

REMI model: Bottom-up emissions inventories for cities with lack of data

Sergio Ibarra-Espinosa¹ and Rita Ynoue¹

Abstract

Emissions inventorying is a complex task with regulatory and/or scientific environmental purposes. In South American cities, when the task is performed, the common denominator is lack of data and documentation, and vehicles are usually the main source of pollutant of emerging and consolidated megacities. Therefore, emissions inventories is becoming more important, especially for mobile sources. In this manuscript we present the model REMI (**R-EM**issions-Inventory) for developing bottom-up emissions inventory for vehicles in cities with lack of data (Ibarra & Ynoue, 2016). The program was written in R (R CORE TEAM 2016) using several libraries. The program consists in several R scripts organized in folders with Inputs& Outputs. For traffic inputs uses counts or simulations, and also, it can be as a top-down method with statistical traffic information. REMI classifies vehicle data by fuel, size of motor, use and gross weight annually up to 50 years, according to EEA/EMEP guidelines and Copert (Ntziachristos, 2014). REMI has several options for emission factors, 1) Emission factors from Ntziachristos (2014), 2) local emission factors or 3) mixed emission factors. In the future REMI will include HBEFA emission factors. REMI also incorporates deterioration factors. Currently REMI estimate hot-engine emissions of 27 pollutants.

Keywords: REMI, vehicular, emissions inventory, R, bottom-up.

1 Introduction

¹ Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão, 1226 - Cidade Universitária São Paulo-SP - Brasil

“Emission Inventories always seem to be available at a very late point in time ... and are easily seen as the scapegoat if a mismatch is found between modeled and observed concentrations of air pollutants” Tim Pulles (2010). So, the task of emissions inventories compilation ideally should be faster enough in order that it can be updated before deadline. Emissions inventories, scientific or regulatory, must to be comprehensive, with quality and documented but all these requirements faces the common difficulty of lack of data. The uncertainty of estimation depends on the type of sources considered, because there are sources with required information more accesible than other. Per example, usually there are legal obligation that indubustries must report their emissions, like the Pollutant Release and Transfer Registry (PRTR) emissions inventory (<http://prtr.ec.europa.eu/>). But in other cases, sources like mobiles are hard estimate because vehicles are in movement in space. Vehicle emissions inventory can have a top-down approach with activity data as registered fleet, or can be bottom-up based on traffic simulations. Either case, there are inherent uncertainties and the agreement between top-down and bottom-up is hard to accomplish. In order to solve this problem, we developed the REMI model, for elaborating emissions inventories of vehicles, which can be adapted to the information that compiler has, and also is fast.

2 Objectives

To develop a model to estimate vehicular emissions based on road network.

3 Method

Structure

REMI works with a working directory placed in any part of the computer and inside it there are folders “functions” and “data” and these folder divide in several sub-folders as showed in figure 1, where black words are folders and blue ones are files.

