IMPROVING VEHICLES' EMISSIONS REDUCTION POLICIES BY TARGETING GROSS POLLUTERS

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

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

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

ABSTRACT

Vehicles' emissions produce a significant share of cities' air pollution, with a substantial impact on the environment and human health. Traditional emission estimation methods use remote sensing stations, missing vehicles' full driving cycle, or focus on a few vehicles. This study uses GPS traces and a microscopic model to analyse the emissions of four air pollutants from thousands of vehicles in three European cities. We discover the existence of gross polluters, vehicles responsible for the greatest quantity of emissions, and grossly polluted roads, which suffer the greatest amount of emissions. Our simulations show that emissions reduction policies targeting gross polluters are way more effective than those limiting circulation based on a non-informed choice of vehicles. Our study applies to any city and may contribute to shaping the discussion on how to measure emissions with digital data.

***K*eywords** data science · human mobility · transportation · GPS data · climate change · GHG emissions · statistical analysis

Introduction

The estimation of air pollutants' distribution over space is a significant challenge that concerns climate change and human health. In urban environments, air pollution generated from vehicles' emissions has become more and more evident, to the point that a temporary interruption of regular traffic flows during the COVID-19's lockdown immediately resulted in a tremendous increase in air quality and decrease of CO_2 emission [13, 4, 62, 32, 23, 36]. Even if this brief period's impact on the epochal challenge of climate change is negligible [16], it helps outline the impact of the emissions related to the transportation sector on our everyday life. In particular, greenhouse gas (GHG) emissions from this sector have more than doubled since 1970, and about 80% of this growth has come from road transport [26]. In 2016, the 11.9% of global GHG emissions were from road transport (60% of which from passenger travel and 40% from road freight) [52]. Moreover, the transport sector emits non-CO_2 pollutants such as nitrogen oxides (NO_x), which lead to the formation of ozone and particulate matter (PM), and volatile organic compounds (VOCs). These pollutants, emitted by internal combustion engines, play a fundamental role in changing climate and are severely dangerous for human health [26]. Among the Sustainable Development Goals to be reached by 2030 [31, 30, 59], the United Nations poses an urgent call for action to reduce "the adverse per capita environmental impact of cities, paying special attention to air quality" [59]. In this regard, understanding how air pollutants emitted by vehicles distribute in a city is fundamental in designing policies to reduce transportation emissions.

GPS data generated by in-vehicle devices describe human mobility in great detail [44, 46, 18, 19, 37] and offer an unprecedented tool to implement strategies such as reducing transport activity and congestions [10, 39, 56, 6], improving vehicle efficiency, encouraging alternative fuels and electrification, and shifting to lower-carbon options [27, 34, 53, 63]. Emissions from vehicles are traditionally studied with the use of two types of data: *(i)* measured traffic data, either coming from sensors [11], official sources [50], or household travel surveys [49]; and *(ii)* simulated traffic data, such as those generated with driving simulators [69], or traffic simulation models [2, 54]. Another strand of research concentrates on estimating people's exposure to air pollution [65], often integrating them with mobile phone records to reach a dynamic assessment of the exposure [41, 67, 12, 33, 48, 66]. Some works adopt this approach to estimate vehicles' emissions. Chong et al.[8] and Luján et al.[38] measure emissions with a Portable Emissions Measurement System (PEMS) to analyze gaseous emissions of a few vehicles driving in real-world conditions. Nyhan et al. [42] and Liu et al. [35] estimate vehicles' emissions with a microscopic emission model using GPS data coming from taxis in Singapore and Hangzhou, respectively. Chen et al. [7] use GPS trajectories from a navigation app to analyse braking emissions of particulate matter in Tokyo. Gately et al. [20] use an emission model that includes vehicle speeds, traffic volumes, fleet information, and weather conditions to assess the effects of traffic congestions on emissions at the road-level, finding critical hotspots where congestions significantly enhance emissions. Yu et al. [68] use GPS data from smartphones to determine the position of customized bus stops and estimate the resulting bus lines' emission reduction potential. Sui et al. [57] find evidence for understanding the advantage of online ride-hailing against traditional taxis concerning fuel consumption and emissions, using taxis' GPS data and a ride-hailing company moving in Chengdu (China). All these works show that data coming from onboard positioning systems allow capturing real-world driving conditions such as acceleration and speed, which play a crucial role in achieving accurate emission estimates (especially for NO_x [38]). Unfortunately, these experiments rely on a small sample of vehicles or specific vehicle types (e.g., taxi fleet, high-duty vehicles), and they are carried out mostly for Asian cities [33, 67, 42, 35, 8, 7, 68, 57]. It is still unclear what statistical patterns characterize the distribution of emissions per vehicle and road, how the road network structure may influence those patterns, and how we can exploit this information to simulate emission reduction scenarios.

Our study analyses the estimated emissions of four air pollutants (carbon dioxide, nitrogen oxides, particulate matter, and volatile organic compounds) from thousands of private vehicles moving in different European cities. We use trajectories generated by onboard GPS devices, match them to the cities' road networks, and exploit information about the vehicles' fuel type to implement a microscopic model that computes the emissions produced by the vehicles. We then study how the emissions distribute across vehicles and roads to discover the statistical patterns that characterise emissions and investigate the relationships between emissions, the vehicles' mobility, and the road network's characteristics. Finally, we simulate two emission reduction scenarios in which a share of vehicles become zero-emissions (as electric or targeted by driving restrictions) or reduce their recurrent mobility (as the drivers being in remote working or given pandemic-driven non-pharmaceutical interventions). These simulations allow us to find strategies to drastically reduce emissions over a city while minimising the share of vehicles targeted.

Our framework applies in principle to any city and provides an efficient way to estimate emissions at the road level. It represents practical support for decision-makers to implement strategies to reduce emissions, improve air quality, citizens' well-being, and design smarter and more sustainable cities [3, 28, 63].

	GPS trajectories			
	vehicles	trips	points	avg sampling rate (std)
London	2,778	244,436	2,415,854	77.4 sec (29.0)
Rome	8,167	161,719	764,082	79.7 sec (21.7)
Florence	3,962	27,117	62,545	74.3 sec (28.5)

Table 1: **Summary statistics of the GPS data.** The number of vehicles, trips, and points, and their mean sampling rate (and standard deviation), for Greater London, Rome, and Florence.

	road networks				
	roads	crossroads	roads density	crossroads/km^2	avg road length (std)
London	256,391	145,441	12,152	92.5	74.5 m (85.7)
Rome	84,231	49,213	6,664	38.3	101.5 m (155.8)
Florence	12,366	7,357	10,498	71.8	86.9 m (100.4)

Table 2: **Summary statistics of the road networks.** The number of roads and crossroads, their densities, and the mean road length (and standard deviation) of the road networks of Greater London, Rome, and Florence. The roads density is the total length of all the roads divided by the land area of the city. The crossroads density (crossroads/km^2) is the number of crossroads divided by the land area of the city.

Results

Computation of emissions

Our data consist of anonymous GPS trajectories describing 433,272 trips from 14,907 private vehicles moving during January 2017 in Greater London, Rome, and Florence throughout January 2017 (see Table 1). The trajectories are produced by on-board GPS devices, which automatically turn on when the vehicle starts, transmitting a point every minute to the server via a GPRS connection [44, 46, 18, 19]. When the vehicle stops, no points are logged nor sent. The GPS traces are collected by a company that provides a data collection service for insurance companies. The market penetration of this service is variable, but in general covers at least 2% of the total registered vehicles, and it is representative of the overall amount of vehicles circulating in a city [44]. Figure 1a shows a sample of trajectories for 20 vehicles in Rome.

As a first step, we filter GPS trajectories so that the time between consecutive points is lower than a certain time threshold (see Methods and Supplementary Information A.1). For each vehicle, we estimate instantaneous speed and acceleration in each point of its trajectory and filter out points with unrealistic values (see Methods). To match the trajectories to each city's road, we download road networks from OpenStreetMap (see Methods).

London has about three times the number of roads than Rome and a smaller average road length (see Table 2). Florence is the smallest city in terms of land area, the number of roads, and crossroads. However, in terms of the number of crossroads per km^2, Florence (~ 72) is almost double as Rome (~ 38) and not too far from London (~ 92). Also, in terms of road density (total length of all the roads divided by the land area of the city), Florence (10,498) is 1.6 times denser than Rome (6,664) and not far from London's density (12,152). In Rome, 0.4% of roads are motorways (usually long and large roads without traffic signals), while both in London and Florence represent only 0.02%. In summary, we analyse three cities that are heterogeneous with respect to their road networks: Rome is huge but with the sparsest network; London is huge but with the densest network; Florence is small ($\sim 1/12$ of Rome and $\sim\ 1/15$ of London in terms of land area) but with a dense road network.

We map the points to the city's road network using a map-matching algorithm (see Methods, Figure 1b). We use a microscopic emissions model [42] that uses speed, acceleration, and fuel type to estimate the vehicles' instantaneous emissions of carbon dioxide (CO_2), nitrogen oxides (NO_x), particulate matter (PM), and volatile organic compounds (VOC) (see Methods, Figure 1c). Finally, we compute each vehicle's overall emissions as the sum of all its instantaneous emissions during the study period. Analogously, we compute the overall amount of air pollutants on each road by summing up all the instantaneous emissions from any vehicle passing through that road during the same period.

Patterns of emissions

We find that, for all three cities, emissions distribute across vehicles in a heterogeneous way: a few vehicles, that we call *gross polluters* [22, 24], are responsible for a tremendous amount of emissions. At the same time, most of them emit significantly less (Figure 2 and Supplementary Information A.3). For example, the top 10% gross polluters in Florence, Rome, and London are responsible for 47.5%, 50.5%, and 38.5% of the total CO_2 emitted during the month, respectively. We find that the distributions

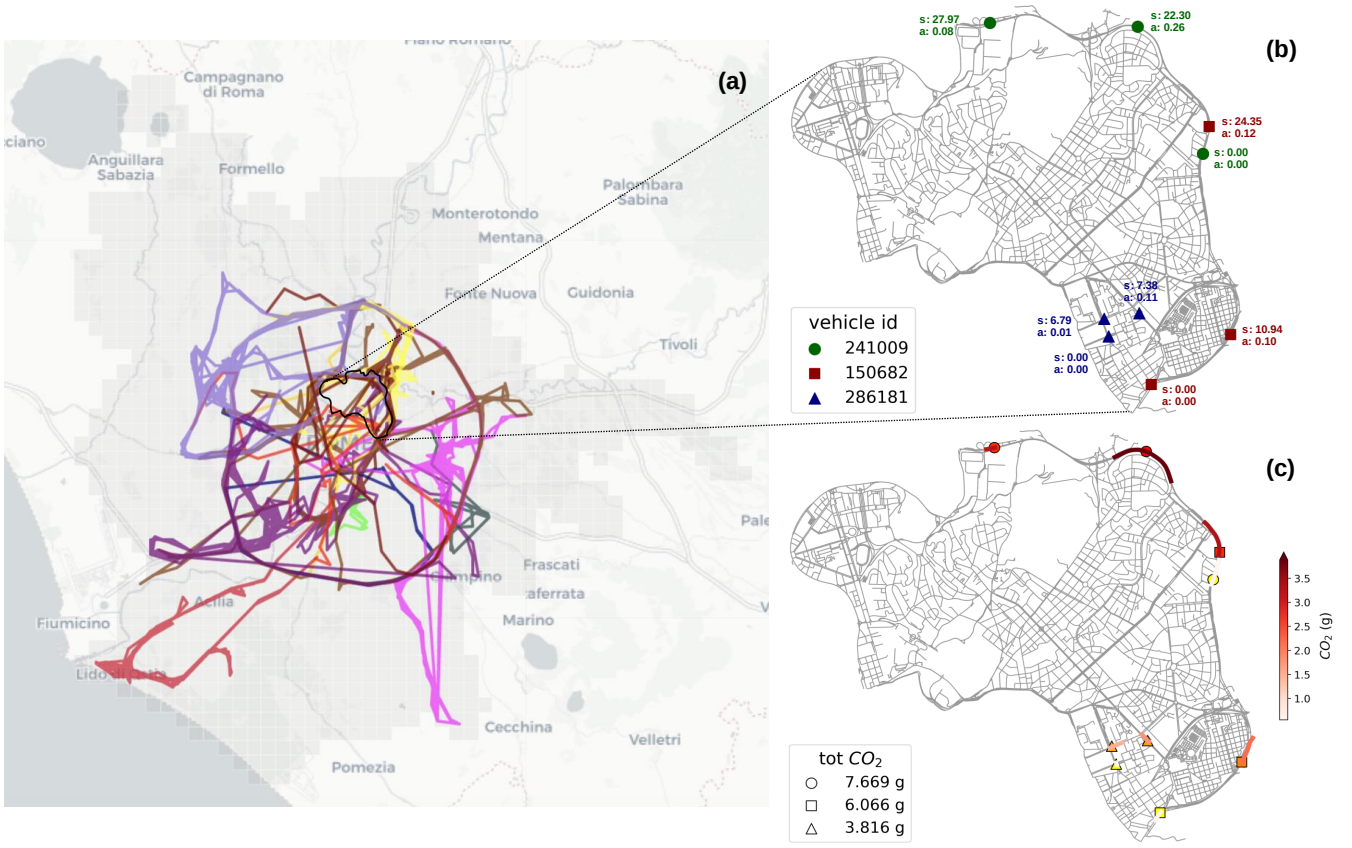

Figure 1: **Computation of emissions from GPS trajectories.** (a) Visualisation of the trajectories of 20 vehicles travelling during January 2017 in Rome. Each colour indicates a different trajectory. The grey area indicates the territory of the municipality of Rome. Plot generated with Python library scikit-mobility [45]. (b) Visualisation of the GPS points of three vehicles passing through a neighbourhood in Rome. Each symbol and colour indicates a different vehicle; for each point we show the corresponding instantaneous speed and acceleration. The point with zero speed and acceleration is the first point of the trajectory sample. (c) The instantaneous emissions of CO_2 of each GPS point and for each road crossed. The points and the roads are coloured in a gradient from white (low emission) to red (high emission). The legend shows the overall quantity of emissions of each vehicle.

of CO_2 emissions per vehicle of Rome and Florence are well approximated by a truncated power law $p(x) \propto x^{-\alpha}e^{-\lambda x}$, with $\alpha = 1.13$, $\lambda = 1.04 \times 10^{-3}$ (Rome, Figure 2e), and $\alpha = 2.12$, $\lambda = 1.45 \times 10^{-3}$ (Florence, Figure 2h). In contrast, London's distribution is well approximated by a stretched exponential $p(x) \propto x^{\beta-1}e^{-\lambda x^{\beta}}$, with parameters $\lambda = 5.7 \times 10^{-4}$ and $\beta = 1.26$ (Figure 2b). These results are consistent with those we find for the other three pollutants (NO_x, PM, VOC): a truncated power law well approximates the distribution for Rome and Florence, and a stretched exponential well approximates that for London (see Supplementary Information A.3). See Supplementary Information A.2 for details on curve fitting.

Regarding the distribution of CO_2 emissions per road, we find that it is well approximated by a truncated power law for all three cities, with exponents $\alpha = 1.55$ and $\lambda = 1.08 \times 10^{-4}$ (Rome, Figure 2f), $\alpha = 1.52$ and $\lambda = 1.30 \times 10^{-4}$ (Florence, Figure 2i), $\alpha = 2.59$ and $\lambda = 2.88 \times 10^{-4}$ (London, Figure 2c). A few *grossly polluted* roads suffer from a significant quantity of emissions, while most of them suffer significantly fewer emissions. London has both the exponents α and λ significantly higher than Rome and Florence, denoting a fairer distribution of the emissions per road (see Figure 8a,c,e,g). Both in Florence and Rome, the top 10% grossly polluted roads are associated with more than 90% of the CO_2 emitted during the period. In London, this quantity is lower (56.7%) but still above half of the city's total emissions of CO_2. Again, we find similar results for the other pollutants (see Supplementary Information A.3).

