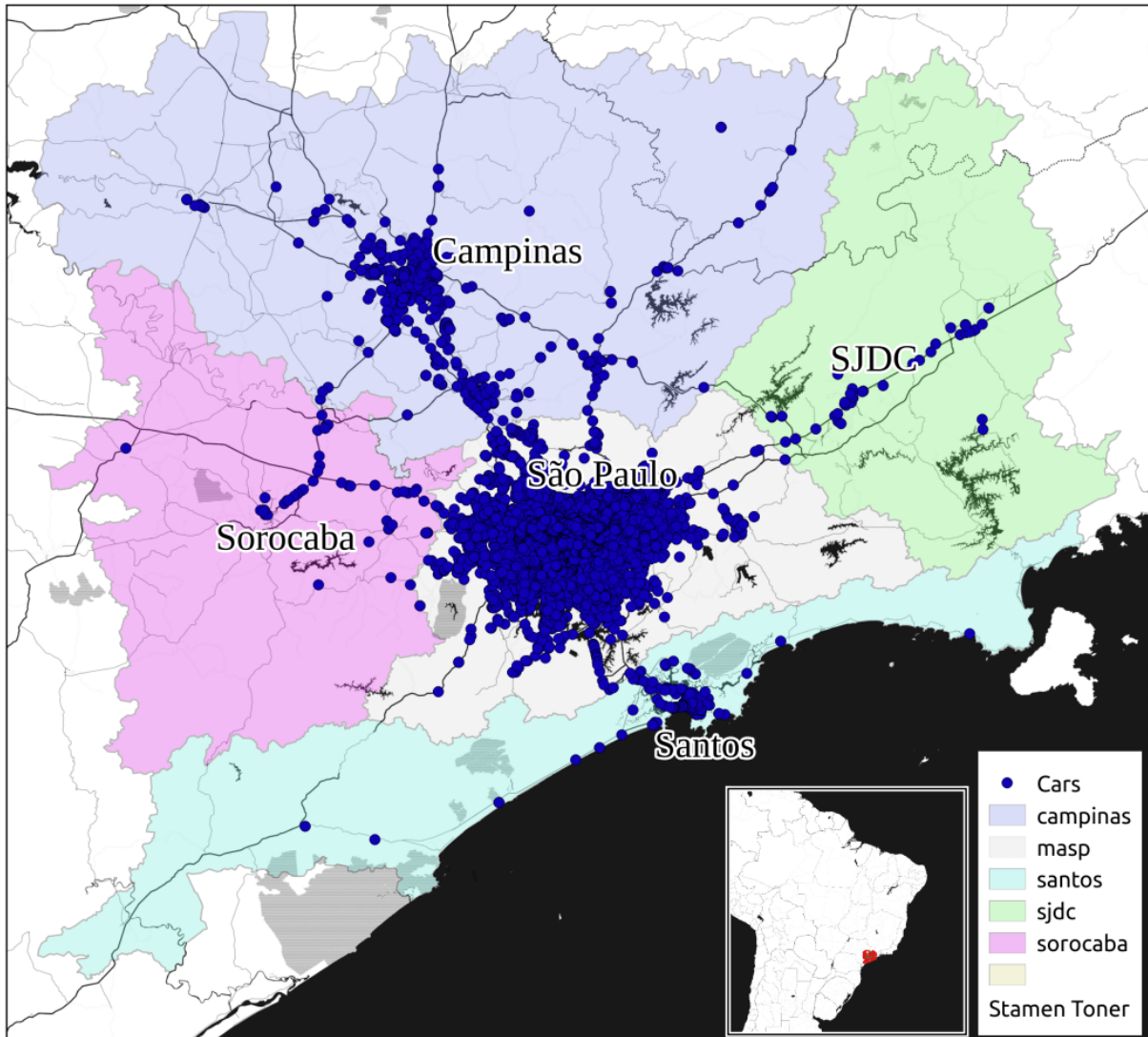


Figure 6 – Cars positions derived from GPS recordings on Monday 2014-10-06 09:00 LT.

the R package **sf** (PEBESMA, 2016). In order to avoid the use of all RAM the data was divided by hour and then created 10 m buffers to have one polygon for each traffic recording position. Then an intersection between the polygons and road network was made. Finally, the total number of vehicles and the average and quantiles by road at each hour were aggregated. To ensure that the calculation of the speed was correct it was calculated the average speed by type of street for Monday Midnight. The resulting average speeds by type of street for Cars, Taxi, Trucks were low, as shown in table (2). This table also shows the maximum Taxi, Trucks and All speeds. It can be seen that the average speed does not change abruptly among the hierarchy of the street, except for *motorways*. Ideally, in order to represent Cars only the recordings of cars should be considered, which shows an average speed of  $74 \text{ km} \cdot \text{h}^{-1}$  at *motorways* and lower speeds at other type of streets. However, the number of streets with recordings was only 64 and the total network have 56966 streets. Therefore, it was used the speed

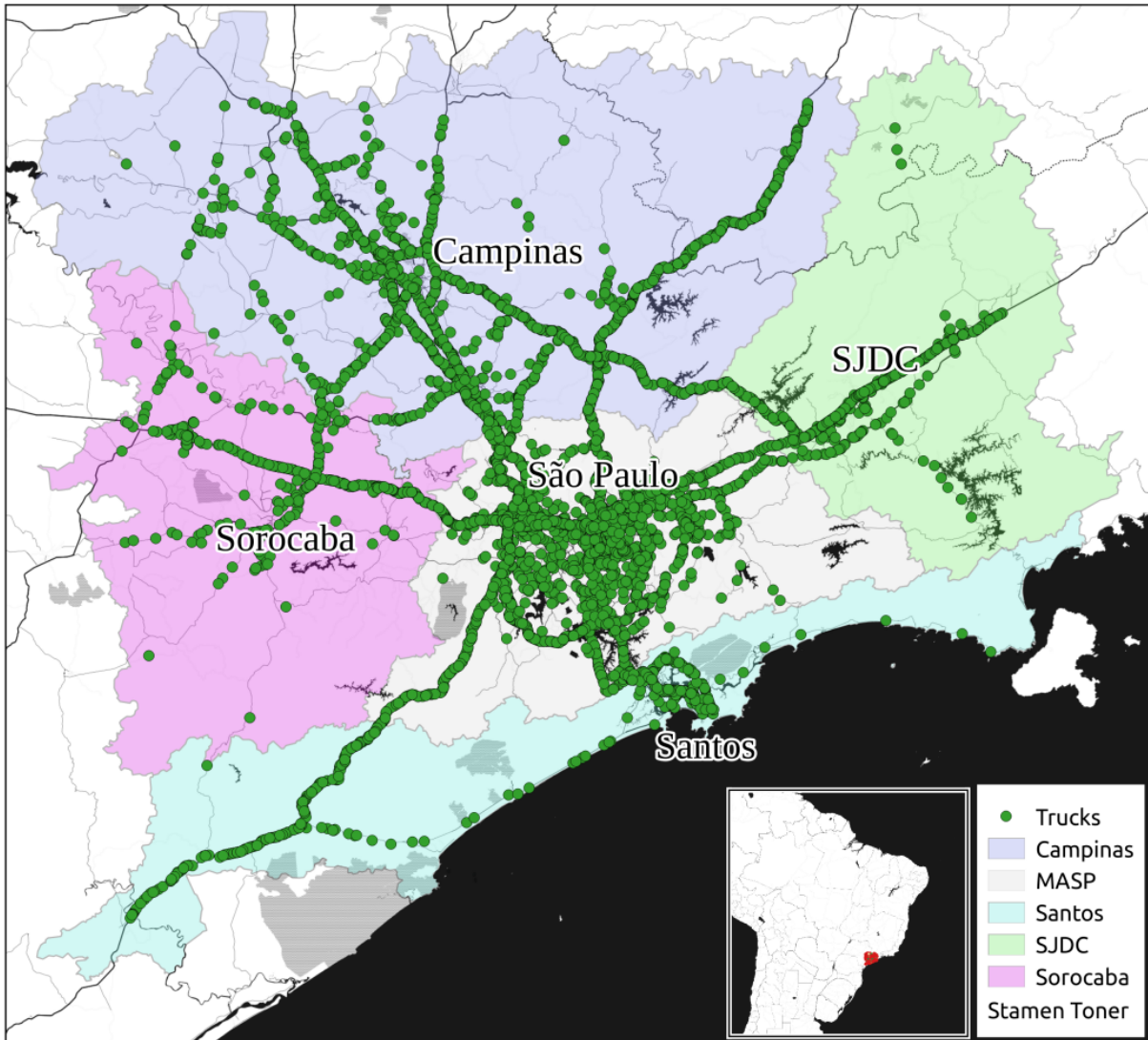


Figure 7 – Trucks positions derived from GPS on Monday 2014-10-06 09:00 LT.

of Taxis as proxy of Cars traffic flow. The number of streets with recording of Taxi is larger than Cars, with 5679 streets. The average and maximum speeds of Taxis are lower than Cars for *motorway*, and higher for the other type of streets. Trucks present higher speeds than Cars and Taxis at all type of streets but there are only 147 street with recordings. Lastly, the Max average speed considering all types (Cars, Taxi, Trucks and Undefined) is higher than the other type of vehicles but still lower than *motorways* of Cars, also, the number of streets with recordings are 9116. In order to produce traffic flows coherently despite the limitations of the data, it was decided to use the average max speed of taxis as proxy for traffic flow of Cars. For the Trucks, it was considered first its average max speed and when there was no data available, the average max speeds of all vehicles.

The spatial distribution of the speeds on Monday at 00:00 LT is show in figure 8. It can be seen that motorways have higher speeds with values between 88 and

Table 2 – Speeds Monday 2014-10-06 00:00 by type of street  $km \cdot h^{-1}$ .

Type of street	$Cars_{mean}$	$Taxi_{mean}$	$Truck_{mean}$	$Taxi_{max}$	$Trucks_{max}$	$All_{mean}$
Motorway	73.98	40.40	55.09	51.45	57.14	65.63
Motorway_link	-	36.99	56.96	41.53	57.83	52.44
Trunk	14.80	22.73	43.49	26.39	43.50	37.84
Trunk_link	-	22.09	-	23.26	-	34.99
Primary	16.09	16.38	44.12	23.26	45.22	27.12
Primary_link	22.00	18.44	-	21.30	-	28.04
Secondary	7.65	14.69	50.80	19.21	60.61	22.24
Secondary_link	0.06	18.64	-	21.00	-	25.39
Tertiary	4.17	12.59	42.32	16.54	43.31	19.59
Tertiary_link	-	16.00	59.65	16.54	72.20	24.49
Mean	10.31	17.76	54.19	23.46	55.96	31.37
n	64	5679	147	5679	147	9116

110  $km \cdot h^{-1}$  shown in red. At major streets near the cities the average speeds oscillate between 44 and 88  $km \cdot h^{-1}$  shown in yellow. The lower speeds are found in the center of the cities shown in green.

The temporal distribution of speeds by type of street for the 24 hours of the Monday 2014-10-06 is shown on Fig. (9). Higher speeds are found in motorways and the speeds diminish as descending on the hierarchy of the type of street from *motorway* to *tertiary* which seems reasonable. The interquartile range of speeds are lower as descending on the hierarchy of the type of street. Regarding the hourly distribution, only *motorway* shows a profile *trunk* of speeds increasing in early morning, and lowering at morning and evening rush hours, followed by an increase in speeds at night. *Primary*, *secondary* and *tertiary* do not present a noticeable variation at different hours.

It is important to have a well represented speed because not all the vehicles count with GPS devices, which could result in a spatial bias. This means that it is possible that specific regions with more vehicles with GPS devices will present higher number of registries, but other regions with higher traffic flows could not appear in the results if there were fewer vehicles with GPS devices. This is most probable for Cars than for Trucks because the spatial behavior of Trucks depends on the transport of goods, following therefore a more fixed route than Cars. One way to solve this problem is to use the speed as proxy of traffic flow. In this way, it is not crucial to have all streets monitored with GPS, but at least know the average speed on most of the streets.

Besides Traffic counts, described in section 2.3.3, CET also provides a database of simultaneous speed and traffic flow measurements, available for the public on <http://www.cetsp.com.br/sobre-a-cet/relatorios-corporativos.aspx>. This data was then used to correct the spatial bias from the GPS recordings. The procedure consisted in creating ratios between traffic flow of LDV and Trucks and speed per type of street and



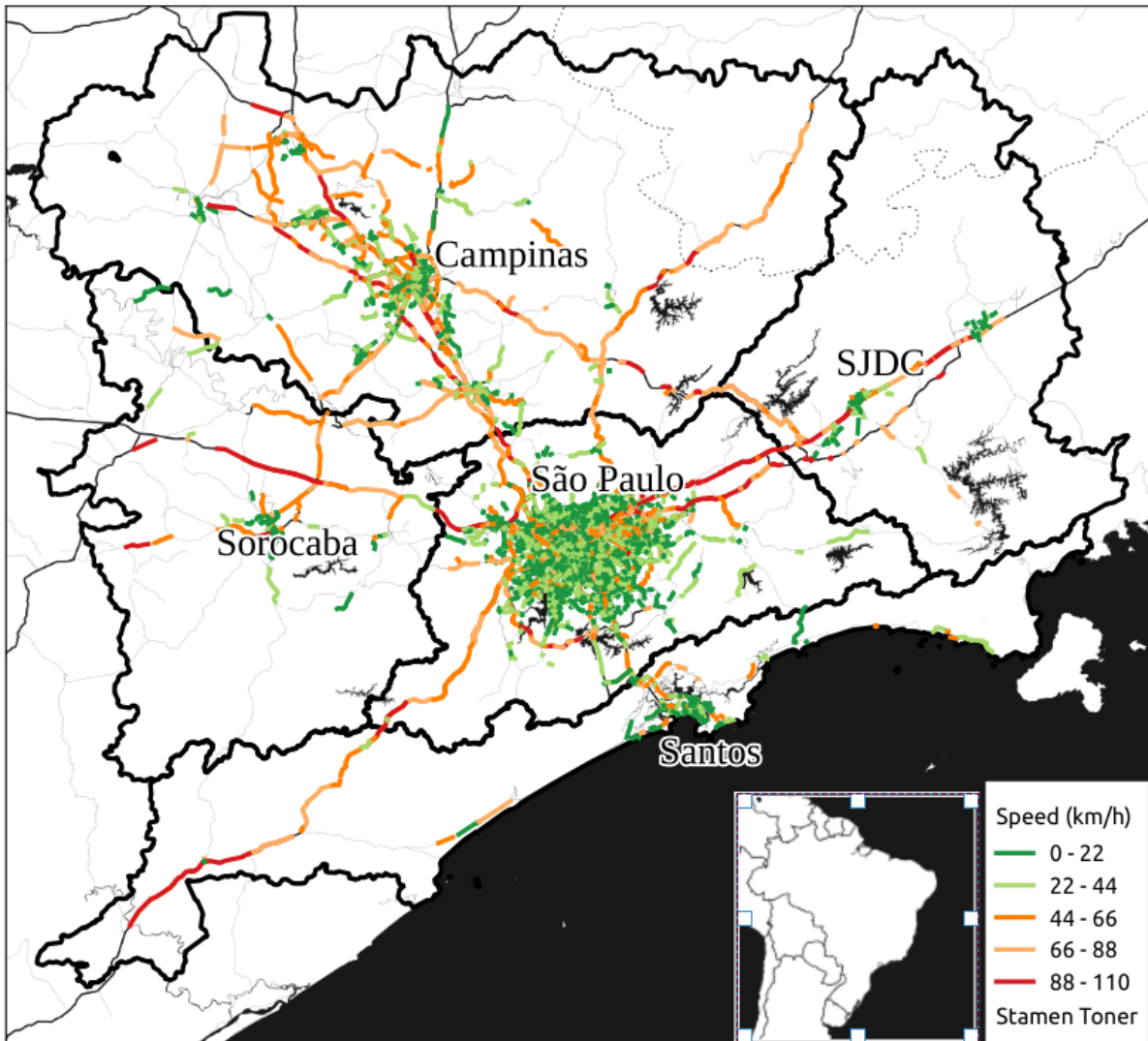


Figure 8 – Speed ( $km \cdot h^{-1}$ ) on Monday at 00:00 LT.

number of lanes. Then, the speed of LDV and Trucks from the GPS records was then multiplied with these ratios when type from street and number of lanes were similar. A formulation of this procedure is shown in Eq. (2.2).

$$FC_{ts,nl} = SGPS_{ts,nl} \cdot \left( \frac{FR_{ts,nl}}{SR_{ts,nl}} \right) \quad (2.2)$$

Where  $FC_{ts,nl}$  is the traffic flow corrected for the type of street  $ts$ , number of lanes  $nl$ .  $SGPS_{ts,nl}$  is the average speed obtained from the GPS recordings at streets indicating the type and number of lanes.  $FR_{ts,nl}$  and  $SR_{ts,nl}$  are the Traffic flow and Speed recorded during CET measurements made between 08:00 and 09:00LT during 2012. The values used as reference are shown in Table (3), as well as the number of lanes, which was also taken from the Traffic Counts Interpolation Section. This dataset is part of the paper in review (IBARRA-ESPINOSA et al., 2017) but the data set was also uploaded into Mendeley Data web services, doi: 10.17632/rz2cymv6c2.1 (IBARRA-



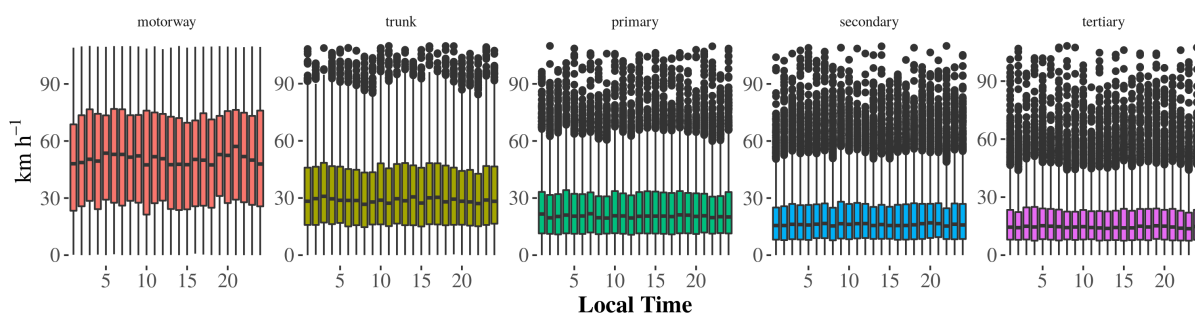


Figure 9 – Temporal distribution of average max speed of Taxi ( $km \cdot h^{-1}$ ) by type of street for Monday 2014-10-06.

Table 3 – Average traffic and speeds by type of street and number of lanes at 08:00-09:00 in São Paulo, 2012

Type of street	Lanes	PC $veh \cdot h^{-1}$	Speed $km \cdot h^{-1}$	PC/Speed
Motorway and Trunk	$\geq 10$	9615.500	69.15000	139.05278
Motorway and Trunk	$< 10 \& \geq 8$	4709.071	60.62143	77.67998
Motorway and Trunk	$< 8 \& \geq 6$	4416.850	54.01000	81.77837
Motorway and Trunk	$< 6 \& \geq 4$	3282.500	51.53333	63.69664
Motorway and Trunk	$< 4$	786.500	16.00000	49.15625
Primary	$\geq 8$	4486.500	61.68750	72.72948
Primary	$= 7$	2664.000	30.80000	86.49351
Primary	$= 6$	3233.250	37.30357	86.67401
Primary	$< 6 \& \geq 4$	3311.200	35.88000	86.67401
Primary	$< 4$	2697.000	29.50000	91.42373
Secondary	$\geq 8$	1671.000	26.00000	64.26923
Secondary	$< 8 \& \geq 6$	2388.000	36.64286	65.16959
Secondary	$= 5$	1295.000	28.90000	44.80969
Secondary	$= 4$	2527.000	26.43333	95.599
Secondary	$< 4$	626.500	27.50000	22.78182
Tertiary	$\geq 4$	1600.000	22.00000	72.72727
Tertiary	$< 4$	804.000	15.90000	50.56604

ESPINOSA, 2017). There are no traffic counts in Motorways, therefore it was assigned the same values of Trucks to Motorways. This table shows a tendency that the higher the hierarchy of the type of streets, the more lanes per streets, and the more vehicles circulating. However, the speed does not increase linearly. The traffic was corrected in 'link' streets, dividing by two, producing a smoother change of flow. It has to be noticed that since these ratio were obtained with Traffic counts, they might not be applicable to other regions.

The specification presented applies only for traffic flow generated from GPS recordings of Cars. In the case Trucks only the traffic circulating in *motorways* and *trunks* were corrected, using the average number of trucks in *trunk* streets  $101 \text{ veh} \cdot h^{-1}$  and the average speed for the same type of streets  $50 \text{ km} \cdot h^{-1}$ . The correction ratio

obtained was 2 and it was applied for trucks flow circulating only in *motorway* and *trunk*.

### 3 VEIN v0.2.2: an R package for bottom-up Vehicular Emissions Inventories

Emissions inventory is a quantification of pollutants discharged into the atmosphere by different sources (PULLES; HESLINGA, 2010). This quantification is vital for regulatory and scientific purposes, because it allows to monitor the state of the Earth's atmosphere and climate. 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 however do not provide detailed characterizations of high resolution applications which 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 Transboundary 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). However, there are many countries and cities that do not count with 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 in movement and the emissions process is complex. 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 that in most cases, emissions inventories are usually seen as a scapegoat when simulations do not match observation (PULLES; HESLINGA, 2010).

Vehicular emissions inventories are classified according to top-down and bottom approaches. Top-down are based on statistics of vehicle composition, representative speeds and country balances, while bottom-up are based on traffic counts, vehicle composition and speed recording (NTZIACHRISTOS; SAMARAS, 2016). The accuracy of the emissions inventory will reflect on 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; NTZIACHRISTOS; BOULTER, 2010) concluded that there is no evidence that the more complex models perform better than the less complex ones. An emissions inventory must be comprehensive, including all the important sources and aspects regarding the emissions. All of these complexities were addressed by the Laboratório de Processos Atmosféricos (LAPAt) from the Departamento de Ciências Atmosféricas (DCA) of the Universidade de São Paulo (USP) when modeling the atmospheric chemistry over Brazilian cities using a top-down vehicular emissions inventory with an on-line atmospheric model (ANDRADE et al., 2015). The Metropolitan Area of São Paulo (MASP), is the most populated megacity in Latin America (IBGE, 2014b) and its most important source of pollution comes from the 11 million vehicles that circulate within the Region (CETESB, 2013b; DENATRAN, 2015; ANDRADE et al., 2017). Furthermore, half of all emissions of  $CO$ ,  $HC$  and  $NO_x$  in the MASP are from vehicles that are more than 10 years old (ANDRADE et al., 2017). Vehicular emissions for these air quality studies were made with too many assumptions and approximations, reinforcing the need to develop a bottom-up vehicular emissions inventory model. This model will enable to generate scientific estimates and provide useful information to decision-makers and urban/environmental planners.

One of the main goals of this project was to develop a high spatial and temporal resolution vehicular emissions inventory model. This chapter presents the final version of the VEIN model, which follows the bottom-up approach. Section 3.1 presents the methodology used by VEIN. The model design, as an R-package, is described in section 3.2. An example of application of VEIN to MASP is given in section 3.3, but only for  $CO$  and  $LDV$ . Full application of VEIN to the previous chapter traffic activity data is described in detail in section 3.4. In section 3.4, a discussion about the VEIN model is presented.

### 3.1 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; HESLINGA, 2010), as shown in Eq. (3.1).

$$Emission_{pollutant} = \sum_{activity} (AR_{activity} \cdot EF_{pollutant,activity}) \quad (3.1)$$

where  $Emission_{pollutant}$  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,  $AR_{activity}$  represents the number of

vehicles times the distance (km) that they travel.  $EF_{pollutant,activity}$  is the emission factor ( $g \cdot km^{-1}$ ) for pollutants of the vehicles.

For a bottom-up estimation of vehicular emissions, a large number of parameters are involved with the activity (traffic flow, vehicle composition, speed recording, length of road) and emission factors (speed or acceleration dependent including deterioration) (NTZIACHRISTOS; 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.

### 3.1.1 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 it, VEIN arranges and organizes the data by vehicle composition, according to Eq. (3.2).

$$F_{i,j,k}^* = Q_i \cdot VC_{i,j} \cdot Age_{j,k} \quad (3.2)$$

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 definitions of (CORVALÁN; OSSES; URRUTIA, 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 (LDV) and it is known that 5% of the  $Q$  are passenger cars (PC), with engine lesser 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 it is usually available only 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; KOCKELMAN, 2009; LAM; 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 hour of the week  $l$  ( $F_{i,j,k,l}$ ; see Eq. 3.3).

$$F_{i,j,k,l} = F_{i,j,k}^* \cdot TF_{j,l} \quad (3.3)$$

where  $TF_{j,l}$  are the temporal factors varying according to each hour of  $l$  and type of vehicle  $j$ . For instance,  $TF$  is defined as a matrix with 24 lines and numbers of columns to each day considered, from Monday to Sunday. In order to expand traffic to other hours,  $TF$  matrices must be normalized to the hour that represents the traffic data. It means that  $TF$  values at morning peak hour must be 1 and the respective proportion must be assigned to the other hours. For example,  $TF$  values can be obtained from automatic traffic count stations.

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. (3.4). 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  $l$ .

$$T_{i,l} = T_{o_i} \cdot \left( 1 + \alpha \cdot \left( \frac{Q_{i,l}}{C_i} \right)^\beta \right) \quad (3.4)$$

In Eq. 3.4,  $T_i$  is the travel time per street link  $i$  at each hour of the week  $l$ .  $T_{o_i}$  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  $\alpha$  and  $\beta$  are adjustments with default values of 0.15 and 4, respectively. These are recommendations by the (Bureau of Public Roads, 1964). However, the user can use other values. When there is no available information for these calculations such as capacity at each street link  $i$ , it is possible to apply a simple average between peak and free flow speeds, in order to obtain the average speeds at different hours.

### 3.1.2 Selection of the emissions factors

The emissions factors describe the relationship intensity of activity and emissions for a given technology (PULLES; HESLINGA, 2010). In the case of the model, an emission factor is the mass of pollutant emitted by the vehicular type, technology and years of use. VEIN counts with 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; 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 km ( $g \cdot km^{-1}$ ) per specific type of vehicle, including fuel type, size and weight, by age of use.
- 3) Scaled local emission factors with Copert in order to incorporate speed variation for local factors, as shown in Eq. (3.5). This produces a specific speed dependent emission factor by age of use for the vehicle.

$$EF_{scaled}(V_{i,l})_{j,k,m} = EF(V_{i,l})_{j,k,m} \cdot \frac{EF_{local_{j,k,m}}}{EF(Vdc_{i,l})_{j,k,m}} \quad (3.5)$$

where  $EF_{scaled}(V_{i,l})_{j,k,m}$  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  $m$ .  $EF_{local_{j,k,m}}$  represents the constant emission factor (not as speed functions).  $EF(Vdc_{i,l})_{j,k,m}$  are Copert emission factors with average speed value of the respective driving cycle for the vehicular category  $j$ . The São Paulo emission factors data includes recordings of Federal Test Procedure (FTP-75) driving cycle for LDV with an average speed of  $34.12 \text{ km} \cdot \text{h}^{-1}$ , as shown in BARLOW et al., 2009.

By default, VEIN includes deterioration factors from Copert (NTZIACHRISTOS; SAMARAS, 2016). However, it is possible to include other sources, such as from (CORVALÁN; VARGAS, 2003).

### 3.1.3 Emissions estimation

VEIN estimates type of emissions including hot exhaust (EH; Eq. 3.6), cold start exhaust (EC; Eq. 3.7), evaporative (EV; Eq. 3.8), deterioration factors and speciation. The total vehicular emission is the sum of all types of emissions.

#### Hot exhaust emission

The VEIN process of emissions estimation is performed per street link, vehicle type, hour of week, and pollutant. Eq. (3.6) shows the hot exhaust estimation:

$$EH_{i,j,k,l,m} = F_{i,j,k,l} \cdot L_i \cdot EF(V_{i,l})_{j,k,m} \cdot DF_{j,k} \quad (3.6)$$

In Eq. (3.6),  $EH_{i,j,k,l,m}$  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  $k$ .

#### 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. Several studies have shown the significant impact for these types of emissions (CHEN et al., 2011b) (WEILENMANN; FAVEZ; ALVAREZ, 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, studies report that when ambient temperature is  $-7^{\circ}\text{C}$ , emissions are one order of magnitude higher than at  $22^{\circ}\text{C}$  (LUDYKAR; WESTERHOLM; ALMEN, 1999).

The VEIN model caters to these emissions by using the approach outlined in Copert (NTZIACHRISTOS; SAMARAS, 2016), as shown in Eq. (3.7).

$$EC_{i,j,k,l,m} = \beta_j \cdot F_{i,j,k,l} \cdot L_i \cdot EF(V_{i,l})_{j,k,m} \cdot DF_{j,k} \cdot (EF_{\text{cold}}(ta_n, V_{i,l})_{j,k,m} - 1) \quad (3.7)$$

This approach adds two terms to Eq. 3.6. 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  $l$  and pollutant  $m$  and monthly average temperature  $n$ . (NTZIACHRISTOS; SAMARAS, 2016) suggest using monthly average temperature. This is an important aspect that will be reviewed in future versions of VEIN.

The second term  $\beta_j$  is defined as the fraction of mileage driven with a cold engine/catalyst (NTZIACHRISTOS; 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.

## 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; 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; NTZIACHRISTOS, 2016). This approach is shown in Eqs. (3.8), (3.9) and (3.10).

$$EV_{j,k} = \sum_s D_s \cdot \sum_j F_j \cdot (HS_{j,k} + de_{j,k} + RL_{j,k}) \quad (3.8)$$

where  $EV_j$  are the volatile organic compounds (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 according 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 losses evaporative emissions ( $\text{g} \cdot \text{day}^{-1}$ ), 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 (MELLIOS; NTZIACHRISTOS, 2016):

$$HS_{j,k} = x_{j,k} \cdot (c \cdot (p \cdot e_{shc} + (1 - p) \cdot e_{swc}) + (1 - c) \cdot e_{shfi}) \quad (3.9)$$

where  $x$  are 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 hot engine, for example, an engine that has reached its normal operating temperature and the catalyst has reached its light-off temperature (NTZIACHRISTOS; 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 ( $\text{g} \cdot \text{parking}^{-1}$ ).  $e_{shfi}$  is the average hot-soak emission factors for gasoline vehicles equipped with fuel injection and non-return fuel systems ( $\text{g} \cdot \text{parking}^{-1}$ ).

$$RL_{j,k} = x_{j,k} \cdot (c \cdot (p \cdot e_{rhc} + (1 - p) \cdot e_{rwc}) + (1 - c) \cdot e_{rhfi}) \quad (3.10)$$

$x$  and  $p$  have the same meanings of Eq. 3.9.  $e_{rhc}$  and  $e_{rwc}$  are average hot and warm running losses emission factors for gasoline vehicles with carburettor or fuel return systems ( $\text{g} \cdot \text{trip}^{-1}$ ) and  $e_{rhfi}$  are average hot running losses emission factors for gasoline vehicles equipped with fuel injection and non-return fuel systems ( $\text{g} \cdot \text{trip}^{-1}$ ). It is recommended to estimate the number of trips per day (MELLIOS; NTZIACHRISTOS, 2016),  $x$ , as the division between the mileage and 365 times the length of trip:  $x = \frac{\text{mileage}_j}{(365 \cdot \text{trip})}$ . However, the mileage of a vehicle is not constant throughout the years. Therefore, VEIN incorporates a dataset of equations to estimate mileage of different types of vehicles by age of use (BRUNI; BALES, 2013).

### 3.1.4 Speciation of emissions in chemical sub-components

Particulate matter and hydrocarbons are a mixture of several chemical compounds that play an important role in atmospheric chemistry (SEINFELD; PANDIS, 2016). VEIN includes speciation profiles for hydrocarbons and particulate matter from (NTZIACHRISTOS; SAMARAS, 2016) and (RAFEE, 2015). These profiles are percentages of the emissions by vehicle type, fuel, emission standard and other characteristics. Also included are speciations of particulate matter in black carbon and organic matter, particulate matter fractions for tire, brake and road wear, non-methanic hydrocarbons and nitrogen oxides.

### 3.1.5 Spatial allocation and data-bases

VEIN provides functions to generate grids and spatially allocate emissions into grids. This is helpful for 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 3.3.4 provides details and examples about the emissions grids and databases.

## 3.2 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 language such as Lisp and syntax of S, as described by (IHAKA, 1998).