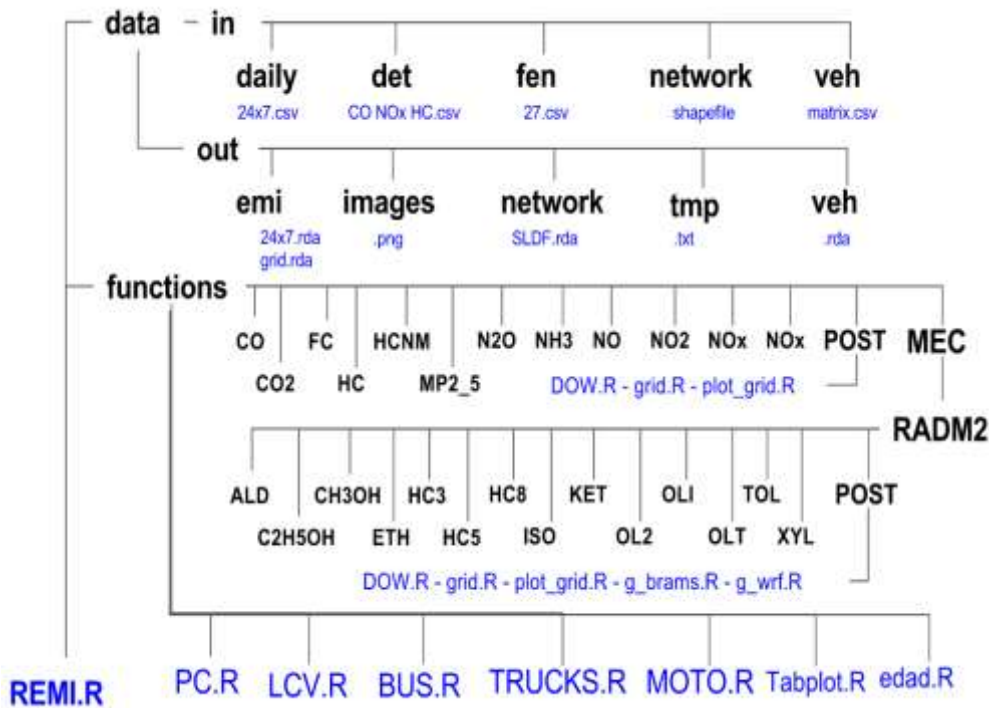


Figure 1: Structure of REMI.

This work consists in present the model REMI for developing bottom-up emissions inventory for vehicles in cities with lack of data (Ibarra & Ynoue, 2016). The program was written in R (R CORE TEAM 2016) using libraries `sp` (Bivand et al, 2013), `rgeos` (Bivand & Rundel, 2015), `maptools` (Bivand & Lewin-Koh, 2015), `ggplot2` (Wickham, 2009) and `ggmap` (Kahle & Wickham, 2013). Activity: Reads traffic simulations, traffic counts and interpolates, or road network and assigns interpolation. Splits categories by fuel, size of motor, use and gross weight annually up to 50 years. Data classification is according to EEA/EMEP guidelines and Copert (Ntziachristos, 2014). The pollutants covered Are CO, NO_x, NO₂, NO, HC, HCNM, PM_{2.5}, N₂O, NH₃, SO₂, CO₂, Fuel Consumption; and speciate HCNM according experimental analysis of tunnel emissions factors in São Paulo for Aldehyde (ALD), C₂H₅OH, CH₃OH, ETH, HC₃, HC₅, HC₈, HCHO, ISO, ketone (KET), OL₂, OLI, OLT, TOL and xylenos (XYL).

The file `REMI.R` works like a namelist where all scripts are called and include options to type of estimation, grid resolution for output, title of images, etc. `REMI.R` has an order, first it calls `NETWORK.R` which produce “data/out/network/red.rda”, then the activity functions `PC.R`, `LCV.R` and so on. Each of these functions produce graphics with `ggplot` saved in “data/out/images/*.png”, tables saved in “data/out/tmp/*.txt ” and the most important, the activity saved in “data/out/veh/na_*.rda”. An advantage of

working with .rda files instead of .csv or shapefiles is that it's possible save different objects with the same extension, and it's light and faster to load in R.

data

Input data

The input data are locate in “data/in” and the folders are daily, fen, det, network and veh.

daily: it contains files with a normalized profile at 08:00-09:00 of the traffic counts 24 hours from monday to sunday. There profiles for Passenger Cars (PC), Light Commercial Vehicles (LCV), Trucks and Buses. It depends on the available data, Ibarra & Ynoue (2016) used traffic counts of a near downtown toll station.

veh: it's a comma-delimited .csv file of 25 vehicles in circulation, that is, vehicles sales after applying survival functions for 50 years. It's a dataframe (R object to store data tables) with header indicating the type of vehicle.

fen: it's a .csv file, a dataframe very similar to veh.csv, but it counts emission factors to the 25 type of vehicles over 50 years of use. There is one file per pollutant.

det: it's a .csv file for the 25 vehicles over 50 years indicating deterioration at year of use. There deterioration factors for CO, HC and NOx.