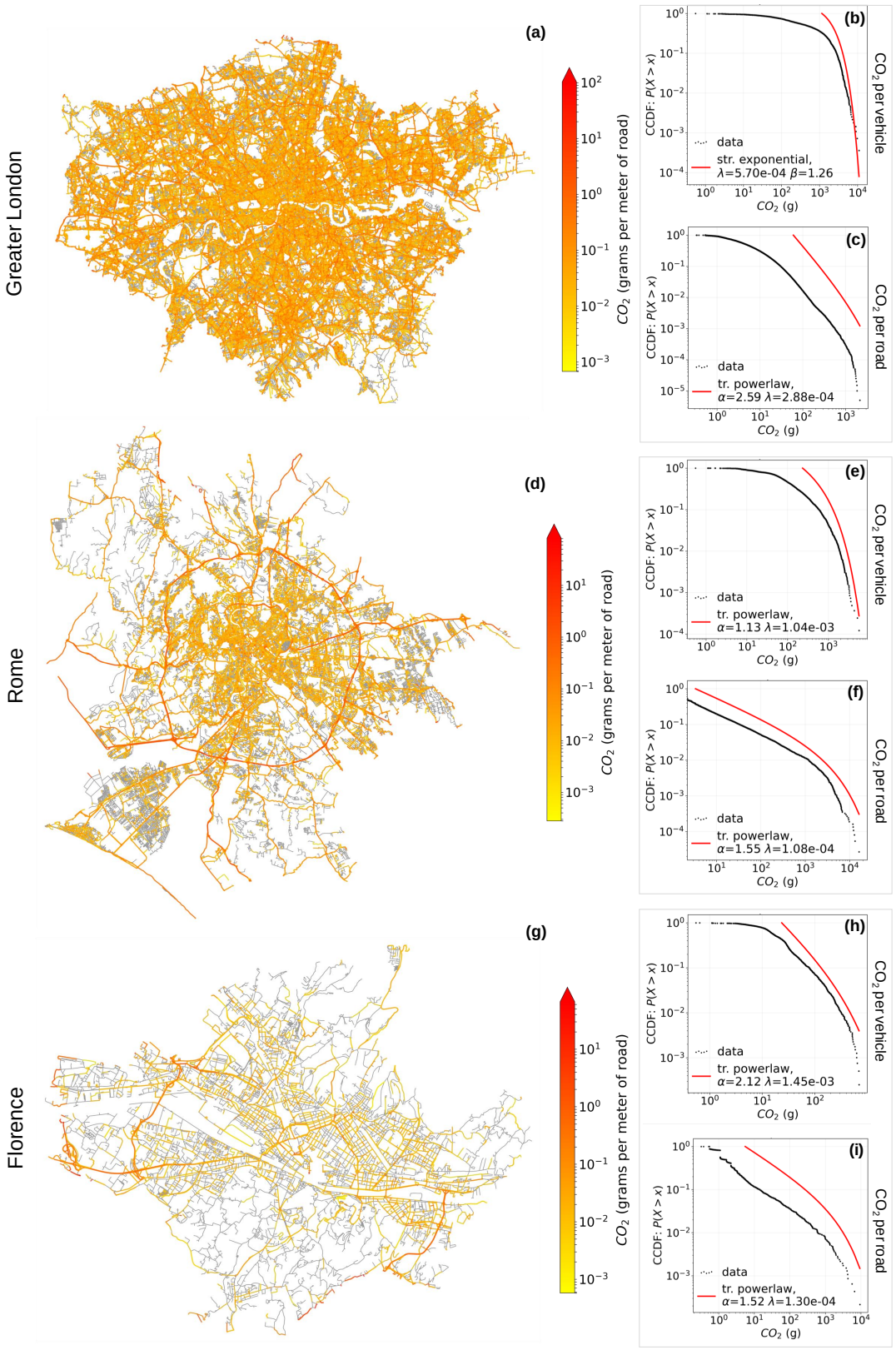

Figure 2: **Distributions of emissions.** The amount of CO_2 emissions (expressed as grams per meter of road emitted during January 2017) on road networks of Geater London (a), Rome (d), and Florence (g). Roads are coloured in a gradient from yellow (low emission) to red (high emission). Panels (b), (e), and (h) show in log-log scale the Complementary Cumulative Distribution Function (CCDF, black dots) of the CO_2 emissions per vehicle, together with the best fit (red curve) in Greater London, Rome, and Florence, respectively. Panels (c), (f), and (i) show the CCDF (black dots) of the CO_2 emissions per road, together with the best fit (red curve), in Greater London, Rome, and Florence, respectively.

	mobility metrics				**roads' features**	
	radius	entropy	max dist.	str. line dist.	betweenness centrality	length
London	0.09	-0.72	0.25	0.95	0.27	0.22
Rome	0.60	-0.51	0.68	0.90	0.30	0.35
Florence	0.32	-0.26	0.48	0.77	0.10	0.25

Table 3: **Correlations between vehicles' emissions, mobility metrics, and road features.** Spearman's correlation coefficients between (left) CO_2 emissions per vehicle and vehicles' mobility metrics (their radius of gyration, uncorrelated entropy, maximum distance and distance travelled straight line), and (right) CO_2 emissions per road and roads' features (betweenness centrality and length).

Relations with mobility and roads' features

To investigate the relationship between a vehicle's emissions and its mobility patterns, we compute the Spearman's correlation coefficient between the amount of CO_2 emissions and four mobility quantities (see Methods): *(i)* the radius of gyration, indicating the characteristic distance travelled by an individual [21, 46, 44]; *(ii)* the mobility entropy [55, 14, 47], characterising the predictability of their visitation patterns; *(iii)* the maximum distance travelled during the period of observation; and *(iv)* the straight line distance travelled during the same period. Radius of gyration and maximum distance show positive and strong correlations with CO_2 emissions for Rome (radius 0.60, max distance 0.68) and slightly weaker for Florence (radius 0.32, max distance 0.48), while they are weak for London (radius 0.09, max distance 0.25). Straight line distance also shows a positive correlation with emissions, and it is strong for all three cities (0.90 for Rome, 0.77 for Florence, 0.95 for London). As one could expect, the more a vehicle travels, the more emissions it produces.

However, vehicles with more regular and predictable behaviour generate the highest emissions, not those with more erratic behaviour. Indeed, the mobility entropy correlates negatively with CO_2 emissions (see Table 3), for all three cities (-0.72 for London, -0.51 for Rome, and -0.26 for Florence). Mobility entropy is low when a vehicle performs a high number of recurring trips, indicating high predictability of its travelling patterns. In contrast, mobility entropy is high when the vehicle performs trips from various origins and destinations, denoting a less predictable travelling behaviour. The observed negative correlations suggest that gross polluters tend to be more regular and predictable than low-emitting vehicles.

We also investigate the relationship between the amount of CO_2 emissions suffered by a road and its network features: *(i)* the betweenness centrality of the edge representing each road in the network; and *(ii)* the road length. The betweenness centrality is based on the frequency with which a road falls on the shortest paths connecting two crossroads (see Methods). Therefore, a road with a high centrality is more likely to be crossed. We find a positive correlation between the betweenness centrality and the amount of emissions in the roads for all the three cities (see Table 3), partially proving that our emissions estimates are consistent with the roads' characteristics, as congested roads are more likely to host a greater quantity of emissions. Similarly, we find a positive correlation with the road length (see Table 3), as longer roads capture more points and, thus, greater quantities of emissions. Similar results hold for the emissions of the other three pollutants in all the three cities (see Supplementary Information A.4). The Figures 2a,d,g show the entire road networks of the Greater London, Rome, and Florence, respectively, with the amount of CO_2 emitted on each road, normalised by the length of the road in order to better appreciate the differences between the roads, and between the cities.

Simulation scenarios

Reducing emissions is a growing concern for cities, and estimating the impact of policies targeting vehicles to reduce their footprint on the city's environment is crucial. We investigate the impact of the vehicles' electrification on the total amount of emissions and the distribution of emissions across the roads. In particular, we study how much the electrification of a certain share of vehicles in our sample would change the amount of CO_2 emissions on the three cities' roads.

We find that the electrification of just the top 1% gross polluters would lead to the same reduction of the CO_2 emitted overall as electrifying 10% random vehicles (Figure 3e,f). We take a neighbourhood in Rome as a case study to investigate the impact of massive electrification on the amount of CO_2 emissions. As the share of gross polluters that shift to electric engines grows, the impact on the roads in reducing emissions becomes more and more evident. In particular, if the top 10% gross polluters shift to an electric engine, 107 roads have a significant reduction on the grams of CO_2 per meter (at least equal to 0.01 g/m), see Table 4 and Figures 3d. In contrast, if 10% of the vehicles that shift to electric engines are chosen at random, only 18 roads have a significant reduction of emissions (see Figure 3b). These results hold for both single neighbourhoods and the entire city (see Figure 3e,f for Rome and Supplementary Information A.5 for London and Florence).

The percentage reduction of the overall emissions of CO_2 grows almost linearly when the share of electric vehicles is chosen at random (i.e., an x% of electric vehicles results in reducing emissions of x%). In contrast, a *Generalised Logistic Function* (also

CO_2 reduction (g/m)	10% random	3% most polluting	10% most polluting
$\geq 10^{-4}$	232	385	665
$\geq 10^{-3}$	137	258	469
$\geq 10^{-2}$	18	53	107
$\geq 10^{-1}$	0	6	10

Table 4: **The roads and their CO_2 reduction in three different scenarios.** Number of roads that experience certain levels of CO_2 reduction in three different scenarios of vehicles' electrification: *(i)* when electrifying 10% random vehicles (left), *(ii)* when electrifying the top 3% most polluting ones (center), and *(iii)* when electrifying the top 10% most polluting ones (right).

known as *Richard's curve* [51], [15]) approximates the growth rate when the vehicles to be electrified are chosen starting from the gross polluters. We use non-linear least squares to fit a *Generalised Logistic Function* $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$, where α represents the upper asymptote, β the growth range, r the growth rate, and ν the slope of the curve. The model gives $R^2 = 0.99$ both for the selected neighbourhood of Rome (Figure 3e) and the whole city (Figure 3f). The estimated growth rates r are 4.84×10^{-2} (the neighbourhood) and 3.96×10^{-2} (Rome), with similar slopes ν (-1.55 and -1.56, respectively); α and β are ≈ 100 and -1 for both the neighbourhood and the city. Similar results hold for Florence (Supplementary Figure 17b, Supplementary Table 13). When considering the whole of London (Supplementary Figure 17a), the growth starts slowly ($\nu = -0.86$): there are less vehicles associated to high levels of emissions, thus electrifying the most polluting vehicles is slightly less effective in reducing CO_2 emissions than in the other two cities (Supplementary Table 13).

Given the increasing importance of remote working, especially during the COVID-19 pandemic [64, 40], we simulate the impact of a massive shift to remote working on the reduction of vehicles' emissions. We identify the vehicles' home and work locations (see Methods) and study the emissions generated from their commuting patterns. We perform a simulation in which a growing share of these commuters become home workers, i.e., they do not travel anymore between their home and work locations. Again, we find that the reduction of CO_2 emissions is more effective when the home workers are gross polluters: in this case, the remote working of the top 1% gross polluters lead to the same reduction reached if they were $\sim 4\%$ random vehicles (see 18).

We fit a *Generalised Logistic Function* to the reduction of CO_2 emissions when the gross polluters become home workers, finding that it well describes its growth, and obtaining estimates for ν (the slope of the curve) that are similar for Rome and Florence (-1.30 and -1.35, respectively), and lower for London (-0.72); see Supplementary Information A.5 and Figure 18.

Overall, the results of both experiments demonstrate that targeting specific profiles of vehicles has the potential to improve significantly emission reduction policies.

Discussion

In this paper, we show the existence of gross polluters, vehicles that are responsible for the greatest quantity of emissions, and grossly polluted roads, which suffer the greatest amount of emissions. In particular, gross polluters are the vehicles that travel the most and, at the same time, whose trajectories are the most predictable. Although some works stated that vehicles have heterogeneous emissions [22, 24], their methodology has been questioned [25] because they use remote sensing measurement sites, which cannot fully represent a vehicle's overall emission level. In contrast, the use of GPS traces, which describes a vehicle's full driving cycle, allows us to estimate the vehicle's instantaneous emissions, and not only emissions at one single point.

Our findings hold for all the considered pollutants (CO_2, NO_x, PM, VOC) and mirror heterogeneous patterns in the distribution of emissions across vehicles and roads, which are well approximated by heavy-tailed distributions with exponents that vary with the city. These peculiar exponents may depend on the characteristics of the city's road network and people's commuting behaviour. For example, in Greater London, which has a huge and dense road network, people use private vehicles less intensively [58] than in Rome and Florence. Londoners' mobility behaviour leads to a distribution of emissions per vehicle that is fairer than people's behaviour in Rome, which is characterised by a huge and sparse road network and an intensive use of private vehicles [17].

Our study can be applied in principle to any city, provided the availability of vehicles' GPS trajectories and road networks data, and may help in finding more effective strategies to reduce emissions. For example, our study clearly demonstrates that blocking the circulation based on a non-informed choice of the vehicles' (e.g., blocking vehicles with odd or even number plates) has less impact on reducing emissions than identifying and targeting a small share of gross polluters. Moreover, we provide a precise model to estimate the overall reduction of CO_2 emissions generated by the electrification of a certain share of vehicles or by a reduction in the number of commuting patterns travelled by the vehicles (caused for example by a transition to remote working of their drivers).

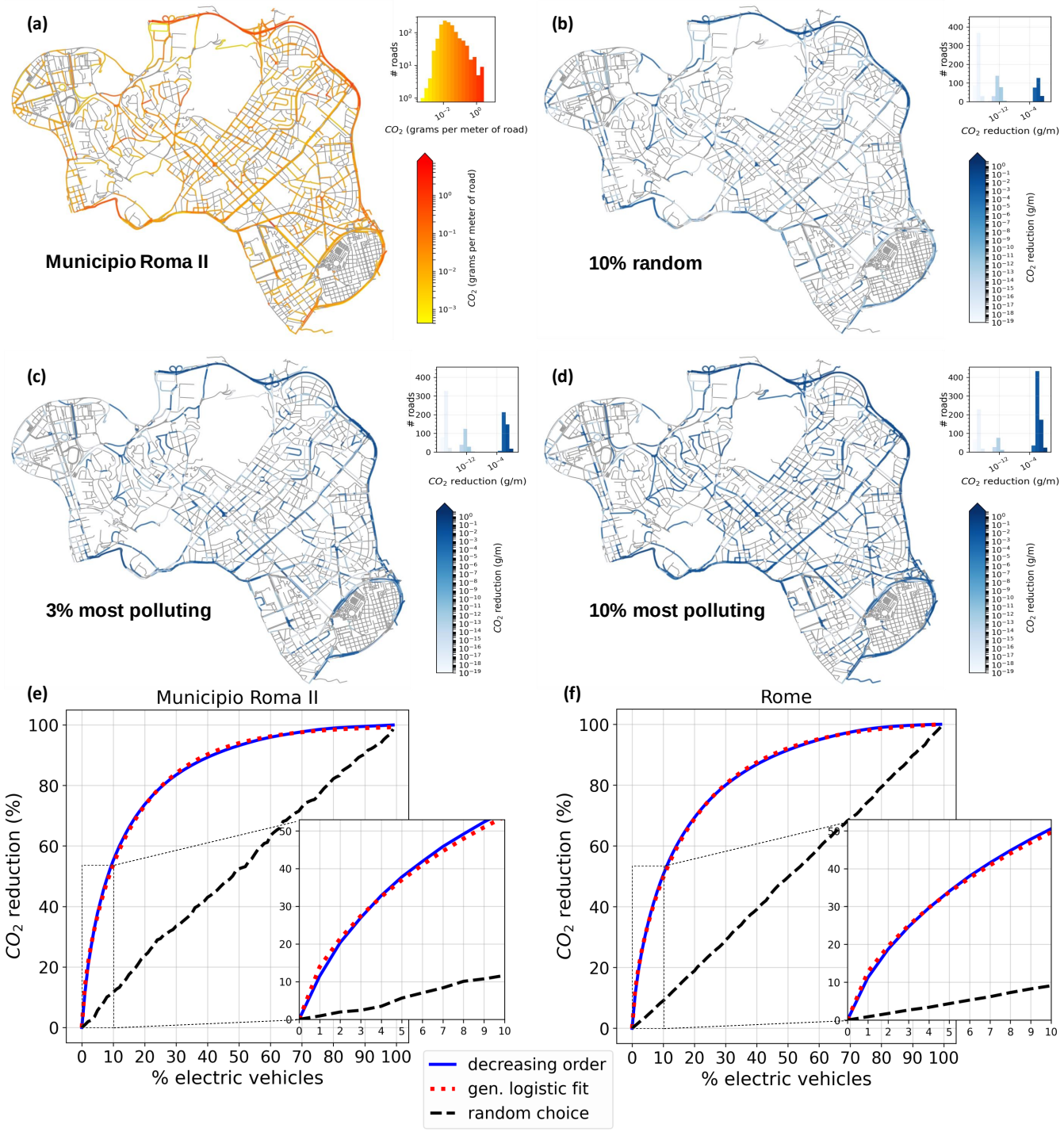

Figure 3: **Simulation of electrification.** Analysis of the impact on CO_2 emissions of the electrification of a sample of vehicles moving within a neighbourhood (the Second Municipality of Rome) during January 2017. (a) Distribution of emissions of CO_2 across the roads. (b) Reduction of emissions if a random 10% of the vehicles shift to electric engines. (c) Reduction of emissions if top 3% of gross polluters shift to electric engines. (d) Reduction of emissions if top 10% gross polluters shift to electric engines. (e) Percentage reduction of CO_2 emissions corresponding to a certain share (0-100%) of electric vehicles for the Second Municipality of Rome. The inset plots zoom on the first tenth share of electric vehicles. The blue solid line is when the vehicles to be electrified are chosen from the most polluting to the least polluting. The black dashed line indicates the emission reduction when the vehicles to be electrified are chosen at random. The dotted red line is the *Generalised Logistic Function* (*Richard's curve*) fit: $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$, where α represents the upper asymptote, β the growth range, r the growth rate, and ν the slope of the curve. (f) Same plot for the whole city of Rome.

The framework proposed in this paper can be improved by including in the emissions model more information about the vehicles, such as emission category (e.g., Euro 2, Euro 3) and weight type (e.g., light-duty, heavy-duty). Moreover, our analysis can be deepened by investigating how the estimation of emissions changes in time: how does the amount of emissions change between weekdays and weekend days? And how does it changes with the season or the weather?

In the meanwhile, our study may contribute to shape the discussion on how to measure emissions with digital data that are massively available everywhere on earth, such as GPS traces, and how to use such measurements to simulate emission reduction scenarios. If we learn how to use such a resource, we have the potential of monitoring in real-time the level of emission in our urban environments and to take immediate and informed actions when they overcome a certain tolerance threshold. This is crucial because the decisions of policy makers depend on what we measure, how good our measurements are, and how promptly we react to these measurements.