The VEIN R package depends on the package **sp** (BIVAND; PEBESMA; GOMEZ-RUBIO, 2013), as it uses several of its classes. VEIN imports some functions from the package **rgeos** (BIVAND; RUNDEL, 2016), which is an interface for the Geometry Open Source (GEOS) library (<https://trac.osgeo.org/geos/>). It also imports functions from **rgdal** (BIVAND; KEITT; ROWLINGSON, 2016), which provides bindings to the Geospatial Data Abstraction Library (GDAL; <http://www.gdal.org/>). 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; 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 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> or from <https://github.com/ibarraespinosa/vein> and the User Manual is available at <https://CRAN.R-project.org/package=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("ibarraespinosa/vein")

library(vein)
demo(VEIN)
```

The diagram process for estimating emissions is shown in Fig. 10. The green circles in this Figure refer to the data and the blue boxes refer to the functions inside the model. The VEIN model diagram starts at the green circle traffic, which represents the morning rush hour traffic data for each street link. Then the *age* functions (*age\_ldv*, *age\_hdv*, *age\_moto* or *my\_ldv*) determine the vehicular composition by age of use as shown in Eq. (3.2). The data *profile* allows to temporally extrapolate traffic data to the other hours and this allows to estimate the average vehicular speed to any hour and link using the function *netspeed*. Emission factor selections start by adding the deterioration effect with the function *emis\_det* into local, speed dependent emission factors from (NTZIACHRISTOS; SAMARAS, 2016), denoted as *speed\_ef* or scaled emission factors denoted as *scaled\_ef* in Fig. 10. Besides including speed dependent emission factors from (NTZIACHRISTOS; SAMARAS, 2016), VEIN also includes local emission factors from (CETESB, 2015). Once the input data is 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. 10 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 grid with format `SpatialPolygonsDataFrame` or a `SpatialGridDataFrame` 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 create an input for WRF-Chem model (GRELL et al., 2005). This data-frame must be exported as a .txt file and could be used as an input into other atmospheric models such as BRAMS (FREITAS et al., 2005).

### 3.2.1 Functions and classes

VEIN uses objects of class **Spatial** (PEBESMA; BIVAND, 2005), to represent road segments. To read geospatial data, there are several packages, such as **rgdal** (BIVAND; KEITT; ROWLINGSON, 2016) or **mapproj** (BIVAND; LEWIN-KOH, 2015). The main requirement is that the network must be a `SpatialLinesDataFrame`, class of **sp** (PEBESMA; BIVAND, 2005).

It was 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 4. These functions implement the equations shown in Section 3.1.

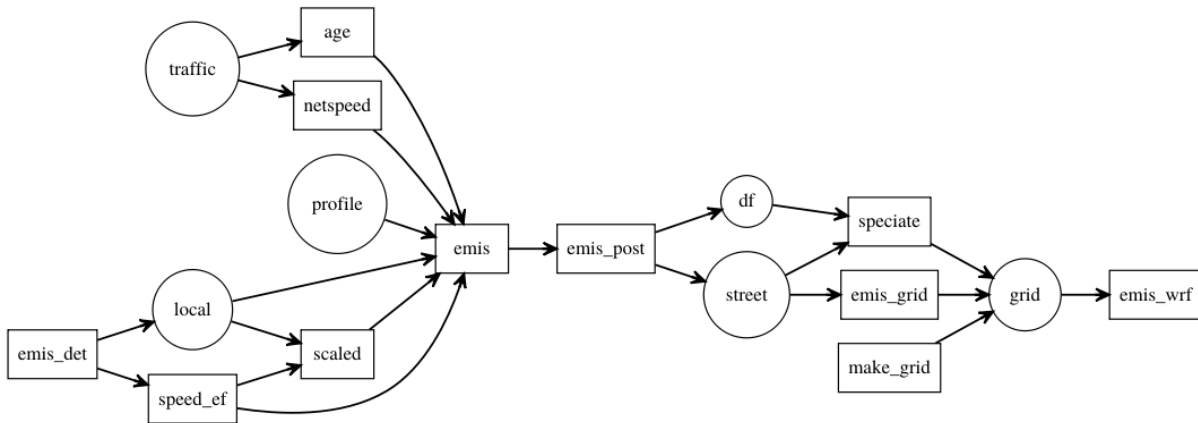


Figure 10 – Representation of the VEIN model. Boxes and circles represent functions and data, respectively.

VEIN incorporates 8 classes (see Table 4), 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 errors and improve the usability. For this task, VEIN imports some functions of the package **units** (PEBESMA; MAILUND; HIEBERT, 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 4 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 that the output of VEIN has a class. When the constructor function 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  $\text{g} \cdot \text{h}^{-1}$  to the numeric vector.

### 3.3 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).

Table 4 – Summary of the VEIN classes, functions and internal data.

Function	Description	Reference
age_hdv	Distribution of HDV by age of use	(Ministerio do Meio Ambiente, 2011)
age_ldv	Distribution of LDV by age of use	(Ministerio do Meio Ambiente, 2011)
age_moto	Distribution of Motorcycle by age of use	(Ministerio do Meio Ambiente, 2011)
ef_evap	Evaporative emission factors	(MELLIOS; NTZIACHRISTOS, 2016)
ef_hdv_scaled	List of scaled emission factors for HDV	(NTZIACHRISTOS; SAMARAS, 2016)
ef_hdv_speed	HDV Emission factors	(NTZIACHRISTOS; SAMARAS, 2016)
ef_ldv_cold	LDV cold start emission factors	(NTZIACHRISTOS; SAMARAS, 2016)
ef_ldv_cold_list	List of LDV cold start emission factors	(NTZIACHRISTOS; SAMARAS, 2016)
ef_ldv_scaled	List of scaled emission factors for LDV	(NTZIACHRISTOS; SAMARAS, 2016)
ef_ldv_speed	LDV Emission factors	(NTZIACHRISTOS; SAMARAS, 2016)
ef_wear	Tyre and break wear, and road abrasion	(NTZIACHRISTOS; BOULTER, 2009)
EmissionFactors	Creates class <code>EmissionFactors</code> ( $g \cdot km^{-1}$ )	
EmissionFactorsList	Creates class <code>EmissionFactorsList</code> ( $g \cdot km^{-1}$ )	
Emissions	Creates class <code>Emissions</code> ( $g \cdot h^{-1}$ )	
EmissionsArray	Creates class <code>EmissionsArray</code> ( $g \cdot h^{-1}$ )	
EmissionsList	Creates class <code>EmissionsList</code> ( $g \cdot h^{-1}$ )	
emis	Estimation to hour and day of the week	
emis_cold	Cold start estimation	(NTZIACHRISTOS; SAMARAS, 2016)
emis_det	Deterioration factors	(NTZIACHRISTOS; SAMARAS, 2016)
emis_evap	Evaporative estimation	(MELLIOS; NTZIACHRISTOS, 2016)
emis_grid	Allocation on rectangular grid	
emis_paved	Resuspension of paved roads	(USA-EPA, 2016)
emis_post	Post processing of emissions	
emis_wear	Estimation of wear emissions	(NTZIACHRISTOS; BOULTER, 2009)
emis_wrf	Creating data-frame to WRF-Chem	(VARA-VELA et al., 2016)
Evaporative	Creates class <code>Evaporative</code> ( $g \cdot d^{-1}$ )	
fe2015	Data of CETESB emission factors	(CETESB, 2015)
fkm	Data of mileage functions by vehicle	(BRUNI; BALES, 2013)
hot_soak	Hot soak evaporative	(MELLIOS; 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; NTZIACHRISTOS, 2016)
speciate	Split by species	(NTZIACHRISTOS; SAMARAS, 2016), (RAFEE, 2015)
Speed	Creates class <code>Speed</code> ( $km \cdot h^{-1}$ )	
temp_fact	Expand hourly traffic	
Vehicles	Creates class <code>Vehicles</code> ( $l \cdot h^{-1}$ )	
vkm	Determination of vehicle-kilometers	

### 3.3.1 Traffic data for MASP

Hourly traffic is a requirement for this data. This data can be represented as only one hour 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 morning rush hour traffic simulation for MASP from (CET, 2014) 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 HDV. Fig. (11) 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  $24708767 \text{ veh} \cdot h^{-1}$  and the number of streets are 34733 with mean of  $711 \text{ veh} \cdot h^{-1} \cdot street^{-1}$ . It is important to note that the VEIN model available at Github provides an extraction of the traffic simulation for the western part of São Paulo. The traffic simulation for MASP weights 61.6 Mb and the extraction for the western part of São Paulo 3.1 Mb. It was provided the extraction and not the whole traffic simulation in

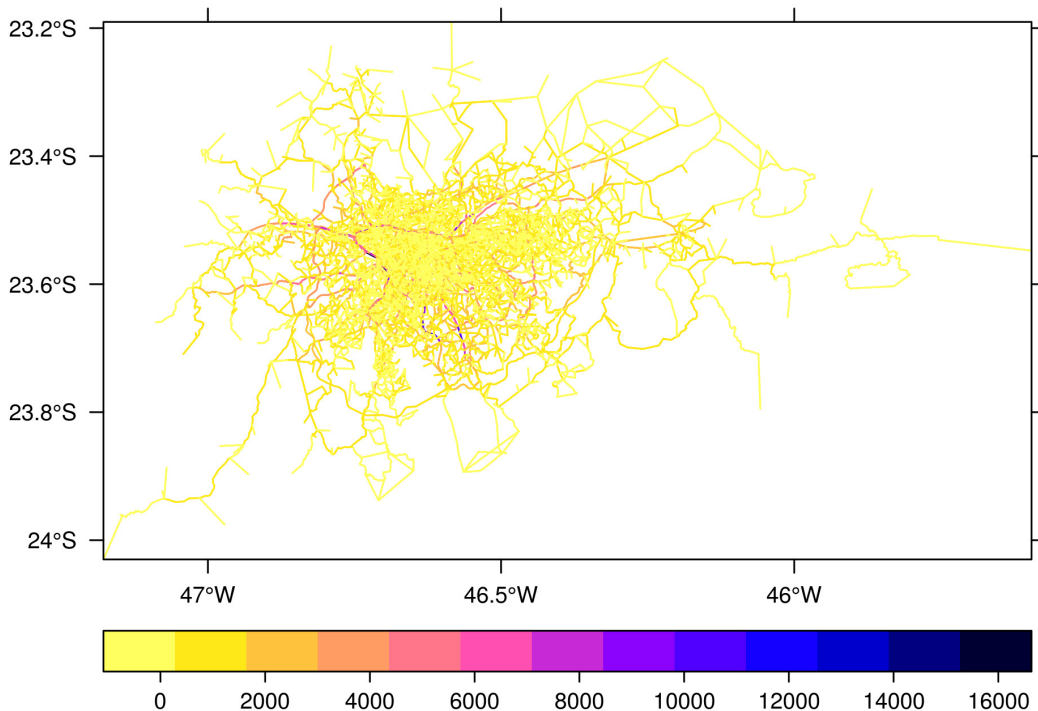


Figure 11 – Traffic flow simulation for LDV ( $\text{veh} \cdot \text{h}^{-1}$ ) at 08:00-09:00 LT for MASP.

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, it was created the function *netspeed*, which applies the function BPR (Bureau of Public Roads, 1964) curves, according to the Eq. 3.4. To use BPR, a data-frame is required with total traffic at all hours and the morning rush parameters capacity, peak speed, free flow speed, length of the road, and with BPR parameters alpha and beta. The argument *scheme* produces a 24 hour speed data-frame, based only on peak and free flow speed with a profile of free flow speeds at early mornings, peak speeds, 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 number of hours.

```
data(net)
data(pc_profile)
pcw <- temp_fact(net$ldv+net$hdv, pc_profile)
speedspeed <- netspeed(pcw, net$ps, net$ffs, net$capacity,
```

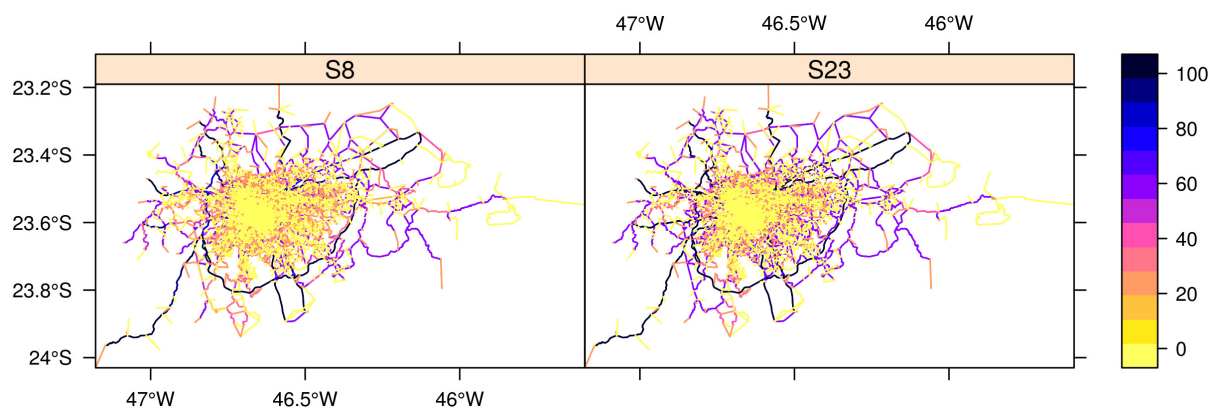


Figure 12 – Traffic speeds (colored lines;  $\text{km} \cdot \text{h}^{-1}$ ) for LDV fleet at 08:00 LT (left panel) and 23:00 LT (right panel) in MASP.

```
net$1km, alpha = 1)
```

For example purposes, the resulting speeds can be observed in Fig. 12, which shows two different speed maps: one for 08:00 LT (left panel) and the other for 23:00 LT (right panel). This 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 (left and right), is that in the late 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 shows Fig. 13. The type of streets comes from São Paulo traffic simulation and they were translated to English and are defined as Motorway: roads with speed limits above  $80 \text{ km} \cdot \text{h}^{-1}$  without physical intersections. Arterial: roads with speed limit of  $60 \text{ km} \cdot \text{h}^{-1}$  with intersection such as traffic lights. Collector: roads with speed limit of  $40 \text{ km} \cdot \text{h}^{-1}$  that collect and distribute traffic between Arterial streets. Local: roads with speed limit of  $30 \text{ km} \cdot \text{h}^{-1}$  that access restricted zones. Figure 13 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; SAMARAS, 2016). On the contrary, maximum average speeds for each type of road are obtained at 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 the Eq. 3.2, where they split the vehicular flow at street link  $Q_i$  by type

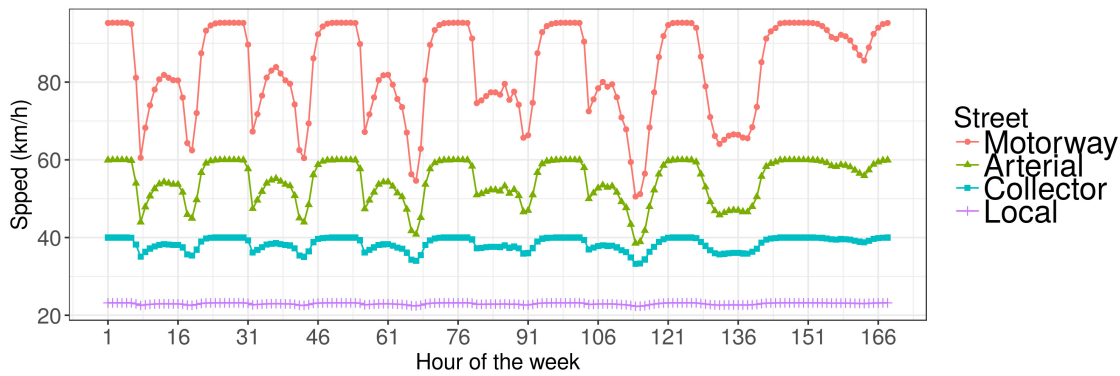


Figure 13 – Traffic average speeds for LDV fleet by type of street (colored lines) at 08:00-09:00 LT in MASP.

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 (Ministerio do Meio Ambiente, 2011) and 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, 2015), 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 14 (a) shows three age distributions, each one sums  $24708767 \text{ veh} \cdot \text{h}^{-1}$ , and each have different average age. It represents 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 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



Table 5 – Proposed equivalence of emission standards used in São Paulo study.

Vehicle	Brazilian Standard	Euro Standard	Year
LDV	L - 1	Pre ECE	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 <http://transportpolicy.net/index.php?title=Category:Brazil>

the number of streets. This allows different age distributions within the same road network and it is particularly useful for areas with less information about the vehicles' age of use.

### 3.3.2 Emission Factors

Once it was obtained the traffic flow for the desired type of vehicles (in 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 it was can proceed to the emissions calculation by itself.

The VEIN package includes a database titled *fe2015* with emission factors for PC and light trucks by age of use from the São Paulo official vehicular emissions inventory (CETESB, 2015). This inventory was performed using a top-down approach and the pollutants estimated were CH<sub>4</sub>, CO, CO<sub>2</sub>, HC, N<sub>2</sub>O, NMHC, NO<sub>x</sub>, and PM. This data includes national and equivalent Euro Emission Standards by year and age. The equivalence among Brazilian (CONAMA, 1986), (Ministerio de Medio Ambiente, 2011) and Euro (DIRECTIVE70/220/EEC, 1991) was added into this database in order to choose the corresponding matching vehicle and emissions standard. The equivalence can be seen in Table 5.

*fe2015* emission factors do not include the deterioration effect due to accumulated age of vehicle and it must be included. This is performed with the the deterioration factor function *emis\_det* which has the arguments: pollutant, size of engine, Euro Stan-

standard and mileage in km. 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; BALES, 2013).

Emission factors for PC, LCV and motorcycles are called with the function *ef\_ldv\_speed*. In the case of trucks and buses they use the function *ef\_hdv\_speed*. 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, 2015) emission factors. The age of LDV shown in Fig. 14 (a) has a length of 41 years. This means that it needs 41 emission factors, one per each age of use. It calls the function *emis\_det*, which requires the accumulated mileage, obtained from the list of mileage equations *fkm*. The Fig. 14(c) shows the emission factors from CETESB with and without deterioration by age of use. It was used deterioration factors from (NTZIACHRISTOS; SAMARAS, 2016) that affect only vehicles with a catalytic system. The base year of this emissions estimation is 2015 and the vehicles with catalytic system started in 1992 in Brazil (23 years before 2015). Therefore, the vehicles that entered into 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, it was repeated the oldest emission factors to have 41 emission factors. Here it was assumed that the emission factors of vehicles 36 years of use is the same as the vehicles till 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(pc_profile)
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 into the Brazilian emission factors (CETESB, 2015), 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  $34.12 \text{ km} \cdot \text{h}^{-1}$  for FTP-75. The default speed value is  $34.12 \text{ km} \cdot \text{h}^{-1}$ , 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)
```

### 3.3.3 Emission estimation

After inputting the database of vehicles and their respective emission factors, VEIN is ready to use the *emis* function. The VEIN package counts with 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 4 dimensions: 1) number of streets; 2) max 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 and array = TRUE. The values can be changed accordingly.

For example, the estimation of the traffic simulations shown in Fig. 11 has 34,733 streets, fleet of 41 years age distribution, 24 hours of the day, and 7 days of the week. Therefore, it will produce an emissions array with dimensions 34,733, 41, 24 and 7. The vehicle fleet used to produce the age distribution is shown in green in Fig. 14(a) and it has 41 years of length.

```
data(pc_profile)
E_CO <- emis(veh = pc1, lkm = net$lkm, ef = lef,
            speed = speed, profile = pc_profile)
```

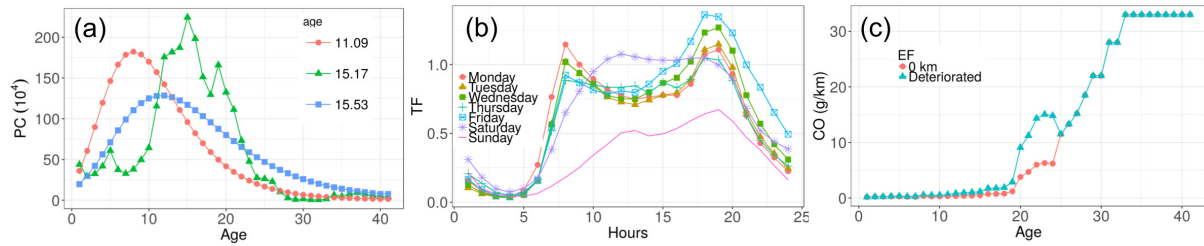


Figure 14 – (a) Distribution of LDV composition by age of use, (b) temporal factor for expanding morning rush hour traffic data and (c) CO emission factors used in the estimation presented in this manuscript.

This emissions array output for 34,733 streets and vehicle fleet with 41 years age distribution, 24 hours 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; BIVAND, 2005) and **ggplot2** (WICKHAM, 2009). VEIN outputs could also be used with the package **openair** (CARSLAW; 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 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. 15a. Higher emissions are found at 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 high-emitting vehicles. Fig. 15b 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. 14 (a). The total number of vehicles is  $24708767 \text{ veh} \cdot \text{h}^{-1}$  (8:00-9:00 LT on a Monday). The total CO emissions is  $233095 \text{ t} \cdot \text{y}^{-1}$ , 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 15% of the fleet, emitting 64% ( $148712 \text{ t} \cdot \text{y}^{-1}$ ) 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 5), and also the introduction of the catalytic system in Brazil. Therefore, the high emissions are due to vehicles with a deteriorated catalytic converter. This information is useful

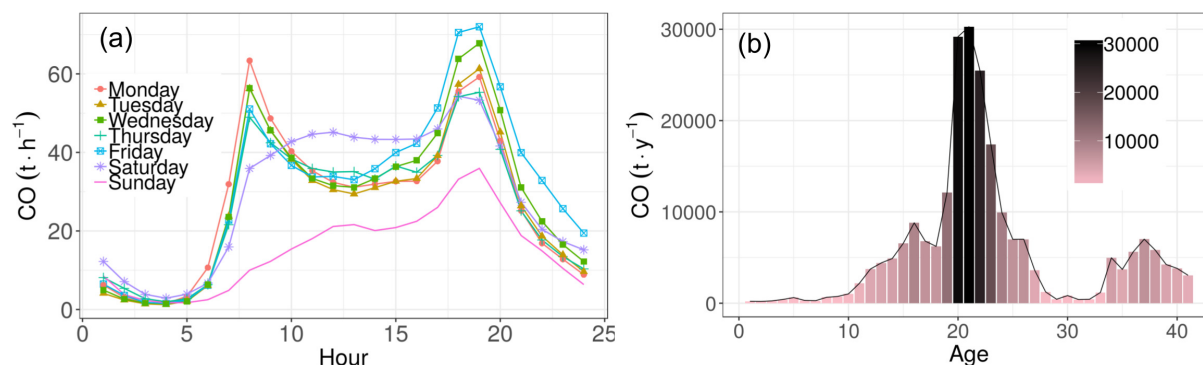


Figure 15 – (a) CO Emissions ( $t \cdot h^{-1}$ ) per hour of the day and day of the week (colored and shaped lines) for LDV from MASP, (b) CO Emissions ( $t \cdot y^{-1}$ ) according to the age of use of the LDV from MASP.

for reducing air pollution, thus supporting the aims of environmental planners and local authorities.

### 3.3.4 Post estimation

The spatial dimensions of the emissions estimation is an important feature of VEIN because it allows the representation of the streets into 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 34733 streets, 24 hours and 7 days of the week, it returns a data-frame with 5835144 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; 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 number of rows matching number of streets for the domain and number of columns as hours. Fig. 16 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(arrs = E_CO, pollutant = "CO",
                          by = "streets_wide")
```

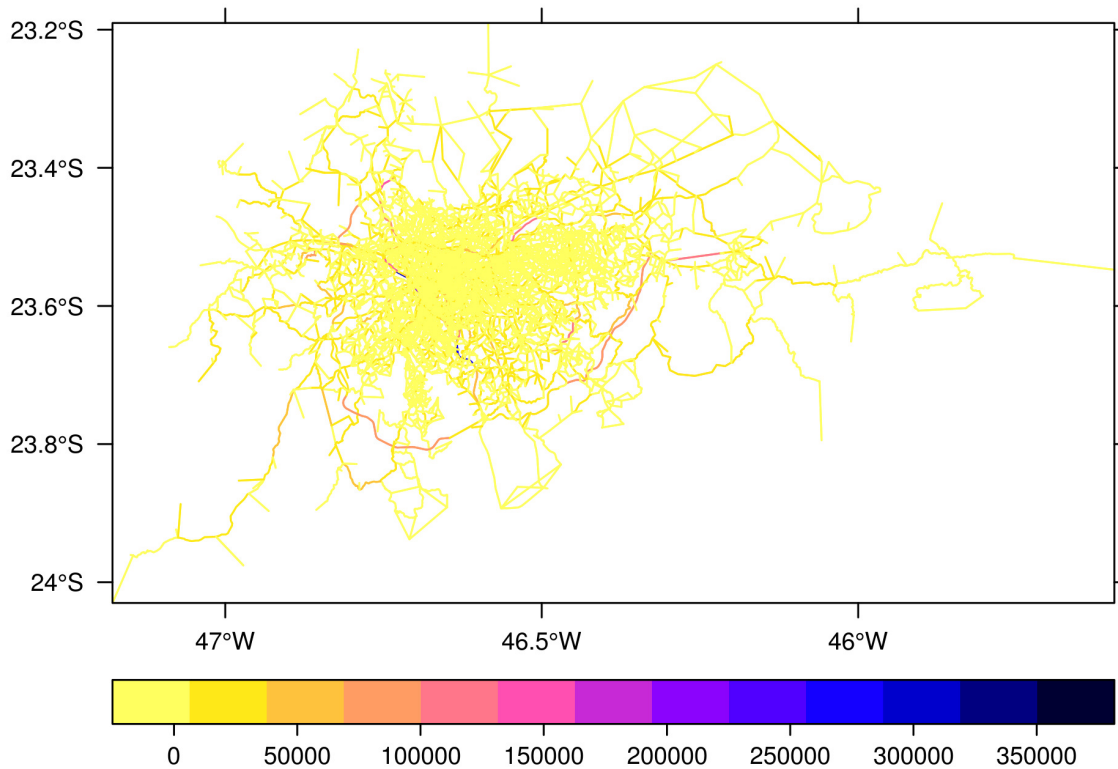


Figure 16 – CO Emissions (colored lines;  $\text{g} \cdot \text{h}^{-1}$ ) for LDV at Friday 19:00 LT over MASP.

```
net@data <- cbind(net@data, E_CO_STREETS)
```

The emissions shown in Fig. 16 are concentrated in two streets, a motorway and a trunk street at 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 the grid is very important for visualization and for inputs to air quality models. It was 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. `SpatialLinesDataFrame` objects 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; 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 `intersect` 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 id of grid and adds these emissions by grid id into the grid. The results are in emissions grid with 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 package `sf` (PEBESMA, 2016). In the following code, it is shown the use of the function `make_grid` only for example purposes. It is recommended that 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. 17. 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")
```

The Fig. 17 (a) shows a `SpatialPolygonsDataFrame` of emissions of CO with resolution of 1 km representing the base year 2015. This emissions grid was built using the package `sf`. Fig. 17 (b) 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.

### 3.3.5 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' (NTZIACHRISTOS; SAMARAS, 2016), 'tyre', 'break', and 'road' (NTZIACHRISTOS; BOULTER, 2009). However, there is only one type of VOC speciation for MASP, the 'iag' (RAFEE, 2015). For PM, the default speciation is 'bcom', which splits the exhaust emissions into black carbon and organic matter.

### 3.3.6 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 ("on-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 conception (2005 to 2006).

WRF-Chem, as another regional atmospheric model, requires a superficial layer of 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.

Fig. 17 shows a comparison for VEIN (Fig. 17a) and EDGAR (Fig. 17b), using emissions inventory for the CO in MASP. One can notice that CO is spatially well represented for VEIN, by comparison with EDGAR. Furthermore, VEIN offers much more details about the emission of this pollutant, which occurs mainly in the urban motorways due to the high volume of traffic in these roads. The total emissions of CO emissions using VEIN are  $1.73e-06$  ( $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ), considering the first second of a typical Monday at 00:00 LT and EDGAR,  $8.46e-08$  ( $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ). 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 manuscript is illustrative, and that more detailed emissions inventories should be made when comparing to others. For example, the



inventory for this manuscript includes estimates only for LDV assuming that all are PC. It does not include other types of vehicles as the total amount of vehicles were not calibrated with fuel consumption. (NTZIACHRISTOS; SAMARAS, 2016) recommends to compare bottom-up estimates with fuel consumption in order to calibrate inputs of emissions inventory (traffic data in this case). These differences highlight the needs for development, inter-comparison 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.

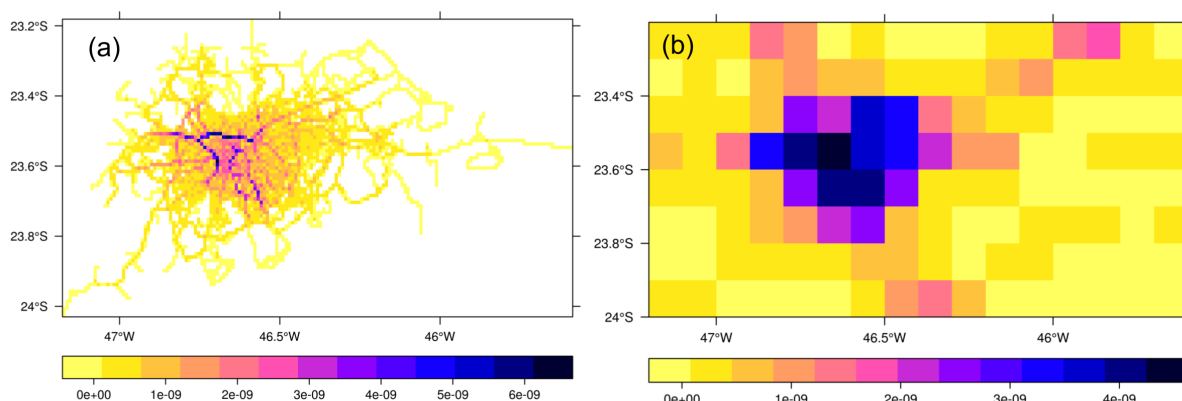


Figure 17 – CO emissions ( $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ ) in MASP for (a) Monday at 00:00:00 LT, estimated with VEIN for a 1 km x 1 km grid spacing, and (b) the emissions of road transport for the same area from EDGAR.

The VEIN model provides functions to transform the emissions grids into inputs for atmospheric models. (VARA-VELA et al., 2016) presented a system for assimilating anthropogenic emissions (AAS4WRF) into the WRF-Chem model. AAS4WRF consists of an NCL (BOULDER, 2017) script that reads a text file in long format and a WRF input header for the desired domain. The VEIN model provides the function *emis\_wrf* to automatically create a data-frame in the correct format with columns longitude, latitude, id of grid cell, pollutants, local time and GMT time in format POSIXct. The arguments of *emis\_wrf* are *sdf*, which is a list of `SpatialPolygonsDataFrames`, each per pollutant. *nr* indicates how many times the 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 if the first argument, *df* is a list (*islist* = TRUE) or a data-frame (*islist* = FALSE). The output of VEIN emissions are in  $g \cdot h^{-1}$ , and in the following code chunk, emissions were converted to mol.

```
E_CO_g@data <- E_CO_g@data[, -1] / (12 + 16)
ldf <- list("co" = E_CO_g)
```

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

## 3.4 VEIN model applied to different traffic activity data

The previous sections of this chapter presented the VEIN model R package focusing on LDV as example. In the next sections, the application of VEIN to different traffic activity data is written in more detail.

### 3.4.1 Conciliating bottom-up and top-down emissions estimation

Once the traffic flows are generated, as explained in Chapter 2, it is time to arrange the traffic in an harmonized way and design the scenarios before the emission estimation. The European guidelines for vehicular emissions inventories shows that there are differences using the two approaches for estimating emissions, bottom-up and top-down (NTZIACHRISTOS; SAMARAS, 2016). The guidelines say that the estimations with both approaches must be conciliated using fuel sales statistics as criteria. In other words, the fuel consumption estimated with any method should result in the same amount of fuel sold for the region and period of time considered.

In the case of bottom-up approaches, the only way of calibrating the fuel consumption is by factoring the traffic flow because the length of the road is constant. Therefore, in order to calibrate the traffic flows generated for this study, it was compared the fuel consumption estimated for these vehicles with fuel sales for each metropolitan region. The base year is 2014 and the fuel sales comes from the Yearbook of Energy and Statistics for São Paulo State 2014 (ELECTRICA, 2014). This technical report is made by the State Government and includes fuel sales for each municipality of the Estate of São Paulo. There are several types of fuel for automotive use. The most used automotive fuels used in Brazil are Gasohol or Gasoline with 25%-27% of ethanol (E25), ethanol (E100) and diesel with 5% of biodiesel (B5). The total amount of fuel sold during 2014 for automotive use and for each metropolitan region is shown on Fig. (18). Most of fuel sales are in MASP because it is the bigger area with most inhabitants and number of vehicles with 3383973.7 t of E25, 2845791.6 t of E100 and 2274181.4 t of B5. This is the only region where sales of E25 are higher than E100. This could have some implications regarding the atmospheric chemistry and transport of pollutant between these regions.

The fuel sold for automotive uses can sometimes have different use than for vehicle itself. For example, many electric generators use diesel or gasoline but they enter into the statistics as automotive fuel sales. Therefore, it is expected that a fraction of the fuel sales could be used for other purposes. In order to consider this, it was adopted

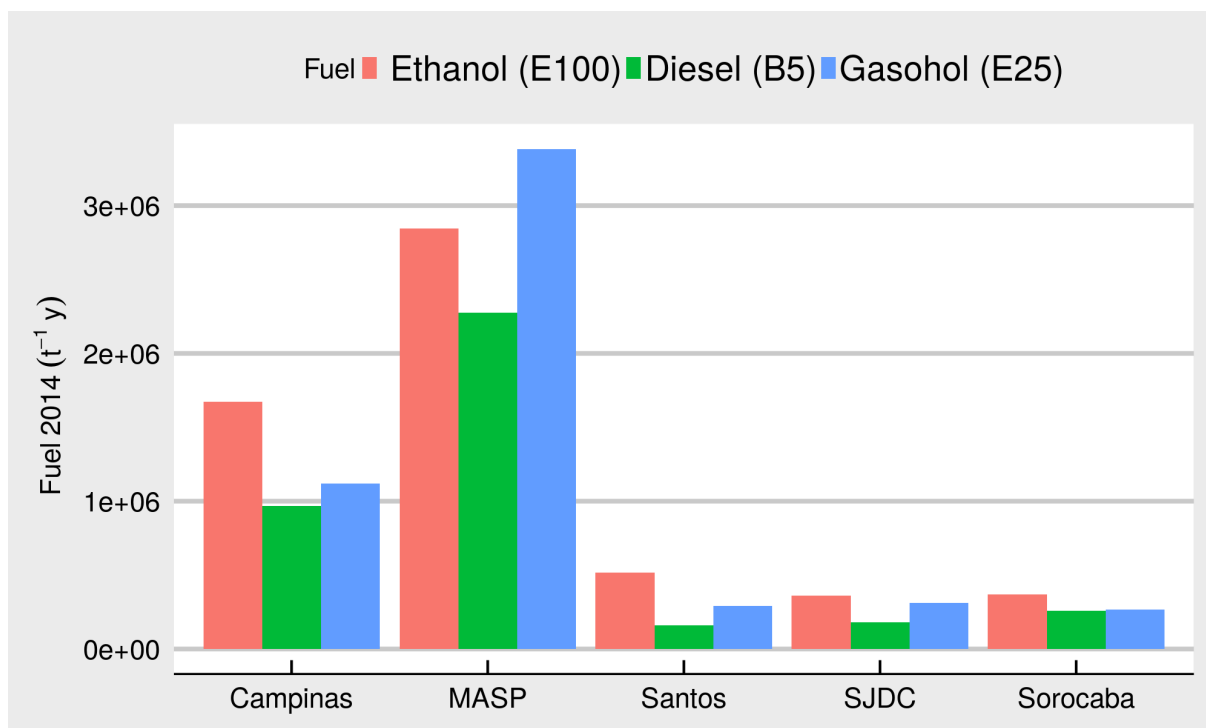


Figure 18 – Fuel sales for 2014 for metropolitan areas Campinas, MASP, Santos, SJDC and Sorocaba.

a criteria for evaluating the fuel consumption: 20% of B5 was used for other purposes than automotive. This means that the fuel consumption estimated from the traffic must be 80% of B5 and 100% of E25 and E100. This value is a conservative number because there are no statistics of diesel fuel sold for purposes such as electricity generation.

Once the amount of fuel consumed with the different traffic flows are estimated they must be compared to the fuel sales.

### 3.4.2 Running VEIN model for the different activity data scenarios

Traffic data can be either, only a morning rush hour data or hourly data for different hours. When inputting only morning rush hour data VEIN provides functions to expand traffic and speeds to other hours. Then, hourly traffic data must be split into the vehicular composition, which is a vehicular classification of vehicles by type, engine size, weight and fuel, without considering the emission standards. For example, Passenger Cars using Gasoline with engine lesser than 1400 cc is one member of the vehicular composition. Then, it is applied an age distribution to each member of the vehicular composition. This allows to have the vehicular composition by age of use, for example: Passenger Cars using Gasoline with engine lesser than 1400 cc with one year of use. The age distribution is based on statistics of vehicles in circulation such as licensing. It is possible to use also survival functions built inside VEIN.

## Vehicular composition

Vehicles classification is an important step when developing emissions inventories. The first is the family classification where vehicles are classified in one of the 5 families, Passenger Cars (PC), Light Commercial Vehicles (LCV), Heavy Good Vehicles (HGV) or Trucks, Buses and Motorcycles (MC). Traffic data used in this study was divided into family members. Traffic of Cars from GPS recordings and LDV from traffic simulation were divided into 73.70 % PC and 13.90 % LCV and 12.40 % MC. The second level of classification is the vehicular composition, defined as sub-classification of vehicles without including the technological improvement due to emission standards or technological composition. The concept of vehicular composition was introduced by (CORVALÁN; OSSES; URRUTIA, 2002) in the development of a vehicular emission inventory for the city of Santiago, Chile. The vehicular composition is the classification of the fleet by type of motor, fuel, size of engine and weight capacity. In Brazil there are mostly three types of motors, 4-strokes, flex 4-strokes and diesel-compression. 4-strokes engines uses gasohol, gasoline with 25% of ethanol (E25) and compression that uses diesel with 5% of biodiesel (B5). Flex engines can use gasoline with any mixture ethanol between 25% and 100%, therefore, the driver can choose between E25 or E100. Fleet statistics of engine size, weight and type of fuel were taken from the report of vehicular emissions inventory for São Paulo (CETESB, 2015). PC was divided by size of the engine with sales statistics of the National Association of Car Manufacturers (ANFAVEA, 2014) by: 38.13%  $\leq 1400$  cc, 58.12%  $1400 < cc \leq 2000$  and 3.54%  $> 2000$  cc. MC was divided with sales statistics of the National Association of Motorcycles Manufacturers (ABRACICLO, 2014) by: 86.23%  $\leq 150$  cc, 12.32%  $150 < cc \leq 500$  and 3.43%  $> 500$  cc. Resulting vehicular composition for this study is shown in Table (6). This table shows the category of vehicular composition, the percentage in each family, average age and max age. Each name of the category was built so that it is self explanatory, for example: PC\_FE25\_1400 are Passenger Cars with flex engine using gasohol E25 with engine size of 1400 cc or less. It was assumed that the oldest vehicle has 40 years of use. Vehicles with flex engine have lower maximum age because these engines entered into the market in 2003 for PC and in 2010 for MC and the base year of the presented inventory is 2014. Also, vehicles designed to work only with E100 have an average age of approximately 20 years of use because these vehicles are no longer produced since 2006. The vehicular categories 13th to 27th include number of the gross weight (GW) at the end of the name being: GW  $\leq 3.5t$  for LCV; GW  $\leq 7.5t$ , GW between 12t and 14t, GW between 14 and 16 t, GW between 16t and 20t, GW  $\geq 32t$  for HGV; and GW  $\leq 15t$ , GW between 15t and 18t and GW  $\geq 18t$  for Buses. The vehicular categories 23 is Small UB, the 25, 26 and 27 are UB, Articulated UB and Small UB of SPtrans for public transportation. According to SPtrans (<http://www.sptrans.com.br/indicadores/>) the average age of UB was 5.2 in

Table 6 – Vehicular composition used in this study.

	Categories	Perc	Mean	Max
1.	PC_E25_1400	14.21	14.25	40
2.	PC_E25_1400_2000	21.65	14.25	40
3.	PC_E25_2000	1.39	14.25	40
4.	PC_FE25_1400	8.50	5.52	12
5.	PC_FE25_1400_2000	12.93	5.52	12
6.	PC_FE25_2000	0.83	5.52	12
7.	PC_FE100_1400	14.48	5.52	12
8.	PC_FE100_1400_2000	22.07	5.52	12
9.	PC_FE100_2000	1.42	5.52	12
10.	PC_E100_1400	0.93	19.83	40
11.	PC_E100_1400_2000	1.42	19.83	40
12.	PC_E100_2000	0.09	19.83	40
13.	LCV_E25_3_5	39.13	13.97	40
14.	LCV_FE25_3_5	15.21	5.01	12
15.	LCV_FE100_3_5	25.90	5.01	12
16.	LCV_E100_3_5	1.18	20.34	40
17.	LCV_B5_3_5	18.56	13.97	40
18.	HGV_B5_7_5	8.38	15.72	40
19.	HGV_B5_12_14	25.50	15.72	40
20.	HGV_B5_14_16	15.28	15.72	40
21.	HGV_B5_16_20	24.98	15.72	40
22.	HGV_B5_32	25.85	15.72	40
23.	SUB_B5_15	9.07	9.28	40
24.	Coach_B5_18	13.5	9.28	40
25.	UB_SP_B5_15_18	28.81	4.94	10
26.	UBA_SP_B5_15_18	14.07	4.94	10
27.	SUB_SP_B5_15	17.23	4.94	10
28.	MC_E25_150	70.04	7.58	40
29.	MC_E25_150_500	10.83	7.58	40
30.	MC_E25_500	3.02	7.58	40
31.	MC_FE25_150	3.77	2.75	5
32.	MC_FE25_150_500	0.55	2.75	5
33.	MC_FE25_500	0.15	2.75	5
34.	MC_FE100_150	6.42	2.75	5
35.	MC_FE100_150_500	0.94	2.75	5
36.	MC_FE100_500	0.26	2.75	5

2014, therefore, it was adjusted the max age of this category to 10 years of use so that the average age is 4.94. MC was distributed using fleet statistics of (CETESB, 2015).

