Quantifying the presence of air pollutants over a road network in high spatio-temporal resolution

Matteo Böhm
Sapienza University
Rome, Italy
bohm@diag.uniroma1.it

Mirco Nanni
ISTI - CNR
Pisa, Italy
mirco.nanni@isti.cnr.it

Luca Pappalardo
ISTI - CNR
Pisa, Italy
luca.pappalardo@isti.cnr.it

Abstract

Monitoring air pollution plays a key role when trying to reduce its impact on the environment and on human health. Traditionally, two main sources of information about the quantity of pollutants over a city are used: monitoring stations at ground-level (when available), and satellites' remote sensing. In addition to these two, other methods have been developed in the last years that aim at understanding how traffic emissions behave in space and time at a finer scale, taking into account the human mobility patterns. We present a simple and versatile framework for estimating the quantity of four air pollutants (CO_2, NO_x, PM, VOC) emitted by private vehicles moving on a road network, starting from raw GPS traces and information about vehicles' fuel type, and use this framework for analyses on how such pollutants distribute over the road networks of different cities.

1 Introduction

The estimation of the distribution of pollutants over space and time is a major challenge, that concerns both climate change and human health. In some cities, air pollution generated from vehicles' emissions has become more and more evident, to the point that a temporary interruption of the normal traffic flows like during the COVID-19's lockdown immediately resulted in an outstanding increase in air quality [1, 2, 3, 4]. Even if some noticed that this relatively brief period will basically do nothing to slow down climate change [5], it helped in outlining how evident the impact of the emissions related to the transportation sector is in everyday life. At national level, the share of total GHG (Greenhouse Gases) emissions coming from transportation arrives at 30% in high income economies, and 72% of the direct GHG emissions caused by the transport sector worldwide in 2010 was from road travel [6]. Moreover, the transport sector also emits non-CO_2 pollutants such as nitrogen oxides (NO_x), that lead to the formation of ozone and particulate matter (PM), and volatile organic compounds (VOCs); these pollutants, emitted by internal combustion engines, also play a fundamental role in changing climate and are dangerous for human health [6]. For these reasons, understanding the way pollutants emitted by vehicles spread over a city is extremely important in designing strategies to reduce the share of emissions coming from transportation. As suggested in the literature [7, 8], big data and machine learning offer the tools to implement crucial strategies such as reducing transport activity, improving vehicle efficiency, alternative fuels and electrification, and shifting to lower-carbon options. It is also clear that they are strictly related to the way vehicles move through a road network, thus it

Tackling Climate Change with Machine Learning workshop at NeurIPS 2020.

becomes fundamental to conceive tools that can help in studying how the resulting emissions patterns evolve in space and time.

Emissions from vehicles have traditionally been studied with the use of measured (either coming from sensors [9], from official sources like Ministries of Transportation [10], resulting from household travel surveys [11], or from experiments using driving simulators [12]) or simulated [13, 14] traffic data. A considerable number of works concentrated on estimating people's exposure to air pollution using a wide range of models (see [15] for a review), often integrating them with mobility data coming from mobile phones in order to reach a dynamic assessment of the exposure [16, 17, 18, 19, 20]. This dynamic approach has also been adopted in some cases in estimating vehicles' emissions. In [21, 22] emissions are directly measured with a Portable Emissions Measurement System (PEMS) including a GPS to analyse gaseous emissions of a few vehicles driving in real-world conditions, while both Nyhan et al. [23] and Liu et al. [24] estimate vehicles' emissions with a microscopic emissions model using taxis' GPS trajectory data coming respectively from Singapore and Hangzhou. Chen et al. [25] use GPS trajectories coming from a navigation app and focus the analysis on braking emissions of particulate matter in Tokyo. Yu et al. [26] use mobile phone GPS data to determine the position of customised bus stops and to estimate the emission reduction potential of the resulting bus lines. Sui et al. [27] find evidence for understanding the advantage of online ride-hailing against traditional taxis w.r.t. fuel consumption and emissions, using GPS data provided by taxis and a ride-hailing company moving in Chengdu (China). The use of data coming from positioning systems directly installed on the vehicles allows to include in the models factors coming from real-world driving conditions, like acceleration and speed, that have been proved to play a fundamental role especially when estimating emissions of NO_x [22], thus reaching more accurate results. However, these experiments often are conducted on a small sample of vehicles, or on biased ones (e.g., taxi fleet, high-duty vehicles). Moreover, they also seem to suffer from a geographical bias, as almost all of them are carried out in Asian cities [19, 17, 23, 24, 21, 25, 26, 27] (mostly in China and Japan).