Methods

Filtering, speed and acceleration. Each trajectory point is associated with a vehicle identifier, a trajectory identifier, a timestamp, and a latitude/longitude pair. The sampling rate of the trajectory points may affect the estimate of instantaneous speed and acceleration. Since the mean time interval between trajectory points varies with the city (it is around one minute for London and around four minutes for Rome and Florence), we retain only the sub-trajectories composed by consecutive points in an interval of 120 seconds, ending up with a mean time interval between the points that is around one minute for all the cities. We then compute, for each vehicle, the speed and acceleration in each point, and retain only points whose speed is lower than 300 km/h and whose acceleration is in the range $[-10, +10]$ m/s^2 (as suggested by Nyhan et al.[42]). Different works use different values for these filtering parameters (e.g., the time interval between the points is set to 5 seconds in Nyhan et al.[42], 5 to 50 seconds in Liu et al.[35], 1 second in Chen et al.[7], 3 seconds in Sui et al.[57]). Our choice for the parameters derives from the sampling rate of our data and are a trade-off between the reliability of the results and the data coverage.

Map-matching. The road network of each city is extracted from OpenStreetMap (OSM), a collaborative project to create a free editable map of the world and provide the geodata underlying the map. In particular, the road network is as a multigraph in which edges represent public roads accessible to vehicles (including service roads). Each edge e is a couple of two identifiers, indicating the starting and ending OSM nodes. To download, compute statistics, and visualise the road networks we use the Python library OSMnx [5]. Our map-matching step consists in a ball tree nearest-neighbour algorithm that maps each point of a GPS trajectory to its nearest edge in the OSM network.

Computing emissions. We implement a microscopic emissions model [42] to compute the instantaneous emissions associated to each trajectory point p. We denote the quantity of pollutant $j \in \{\text{CO}_2, \text{NO}_x, \text{PM}, \text{VOC}\}$ emitted at point p as E_j^p and the instantaneous speed and acceleration of the vehicle in p as v_p and a_p, respectively. For each vehicle, information about its engine type (whether it is a petrol, diesel, or LPG vehicle) is available. This information determines, together with the type of pollutant, the emission factors f. We use the following equation to compute the instantaneous emissions E_j^p 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). In Supplementary Table 15 we show the value of factors $f_1, \ldots, f_6$ varying the vehicle's type and acceleration.

Mobility measures and roads' centrality. We use four quantities to describe the mobility of a vehicle u: *(i)* the radius of gyration [21, 46, 44] $r_g(u) = \sqrt{\frac{1}{n} \sum_{i \in P} dist(\mathbf{r}_i(u) - \mathbf{r}_{cm}(u))^2}$, where P is the set of n points recorded for u, $\mathbf{r}_i(u)$ indicates the coordinates of trajectory point $i \in P$, and $\mathbf{r}_{cm}(u)$ is the centre of mass of u; *(ii)* the temporal-uncorrelated entropy [55, 14, 47] $S(u) = -\sum_{i=1}^{N_u} p_u(i) \log_2 p_u(i)$, where N_u is the number of distinct locations visited by u, and $p_u(i)$ is the probability that u visits location i; *(iii)* the maximum distance travelled by u, defined as $d_{max}(u) = \max_{1 \leq i < j \leq n} dist(\mathbf{r}_i(u), \mathbf{r}_j(u))$, with $\mathbf{r}_i(u), \mathbf{r}_j(u)$ are the coordinates of two locations i and j visited by u; *(iv)* the straight line distance travelled by u, computed as the sum of all the distances travelled by u. We measure the centrality of a road in a road network, a proxy of its traffic volume in the city, as its betweenness centrality. In network science, the betweenness centrality of an edge e (i.e., a road in our case) is defined as $C_b(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}$, where V is the set of nodes in the network, $\sigma(s,t)$ is the number of shortest paths between s and t, and $\sigma(s,t|e)$ is the number of those shortest paths passing through edge e.

Home and work locations. We identify a vehicle's home and work locations selecting the starting and ending points of its trajectories. We then spatially cluster the points within a radius of 250 meters and take the centroid of each cluster as the vehicle's

stop locations. We perform the clustering because the position of the starting (or ending) points of trajectories that start from (or end to) the same semantic location may not exactly coincide because: (a) a driver can park the vehicle within a certain radius from the location; (b) the first point sent by the onboard GPS device often lacks of precision and it is thus discarded, tacking as starting point of the trajectory the second point sent. To identify a vehicle's home and work locations, we use a principle commonly used in the literature [61, 60, 43]: the home location is the stop location corresponding to the most frequent cluster; the work location is the stop location corresponding to the second most frequent cluster. We discard the vehicles for which it is not possible to identify the most frequent stop location(s) (as for example the vehicle visited each location only once). We successfully identify both the home and work locations for 55%, 31%, and 16% of the vehicles moving in London, Rome, and Florence, respectively.

References

[1] J. Alstott, E. Bullmore, and D. Plenz. powerlaw: A python package for analysis of heavy-tailed distributions. *PLOS ONE*, 9(1):1–11, 01 2014.

[2] H. M. A. Aziz and S. 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. *International Journal of Sustainable Transportation*, 12(3):192–204, 2018.

[3] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali. Smart cities of the future. *The European Physical Journal Special Topics*, 214(1):481–518, 2012.

[4] J. D. Berman and K. Ebisu. Changes in u.s. air pollution during the covid-19 pandemic. *Science of The Total Environment*, 739:139864, 2020.

[5] G. Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126 – 139, 2017.

[6] C. Q. Camargo, J. Bright, G. McNeill, S. Raman, and S. A. Hale. Estimating traffic disruption patterns with volunteered geographic information. *Scientific reports*, 10(1):1–8, 2020.

[7] J. Chen, W. Li, H. Zhang, W. Jiang, W. Li, Y. Sui, X. Song, and R. Shibasaki. Mining urban sustainable performance: Gps data-based spatio-temporal analysis on on-road braking emission. *Journal of Cleaner Production*, 270:122489, 2020.

[8] H. S. Chong, S. Kwon, Y. Lim, and J. Lee. Real-world fuel consumption, gaseous pollutants, and co2 emission of light-duty diesel vehicles. *Sustainable Cities and Society*, 53:101925, 2020.

[9] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[10] S. Çolak, A. Lima, and M. C. González. Understanding congested travel in urban areas. *Nature communications*, 7(1):1–8, 2016.

[11] P. deSouza, A. Anjomshoaa, F. Duarte, R. Kahn, P. Kumar, and C. Ratti. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60:102239, 2020.

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

[13] F. Dutheil, J. Baker, and V. Navel. Covid-19 as a factor influencing air pollution? *Environmental Pollution*, 263:114466, 04 2020.

[14] N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[15] D. Fekedulegn and J. Colbert. Parameter estimation of nonlinear growth models in forestry. *Silva Fennica*, 33, 01 1999.

[16] P. Forster, H. Forster, M. Evans, M. Gidden, C. Jones, C. Keller, R. Lamboll, C. Quéré, J. Rogelj, D. Rosen, C.-F. Schleussner, T. Richardson, C. Smith, and S. Turnock. Current and future global climate impacts resulting from covid-19. *Nature Climate Change*, 2020.

[17] A. Fuschiotto, C. Severini, R. Gigli, A. Picardi, D. Ciavatta, and G. Mancinelli. Rapporto mobilità 2019. Technical report, Dipartimento Mobilità e Trasporti Roma Capitale, 2019. Available at `https://romamobilita.it/it/media/pubblicazioni/rapporto-mobilita-2019`.

[18] R. Gallotti, A. Bazzani, and S. Rambaldi. Towards a statistical physics of human mobility. *International Journal of Modern Physics C*, 23(09):1250061, 2012.

[19] R. Gallotti, A. Bazzani, S. Rambaldi, and M. Barthelemy. A stochastic model of randomly accelerated walkers for human mobility. *Nature Communications*, 7(1):12600, 2016.

[20] C. K. Gately, L. R. Hutyra, S. Peterson, and I. Sue Wing. Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone gps data. *Environmental Pollution*, 229:496–504, 2017.

[21] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[22] P. Guenther, G. Bishop, J. Peterson, and D. Stedman. Emissions from 200 000 vehicles: a remote sensing study. *Science of The Total Environment*, 146-147:297 – 302, 1994.

[23] G. He, Y. Pan, and T. Tanaka. The short-term impacts of covid-19 lockdown on urban air pollution in china. *Nature Sustainability*, 3, 12 2020.

[24] Y. Huang, B. Organ, J. L. Zhou, N. C. Surawski, G. Hong, E. F. Chan, and Y. S. Yam. Remote sensing of on-road vehicle emissions: Mechanism, applications and a case study from hong kong. *Atmospheric Environment*, 182:58 – 74, 2018.

[25] Y. Huang, N. Surawski, Y. S. Yam, C. Lee, J. Zhou, B. Organ, and E. Chan. Re-evaluating effectiveness of vehicle emission control programs targeting high-emitters. *Nature Sustainability*, 3, 11 2020.

[26] Intergovernmental Panel on Climate Change. *Transport*, page 599–670. Cambridge University Press, 2015.

[27] A. Jenn. Emissions benefits of electric vehicles in uber and lyft ride-hailing services. *Nature Energy*, 5(7):520–525, 2020.

[28] R. Kitchin. The real-time city? big data and smart urbanism. *GeoJournal*, 79(1):1–14, 2014.

[29] A. Klaus, S. Yu, and D. Plenz. Statistical analyses support power law distributions found in neuronal avalanches. *PLOS ONE*, 6(5):1–12, 05 2011.

[30] C. Kroll, A. Warchold, and P. Pradhan. Sustainable development goals (sdgs): Are we successful in turning trade-offs into synergies? *Palgrave Communications*, 5(1):1–11, 2019.

[31] D. Le Blanc. Towards integration at last? the sustainable development goals as a network of targets. *Sustainable Development*, 23(3):176–187, 2015.

[32] C. Le Quéré, R. Jackson, M. Jones, A. Smith, S. Abernethy, R. Andrew, A. De-Gol, D. Willis, Y. Shan, J. Canadell, P. Friedlingstein, F. Creutzig, and G. Peters. Temporary reduction in daily global co2 emissions during the covid-19 forced confinement. *Nature Climate Change*, 10:1–7, 07 2020.

[33] M. Li, S. Gao, F. Lu, H. Tong, and H. Zhang. Dynamic estimation of individual exposure levels to air pollution using trajectories reconstructed from mobile phone data. *International Journal of Environmental Research and Public Health*, 16(22):4522, 2019.

[34] X. Liang, S. Zhang, Y. Wu, J. Xing, X. He, K. Zhang, S. Wang, and J. Hao. Air quality and health benefits from fleet electrification in china. *Nature Sustainability*, 2:962–971, 10 2019.

[35] J. Liu, K. Han, X. M. Chen, and G. P. Ong. Spatial-temporal inference of urban traffic emissions based on taxi trajectories and multi-source urban data. *Transportation Research Part C: Emerging Technologies*, 106:145 – 165, 2019.

[36] Z. Liu, P. Ciais, Z. Deng, R. Lei, S. J. Davis, S. Feng, B. Zheng, D. Cui, X. Dou, B. Zhu, et al. Near-real-time monitoring of global co 2 emissions reveals the effects of the covid-19 pandemic. *Nature communications*, 11(1):1–12, 2020.

[37] M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo. Deep Learning for Human Mobility: a Survey on Data and Models. *arXiv e-prints*, page arXiv:2012.02825, Dec. 2020.

[38] J. M. Luján, V. Bermúdez, V. Dolz, and J. 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). *Atmospheric Environment*, 174:112 – 121, 2018.

[39] H. Lwin and T. Naing. Estimation of road traffic congestion using gps data. *IJARCCE*, 4:1–5, 12 2015.

[40] L. Nagel. The influence of the covid-19 pandemic on the digital transformation of work. *International Journal of Sociology and Social Policy*, 2020.

[41] M. Nyhan, I. Kloog, R. Britter, C. Ratti, and P. Koutrakis. Quantifying population exposure to air pollution using individual mobility patterns inferred from mobile phone data. *Journal of Exposure Science & Environmental Epidemiology*, 29, 2018.

[42] M. Nyhan, S. Sobolevsky, C. Kang, P. Robinson, A. Corti, M. Szell, D. Streets, Z. Lu, R. Britter, S. R. Barrett, and C. Ratti. Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmospheric Environment*, 140:352 – 363, 2016.

[43] L. Pappalardo, L. Ferres, M. Sacasa, C. Cattuto, and L. Bravo. An individual-level ground truth dataset for home location detection. *arXiv e-prints*, page arXiv:2010.08814, Oct. 2020.

[44] L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, and F. Giannotti. Understanding the patterns of car travel. *European Physical Journal: Special Topics*, 215(1), 2013.

[45] L. Pappalardo, F. Simini, G. Barlacchi, and R. Pellungrini. scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data. *arXiv e-prints*, page arXiv:1907.07062, July 2019.

[46] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 2015.

[47] L. Pappalardo, M. Vanhoof, L. Gabrielli, Z. Smoreda, D. Pedreschi, and F. Giannotti. An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics*, 2(1):75–92, 2016.

[48] M. Picornell, T. Ruiz, R. Borge, P. García-Albertos, D. Paz, and J. Lumbreras. Population dynamics based on mobile phone data to improve air pollution exposure assessments. *Journal of Exposure Science & Environmental Epidemiology*, 29, 2018.

[49] M. N. Rahman and A. O. Idris. Tribute: Trip-based urban transportation emissions model for municipalities. *International Journal of Sustainable Transportation*, 11(7):540–552, 2017.

[50] A. Reznik, M. Kissinger, and N. Alfasi. Real-data-based high-resolution ghg emissions accounting of urban residents private transportation. *International Journal of Sustainable Transportation*, 13(4):235–244, 2019.

[51] F. J. Richards. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10(29):290–300, 1959.

[52] H. Ritchie. Sector by sector: where do global greenhouse gas emissions come from?, 2020. Available at `https://ourworldindata.org/ghg-emissions-by-sector`.

[53] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. Slavin Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. Luccioni, T. Maharaj, E. D. Sherwin, S. Karthik Mukkavilli, K. P. Kording, C. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. Chayes, and Y. Bengio. Tackling Climate Change with Machine Learning. *arXiv e-prints*, page arXiv:1906.05433, June 2019.

[54] J. J. So, N. Motamedidehkordi, Y. Wu, F. Busch, and K. Choi. Estimating emissions based on the integration of microscopic traffic simulation and vehicle dynamics model. *International Journal of Sustainable Transportation*, 12(4):286–298, 2018.

[55] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[56] J. Stipancic, L. Miranda-Moreno, A. Labbe, and N. Saunier. Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected gps data. *Transportation letters*, 11(7):391–401, 2019.

[57] Y. Sui, H. Zhang, X. Song, F. Shao, X. Yu, R. Shibasaki, R. Sun, M. Yuan, C. Wang, S. Li, and Y. Li. Gps data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. *Journal of Cleaner Production*, 227:495 – 505, 2019.

[58] Transport for London. Travel in london, report 12. Technical report, 2019. Available at `http://content.tfl.gov.uk/travel-in-london-report-12.pdf`.

[59] United Nations General Assembly. Transforming our world: the 2030 agenda for sustainable development. Technical report, 2015. Accessed: 2021-02-23.

[60] M. Vanhoof, C. Lee, and Z. Smoreda. *Performance and Sensitivities of Home Detection on Mobile Phone Data*, chapter 8, pages 245–271. John Wiley & Sons, Ltd, 2020.

[61] M. Vanhoof, F. Reis, T. Ploetz, and Z. Smoreda. Assessing the quality of home detection from mobile phone data for official statistics. *Journal of Official Statistics*, 34(4):935 – 960, 2018.

[62] Z. S. Venter, K. Aunan, S. Chowdhury, and J. Lelieveld. Covid-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences*, 117(32):18984–18990, 2020.

[63] V. Voukelatou, L. Gabrielli, I. Miliou, S. Cresci, R. Sharma, M. Tesconi, and L. Pappalardo. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics*, 06 2020.

[64] L. Vyas and N. Butakhieo. The impact of working from home during covid-19 on work and life domains: an exploratory study on hong kong. *Policy Design and Practice*, pages 1–18, 2020.

[65] X. Xie, I. Semanjski, S. Gautama, E. Tsiligianni, N. Deligiannis, R. Rajan, F. Pasveer, and W. Philips. A review of urban air pollution monitoring and exposure assessment methods. *ISPRS International Journal of Geo-Information*, 6(12):389, 2017.

[66] E.-h. Yoo, Q. Pu, Y. Eum, and X. Jiang. The impact of individual mobility on long-term exposure to ambient pm2.5: Assessing effect modification by travel patterns and spatial variability of pm2.5. *International Journal of Environmental Research and Public Health*, 18(4), 2021.

[67] H. Yu, A. Russell, J. Mulholland, and Z. Huang. Using cell phone location to assess misclassification errors in air pollution exposure estimation. *Environmental Pollution*, 233:261 – 266, 2018.

[68] Q. Yu, H. Zhang, W. Li, X. Song, D. Yang, and R. Shibasaki. Mobile phone gps data in urban customized bus: Dynamic line design and emission reduction potentials analysis. *Journal of Cleaner Production*, 272:122471, 2020.