network: it contains the road network in a spatial format. It can read virtually any spatial format with rgdal library being shapefile one of the most popular, but as it's called on REMI.R, it can be modified as the user prefers. The road network has mandatory fields depending if it's from openstreetmap or a traffic simulation. As traffic simulation varies greatly, here is presented the mandatory fields when performing an emissions inventory with traffic counts and openstreetmap:

- “highway”: field with type of streets: “motorway”, “motorway_link”, etc.
- “xlong1”: first longitude coordinate in EPSG:4326.
- “xlong2”: last longitude coordinates in EPSG:4326
- “ylat1”: first latitude coordinate in EPSG:4326.
- “ylat2”: last latitude coordinate in EPSG:4326.
- “longkm”: Length of the link in km.

Output data

The input data are locate in “data/out” and the folders are emi, images, tmp, network and veh.

emi: it contains all the resulting emissions with 4 types:

1) emissions inventory tables according 5, 20 and 142 types of vehicles, df.rda, ta_cetesb.rda and E_df.rda, respectively. Also the same for RADM2 speciation: df_hcx.rda, ta_cetesb_hcx.rda and E_df_hcx.rda.

- 2) Hourly emissions by each link from Monday to Sunday to each pollutant,
- 3) Gridded emissions with resolution selected by compiler, hourly of a Monday (ex; g_NO2.rda),
- 4) Two text files .txt of the gridded hourly emissions of a Monday. Each file has all the pollutants considering a RADM2 estimation. These files are used one generate WRF-Chem (DF_wrf.txt) (Grell *et al.*, 2005) input data, and the other to BRAMS (DF_brams.txt) input data. These files require that the user indicate the time zone of the area where the emissions were estimated because it stores the emissions in GMT.

images: Here are stored all the images generated by different scripts. There are images of fleet distribution, emissions plot in bars, gridded emissions maps and network maps.

tmp: In this folder there are all the .txt tables generated indicating fleet, accumulated fleet, fleet by fuel standard, and also an emissions table ta_cetesb.csv.

network: It contains the file red.rda of the network which is a SpatialLineDataFrame (SLDF) format, in other words, an sp object. The following fields were added: "ID", "V" the rush hour speed, "Vlibre" free flow speed, "Vmedia" the average speed between "V" and "Vlibre", "dcentro" which is the distance between the mid point of each link and the chosen center of the city, and "dcentroC" which is the same as "dcentro" with a limit in the distance calculation to avoid overestimation of trucks in streets far outside the city.

veh: it includes 5 files of activity level: na_bus.rda, na_lcv.rda, na_m.rda, na_pc.rda and na_trucks.rda. Each file contains the amount of vehicle in each link at 08:00-09:00 of a Monday to each type year of use, from 0 to 50 and by technological composition: na_pc.rda has 348, na_lcv.rda 171, na_trucks.rda 255, na_bus.rda 102 and na_m.rda 171.

functions

Here there are the vehicles function PC.R, LCV.R, TRUCKS.R, BUS.R and MOTO.R. These functions assign vehicles to the road network applying a survival function over 50 years of use and classifies by fuel, size, weight and emissions standard. These functions returns dataframes called na_pc.rda, na_lcv.rda, na_trucks.rdam na_bus.rda and na_m.rda as showed in last section. These functions also create fleet plots and dataframes saved as text files.

Other functions are Tabplot.R and edad.R. edad.R stores a wide range of survival functions resulting in different average years for each type of vehicle (PC, LCV, TRUCKS, BUS and MOTO). The user must choose the respective average age of the fleet. Tabplot.R is to plot emissions bar later after the estimation process ended, it reads df.rda.

There are also pollutant functions, that are the core functions of REMI. Each pollutant has a folder with scripts for the emission factors of each type of vehicle and also the estimations procedure, described in following section.