The aim of this study is to present a versatile framework to estimate the quantity of four diverse pollutants emitted by private vehicles moving on a road network over time, starting from raw GPS traces and information about vehicles' fuel type. Our method can make use of a considerable number of trajectories coming from real-world vehicles, and it is able to provide a precise idea of the way air pollutants spread over space and time with a very fine resolution. Indeed, it makes it possible to visualise how much each road of a network is polluted during a certain period, to study the change in pollution eventually caused by changes in vehicular mobility over time (e.g., during a lockdown, or caused by a new infrastructure built in a certain area), to compare vehicles' behaviour in terms of emissions across space and time, to analyse the distribution of the quantity of emissions per vehicle and/or per road across different cities and countries. In a few words, it represents a useful tool for decision-makers in implementing those strategies aiming at reducing emissions coming from road travel.

2 Methods

This section briefly describes the proposed framework for estimating the quantity of CO_2, NO_x, PM and VOC emitted by a fleet of vehicles moving in a road network starting from their GPS trajectories and information about their fuel type. In Fig.1 the framework's pipeline is shown: firstly, instantaneous speed and acceleration of the vehicles are computed in each point of their GPS trajectories, and some filtering technique is performed on them. Then, the points are mapped to the edges (i.e. roads) of the network. Finally, a microscopic emissions model is used to estimate the vehicles' emissions in each point. Some details of these steps are briefly described in the next paragraphs.

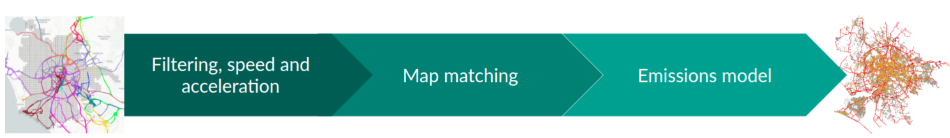


Figure 1: The steps from raw GPS trajectories of vehicles (a sample of vehicles moving in Rome is shown on the left) to their emissions over the road network (right).

Filtering, speed and acceleration The trajectories are stored in a `TrajectoryDataFrame`, a data structure from `scikit-mobility`[1] [28]. Each point p is associated with the vehicle id (`uid`) and the trajectory id (`tid`) to which it belongs, with a `timestamp` t, and with its latitude/longitude position at that time. Note that, in such a way, each point p is uniquely identified by `(uid, tid, timestamp, lat, lon)`. A first filtering retains only the sub-trajectories composed by points that are distant not more than a certain time interval from each other; this is intended to make the following computation of the instantaneous speed and acceleration of the vehicle in each point as reliable as possible. Then, once computed, a filtering on these values is performed: only points in which the vehicle's speed is not greater than 300 km/h and its acceleration does not exceed 10 m/s^2 are retained (following [23]). It is important to outline that different works use very different values for these filtering parameters (e.g. the time interval between the points is set to 5 seconds in [23], 180 seconds in [24], 1 second in [25], 3 seconds in [27]), so these choices are not fixed here and they can easily be changed.