[69] S. Zhu, I. Kim, and K. Choi. High-resolution simulation-based analysis of leading vehicle acceleration profiles at signalized intersections for emission modeling. *International Journal of Sustainable Transportation*, 0(0):1–11, 2020.

Appendices

A Supplementary Information

A.1 Trajectory sampling

The sampling rate of the trajectory points (i.e., the mean time interval between consecutive points forming a trajectory) in our data is smaller for London (around one point every minute on average) than for Rome and Florence (around one point every 4 minutes on average), we retain only the sub-trajectories composed by consecutive points in an interval of 120 seconds before computing speed and acceleration. Thus, after this preprocessing step, London's sampling rate results higher than the one of the other two cities. Thus, after the estimation of the emissions, we perform a sub-sampling of London's data, filtering out points so that the final mean sampling rate per trajectory is similar to that of Rome and Florence (Table 1). However, one can argue that, still, both London's number of trips per vehicle and number of points per trajectory (~ 88 and ~ 9.9, respectively) are considerably bigger than the Roman (~ 20 and ~ 4.7, respectively) and Florentine (~ 7 and ~ 2.3, respectively) ones. As the empirical Complementary Cumulative Distribution Functions of the data show (Figures 4 and 5), this fact does not influence the shape of the distributions, that is reasonably similar across the three cities.

A.2 Fitting of distributions

We fit the distributions of emissions of the four pollutants (CO_2, NO_x, PM, VOC) across vehicles and roads for the three cities (Florence, Rome, Greater London) using the statistical methods developed in Clauset et al. [9] and Klaus et al. [29] and implemented by Alstott et al. 2014 [1]. These methods use a maximum-likelihood fitting and evaluate the goodness-of-fit using the Kolmogorov-Smirnov distance, comparing different models with a log-likelihood ratio test. This fitting procedure is used on 4x2x3=24 distributions in total.

We fit five models – power law, truncated power law, log-normal, exponential, and stretched exponential – to the data and compare pairwise their goodness-of-fit with a log-likelihood ratio test. We give particular attention to the comparison with the exponential model, as it is generally considered the minimum alternative candidate for evaluating the heavy-tailedness of a distribution. Indeed, if the exponential is the best fit, one should reconsider the hypothesis of heavy-tail distribution for the data. For each model, we use log-likelihood ratio tests for comparing its goodness-of-fit to the data with all the others', and we choose the best fitting model as the one (if any) that wins the higher number of comparisons.

As the power law behaviour of all our distributions starts from a certain value x_{min} (as it happens for the majority of the phenomena that obey power laws), we estimate x_{min} as follows. We compute the power law fit for each possible value of x_{min} in the range of the data, and evaluate the goodness of fit with the Kolmogorov-Smirnov statistic (D), that is the maximum distance between the empirical Cumulative Distribution Function (CDF) and that of the fitted model (the lower D, the better the fit). We choose the x_{min} corresponding to the global minimum of D. However, the choice of x_{min} influences the precision of the estimator we use. Indeed, the Maximum Likelihood Estimator $\hat{\alpha}$ of the exponent of the power law is asymptotically Gaussian with variance $\sigma = \frac{(\alpha-1)}{n}$ (see Proposition B.4 in Clauset et al. [9]), where n is the number of data points $x_i \geq x_{min}$. As σ grows with x_{min} (as n decreases), the trade-off is given by the choice of the best power law fit (i.e., the x_{min} with minimum D) with a low variability for $\hat{\alpha}$. As Figures 6 and 7 show for Rome and Florence, the global minimum of D also correspond to very low σ, with the only exception of the distributions of CO_2 and VOC emissions per vehicle in Rome [1]. This is not the case for London, for which the global minima for D are reached at high values of x_{min}, which correspond to σ higher than 0.4. At the same time, pretending a lower σ would lead to a solution with high values of D (around 1.5). We conclude that a power law fit is not reasonable for London. Instead, the distributions for London are better approximated by a stretched exponential.

A.3 Distribution of emissions

Emissions distribute across both vehicles and roads in a heterogeneous way. In Figure 8 we show the Lorenz curves representing the distributions of the four pollutants per vehicle (Figures 8b,d,f,h) and road (Figures 8a,c,e,g). These curves, together with the associated Gini coefficients, highlight the inequality of our distributions. The Gini coefficient varies between 0 and 1, with high values corresponding to unequal distributions. We obtained, for all the cities and distributions, a Gini coefficient that is always above 0.55. The most unequal distributions are those across the roads of Rome and Florence (Figures 8a,c,e,g), whose Gini coefficient is nearly always above 0.8, while for London these distributions are less unequal (Gini coefficient around 0.65).

[1] Nevertheless, the local minimum of D found when $\sigma < 0.05$ are not far from the global minimum (there is a difference of ~ 0.02), so we can choose those x_{min} with a reasonable loss in terms of goodness of fit.

Instead, the emissions of CO_2 (Figure 8a) and VOC (Figure 8h) per vehicle are distributed in a similar way across the cities (with a maximum difference between the Gini coefficients equal to 0.08). This is not the case for the emissions of NO_x (Figure 8d) and PM (Figure 8f), whose inequality in their distributions across the vehicles change with the city, being those in London more equal (Gini coefficient around 0.62) and those in Rome (in the case of NO_x) or Florence (in the case of PM) more unequal.

Regarding the results of fitting the distributions, all of them are well approximated by at least one of the models we take into consideration (namely, power law, truncated power law, log-normal, stretched exponential), and at least one of these models is always a better fit than the exponential one. Considering the distributions of emissions across vehicles (Figure 9) and London as a first case, they are always well approximated by a stretched exponential model, with λ varying with the pollutant, and β in the interval $[1.22, 1.30]$ for all the four pollutants (Figure 9a,d,g,j). Moreover, in the case of NO_x and VOC, also the log-normal model is a good fit for the distribution of their emissions across London's vehicles, with parameters respectively ($\mu = 0.31$, $\sigma = 0.53$) for NO_x emissions and ($\mu = 2.02$, $\sigma = 0.52$) for VOC emissions (Figure 9d,j). Then, considering the case of Rome, the distributions of emissions of both CO_2 and VOC are well approximated by a truncated power law, with similar exponents α but different λ, as they are respectively ($\alpha = 1.13$, $\lambda = 1.04 \times 10^{-3}$) for CO_2 emissions and ($\alpha = 1.21$, $\lambda = 0.51$) for VOC emissions (Figure 9b,k). For the distributions of NO_x and PM emissions the log-likelihood ratio tests give no overall best fitting model (Figure 9e,h). Indeed, in the case of NO_x emissions, the power law, the truncated-power law and the log-normal all are better fits than the exponential and stretched exponential models (Table 5), but the tests give no evidence for concluding that one of the three has a better fit than the others. For the PM emissions, all the four heavy-tailed models have better fits than the exponential one, but none of them is better then the others in terms of goodness of fit (Table 6). Finally, the distributions of CO_2, NO_x, PM, VOC emissions per vehicle in the city of Florence are all well approximated by a truncated power law (Figure 9c,f,i,l), with estimated parameters respectively: ($\alpha = 2.12$, $\lambda = 1.45 \times 10^{-3}$), ($\alpha = 1.97$, $\lambda = 0.90$), ($\alpha = 1.76$, $\lambda = 25.20$), ($\alpha = 1.73$, $\lambda = 1.39$).

The distributions of emissions across the roads of the three cities are more homogeneous than those across the vehicles w.r.t. their best fitting model (Figure 10). Indeed, for both Rome and Florence these distributions are all well approximated by a truncated power law with parameters α, λ that are similar for each pollutant (Figure 10b,c,e,f,h,i,k,l): for the emissions of CO_2 per road they are respectively ($\alpha = 1.55$, $\lambda = 1.08 \times 10^{-4}$) in Rome and ($\alpha = 1.52$, $\lambda = 1.30 \times 10^{-4}$) in Florence; for the NO_x, ($\alpha = 1.65$, $\lambda = 0.27$) in Rome and ($\alpha = 1.56$, $\lambda = 0.29$) in Florence; for the PM, ($\alpha = 1.61$, $\lambda = 42.25$) in Rome and ($\alpha = 1.52$, $\lambda = 35.24$) in Florence; for the VOC, ($\alpha = 1.66$, $\lambda = 0.06$) in Rome and ($\alpha = 1.57$, $\lambda = 0.10$) in Florence. These results reflect the similarity of these distributions in Rome and Florence that we observed in their Lorenz curves (Figures 8a,c,e,g). In the case of London's distributions of emissions per road, for the CO_2 the best fit is again represented by the truncated power law (Figure 10a), with exponents $\alpha = 2.59$ and $\lambda = 2.88 \times 10^{-4}$, while for the other three pollutants the comparisons give no overall best-fitting model. Indeed, regarding NO_x, PM and VOC emissions across London's roads (Figure 10d,g,j), all the four heavy-tailed models have better fits than the exponential one, and at the same time the log-normal, the truncated power law and the stretched exponential models all have better fits than the simple power law, with no overall winning model between the three (Tables 7, 8 and 9).

A.4 Relations with vehicles' mobility and roads' features

The relationship between the emissions of the four pollutants and, from one side, the vehicles' mobility behaviour and, from the other, the roads' characteristics, is often non-linear (Figures 11-16). Thus, we used the Spearman's correlation coefficient to investigate these relationships. Considering the relation between the quantities of air pollutants emitted per vehicle and four quantities describing its mobility (the radius of gyration, temporal-uncorrelated entropy, maximum distance and straight line distance), the strongest (and almost linear) correlation is that with the distance travelled straight line by the vehicles (above 0.77, with the exception of PM emissions in Florence), for the emissions of all the air pollutants and in all the three cities (Figures 11, 13, 15, and Tables 10, 11, 12): the more a vehicle travels, the more emissions it produces. Similarly, we find positive correlations with the maximum distance travelled by a vehicle; in London, these coefficients are in the range $[0.18, 0.25]$ for the four air pollutants, while in Rome the coefficients are always above 0.45, and in Florence they are positive but slightly weaker (0.48 for the CO_2, 0.34 for the NO_x, 0.33 for the VOC) with the only exception of the PM emissions (see Table 11), that exhibit a negative correlation (-0.27). This Florentine PM exception (see Table 11) is also confirmed by a weaker correlation – compared to the other two cities – with the straight line distance (0.53), and a negative correlation with the radius of gyration (-0.45). The correlations with the vehicles' mobility entropy are strong and negative for all air pollutants and in all cities, with the highest coefficients found in London (-0.72 for CO_2 and PM, -0.75 for NO_x and VOC), and the lower in Florence (-0.26 for CO_2, -0.39 for NO_x, -0.60 for PM, -0.45 for VOC). As a low mobility entropy indicates that a vehicle performs a high number of recurring trips, these negative correlations suggest that gross polluters tend to be more regular and predictable than low-emitting vehicles. Finally, the correlations with the radius of gyration are almost null in London, while they are positive and strong for Rome (CO_2 0.60, NO_x 0.39, PM 0.32, VOC 0.53), and weaker for Florence (CO_2 0.32, NO_x 0.17, PM -0.45, VOC 0.13). In conclusion, as a general rule, even if with a few exceptions, the gross polluters are those vehicles that travel the most, and whose travelling behaviour is more predictable (e.g. the commuters).

Concerning the relation between the quantities of air pollutants emitted on the roads and *(i)* their (betweenness) centrality w.r.t. the entire network and *(ii)* their length, we find positive correlation coefficients for all air pollutants and cities (Figures 12, 14, 16, and Tables 10, 11, 12). As the betweenness centrality of a road is based on the frequency with which it falls on the shortest paths connecting two crossroads in the network, and, thus, it can be considered as a proxy of the traffic volume it hosts, these results suggest that the grossly polluted roads are more likely to be the longer and more congested ones in the road network.

A.5 Emissions reduction simulations

The Figure 17 shows the reduction of the CO_2 emitted overall by the vehicles in London and Florence when a certain share of them are zero-emitting vehicles (e.g. electric vehicles). We compute this reduction in two cases: when the vehicles to be electrified are chosen starting from the most polluting ones (and proceeding in decreasing order), and when they are chosen at random. In London, the reduction of CO_2 emissions in the first case is ~ 6 times more effective than with the random choice (Figure 17a); this means that electrifying the top 1% most polluting vehicles leads to a $\sim 6\%$ reduction of the overall emissions of CO_2, while a random 1% electric vehicles leads to a reduction of $\sim 1\%$ in the CO_2 emissions. In Florence, this difference is even bigger, with the electrification of the top 1% most polluting vehicles being ~ 10 times more effective than the electrification of 1% vehicles chosen at random (Figure 17b). We find that the growth of this reduction is well described by a *Generalised Logistic Function* (*GLM*, or *Richard's curve*). Indeed, we use non-linear least squares to fit a GLM $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$ to the reduction of the overall emissions of CO_2 when the electrified vehicles are chosen in decreasing order starting from the most polluting ones. We set the starting values for the parameters following Fekedulegn et al. [15]. The parameters of the curve change with the city (Table 13); in particular, its slope is -0.86 for London and -1.91 in Florence, as this latter experiences a faster growth in the reduction of overall CO_2 emissions obtained by electrifying the most polluting vehicles.

We also simulate the impact of a massive shift to remote working on the reduction of vehicles' CO_2 emissions in the three cities (Figure 18), again finding that the reduction of the emissions is more effective when the home workers are gross polluters: in this case, the remote working of the top 1% gross polluters lead to the same reduction reached if they were $\sim 4\%$ random vehicles. Again, the growth of this reduction is well described by a *Generalised Logistic Function*, whose parameters slightly change with the city and are shown in Table 14.

B Supplementary Figures

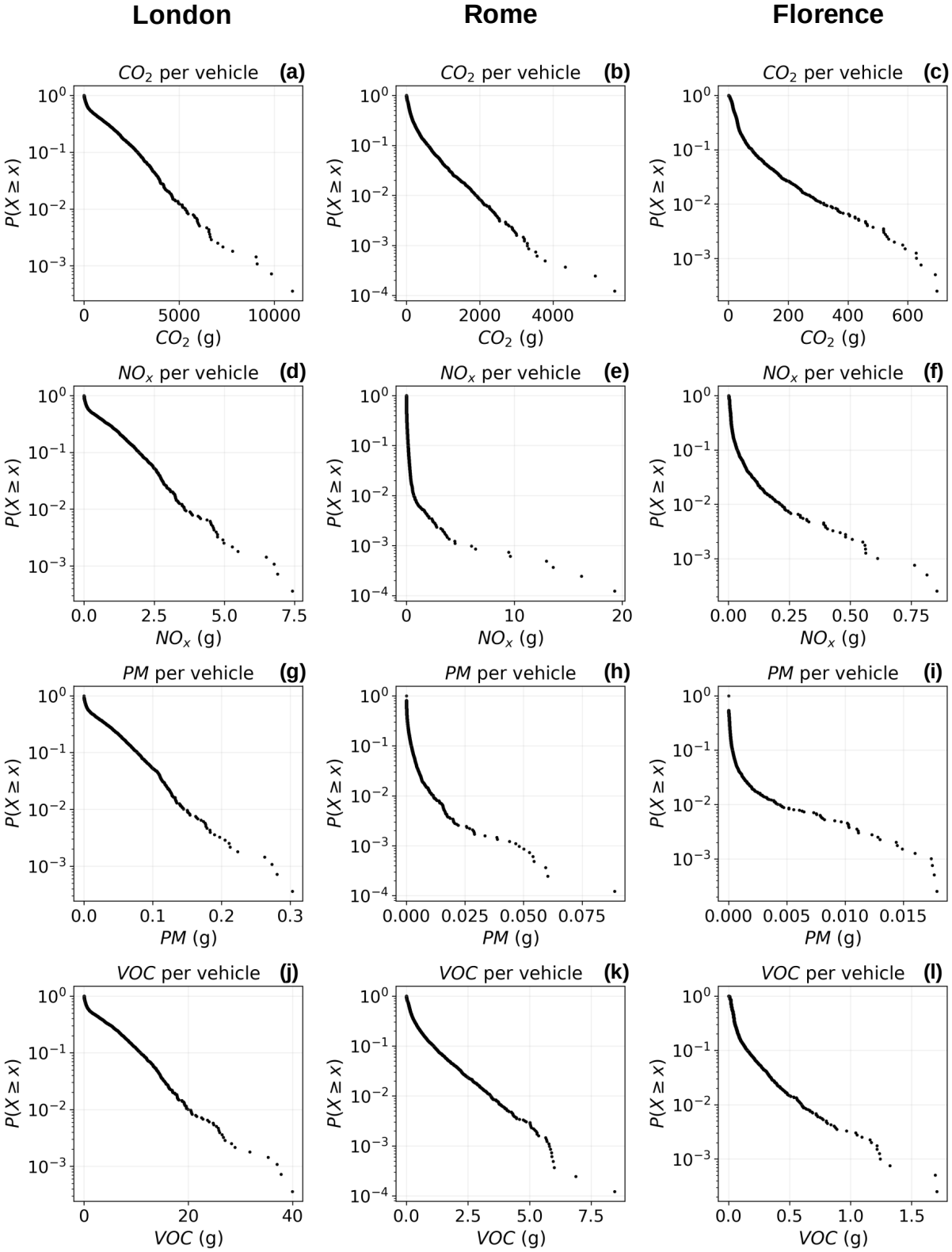

Figure 4: **Empirical Complementary Cumulative Distribution Function (CCDF) of the emissions per vehicle**. For each pollutant and city, we show the empirical CCDF of the data. For each quantity x emitted by a vehicle, the CCDF is the probability of getting a vehicle that emitted more than x. The scale is logarithmic in the y axis and linear in the x axis. The shape of the distribution is similar across the cities and air pollutants.