There is another element that is related to the average age of each category of the vehicular composition. This element is the statistical distribution of the vehicles by age of use. Each category of the vehicular composition has its own distribution by age of

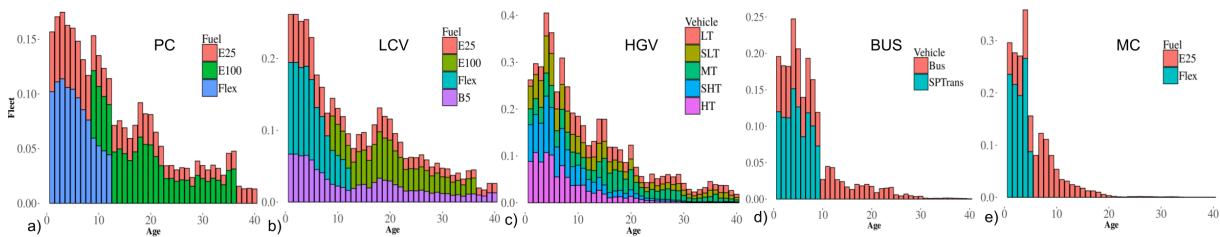


Figure 19 – Normalized fleet of PC, LCV, HGV, Bus and MC by age of use.

use. Each distribution comes from annual time series of vehicles in circulation. (CETESB, 2015) reports a description of the fleet by type of vehicle and fuel. It was used this data to distribute each category of the vehicular composition. The normalized licensing curves are shown in Fig. (19). It can be seen an increment in licensing over the last ten years and a small declination in 2010. The distribution of vehicular composition by age of use allows to associate each vehicle-age with the respective emission standard. The Brazilian emissions standard Proconve varies between 0 (Pre-Proconve) and 7 for HGV and Bus and 0 and 5 for PC and LCV. Table 10 shows the equivalence of Brazilian and European Standards.

### Emissions Factors

VEIN model allows to choose between the type of emission factors, local sources, speed functions from (NTZIACHRISTOS; SAMARAS, 2016) or scaled factors to incorporate speed variation into local factors. In the case of São Paulo, the vehicular emissions inventory reports the emission factors of new vehicles that entered into the market each year (CETESB, 2015).

These factors are the average of the emission certifications with driving cycle Federal Test Procedure FTP-75 for 4-strokes light duty vehicles, motorcycles with World Motorcycle Test Cycle (WMTC) and HDV with European Stationary Cycle (ETC). Therefore, the data consists in yearly emission factors by type of vehicles: PC are 1) PC\_E25, 2) PC\_FE25, 3) PC\_FE100 and 4) PC\_E100, LCV are 5) LCV\_E25, 6) LCV\_FE25, 7) LCV\_FE100, 8) LCV\_E100 and 9) LCV\_B5. Trucks are 10) Light Truck  $GW < 6t$ , 11) Semi Light Truck  $GW$  between  $6t$  and  $10t$ , 12) Medium Truck  $GW$  between  $10t$  and  $15t$ , 13) Semi Heavy Truck are rigid trucks with  $GW$  less than  $15t$  and articulated Trucks with  $GW$  less than  $40t$ , and 14) Heavy Trucks which are rigid Trucks with  $GW$  less than  $15t$  and articulated trucks with  $GW$  bigger than  $40t$ . Buses are 15) UB, 16) Small UB, and 17) Coach. MC are 18) MC\_E25, 21) MC\_FE25 and 24) MC\_FE100 for engines less than  $150cc$ , between  $150cc$  and  $500cc$  and bigger than  $500cc$  with 26 types of vehicles in total. The emission factors of these categories were assigned into the vehicular composition shown in Table (6), for example, the CETESB emission factors PC\_E25 was assigned equally to VEIN categories PC\_E25\_1400, PC\_E25\_1400\_2000

and PC\_E25\_2000.

The CETESB emission factors covers the following pollutants of hot exhaust and evaporative emissions:  $CO$ ,  $NO_x$ ,  $PM$ ,  $SO_2$ ,  $HC$ ,  $NMHC$ ,  $Aldehyde$ ,  $CO_2$  and  $N_2O$ . The emission factors are numerous so only the  $CO$ ,  $NO_x$ ,  $HC$  and  $PM$  Emission factors (in  $g \cdot km^{-1}$ ) for PC, LCV, Medium Trucks, UB and MC from (CETESB, 2015) are shown in Fig. (20). The figure is in logarithmic scale so that it is easier to identify small variations. The highest emission factors of  $CO$  are due older PC with  $33 g \cdot km^{-1}$ , with  $NO_x$  the older UB emit  $17 g \cdot km^{-1}$ , in the case of  $HC$  the older PC emit  $1.16 g \cdot km^{-1}$  and in the case of  $PM$ , the older UB emit  $1.07 g \cdot km^{-1}$ . In all cases newer vehicles emit less due to technological improvements resultant from the stringent emission standards. It is interesting to note that newer MC emit more  $CO$  and  $HC$  than the other type of vehicles.

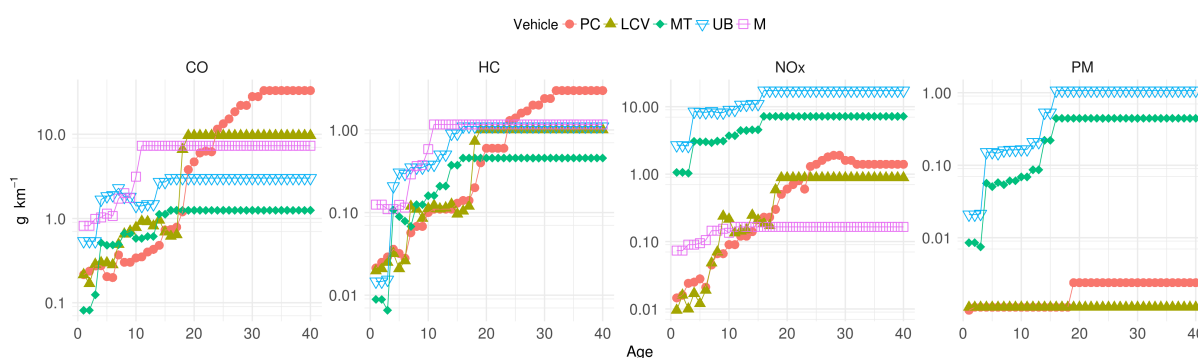


Figure 20 –  $CO$ ,  $NO_x$ ,  $HC$  and  $PM$  Emission factors ( $g \cdot km^{-1}$ ) for PC, LCV, Medium Trucks, UB and MC (CETESB, 2015).

It is important to mention that there is no formal definition of vehicular composition in VEIN, however, the vehicular composition followed the type of vehicles in the European Emissions Guidelines (NTZIACHRISTOS; SAMARAS, 2016). This allowed to create a scaled emission factor between the local source of CETESB emission factors and the speed dependent functions from NTZIACHRISTOS; SAMARAS (2016). As result, speed variation could be incorporated into the local emission factors. The (CETESB, 2015) emission factors are average values for the emissions tests FTP75 and WMTC. Each has its own average speed. The scaling process consists in identifying speed functions so that, at the speed of the driving cycles, will return exactly the same value of the local emission factors, (CETESB, 2015) in this study. ESC driving cycle does not have driven speed and instead it has motor speed, so the speed of FTP75, which is  $34.2 km \cdot h^{-1}$ , was used. In the case of WMTC, the speed is  $54.7 km \cdot h^{-1}$  according to the reference report on driving cycles from the Transport and Research Laboratory (TRL) (BARLOW et al., 2009).

The speed functions emission factors of (NTZIACHRISTOS; SAMARAS, 2016)

depend on the emission Euro standard. Therefore, in order to select the right speed function, it is necessary firstly do an equivalence between Brazilian and European Emission standards. Both emission standards are relatively similar but in Brazil they are delayed. Brazilian emission standards are named Proconve for LDV and HDV and Promot for motorcycles (Ministerio de Medio Ambiente, 2011). Proconve L6 is equivalent with Euro V, L3 Euro II, L2 with Euro I and L1 with Pre-Euro. In the case HDV, Proconve P7 is equivalent with Euro V, P5 with Euro III, P4 with Euro II, P3 with Euro I and the rest with Pre-Euro. Motorcycles are directly equivalent, Promot M3 with Euro III, M3 with Euro II, M1 with Euro I. This equivalence was made based on the site <http://www.transportpolicy.net/region/south-america/brazil/> and is shown in Table 10.

To visualize an example of this new scaled factors a plot of the emission factor that depends on age of use and speed is shown as a 3d surface in Fig. (21) and (22). Fig. (21) shows the CO emission factors of PC using E25 as function on speed and includes an horizontal black line when the emission factor is  $33 \text{ g} \cdot \text{km}^{-1}$ . There is also a vertical line at  $34.2 \text{ km} \cdot \text{h}^{-1}$  which is the average speed of the emission test. Therefore, at  $34.2 \text{ km} \cdot \text{h}^{-1}$  the value  $33 \text{ g} \cdot \text{km}^{-1}$  is the same value of (CETESB, 2015) and the values at other speeds belongs to the speeds variation form (NTZIACHRISTOS; SAMARAS, 2016). In other words, the speed function must be multiplied with a factor so that the speed function at  $34.2 \text{ km} \cdot \text{h}^{-1}$  must be  $33 \text{ g} \cdot \text{km}^{-1}$ . VEIN provides functions to automatically scale local factors by age of use if the user only knows the Euro equivalence of each local emission factor. Fig. (22) shows a more general view to see the interaction between emission factors by age of use and speed. The diagonal line shows the speeds ( $\text{km} \cdot \text{h}^{-1}$ ), the horizontal line shows the distribution by age of use and the vertical the emission factors ( $\text{g} \cdot \text{km}^{-1}$ ). At lower speeds and higher age of use of vehicles emission factors are higher. However, at higher age of use and higher speeds the emission factors do not grow so abruptly.

The discussed emission factors were only for hot exhaust conditions without deterioration. The deterioration of vehicles has been studied and it is more notorious in 4-strokes vehicles whose catalytic system is not working or malfunctioning. In this study the deterioration factors from (NTZIACHRISTOS; SAMARAS, 2016) were used, but they are available for 4-strokes vehicles only. These factors depend on the accumulated mileage and Euro standard and are applied to vehicles with catalytic converter only. For vehicles without catalytic converter there is no deterioration according to this method. This method was chosen since there is no Brazilian studies of deterioration factors.

Another type of emissions considered in this study is the evaporative. This type of emissions can be important, they are basically the vaporization of fuel due to variations in ambient temperatures. VEIN count the with (MELLIOS; NTZIACHRISTOS, 2016)' methods for estimating evaporative emissions. The emission factors comes from



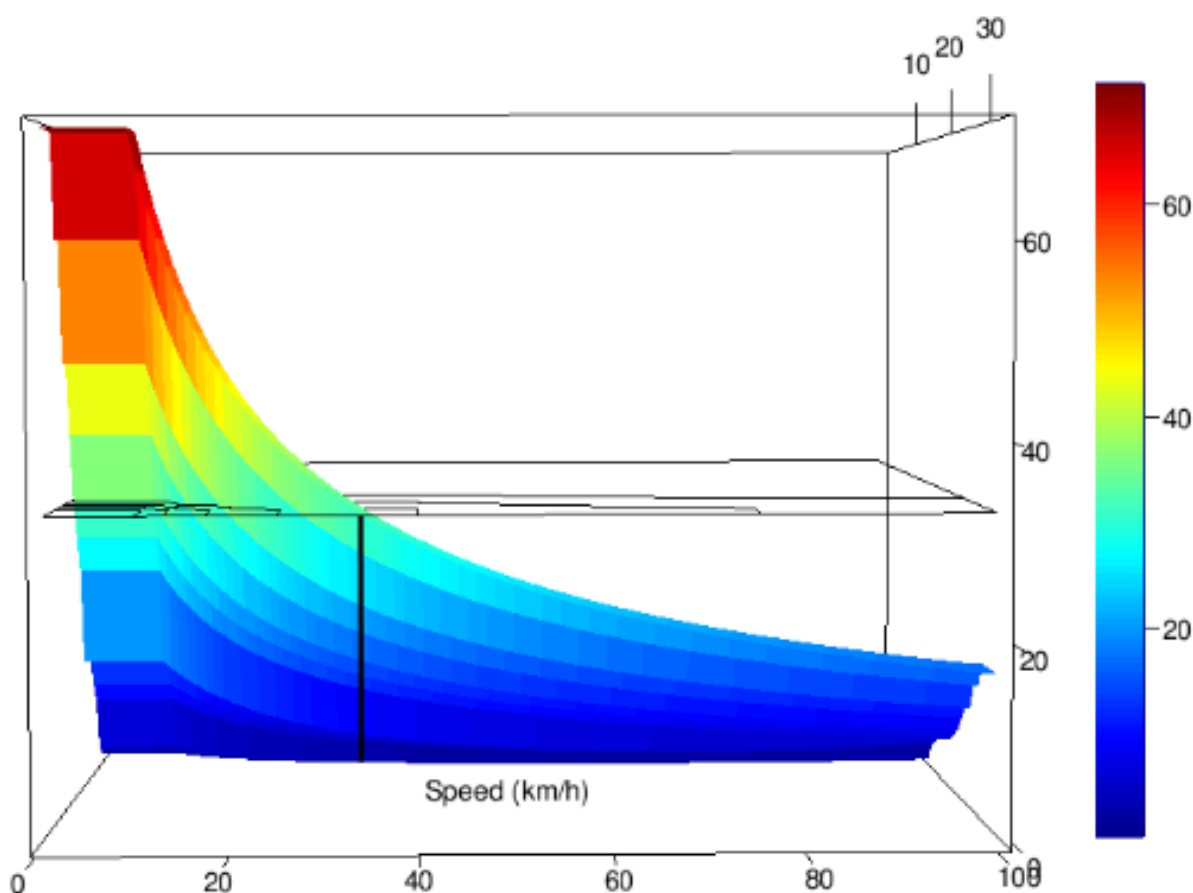


Figure 21 – CO emission factors ( $g \cdot km^{-1}$ ) for PC as a function of speed ( $km \cdot h^{-1}$ ).

(CETESB, 2015) average measurements of evaporative emissions by type of vehicle and year. This database includes factors of diurnal ( $g \cdot day^{-1}$ ), hot\_soak ( $g \cdot trip^{-1}$ ) and running\_losses ( $g \cdot trip^{-1}$ ) evaporative emissions, for three ranges of temperature, between 0 and 10, 10 and 20 and 20 and 35 °C and available for 4-strokes PC and LCV. The emission factors for PC are shown in Fig. (23). The three types of emissions have a similar behavior with less emissions in newer vehicles. Older vehicles with more than 28 years presents higher emissions and when incrementing the temperature, higher evaporative emissions. As the period of study is the week between 2014-10-06 and 2014-10-12, the monthly average temperature for this period, 21 °C, was used to calculate the emission factors between 20 and 35 °C.

To incorporate these emission factors into VEIN, each emission factor was converted into  $g \cdot km^{-1}$ . The emission factors of diurnal evaporative emissions ( $g \cdot day^{-1}$ ) were divided by the daily mileage ( $km \cdot day^{-1}$ ) according to the age of use of the type of vehicle. There are mileage functions inside VEIN, based on odometer readings of more than 1.6 million vehicles (BRUNI; BALES, 2013), these functions return the annual mileage which was divided by 365 to obtain the daily mileage. (BRUNI; BALES, 2013) mileage functions are based on age of use, as VEIN works also with vehicular

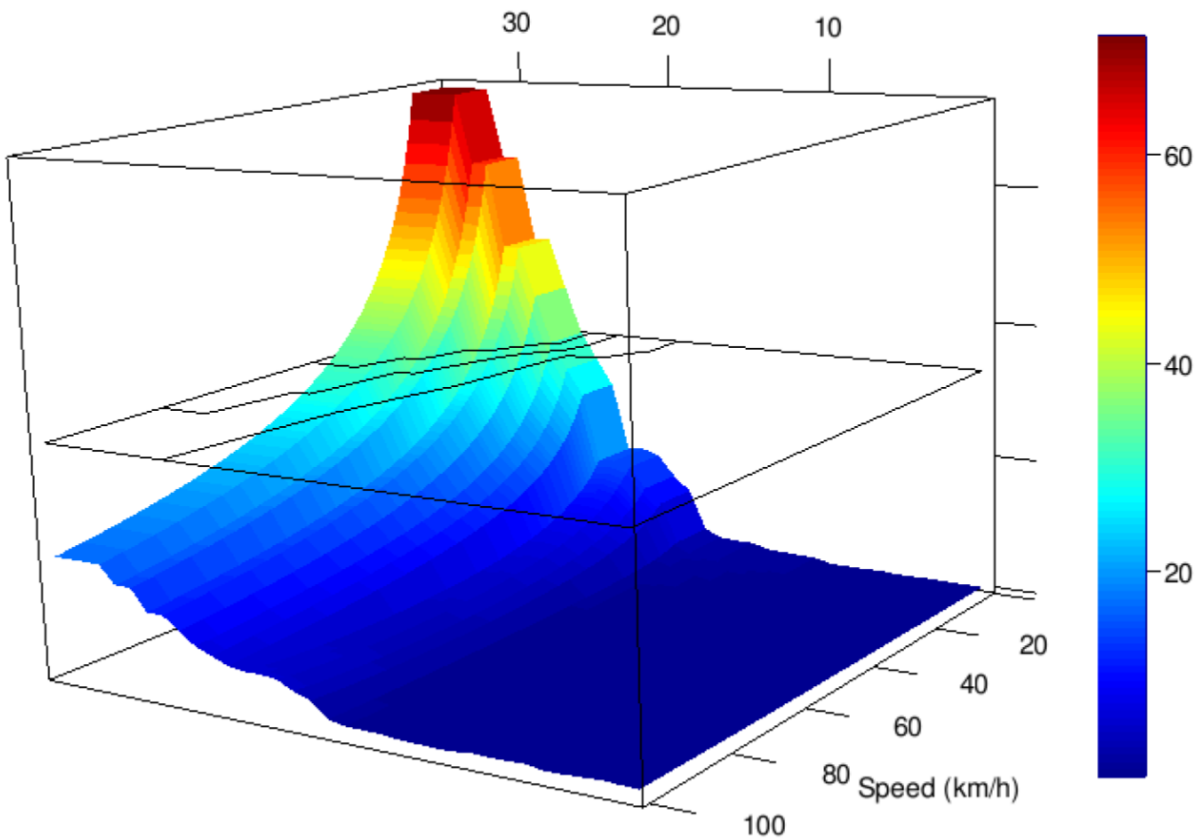


Figure 22 – CO emission factors ( $g \cdot km^{-1}$ ) for PC as a function of speed ( $km \cdot h^{-1}$ ) and age of use.

composition by age of use, the fleet characterization is consistent. The other emission factors in  $g \cdot trip^{-1}$  were converted to  $g \cdot day^{-1}$ , by multiplying by the average number of  $4.6 \text{ trips} \cdot day^{-1}$  and then divided by the daily mileage  $km \cdot day^{-1}$  according to the age of use of the type of vehicle. There are no statistics with the number of trips per day in São Paulo, therefore this value came from the statistics of the (Bureau of Transportation Statistics, 2017), U.S. Department of Transportation.

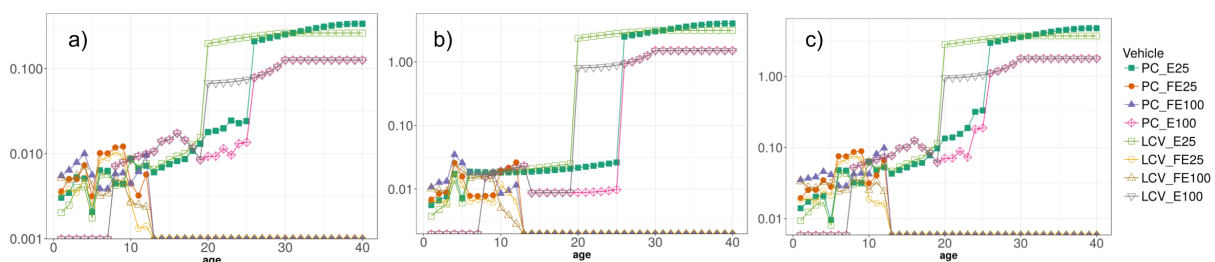


Figure 23 – Evaporative emission factors for a) Diurnal, b) Hot-Soak and c) Running Losses  $g \cdot km^{-1}$  by age of use for average monthly temperature between 20 and 35 Celsius based on (CETESB, 2015).

### Area source emissions

According to (CETESB, 2015) report, evaporative emissions from leak at the moment the car is filled are becoming increasingly important. They even claim that this will be the most important source of hydrocarbons from year 2022. The emission factors proposed by (CETESB, 2015) depend are  $1.14 \text{ g} \cdot \text{l}^{-1}$  for E25 and  $0.37 \text{ g} \cdot \text{l}^{-1}$  for E100. The inventory presented in this study considers fuel consumption with factors in  $\text{g} \cdot \text{km}^{-1}$ , therefore, the fuel consumption factors (in  $\text{g} \cdot \text{km}^{-1}$ ) was divided by the density of each fuel,  $0.754 \text{ t} \cdot \text{m}^{-3}$  for E25 and  $0.809 \text{ t} \cdot \text{m}^{-3}$  for E100. The resulting factor for evaporative leaked emissions are  $1.512\text{e-}03 \text{ g} \cdot \text{km}^{-1}$  for E25 and  $4.57\text{e-}04 \text{ g} \cdot \text{km}^{-1}$  for E100. These emissions are produced in fuel stations but only the spatial distribution for MASP was available, as seen in Fig.(24). For the other regions, it was assumed that these emissions are distributed on streets.

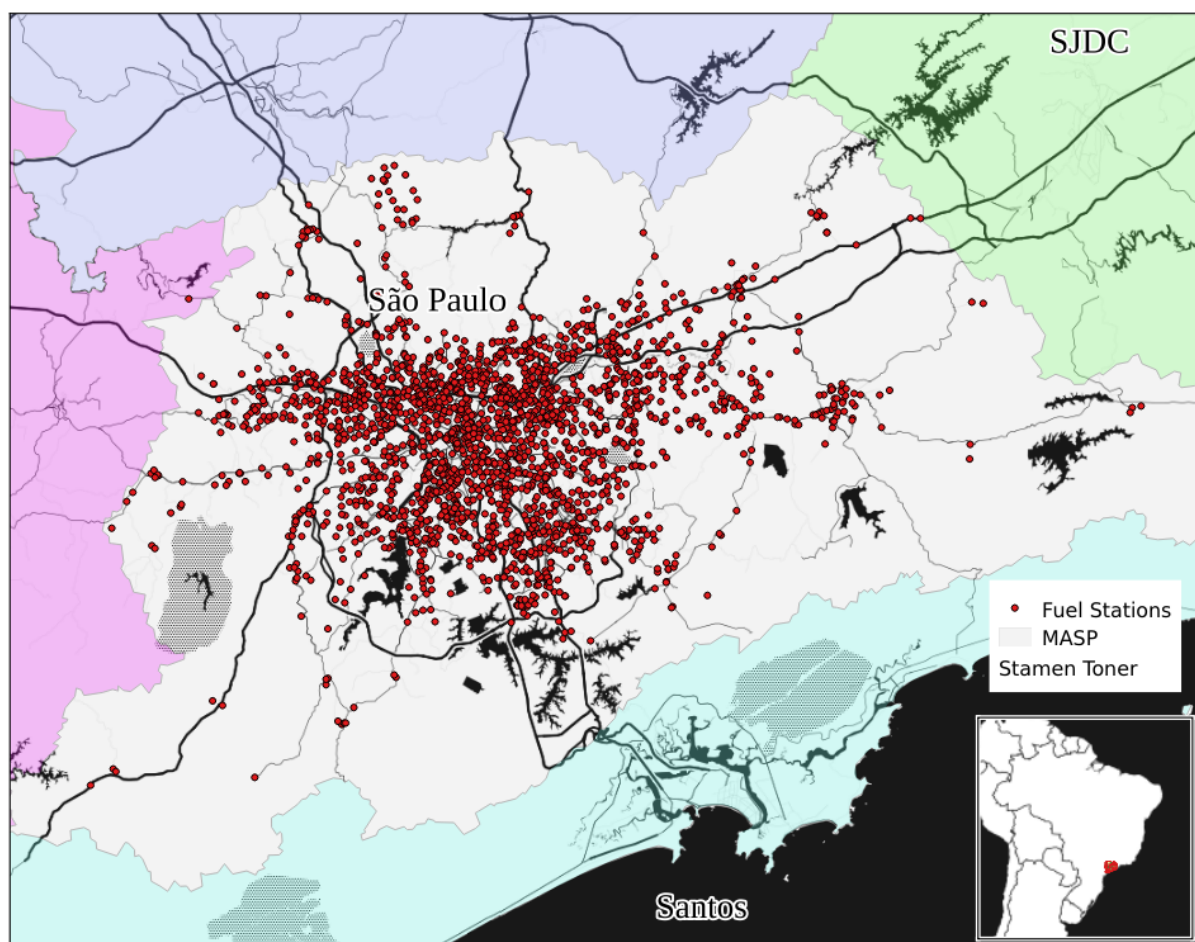


Figure 24 – Fuel stations on MASP, obtained from CETESB.

One of the objectives of this study is to evaluate the impact of the emissions on the air concentrations of several pollutant including  $O_3$ . This is a secondary pollutant that in the troposphere, formed due to concentrations of hydrocarbons, oxides of nitrogen ( $NO_x$ ) and solar radiation. Therefore, it is important to include all relevant anthropogenic

sources of hydrocarbons and  $NO_X$ . According to the European Emission Guidelines, coating applications including painting are important sources of hydrocarbons. Therefore, this source was also included with activity data from the Brazilian Association of Paints Manufacturers (ABRAFATI, 2017). The activity data consists in sales of paints for  $1119 \cdot 10^6 \text{ lt} \cdot \text{year}^{-1}$  for buildings,  $217 \cdot 10^6 \text{ lt} \cdot \text{year}^{-1}$  for cars and industry, and  $61 \cdot 10^6 \text{ lt} \cdot \text{year}^{-1}$  for other, all sales in Brazil during 2014. According to (KUENEN; TROZZI, 2016), emissions factors of NMHC from decorative coating application is  $150 \text{ g} \cdot \text{kg}^{-1}$ , from industrial is  $400 \text{ g} \cdot \text{kg}^{-1}$  and from other is  $200 \text{ g} \cdot \text{kg}^{-1}$ . The densities of painting for buildings is  $1.25 \text{ g} \cdot \text{cm}^{-3}$ , and industry and automotive,  $0.75 \text{ g} \cdot \text{cm}^{-3}$ . The resulting emissions are  $209812.5 \text{ t} \cdot \text{y}^{-1}$  from building paintings and  $65751 \text{ t} \cdot \text{y}^{-1}$  from industry and cars paintings. There is no statistics of the volume of paints sold per region, therefore, it was approximated the regional value with population proportions. The population in Brazil in 2014 was  $202.77 \cdot 10^6$  and for the metropolitan regions covered in this study: MASP  $20.93 \cdot 10^6$ , Campinas  $3.04 \cdot 10^6$ , SJDC  $2.43 \cdot 10^6$ , Sorocaba  $2.06 \cdot 10^6$  and Santos  $1.78 \cdot 10^6$  (IBGE, 2014a). The proportional emissions of NMHC are therefore  $28443.77 \text{ t} \cdot \text{y}^{-1}$  in MASP,  $4131.35 \text{ t} \cdot \text{y}^{-1}$  in Campinas,  $3302.36 \text{ t} \cdot \text{y}^{-1}$  in SJDC,  $2799.53 \text{ t} \cdot \text{y}^{-1}$  in Sorocaba and  $2419.01 \text{ t} \cdot \text{y}^{-1}$  in Santos. The spatial distribution of these emissions was allocated in non-motorway streets in each region.

### Speciation of NMHC

Once all the required data was gathered, the exhaust and evaporative emissions were estimated. The method for estimating evaporative emissions is top down, which implies that total emissions per day are generated. As hourly emissions are needed the evaporative emissions were distributed between 06:00 and 16:00LT when, according to (MELLIOS; NTZIACHRISTOS, 2016), most of evaporative emissions are produced due to increasing ambient temperature.

The emissions estimates were aggregated into amount of pollutant per each street and then in gridded emissions with a 1 km x 1 km horizontal resolution. This grid spacing was chosen in order to take the advantage of the high resolution emissions inventories. After estimating all the emissions of NMHC, it was used the speciation presented in (RAFEE et al., 2017). This speciation is based on tunnel measurements in São Paulo, depends on the type of fuel (E25, E100 and B5) and provides the mass of each chemical compound as  $\text{mol} \cdot (100\text{g})^{-1}$ . This speciation splits the NMHC from evaporative, liquid and exhaust emissions of E25, E100 and B5, into minimum compounds required for the Carbon Bond Mechanism (CBMZ) (ZAVERI; PETERS, 1999). The compounds presented in (RAFEE et al., 2017) speciation are: Alkanes (Ethane (ETH), Alkanes with HO rate constant between  $2.7 \cdot 10^{-13}$  and  $3.4 \cdot 10^{-12}$  (HC3), Alkanes with HO rate constant between  $3.4 \cdot 10^{-12}$  and  $6.8 \cdot 10^{-12}$  (HC5) and Alkanes with HO rate constant bigger than  $6.8 \cdot 10^{-12}$  (HC8)); Alkenes (Ethene (OL2), Terminal

Alkenes (OLT), Internal Alkenes (OLI) and Isoprene (ISO)); Aromatics (Toluene and less reactive aromatics (TOL) and Xylene and more reactive aromatics (XYL)); Carbonyls (Formaldehyde (HCHO), Acetaldehyde and higher aldehydes (ALD)); and ethanol (C<sub>2</sub>H<sub>5</sub>OH). All of these hydrocarbons are *O*<sub>3</sub> precursors. A common measure of the reactivity of hydrocarbons for producing *O*<sub>3</sub> is the Maximum Incremental Reactivity (MIR). (CARTER, 1994) presented a study with the MIR for several species of hydrocarbons based on scenarios of environmental conditions of cities of United States of America. In order to produce a simulation adjusted for São Paulo conditions, the speciation of (RAFEE et al., 2017) was updated in order to represent the average MIR value for Brazilian fuels E25 and E100: 4.7 and 3.2, respectively. The MIR values for the fuels were obtained with a Gas Chromatography with Flame Ionization Detector (GC-FID) analysis made by (Branco G. and Branco F., 2016). The MIR values for the atmospheric species were obtained by (MARTINS, 2006) and in the case of *CH*<sub>3</sub>*OH* and *KET*, average MIR values come from (CARTER, 2009). The MIR weighted mean between for E25 exhaust NMHC is 3.8 in (RAFEE et al., 2017) which is below the E25 MIR of 4.7. Therefore, the NMHC emission factors were multiplied by 1.7 when their MIR was below 4.7. In the case of E100 exhaust, the weighted mean is 3.44 which shows that these speciation is appropriate for the fuel burnt in São Paulo. The adjusted speciation and MIR for the CBMZ mechanism are shown in table 7. The emission factors of the NMHC compounds are expressed as *mol* · (100*g*)<sup>-1</sup> of fuel from exhaust, evaporative or liquid emissions. The chemical composition of fuel presents differences in various countries, therefore, this method could be used in different regions when there is no more available information.

The NMHC produced by painting activity was speciated using EPA tool Speciate (SIMON et al., 2010). Specifically, the activity paintings were selected and calculated an average based on technical reports and papers of NMHC measurements, such as (SEXTON; WESTBERG, 1980), where NMHC and other pollutants were measured downwind a plant of car paintings. USEPA-Speciate counts with several profiles for paintings. The profile numbers 2546, 4661 and 4662 were used for Industrial painting emissions and profile numbers 7106, 7113, 7125, 7133, 7139, 7141, 7163 and 7173 for building painting. Then an average percentage for each compound was calculated and normalized. Most of pollutants are included in CBMZ speciation such as Toluene or Xylenes, but other were aggregated as alkanes using the categorization from VOC Reactivity Scales (CARTER, 2000) for CBMZ.

The chemical mechanism CBMZ needs more pollutants and most of them are covered in normal VEIN estimation based on (CETESB, 2015) and (NTZIACHRISTOS; SAMARAS, 2016) emission factors: Methane (*CH*<sub>4</sub>), Carbon Monoxide (*CO*), Carbon Dioxide (*CO*<sub>2</sub>), Total Hydrocarbons (HC), Nitrous Oxide (*N*<sub>2</sub>*O*), Ammonia (*NH*<sub>3</sub>), Non-Methanic Hydrocarbons (*NMHC*), Nitric Oxide (*NO*), Nitrogen Dioxide (*NO*<sub>2</sub>),

Table 7 – Speciation for NMHC by type of fuel and process  $mol \cdot 100g^{-1}$ . MIR values from (MARTINS, 2006) and speciation values from (RAFEE et al., 2017)

NMHC	MIR	NMHC with E25			NMHC with E100		NMHC with B5	Painting	
		Exhaust	Evap	Liq	Exhaust	Evap & Liquid	Exhaust	Buildings	Industry
<i>ETH</i>	0.28	0.2826	0.2317	0.0250	0	0	0	0	0
<i>HC<sub>3</sub></i>	1.33	0.4352	0.3567	0.2400	0	0	0.0490	0	0.0006
<i>HC<sub>5</sub></i>	1.58	0.1586	0.1300	0.4500	0	0	0.0577	0.5660	0.2198
<i>HC<sub>8</sub></i>	1.01	0.0765	0.0627	0	0.0490	0	0.2967	0	0
<i>OL<sub>2</sub></i>	9.07	0.5807	0.2800	0.0382	0.1260	0	0.3189	0	0
<i>OLT</i>	8.91	0.2434	0.1174	0.200	0.0018	0	0.3853	0	0
<i>OLI</i>	1.94	0.1614	0.1323	0.4600	0.0013	0	0	0	0
<i>ISO</i>	10.68	0.0077	0.0037	0	0	0	0	0	0
<i>TOL</i>	2.5	0.1405	0.1152	0.0850	0.0050	0	0.2351	0.4534	0.0821
<i>XYL</i>	8.15	0.2677	0.1291	0	0.0149	0	0.0084	0.1805	0.0409
<i>C<sub>2</sub>H<sub>5</sub>OH</i>	1.69	0.3082	0.3082	0.3500	1.5226	2.1706	0	0	0
<i>HCHO</i>	6.98	0.1127	0.0508	0	0.3605	0	0.3225	0	0
<i>ALD</i>	8.96	0.0864	0.0663	0	0.1317	0	0.0751	0	0.0246
<i>CH<sub>3</sub>OH</i>	0.65	0	0	0	0	0	0	0.2561	0.1436
<i>KET</i>	1.65	0	0	0	0	0	0	0.0818	0.2558

Particulate Matter from engine combustion (*PM*) and Sulfur Dioxide (*SO<sub>2</sub>*). Atmospheric simulations using the same pollutants in Brazil have resulted in good agreement with observations (ANDRADE et al., 2015).

### 3.5 Partial Conclusions

This section described 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 for 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. A recent study used VEIN estimates with a grid of 10 m resolution to determine the relationship between vehicular emissions and birth outcomes in the western area of São Paulo (FINK et al., 2017). 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 package 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. 17, 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 degree, approximately 12 km and it is possible to have 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 meso-scale interactions, including feedbacks. However, with the computational advances, it would be possible to perform air pollution modeling for larger domains with 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 one hour of traffic data for each street considered in the estimation. This data can be based on traffic simulation or traffic counts, however, most of cities do not count with 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 smart-phones such as Waze (c) or Uber (c), 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>) cover several countries and this 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 spatial dependencies into the new spatial features **sf** package. This package provides S3 classes for handling spatial data faster than its predecessor, the package **sp** (PEBESMA; 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-MARTINEZ 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; 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 into these functions, such as those in the EMEP/EEA guidelines (NTZIACHRISTOS; SAMARAS, 2016). In the case of Brazil, there are several studies of tropospheric ozone, which use speciation of VOC emissions as an input (VARA-VELA et al., 2016; RAFEE et al., 2017).





## 4 Weather and Research Forecast model with Chemistry (WRF-Chem)

WRF-Chem is a fully coupled online meteorological and chemical transport model (GRELL et al., 2005), available at <http://www.wrf-model.org>. This model has been widely used in different parts of the world including Brazil (ANDRADE et al., 2015; VARA-VELA et al., 2016; RAFEE et al., 2017; HOSHYARIPOUR et al., 2016).

### 4.1 Model configuration

The objective of the present study is to compare and evaluate different bottom-up emissions inventories. Consequently, the model configuration of WRF-Chem must be set in order to take fully advantage of the detailed and high resolution emissions fluxes. Some previous sensitivity studies regarding model horizontal resolution show that not always a higher resolution will imply in better agreement with observations. For instance, a study of ozone formation in Mexico compared observation with simulations with model resolutions of 3, 6, 12 and 24 km of grid spacing (TIE; BRASSEUR; YING, 2010). Although their analyses were based solely on two days of simulation, they concluded that better results were obtained with 6 km of grid spacing. The authors, however, did not evaluate the impact of the emissions resolution, since their emissions were only spatially aggregated. It remained unclear if the methods adopted to calculate the emissions inventory with different resolutions would imply in a better agreement with observations. Another study more focused on the planetary boundary layer (PBL) schemes also analyzed WRF-Chem performance at 12 and 4 km of resolution (MISENIS; ZHANG, 2010). Their results indicate the need of improving WRF-Chem for higher resolutions. Therefore, although it could be made a 1 km grid resolution for emissions, uncertainties in other physical parameterizations could jeopardize the analysis of using different inventories.

The model configuration consisted in three domains shown in Fig. (25). Only the inner domain included the emissions in a grid of 3 km of grid spacing, with 85 grid points in the zonal direction and 96 points in the meridional direction, centered at the city of São Paulo, with latitude -23.550391 and longitude -46.633949. Fig. (1) shows that the Santos Metropolitan Area has lower elevations because it is a coastal region, but there is an abrupt increase in elevations at the Serra do Mar, between 600 and 900 m above sea level. The vertical grid consisted in 35 levels with the top at 50 *hPa*. The outer domains were run for meteorological fields only, with 9 and 27 km of grid

spacing. The initial and boundary meteorological conditions were obtained from the National Center for Environmental Prediction's Final Operational Global Analysis with 0.5 degrees of grid spacing, 26 vertical levels, every 6 hours: 00:00, 06:00, 12:00, and 18:00 UTC from <https://rda.ucar.edu/datasets/ds335.0/>. The set of parameterizations for different physical process is shown in Table (8). It was included cumulus parametrization only in outer domains.