Map matching The road network is took from OpenStreetMap. Each edge (namely, an OSM element `way`) is represented as the couple of OSM `node` ids of its starting and ending nodes. Each point p of the trajectory dataset is mapped to its nearest edge in the network with a ball tree nearest-neighbour algorithm for fast haversine search.

Computing emissions Following Nyhan et al. [23], a microscopic emissions model is implemented to compute the instantaneous emissions associated to each point p. Some notation is introduced. The quantity of pollutant j, $j \in \{\mathrm{CO_2}, \mathrm{NO}_x, \mathrm{PM}, \mathrm{VOC}\}$, emitted at point p is noted as E_j^p. The instantaneous speed and acceleration of the vehicle in p are respectively noted as v_p and a_p. For each of the considered vehicles, the information about its engine type (whether they are petrol, diesel or LPG vehicles) is gathered, as it defines, together with the type of pollutant and the type of vehicle, the emission factors f. This data about the vehicle is, just for the sake of notation, resumed by the letter u. Then, the following equation is used for computing the instantaneous emissions of pollutant j from vehicle u in point p:

$$E_{p,j} = f_1^{j,u} + f_2^{j,u} v_p + f_3^{j,u} v_p^2 + f_4^{j,u} a_p + f_5^{j,u} a_p^2 + f_6^{j,u} v_p a_p \tag{1}$$

where for NO_x and VOC emissions the factors $f_1, ..., f_6$ change with acceleration (based on whether $a_p \geq -0.5\ m/s^2$ or $a_p < -0.5\ m/s^2$).

3 Case study

This section aims at exploring some possible analyses that can be done using the proposed framework, as well as how it can be improved and extended. The data used here consists in GPS traces from private vehicles moving in London, Rome and Florence throughout January 2017.

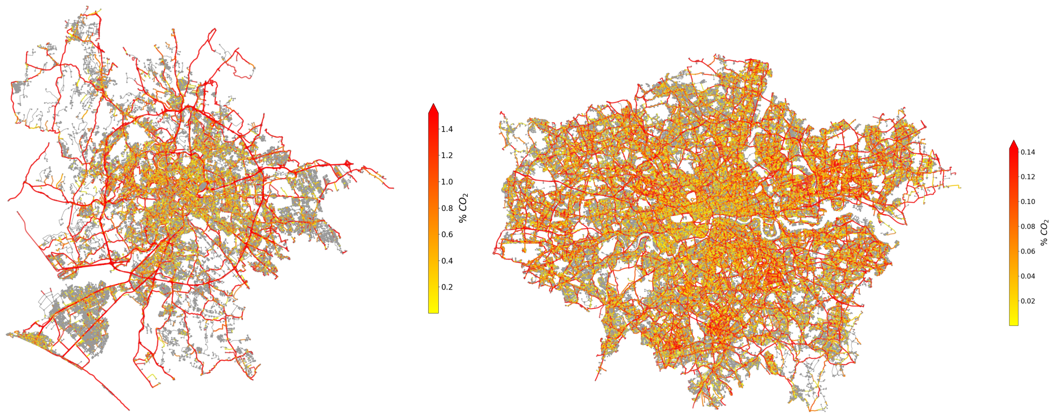
3.1 The distribution of air pollution

With the proposed framework it is very easy to visualise the total quantity of pollutant emitted during a certain period across a road network. As an example, Fig. 2 shows how much the intrinsic differences of the road networks of two big cities like London and Rome can influence how the emissions of $\mathrm{CO_2}$ spread. Indeed, our analyses show that the road network of London has about 3 times the number of streets of that of Rome, and these streets have an average length of 77 meters (about 25 meters less than those of Rome): London is about two times denser than Rome both in terms of street density (i.e. the sum of all the streets' length divided by the area covered by the network) and intersection density (i.e. the number of intersections divided by the same area). Obviously, the types of roads also play a role: the road network of Rome includes a 0.4% of roads tagged as motorways[2] (which usually are quite long and large roads), while in London they represent only the 0.02% of the total. Consequently, air pollutants from vehicular traffic seem to distribute between the roads in a more equal manner in London than in Rome.