This folder also contains subfolders POST and MEC. POST include functions to produce SpatialLineDataFrames with 24 emissions for each day of the week, to generate grid with spatial resolution chosen by the user, and functions to plot grids with ggmap library. MEC the chemical mechanism, so far it has RAMD2 functions which speciate the Non-Methane HydroCarbons (NMHC) in fifteen pollutants Aldehyde (ALD), C₂H₅OH, CH₃OH, ETH, HC₃, HC₅, HC₈, HCHO, ISO, ketone (KET), OL₂, OLI, OLT, TOL and xylenos (XYL), according experimental observations performed at Departament of Atmospheric Sciences of IAG/USP (<http://www.iag.usp.br/>). This folder includes a subfolder also called POST with same functions to the respective pollutants, but also includes two special functions to generate text files used to generate WRFR-Chem and BRAMS emissions inputs.

Finally, each folder has a delete.R function used to exclude temporary files.

Emissions estimation process

One all the activity .rda files were generated starts the emissions estimation process. In oder to clarify we present a small part of REMI.R

First, define the working directory and call the libraries.

Then it reads the road network with the respective spatial reference. It's important to note that the original road network could be in any spatial reference, but in order to run in REMI, it must be in EPSG:4326, which is WRG84 LAT LON.

Then are defined the considered latitude and longitude for center in order to calculate its euclidean distance with midpoint of each link (in future it will calculate the geographical (geodesic) distance) .when used the method of interpolation of traffic counts.

```
R> setwd("/your/path")
R> library(maptools); library(rgeos); library(ggmap);
R> red <- readShapeLines("data/in/network/sp.shp", proj4string=CRS("+init=epsg:4326"))
R> xlongCENTRO <- -46.633949
R> ylatCENTRO <- -23.550391
R> system.time(source("functions/NETWORK.R"))
R> rm(list = ls())
```

The activity functions are called following. They assign the vehicles to the road network if it is an interpolation, or read from traffic simulation, and then split it by type of fuel, size of motor and gross weight in a 50 years distribution reading the file data/in/veh/veh.csv. As a result, we have the amount of vehicles at each link with vehicles age distribution and type of fuel, so it's possible to assign

respective emissions standard, resulting in a dataframe names `data/out/veh/na*.rda`. It also produces different fleet plots and tables. The procedure, for Passenger Cars, is just:

```
R > system.time(source("functions/PC.R")) # na_pc.rda
```

The emissions estimations starts loading the emission factors. REMI uses local emissions factors from measurements made by the manufacturers in Brazil with FTP-75 driving cycle (CETESB, 2013), and uses (Ntziachristos, 2014). to incorporate the speed variation at other speeds than 34.12 km/h, the mean speed at this driving cycle. As the emission factors are available to each year of use, it incorporates deterioration factor identifying 50 years. The following lines show the procedure in R:

```
R> det <- read.csv("data/in/det/detco.csv", h=T)
R> fen <- read.csv("data/in/fen/fenco.csv", h=T)
R> F_CO_A_2012_G_L5_1000 <- function(V) {
R> ((1.36E-01 + -8.91E-04*V + 0*V^2)/(1 + -1.41E-02*V + 4.99E-05*V^2))
R> }
R> FFE_1 <- fen[1,1]/F_CO_A_2012_G_L5_1000(34.12)
R> F_CO_A_2012_G_L5_1000 <- function(V) {
R> ((1.36E-01 + -8.91E-04*V + 0*V^2)/(1 + -1.41E-02*V + 4.99E-05*V^2))*FFE_1*det[1,1]
R> }
```

Then the emissions are then calculated loading the activity `na*.rda` file and `red.rda`. REMI considers three mean speeds for rush hour (V), free flow (Vlibre) and an averaged speed (Vmedia) distributed over 24 hours. The emissions are calculated for 50 years of use calculating first to each hour, as follows:

```
R> E_CO_OU_2012_D_P7_seg_8 <- OU_0_D_P7 * F_CO_OU_2012_D_P7 (red$V)*red$longkm
...
R> E_CO_OU_2012_D_P7_seg_8_vl <- OU_0_D_P7 * F_CO_OU_2012_D_P7 (red$Vlibre)*red$longkm
...
R> E_CO_OU_2012_D_P7_seg_8_vm<-OU_0_D_P7*F_CO_OU_2012_D_P7 (red$Vmedia)*red$longkm
...
```