The gas-phase chemistry mechanism used was Carbon Bond Mechanism (CBMZ), as previously mentioned. Regarding the chemical boundary conditions, according the GAVIDIA-CALDERON's thesis, including chemical dynamic boundary conditions not necessary will improve  $O_3$  simulations at superficial level in São Paulo. Therefore, it was used the idealized profile for default chemical boundary conditions of WRF-Chem (LIU et al., 1996).

## WPS Domain Configuration

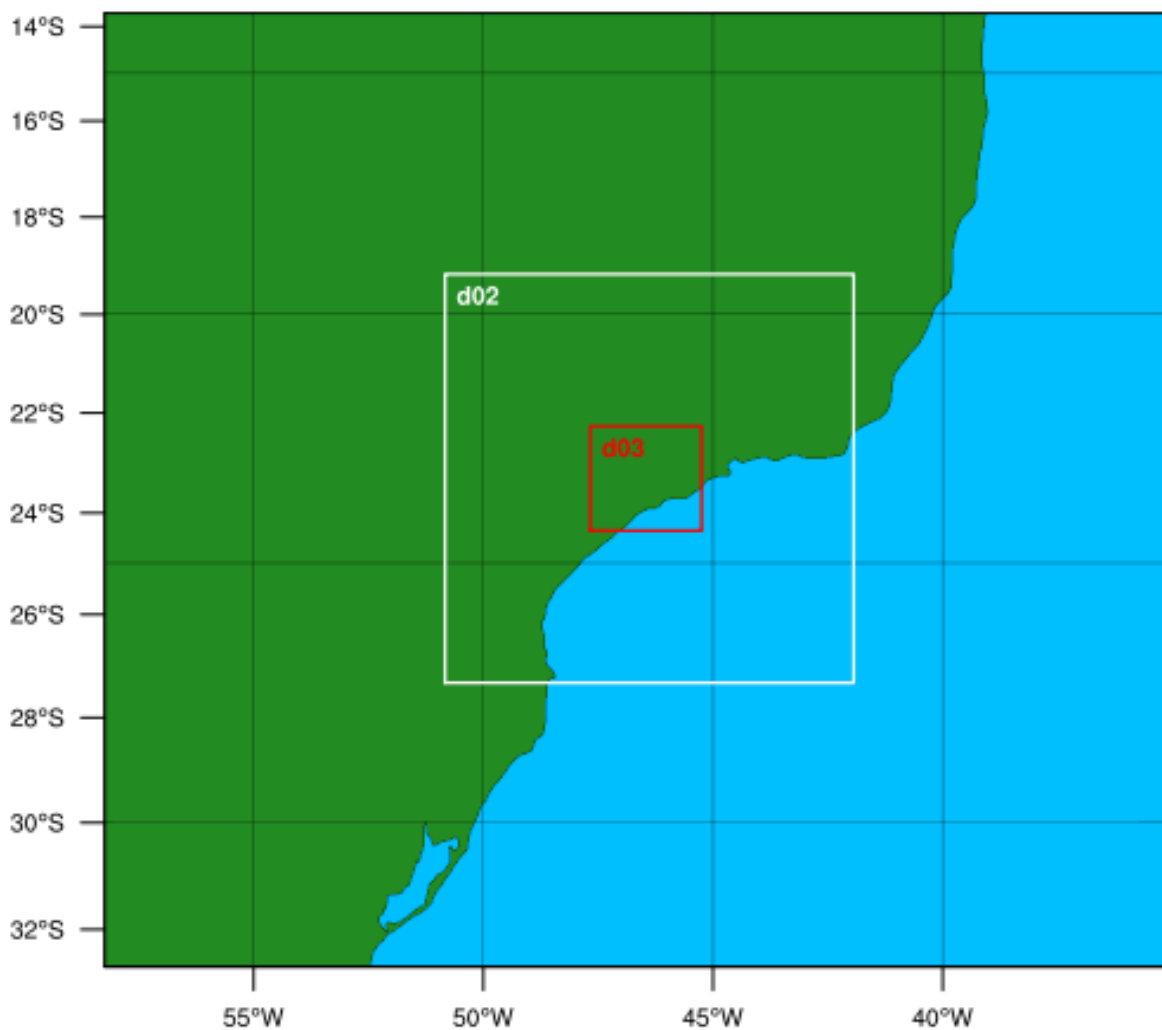


Figure 25 – Domains of 27, 9 and 3 km of grid spacing.

Table 8 – Parameterizations used in WRF-Chem

Physics options	Parametrization
Microphysics	Lin (LIN; FARLEY; ORVILLE, 1983)
Boundary Layer	YSU (HONG; NOH; DUDHIA, 2006)
Cumulus	(GRELL; DÉVÉNYI, 2002)
Shortwave	Goddard (CHOU; SUAREZ, 1999)
Longwave	RRTM (MLAWER et al., 1997)
Land surface	Noah (TEWARI et al., 2004)
Surface layer	MM5 (ZHANG; ANTHES, 1982)
Urban Surface	UCM (CHEN et al., 2011a)
Gas-phase chemistry	CBMZ (ZAVERI; PETERS, 1999)
Photolysis	Fast-J (FAST et al., 2006)

## 4.2 Meteorological conditions

In order to understand air pollution concentration with simulations, it is important to choose periods of time according to the objective of the study. In this case, it was of interest evaluating the emissions, so it is important to check if there was precipitation in the smaller domain during the period of study. During October 2014 there was a severe drought in Southeast Brazil (COELHO et al., 2016) which produced risk in water availability for different uses. According to the Group of Climatic Studies (GREC, 2014) of the Universidade de São Paulo (<http://www.grec.iag.usp.br/>), the climatological mean of precipitation for October in South and Southeast Brazil are 107 *mm* and 173 *mm* respectively. However, during October of 2014, there were anomalies of -51.7 and -56.7 *mm* in each region. Regarding the synoptic conditions, the climatic bulletin of the Center for Weather Forecasting and Climate Research (CPTEC) synoptic analyses (CPTEC, 2014) show that for almost the whole period of simulation there was anticyclone with center on the Atlantic Ocean which was blocking southern fronts from reaching São Paulo, as shown in Fig. (27). The only day with presence of clouds was Tuesday after a weak front reached São Paulo. The closest precipitation system near São Paulo was a stationary front acting in south of Rio Grande do Sul, the Southeast state of Brazil between 12 and 13 of October of 2014, with a distance of approximately 1000 km of São Paulo. In addition, it was investigated the presence of precipitation with data of rain gauge collectors of the National Institute of Meteorology located in São Paulo. Fig. (26) shows daily accumulated precipitation of the station Mirante de Santana, located at latitude -23.49 and longitude -46.62. It can be seen that there was no precipitation during the study period, shown between red lines on Fig (26).

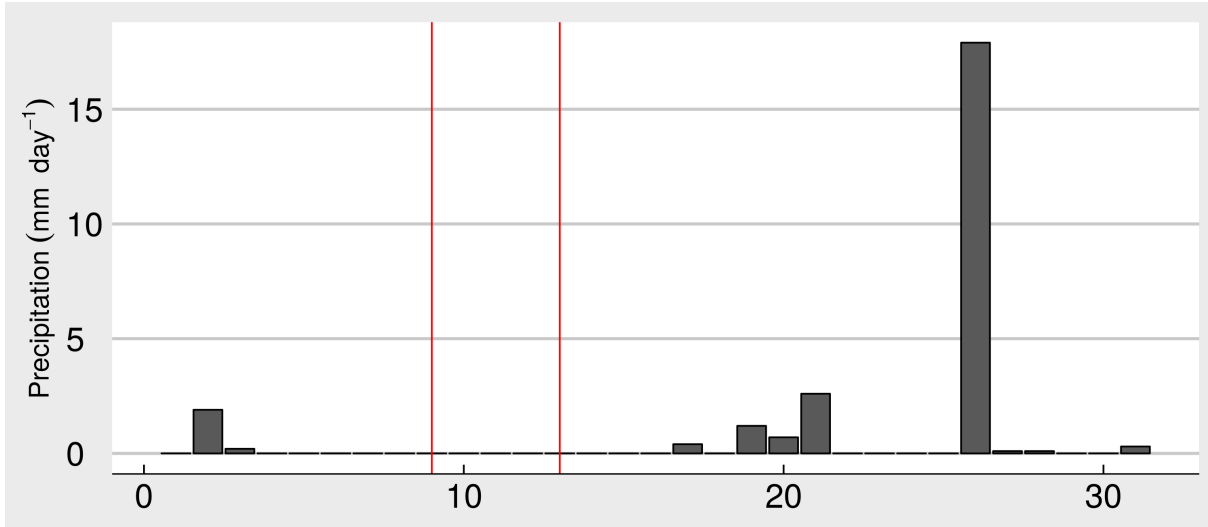


Figure 26 – Hourly accumulated precipitation at Mirante de Santana station, October 2014, (INMET, 2016)

### 4.3 Evaluation

In order to evaluate the performance of the atmospheric and air quality simulations, it was first made a visual comparison between observed and simulated values. Then, it was calculated the statistics Mean bias (MB), Mean fractional bias (MFB), Mean fractional error (MFE), Unbiased root mean square error (RMSE) and Correlation coefficient (R). The evaluation is made by pairs of simulations from WRF-Chem ( $M_i$ ) and observations ( $O_i$ ) from the air quality and meteorology network stations from CETESB, shown on Table 9. This table shows the column Reg which means Regions 1: Campinas, 2: MASP, 3: Santos, 4: SJDC and 5: Sorocaba.

$$MB = \frac{1}{n} \sum_{i=1}^n (M_i - O_i) \quad (4.1)$$

$$MFB = \frac{1}{n} \sum_{i=1}^n \frac{2 * (M_i - O_i)}{M_i + O_i} * 100\% \quad (4.2)$$

$$MFE = \frac{1}{n} \sum_{i=1}^n \frac{2 * |M_i - O_i|}{M_i + O_i} * 100\% \quad (4.3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(M_i - \bar{M}_i) - (O_i - \bar{O})]^2} \quad (4.4)$$

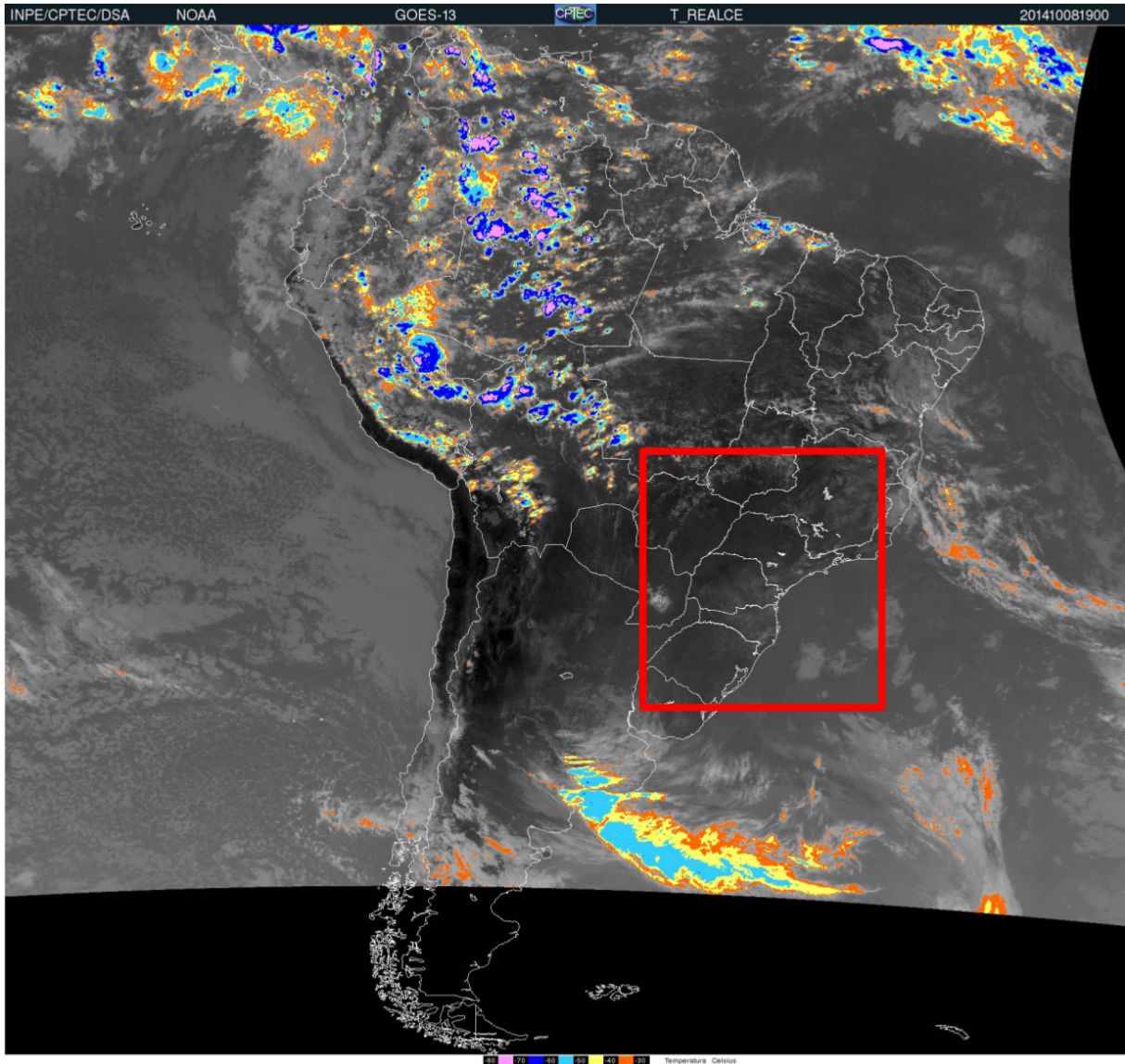


Figure 27 – GOES 13 satellite IR image for October 8th, at 15:00 LT.

$$r = \frac{\sum_{i=1}^N (M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N [(M_i - \bar{M}) - (O_i - \bar{O})]^2}} \quad (4.5)$$

Lastly, it was compared simulations at observations sites to see if there is significant difference between scenarios. Therefore, it was used the Wilcoxon (BAUER, 1972) tests inside the package "stats" of R language programming (R Core Team, 2017).

Table 9 – Air quality and meteorology stations from CETESB used in this study

Station	Latitude	Longitude	Reg	Parameters
Americana	-22.715	-47.339	1	<i>PM<sub>10</sub>, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Paulinia	-22.763	-47.155		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Campinas	-22.902	-47.057		<i>PM<sub>10</sub>, CO, UR, TEMP</i>
Jundiai	-23.192	-46.897		<i>PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Cerqueira Cesar	-23.553	-46.672	2	<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO</i>
Congonhas	-23.616	-46.663		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO</i>
Ibirapuera	-23.591	-46.660		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Sao Caetano do sul	-23.609	-46.556		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Pedro II	-23.545	-46.628		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Parelheiros	-23.776	-46.698		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP</i>
Pinheiros	-23.561	-46.702		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Taboa da Serra	-23.600	-46.758		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, UR, TEMP</i>
Interlagos	-23.671	-46.675		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Carapicuiaba	-23.522	-46.836		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Guarulhos Municipal	-23.446	-46.518		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Capao Redondo	-23.659	-46.780		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Ponte Remedios	-23.509	-46.743		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Itaim Paulista	-23.502	-46.421		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Capuava	-23.630	-46.492		<i>PM<sub>10</sub>, SO<sub>2</sub>, O<sub>3</sub>, WS, WD</i>
Mooca	-23.550	-46.600		<i>PM<sub>10</sub>, CO, O<sub>3</sub>, WS, WD</i>
Nossa Senhora de O	-23.480	-46.692		<i>PM<sub>10</sub>, O<sub>3</sub>, UR, TEMP</i>
Santana	-23.506	-46.629		<i>PM<sub>10</sub>, O<sub>3</sub>, UR, TEMP</i>
Santo Amaro	-23.646	-46.710		<i>PM<sub>2.5</sub>, O<sub>3</sub>, WS, WD</i>
Diadema	-23.686	-46.612		<i>PM<sub>10</sub>, O<sub>3</sub></i>
Osasco	-23.527	-46.792	<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, WS, WD</i>	
Santo Andre	-23.640	-46.492	<i>PM<sub>10</sub>, CO, WS, WD</i>	
Maua	-23.569	-46.465	<i>PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub></i>	
Vila Parisi	-23.840	-46.389	3	<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, WS, WD</i>
Cubatao Centro	-23.879	-46.418		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Santos	-23.963	-46.321		<i>PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub>, UR, TEMP, WS, WD</i>
Ponta Praia	-23.972	-46.300		<i>PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub>, UR, TEMP, WS, WD,</i>
Jacarei	-23.294	-45.968	4	<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, CO, O<sub>3</sub>, UR, TEMP, WS, WD</i>
SJDC	-23.188	-45.871		<i>PM<sub>10</sub>, SO<sub>2</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub>, UR, TEMP, WS, WD, BEN, TOL</i>
Sorocaba	-23.502	-47.479	5	<i>PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>X</sub>, O<sub>3</sub>, UR, TEMP, WS, WD</i>

## 5 Results

In urban centers, vehicles are the most important source of air pollution (MOLINA; MOLINA, 2004b). In the mega-city of São Paulo, in Southeast of Brazil, there were 7940713 vehicles on June 2017, with 70 % as Passenger Cars (PC) according to data from the Brazilian Department of Transit (DENATRAN, 2017). Besides, the city of São Paulo is part of a bigger conurbation called Metropolitan Region of São Paulo (MASP) located near other 4 metropolitan areas: Baixada Santista (hereafter identified as Santos), Vale do Paraíba (in this work will be identified as the name of its main city São José Dos Campos - SJDC), Sorocaba and Campinas. Altogether these 5 metropolitan areas accounted for more than 30 million of inhabitants at year 2016 according the Brazilian Institute of Geography and Statistics (IBGE, 2016). Emissions inventories made for these regions show that vehicles are the most important source of air pollution (CETESB, 2015) remarking the importance of a correct characterization of the emission of these pollutants, which would help to evaluate impacts on the climate, ecosystem and population exposure.

Different approaches have been used to generate emissions inputs for air quality models for decades generating good results. For example, in Santiago, Chile, a city with steep terrain and stable nocturnal conditions, air quality simulations using traffic simulations among other input data could forecast critical episodes of air pollution (SAIDE et al., 2011). However, the results are still subject to the quality of the input data. In other words, air quality simulations depend on the quality of the data used to generate the emissions fluxes.

In this study four types of traffic data described in Chapter 2 were used in developing bottom-up vehicular emissions inventories with VEIN, which produced inputs for WRF-Chem model using the Assimilation System for WRF (AS4WRF) presented by (VARA-VELA et al., 2016). This system reads the gridded emissions and create WRF-Chem input files for the desired domain.

This chapter presents the results from the different traffic data sources, in section 5.1. From these traffic flows, 3 emissions scenarios were designed. Section 5.2 shows the emissions inventories from VEIN for the 3 scenarios and for the week from the 06th to the 12th of October, 2014. The gridded emissions for these scenarios produced by VEIN were then used as input for WRF-CHEM, and the air quality simulation results are discussed in section 5.3. Final remarks are given in section 5.4.

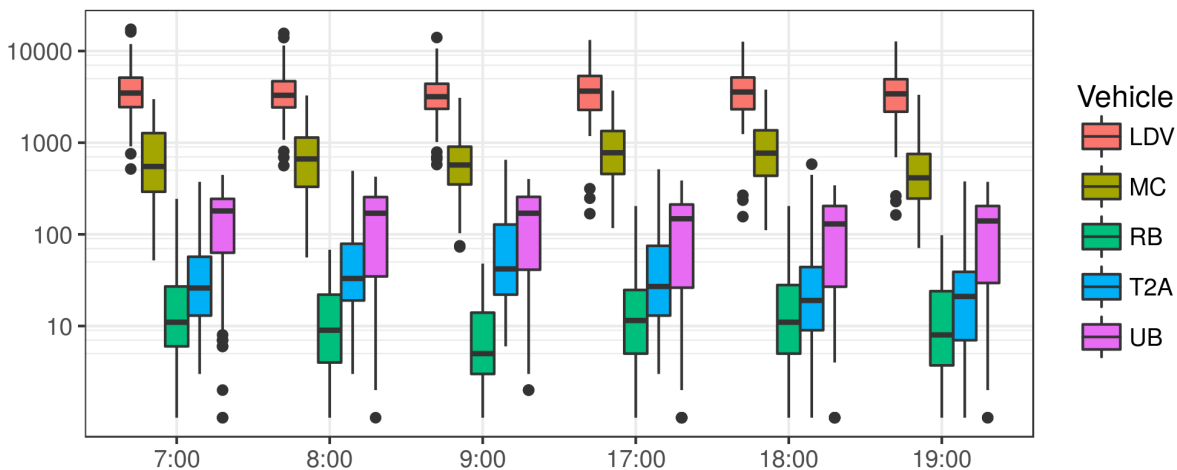


Figure 28 – Box-plots for traffic counts.

## 5.1 Traffic data

### 5.1.1 Traffic flow from interpolation

The dataset consisted of 113 points counting traffic data in both ways. The total amount of vehicles at all hours was 3266323 with 78.2 % LDV, 2.82 % UB, 0.36 % RB , 1.25 % T2A, 0.47 % T3A, 0.25 % T4A and 16.63 % MC. T3A had 86 traffic counts with zeros, representing 76.6% of all hours. T4A had 85 traffic counts with zero, representing 79.4%. There is a restriction in circulation that affects trucks with 3 and 4 axes at morning and evening rush hours at city of São Paulo (<http://www.cetsp.com.br/consultas/rodizio-municipal/como-funciona.aspx>). Since these categories represent less than 1% of traffic, they were not considered for further analyses.

Figure 28 shows histograms in logarithmic scale for each type of vehicle at different hours. Each box-plot does not show high variation for each type of vehicle at different hours. The highest median is approximately  $4000 \text{ veh} \cdot \text{h}^{-1}$  for LDV, then  $500 \text{ veh} \cdot \text{h}^{-1}$  for MC,  $150 \text{ veh} \cdot \text{h}^{-1}$  for UB,  $10 - 50 \text{ veh} \cdot \text{h}^{-1}$  for T2A and  $10 \text{ veh} \cdot \text{h}^{-1}$  for RB. LDV does not show upper outliers at evening times, contrary to morning times. UB shows a marked median near the quantile 75.

As previously discussed, traffic counts were geo-referenced and plotted over an OpenStreetMap road network. Table 10 shows a summary of statistics about the traffic counts for LDV, UB, RB, T2A and MC by type of street at hours 7:00, 8:00, 9:00, 17:00, 18:00 and 19:00. Trunk roads have more vehicles than Primary, Secondary and Tertiary, except for UB, with more vehicles in Secondary roads. This is because urban buses must follow planned routes collecting passengers at residential neighborhoods. All types of vehicles show great variability at each hour.

The box plots of the traffic counts by type of street in Figure (29 a) show that



Table 10 – Summary of statistics for traffic counts by type of street.

Vehicle	Highway	7:00	8:00	9:00	17:00	18:00	19:00
LDV	Trunk	5303	4859	4548	5303	5041	4995
	Primary	3764	3450	3292	3796	3703	3530
	Secondary	2085	2037	1794	1736	1733	1397
	Tertiary	1553	1202	1795	1946	2006	1873
	mean	4193	3855	3631	4168	4014	3877
	sd	2801	2527	2132	2517	2312	2437
UB	Trunk	124	130	126	111	103	108
	Primary	177	183	183	164	149	152
	Secondary	200	203	203	134	128	125
	Tertiary	127	117	140	116	104	110
	mean	155	160	159	135	125	128
	sd	118	120	118	107	99	102
RB	Trunk	33	18	11	29	29	22
	Primary	18	13	9	19	19	13
	Secondary	12	5	3	8	8	7
	Tertiary	2	2	4	2	4	1
	mean	24	14	9	22	22	16
	sd	38	15	10	34	33	20
T2A	Motorway[*]	246	379	305	449	349	272
	Trunk	71	97	141	81	60	53
	Primary	28	37	79	28	21	17
	Secondary	29	38	53	103	93	60
	Tertiary	20	10	16	6	6	1
	mean	62	98	124	103	80	64
sd	112	169	171	209	182	127	
MC	Trunk	1118	1047	891	1189	1173	639
	Primary	726	814	719	927	908	497
	Secondary	227	277	292	884	805	613
	Tertiary	124	192	261	274	238	130
	mean	826	837	732	1027	1001	569
	sd	693	679	561	734	740	473

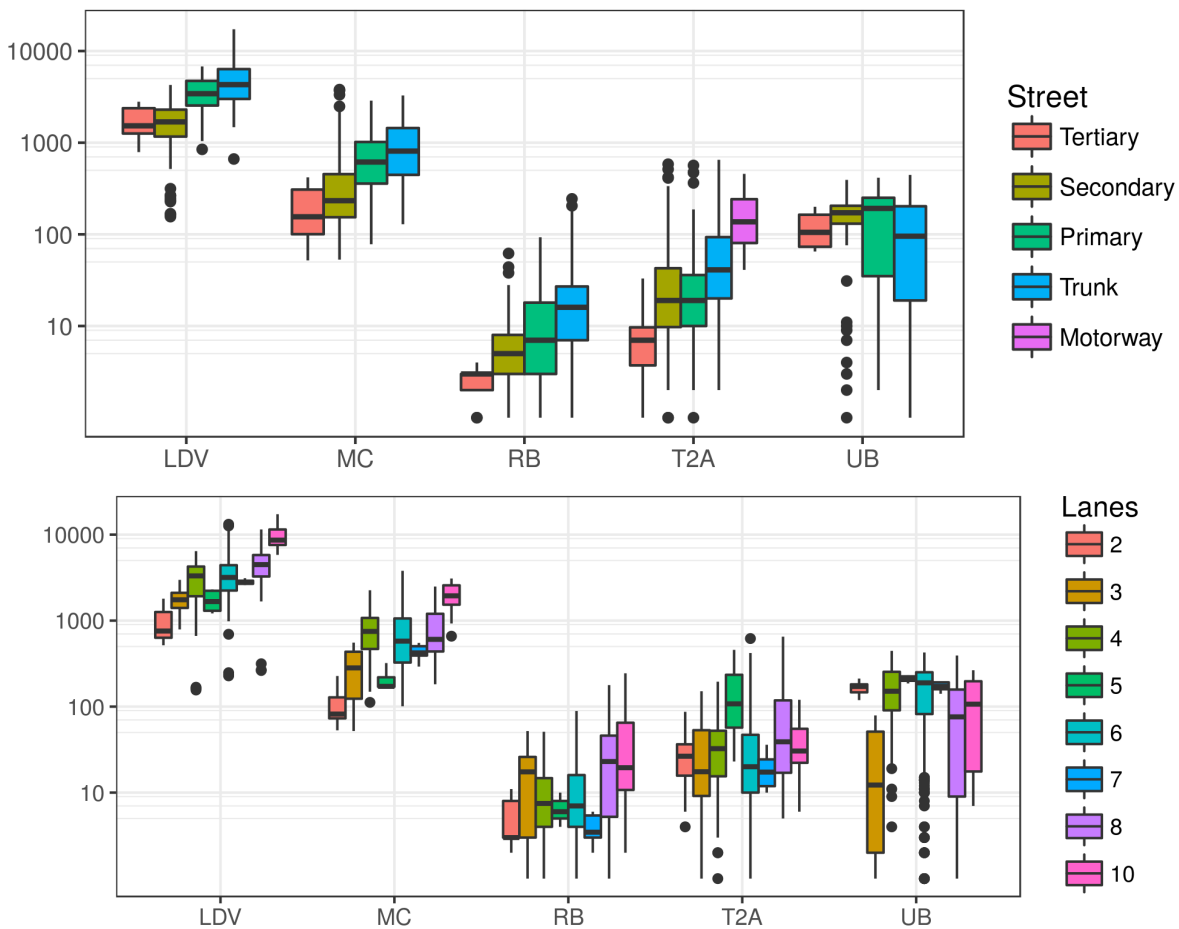


Figure 29 – Box-plots for traffic counts of vehicles ( $veh \cdot h^{-1}$ ) for (a) type of streets and (b) number of lanes.

the median values increase as the hierarchy of the streets increases (from Tertiary to Trunk). This observation is clear for all types of vehicles except for UB, which shows higher median at Primary roads.

Type of street  $ts$  is a categorical variable and, after some tests, it was assigned the following values: Motorway = 0, Trunk = 1, Primary = 2, Secondary = 3, Tertiary = 4 and Residential = 20. These values considered the fact that Trunk roads and Motorway roads should have a higher amount of vehicles. These assumptions were made after analyzing Figure 29a, where there is a clear tendency to have more vehicles with higher road hierarchy. The road type, Residential, was not considered in the traffic counts. This presents a problem for the regressions with type of street because there are many residential types of streets in the city road network. After testing different values, it was observed that using values of 5 or 10 could overestimate the vehicles in the regression. Therefore, this study used the value of 20 for the type of road Residential.

The number of lanes  $lan$  in each traffic count were also identified. This number refers to the total number of lanes in both ways. This variable is present in Open-

StreetMap data. Capacity is the maximum amount of vehicles that can circulate on a street. A street with more lanes will have higher capacity than another with fewer lanes. Figure 29b shows a box-plot for traffic counts and lanes. There is a clear tendency to have more vehicles with more lanes, especially for LDV, MC RB, and T2A. Median values for UB vary around 100 vehicles for lanes 4 to 10. UB have defined routes and frequency, which is reflected in the amount of traffic per lanes.

It was tested regressions for each vehicle types(LDV, UB, RB, T2A and MC) for the hours 07:00, 08:00, 09:00, 17:00, 18:00 and 19:00. The total amount of regressions was 168. In this section the best the regressions were chosen such that: 1) all predictors that were statistically significant; 2) produced least AIC; and 3) the highest correlation between observed and modeled outputs. In general, the most significant variables were type of street (*ts*) and number of lanes (*lan*).

LDV regressions are shown in Table 11 with the coefficient (Coef,  $\beta$  from Equation 2.1), estimate (Est), standard error (SE), z value and p-value  $P(z)$  to test the null hypothesis that the coefficient is 0. The quasi-poisson approach performed better than the negative binomial at all hours. The maximum correlation obtained was 0.72 for morning rush hours and the lowest was 0.67 at 18:00. For most hours, the inclusion of the variable distance from center *dfc* obtained better results than the geographical coordinates *lat* and *lon*. The exception is for 18:00 when geographical coordinates were more significant. All variables were statistically significant. When analyzing the temporal tendencies, the study obtained better results for morning rush hours including higher correlations with more significant predictors.

UB regressions are shown in Table 12. As with LDV, the quasi-poisson approach performed better than negative binomial at all hours. The maximum correlation obtained was 0.54 at 07:00 and the lowest was 0.43 at 17:00. Contrary to LDV, here geographical coordinates appears to achieve better results than *dfc*. This could be due to the fact that UB is a fleet with planned frequency involving variables other than solely geographical aspects.

The results for RB regressions obtained correlations below 0.5. The maximum correlation obtained was only 0.46 at 08:00, but the lowest was 0.41 at 18:00. Since this category represents only 0.36 % of traffic counts, it was discarded from the analysis.

The T2A regressions obtained correlations up to 0.73 and 0.74 for morning rush hours 07:00 and 08:00. However, the correlations fall below 0.5 at 18:00. When considering correlations, all regressions show that quasi-poisson performed better than the negative binomial approach. The morning rush hour may have more accurate results due to the fact that this category relates to small trucks delivering goods inside the city and these vehicles are not subject to restriction applied to T3A and T4A. This allows the circulation of a greater number of vehicles. Regarding the geographical approach,

Table 11 – Regressions coefficients for LDV.

Hour	Coef	Est	SE	z	P(z)	COR
07:00	Int	8.00	0.25	31.67	0.00	0.72
	ts	-0.33	0.07	-4.77	0.00	
	lan	0.17	0.03	6.66	0.00	
	dfc	-0.04	0.01	-3.65	0.00	
08:00	Int	8.09	0.25	32.79	0.00	0.72
	ts	-0.35	0.07	-5.10	0.00	
	lan	0.15	0.02	6.20	0.00	
	dfc	-0.04	0.01	-4.47	0.00	
09:00	Int	8.03	0.23	35.47	0.00	0.72
	ts	-0.33	0.06	-5.31	0.00	
	lan	0.14	0.02	6.17	0.00	
	dfc	-0.03	0.01	-3.80	0.00	
17:00	Int	8.27	0.25	33.37	0.00	0.69
	ts	-0.37	0.07	-5.43	0.00	
	lan	0.13	0.03	5.18	0.00	
	dfc	-0.03	0.01	-2.71	0.01	
18:00	Int	65.57	25.79	2.54	0.01	0.67
	ts	-0.31	0.06	-4.71	0.00	
	lan	0.14	0.03	5.71	0.00	
	lat	2.45	1.10	2.24	0.03	
19:00	Int	8.13	0.26	31.10	0.00	0.70
	ts	-0.39	0.07	-5.37	0.00	
	lan	0.14	0.03	5.35	0.00	
	dfc	-0.02	0.01	-2.35	0.02	

distance from the center achieved better results in all hours except at 09:00.

Despite the relatively high amount of motorcycles in the dataset, the results shown in Table 14 were worst than LDV. It is interesting to notice that negative binomial distribution appears to produce better results than quasi-poisson, except at 08:00. Therefore, this study compared results with both distributions. The highest correlation was obtained at 09:00 with both distributions, 0.59 negative binomial and 0.53 quasi-poisson. Furthermore, with negative binomial distribution all coefficients were statistically significant, and with quasi-Poisson only *ts* and *lan* were significant. In general, the negative binomial results achieved more accurate results than quasi-poisson for almost all hours, except 08:00, where quasi-poisson correlation was 0.45 and negative binomial was 0.44. Nevertheless, the correlation values are still low. There is no significant difference in using one distribution or another.

Regarding the spatial approach, only *dfc* was statistically significant at 09:00. These results suggest that motorcycles follow a pattern more related to the road

Table 12 – Regressions coefficients for UB.

Hour	Coef	Est	SE	z	P(z)	COR
07:00	Int	-303.46	59.52	-5.10	0.00	0.54
	ts	0.23	0.09	2.65	0.01	
	lon	-4.64	1.51	-3.07	0.00	
	lat	-3.88	1.85	-2.10	0.04	
08:00	Int	-271.67	61.23	-4.44	0.00	0.48
	ts	0.18	0.09	2.05	0.04	
	lon	-5.93	1.31	-4.52	0.00	
09:00	Int	-268.49	60.30	-4.45	0.00	0.50
	ts	0.20	0.08	2.38	0.02	
	lon	-5.86	1.29	-4.53	0.00	
17:00	Int	-260.51	62.25	-4.18	0.00	0.43
	lon	-5.69	1.33	-4.26	0.00	
18:00	Int	-274.84	59.80	-4.60	0.00	0.50
	lon	-4.00	1.52	-2.63	0.01	
	lat	-3.94	1.87	-2.11	0.04	
19:00	Int	-288.51	59.62	-4.84	0.00	0.52
	lon	-4.13	1.52	-2.71	0.01	
	lat	-4.27	1.85	-2.31	0.02	

Table 13 – Regression coefficients for T2A.

Hour	Coef	Est	SE	z	P(z)	COR
07:00	Int	1.84	0.38	4.85	0.00	0.73
	lan	0.16	0.05	3.24	0.00	
	dfc	0.12	0.01	9.33	0.00	
08:00	Int	2.03	0.39	5.23	0.00	0.74
	lan	0.16	0.05	3.32	0.00	
	dfc	0.13	0.01	9.66	0.00	
09:00	Int	298.37	51.33	5.81	0.00	0.54
	lon	6.28	1.10	5.70	0.00	
	lat	-0.55	0.15	-3.69	0.00	
17:00	Int	2.12	0.59	3.59	0.00	0.60
	lan	0.15	0.08	1.99	0.05	
	dfc	0.13	0.02	6.25	0.00	
18:00	Int	2.87	0.32	8.83	0.00	0.49
	lan	0.12	0.02	5.00	0.00	
19:00	Int	1.21	0.58	2.09	0.04	0.70
	lan	0.20	0.07	2.69	0.01	
	dfc	0.14	0.02	7.27	0.00	

Table 14 – Regression coefficients for MC with Binomial Negative and Quasi-Poisson distributions.

Hour	Coef	Negative Binomial					Quasi-Poisson				
		Est	SE	z	P(z)	COR	Est	SE	z	P(z)	COR
07:00	Int	7.11	0.36	19.52	0.00		7.66	0.18	43.23	0.00	
	ts	-0.65	0.10	-6.73	0.00	0.47	-0.60	0.11	-5.29	0.00	0.44
	lan	0.10	0.04	2.26	0.02						
08:00	Int	6.73	0.35	19.34	0.00		6.59	0.37	17.89	0.00	
	ts	-0.47	0.09	-5.08	0.00	0.44	-0.37	0.11	-3.36	0.00	0.45
	lan	0.12	0.04	2.84	0.00		0.11	0.04	2.76	0.01	
09:00	Int	6.12	0.34	17.74	0.00		5.94	0.34	17.39	0.00	
	ts	-0.34	0.08	-4.23	0.00		-0.26	0.10	-2.63	0.01	
	lan	0.16	0.03	4.54	0.00	0.59	0.16	0.04	4.41	0.00	0.53
	dfc	-0.03	0.01	-2.21	0.03						
	dtr	0.00	0.00	1.97	0.05						
17:00	Int	6.00	0.23	25.63	0.00		5.94	0.26	23.04	0.00	
	lan	0.15	0.04	4.02	0.00	0.40	0.15	0.04	4.11	0.00	0.40
18:00	Int	6.51	0.35	18.86	0.00		6.05	0.27	22.18	0.00	
	ts	-0.20	0.09	-2.16	0.03	0.35	0.13	0.04	3.35	0.00	0.33
	lan	0.11	0.04	2.77	0.01						
19:00	Int	5.57	0.26	21.70	0.00		5.56	0.31	17.89	0.00	
	lan	0.12	0.04	3.02	0.00	0.27	0.12	0.05	2.68	0.01	0.27

characteristics than the spatial characteristics. At this time, only length of street segment *dtr* appears to be statistically significant. However, the value of the coefficient was 0.0005 and since the tables are showing rounded numbers with two digits, it appears as zero, denoting a lower significance of this variable. Therefore, this suggests that *dtr* is not significant for most hours. It must be considered that there will be difficulties when modeling with the type of vehicle due to limitations in data access, especially in Brazil. In Brazil, motorcycles are allowed to share lanes with other types of vehicles. As a consequence, they reach higher speeds and it is expected that they would prefer streets with higher numbers of lanes. This aspect makes modeling motorcycles a challenging task.