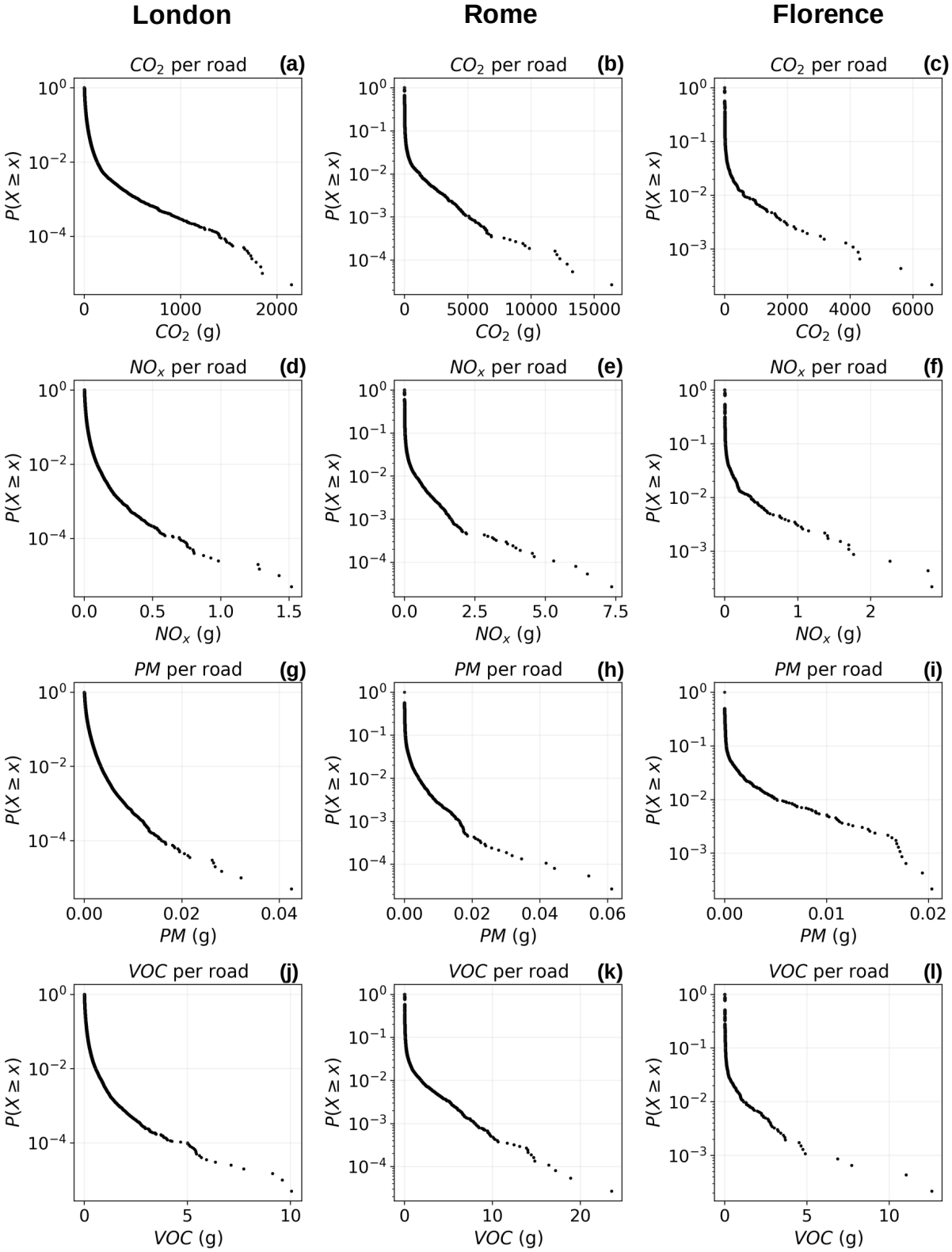

Figure 5: **Empirical Complementary Cumulative Distribution Function (CCDF) of the emissions per road**. For each pollutant and city, we show the empirical CCDF of the data. For each quantity x emitted on a road, the CCDF is the probability of getting a road that has more emissions than x. The scale is logarithmic in the y axis and linear in the x axis. The shape of the distribution is similar across the cities and air pollutants.

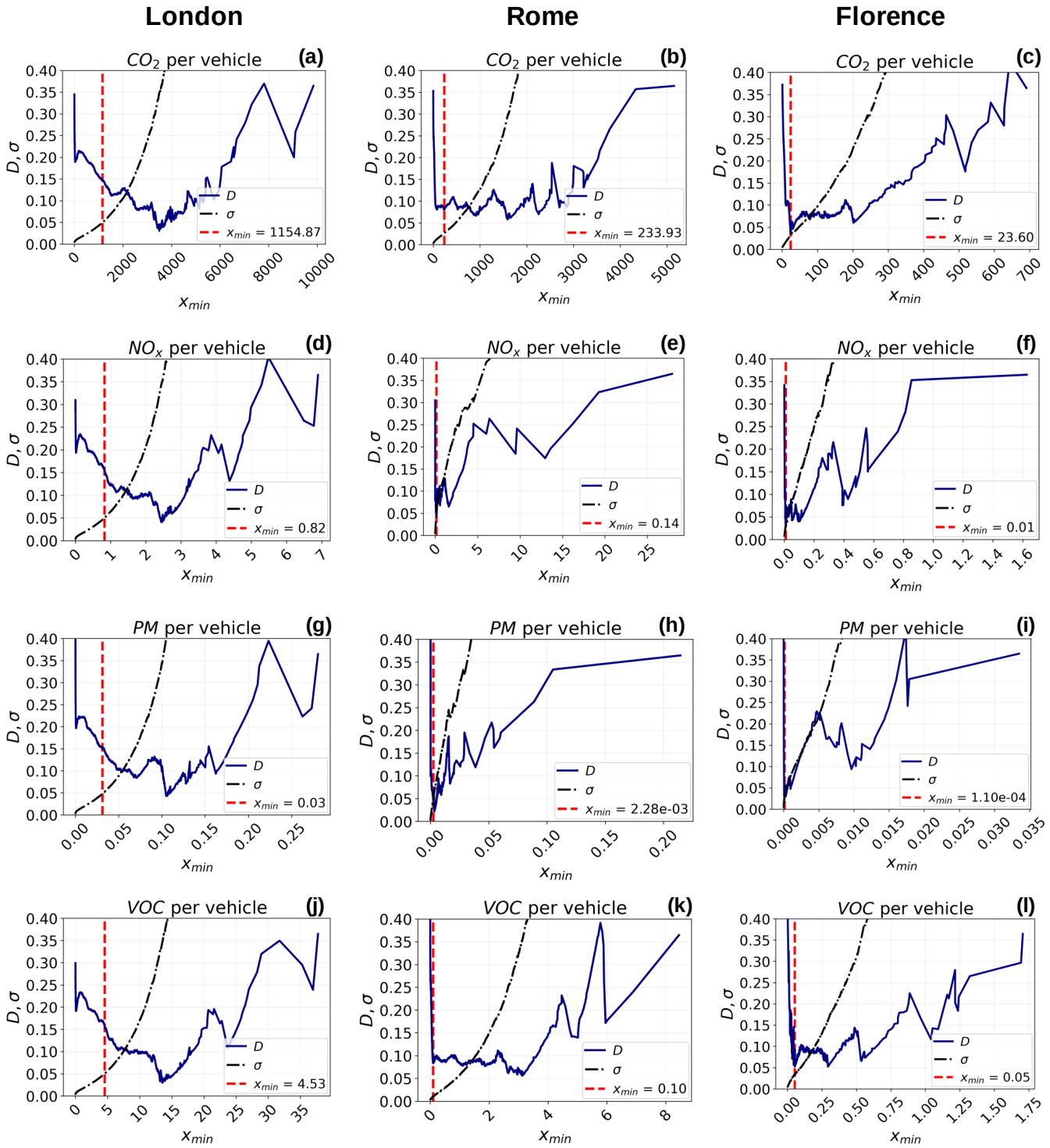

Figure 6: **Choice of x_{min} when fitting the distributions of emissions per vehicle**. For each pollutant and city, we show the values of the Kolmogorov-Smirnov distance D between the data and the power law fit (blue solid line) and the standard error σ of the power law exponent (black dashed line) for each possible choice of x_{min}. The value of x_{min} that minimises D under the constraint $\sigma < 0.05$ is indicated by the red dashed vertical line.

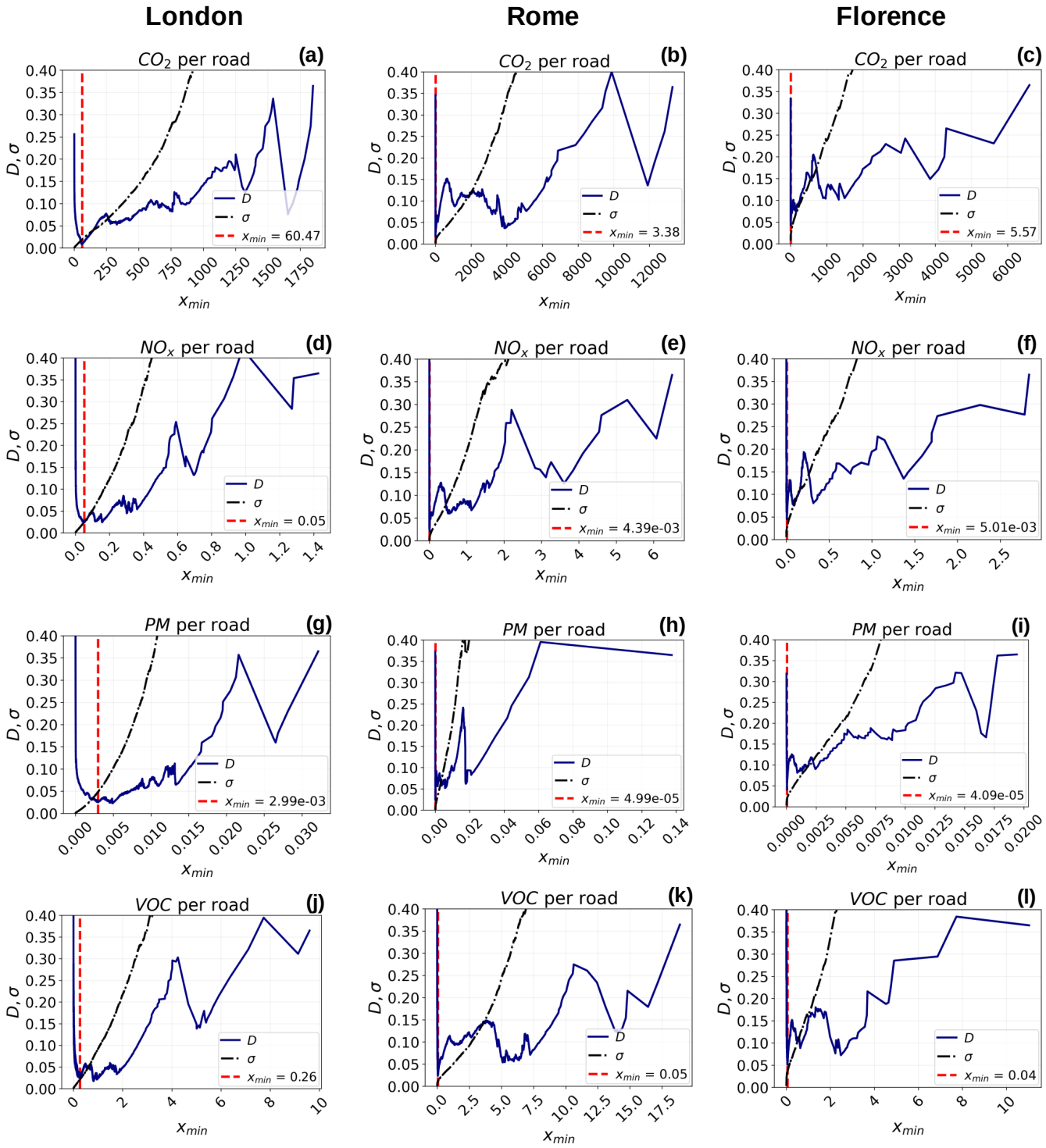

Figure 7: **Choice of x_{min} when fitting the distributions of emissions per road**. For each pollutant and city, we show the values of the Kolmogorov-Smirnov distance D between the data and the power law fit (blue solid line) and the standard error σ of the power law exponent (black dashed line) for each possible choice of x_{min}. The value of x_{min} that minimises D under the constraint $\sigma < 0.05$ is indicated by the red dashed vertical line.

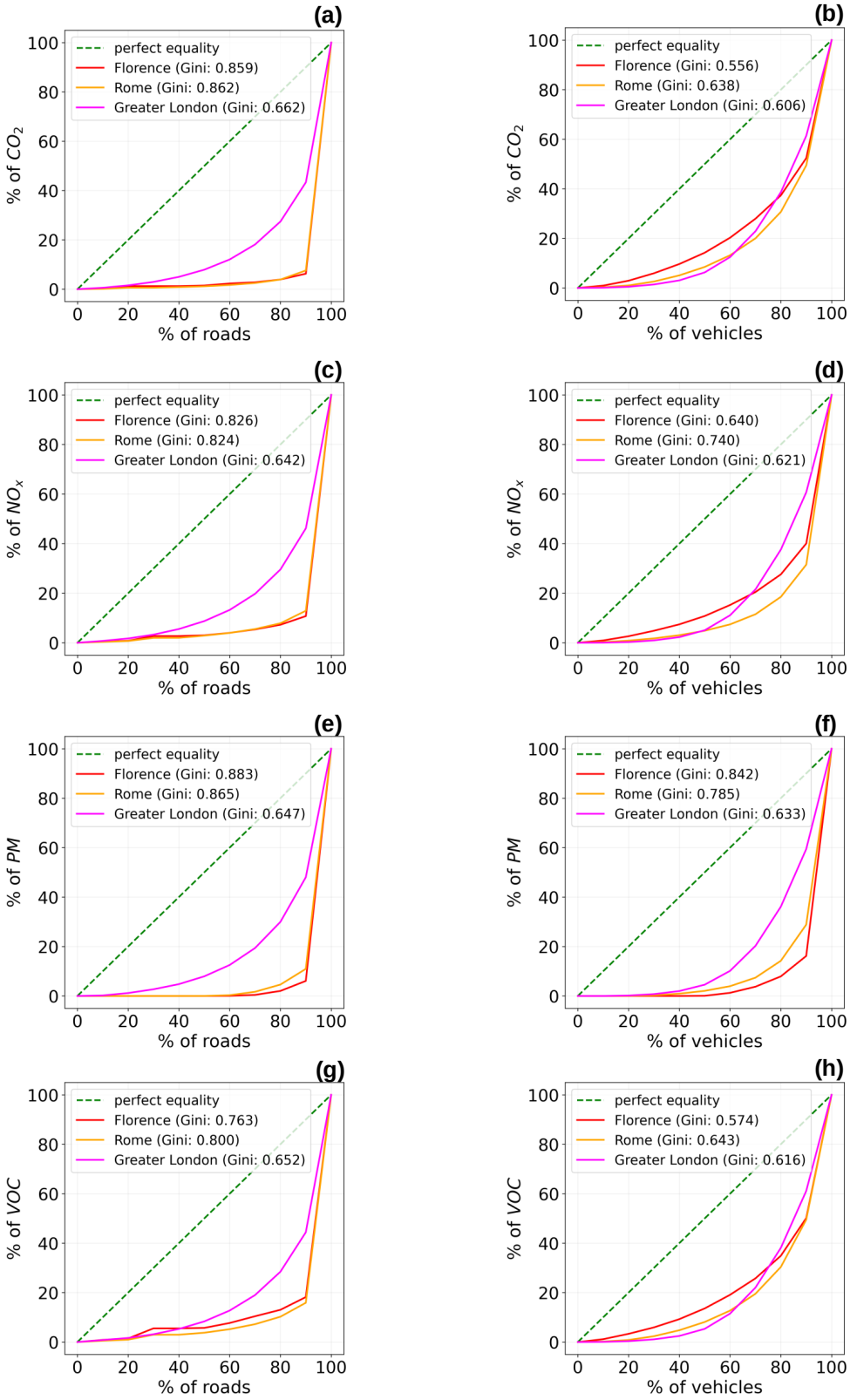

Figure 8: **Distributions of emissions across vehicles and roads**. Lorenz curves showing the share of overall emissions associated respectively to the bottom $x\%$ of the roads (left) and to the bottom $x\%$ of the vehicles (right), for Greater London (violet curve), Rome (orange curve), and Florence (red curve), and for CO_2 (a, b), NO_x (c, d), PM (e, f), and VOC (g, h). The green dashed curve indicates a uniform distribution. We show in the legend the Gini coefficient for each curve.