In this case the base-year is 2012, so the emissions are calculated till year 1962 and summed by the emission standard classification of type of vehicle. Then comes the part of extrapolate the emissions to the hours and days of the week. The emissions s are calculated with traffic flow of morning peak hour of 08:00-09:00 and with three average speeds, rush hour, free flow and averaged, and the extrapolation process uses a profile of 24 hours and each day of the week normalized to 08:00-09:00. So, three type of emissions (by each mean speed)

were allocated to 24 hours and then multiplied with this profile at 24 hours and each day of the week. The allocation assumes the following speed distribution:

- 00:00-06:00 Free Flow mean speed (Vlibre).
- 06:00-07:00 Averaged mean speed (Vmedia).
- 07:00-10:00 Rush Morning hour (V).
- 10:00-17:00 Averaged mean speed (Vmedia).
- 17:00-20:00 Rush Evening hour (V).
- 20:00-21:00 Averaged mean speed (Vmedia).
- 21:00-00:00 Free Flow mean speed (Vlibre).

So, the emission are multiplied with the normalized profile at each respective hour. Per example, the 00:00 hour emissions are the emissions calculated with free flow speed and multiplied with the profile of 00:00 hour. The following lines show the process:

```
#Monday
R> E_CO_OU_D_P7_seg_0 <- E_CO_OU_D_P7_seg_8_vl *profile[1,1]
...
R> E_CO_OU_D_P7_seg_6 <- E_CO_OU_D_P7_seg_8_vm *profile[7,1]
...
R> E_CO_OU_D_P7_seg_7 <- E_CO_OU_D_P7_seg_8 *profile[8,1]
...
# Tuesday
R> E_CO_OU_D_P7_ter_0 <- E_CO_OU_D_P7_seg_8_vl *profile[1,2]
...
```

Once the emissions are calculated to all the hours of the week, are summed to tons/year and stored as dataframe E_df.rda. Then, the emissions are summed to each hour and day of the week at each link and also stored as dataframe.

```
R>E_CO_BUS_seg_0<E_CO_OU_D_P7_seg_0 + E_CO_OU_D_P5_seg_0 + ...
```

Benchmark

REMI runs in a machine with GNU/Linux operational system Debian 8.2 Jessie, kernel x_86_64 Linux 3.15.0-4-amd64, CPU of Intel Core i7-4770 3.9GHz and 16 Gb RAM. When calculating the emissions, including gridding and all the plots takes:

Table 1: Performance of REMI

REMI for Metropolitan Area of São Paulo	Time
Count interpolation 105468 links 27 pollutants 24 hours, 7 days of the week	4 hours
Traffic Simulation 34733 links 27 pollutants 24 hours, 7 days of the week	1 hour

4 Discussion

In order to run this model properly it's needed with traffic interpolation it's needed traffic counts, local emissions factors and fleet statistics over 50 years. But this information is not always available when performing an inventory. In REMI model with traffic count interpolation, the amount of vehicles basically depends on road network of open street map (<http://www.openstreetmap.org/>). So, a bigger city will have a bigger road network and a smaller city a smaller network. Keeping that in mind, it's possible to assign the interpolation of one city into another, while those cities shares relatively the same level of congestion per street. As a result, a bigger city will have more vehicles circulating and a smaller one less. It's always important take into account the peculiarities of each city, per example, a city near an important economic activity that requires circulation of heavy duty vehicles, like ports of mining industry. Also, the fleet distribution by type of vehicle, per example, cities with larger amount of motorcycles, like asian cities. Another important aspect is the average age of the fleet of the city, because, despite that the city could have high emissions standards, if the average age of th fleet is old, that city will have more older vehicles circulating, resulting in higher emissions.

With those considerations it was applied REMI model (called REMI in that time) to estimate the vehicular emissions of 58 urban centers of South America (Ibarra *et al*, 2015a). The pollutants covered were CO, PM2.5, HC and NOx. The figure 2 and 3 shows the results by country for CO and by city for PM2.5. In this study were corrected the different type of fuels used in each city, and it was considered the average age of fleet in each country with a statistically significant association with GDPpp for 2012.