In order to obtain an idea of the power of prediction for the regressions, it was estimated the amount of vehicles to the OSM road network of the Metropolitan Area of São Paulo (MASP). Despite the limitations of the data, it is important to test if this method could be applied to a mega-city. The resulting amount of vehicles was later adjusted with fuel sales. The sold mass for each type of fuel was 3604133 *t* for gasohol, 1778825 *t* for ethanol and 2842849 *t* for diesel.

The initial amount of vehicles circulating over the road network was estimated to

be:  $LDV = 22.3 \cdot 10^6$ ,  $T2A = 10.91 \cdot 10^6$ ,  $UB = 1.9 \cdot 10^6$  and  $MC = 3.7 \cdot 10^6$ . These numbers were calculated as the sum of vehicles over the road network. The amount of vehicles were classified according to the official vehicular emissions inventory for São Paulo City (CETESB, 2013a). This report provides the fuel consumption by type of vehicle. The VEIN model was used to estimate fuel consumption. The estimated consumed fuel for 2012 was: 1582597 t of gasohol, 1444188 t of ethanol and 4208099 t of diesel. The ratios between the consumed fuel estimated by VEIN and fuel sales were 0.55 for gasohol, 0.81 for ethanol and 1.48 for diesel. These discrepancies indicate that vehicles which consume gasohol and ethanol are sub-estimated and vehicles that consume diesel are over-estimated. Therefore, it was necessary to re-estimate the fuel multiplying vehicles that consume gasohol by 2, ethanol by 1.2 and diesel by 0.5. After running the model, the new values of fuel consumption were 3165194 t for gasohol, 1733026 t for ethanol and 2129871 t for diesel. The ratios between the fuel estimated by VEIN and fuel sales were 0.88 for gasohol, 0.97 for ethanol and 0.75 for diesel. These new ratios seem more plausible and realistic. It is not expected to have a ratio equal to 1 because some fuel is consumed for purposes other than automotive. Moreover, some vehicles may purchase fuel inside MASP but consume the fuel outside of this region, especially in the case of trucks. On the other hand, it is possible that vehicles circulate inside MASP with fuel purchased from outside of MASP. Therefore, under these circumstances the new ratios are satisfactory.

The new amount of vehicles circulating over the road network was estimated to be:  $LDV = 36.78 \cdot 10^6$ ,  $T2A = 5.45 \cdot 10^6$ ,  $UB = 0.92 \cdot 10^6$  and  $MC = 7.11 \cdot 10^6$ .

The road network with the traffic interpolation is shown in Figure 30 for selected hours which produced better results. On panel (a), LDV shows higher concentrations of vehicles towards the center of the network and LDV is concentrated on main roads. On panel (b), UB shows spatial bias to the western part of the city. This could be due to *lon* being the variable with the highest value, -4.64 at 07:00 while the others have lower values. On panel (c), T2A shows overestimation towards the easternmost and western part of the network. Figure 5 shows traffic counts with higher values at the east and west sides. It also seems that the regressions over-predicted traffic at these regions. On panel (d), MC shows a distribution similar to LDV on panel (a), but with lower values. It seems that after all the analyses, LDV and MC were more accurately predicted than the other types of vehicles.

#### 5.1.1.1 Discussion

The dataset of traffic counts consisted of 107 points located inside the city and 6 toll stations near the city which added trucks, resulting in a dataset of 113 points in total. Most of traffic in circulation is due LDV with 78.2%, then MC with 16.63% and the

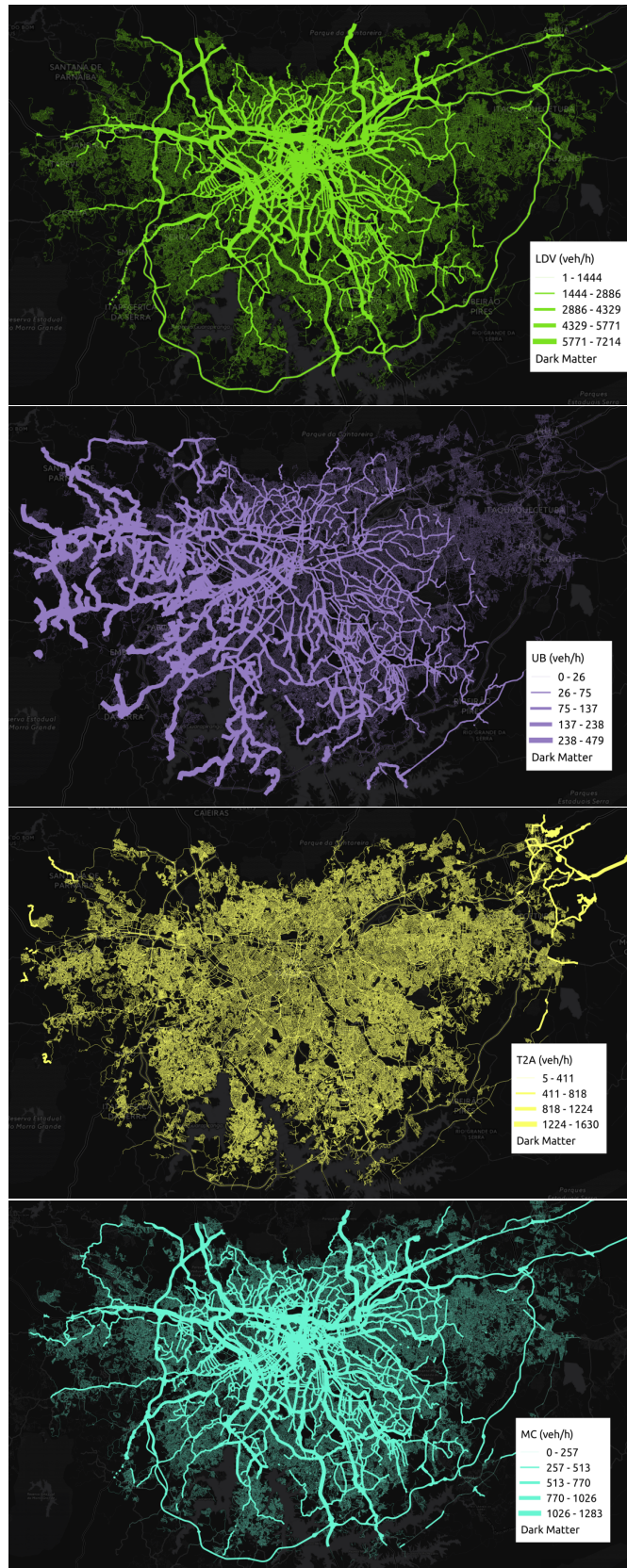


Figure 30 – Traffic count interpolation for (a) LDV, (b) UB, (c) T2A and (d) MC



remainder of vehicles with only 5.17%.

Each type of vehicle was associated with a specific set of variables, at almost all hours. For example, the statistically significant variables for LDV were  $ts$ ,  $lan$  and  $dfc$  for all hours, except for 18:00. The study obtained highest correlation values at morning rush hours with very similar values to each coefficient. This means that LDV traffic is very homogeneous at morning rush hours.

A spatial pattern of high numbers of light vehicles were observed towards the center of the city with trucks observed towards the outer regions of the city. This pattern is due to the fact that the variable  $dfc$  is significant in all regressions of LDV, T2A, and also in some hours of MC. It can be expected that this pattern could be replicated in other cities with similar characteristics as São Paulo. Among the characteristics of São Paulo is that it is a large urban center with a small percentage of heavy vehicles. This can differ to coastal cities which usually have ports for transport of goods by trucks.

The variable  $dfc$  implicitly relates to the coordinates of the mid-point for each link at a defined center of the city. Hypothetically speaking, in any town there is a center with denser traffic. However, the strength of this variable was very low compared with type of street  $ts$  and number of lanes  $lan$ . For example, Table 11 shows that  $dfc$  of LDV is -0.04, but the other variables  $ts$  and  $lan$  have higher magnitude values of -0.33 and 0.17 respectively. In some cases, the explicit use of geographical coordinates achieved better results, with UB for example. In other situations, nor  $dfc$  or geographical coordinates were significant. For instance, quasi-poisson MC at 09:00 was associated only with intercept  $Int$ ,  $ts$  and  $lan$ . Furthermore, it would be interesting to know if  $dfc$  is significant in other cities.

In the case of UB, the variable  $lon$  was included in all hours. The results also show that  $ts$  is significant at morning rush hours. However, in the afternoon the variable  $lat$  was significant at 18:00 and 19:00. Therefore, it can be interpreted that this method is not adequate to properly model UB, especially considering the low values of correlation. Urban Buses is a type of vehicle which has a planned frequency and routes, as a consequence its modeling requires other information.

The category of trucks T2A showed that, in most hours, the only significant variables were  $lan$  and  $dfc$ . The only hour with different variables was 09:00 with  $lat$  and  $lon$ , and also with lower correlation. Higher correlations were found with the other types of vehicles. Figure 5 (b) helps to understand that distance from center  $dfc$  was present at almost all hours, due to the high values at the eastern and western parts of the city.

Motorcycles are the only type of vehicle that presented higher correlation with negative binomial than quasi-poisson. In general, negative binomial regressions include more variables. On the other hand, results with quasi-poisson show that  $ts$  and  $lan$  are

present in many hours. Regressions shows that generally geographical aspects do not play an significant role with the exception of results at 09:00 where  $dfc$  was significant. At this hour the highest correlation found was only 0.59.

Spatial interpolation on traffic performed better for LDV and MC, as shown on panel (a) and (d) of Figure 30.

For example, "Trucks" on the Santos Region are important because this city is a port, attracting and generating several "trucks" trips. Traffic data from GPS in Santos identified streets with high number of "Trucks" near the port and also, in the streets at south est of the study area. It was corroborated this findings with Google image satellites (not shown here). Traffic speeds from GPS was used for generating traffic flow based on a set of traffic and speed measurements for MASP during 2012, (IBARRA-ESPINOSA, 2017). Unequivocally, updating and enhancing this information with new data sets could improve traffic flow generation. Therefore, traffic data from Internet GPS recordings is a promising new traffic data source for emissions inventories.

### 5.1.2 Traffic flow from GPS

Traffic flow speed provides uncertainty for generating traffic flow because it depend on the quality and frequency of GPS recordings. The main difficulty with this approach resided in the frequency of consecutive observations for the same vehicle, which restricted the analyses. The three types of vehicles identified "Cars", "Taxi" and "Trucks", and the vehicles "Undefined" provided information for calculating speeds for each type of vehicle and for all. The speed later served as proxy for obtaining the traffic flow.

The process of generating Trucks consisted in using the speed of Trucks and when there is no available information, the speed of all vehicles only for *motorways* and *trunks*. This process seems appropriate for "Cars" but in the case of "Trucks" could lead to assignment of in the same streets of "Cars" which could not necessary be truth. Circulation of "Trucks" deliver goods and this determine the route they follow. The initial traffic flow including the spatial bias is shown in Fig. (31). The maximum amount of vehicles per lane varies between one and two  $veh \cdot h^{-1}$  and the traffic is concentrated in just some streets. The spatial bias occurs because there are more GPS recordings concentrated in some parts of the city. However, after correcting with speeds as proxy, the traffic flow looks more like reality as shown in Fig. (32). Now the traffic has reached a volume that exceeds the 15000  $veh \cdot h^{-1}$ . Also, the traffic is condensed in urban *motorways* near the center of the city. Considering the spatial distribution of the traffic, it seems that the spatial bias was removed. In addition, the amount of vehicles per lane also seems reasonable.

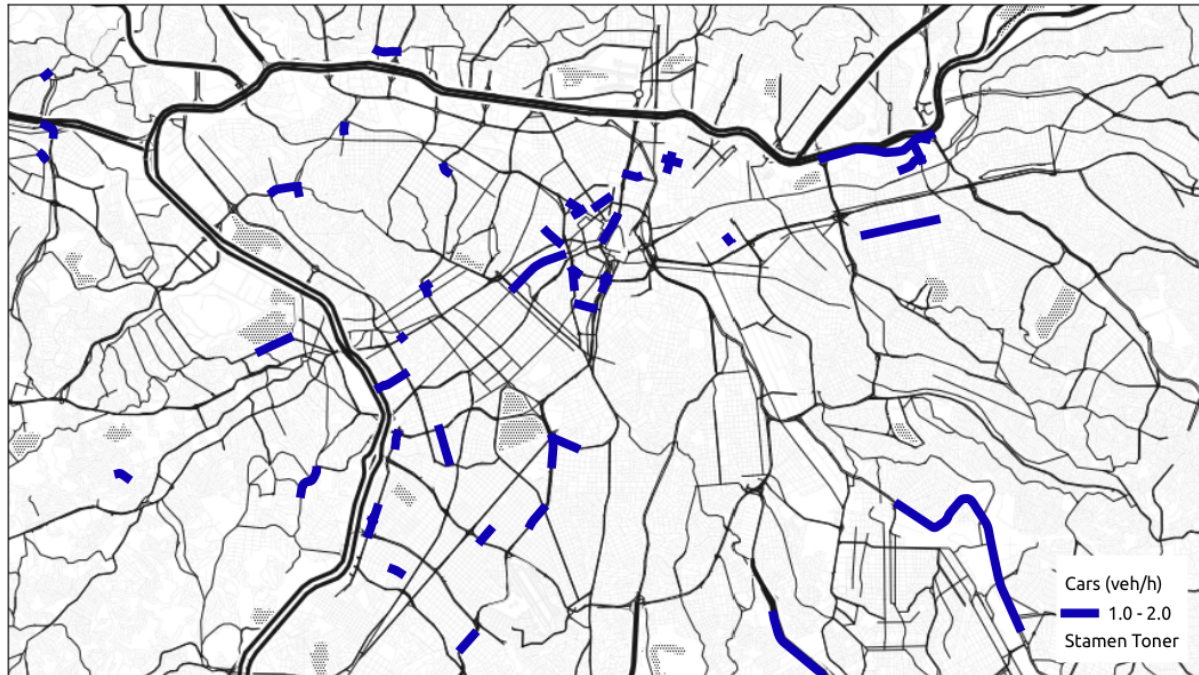


Figure 31 – Traffic flow of Cars generated from GPS data ( $veh \cdot h^{-1}$  per lane).

It was presented these aspects in Fig. (31) and in Fig. (32) for the center of the city of São Paulo so that it can be seen very clearly the spatial bias and also, the effect of its correction. The resulting traffic flow for all the metropolitan areas is shown in Fig. (33) for Cars and Fig. (34) for Trucks. The resulting traffic flow of Cars is concentrated in main roads near the center of each region. For instance, in center of Campinas, São Paulo and Santos traffic is more density. There are streets where traffic from Cars can overpass  $15000 \text{ veh} \cdot h^{-1}$ . However, the total amount of Cars is  $18413811 \text{ veh} \cdot h^{-1}$  and as the number of streets with Cars is 8459, there are on average  $2176.8 \text{ veh} \cdot h^{-1} \cdot street^{-1}$ . In the case of Trucks, they are concentrated in motorways. The volume can overpass  $200 \text{ veh} \cdot h^{-1}$  in most congested streets. The total amount of Trucks is  $296584.4 \text{ veh} \cdot h^{-1}$  and as the number of streets with Trucks is 2774, the average number per street is  $106.9 \text{ veh} \cdot h^{-1} \cdot street^{-1}$ .

### 5.1.3 Traffic flow from simulations

The traffic from traffic simulation from CET contained LDV and HGV simulation for morning rush hour, and simulation from SPtrans Urban buses. Each simulation have an uncertainty in the exact number of vehicles circulating. CET performs traffic counts to calibrate traffic simulations (IBARRA-ESPINOSA, 2017) in the mains routes at MASP. Both simulations include a limited number of local streets as virtual links located at the centroid of each zone of Origin-Destination. These links represent real traffic over virtual links, therefore, provides an associated element of uncertainty. However, most of traffic



Figure 32 – Traffic flow of Cars generated from GPS data with spatial correction using speed as proxy ( $veh \cdot h^{-1}$ ).

occurs in main streets, such as *motorways*, *trunks* and *primary*.

Traffic simulations from CET and SPtrans is data that is ready to be processed by the model VEIN, and therefore, does not need a description here. More details in section 2.3.

#### 5.1.4 Emissions scenarios

The hourly traffic flow generated using GPS recordings was for Cars (LDV) and Trucks (HGV) for the period between 2014-10-05 00:00 and 2014-10-11 00:00 UTC. The initial data set of 145 hours was filled with the same initial 23 hours (Sunday) to add up to 168 hours of a week of data. The spatial coverage includes all the metropolitan areas for this study. The traffic simulations from CET and SPtrans includes one morning rush traffic data for a typical working day of 2014, including volume of LDV, HGV and Urban Buses (UB) only for MASP. The traffic interpolation includes volume of LDV and MC using count data of 2012 but it was assumed that it could represent 2014. Traffic interpolation was made only for MASP. With this information it was configured the scenarios that cover all regions:

- A Traffic from GPS records in all regions. MASP with traffic simulation of UB from SPtrans.



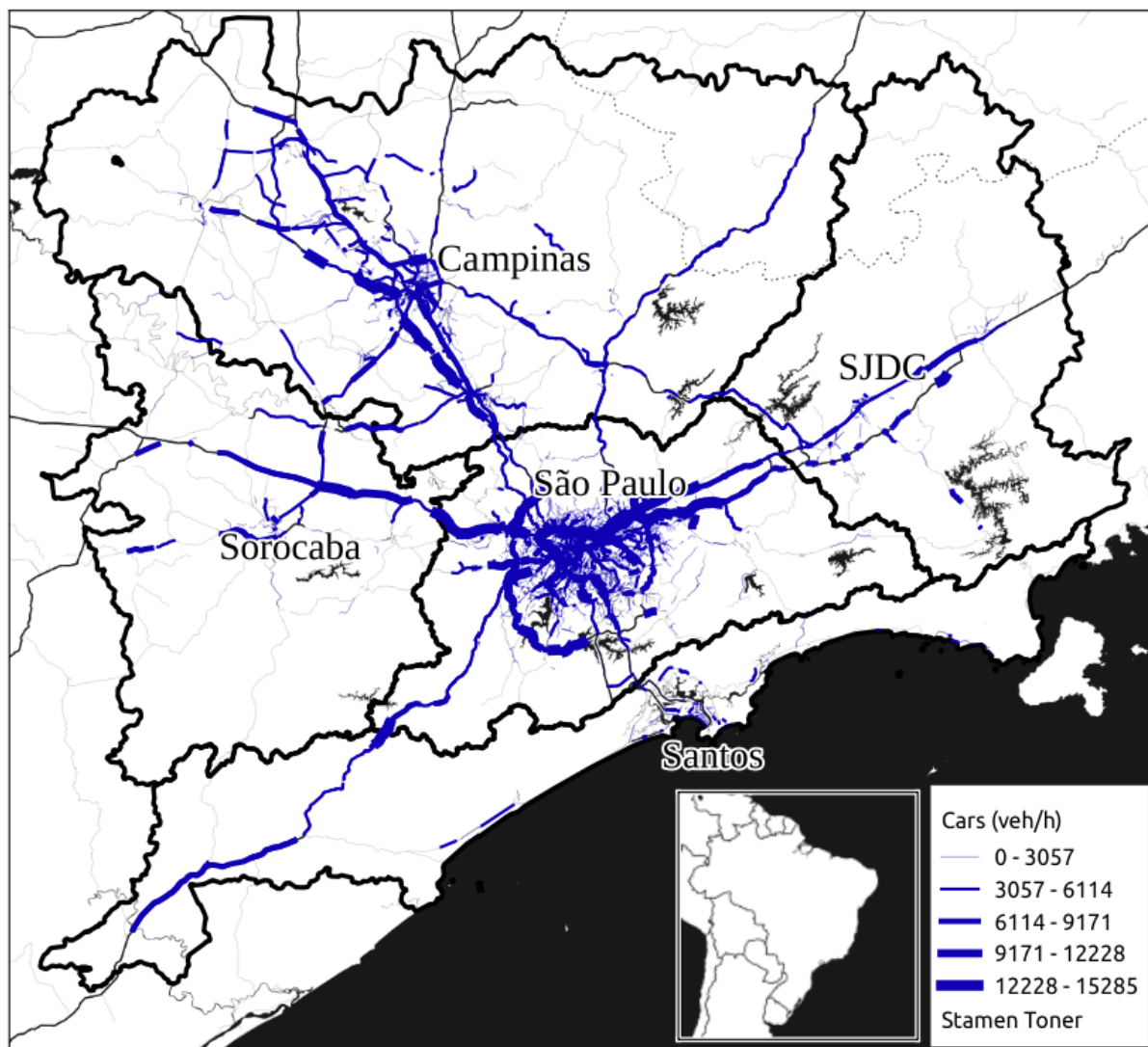


Figure 33 – Traffic flows derived from GPS recordings for Cars ( $veh \cdot h^{-1}$ ) on Monday 2014-10-06 08:00 - 09:00 LT.

- B Traffic from GPS records except MASP. MASP with traffic simulation from CET and SPtrans of LDV, HGV and UB.
- C Traffic from GPS records except MASP. MASP with traffic interpolation for LDV and MC, HGV from traffic simulation of CET and UB from SPtrans.

The scenario A considers not only GPS records but also traffic simulation of UB from SPtrans. It was included UB because GPS data include only Cars and Trucks. Also, the spatial distribution of UB and Trucks is different as shown Fig. (34) and Fig. (4). The average number of vehicles per streets is quite different, with more UB. The traffic will be later calibrated but before that, the proportions of UB and Trucks need to be adjusted. According to the traffic simulations from CET and SPtrans, the volume of trucks is  $2809297 \text{ veh} \cdot h^{-1}$  and UB is  $763097 \text{ veh} \cdot h^{-1}$  meaning that the proportion of UB to Trucks is 0.2716328. To obtain this proportion it would be possible to use other

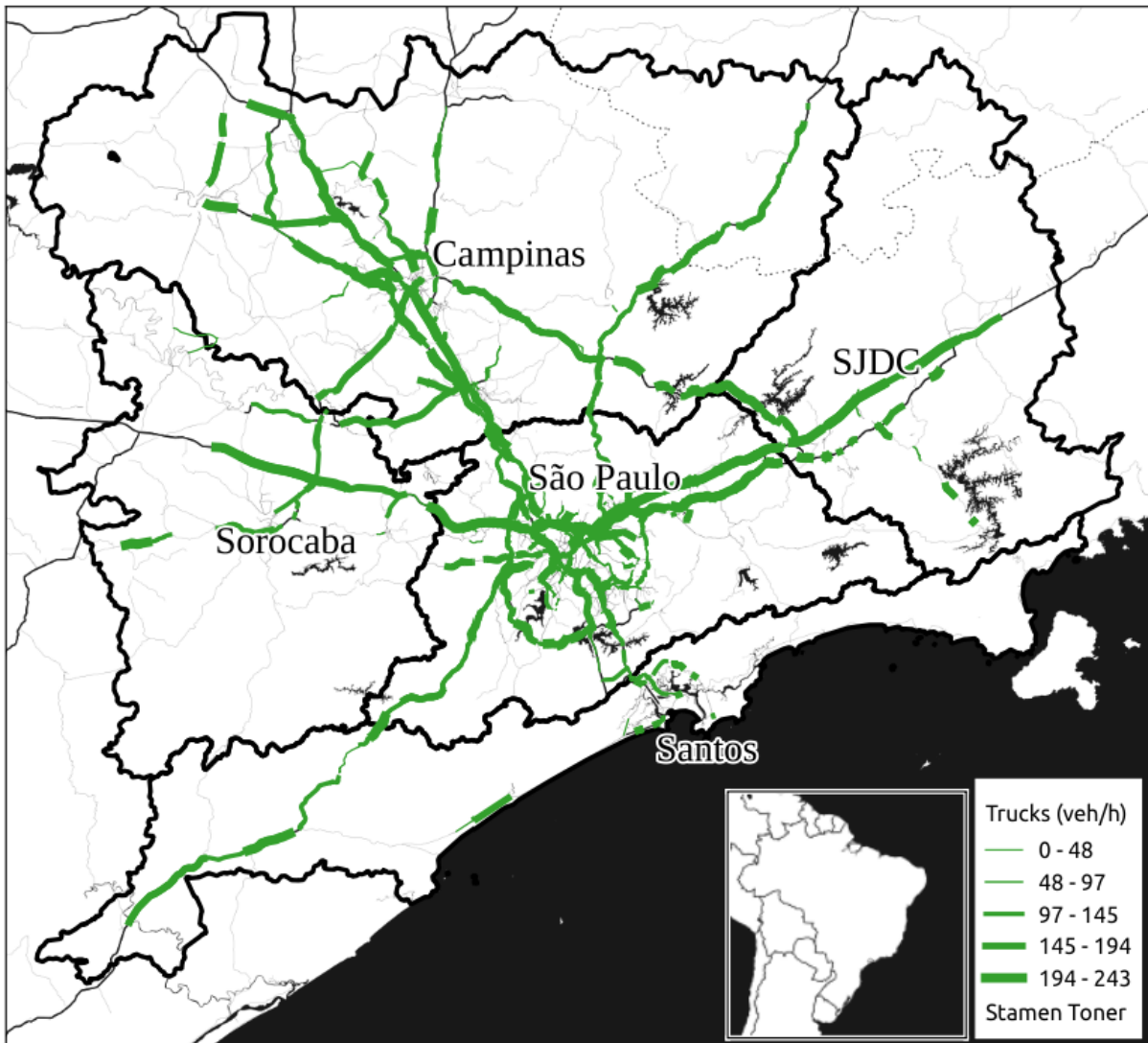


Figure 34 – Traffic flows derived from GPS recordings for Trucks ( $veh \cdot h^{-1}$ ) on Monday 2014-10-06 08:00-09:00 LT.

source of data but it was used traffic simulations because both have from the same origin: ODS. This proportion is kept by multiplying the number of UB with a factor. The volume of Trucks from GPS between 08:00 and 09:00 in MASP  $474753.3 \text{ veh} \cdot h^{-1}$ . Therefore, the multiplication factor of UB is  $0.2716328 \cdot \frac{474753}{763097.3} = 0.1689935$

The scenario B uses GPS records for all regions with exception of MASP where traffic simulations are used instead. The scenario C uses LDV and MC from interpolations. Scenarios B and C does not need to multiply traffic to keep proportions.

Table 15 – Calibration factors for traffic by type of fuel in each region and scenario

Region	E25		E100		B5	
	Factor	%	Factor	%	Factor	%
MASP (GPS)	0.495	101.22	0.393	100.05	0.0124	84.14
Campinas (GPS)	0.2695	101.01	0.262	100.61	4.63	84.39
SJDC (GPS)	0.362	100.27	0.238	100.16	3.7	85.40
Santos (GPS)	0.265	100.09	0.164	100.73	5.15	85.61
Sorocaba (GPS)	0.181	100.50	0.199	100.55	3.5	85.56
MASP (Simulation)	1.135	100.41	0.905	100.47	0.44	84.47
MASP (Interpolation)	0.795	100.09	0.595	100.43	0.55	85.03

## 5.2 Emissions inventories

### 5.2.1 Traffic flow calibration with fuel consumption

Traffic from all sources was calibrated in order to match the fuel sales in each region. A calibration factor must be multiplied to the traffic flow in order to match 100% of sold E25 and E100, and 85% of B5. Table (15) shows the calibration factors for traffic by type of fuel and region. Traffic data was overestimated for E25 and E100 in all regions with the exception of MASP (Simulation) consuming E25. The region that required a smaller factor for adjusting the traffic consuming E25 was Sorocaba, which means that the traffic was slightly over predicted in this region. In the case of E100, all regions presented over prediction, the highest factor was obtained with MASP (Simulation) and the smaller for Santos. Lastly, the B5 presented values above 1 for all regions with GPS traffic data, except MASP. In this case, MASP (GPS) contains traffic simulation for buses, and MASP (Simulation) and MASP (Interpolation) have traffic simulation for trucks and buses. This means that traffic was over predicted in this area, however, in the case MASP (GPS) it was excessively over predicted. This is due to the fact the speed was used as proxy for estimating traffic data from GPS recordings, and this area has more *motorways* and *trunks*. Another reason is that more vehicles are recorded in MASP contrary to the other regions, and therefore, this region has more streets with speeds. MASP (simulation) produced the values closer to one in all fuels, which means that this is the most representative data before adjustment by traffic sales.

### 5.2.2 Emissions in each region by type of vehicle

Table (16) shows the emissions inventories for each region;  $CO_2$  is expressed as  $10^6 t \cdot y^{-1}$  and Vehicles times kilometers (VKM) as  $10^{12} km \cdot y^{-1}$ . The total NMHC includes also the regional total for paintings, as explained in section 3.4.2. The region with the greatest amount of emissions is MASP, however, the quantities varies depending on the input of the activity data. In the case of  $PM$ ,  $CO$ ,  $NO_x$ ,  $NMHC$ ,  $CH_4$  and  $CO_2$

the highest emissions are obtained with MASP GPS, however, this regions produces less activity VKM in comparison with the other MASP regions. The reason is that the average speed in MASP GPS is lower than the other regions. This can be seen in Fig. (35) where the average speeds in MASP for the 3 data sources are presented for the 168 hours of the week. The three speeds curves shows a similar behavior with lower speeds at rush hours and higher speeds when the traffic is not so intense. As Fig. (21) shows, emissions increases when the average speeds lowers. Therefore, this explains MASP GPS higher emissions for some pollutants despite having less activity (VKM) and highlights the importance of including kinematic parameters in vehicular emissions inventories. Another region of interest is Campinas, because it also has lower VKM than MASP Simulation and MASP Interpolation, but PM emissions are higher, which is due to the relatively high VKM values for HGV and also slower speeds, which is a characteristic of the set GPS data presented in this study. Regarding the Categories of vehicles, there is a consistent estimation by type of vehicles for all the regions and scenarios. In other words, vehicles consuming diesel emits more  $NO_x$ ,  $SO_2$  and PM, and LDV emits more CO and NMHC.

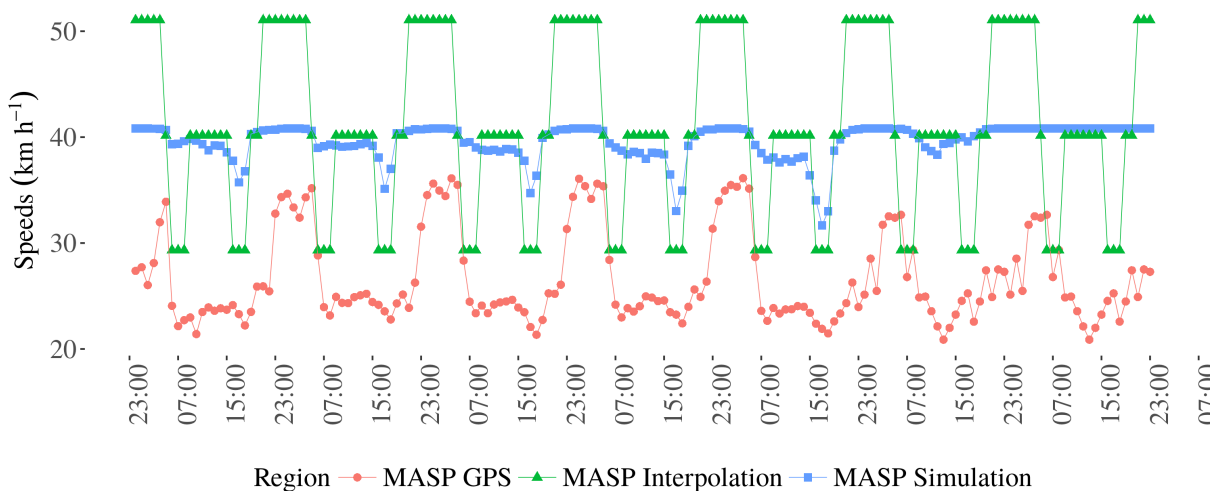


Figure 35 – Hourly speeds in MASP with data from GPS, simulations and interpolations ( $km \cdot h^{-1}$ ).

### 5.2.3 Emissions by age of use of vehicle

The vehicular emissions inventories presented in this study shares a common characteristic of the age of use of vehicle. This allows to identify a *signature* of the pollutant in the region, which is defined as the total amount of emissions by age of use and type of vehicle shown in 36. These figures comes from the GPS traffic data for the 5 metropolitan regions. The emissions with traffic data from traffic simulation and interpolations produce a similar signature with smaller emissions, therefore, they are not



Table 16 – Emissions by each regions ( $t \cdot y^{-1}$ ) 2014

Region	Cat	PM	CO	NO <sub>x</sub>	NMHC	NH <sub>3</sub>	SO <sub>2</sub>	CH <sub>4</sub>	CO <sub>2</sub>	VKM
Campinas GPS	PC	20.80	72743.70	7214.00	25056.00	741.80	174.50	1741.90	4.06	61.67
	LCV	52.50	7033.50	1752.50	4480.60	103.90	32.00	233.00	0.83	10.41
	HGV	2424.40	6660.30	42079.70	1830.30	24.40	942.20	23.10	3.98	9.78
	MC	0.00	11636.10	780.00	1151.50	8.70	10.30	224.30	0.28	10.47
	<b>Total</b>	<b>2497.70</b>	<b>98073.60</b>	<b>51826.20</b>	<b>35820.76</b>	<b>878.80</b>	<b>1159.00</b>	<b>2222.30</b>	<b>9.14</b>	<b>92.34</b>
Santos GPS	PC	5.40	17640.30	1868.00	6159.00	173.50	36.90	417.10	0.90	2.40
	LCV	13.70	1846.40	442.60	1134.60	25.20	7.30	56.00	0.20	0.43
	HGV	775.90	2052.50	13191.80	546.20	7.80	296.90	7.40	1.25	0.61
	MC	0.00	3275.00	222.30	297.80	2.20	2.70	57.50	0.07	0.46
	<b>Total</b>	<b>795.00</b>	<b>24814.20</b>	<b>15724.70</b>	<b>10556.61</b>	<b>208.70</b>	<b>343.80</b>	<b>538.00</b>	<b>2.43</b>	<b>3.89</b>
SJDC GPS	PC	5.80	19148.50	1988.90	6630.60	187.30	40.60	449.00	0.99	2.73
	LCV	14.60	2025.70	474.40	1218.40	27.10	8.00	60.20	0.22	0.48
	HGV	513.30	1434.30	8965.40	393.50	5.20	202.00	4.90	0.85	0.39
	MC	0.00	3440.30	232.40	319.40	2.40	2.90	61.40	0.08	0.52
	<b>Total</b>	<b>533.70</b>	<b>26048.80</b>	<b>11661.10</b>	<b>11864.26</b>	<b>222.00</b>	<b>253.50</b>	<b>575.50</b>	<b>2.13</b>	<b>4.12</b>
Sorocaba GPS	PC	5.00	17708.90	1856.20	6061.70	186.00	43.80	433.60	1.00	1.97
	LCV	12.70	1834.00	444.00	1068.10	25.70	7.80	57.90	0.20	0.33
	HGV	556.80	1409.60	9254.60	364.10	5.60	208.40	5.30	0.88	0.36
	MC	0.00	3146.30	215.70	275.00	2.10	2.60	54.60	0.07	0.32
	<b>Total</b>	<b>574.50</b>	<b>24098.80</b>	<b>11770.50</b>	<b>10568.43</b>	<b>219.40</b>	<b>262.60</b>	<b>551.40</b>	<b>2.16</b>	<b>2.98</b>
MASP GPS	PC	58.40	216870.40	19458.40	70080.00	1974.60	477.80	4689.70	11.40	493.06
	LCV	147.50	24418.70	4851.90	12983.20	281.60	92.60	628.50	2.43	101.06
	HGV	2555.70	7865.30	47567.40	2349.50	25.80	1071.60	24.40	4.49	70.24
	Bus	259.70	3364.00	15624.40	615.30	4.40	349.20	51.90	2.63	5.56
	MC	0.00	30804.20	1960.50	3457.40	24.00	28.60	624.20	0.77	105.36
<b>Total</b>	<b>3021.30</b>	<b>283322.60</b>	<b>89462.60</b>	<b>117928.93</b>	<b>2310.40</b>	<b>2019.80</b>	<b>6018.70</b>	<b>21.74</b>	<b>775.27</b>	
MASP Simu	PC	53.70	127637.80	16307.70	30435.90	3422.40	278.20	3069.50	11.01	1932.57
	LCV	136.8	14021.70	3810.90	5692.70	339.30	178.50	433.9	2.63	356.35
	HGV	1815.3	6785.80	42215.50	1752.10	30.30	1232.80	17.70	5.18	158.21
	Bus	109.6	1708.20	7574.80	268.30	2.10	260.90	20.50	1.43	13.14
	MC	0.00	28622.60	1863.30	3382.80	25.20	11.90	651.40	0.76	336.43
<b>Total</b>	<b>2111.54</b>	<b>178776.1</b>	<b>71772.20</b>	<b>69975.60</b>	<b>3819.30</b>	<b>1701.14</b>	<b>4220.00</b>	<b>21.01</b>	<b>2796.70</b>	
MASP Inter	PC	62.10	137867.80	18109.90	33002.60	3756.20	480.00	3479.70	11.57	2006.11
	LCV	158.10	14590.90	4152.80	6242.90	372.30	227.00	492.20	2.79	372.76
	HGV	1815.30	6785.80	42215.50	1752.10	30.30	1232.80	17.70	5.18	115.52
	Bus	109.60	1708.20	7574.80	268.30	2.10	260.90	20.50	1.43	13.14
	MC	0.00	6064.50	403.20	660.00	5.30	91.50	136.30	0.15	63.31
<b>Total</b>	<b>2145.10</b>	<b>167017.20</b>	<b>72456.20</b>	<b>70369.43</b>	<b>4166.20</b>	<b>2206.40</b>	<b>4146.40</b>	<b>21.12</b>	<b>2570.85</b>	

shown here. To understand these figures it is necessary to consider that the emission factors are higher for older vehicles as shown on Fig. (20), and also the deterioration of vehicles and fleet distribution as shown in Fig. (19). Emissions of  $CH_4$ ,  $CO$  and  $NMHC$  show three peaks, at vehicles newer than 10, at 20 and older than 30 years of use. The reason of the first peak is the rapid increment of the fleet between 2000 and 2011 including Motorcycles (MC). The second peak (vehicles between 19 and 23 years of use) is due to relative equilibrium between fleet and deteriorated emission factors. According to the Manual of Air Pollution Control Program by Motor Vehicles (MACEDO et al., 2011), vehicles with catalytic system entered into market with the Brazilian emissions standard Proconve L2, in 1992. Vehicles of 1992 have 23 years of use in 2014, therefore, the second peak is due to vehicles with deteriorated catalytic system. The third peak is due to relatively small fleet with higher emission factors. Emissions of  $CO_2$ ,  $N_2O$  and  $SO_2$  share a pattern following the increment of the fleet.