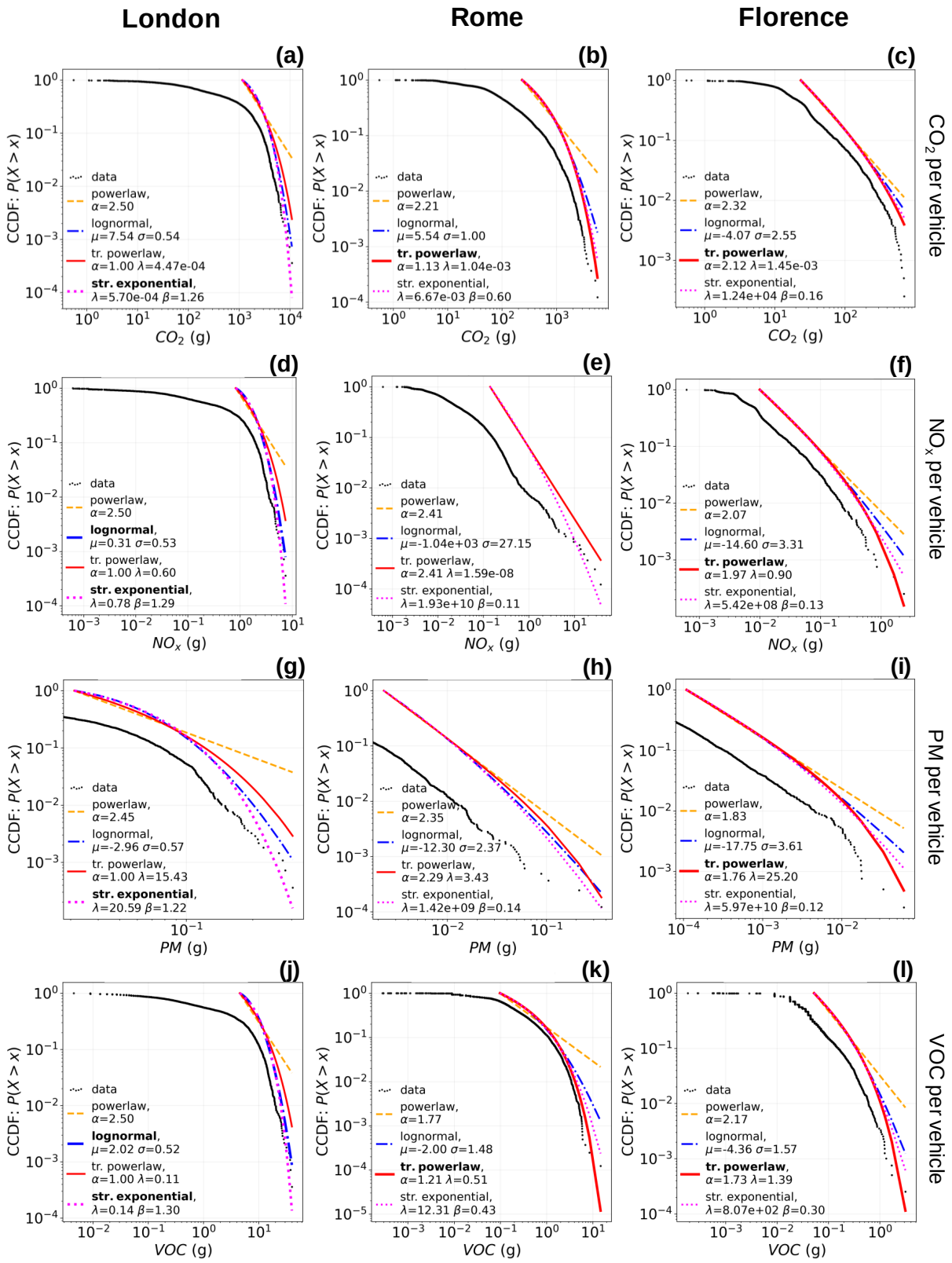

Figure 9: **Fitting of the distributions of emissions per vehicle**. For each pollutant and city, we show the Complementary Cumulative Distribution Function (CCDF) of the emissions per vehicle (black dots), together with the fitted distributions: power law (dashed orange curve), log-normal (dashed blue curve), truncated power law (solid red curve), stretched exponential (dotted purple curve). In the legend, we highlight the best fit(s), if any, in bold. We use a maximum-likelihood fitting method and evaluate the goodness-of-fit using the Kolmogorov-Smirnov distance, comparing different models with a log-likelihood ratio test.

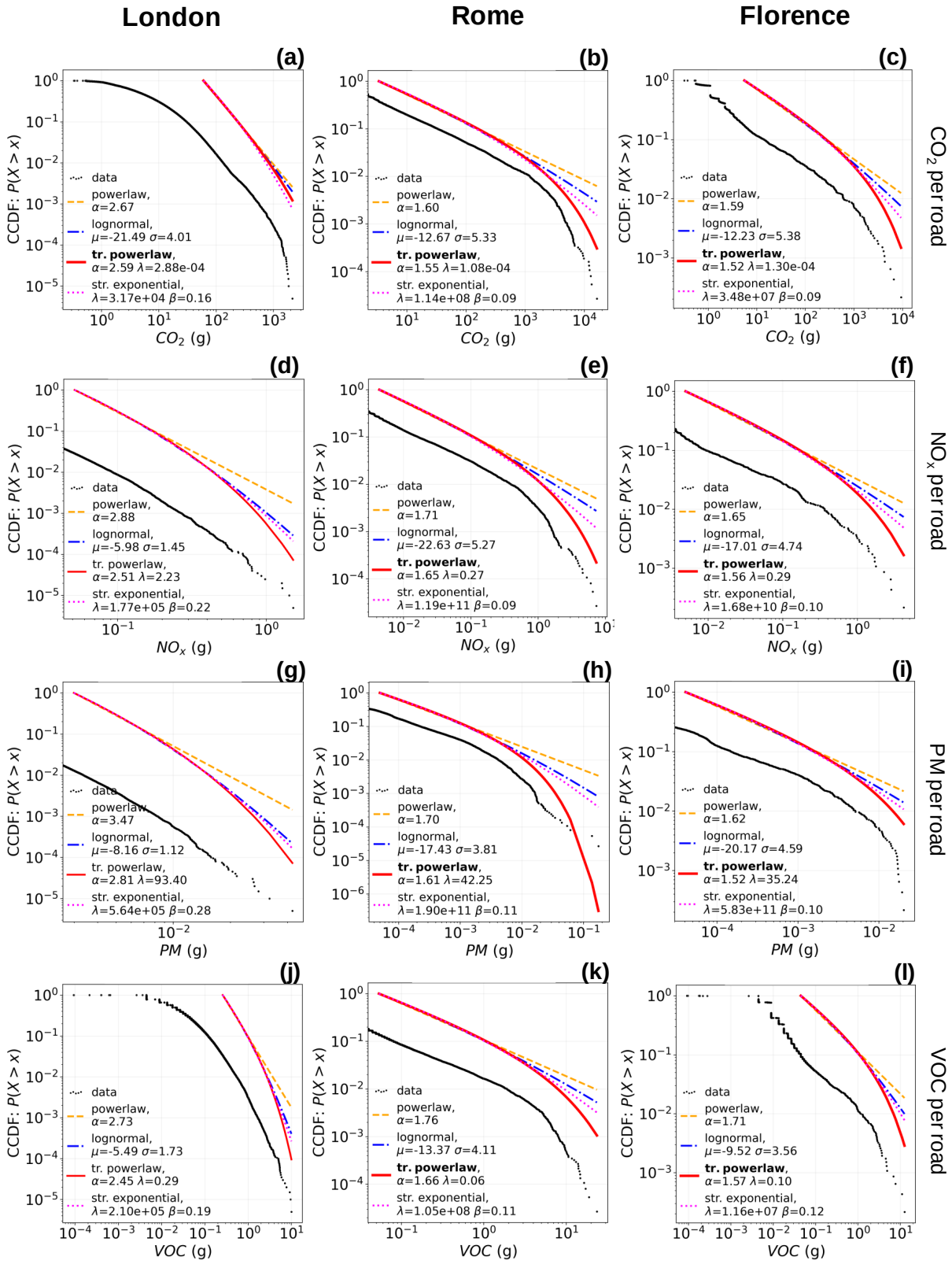

Figure 10: **Fitting of the distributions of emissions per road**. For each pollutant and city, we show the Complementary Cumulative Distribution Function (CCDF) of the emissions per road (black dots), together with the fitted distributions: power law (dashed orange curve), log-normal (dashed blue curve), truncated power law (solid red curve), stretched exponential (dotted purple curve). In the legend, we highlight the best fit(s), if any, in bold. We use a maximum-likelihood fitting method and evaluate the goodness-of-fit using the Kolmogorov-Smirnov distance, comparing different models with a log-likelihood ratio test.

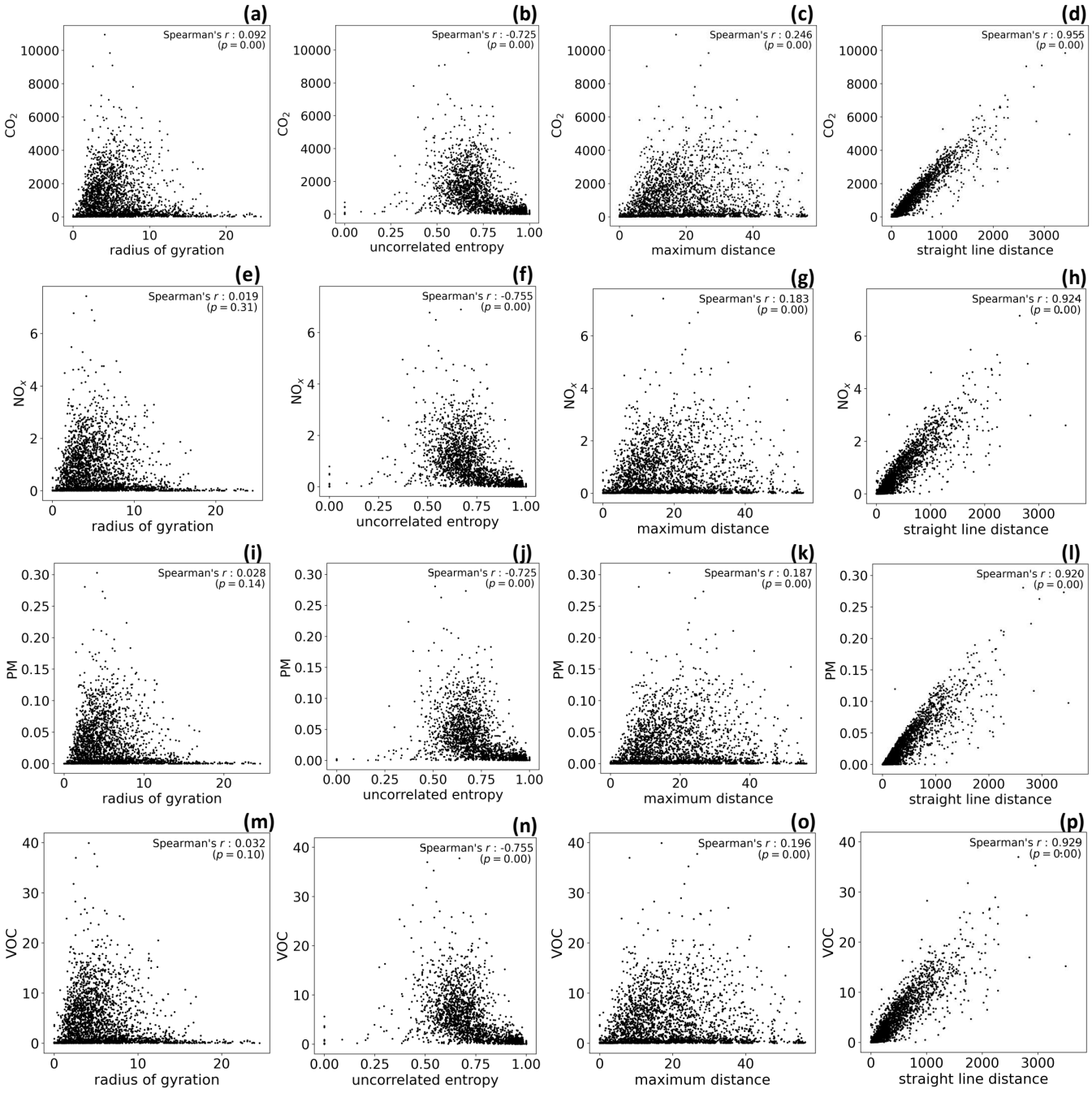

Figure 11: **Correlations between emissions and mobility metrics for London.** The scatter plots show the relationship between the emissions of the four pollutants (rows) and four mobility metrics of the vehicles: the radius of gyration, temporal-uncorrelated entropy, maximum distance and straight line distance. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

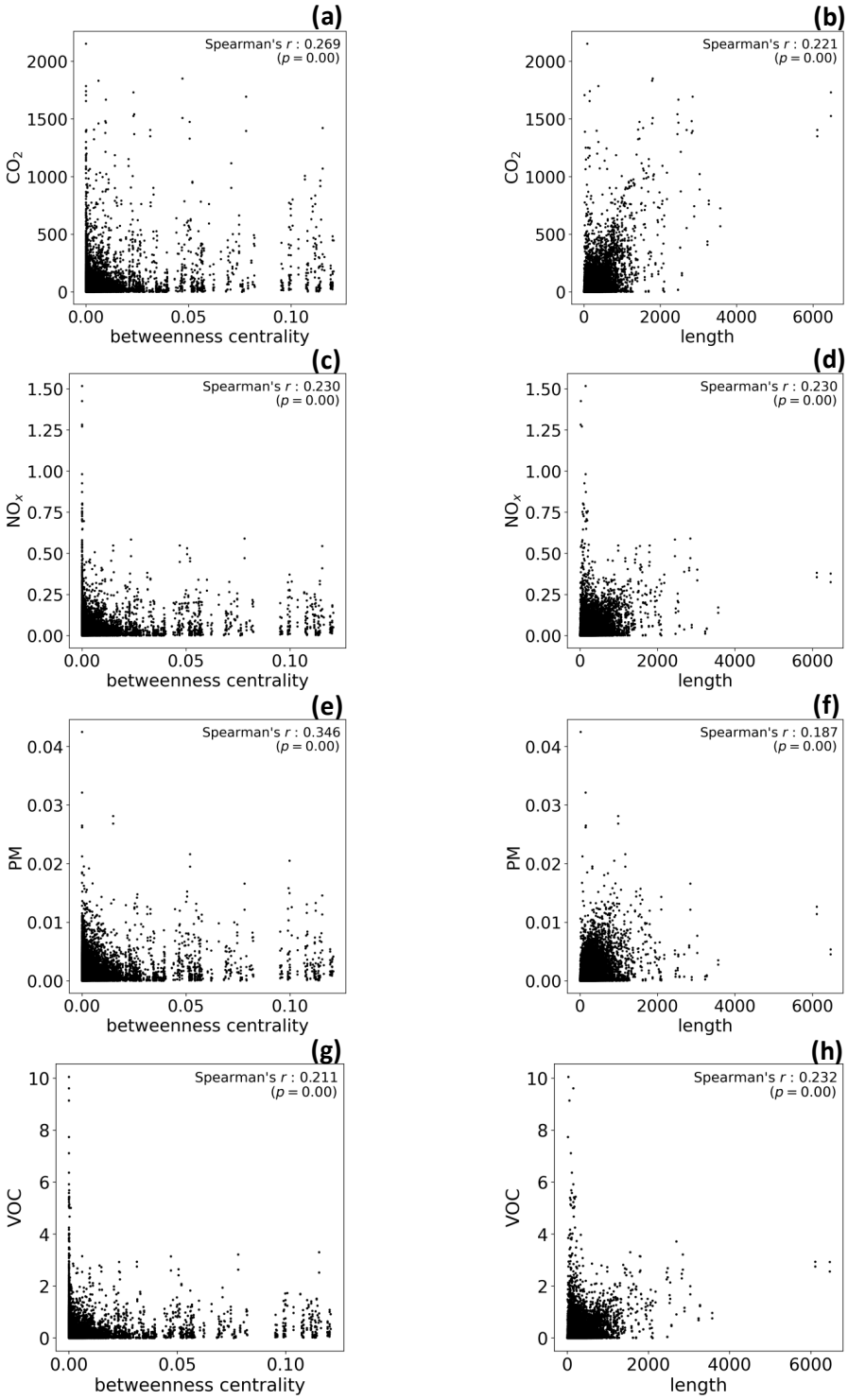

Figure 12: **Correlations between emissions and roads features for London.** The scatter plots show the relation between the emissions of the four pollutants (rows) and two roads' features: the betweenness centrality and length. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