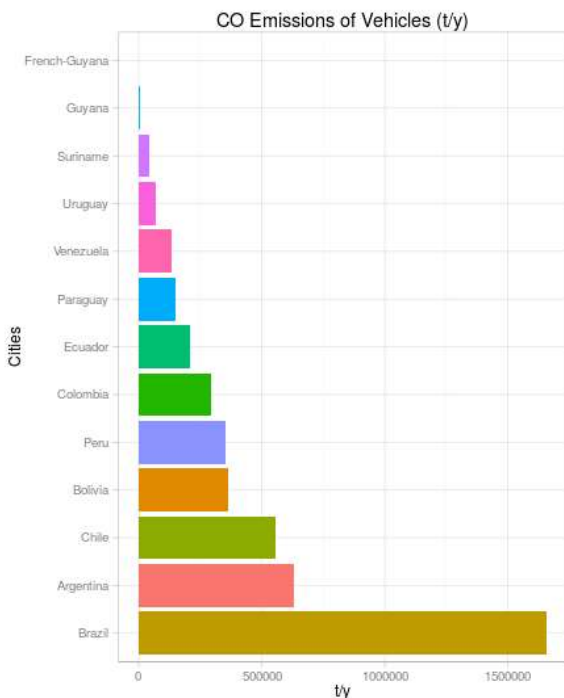


Figure 2: Vehicle emissions of CO by country in 2012 (Ibarra *et al.*, 2015a).

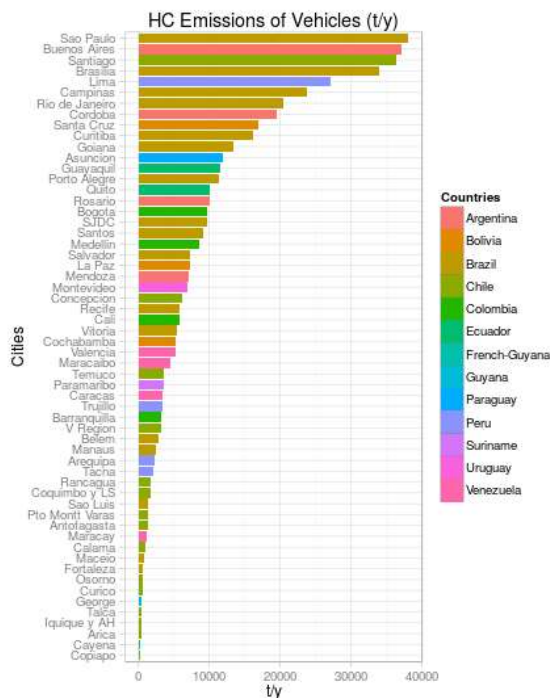


Figure 3: Vehicle emissions of HC by city in 2012 (Ibarra *et al.*, 2015a).

This model was also used to estimate emissions and simulate air pollutant concentrations for the city of Porto Alegre, Brazil (Ibarra *et al.*, 2015b). The air pollutant concentrations could not be compared with observations due to lack of data, but the results were meteorological and chemically consistent. Figure 4 and 5 show results of CO emissions and concentrations.

The model is still in development and a new version it's planned to produce an R package which must be very versatile. This characteristic is for adequate to the reality of different cities. Also, as it's this packages will have different emission factors, allowing projections and estimation of past scenaries. REMI will also include a more type of vehicles, emission factors and it would be applicable to region in the world.

5 Conclusion

REMI model is a program to estimate bottom-up vehicular emissions inventory which, unders certain assumptions, can be adapted to availability of data to any city. Right now it represents Brazilian conditions, but it can be adapted to other realities, including emission factors. SO far, it does not include evaporative and

cold start emissions, but it's planned to include them near in future. This work is part of the PhD in meteorology program of the University of São Paulo, Brazil.

Acknowledgments

We would like to thank María de Fátima Andrade of Laboratório de Processos Atmosféricos (LAPAT) for all the support, as scholarships programs CAPES of Ministry of Education, Brazil, and Becas Chile, from Conicyt, Chile.