Emissions of  $NO_X$  and  $PM$  have a pattern slightly similar with a central peak, however, there are more  $NO_X$  emissions in newer vehicles due to the participation of Buses. Evaporative emissions have a clear pattern of higher emissions in older vehicles, for this graph diurnal, hot-soak, running losses and also, leaking evaporative fuel emissions produced when buying fuel were aggregated. Regarding the participation of each type of vehicle, PC is the main responsible on emissions of 78.04% of  $CH_4$ , 75.40%  $CO$ , 48.81% of  $CO_2$ , 82.92% of Evaporative, 63.12% of  $N_2O$  and 64.74% of  $NMHC$ . In the case of HGV, they were responsible of 67.09 % of  $NO_X$ , 91.96%  $PM$  and 67.37 % of  $SO_2$ .

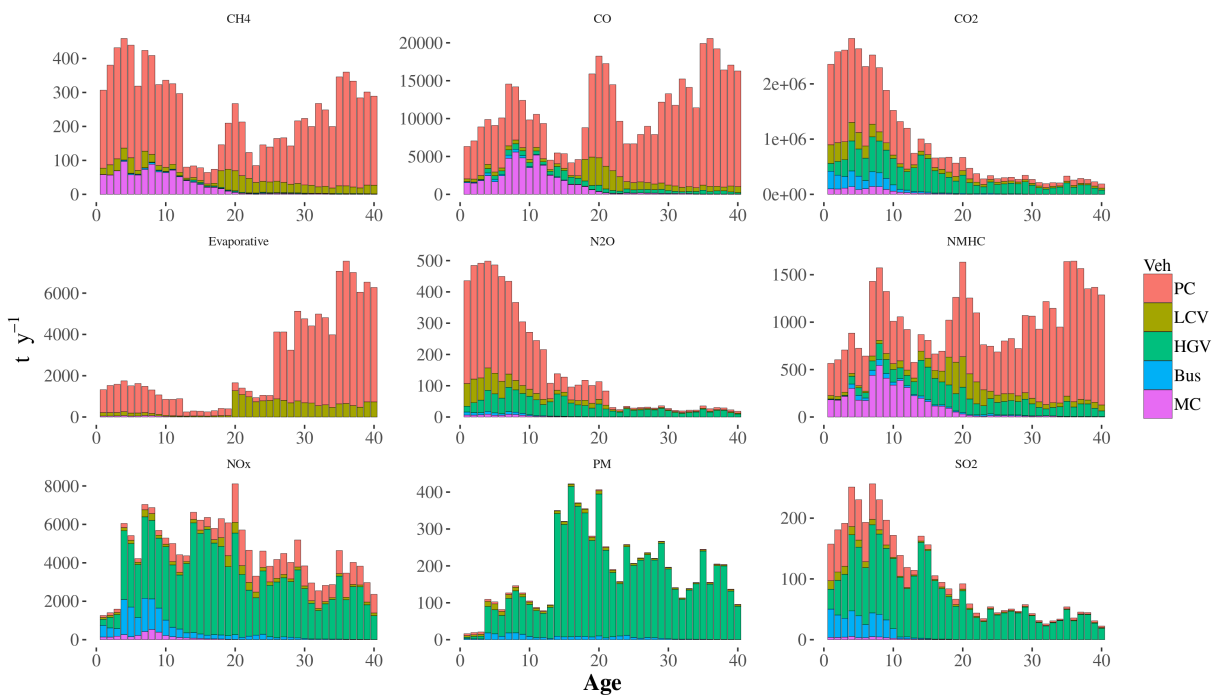


Figure 36 – Signature (vehicular emissions by age of use and type of vehicle) in all metropolitan regions with GPS traffic data for different pollutants ( $t \cdot y^{-1}$ ).

#### 5.2.4 Emission maps

The emission maps of  $CO$ ,  $NO_X$ ,  $NMHC$  considering exhaust, evaporative, evaporation at fuel stations and paintings, and  $PM$  for traffic data from GPS, Simulations and Interpolations are shown in Fig. (37). All emissions outside the metropolitan region of São Paulo comes from GPS data, therefore, the spatial difference is found only at MASP.  $CO$  have a spatial distribution very similar for the data with different sources. However, high MASP emissions are broader with data from Interpolation, followed by GPS. In the case of  $NO_X$ , data from Interpolation and Simulation show a similar pattern, clearly identifying the motorway ring at south of MASP. This is not the case for  $NO_X$  emissions from GPS data because trucks uses speeds from *trunks* and *motorways* as proxy for traffic flow, the same streets used for Cars traffic generation. Therefore, Cars

and Trucks share, in some cases, the same streets. The emissions of PM follows the pattern of  $NO_x$ . There is a peak emission in a motorway at the southwest part of the region, at longitude 47.5 W and latitude 24.4 S which is due to a high activity of trucks in this area (Rodovia Regis Bittencourt). Other motorways (Rodovia dos Bandeirantes and Rodovia Anhanguera) at the northwest part also show higher emissions of PM. NMHC shows a different pattern than the other pollutants with more streets with low emissions in red, due to painting emissions distributed in non *motorway* streets. The higher density of NMHC emissions is similar to CO, with hot-spots in the center of the cities.

Fig. (38) shows the LAPAt emissions generated according to ANDRADE et al. (2015), for NO for a 2 km x 2 km grid. VEIN emission maps show details that could not be done using the top-down approach.

### 5.2.5 Emissions by hour of the week

The hourly emissions for the 168 hours of the period study are shown on Fig. (39). The pollutants shown are  $CO$ ,  $NO_x$ ,  $NMHC$  and  $PM$  estimated with traffic data from GPS recordings, Interpolation and traffic Simulations for MASP. Emissions estimated with GPS have a similar pattern for different pollutants, and in the case of data from Interpolations and Simulations, both set of hourly estimations show a similar pattern because the traffic from one morning rush traffic flow is extrapolated using temporal factors of the same normalized traffic counts from the nearest toll stations from São Paulo. Therefore, emissions dominated by light vehicles are very similar with traffic data from Interpolations and Simulations, with some small differences in the totals. Emissions from GPS data presents profiles relatively constants for  $CO$  and  $NMHC$ , with lower emissions at night and early morning, and lower emissions on weekends. In the case of  $NO_x$  the profile is more marked and with PM not so marked, this is due to the profile of Buses which comes from traffic simulation of SPTrans. On contrary, emissions from Interpolation and Simulations have marked profiles in all emissions.

## 5.3 Atmospheric simulation

Modeling air quality is a complex process that depends on emissions and the meteorological simulation. This subsection presents first the comparison between observed and simulated meteorology parameters and then the comparison for air pollutant concentrations for the inner grid only. Observation data are from the Air Quality and Meteorological Network (Qualar, <http://qualar.cetesb.sp.gov.br/>) from the Environmental Agency of São Paulo, CETESB. This system provides information of air pollutant concentration and measurements of meteorological parameters for 62 superficial stations. It was chosen the stations inside the metropolitan areas of this study,

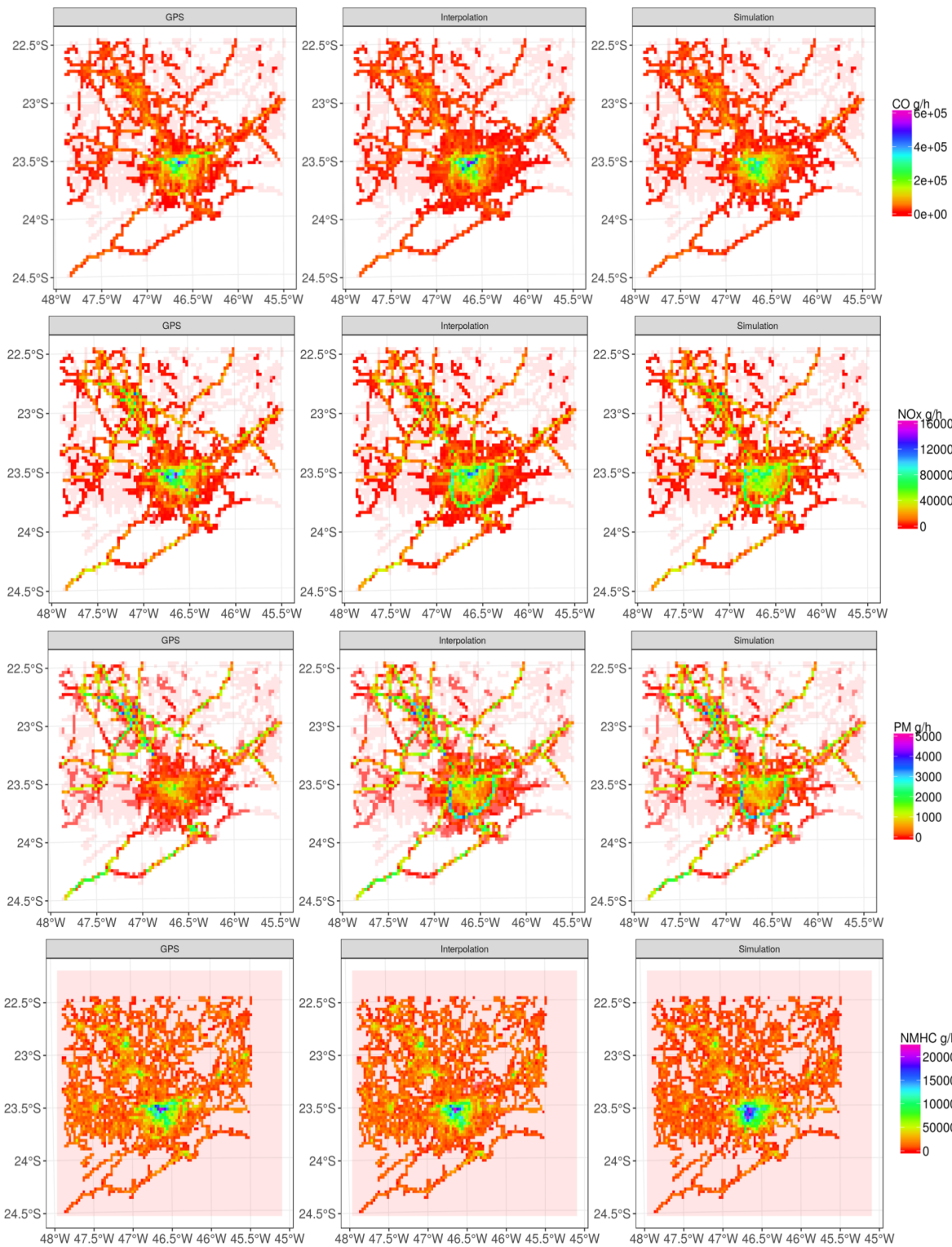


Figure 37 – Emission maps of  $CO$ ,  $NO_x$ ,  $PM$  and  $NMHC$  due traffic activity from GPS, Interpolation and Simulations ( $g \cdot h^{-1}$ ) for 08:00-09:00LT Monday 2014.

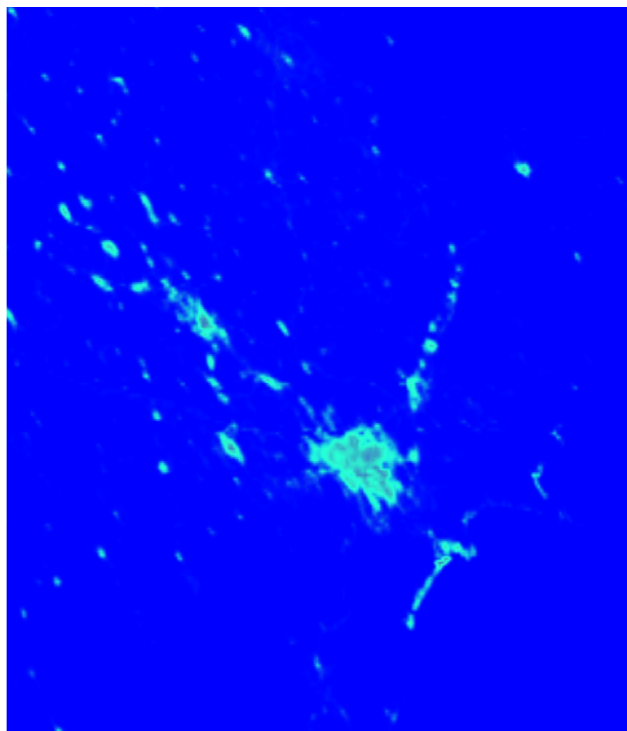


Figure 38 – Emission maps of NO using LAPAt top-down approach ( $g \cdot h^{-1}$ ) for 08:00-09:00LT.

shown in Fig. (40). It was used the information of 41 stations located in 5 metropolitan areas.

The emissions in Campinas, SJDC, Santos and Sorocaba were the same for all the emissions scenario, using only activity data from GPS recordings. As a consequence, the air pollutant concentrations and meteorological parameters did not present too much variation. Therefore, for these 4 areas, the analyses presented were those from the GPS scenario, whereas, for MASP, it was presented the 3 scenarios: GPS, traffic interpolation and traffic simulation. The analyses consisted in hourly mean for each metropolitan region comparing observed and simulated variables.

### 5.3.1 Meteorological simulation

WRF underestimates maximum temperature and overestimates minimum temperature, Fig. (41). However, the diurnal pattern (with minimum temperatures around 06LT and maximum temperatures before 15LT) and the increase of temperatures over the week are well simulated. Temperature simulated at Santos presents higher variance, with interquartile ranges reaching almost 10 degrees. This was caused because the stations at Santos varies in land use and altitude: the city of Cubatão is an industrial district, farther from the coast and at higher altitudes than the station at the city of Santos, which is very near the beach.



Figure 39 – Hourly emissions of  $CO$ ,  $NO_x$ ,  $NMHC$  and  $PM$  due traffic activity from GPS, Interpolation and Simulations in MASP ( $g \cdot h^{-1}$ ).

WRF overestimates wind speeds at all regions, as shown in Fig. (42). However, the simulation follows the tendency of the observations more clearly in some regions, such as Santos and Campinas. On the other hand, it is also in Santos where the largest standard deviations are found resulting in different magnitude of winds. This means that simulated wind is simulated at higher speeds than observed.

### 5.3.2 Simulation of $CO$ , $NO$ , $NO_2$ , $SO_2$ and $O_3$

Fig. (43) shows the comparison for  $CO$  concentrations. There are no  $CO$  data in Santos and SJDC, therefore it was presented only the simulated concentrations in those regions. Simulated  $CO$  for MASP GPS stands out for its much higher values than any other simulation and region. For all the other simulations, inside each metropolitan area, simulated  $CO$  are in general lower than observed. This could be due to higher simulated wind speed. However, they all simulate the morning and evening peaks associated to rush-hours higher emissions. Simulated  $CO$  for MASP Interpolation and MASP Simulation show good agreement with observations, the morning peak for week days are well represented, with simulated values very near from observed ones. The

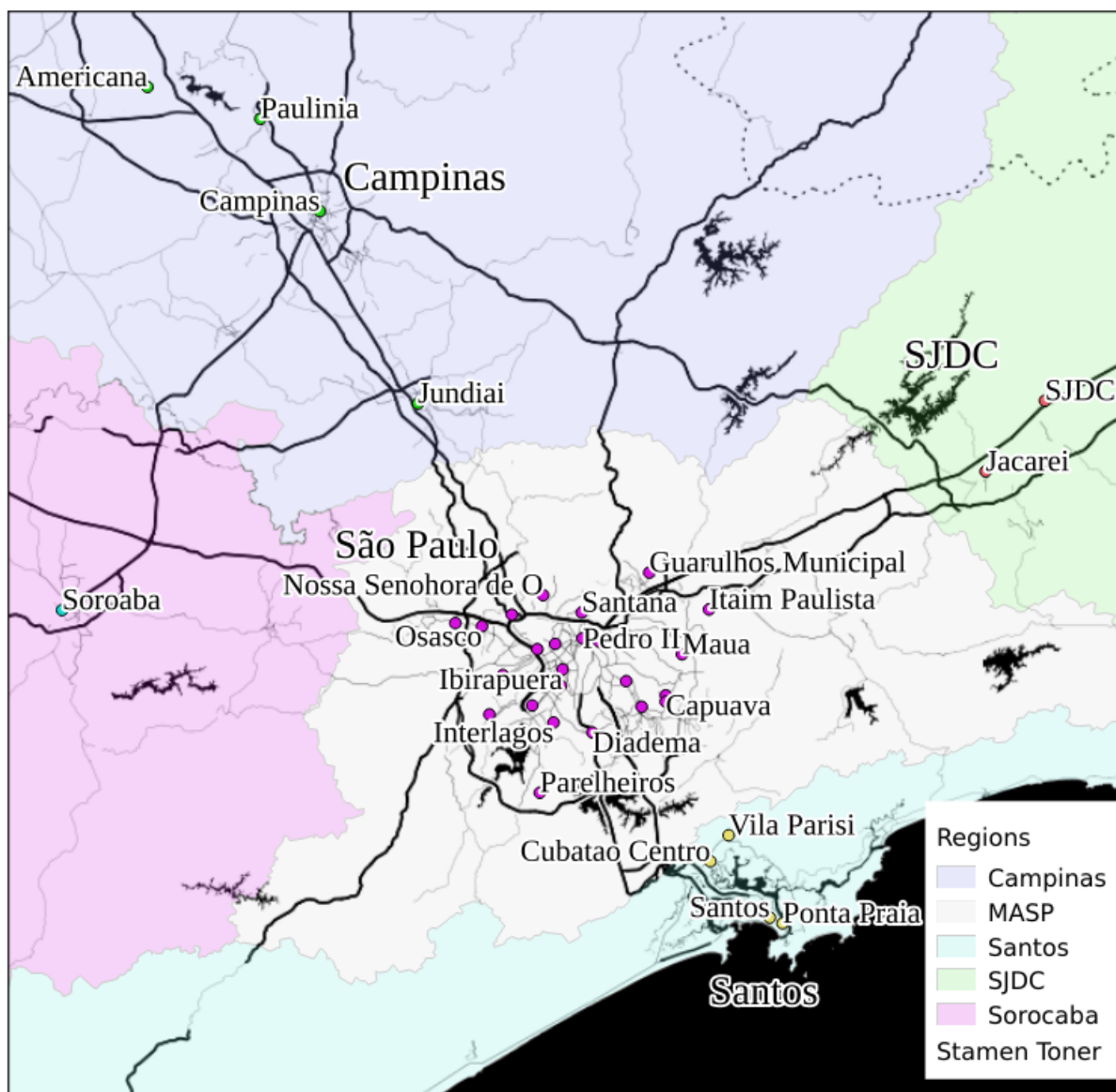


Figure 40 – Meteorological and Air Quality Stations from CETESB used in this study.

evening peak is also well represented, however, the higher observed concentrations at night, could not be simulated, specially from Friday to Saturday.

*NO* observations and simulations are shown on Fig. (44). Campinas observations are well accompanied by simulations in most hours, except in Friday and Saturday peaks. Result with MASP Interpolation and MASP Simulation are very similar, observations are higher than simulations but there is an agreement in general tendency. In the case of MASP GPS, it seems that there is a better agreement with observation. The morning peak is usually somewhat simulated, but the higher observed concentrations during the evening and early hours of each day could not. In the case of Santos, simulations barely accompanied observations that were higher in all hours. Santos is a region that has the biggest port in South America, therefore, since these maritime and stationary sources



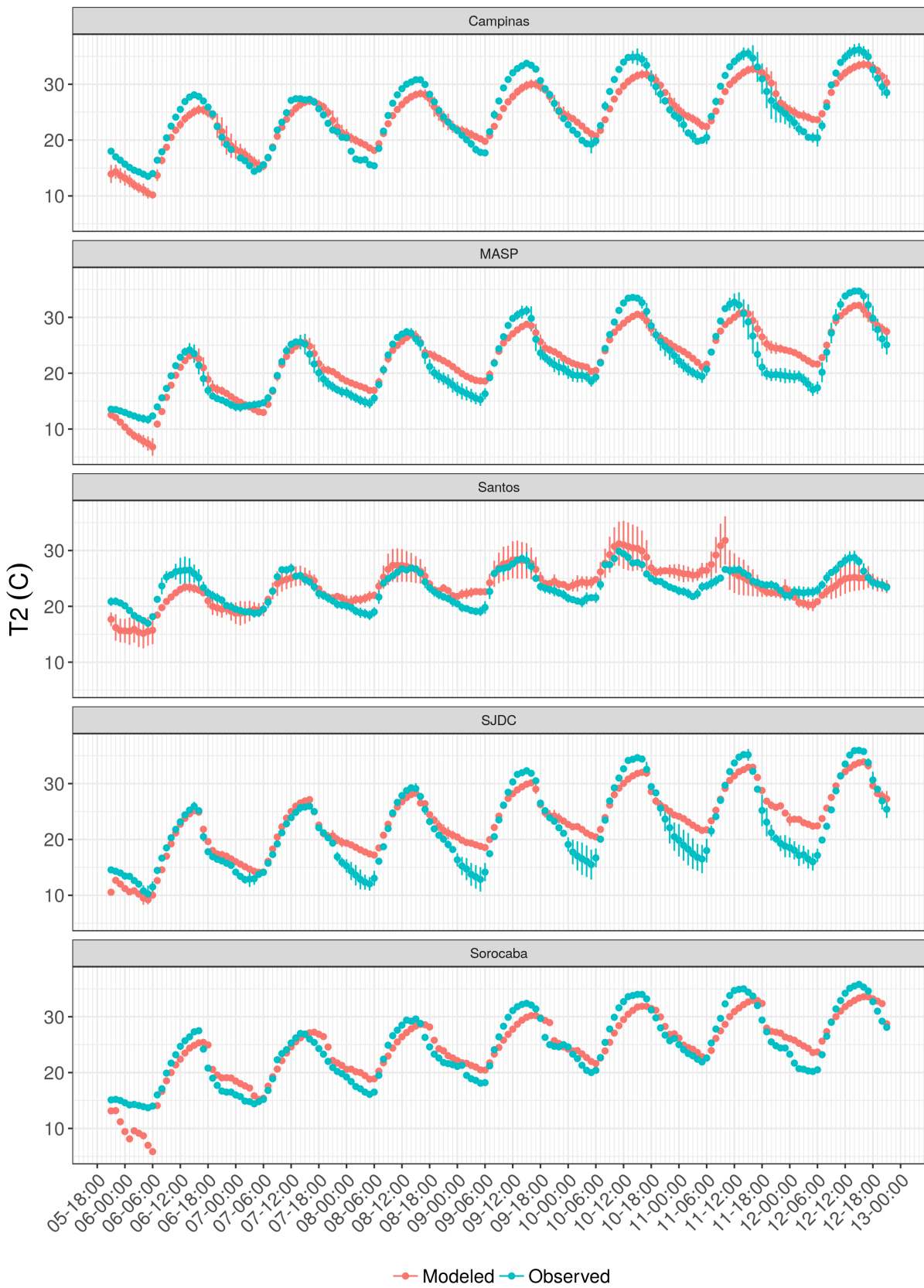


Figure 41 – Hourly mean of observed and simulated temperature (Celsius) by Metropolitan Area.



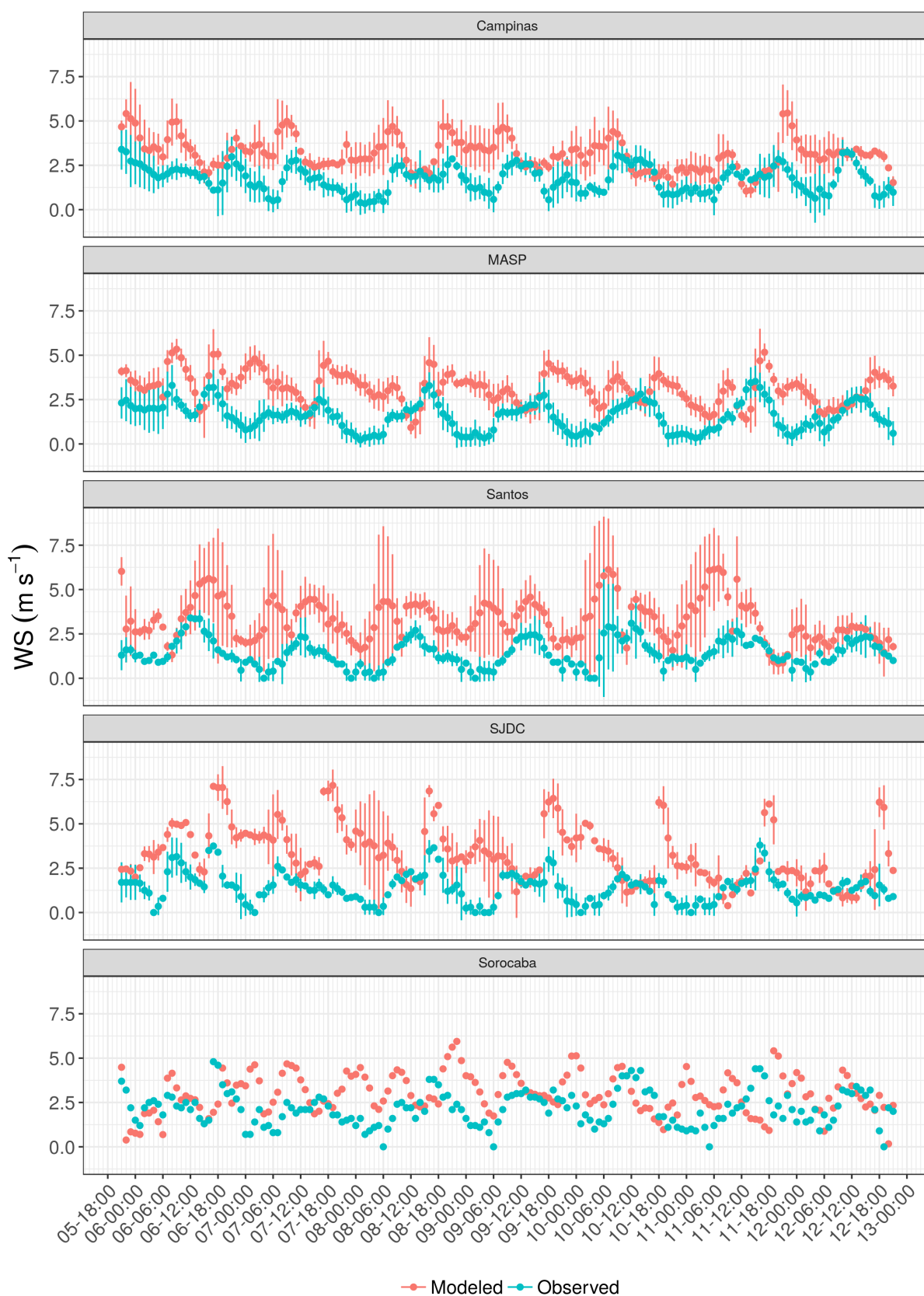


Figure 42 – Hourly mean and standard deviation of observed and simulated wind speed ( $m \cdot s^{-1}$ ) by Metropolitan Area.

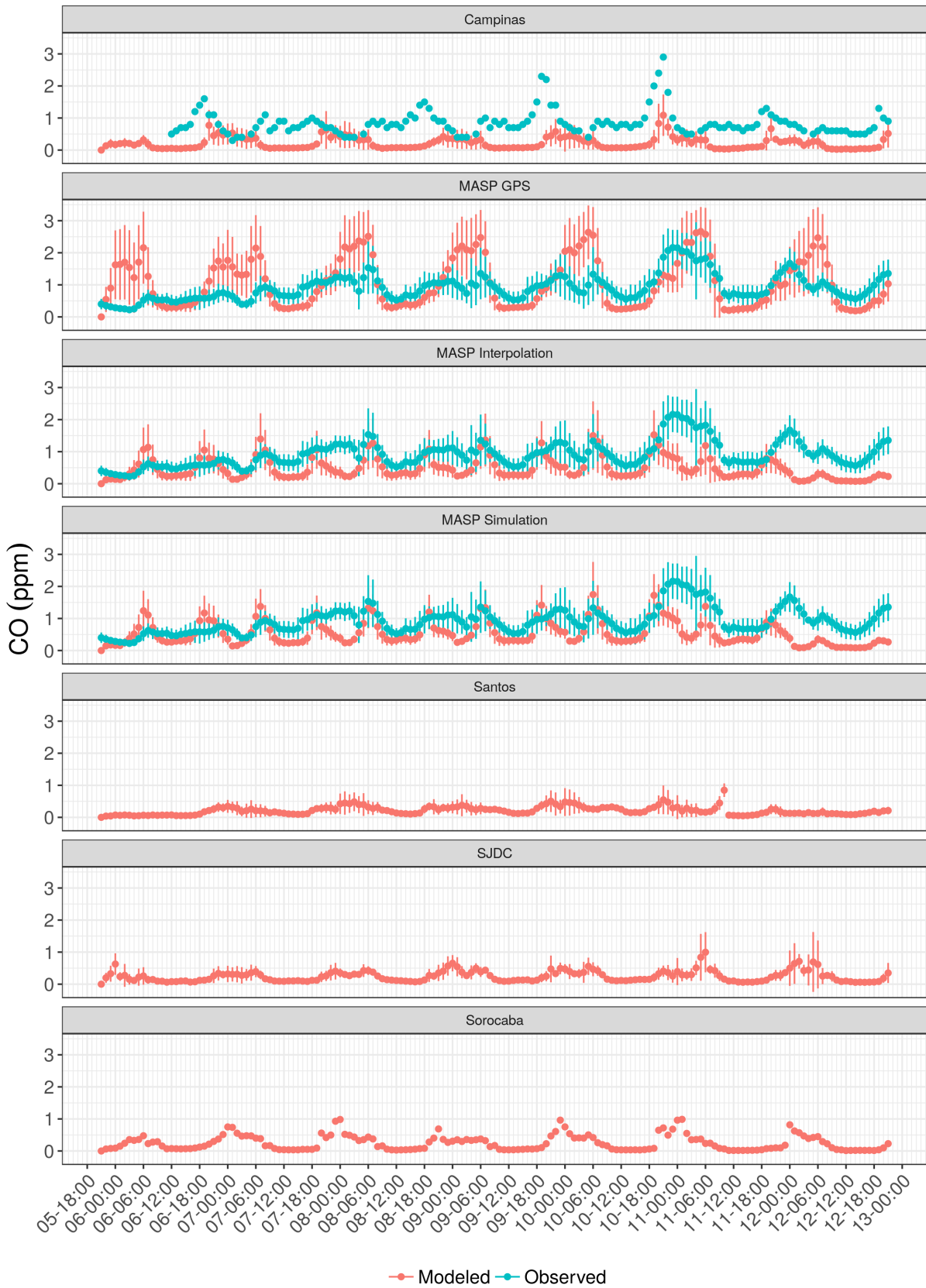


Figure 43 – Hourly mean and standard deviation of  $CO$  (ppm) by region and scenario.

at the port were not included in the emissions, it is reasonable that the simulated values were so low. SJDC and Sorocaba show better agreement for low concentrations, peak observations are not well accompanied by simulations. Since NO is also a primary pollutant, its simulated concentration was also impacted by the higher wind speeds. Sorocaba simulated wind speed were more similar to observed values which could explain the better agreement of observed and simulated NO.

$NO_2$  observations and simulations are shown on Fig. (45). In all regions observations are higher than observations. Simulated NO Fig. (44) was also much lower than observed, which would mean less  $NO_2$  photochemical production, and thus, lower simulated  $NO_2$ . However, there also seems that in Sorocaba, the photochemical production of  $NO_2$  is consistent with diurnal values. The accumulation of this pollutant during the evening, however, could not be reproduced by the model.

Hourly mean and standard deviation of  $SO_2$  observations and simulations are shown on Fig. (46). Campinas and Sorocaba do not present observations and both regions present a similar profile, with higher concentrations in Campinas. Regarding the MASP and SJDC regions, simulations are in the same order of magnitude with observations, and the standard deviation present small values. Santos is the only region in which observations were much above atmospheric simulations. This could mean that emissions produced by diesel combustion are underestimated, however, as already mentioned, a detailed characterization of the sources is required for this region. There is a peak of  $SO_2$  at early morning of Sunday which is not accompanied by any simulations, with exception of MASP GPS. This could be due to trucks activity at that time, and normalized profiles to expand hourly morning rush hour did not capture this variation. MASP Simulation and Interpolation  $SO_2$  simulated values have almost the same peaks for morning and evening rush hours. However, as with the other primary pollutants, there seems to be an issue with the Planetary Boundary Layer parameterization, since at evening higher observed values could not be reproduced. Simulated CO and NO seemed to be more diluted by higher winds; thus, if the same hypothesis were applied here, it would mean that  $SO_2$  emissions are overestimated for MASP.

Hourly mean and standard deviation of  $O_3$  concentrations and simulations are shown in Fig. (47). Concentrations in MASP regions seems to have a reasonably good agreement with observations. However, the diurnal peak of simulated  $O_3$  occurs a few hours later than observations. The early morning observed peak (due to recirculation) could not be reproduced by the model. It is interesting to notice that the diurnal peak of simulated  $O_3$  occurs a few hours before the observed one. In SJDC, this pattern varies throughout the week.  $O_3$  is underestimated at all the regions, although in Santos the values are more similar. In Campinas simulated  $O_3$  production was almost absent.