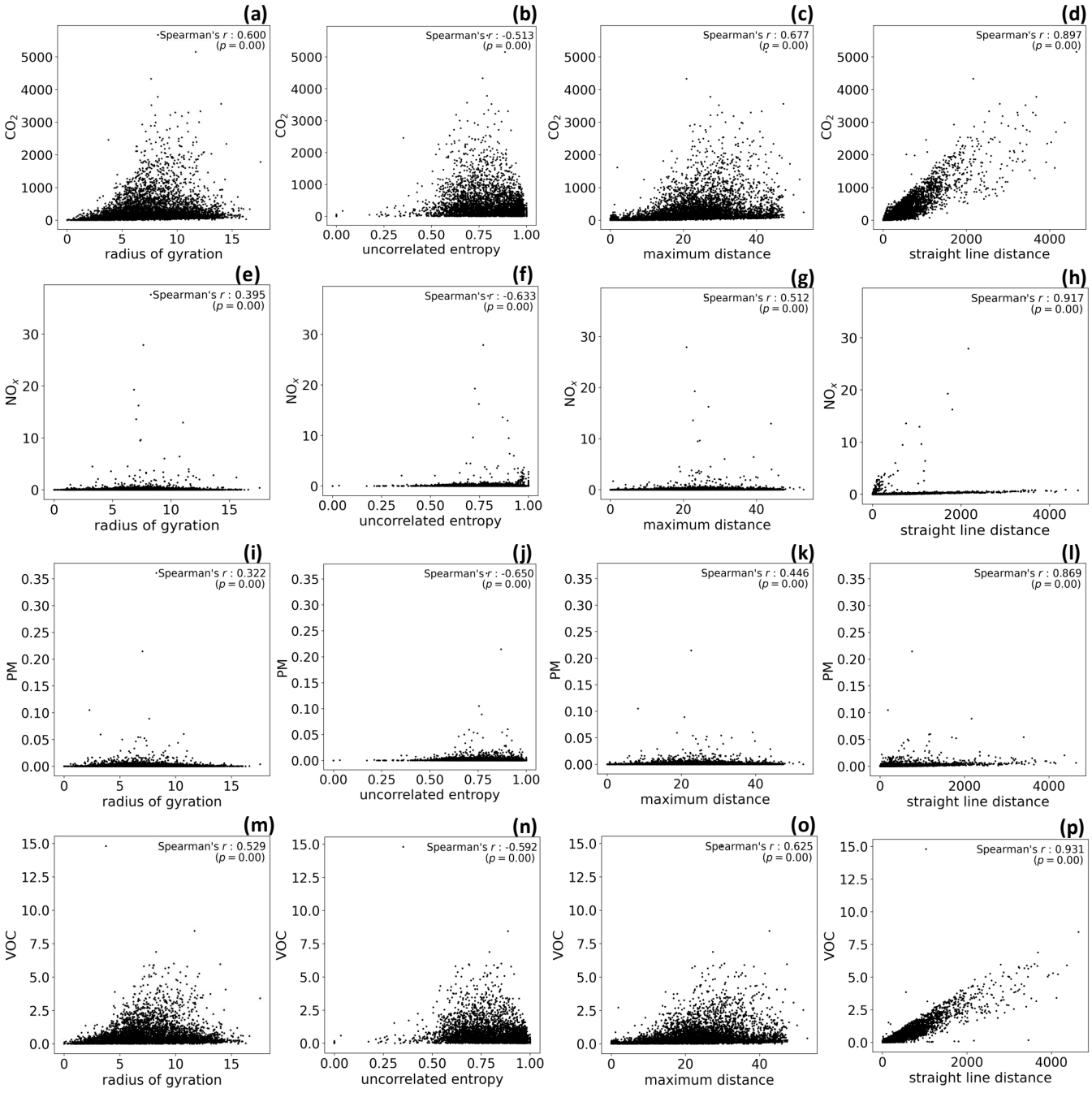

Figure 13: **Correlations between emissions and mobility metrics for Rome.** The scatter plots show the relation between the emissions of the four pollutants (rows) and four mobility metrics of the vehicles: the radius of gyration, temporal-uncorrelated entropy, maximum distance and straight line distance. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

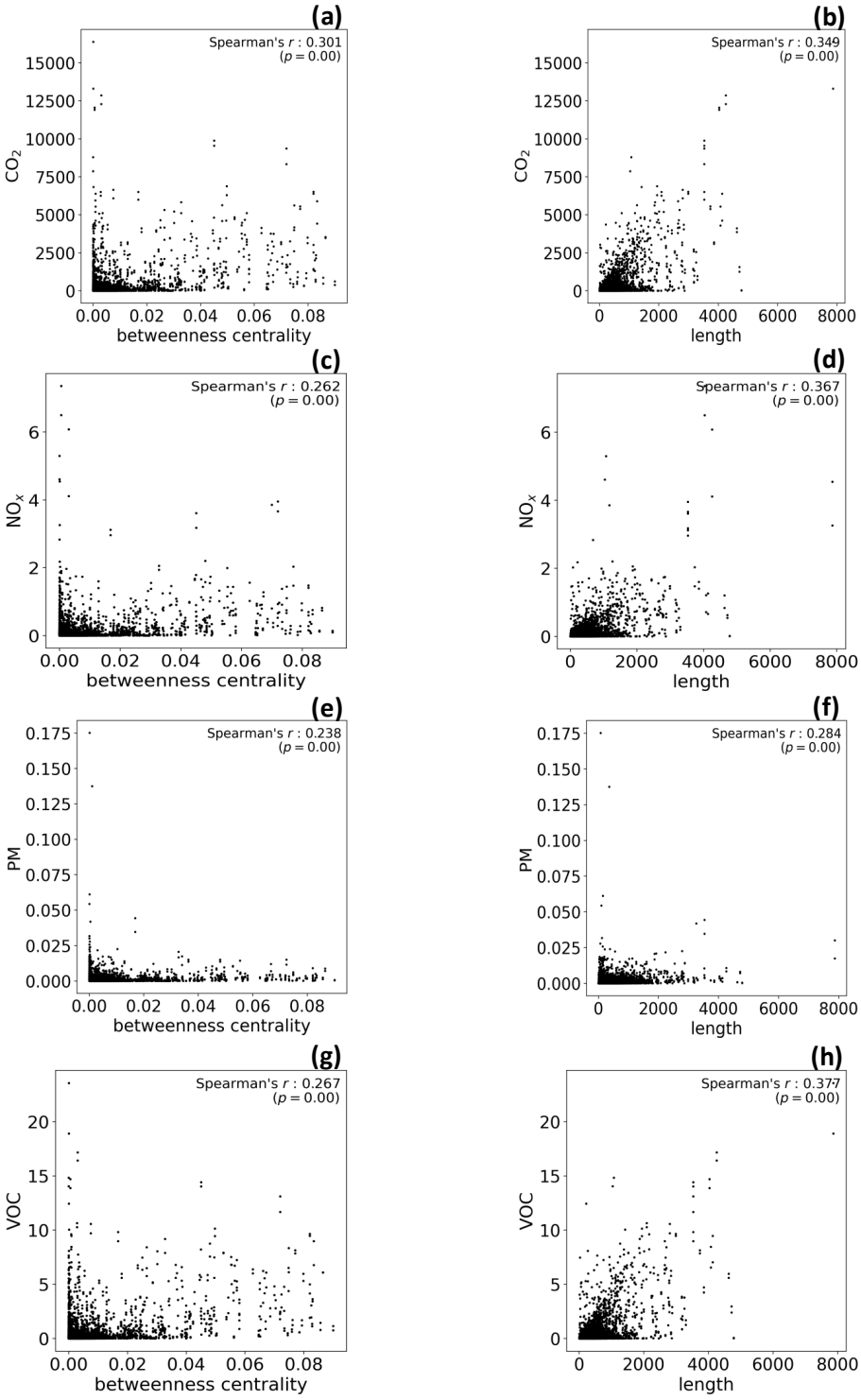

Figure 14: **Correlations between emissions and roads features for Rome.** The scatter plots show the relation between the emissions of the four pollutants (rows) and two roads' features: the betweenness centrality and length. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

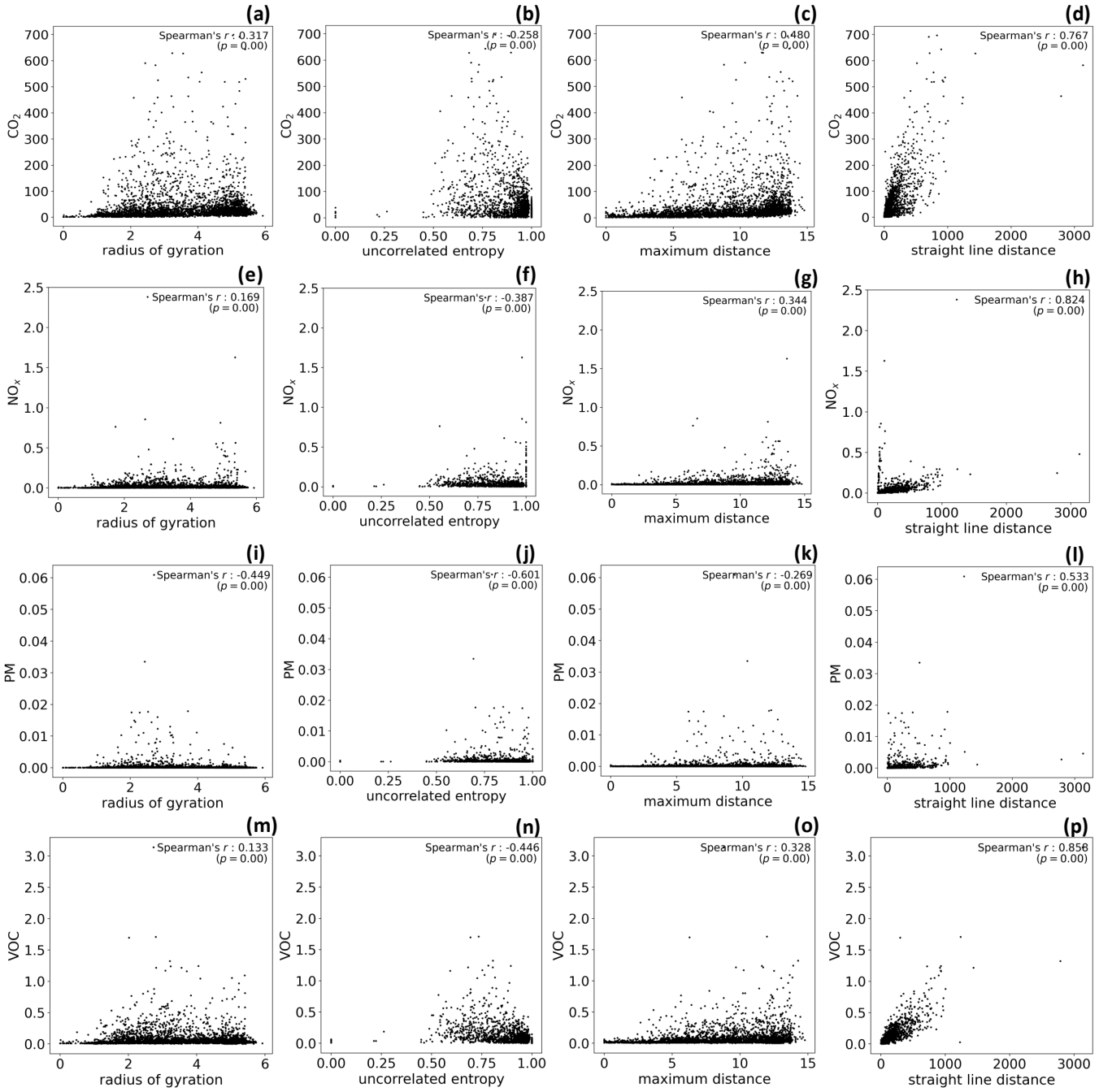

Figure 15: **Correlations between emissions and mobility metrics for Florence.** The scatter plots show the relation between the emissions of the four pollutants (rows) and four mobility metrics of the vehicles: the radius of gyration, temporal-uncorrelated entropy, maximum distance and straight line distance. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

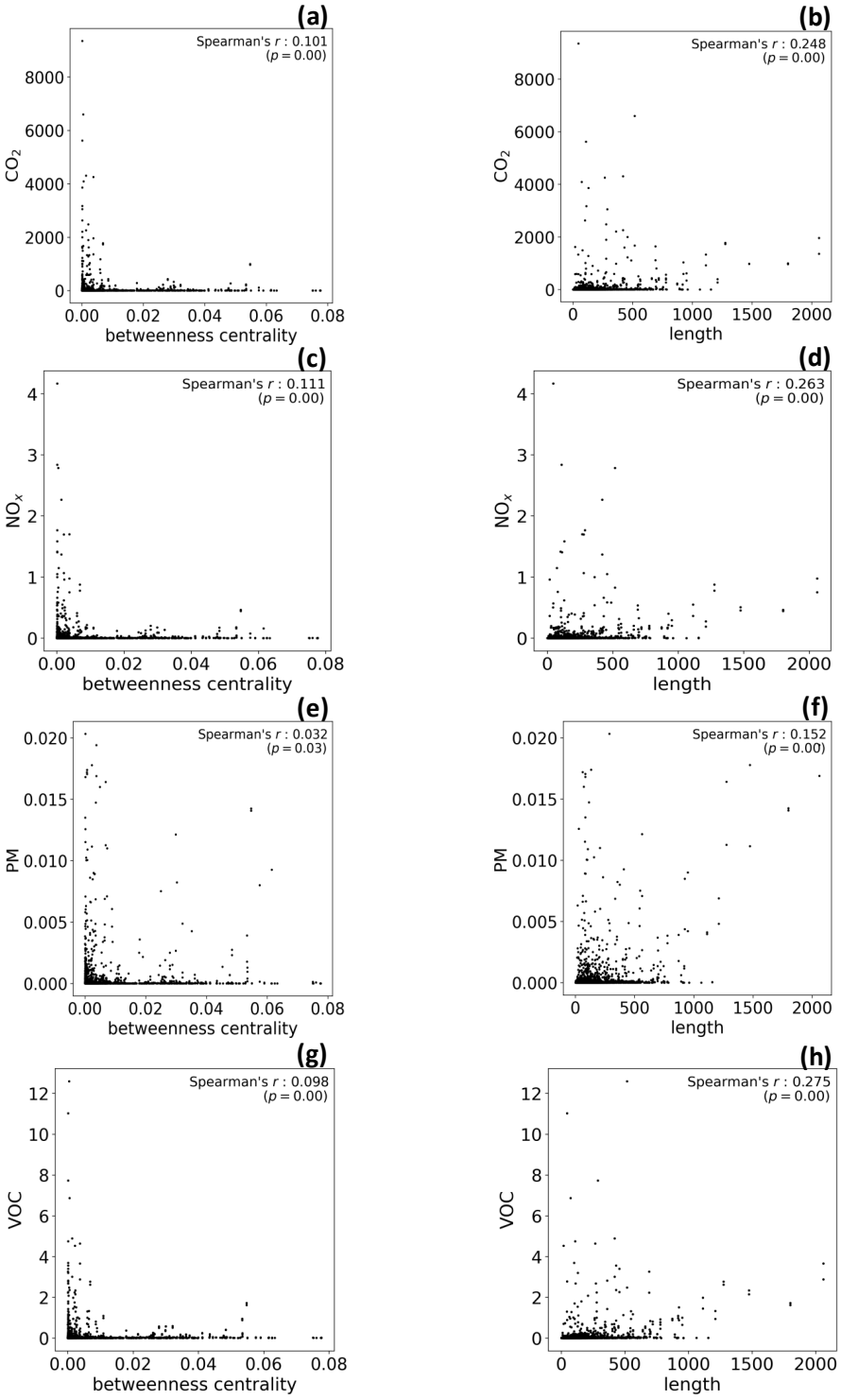

Figure 16: **Correlations between emissions and roads features for Florence.** The scatter plots show the relation between the emissions of the four pollutants (rows) and two roads' features: the betweenness centrality and length. On the upper right corner of each figure we show the Spearman's correlation coefficient, and the corresponding p-value.

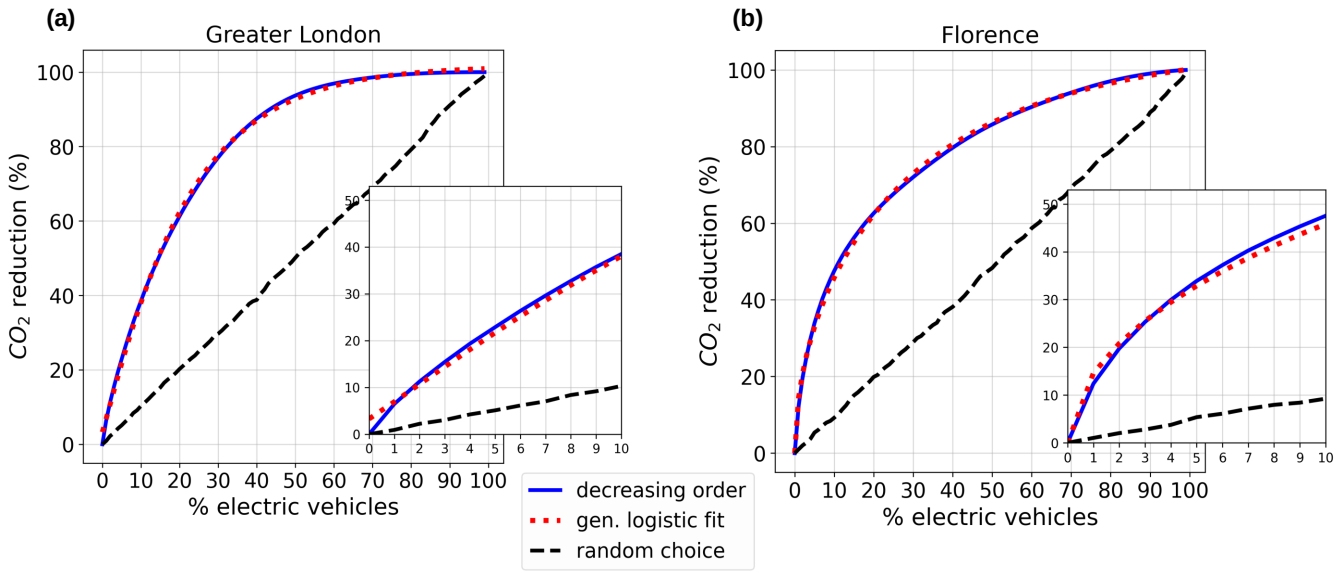

Figure 17: **CO_2 emissions reduction from vehicles electrification in Greater London and Florence**. Percentage reduction of CO_2 emissions corresponding to a certain share (0-100%) of electric vehicles in (a) Greater London and (b) Florence. The inset plots zoom on the first tenth share of electric vehicles. The blue solid line is when the vehicles to be electrified are chosen from the most polluting to the least polluting. The black dashed line is when the vehicles to be electrified are chosen at random. The dotted red line is the *Generalised Logistic Function* (*Richard's curve*) fit: $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$, where α represents the upper asymptote, β the growth range, r the growth rate, and ν the slope of the curve.