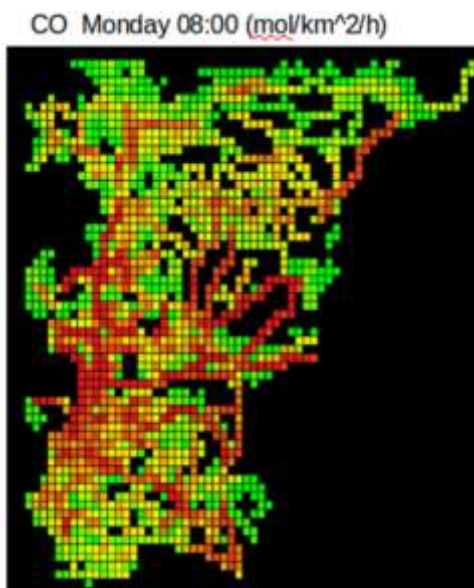


Figure 4: Vehicle emissions of CO Porto Alegre 2012 (Ibarra *et al.*, 2015b).

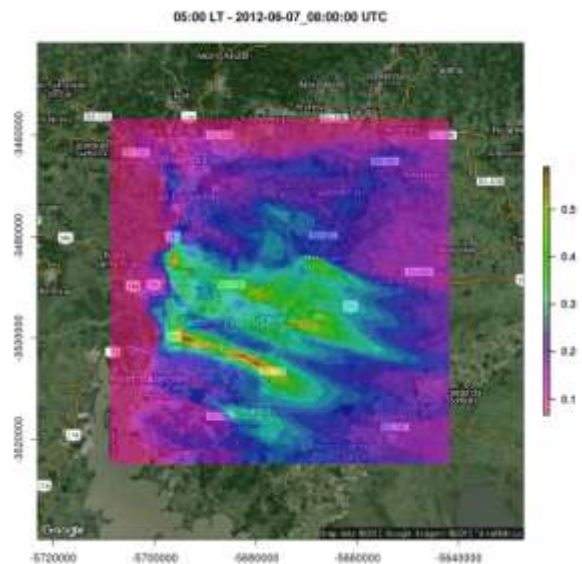


Figure 5: CO concentration (ppm) Porto Alegre 2012 (Ibarra *et al.*, 2015b).

References

- [1] CETESB, 2012. Emissões veiculares no estado de são paulo. <http://veicular.cetesb.sp.gov.br/relatorios-e-publicacoes/>.
- [2] D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144-161. URL <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- [3] Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled “online” chemistry within the WRF model. *Atmospheric Environment*, 39(37), 6957-6975.
- [4] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

- [5] Ibarra S. and Ynoue R. (2016) High resolution vehicular emissions inventory for the megacity of São Paulo. Manuscript to be sent to Atmospheric Environment.
- [6] Ibarra S., Ynoue R., Vela A., Vehicular Bottom-up Emissions Inventory And Atmospheric Simulation For 58 Urban Centers Of South America. 11TH International Conference On Southern Hemisphere Meteorology And Oceanography. October 5TH TO 9TH 2015, Santiago, Chile.
- [7] Ibarra S., Vela A., Rehbein A., Ynoue R., High Resolution Air Pollutant Simulation For The Metropolitan Region Of Porto Alegre. IX Workshop Brasileiro De Micrometeorologia. 11-13 Novembro 2015, Santa Maria, RS, Brasil.
- [8] Ntziachristos, L., and Z. Samaras. "EMEP/EEA emission inventory guidebook." European Environment Agency, Copenhagen" (2009).
- [9] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [10] Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with R, Second edition. Springer, NY. <http://www.asdar-book.org/>
- [11] Roger Bivand and Colin Rundel (2015). rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-15. <https://CRAN.R-project.org/package=rgeos>
- [12] Roger Bivand and Nicholas Lewin-Koh (2015). maptools: Tools for Reading and Handling Spatial Objects. R package version 0.8-37. <https://CRAN.R-project.org/package=maptools>
- [13] Tim Pulles and Dick Heslinga. "The art of emission inventorying." TNO, Utrecht (2010).