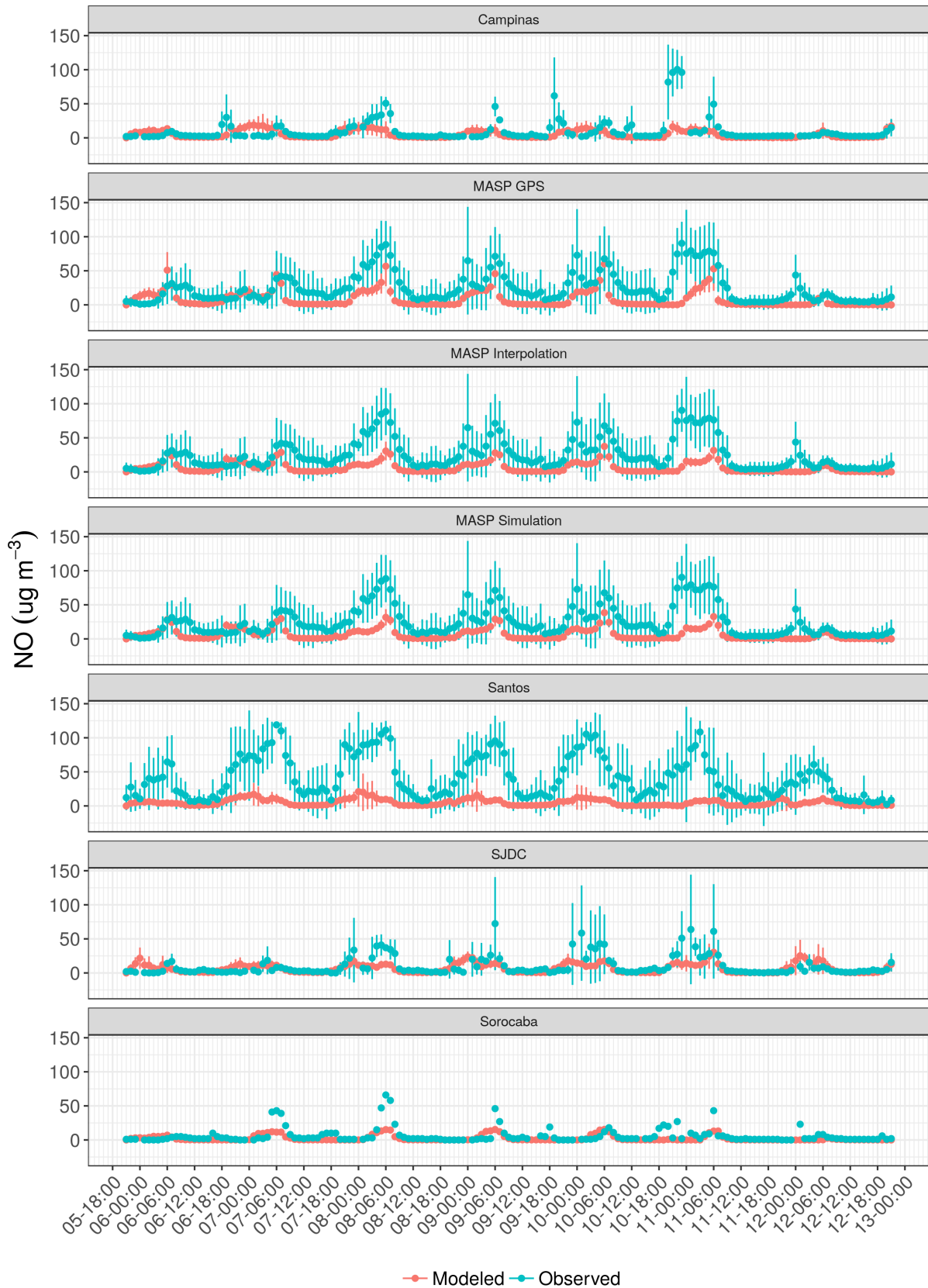
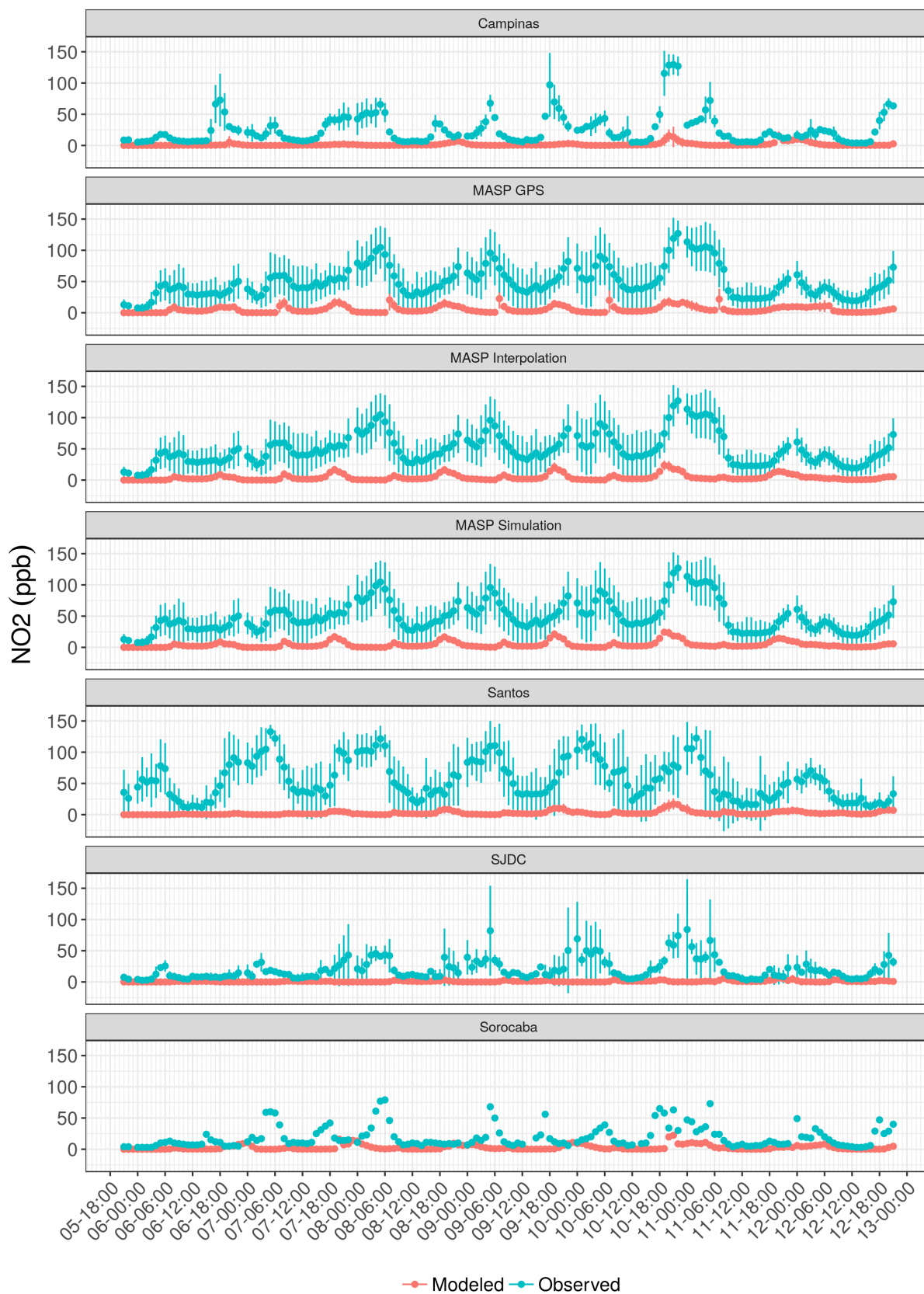


Figure 44 – Hourly mean and standard deviation of  $NO$  ( $\mu g \cdot m^{-3}$ ) by region.

Figure 45 – Hourly mean and standard deviation of  $NO_2$  ( $ug \cdot m^{-3}$ ) by region.

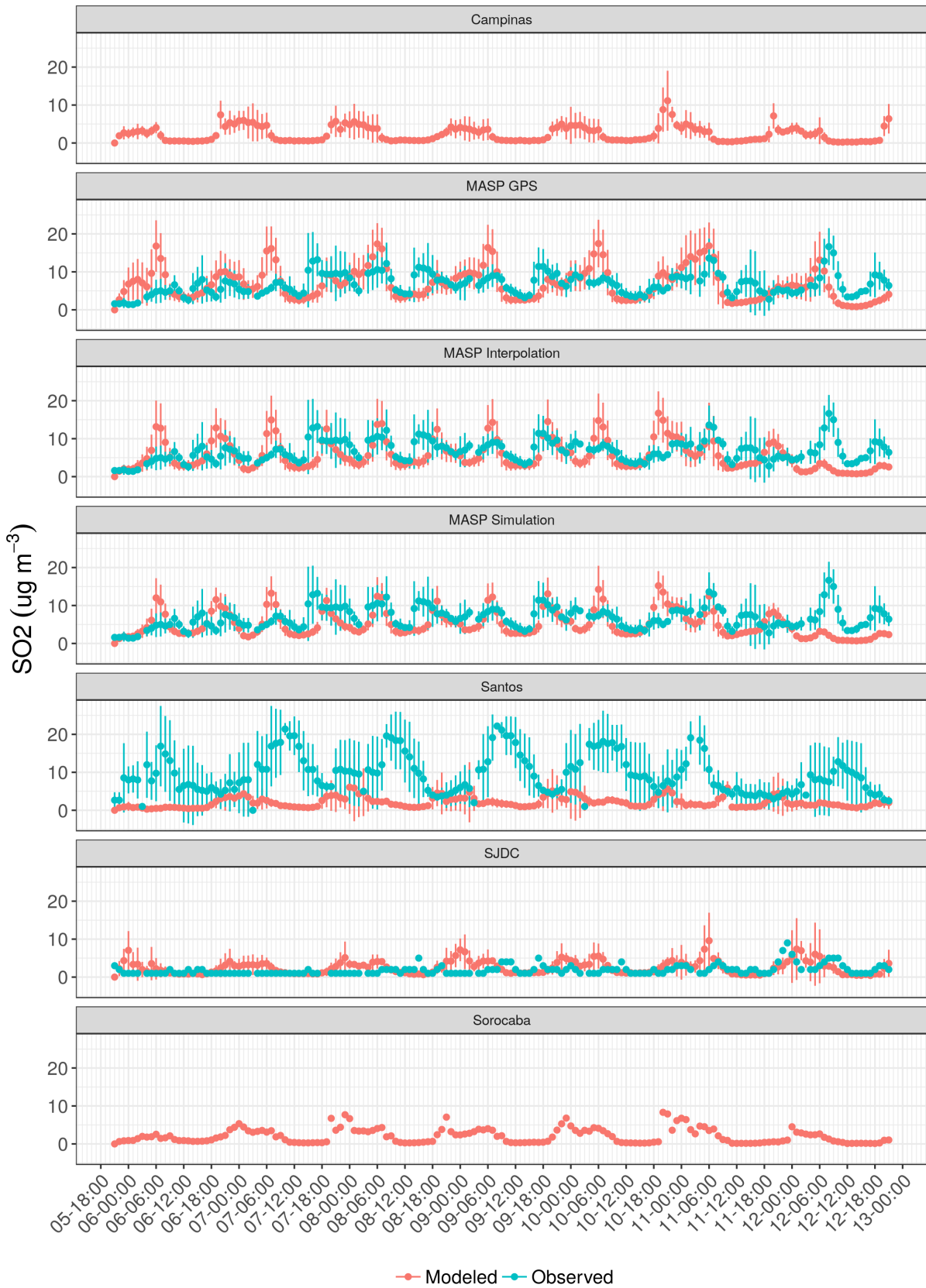


Figure 46 – Hourly mean and standard deviation of  $SO_2$  ( $ug \cdot m^{-3}$ ) by region.

Overall, the under prediction of  $O_3$  could be due to under prediction of  $NO$ . However, NMHC should also be evaluated. Unfortunately, there are no NMHC observations available for this period. It should also be noted that simulated  $O_3$  diurnal variation is driven by temperature, which was very well simulated.

Another important point that affects all scenarios is the absence of other sources of emissions such as industrial or other that combustion wood such as Pizzerias. Therefore, air quality modeling do not match observations not only by over prediction of wind speed, but also missing sources.

### 5.3.3 Evaluation of the WRF-Chem simulation in each region

The last part of this section shows the evaluation with the statistics indexes shown on Table (17). Temperature was the best forecast variable. It was well modeled in most of regions, except in SJDC, where the errors were bigger. In the case of wind speed, the better correlations were obtained in the region of SJDC and Santos. CO correlation was slightly better for MASP than for Campinas, however the MFB and MFE had higher magnitudes in Campinas than in MASP. Regarding the NO, this pollutant has observations in all regions and better correlation in Santos followed by MASP, however, the errors were smaller in MASP. In the case of  $SO_2$  and  $NO_2$  the simulations were very distant from the observations and the correlation was very low. The  $O_3$  correlations reached mean values above 0.62 in MASP and 0.61 in Santos. There were even stations such as Itaim Paulista where correlations reached 0.82. The results for traffic data from Interpolation and Simulation were similar, but higher correlations were found for  $NO$  and  $O_3$ .  $NO_2$  is systematically underestimated in all regions, which suggest that the fraction  $NO/NO_2$  is not well represented in the emission factors.

## 5.4 Discussion and uncertainty

This study faced the complex problem of air pollution, from the development of detailed bottom-up vehicular emissions inventories, inclusion of speed in constant emission factors by age of use, evaluation with fuel sales in each region and air pollutant simulation with an "online" air quality model. Each stage included difficulties, challenges, assumptions and uncertainties discussed in this section, that provides opportunities to improve in future researches.

Developing a bottom-up vehicular emissions inventory with data from massive GPS recordings is a novel approach to my knowledge, a new approach approach that requires further improvements and this study provided an exploratory analyses. Traffic flow speed provides uncertainty for generating traffic flow because it depends on the quality and frequency of GPS recordings. The main difficulty with this approach resided



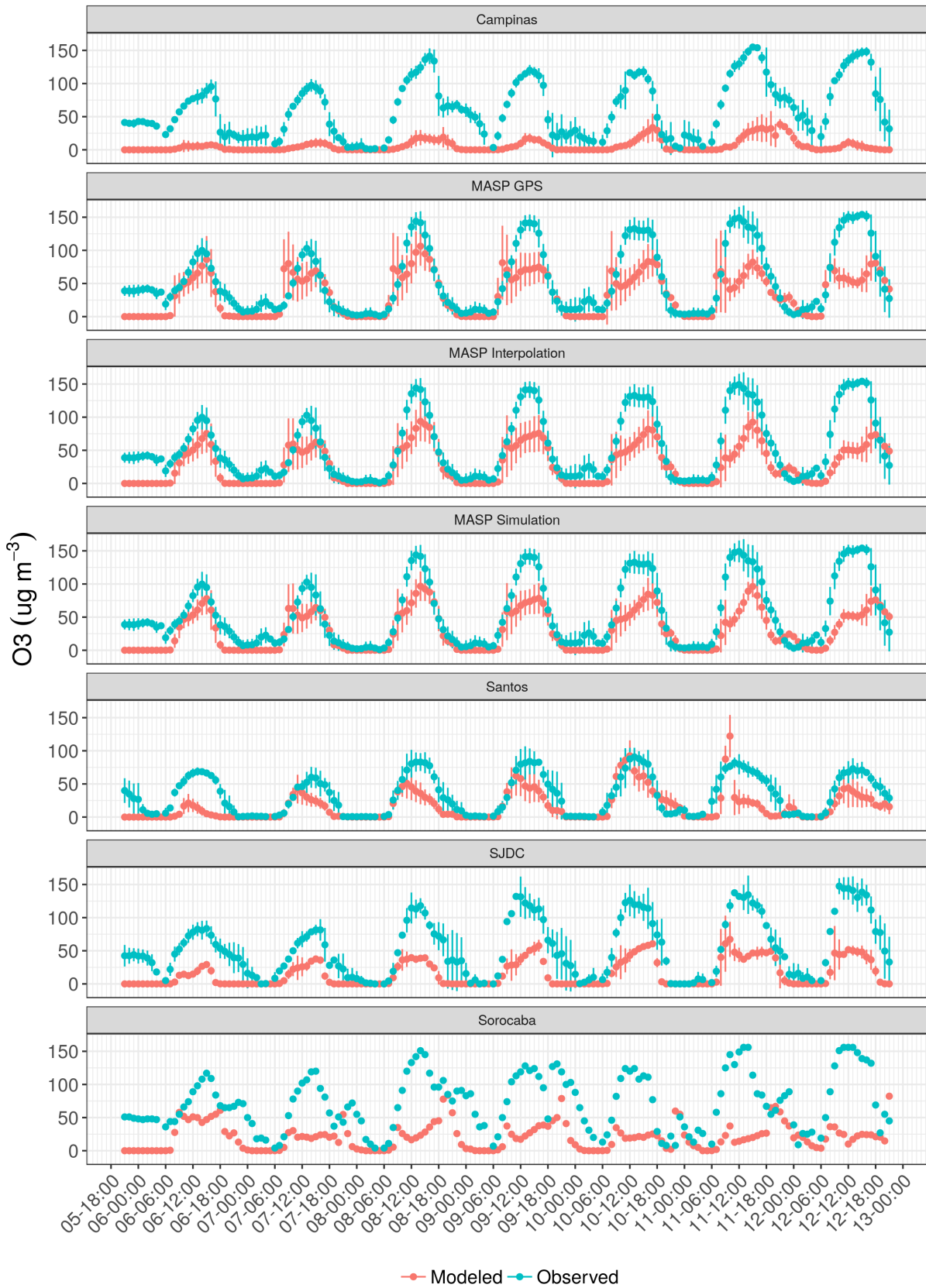


Figure 47 – Hourly mean and standard deviation of  $O_3$  ( $ug \cdot m^{-3}$ ) by region.



Table 17 – Evaluation of atmospheric simulation in each region

Region	MB	MFB	MFE	RMSE	COR	Parameter	Data
Campinas	0.34	2.13	6.81	11.62	0.92	Temperature	
MASP	0.65	2.78	11.39	0.35	0.9		
Santos	-0.71	-2.59	9.14	0	0.59		
SJDC	-3.21	17.23	104.75	5.04	0.39		
Sorocaba	0.41	2	10.14	0	0.92		
Campinas	1.06	41.59	57.74	0.03	0.12	Wind Speed	
MASP	1.24	54.26	74.34	0.01	0.1		
Santos	1.99	80.3	88.89	0	0.26		
SJDC	-3.21	17.23	104.75	5.04	0.39		
Sorocaba	0.53	19.91	49.72	0	0.05		
Campinas	-0.42	-75.92	84.09	0.18	0.29	<i>CO</i>	
MASP	0.59	7.59	72.27	0.81	0.30		
Campinas	-0.64	46.82	113.3	4.15	0.25	<i>NO</i>	
MASP	-19.11	-63.96	126.16	2.87	0.44		
Santos	-39.61	-92.19	140.32	1.7	0.55		
SJDC	-3.21	17.23	104.75	5.04	0.39		
Sorocaba	-1.03	-5.34	142.23	1.87	0.37		
Campinas	-17.23	-152.01	152.80	3.78	0.16	<i>NO<sub>2</sub></i>	GPS
MASP	-45.80	-146.67	146.96	1.05	0.25		
Santos	-52.65	-159.30	159.88	0.60	0.02		
SJDC	-17.98	-142.65	148.96	0.34	-0.14		
Sorocaba	-10.18	-124.31	137.12	3.20	0.05		
MASP	1.82	6.34	66.03	1.05	0.2	<i>SO<sub>2</sub></i>	
Santos	-11.95	-102.48	117.18	0.42	0.06		
SJDC	1.66	40.49	75.88	0.04	0.02		
Campinas	-56.83	-164.75	166.18	2.44	0.45	<i>O<sub>3</sub></i>	
MASP	-23.93	-75.33	108.53	15.57	0.62		
Santos	-17.45	-82.32	122.13	7.97	0.61		
SJDC	-3.21	17.23	104.75	5.04	0.39		
Sorocaba	-57.63	-126.83	129.22	7.8	0.22		
MASP	-0.25	-41.44	69.72	0.07	0.28	<i>CO</i>	Simulation
	-21.35	-66.76	123.79	1.73	0.48	<i>NO</i>	
	-45.80	-146.67	146.96	1.05	0.25	<i>NO<sub>2</sub></i>	
	34.40	112.67	120.07	11.68	0.15	<i>SO<sub>2</sub></i>	
	-28.16	-89.43	113.87	13.33	0.71	<i>O<sub>3</sub></i>	
	-0.29	-47.87	75.13	0.08	0.26	<i>CO</i>	Interpolation
	-21.38	-66.84	123.79	1.76	0.48	<i>NO</i>	
	-47.24	-154.10	154.14	0.56	0.20	<i>NO<sub>2</sub></i>	
	34.93	111.99	119.19	12.12	0.15	<i>SO<sub>2</sub></i>	
	-28.13	-89.17	114.06	13.36	0.71	<i>O<sub>3</sub></i>	

in the frequency of consecutive observations for the same vehicle, which restricted the analyses. The three types of vehicles identified "Cars", "Taxi" and "Trucks", and the vehicles "Undefined" provided information for calculating speeds for each type of vehicle. The speed later served as proxy for obtaining the traffic flow for "Cars". The process of generating traffic flow for "Trucks" also consisted in using the speed of Trucks, but when there was no available information, the average speed of all vehicles was used. This process seems appropriate for "Cars" but in the case of "Trucks", it could lead to assignment in the same streets of "Cars" which could not necessarily be true. Circulation of "Trucks" is driven by delivering goods and this determines the route "Trucks" follow. Therefore, this traffic flow could not have an exact spatial representation. This could be one reason that GPS generated PM and  $NO_x$  emissions have different spatial distribution on Fig. (37) from Simulations and Interpolation emissions. This figure shows clearly a ring of  $NO_x$  and PM emissions surrounding the city of São Paulo, and in the case of the emissions generated with traffic from GPS, this ring does not appear. On the other hand, this data provided insights of "Trucks" flow that hardly could be obtained from other sources. For example, "Trucks" on Santos Region are important because this city is a port, attracting and generating several "trucks" trips. Traffic data from GPS in Santos identified streets with high number of "Trucks" near the port and also, in the streets at southwest of the study area. Unequivocally, updating and enhancing this information with new data sets could improve traffic flow generation. Therefore, traffic data from Internet GPS recordings is a promising new traffic data source for emissions inventories.

The traffic simulation from CET contained LDV and HGV simulation for morning rush hour, and the simulation from SPtrans contained Urban buses. Each simulation has an uncertainty in the exact number of vehicles circulating. CET performs traffic counts to calibrate traffic simulations (IBARRA-ESPINOSA, 2017) in the main routes at MASP. Both simulations include a limited number of local streets as virtual links located at the centroid of each zone of Origin-Destination. These links represent real traffic over virtual links, therefore, provides an associated element of uncertainty. However, most of traffic occurs in mains streets, such as *motorways*, *trunks* and *primary*, therefore the uncertainty of virtual links should be small for a emissions grid with a spacing of 3 km.

The interpolation of LDV and MC from traffic counts produced results very similar to the traffic simulation. This means that traffic interpolation is a suitable method for cities that do not have traffic simulations. Regarding the speed, it was also interpolated from morning rush hour but the hourly profile was generated using a simple scheme of 3 hours of rush hour speeds at morning and evening, free flow speeds from speeds limits registered at OpenStreetMap and an average in the hours in-between. This method leads to speeds at free flow hours higher than the average of the speeds of GPS and traffic simulation, on average, as shown on Fig. (35).

Traffic calibration using fuel sales resulted in consistent traffic flows regarding the origin of the data. This can be seen in the  $CO_2$  emissions which resulted in  $21 \cdot 10^6 t \cdot y^{-1}$  for the whole MASP. As  $CO_2$  depends more directly on fuel consumption, it should be expected that calibration with fuel sales produced satisfactory results. There is a source of uncertainty in fuel consumption of diesel, because sales statistics (ELECTRICA, 2014) is given for total diesel, without discriminating the fuel used only for automotive purpose. For instance, there are many electricity generators that consumes diesel. In addition, as trucks travel long distances for delivering goods, fuel bought in one region could be consumed outside it, adding uncertainty. The criteria that 85% of diesel sales is used by vehicles may be conservative, thus it is an approach subject to review.

The emissions estimated with VEIN were consistent in each region. The difference in MASP region for each pollutant presents differences associated with the speeds and amount of vehicles. As Table (16) shows, the VKM in MASP Interpolation is  $2570 \cdot 10^{12} km \cdot y^{-1}$  and MASP Simulation  $2796 \cdot 10^{12} km \cdot y^{-1}$ , which are very similar, but in MASP GPS this value is  $775 \cdot 10^{12} km \cdot y^{-1}$ , approximately 3.4 times lower. However, the emissions in MASP GPS are higher for pollutants emitted from gasohol and ethanol vehicles. These higher emissions are due to lower speeds, as shown in Fig. (35), and then emission factors are higher, as shown in Fig. (20). This highlights the importance of including kinematic parameters in the estimation of vehicular emissions. Regarding the emission factors, there are elements of uncertainty in the emission factors measurement. LDV and MC emission factors informed by (CETESB, 2015) come from dynamometer measurements and it has been discovered that these emissions are not representative of real world emissions (PELKMAN; DEBAL, 2006). In addition, car manufacturers incorporated software in the vehicles to cheat emission certification tests. In Brazil, the Institute of Natural Resources and Environment (IBAMA) fined Volkswagen with USD  $15 \cdot 10^6$  after tests using portable emissions measurements made by CETESB (IBAMA, 2017). However, the diesel fleet of LDV is small and the impacts on vehicular estimation may not be significant. These examples show the importance to start a program of real world emissions measurements in Brazil, and that (CETESB, 2015) emission factors could be underestimated. In addition, there are uncertainty also in the deterioration factors used in this study, because they were calculated for European vehicles (NTZIACHRISTOS; SAMARAS, 2016) and they may not fully represent Brazilian vehicular deterioration. Lastly, the emission factors of certification tests FTP-75 informed by (CETESB, 2015) only include one value for the all cycle, without discriminating each phase, and it is known that cold start emissions can be very important, and they increase when the temperature decrease (NTZIACHRISTOS; SAMARAS, 2016).

There are other corrections that could improve the emission factors, such as the incorporation of percentage of load in HGV and the effect of the road inclination. This could be important in the streets that connect São Paulo and Santos, because

Table 18 – Comparison with vehicular emissions inventories in different cities ( $t \cdot y^{-1}$ )

City	Year	PM	CO	$NO_x$	NMHC	$CO_2 \cdot 10^6$	Reference
MASP, Brazil	2014	2111	178776	71772	69975	21.01	This study
MASP, Brazil	2014	1484	162896	54334	34824	14.28	(CETESB, 2015)
Rio de Janeiro, Brazil	2010	2917	583667	68400	69689		(SOUZA et al., 2013)
Santos, Brazil	2014	795	24814	15724	10556	2.43	This study
Santos, Brazil	2014	182	13497	6157	2628	1.38	(CETESB, 2015)
Sorocaba, Brazil	2014	574	24098	11770	10568	2.16	This study
Sorocaba, Brazil	2014	257	20203	8832	4070	1.9	(CETESB, 2015)
SJDC, Brazil	2014	533	26048	11661	11864	2.13	This study
SJDC, Brazil	2014	271	27406	10116	5287	2.4	(CETESB, 2015)
Campinas, Brazil	2014	2497	98073	51826	35820	9.14	This study
Campinas, Brazil	2014	377	34890	13851	7180	3.3	(CETESB, 2015)
Porto Alegre, Brazil	2004	2350	195740	34111	23450		(TEIXEIRA; FELTES; SANTANA, 2008)
Manizales, Colombia	2014	800	43400	4900	9600	0.45	(GONZÁLEZ et al., 2017)
Santiago, Chile	2010	1275	199884	40126	16668	7.3	(ESCOBAR et al., 2007)
Buenos Aires, Argentina	2006	6370	569000	81900	69800	11.5	(D'ANGIOLA et al., 2010)

the load could be different in each trip. In addition, the road inclination is significant because Santos is near sea level, whereas São Paulo is over 760 meters above sea level. The streets with the highest inclinations are located near Cubatão, an industrial district between Santos and São Paulo, and including this effect would change the atmospheric chemistry in simulations at the interface between both cities.

It was compared the inventories of MASP simulation and GPS Sorocaba, Campinas, Santos and SJDC with the emissions of different cities in Latin America shown on Table (18). The first comparison is between the official emissions inventory for MASP, which uses a top-down approach with statistics of fleet and representative speeds. In this case, the CO emissions are quite similar, being the 10% overestimated. The differences with the other pollutants is bigger, for example, the  $CO_2$  estimative is 1.5 times higher.

The comparison with the official inventories for the other cities show that this study emissions are higher in all cases. Values for Rio de Janeiro, Porto Alegre and Buenos Aires shows the estimates in those cities are higher, which could be due to the fact the the base years are 2010, 2004 and 2006 respectively and the circulating fleet had older technology allowing more emissions. The comparison with Santiago seems to give reasonable estimates despite that the fleet is older, however, the number of vehicles circulating in Santiago is smaller because of the size of the city and road network. The only city in which the NMHC emissions were significantly higher than  $NO_x$  was Manizales. This seems to be related to the evaporative emissions estimations which is based on an hourly distribution of trips based on a survey with number of times that the engine started. In Manizales, the survey shows that there are 5 trips per day for light vehicles and in the present study, 4.5 based on literature. This small difference in number of trips per day can have a greater impact on evaporative emissions estimations, therefore, improving the knowledge of this parameter could enhance the inventory in São Paulo.

WRF simulations produced good values of temperature, as shown on Fig (41) and correlations in Table (17). However, the simulation of wind speed did not produce satisfactory results, with lower correlations, with simulated wind speeds higher than observed ones, except in SJDC. These higher winds could disperse more the pollutants emitted and thus result in lower simulated concentrations of the primary pollutants, which was exactly what happened in MASP. Thus, it is very important to have meteorological simulations of good quality, because the concentrations, formation and transport of pollutant depend on it. It has been shown the the sea breeze plays an important role in local winds in São Paulo (FREITAS et al., 2007), trapping the pollutants and increasing their concentrations (ANDRADE et al., 2015). Therefore, future simulations should be performed in order to obtain better representations of meteorology.

The air pollutants concentrations obtained with the emissions inventories presented well agreement with observations in some stations across the regions. Specifically, the pollutants with better agreement was  $O_3$ , followed by NO. In both cases, the model did not reach the peak of observations, however, the simulated concentrations were reasonable. The simulation presented in this study covers painting emissions, however, this is not a comprehensive inventory and naturally, **there are missing sources that could help to improve the matching between observations and simulations.** Despite this, air pollutants simulation is satisfactory in the case of  $O_3$ , for example, with higher regional correlation of 0.71 in MASP with traffic data from Simulations and Interpolations. Although the comparison between observed and simulated concentrations present high variations, VEIN emissions maps are definitely a great advance when compared to the maps being used nowadays at LAPAt. The horizontal resolution used in the air quality simulation does not allow to take full advantage of VEIN high spatial resolution. However, this characteristic could suit street level air quality models, such as Aria City model (<http://www.aria.fr/projets/aircity/>).

The apparent similarity was evaluated with the Wilcoxon test as shown on Tables 19 and 20 (Appendix A). This tables indicates the columns Sim for traffic simulation, Int for traffic interpolation and Gps, for traffic from GPS recordings. The difference in air pollutant with each scenario is statistically significant in most of stations. However, the stations whose differences in concentrations between scenarios were not significant had the same result for all the pollutants. This means that results are not too different, independent of the scenario. In the case of  $CO$ , the stations with no significant difference are Congonhas, São Caetano do sul, Cerqueira Cesar and Osasco. In the case of  $NO$ , Santo Amaro, Cerqueira Cesar and Ipen USP. In the case of  $NO_2$ , Santo Amaro. In the case of  $SO_2$ , Santo Amaro, Pedro II and Ipen USP. Lastly, for  $O_3$ , Santo Amaro, Pedro II and Cerqueira Cesar.

The similarity between the atmospheric concentrations simulated in MASP with

data from Simulations and Interpolations, suggest that it is possible to elaborate vehicular emissions inventory using traffic counts interpolations. This is very important because most cities do not count with the base information for developing traffic simulations. Moreover, new technologies are providing increasingly new data that can be used for developing inventories that could enhance and improve inventories and simulations in different parts of the world. In this way, the VEIN model provides flexibility in estimating emissions adapting to different types of traffic data and emissions factors. Also, the semi-automatic conversion to inputs of WRF-Chem using the model AAS4WRF (VARA-VELA et al., 2016) is very helpful.

## 6 Conclusions

This study was conducted aiming at modeling air pollution in São Paulo using bottom-up vehicular emissions inventories. For the Metropolitan Area of São Paulo (MASP), there are some traffic activity data available, such as traffic counts and traffic simulations. However, this information must be transformed into hourly gridded emissions to be used as input in air quality modeling. This "transformation" could be done with the development of a bottom-up vehicular emissions inventory model, the VEIN model R-package.

Traffic simulations were provided by CET (Traffic Engineering Company) for Light Duty Vehicles (LDV) and Heavy Good Vehicles (HGV) and by SPTrans (Secretary of Transport and Mobility of São Paulo) for Urban Buses (UB), and they supplied the traffic flow for a morning rush hour at MASP. Traffic counts were available for morning and evening rush hours. These counts were interpolated using negative binomial and quasi-poisson regressions resulting in traffic flows for the 6 rush hours with best results for LDV and motorcycles (MC). Traffic data from GPS tracking records is a novel and promising approach. GPS data was available for all the metropolitan areas for LDV and HGV from 06th - 11th October, 2014, on an hourly basis. However, the frequency of consecutive GPS records were low, compromising its results. Vehicle speeds were calculated using two consecutive position records and these speeds were used as proxy for traffic flow for LDV and HGV. Traffic flows from different sources were very different and calibration with fuel sales showed to be an effective way for harmonizing them.

VEIN model was then applied to estimate the emissions considering three scenarios: outside MASP, GPS records were used; inside MASP 1) GPS, 2) traffic count interpolation, and 3) CET simulations were considered. SPTrans simulation was also used in all the 3 scenarios. VEIN distributes the traffic flows in different vehicle types, consuming different fuels and with 1 to 40 years of age. The emission factors are functions of these characteristics, but are also dependent on the vehicle speed. Emissions inventories (EI) from VEIN showed that GPS-EI had highest values for LDV emissions (due to lower speeds and, thus, higher emissions). Traffic simulation and interpolation EI's were very similar, and thus, traffic count interpolation may be relied upon as input data. The emissions signature (emissions by age of use) shows that a scrapping policy for taking older vehicles out of streets could cut emissions of *NMHC*, *CO*, *CH<sub>4</sub>*, *PM* and Evaporative. VEIN emissions were overestimated for all the pollutants in all the metropolitan regions when compared to CETESB estimated emissions for the year 2014.

Finally, these emissions scenarios were used as input in the WRF-Chem model. Overall, the simulated primary pollutants concentrations were lower than the observed ones, despite the higher EI's. This could be due to stronger simulated winds. However, the simulated diurnal variation resembled the observed one, indicating that the temporal distribution of the emissions is consistent. Besides, there were unaccounted emission sources such as industries and a seaport at Santos. Despite differences between observed and simulated concentrations, VEIN presented a great advance in emissions maps which could be used in models with higher horizontal resolution.

Therefore, VEIN turned out to be a very effective tool to estimate vehicular emissions. Furthermore, the VEIN model is free of charge and open source, making it suitable for applications in developing countries with lack of traffic and emission factors information. Besides, it is multi platform, and can be run in R with most of operative systems including Windows, Mac, Linux, Solaris, etc.

## 6.1 Suggestions for future works

- New emission factors based on real world measurements with Portable Emissions Measurements Systems must be performed and added to VEIN
- New vehicular emissions inventory should account for effect of load and road inclination.
- Interpolation of hourly traffic of Trucks and Buses needs more information inherent to their activity, such as origin and destinations.
- The frequency of consecutive GPS records of the same vehicle is important and efforts must be made to obtain or generate data with higher frequency.



## 7 Academic literature produced during the Ph.D.

### 7.1 Publications with peer review process

- Ibarra S., Ynoue R. (2017). REMI model: Bottom - up emissions inventories for cities with lack of data. *Journal of earth sciences and geotechnical engineering*.
- Andrade MF., Ynoue R., Freitas E., Todezco E., Vara-Vela A., Ibarra S.m Martins L., Martins J and Carvalho V. (2015). Air quality forecasting system for outtheastern Brazil. *Frontiers in Environmental Science*.

### 7.2 Submitted publications under review process

- Ibarra S. Ynoue R., Gianotti M (2017). A comparison between bottom-up vehicular emissions inventories using traffic data from Internet GPS recordings, traffic simulations and traffic interpolations with VEIN model and atmospheric simulation with WRF-Chem. Paper in preparation for *Science of Total Environment*
- Ibarra-Espinosa, S., Ynoue, R., O'Sullivan, S., Pebesma, E., Andrade, M. D. F., and Osses, M.: VEIN v0.2.2: an R package for bottom-up Vehicular Emissions Inventories, *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2017-193>, in review, 2017.
- Ibarra S., O'Sullivan S., Osses M., Ynoue R. (2017). Negative binomial and quasi-poisson regressions of hourly traffic data with OpenStreetMap. Paper in review *Computers, Environment and Urban Systems*
- Fink, G., O'Sullivan S., Ibarra-Espinosa S., Grisi S., Saldiva, P., Brentani, A. (2017). Exposure to Outdoor Air Pollution and Birth Outcomes – Evidence from São Paulo's Western Region Project. Paper in preparation to *Plos One*.

### 7.3 Software

- Ibarra-Espinosa S (2017). vein: Vehicular Emissions Inventories. <https://CRAN.R-project.org/package=vein>.

## 7.4 Conferences

- Ibarra S., Ynoue R., and Andrade MF. (2017). High Resolution vehicular emissions inventory in Shanghai China: Application of VEIN model. Japan Geoscience Union (JPGU) meeting, Chiba, Japan from May 20 th to 25 th , 2017.
- Ibarra S. and Ynoue R. (2016). REMI model: Bottom-up emissions inventories for cities with lack of data. 21 International Transport and Air Pollution Conference “TAP 2016”. Lyon, France from May 24 th to 26 th , 2016.
- Ibarra S., Vara-Vela A., Rehbein A., Ynoue R. (2015). High resolution air pollutant simulation for the Metropolitan Region of Porto Alegre. In: IX Workshop Brasileiro de Micrometeorologia, 2015, Santa Maria.
- Ibarra S., Vara-Vela A., Ynoue R. (2015). Vehicular bottom-up emissions inventory and atmospheric simulation for 58 urban centers of South America. In: 11th International Conference on Southern Hemisphere Meteorology and Oceanography, 2015, Santiago.
- Ibarra S., Ynoue R. and Vara-Vela A. (2014). Development and evaluation of a vehicular emissions inventory based in traffic counts for Metropolitan Region of São Paulo. Joint 13th IGAC Science Conference and 13th Quadrennial iCACGP Symposium held at Natal Convention Center (NCC), Natal, Brazil, from September 22 to 26, 2014.

## 7.5 Other events

- Invitation to run VEIN model in China by the Laboratory of Tibetan Environment Changes and Land use, Chinese Academy of Sciences
- Workshop on emissions by EMISIA, May 2016, Lyon, France.
- Transport and Air Pollution Meeting, September 2016, IAG-USP.
- IAG Science Day, May 2016, IAG-USP.
- Meteorologia para a terceira idade, September 2016, IAG-USP.
- Meteorologia para a terceira idade, June 2017, IAG-USP.
- Simposio de Iniciação Científica, September 2016, IAG-USP.
- Internship at Transport Research Laboratory, May, 2016, TRL, UK.
- Participation in Virada Científica USP, 2014.

- Participation in Virada Científica USP, 2015.
- Partnership with CET to evaluate the effect of change in speeds on air pollution, 2015



# Bibliography

ABRACICLO. *Anuario 2014*. 2014. <<http://www.abraciclo.com.br/>>. Cited in the page 78.

ABRAFATI. *Números do Sector; Milhões de litros*. [S.l.], 2017. Disponível em: <<http://www.abrafati.com.br/indicadores-do-mercado/numeros-do-setor/>>. Cited in the page 86.

ANDRADE, M. d. F. et al. Air quality in the megacity of são paulo: Evolution over the last 30 years and future perspectives. *Atmospheric Environment*, Elsevier, v. 159, p. 66–82, 2017. Cited 3 times in the pages 53, 54, and 58.

ANDRADE, M. d. F. et al. Air quality forecasting system for southeastern brazil. *Frontiers in Environmental Science*, Frontiers, v. 3, p. 1–12, 2015. Cited 8 times in the pages 30, 31, 54, 73, 88, 91, 117, and 135.

ANDRADE, M. de F. et al. Vehicle emissions and pm<sub>2.5</sub> mass concentrations in six brazilian cities. *Air Quality, Atmosphere & Health*, Springer, v. 5, n. 1, p. 79–88, 2012. Cited in the page 31.