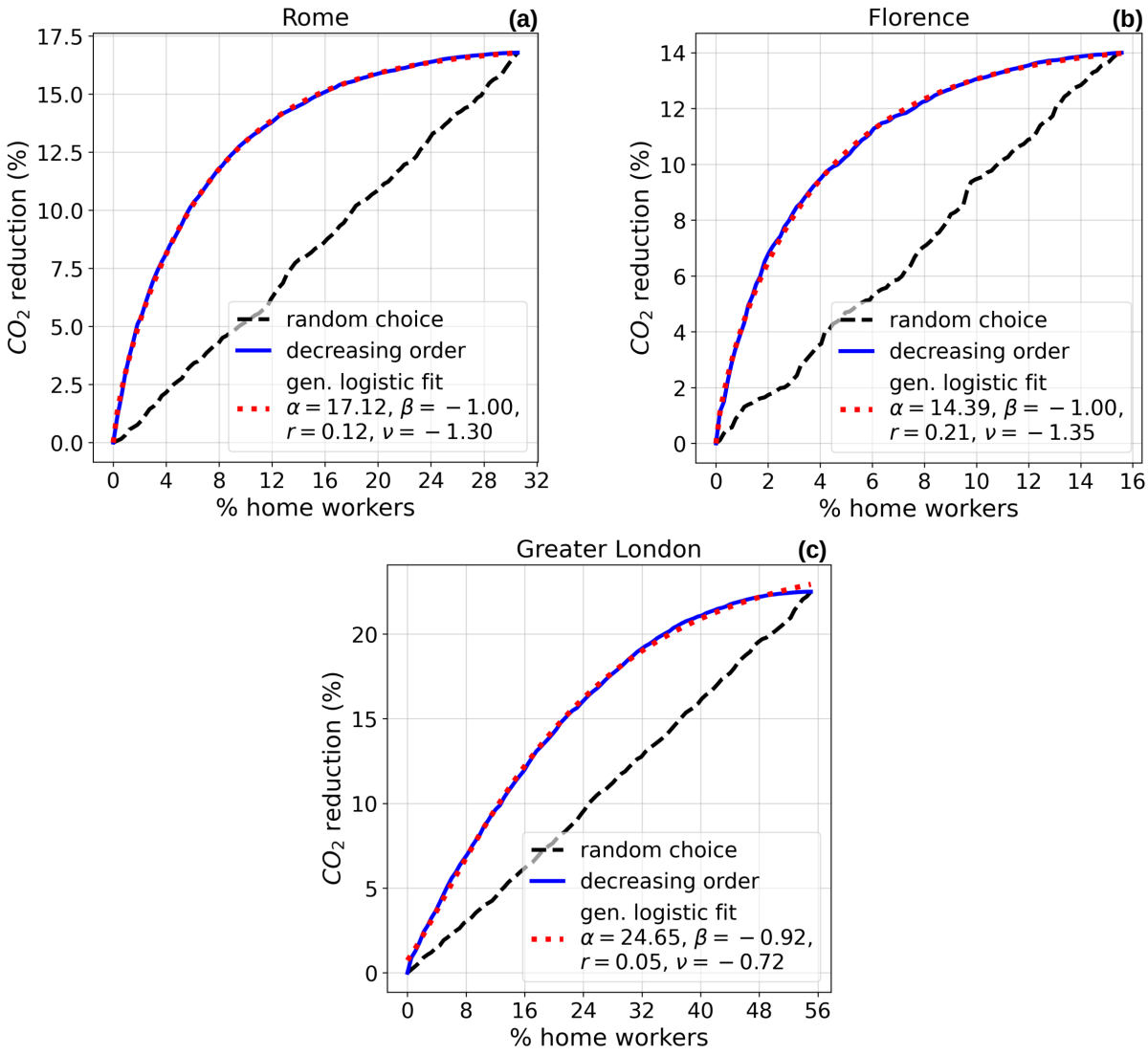

Figure 18: **CO_2 emissions reduction from remote working in Rome, Florence and Greater London**. Percentage reduction of CO_2 emissions corresponding to a certain share of vehicles that stop commuting in (a) Rome, (b) Florence, and (c) Greater London. The maximum share of vehicles that stop commuting in our simulation changes with the city: it is 31% for Rome, 16% for Florence, and 55% for London, as these are the percentages of vehicles for which we can identify both the home and work locations and, thus, the commuting trajectories. The blue solid line is when the vehicles that stop commuting are chosen from the most polluting to the least polluting. The black dashed line is when the vehicles that stop commuting are chosen at random. The dotted red line is the *Generalised Logistic Function* (*Richard's curve*) fit: $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$, where α represents the upper asymptote, β the growth range, r the growth rate, and ν the slope of the curve. We show the estimated parameters in the legend.

C Supplementary Tables

	power law	**log-normal**	**truncated power law**	**stretched exponential**	**exponential**
power law	-	None	None	power law (R=2.59, p=0.01)	power law (R=5.28, p=0.00)
log-normal	-	-	None	log-normal (R=2.56, p=0.01)	log-normal (R=5.27, p=0.00)
truncated power law	-	-	-	tr. power law (R=2.59, p=0.01)	tr. power law (R=5.28, p=0.00)
stretched exponential	-	-	-	-	str. exponential (R=5.34, p=0.00)

Table 5: **Results of log-likelihood ratio tests for comparing different models fitting the distribution of NO_x emissions per vehicle in Rome.** Each (i, j) cell of the table shows the result of a log-likelihood ratio test used for comparing the goodness-of-fit of the model in row i and the one in column j. If the test gives result in favour of a model, its name is shown together with the statistic R, that is negative if the result is in favour of the first model (the one on the row) or positive if in favour of the second (the one on the column), and the corresponding p-value. If the test gives no evidence in favour of one of the two models, the result is signed as None. In this case, there is no model that wins more comparisons than the others.

	power law	**log-normal**	**truncated power law**	**stretched exponential**	**exponential**
power law	-	None	None	None	power law (R=3.21, p=0.00)
log-normal	-	-	None	None	log-normal (R=3.32, p=0.00)
truncated power law	-	-	-	None	tr. power law (R=3.27, p=0.00)
stretched exponential	-	-	-	-	str. exponential (R=3.34, p=0.00)

Table 6: **Results of log-likelihood ratio tests for comparing different models fitting the distribution of PM emissions per vehicle in Rome.** Each (i, j) cell of the table shows the result of a log-likelihood ratio test used for comparing the goodness-of-fit of the model in row i and the one in column j. If the test gives result in favour of a model, its name is shown together with the statistic R, that is negative if the result is in favour of the first model (the one on the row) or positive if in favour of the second (the one on the column), and the corresponding p-value. If the test gives no evidence in favour of one of the two models, the result is signed as None. In this case, there is no model that wins more comparisons than the others.

	power law	**log-normal**	**truncated power law**	**stretched exponential**	**exponential**
power law	-	log-normal (R=-4.32, p=0.00)	tr. power law (R=-4.40, p=0.00)	str. exponential (R=-4.27, p=0.00)	power law (R=7.99, p=0.00)
log-normal	-	-	None	None	log-normal (R=9.48, p=0.00)
truncated power law	-	-	-	None	tr. power law (R=9.54, p=0.00)
stretched exponential	-	-	-	-	str. exponential (R=9.56, p=0.00)

Table 7: **Results of log-likelihood ratio tests for comparing different models fitting the distribution of NO_x emissions per road in London.** Each (i, j) cell of the table shows the result of a log-likelihood ratio test used for comparing the goodness-of-fit of the model in row i and the one in column j. If the test gives result in favour of a model, its name is shown together with the statistic R, that is negative if the result is in favour of the first model (the one on the row) or positive if in favour of the second (the one on the column), and the corresponding p-value. If the test gives no evidence in favour of one of the two models, the result is signed as None. In this case, none of the log-normal, truncated power law and stretched exponential models wins more comparisons than the other two.

	power law	**log-normal**	**truncated power law**	**stretched exponential**	**exponential**
power law	-	log-normal (R=-2.87, p=0.00)	tr. power law (R=-2.87, p=0.00)	str. exponential (R=-2.89, p=0.00)	power law (R=3.70, p=0.00)
log-normal	-	-	None	None	log-normal (R=5.06, p=0.00)
truncated power law	-	-	-	None	tr. power law (R=5.10, p=0.00)
stretched exponential	-	-	-	-	str. exponential (R=5.09, p=0.00)

Table 8: **Results of log-likelihood ratio tests for comparing different models fitting the distribution of PM emissions per road in London.** Each (i, j) cell of the table shows the result of a log-likelihood ratio test used for comparing the goodness-of-fit of the model in row i and the one in column j. If the test gives result in favour of a model, its name is shown together with the statistic R, that is negative if the result is in favour of the first model (the one on the row) or positive if in favour of the second (the one on the column), and the corresponding p-value. If the test gives no evidence in favour of one of the two models, the result is signed as None. In this case, none of the log-normal, truncated power law and stretched exponential models wins more comparisons than the other two.

	power law	**log-normal**	**truncated power law**	**stretched exponential**	**exponential**
power law	-	log-normal (R=-4.10, p=0.00)	tr. power law (R=-4.56, p=0.00)	str. exponential (R=-3.69, p=0.00)	power law (R=10.11, p=0.00)
log-normal	-	-	None	None	log-normal (R=11.29, p=0.00)
truncated power law	-	-	-	None	tr. power law (R=11.47, p=0.00)
stretched exponential	-	-	-	-	str. exponential (R=11.46, p=0.00)

Table 9: **Results of log-likelihood ratio tests for comparing different models fitting the distribution of VOC emissions per road in London.** Each (i, j) cell of the table shows the result of a log-likelihood ratio test used for comparing the goodness-of-fit of the model in row i and the one in column j. If the test gives result in favour of a model, its name is shown together with the statistic R, that is negative if the result is in favour of the first model (the one on the row) or positive if in favour of the second (the one on the column), and the corresponding p-value. If the test gives no evidence in favour of one of the two models, the result is signed as None. In this case, none of the log-normal, truncated power law and stretched exponential models wins more comparisons than the other two.

	mobility metrics				**roads' features**	
	radius	entropy	max dist.	str. line dist.	betweenness centrality	length
London	0.02*	-0.75	0.18	0.92	0.23	0.23
Rome	0.39	-0.63	0.51	0.92	0.26	0.37
Florence	0.17	-0.39	0.34	0.82	0.11	0.26

Table 10: **Correlations between vehicles' emissions of NO_x, mobility metrics, and road features.** Spearman's correlation coefficients between (left) NO_x emissions per vehicle and vehicles' mobility metrics (their radius of gyration, uncorrelated entropy, maximum distance and distance travelled straight line), and (right) NO_x emissions per road and roads' features (betweenness centrality and length). *p-value > 0.05.

	mobility metrics				**roads' features**	
	radius	entropy	max dist.	str. line dist.	betweenness centrality	length
London	0.03*	-0.72	0.19	0.92	0.35	0.19
Rome	0.32	-0.65	0.45	0.87	0.24	0.28
Florence	-0.45	-0.60	-0.27	0.53	0.03	0.15

Table 11: **Correlations between vehicles' emissions of PM, mobility metrics, and road features.** Spearman's correlation coefficients between (left) PM emissions per vehicle and vehicles' mobility metrics (their radius of gyration, uncorrelated entropy, maximum distance and distance travelled straight line), and (right) PM emissions per road and roads' features (betweenness centrality and length). *p-value > 0.05.

	mobility metrics				**roads' features**	
	radius	entropy	max dist.	str. line dist.	betweenness centrality	length
London	0.03*	-0.75	0.20	0.93	0.21	0.23
Rome	0.53	-0.59	0.62	0.93	0.27	0.38
Florence	0.13	-0.45	0.33	0.86	0.10	0.27

Table 12: **Correlations between vehicles' emissions of VOC, mobility metrics, and road features.** Spearman's correlation coefficients between (left) VOC emissions per vehicle and vehicles' mobility metrics (their radius of gyration, uncorrelated entropy, maximum distance and distance travelled straight line), and (right) VOC emissions per road and roads' features (betweenness centrality and length). *p-value > 0.05.

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{r}$	$\boldsymbol{\nu}$	$\boldsymbol{R^2}$
London	101.66	-0.95	5.05×10^{-2}	-0.86	0.99
Rome	101.24	-0.99	3.96×10^{-2}	-1.56	0.99
Florence	106.57	-0.99	2.20×10^{-2}	-1.91	0.99
2^{nd} Municipality (Rome)	99.8	-0.99	4.84×10^{-2}	-1.55	0.99

Table 13: **Estimated parameters of a *Generalised Logistic Function* fitted to the reduction of the overall CO_2 emissions resulting from the electrification of the vehicles starting from the most polluting ones.** The estimated parameters α (the upper asymptote), β (the growth range), r (the growth rate), and ν (the slope) obtained with non-linear least squares fitting of a *Generalised Logistic Function* to the data are shown for London, Rome, Florence and a neighbourhood of Rome (the Second Municipality). We also show the R^2 for each model as a measure of its goodness of fit to the data.

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{r}$	$\boldsymbol{\nu}$	$\boldsymbol{R^2}$
London	22.49	-0.92	0.05	-0.72	0.99
Rome	17.12	-1.00	0.12	-1.30	0.99
Florence	14.39	-1.00	0.21	-1.35	0.99

Table 14: **Estimated parameters of a *Generalised Logistic Function* fitted to the reduction of the overall CO_2 emissions resulting from the shift to remote working of the drivers of the vehicles starting from the most polluting ones.** The estimated parameters α (the upper asymptote), β (the growth range), r (the growth rate), and ν (the slope) obtained with non-linear least squares fitting of a *Generalised Logistic Function* to the data are shown for London, Rome, and Florence. We also show the R^2 for each model as a measure of its goodness of fit to the data.

	fuel type	acceleration	f_1	f_2	f_3	f_4	f_5	f_6
CO_2	petrol	all	5.53×10^{-1}	1.61×10^{-1}	-2.89×10^{-3}	2.66×10^{-1}	5.11×10^{-1}	1.83×10^{-1}
	diesel	all	3.24×10^{-1}	8.59×10^{-2}	4.96×10^{-3}	-5.86×10^{-2}	4.48×10^{-1}	2.30×10^{-1}
	LPG	all	6.00×10^{-1}	2.19×10^{-1}	-7.74×10^{-3}	3.57×10^{-1}	5.14×10^{-1}	1.70×10^{-1}
NO_x	petrol	$\geq -0.5\, m/s^2$	6.19×10^{-4}	8.00×10^{-5}	-4.03×10^{-6}	-4.13×10^{-4}	3.80×10^{-4}	1.77×10^{-4}
	petrol	$< -0.05\, m/s^2$	2.17×10^{-4}	0	0	0	0	0
	diesel	$\geq -0.5\, m/s^2$	2.41×10^{-3}	-4.11×10^{-4}	6.73×10^{-5}	-3.07×10^{-3}	2.14×10^{-3}	1.50×10^{-3}
	diesel	$< -0.05\, m/s^2$	1.68×10^{-3}	-6.62×10^{-5}	9.00×10^{-6}	2.50×10^{-4}	2.91×10^{-4}	1.20×10^{-4}
	LPG	$\geq -0.5\, m/s^2$	8.92×10^{-4}	1.61×10^{-5}	-8.06×10^{-7}	-8.23×10^{-5}	7.60×10^{-5}	3.54×10^{-5}
	LPG	$< -0.05\, m/s^2$	3.43×10^{-4}	0	0	0	0	0
PM	petrol	all	0	1.57×10^{-5}	-9.21×10^{-7}	0	3.75×10^{-5}	1.89×10^{-5}
	diesel	all	0	3.13×10^{-4}	-1.84×10^{-5}	0	7.50×10^{-4}	3.78×10^{-4}
	LPG	all	0	1.57×10^{-5}	-9.21×10^{-7}	0	3.75×10^{-5}	1.89×10^{-5}
VOC	petrol	$\geq -0.5\, m/s^2$	4.47×10^{-3}	7.32×10^{-7}	-2.87×10^{-8}	-3.41×10^{-5}	4.94×10^{-6}	1.66×10^{-6}
	petrol	$< -0.05\, m/s^2$	2.63×10^{-3}	0	0	0	0	0
	diesel	$\geq -0.5\, m/s^2$	9.22×10^{-5}	9.09×10^{-6}	-2.29×10^{-7}	-2.20×10^{-5}	1.69×10^{-5}	3.75×10^{-6}
	diesel	$< -0.05\, m/s^2$	5.25×10^{-5}	7.22×10^{-6}	-1.87×10^{-7}	0	-1.02×10^{-5}	-4.22×10^{-6}
	LPG	$\geq -0.5\, m/s^2$	1.44×10^{-2}	1.74×10^{-7}	-6.82×10^{-9}	-8.11×10^{-7}	1.18×10^{-6}	3.96×10^{-7}
	LPG	$< -0.05\, m/s^2$	8.42×10^{-3}	0	0	0	0	0

Table 15: **Emission functions**. The emission functions for each pollutant, fuel type, and acceleration profile.