The study of the distribution of air pollution per vehicle reveals a well-known result about its characteristics; for all the three cities, it follows a power-law (at least for $\mathrm{CO_2}$, NO_x and PM): a few

[1] https://github.com/scikit-mobility/scikit-mobility

[2] see OpenStreetMap Wiki for more information on road tagging and types.



(a) Rome. (b) London.

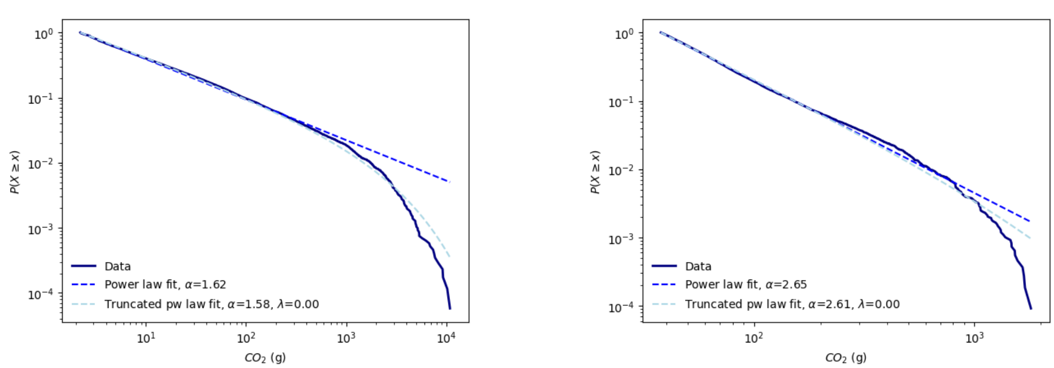
Figure 2: The road networks of Rome and London: percentage of CO_2 emitted in each road in January 2017. There are ∼6.7K vehicles moving in Rome and ∼2.5K in London.

vehicles are responsible for a great quantity of emissions [29, 30]. Within the vehicles moving in Rome, Florence and London, for example, 10% of them are responsible for respectively 50%, 52% and 40% of the total CO_2 emitted in January 2017. A novel result, as far as we know, from this study is represented by the finding that also the distribution of air pollution per road follows the same rule: a few roads have the greatest share of emissions in the network. Both in Rome and Florence, only 10% of the roads are associated to more than 90% of the CO_2 emitted during the period; in London this quantity is lower (57%), but still definitely above the half of the total emissions of CO_2.

In order to support our finding with a statistical validation, a likelihood-ratio test has been used to compare various models. The goodness of fit of a power-law has been firstly compared with that of an exponential, that is the minimum alternative candidate for evaluating the heavy-tailedness of the distribution, giving results in favour of the first one. In Fig. 3, the power-law is compared with a truncated power-law, for the two cases of Rome and London. The results suggest that, despite the outlined differences in the two road networks, in both the cases the distribution of the quantity of CO_2 (but also NO_x and PM) emitted per road are well approximated by a truncated power-law.

4 Conclusions and future work

The proposed framework is quite simple and flexible as it only relies on GPS traces and information about fuel types and its parameters can be easily fixed w.r.t. data set and scopes. It has been used to



(a) Rome. (b) London.

Figure 3: The complementary CDF of the data on the quantity of CO_2 emitted per road, with its best power-law and truncated power-law fits.

explore the laws of vehicular air pollution and its spreading over the road network of three different cities, as well as the relation that this spreading has with the intrinsic characteristics of the network. In a nutshell, this versatile tool permits to study two main factors driving vehicular emissions: how people move within the city (the mobility behaviours) and how the infrastructures of the city (the road network) are conceived to let people move. Thus, the main question that this framework aims to answer is which of the two, from city to city, can have the major impact in reducing emissions.