ANDRÉ, M. The artemis european driving cycles for measuring car pollutant emissions. *Sci. Total Environ.*, Elsevier, v. 334, p. 73–84, 2004. Cited in the page 89.

ANFAVEA. *Estatísticas produção, vendas e exportação de automóveis*. 2014. <<http://www.anfavea.com.br/estatisticas.html>>. Cited in the page 78.

ARTESP. *Toll station count data*. 2012. [Http://www.artesp.com.br](http://www.artesp.com.br). Cited in the page 63.

BARLOW, T. et al. A reference book of driving cycles for use in the measurement of road vehicle emissions. *TRL Published Project Report*, 2009. Cited 2 times in the pages 57 and 81.

BAUER, D. F. Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, Taylor & Francis, v. 67, n. 339, p. 687–690, 1972. Cited in the page 95.

BIVAND, R.; KEITT, T.; ROWLINGSON, B. *rgdal: Bindings for the Geospatial Data Abstraction Library*. [S.l.], 2016. R package version 1.1-10. Disponível em: <<https://CRAN.R-project.org/package=rgdal>>. Cited 2 times in the pages 60 and 61.

BIVAND, R.; LEWIN-KOH, N. *mapproj: Tools for Reading and Handling Spatial Objects*. [S.l.], 2015. R package version 0.8-34. Disponível em: <<http://CRAN.R-project.org/package=mapproj>>. Cited in the page 61.

BIVAND, R.; PEBESMA, E.; GOMEZ-RUBIO, V. *Applied spatial data analysis with R*. Springer, NY, 2013. Disponível em: <<http://www.asdar-book.org/>>. Cited in the page 60.

BIVAND, R.; RUNDEL, C. *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. [S.l.], 2016. R package version 0.3-21. Disponível em: <<https://CRAN.R-project.org/package=rgeos>>. Cited 2 times in the pages 60 and 72.

BOIAN, C.; ANDRADE, M. d. F. Characterization of ozone transport among metropolitan regions. *Revista Brasileira de Meteorologia*, SciELO Brasil, v. 27, n. 2, p. 229–242, 2012. Cited in the page 31.

BOND, T. C. et al. Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, v. 118, n. 11, p. 5380–5552, 2013. ISSN 2169-8996. [Http://dx.doi.org/10.1002/jgrd.50171](http://dx.doi.org/10.1002/jgrd.50171). Cited in the page 73.

BOULDER, C. U. *The NCAR Command Language (NCL)(version 6.4. 0)*. 2017. Disponível em: <<http://dx.doi.org/10.5065/D6WD3XH5>>. Cited in the page 75.

BOULTER, P. *Environmental traffic management: A review of factors affecting cold start emissions*. Old Wokingham Road, Crowthorne, Berkshire, RG45 6AU, UK, 1997. Cited in the page 58.

Branco G. and Branco F. *Determinação dos Hidrocarbonetos no gás de escape de um veículo Total Flex*. [S.I.], 2016. Personal communication. Cited in the page 87.

BRUNI, A. D. C.; BALES, M. P. *Curvas de intensidade de uso por tipo de veículo automotor da frota da cidade de São Paulo*. [S.I.], 2013. [Http://cetesb.sp.gov.br/veicular/relatorios-e-publicacoes/](http://cetesb.sp.gov.br/veicular/relatorios-e-publicacoes/). Cited 4 times in the pages 59, 63, 68, and 83.

Bureau of Public Roads. *Traffic and assignment manual*. Washington D.C., 1964. Cited 2 times in the pages 56 and 64.

Bureau of Transportation Statistics. *Daily Passenger Travel*. [S.I.], 2017. Disponível em: <[https://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/highlights\\_of\\_the\\_2001\\_national\\_household\\_travel\\_survey/html/section\\_02.html](https://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/highlights_of_the_2001_national_household_travel_survey/html/section_02.html)>. Cited in the page 84.

CARSLAW, D. C.; ROPKINS, K. Openair—an r package for air quality data analysis. *Environmental Modelling & Software*, Elsevier, v. 27, p. 52–61, 2012. Cited in the page 70.

CARTER, W. P. Development of ozone reactivity scales for volatile organic compounds. *Air & waste*, Taylor & Francis, v. 44, n. 7, p. 881–899, 1994. Cited in the page 87.

CARTER, W. P. Documentation of the saprc-99 chemical mechanism for voc reactivity assessment. *Contract*, v. 92, n. 329, p. 95–308, 2000. Cited in the page 87.

CARTER, W. P. Updated maximum incremental reactivity scale and hydrocarbon bin reactivities for regulatory applications. *California Air Resources Board Contract*, p. 07–339, 2009. Cited in the page 87.

CET. *Pesquisa de Monitoração da Fluidez; Desempenho do Sistema Viário Principal; Volume e Velocidade - 2012*. 2013. Accessed on 18-07-2017. Disponível em: <["http://cetsp.com.br/sobre-a-cet/relatorios-corporativos.aspx"](http://cetsp.com.br/sobre-a-cet/relatorios-corporativos.aspx)>. Cited in the page 40.

CET. *Simulação de tráfego para Região Metropolitana de São Paulo*. 2014. [Http://www.cetsp.com.br](http://www.cetsp.com.br). Cited in the page 63.

CETESB. *Emissões Veiculares no Estado de São Paulo*. 2013. Accessed on 18-07-2017. Disponível em: <<http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/>>. Cited 3 times in the pages 30, 44, and 105.

CETESB. *Emissões Veiculares no Estado de São Paulo*. 2013. <<http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/>>. Cited in the page 54.

CETESB. *Emissões Veiculares no Estado de São Paulo 2014*. 2015. <<http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/>>. Cited 21 times in the pages 17, 18, 33, 61, 63, 66, 67, 68, 69, 78, 79, 80, 81, 82, 83, 84, 85, 87, 97, 133, and 134.

CHEN, F. et al. The integrated wrf/urban modelling system: development, evaluation, and applications to urban environmental problems. *International Journal of Climatology*, Wiley Online Library, v. 31, n. 2, p. 273–288, 2011. Cited in the page 93.

CHEN, R.-H. et al. Cold-start emissions of an si engine using ethanol–gasoline blended fuel. *Applied Thermal Engineering*, Elsevier, v. 31, n. 8, p. 1463–1467, 2011. Cited in the page 58.

CHOU, M.-D.; SUAREZ, M. J. A solar radiation parameterization (clirad-sw) for atmospheric studies. *NASA Tech. Memo*, v. 10460, p. 48, 1999. Cited in the page 93.

COELHO, C. A. et al. The 2014 southeast brazil austral summer drought: regional scale mechanisms and teleconnections. *Climate Dynamics*, Springer, v. 46, n. 11-12, p. 3737–3752, 2016. Cited in the page 93.

COHEN, A. J. et al. The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health, Part A*, Taylor & Francis, v. 68, n. 13-14, p. 1301–1307, 2005. Cited in the page 29.

CONAMA. *RESOLUÇÃO CONAMA Nº 18, de 6 de maio de 1986*. 1986. <<http://www.mma.gov.br/port/conama/res/res86/res1886.html>>; accessed 2017-10-19. Cited 2 times in the pages 33 and 67.

COOK, J. et al. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, v. 11, n. 4, p. 048002, 2016. Disponível em: <<http://stacks.iop.org/1748-9326/11/i=4/a=048002>>. Cited in the page 29.

CORVALÁN, R. M.; OSSES, M.; URRUTIA, C. M. Hot emission model for mobile sources: application to the metropolitan region of the city of santiago, chile. *Journal of the Air & Waste Management Association*, Taylor & Francis, v. 52, n. 2, p. 167–174, 2002. Cited 3 times in the pages 33, 55, and 78.

CORVALÁN, R. M.; VARGAS, D. Experimental analysis of emission deterioration factors for light duty catalytic vehicles case study: Santiago, chile. *Transportation Research Part D: Transport and Environment*, Elsevier, v. 8, n. 4, p. 315–322, 2003. Cited 2 times in the pages 57 and 89.

CPTEC. *Boletim tecnico*. <http://tempo.cptec.inpe.br/boletimtecnico/pt>, 2014. Cited in the page 93.

D'ANGIOLA, A. et al. On-road traffic emissions in a megacity. *Atmospheric Environment*, Elsevier, v. 44, n. 4, p. 483–493, 2010. Cited in the page 134.

- DAVIS, N. et al. Part 3: Developing countries: development and application of an international vehicle emissions model. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, v. 8, n. 1939, p. 155–165, 2005. Cited in the page 58.
- DENATRAN. *National Department of Traffic Road*. 2015. Online; accessed 04 February 2016. Disponível em: <[Availablein:http://www.denatran.gov.br/frota2015.htm](http://www.denatran.gov.br/frota2015.htm)>. Cited in the page 54.
- DENATRAN. *Frota Nacional (2017)*. 2017. <<http://www.denatran.gov.br/index.php/estatistica/610-frota-2017>>"; accessed 09-August-2017. Cited in the page 97.
- DIRECTIVE70/220/EEC. *Measures to be taken against the emission of gaseous pollutants from diesel engines for use in vehicles*. 1991. <<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31991L0441&from=EN>>. Cited in the page 67.
- DOWLE, M.; SRINIVASAN, A. *data.table: Extension of 'data.frame'*. [S.l.], 2017. R package version 1.10.4. Disponível em: <<https://CRAN.R-project.org/package=data.table>>. Cited in the page 46.
- EEA, E. Emission inventory report 1990–2011 under the unece convention on lrtap. *European Environment Agency, Luxembourg*, 2013. Cited in the page 53.
- EICKER, M. O. de et al. Spatial accuracy of a simplified disaggregation method for traffic emissions applied in seven mid-sized chilean cities. *Atmospheric Environment*, Elsevier, v. 42, n. 7, p. 1491–1502, 2008. Cited in the page 31.
- EJ-JRC/PBL. *Emission Database for Global Atmospheric Research (EDGAR), release EDGAR v4.3.1\_v2 (1970 - 2010)*. 2016. Disponível em: <<http://edgar.jrc.ec.europa.eu>>. Cited 2 times in the pages 73 and 88.
- ELECTRICA, S. de E. *Anuario estatístico de energéticos por municipio no Estado de São Paulo 2014*. [S.l.], 2014. Cited 2 times in the pages 76 and 133.
- EOM, J. et al. Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 1968, p. 20–29, 2006. Cited in the page 40.
- ESCOBAR, J. et al. Actualización del inventario de emisiones de contaminantes atmosféricos en la región metropolitana 2005. *DICTUC, Pontificia Universidad Católica de Chile. Santiago: Comisión Nacional Del Medio Ambiente Región Metropolitana*, 2007. Cited in the page 134.
- FAST, J. D. et al. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 111, n. D21, 2006. Cited in the page 93.
- FINK, G. et al. Exposure to outdoor air pollution and birth outcomes – evidence from são paulo's western region project. *Submitted to Environmental Health Perspectives*, 2017. Cited in the page 88.



- FREITAS, E. et al. A simple photochemical module implemented in rams for tropospheric ozone concentration forecast in the metropolitan area of são paulo, brazil: Coupling and validation. *Atmospheric Environment*, Elsevier, n. 39, p. 6352–6361, 2005. Cited 3 times in the pages 61, 73, and 88.
- FREITAS, E. D. et al. Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of são paulo, brazil. *Boundary-Layer Meteorology*, v. 122, n. 1, p. 43–65, Jan 2007. Cited in the page 135.
- FREITAS, S. et al. Prep-chem-src–1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models. *Geoscientific Model Development*, Copernicus GmbH, v. 4, n. 2, p. 419–433, 2011. Cited in the page 74.
- GAVIDIA-CALDERON, M. *Impacto das condições de contorno químicas na representação do ozônio troposférico na Região Metropolitana de São Paulo com o modelo WRF-Chem*. Tese (Doutorado) — Universidade de Sao Paulo, IAG, 2015. Cited in the page 92.
- GIROLDO, M. B. et al. Development of 1.6l flex fuel engine for brazilian market. In: *SAE Technical Paper*. SAE International, 2005. Disponível em: <<https://doi.org/10.4271/2005-01-4130>>. Cited in the page 33.
- GONZÁLEZ, C. et al. Relative impact of on-road vehicular and point-source industrial emissions of air pollutants in a medium-sized andean city. *Atmospheric Environment*, Elsevier, v. 152, p. 279–289, 2017. Cited in the page 134.
- GONZÁLEZ, C. et al. Relative impact of on-road vehicular and point-source industrial emissions of air pollutants in a medium-sized andean city. *Atmospheric Environment*, Elsevier, v. 152, p. 279–289, 2017. Cited in the page 89.
- GREC. *Relatorio Climatico Mensal. Monitoramento climatico Outubro 2014*. <http://www.grec.iag.usp.br>, 2014. Cited in the page 93.
- GRELL, G. A.; DÉVÉNYI, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, Wiley Online Library, v. 29, n. 14, 2002. Cited in the page 93.
- GRELL, G. A. et al. Fully coupled “online” chemistry within the wrf model. *Atmospheric Environment*, Elsevier, v. 39, n. 37, p. 6957–6975, 2005. Cited 5 times in the pages 61, 73, 74, 88, and 91.
- GROUP., U. S. C. *Regressions models with count data @ONLINE*. 2017. Accessed on 18-07-2017. Disponível em: <<https://stats.idre.ucla.edu/stata/seminars/regression-models-with-count-data/>>. Cited in the page 43.
- HIJMANS, R. J. *raster: Geographic Data Analysis and Modeling*. [S.l.], 2016. R package version 2.5-8. Disponível em: <<https://CRAN.R-project.org/package=raster>>. Cited in the page 72.
- HOEF, J. M. V.; BOVENG, P. L. Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, Wiley Online Library, v. 88, n. 11, p. 2766–2772, 2007. Cited in the page 40.

HONG, S.-Y.; NOH, Y.; DUDHIA, J. A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, v. 134, n. 9, p. 2318–2341, 2006. Cited in the page 93.

HOSHYARIPOUR, G. et al. Prediction of ground-level ozone concentration in são paulo, brazil: Deterministic versus statistic models. *Atmospheric Environment*, v. 145, p. 365 – 375, 2016. ISSN 1352-2310. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1352231016307798>>. Cited in the page 91.

IBAMA. *Ibama confirma multa de R 50 milhoes por fraude em testes de emissao e exige recall da Volkswagen*. [S.l.], 2017. Disponível em: <<http://www.ibama.gov.br/noticias/422-2017/1032-ibama-confirma-multa-de-r-50-milhoes-por-fraude-em-testes-de-emissao-e-exige-recall-da-volkswagen>>. Cited in the page 133.

IBARRA-ESPINOSA, S. *Traffic counts both ways CET 2012 Sao Paulo*. [S.l.], 2017. Accessed on 10-08-2017. Disponível em: <"<http://dx.doi.org/10.17632/rz2cymv6c2.1>">. Cited 4 times in the pages 51, 108, 109, and 132.

IBARRA-ESPINOSA, S. et al. Negative binomial and quasi-poisson regressions of hourly traffic data with openstreetmap. *Journal of Computers, Environment and Urban Systems*, 2017. Under Review. Cited in the page 50.

IBARRA-ESPINOSA, S.; YNOUE, R. Remi model: Bottom-up emissions inventories for cities with lack of data. *Journal of Earth Sciences and Geotechnical Engineering*, Scienpress Ltd, v. 1, p. 277–288, 2017. ISSN 1792-9040. Cited in the page 60.

IBARRA-ESPINOSA, S. et al. Vein v0.2.2: an r package for bottom-up vehicular emissions inventories. *Geoscientific Model Development Discussions*, v. 2017, p. 1–29, 2017. Disponível em: <<https://www.geosci-model-dev-discuss.net/gmd-2017-193/>>. Cited in the page 33.

IBGE. *IBGE divulga as estimativas populacionais dos municípios em 2014*. 2014. <<http://agenciadenoticias.ibge.gov.br/agencia-noticias/2013-agencia-de-noticias/releases/14659-asi-ibge-divulga-as-estimativas-populacionais-dos-municipios-em-2014.html>>; accessed 07-September-2017. Cited in the page 86.

IBGE. *Instituto Brasileiro de Geografia e Estatística – Población estimada 2014*. 2014. Online; accessed 12 December 2014. Disponível em: <Availablein:<http://cod.ibge.gov.br/232IH>>. Cited in the page 54.

IBGE. *IBGE divulga as estimativas populacionais dos municípios em 2016*. 2016. <<http://cgp.cfa.org.br/ibge-divulga-as-estimativas-populacionais-dos-municipios-em-2016/>>; accessed 09-August-2017. Cited in the page 97.

IHAKA, R. R: Past and future history. *Computing Science and Statistics*, v. 392396, 1998. Cited in the page 60.

INMET. *Dados meteorologicos estação de Santana, São Paulo*. [S.l.], 2016. Cited 2 times in the pages 18 and 94.

KAHLE, D.; WICKHAM, H. ggmap: Spatial visualization with ggplot2. *The R Journal*, v. 5, n. 1, p. 144–161, 2013. Disponível em: <<http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>>. Cited in the page 71.

KINNEE, E. et al. Allocation of onroad mobile emissions to road segments for air toxics modeling in an urban area. *Transportation Research Part D: Transport and Environment*, v. 9, n. 2, p. 139 – 150, 2004. ISSN 1361-9209. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1361920903000877>>. Cited in the page 33.

KUENEN, J.; TROZZI, C. Emep/eea emission inventory guidebook; coating applications. *European Environment Agency, Copenhagen*, 2016. Cited in the page 86.

LAM, W. H.; XU, J. Estimation of aadt from short period counts in hong kong—a comparison between neural network method and regression analysis. *Journal of Advanced Transportation*, Wiley Online Library, v. 34, n. 2, p. 249–268, 2000. Cited 2 times in the pages 40 and 55.

LENTS, J. et al. *São Paulo Vehicle Activity Study*. 605 South Palm Street, Suite C, La Habra, CA 90631, USA, 2004. Cited 2 times in the pages 58 and 63.

LIN, Y.-L.; FARLEY, R. D.; ORVILLE, H. D. Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, v. 22, n. 6, p. 1065–1092, 1983. Cited in the page 93.

LIU, S. et al. Model study of tropospheric trace species distributions during pem-west a. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 101, n. D1, p. 2073–2085, 1996. Cited in the page 92.

LOWRY, M. Spatial interpolation of traffic counts based on origin–destination centrality. *Journal of Transport Geography*, Elsevier, v. 36, p. 98–105, 2014. Cited 2 times in the pages 36 and 40.

LUDYKAR, D.; WESTERHOLM, R.; 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. *Science of the Total Environment*, Elsevier, v. 235, n. 1, p. 65–69, 1999. Cited in the page 58.

MACEDO, P. C. de et al. *Air Pollution Control Program by Motor Vehicles*. 3. ed. Brasília, Brazil: Brazilian Institute of Environment and Renewable Natural Resources (IBAMA), 2011. v. 3. ISBN 978-85-7399-358-1. Cited 2 times in the pages 33 and 115.

MARTINS, L. *Sensibilidade da formação do ozônio troposférico às emissões veiculares na Região Metropolitana de São Paulo*. Tese (Doutorado) — Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, Brazil, 2006. Cited 3 times in the pages 21, 87, and 88.

MARTINS, L. D. et al. Emission factors for gas-powered vehicles traveling through road tunnels in são paulo, brazil. *Environ. Sci. Technol.*, ACS Publications, v. 40, n. 21, p. 6722–6729, 2006. Cited in the page 89.

MELLIOS, G.; NTZIACHRISTOS, L. *EMEP/EEA emission inventory guidebook; Gasoline evaporation from vehicles*. [S.l.], 2016. Cited 5 times in the pages 58, 59, 63, 82, and 86.

METRO. *Pesquisa Origem e Destino*. 2017. Accessed on 10-08-2017. Disponível em: <"<https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino>">. Cited 2 times in the pages 37 and 38.

Ministerio de Medio Ambiente. *Air Pollution Control Program by Motor Vehicles PROCONVE PROMOT*. 2011. Disponível em: <[http://www.ibama.gov.br/phocadownload/veiculosautomotoresmanual%20proconve%20promot\\\_english.pdf](http://www.ibama.gov.br/phocadownload/veiculosautomotoresmanual%20proconve%20promot\_english.pdf)>. Cited 3 times in the pages 33, 67, and 82.

Ministerio do Meio Ambiente. *Primeiro Inventario Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários*. 2011. Disponível em: <[Availablein:http://www.mma.gov.br/estruturas/163/\\_publicacao/163\\_publicacao27072011055200.pdf](http://www.mma.gov.br/estruturas/163/_publicacao/163_publicacao27072011055200.pdf)>. Cited 2 times in the pages 63 and 66.

MISENIS, C.; ZHANG, Y. An examination of sensitivity of wrf/chem predictions to physical parameterizations, horizontal grid spacing, and nesting options. *Atmospheric Research*, Elsevier, v. 97, n. 3, p. 315–334, 2010. Cited in the page 91.

MLAWER, E. J. et al. Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 102, n. D14, p. 16663–16682, 1997. Cited in the page 93.

MOLINA, M. J. Polar ozone depletion (nobel lecture). *Angewandte Chemie International Edition*, Wiley Online Library, v. 35, n. 16, p. 1778–1785, 1996. Cited in the page 29.

MOLINA, M. J.; MOLINA, L. T. Megacities and atmospheric pollution. *Journal of the Air & Waste Management Association*, Taylor & Francis, v. 54, n. 6, p. 644–680, 2004. Cited in the page 30.

MOLINA, M. J.; MOLINA, L. T. Megacities and atmospheric pollution. *Journal of the Air & Waste Management Association*, Taylor & Francis, v. 54, n. 6, p. 644–680, 2004. Cited in the page 97.

NOGUEIRA, T. et al. On-road emissions of carbonyls from vehicles powered by biofuel blends in traffic tunnels in the metropolitan area of sao paulo, brazil. *Atmospheric Environment*, Pergamon, v. 108, p. 88–97, 2015. Cited in the page 53.

NTZIACHRISTOS, L.; BOULTER, P. Emeq/eea emission inventory guidebook; road transport: Automobile tyre and break wear and road abrasion. *European Environment Agency, Copenhagen*, 2009. Cited 2 times in the pages 63 and 74.

NTZIACHRISTOS, L.; SAMARAS, Z. Emeq/eea emission inventory guidebook; road transport: Passenger cars, light commercial trucks, heavy-duty vehicles including buses and motorcycles. *European Environment Agency, Copenhagen*, 2016. Cited 22 times in the pages 31, 33, 44, 53, 55, 56, 57, 58, 59, 61, 63, 65, 68, 74, 75, 76, 80, 81, 82, 87, 89, and 133.

OLIVIER, J. et al. Emission database for global atmospheric research (edgar). *Environmental Monitoring and Assessment*, Springer, v. 31, n. 1-2, p. 93–106, 1994. Cited in the page 29.

OLIVIER, J. G. J. et al. 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 x 1 degree grid. *Rijksinstituut voor Volksgezondheid en Milieu RIVM*, 1996. Cited 2 times in the pages 53 and 74.

- OpenStreetMap contributors. *Planet dump* retrieved from <https://planet.osm.org> . 2017. <<https://www.openstreetmap.org>>. Cited in the page 36.
- ORTUZAR, J. d. D.; WILLUMSEN, L. G. *Modelling transport*. [S.l.]: Wiley, 2002. v. 3. Cited in the page 37.
- ORTUZAR, J. de D.; WILLUMSEN, L. G. et al. *Modelling transport*. [S.l.]: Wiley New Jersey, 1994. Cited in the page 37.
- PANIS, L. I.; BROEKX, S.; LIU, R. Modelling instantaneous traffic emission and the influence of traffic speed limits. *Science of The Total Environment*, v. 371, n. 1, p. 270 – 285, 2006. ISSN 0048-9697. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S004896970600636X>>. Cited in the page 33.
- PAULO", G. do Estado do São. *Anuário Estatístico de Energéticos por Município no Estado de São Paulo*. 2013. Accessed on 18-07-2017. Disponível em: <<http://www.energia.sp.gov.br/>>. Cited in the page 45.
- PEBESMA, E. *sf: Simple Features for R*. [S.l.], 2016. R package version 0.4-3. Disponível em: <<https://github.com/edzer/sfr/>>. Cited 2 times in the pages 47 and 73.
- PEBESMA, E.; BIVAND, R. *Classes and methods for spatial data in R*. 2005. Disponível em: <<http://cran.r-project.org/doc/Rnews/>>. Cited 3 times in the pages 61, 70, and 89.
- PEBESMA, E.; MAILUND, T.; HIEBERT, J. Measurement units in R. *The R Journal*, v. 8, n. 2, p. 486–494, december 2016. Disponível em: <<https://journal.r-project.org/archive/2016-2/pebesma-mailund-hiebert.pdf>>. Cited in the page 62.
- PELKMANS, L.; DEBAL, P. Comparison of on-road emissions with emissions measured on chassis dynamometer test cycles. *Transportation Research Part D: Transport and Environment*, Elsevier, v. 11, n. 4, p. 233–241, 2006. Cited in the page 133.
- PÉREZ-MARTINEZ, P. et al. Emission factors of air pollutants from vehicles measured inside road tunnels in São Paulo: case study comparison. *Int. J. Environ. Sci. Te.*, Springer, v. 11, n. 8, p. 2155–2168, 2014. Cited in the page 89.
- PULLES, T.; HESLINGA, D. The art of emission inventorying. *TNO, Utrecht*, 2010. Cited 4 times in the pages 29, 53, 54, and 56.
- QGIS Development Team. *QGIS Geographic Information System*. [S.l.], 2017. Disponível em: <<http://qgis.osgeo.org>>. Cited 2 times in the pages 36 and 41.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Cited 4 times in the pages 41, 43, 60, and 95.
- RAFEE, S. A. A. *Estudo numerico do impacto das emissões veiculares e fixas da cidade de Manaus nas concentrações de poluentes atmosféricos da região amazônica*. Tese (Doutorado) — Londrina: Universidade Tecnológica Federal do Parana, 2015. Cited 3 times in the pages 59, 63, and 74.



RAFEE, S. A. A. et al. Mobile and stationary sources of air pollutants in the amazon rainforest: a numerical study with wrf-chem model. *Atmos. Chem. and Phys.*, p. 1–35, 2017. Disponível em: <<http://www.atmos-chem-phys-discuss.net/acp-2016-1190/>>. Cited 6 times in the pages 21, 86, 87, 88, 89, and 91.

ROYSTON, J. An extension of shapiro and wilk's w test for normality to large samples. *Applied Statistics*, JSTOR, p. 115–124, 1982. Cited in the page 41.

SAIDE, P. et al. Spatial disaggregation of traffic emission inventories in large cities using simplified top–down methods. *Atmospheric Environment*, Elsevier, v. 43, n. 32, p. 4914–4923, 2009. Cited in the page 31.

SAIDE, P. E. et al. Forecasting urban pm10 and pm2.5 pollution episodes in very stable nocturnal conditions and complex terrain using wrf–chem co tracer model. *Atmospheric Environment*, v. 45, n. 16, p. 2769 – 2780, 2011. ISSN 1352-2310. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1352231011001178>>. Cited in the page 97.

SANHUEZA, P. A. et al. Particulate air pollution and health effects for cardiovascular and respiratory causes in temuco, chile: a wood-smoke-polluted urban area. *Journal of the Air & Waste Management Association*, Taylor & Francis, v. 59, n. 12, p. 1481–1488, 2009. Cited in the page 30.

SCHULTZ, M. G. Reanalysis of the tropospheric chemical composition over the past 40 years. 2007. Cited in the page 74.

SEINFELD, J. H.; PANDIS, S. N. *Atmospheric chemistry and physics: from air pollution to climate change*. [S.l.]: John Wiley & Sons, 2016. Cited in the page 59.

SELBY, B.; KOCKELMAN, K. M. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. *Journal of Transport Geography*, Elsevier, v. 29, p. 24–32, 2013. Cited in the page 40.

SEXTON, K.; WESTBERG, H. Ambient hydrocarbon and ozone measurements downwind of a large automotive painting plant. *Environmental science & technology*, ACS Publications, v. 14, n. 3, p. 329–332, 1980. Cited in the page 87.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, Biometrika Trust, v. 52, n. 3-4, p. 591–611, 1965. Cited in the page 41.

SIMON, H. et al. The development and uses of epa's speciate database. *Atmospheric Pollution Research*, Elsevier, v. 1, n. 4, p. 196–206, 2010. Cited in the page 87.

SMIT, R.; NTZIACHRISTOS, L.; BOULTER, P. Validation of road vehicle and traffic emission models—a review and meta-analysis. *Atmospheric environment*, Elsevier, v. 44, n. 25, p. 2943–2953, 2010. Cited in the page 54.

SOUZA, C. D. R. de et al. Inventory of conventional air pollutants emissions from road transportation for the state of rio de janeiro. *Energy Policy*, Elsevier, v. 53, p. 125–135, 2013. Cited in the page 134.

STOCKER, T. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. [S.l.]: Cambridge University Press, 2014. Cited in the page 29.

STREETS, D. G. et al. An inventory of gaseous and primary aerosol emissions in asia in the year 2000. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 108, n. D21, 2003. Cited in the page 53.

TEIXEIRA, E. C.; FELTES, S.; SANTANA, E. R. R. d. Estudo das emissões de fontes móveis na região metropolitana de porto alegre, rio grande do sul. *Química Nova*, SciELO Brasil, v. 31, n. 2, p. 244–248, 2008. Cited in the page 134.

TEWARI, M. et al. Implementation and verification of the unified noah land surface model in the wrf model. In: *20th conference on weather analysis and forecasting 16th conference on numerical weather prediction*. [S.l.: s.n.], 2004. v. 1115, p. 1–6. Cited in the page 93.

TIE, X.; BRASSEUR, G.; YING, Z. Impact of model resolution on chemical ozone formation in mexico city: application of the wrf-chem model. *Atmospheric Chemistry and Physics*, Copernicus GmbH, v. 10, n. 18, p. 8983–8995, 2010. Cited in the page 91.

ULKE, A. G.; ANDRADE, M. F. Modeling urban air pollution in sao paulo, brazil: sensitivity of model predicted concentrations to different turbulence parameterizations. *Atmospheric Environment*, Elsevier, v. 35, n. 10, p. 1747–1763, 2001. Cited in the page 31.

USA-EPA. Ap42 compilation of emission factors; resuspension emissions from paved roads. *Environment Protection Agency, USA*, 2016. Cited in the page 63.

VARA-VELA, A. *Avaliação do impacto da mudança dos fatores de emissão veicular na formação de ozônio troposférico na Região Metropolitana de São Paulo (RMSP)*. Tese (Doutorado) — Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, São Paulo, Brazil, 2013. Cited in the page 31.

VARA-VELA, A. et al. 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. and Phys.*, v. 16, p. 777–797, 2016. Cited 8 times in the pages 30, 31, 63, 75, 89, 91, 97, and 136.

VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S*. Fourth. New York: Springer, 2002. ISBN 0-387-95457-0. Disponível em: <<http://www.stats.ox.ac.uk/pub/MASS4>>. Cited 2 times in the pages 43 and 44.

VIVANCO, M. G.; ANDRADE, M. de F. Validation of the emission inventory in the sao paulo metropolitan area of brazil, based on ambient concentrations ratios of co, nmog and nox and on a photochemical model. *Atmospheric Environment*, Elsevier, v. 40, n. 7, p. 1189–1198, 2006. Cited in the page 31.

WANG, H. et al. On-road vehicle emission inventory and its uncertainty analysis for shanghai, china. *Science of the Total Environment*, Elsevier, v. 398, n. 1, p. 60–67, 2008. Cited in the page 89.

WANG, X.; KOCKELMAN, K. Forecasting network data: Spatial interpolation of traffic counts from texas data. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, v. 1, n. 2105, p. 100–108, 2009. Cited 2 times in the pages 40 and 55.

WEILENMANN, M.; FAVEZ, J.-Y.; ALVAREZ, R. Cold-start emissions of modern passenger cars at different low ambient temperatures and their evolution over vehicle legislation categories. *Atmospheric Environment*, v. 43, n. 15, p. 2419 – 2429, 2009. ISSN 1352-2310. Cited in the page 58.

WICKHAM, H. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009. ISBN 978-0-387-98140-6. Disponível em: <<http://had.co.nz/ggplot2/book>>. Cited 2 times in the pages 70 and 71.

WILSON, S. J. et al. Mapping the spatial distribution of global anthropogenic mercury atmospheric emission inventories. *Atmospheric Environment*, Elsevier, v. 40, n. 24, p. 4621–4632, 2006. Cited in the page 31.

ZAVERI, R. A.; PETERS, L. K. A new lumped structure photochemical mechanism for large-scale applications. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 104, n. D23, p. 30387–30415, 1999. Cited 2 times in the pages 86 and 93.

ZEILEIS, A.; KLEIBER, C.; JACKMAN, S. Regression models for count data in r. *Journal of statistical software*, Foundation for Open Access Statistics, v. 27, n. 8, p. 1–25, 2008. Cited 2 times in the pages 40 and 43.

ZHANG, D.; ANTHES, R. A. A high-resolution model of the planetary boundary layer—sensitivity tests and comparisons with sesame-79 data. *Journal of Applied Meteorology*, v. 21, n. 11, p. 1594–1609, 1982. Cited in the page 93.

ZHAO, F.; CHUNG, S. Contributing factors of annual average daily traffic in a florida county: exploration with geographic information system and regression models. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 1769, p. 113–122, 2001. Cited in the page 40.



# APPENDIX A – Evaluation of tests of Wilcoxon for scenarios of simulation

Table 19 – Test Wilcoxon for  $CO$ ,  $NO$  and  $NO_2$

Station	CO			NO			NO <sub>2</sub>		
	Sim Int	Sim Gps	Int Gps	Sim Int	Sim Gps	Int Gps	Sim Int	Sim Gps	Int Gps
Santo Amaro	4.00E-10	1.96E-04	3.66E-04	2.26E-03	<b>5.67E-01</b>	<b>5.70E-01</b>	8.63E-04	<b>4.18E-01</b>	<b>4.26E-01</b>
Pedro II	1.27E-06	4.99E-02	1.82E-02	1.99E-01	5.50E-02	6.21E-02	5.95E-04	5.54E-11	8.69E-11
Congonhas	<b>5.72E-02</b>	2.55E-04	1.60E-04	2.24E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Ibirapuera	7.24E-06	3.50E-11	1.60E-04	1.09E-08	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Sao Caetano do Sul	<b>1.30E-01</b>	2.31E-04	3.71E-11	3.35E-04	1.78E-10	2.12E-10	1.13E-13	2.20E-16	2.20E-16
Cerqueira Cesar	4.52E-07	1.59E-04	1.11E-04	<b>2.32E-01</b>	<b>5.05E-01</b>	<b>5.15E-01</b>	1.13E-13	9.01E-05	1.04E-04
Ipen USP	1.51E-05	<b>5.12E-02</b>	<b>5.45E-02</b>	3.20E-02	<b>5.42E-01</b>	<b>5.48E-01</b>	7.62E-03	5.60E-08	6.30E-08
Parelheiros	1.30E-02	3.62E-08	2.56E-08	9.98E-10	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Tabao da Serra	5.30E-12	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Osasco	<b>8.72E-02</b>	2.21E-08	1.16E-08	6.46E-13	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Carapicuiaba	3.08E-14	2.20E-16	2.20E-16	2.00E-17	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Marginal TPR	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Capao Redondo	2.20E-16	2.20E-16	2.20E-16	9.63E-12	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Capuava	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Diadema	3.03E-16	2.20E-16	2.20E-16	2.75E-14	2.20E-16	2.20E-16	1.67E-14	2.20E-16	2.20E-16
GRUPM	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Itaim Paulista	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Maua	1.35E-15	2.20E-16	2.20E-16	3.64E-05	5.29E-14	5.28E-14	2.20E-16	2.20E-16	2.20E-16
NSO	3.03E-06	1.99E-10	2.00E-10	1.48E-05	2.12E-10	1.22E-09	9.08E-06	6.90E-10	6.68E-10
Santana	2.20E-16	2.20E-16	2.20E-16	7.47E-16	2.20E-16	2.20E-16	4.02E-16	2.20E-16	2.20E-16

Table 20 – Test Wilcoxon for  $SO_2$  and  $O_3$ 

Station	SO2			O3		
	Sim Int	Sim Gps	Int Gps	Sim Int	Sim Gps	Int Gps
Santo Amaro	4.69E-05	<b>1.50E-01</b>	<b>1.39E-01</b>	<b>1.87E-01</b>	<b>1.47E-01</b>	<b>1.56E-01</b>
Pedro II	<b>1.18E-01</b>	<b>1.56E-01</b>	<b>2.27E-01</b>	<b>3.87E-01</b>	1.68E-02	1.94E-02
Congonhas	4.40E-05	5.87E-11	2.09E-10	2.41E-04	2.10E-09	2.86E-09
Ibirapuera	3.01E-12	2.20E-16	2.20E-16	9.52E-14	2.20E-16	2.20E-16
Sao Caetano do Sul	2.85E-03	6.01E-09	3.09E-08	1.22E-05	1.23E-13	1.35E-13
Cerqueira Cesar	1.45E-03	<b>2.13E-01</b>	<b>1.22E-01</b>	1.90E-03	<b>1.97E-01</b>	<b>1.84E-01</b>
Ipen USP	<b>1.80E-01</b>	<b>2.10E-01</b>	<b>2.53E-01</b>	2.02E-01	3.24E-05	3.46E-05
Parelheiros	3.19E-07	2.20E-16	2.23E-16	5.18E-13	2.20E-16	2.20E-16
Tabao da Serra	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Osasco	6.14E-02	6.74E-10	1.37E-09	6.95E-04	5.31E-13	5.22E-13
Carapicuíba	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Marginal TPR	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Capao Redondo	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Capuava	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Diadema	7.72E-14	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
GRUPM	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Itaim Paulista	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Maua	5.18E-12	2.20E-16	2.20E-16	4.51E-15	2.20E-16	2.20E-16
NSO	1.95E-06	9.45E-10	1.07E-09	1.06E-07	5.09E-15	4.74E-15
Santana	1.08E-13	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16