Acknowledgments and Disclosure of Funding

Mirco Nanni has been partially supported by the European Community H2020 programme under the funding scheme Track & Know (Big Data for Mobility Tracking Knowledge Extraction in Urban Areas), G.A. 780754, https://trackandknowproject.eu/. Matteo Böhm has been partially supported by EU project SoBigData++, G.A. 871042.

References

[1] Fred Dutheil, Julien Baker, and Valentin Navel. "COVID-19 as a factor influencing air pollution?" In: *Environmental Pollution* 263 (Apr. 2020), p. 114466.

[2] Jesse D. Berman and Keita Ebisu. "Changes in U.S. air pollution during the COVID-19 pandemic". In: *Science of The Total Environment* 739 (2020), p. 139864.

[3] Zander S. Venter, Kristin Aunan, Sourangsu Chowdhury, and Jos Lelieveld. "COVID-19 lockdowns cause global air pollution declines". In: *Proceedings of the National Academy of Sciences* 117.32 (2020), pp. 18984–18990.

[4] Corinne Le Quéré et al. "Temporary reduction in daily global CO2 emissions during the COVID-19 forced confinement". In: *Nature Climate Change* 10 (July 2020), pp. 1–7.

[5] P.M. Forster et al. "Current and future global climate impacts resulting from COVID-19". In: *Nature Climate Change* (2020).

[6] Sims R. et al. "Transport". In: *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report*. Cambridge University Press, 2014, pp. 599–670.

[7] David Rolnick et al. "Tackling Climate Change with Machine Learning". In: *CoRR* abs/1906.05433 (2019).

[8] Vasiliki Voukelatou, Lorenzo Gabrielli, Ioanna Miliou, Stefano Cresci, Rajesh Sharma, Maurizio Tesconi, and Luca Pappalardo. "Measuring objective and subjective well-being: dimensions and data sources". In: *International Journal of Data Science and Analytics* (June 2020).

[9] Priyanka deSouza, Amin Anjomshoaa, Fabio Duarte, Ralph Kahn, Prashant Kumar, and Carlo Ratti. "Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned". In: *Sustainable Cities and Society* 60 (2020), p. 102239.

[10] Ariel Reznik, Meidad Kissinger, and Nurit Alfasi. "Real-data-based high-resolution GHG emissions accounting of urban residents private transportation". In: *International Journal of Sustainable Transportation* 13.4 (2019), pp. 235–244.

[11] Md. Nobinur Rahman and Ahmed O. Idris. "TRIBUTE: Trip-based urban transportation emissions model for municipalities". In: *International Journal of Sustainable Transportation* 11.7 (2017), pp. 540–552.

[12] Sicong Zhu, Inhi Kim, and Keechoo Choi. "High-resolution simulation-based analysis of leading vehicle acceleration profiles at signalized intersections for emission modeling". In: *International Journal of Sustainable Transportation* 0.0 (2020), pp. 1–11.

[13] H. M. Abdul Aziz and Satish V. Ukkusuri. "A novel approach to estimate emissions from large transportation networks: Hierarchical clustering-based link-driving-schedules for EPA-MOVES using dynamic time warping measures". In: *International Journal of Sustainable Transportation* 12.3 (2018), pp. 192–204.

[14] Jaehyun (Jason) So, Nassim Motamedidehkordi, Yingjie Wu, Fritz Busch, and Keechoo Choi. "Estimating emissions based on the integration of microscopic traffic simulation and vehicle dynamics model". In: *International Journal of Sustainable Transportation* 12.4 (2018), pp. 286–298.

[15] Xingzhe Xie, Ivana Semanjski, Sidharta Gautama, Evaggelia Tsiligianni, Nikos Deligiannis, Raj Rajan, Frank Pasveer, and Wilfried Philips. "A Review of Urban Air Pollution Monitoring and Exposure Assessment Methods". In: *ISPRS International Journal of Geo-Information* 6.12 (2017), p. 389.

[16] Marguerite Nyhan, Itai Kloog, R. Britter, C. Ratti, and Petros Koutrakis. "Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data". In: *Journal of Exposure Science & Environmental Epidemiology* 29 (2018).

[17] Haofei Yu, Armistead Russell, James Mulholland, and Zhijiong Huang. "Using cell phone location to assess misclassification errors in air pollution exposure estimation". In: *Environmental Pollution* 233 (2018), pp. 261–266.

[18] B. Dewulf, T. Neutens, W. Lefebvre, Gerdy Seynaeve, C. Vanpoucke, C. Beckx, and N. van de Weghe. "Dynamic assessment of exposure to air pollution using mobile phone data". In: *International Journal of Health Geographics* 15 (2016).

[19] Mingxiao Li, Song Gao, Feng Lu, Huan Tong, and Hengcai Zhang. "Dynamic Estimation of Individual Exposure Levels to Air Pollution Using Trajectories Reconstructed from Mobile Phone Data". In: *International Journal of Environmental Research and Public Health* 16.22 (2019), p. 4522.

[20] Miguel Picornell, Tomás Ruiz, Rafael Borge, Pedro García-Albertos, David Paz, and Julio Lumbreras. "Population dynamics based on mobile phone data to improve air pollution exposure assessments". In: *Journal of Exposure Science & Environmental Epidemiology* 29 (2018).

[21] Hwan S. Chong, Sangil Kwon, Yunsung Lim, and Jongtae Lee. "Real-world fuel consumption, gaseous pollutants, and CO2 emission of light-duty diesel vehicles". In: *Sustainable Cities and Society* 53 (2020), p. 101925.

[22] José M. Luján, Vicente Bermúdez, Vicente Dolz, and Javier Monsalve-Serrano. "An assessment of the real-world driving gaseous emissions from a Euro 6 light-duty diesel vehicle using a portable emissions measurement system (PEMS)". In: *Atmospheric Environment* 174 (2018), pp. 112–121.

[23] Marguerite Nyhan et al. "Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model". In: *Atmospheric Environment* 140 (2016), pp. 352–363.

[24] Jielun Liu, Ke Han, Xiqun (Michael) Chen, and Ghim Ping Ong. "Spatial-temporal inference of urban traffic emissions based on taxi trajectories and multi-source urban data". In: *Transportation Research Part C: Emerging Technologies* 106 (2019), pp. 145–165.

[25] Jinyu Chen, Wenjing Li, Haoran Zhang, Wenxiao Jiang, Weifeng Li, Yi Sui, Xuan Song, and Ryosuke Shibasaki. "Mining urban sustainable performance: GPS data-based spatio-temporal analysis on on-road braking emission". In: *Journal of Cleaner Production* 270 (2020), p. 122489.

[26] Qing Yu, Haoran Zhang, Weifeng Li, Xuan Song, Dongyuan Yang, and Ryosuke Shibasaki. "Mobile phone GPS data in urban customized bus: Dynamic line design and emission reduction potentials analysis". In: *Journal of Cleaner Production* 272 (2020), p. 122471.

[27] Yi Sui et al. "GPS data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions". In: *Journal of Cleaner Production* 227 (2019), pp. 495–505.

[28] Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. *scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data*. 2019.

[29] P.L. Guenther, G.A. Bishop, J.E. Peterson, and D.H. Stedman. "Emissions from 200 000 vehicles: a remote sensing study". In: *Science of The Total Environment* 146-147 (1994). Highway Pollution, pp. 297–302.

[30] Yuhan Huang, Bruce Organ, John L. Zhou, Nic C. Surawski, Guang Hong, Edward F.C. Chan, and Yat Shing Yam. "Remote sensing of on-road vehicle emissions: Mechanism, applications and a case study from Hong Kong". In: *Atmospheric Environment* 182 (2018), pp. 58–74.