Gross polluters and vehicle emissions reduction

Matteo Böhm [1,2 ✉], Mirco Nanni [2] and Luca Pappalardo [2 ✉]

Vehicle emissions produce an important share of a city's air pollution, with a substantial impact on the environment and human health. Traditional emission estimation methods use remote sensing stations, missing the full driving cycle of vehicles, or focus on a few vehicles. We have used GPS traces and a microscopic model to analyse the emissions of four air pollutants from thousands of private vehicles in three European cities. We found that the emissions across the vehicles and roads are well approximated by heavy-tailed distributions and thus discovered 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 far more effective than those limiting circulation based on an uninformed choice of vehicles. Our study contributes to shaping the discussion on how to measure emissions with digital data.

Estimating the distribution of air pollutants over space and time is a crucial challenge concerning climate change and human health. In urban environments, the air pollution generated by vehicle emissions has become increasingly evident, to the point that the temporary interruption of regular traffic during the COVID-19 lockdowns resulted in a tremendous decrease in CO_2 emission[1-4]. Even if this brief period's impact on the epochal challenge of climate change is negligible[5], it helps to outline the impact of emissions related to transportation on our everyday life. Greenhouse gas (GHG) emissions from this sector have doubled since 1970, and, in 2016, 11.9% of global GHG emissions were from road transport (60% of which was from passenger travel)[6,7]. Moreover, the transport sector emits non-CO_2 pollutants such as nitrogen oxides, ozone, particulate matter and volatile organic compounds, which play a fundamental role in changing climate and are dangerous for human health[6]. Among the Sustainable Development Goals to be reached by 2030 (ref. [8]), the United Nations posed an urgent call for action to reduce "the adverse per capita environmental impact of cities, including by paying particular attention to air quality"[8]. In this regard, measuring vehicle emissions is primary to designing policies to reduce transportation emissions.

Based on the available data, existing methods to quantify vehicle emissions range between two extremes. On the one hand, some approaches rely on measurements performed on small samples of vehicles (usually less than ten) but with high spatiotemporal resolution, such as those coming from particulate sensors[9] or portable emissions measurement systems (PEMS)[10,11]. These sensors measure emissions in real-world driving conditions, producing accurate estimates, but they are hardly generalizable patterns due to the limited sample size. For example, two studies[10,11] analysed emissions from PEMS of one and three light-duty vehicles, finding that the highest emissions are associated with the urban part of their routes, flat roads and low speed.

On the other hand, some studies cover a region's almost entire fleet, for example, using odometer readings obtained from annual safety inspections. These inspections provide data on the age, fuel type, engine volume as well as the distance travelled for each vehicle, and are used in macroscopic models to estimate annual emissions. For example, two studies[12,13] used odometer readings to compute mean annual emissions for UK postcode areas and explored the built-environment effects (for example, work accessibility) on the annual kilometres travelled by vehicles in Boston. Unfortunately, odometer readings miss critical information such as instantaneous speed and acceleration[14-17], making it challenging to track emissions over time and map them to suburban areas.

Global Positioning System (GPS) traces generated by in-vehicle devices stand as a trade-off between these two extremes. Depending on the provider's market penetration, they can cover a representative fraction of the vehicle fleet[18] and allow the instantaneous speed and acceleration to be computed, which are then used within microscopic models to obtain emissions estimates with high spatiotemporal resolution. GPS traces describe human mobility in great detail[19-22] and offer an unprecedented tool to implement strategies such as reducing congestion, improving vehicle efficiency and shifting to lower-carbon options[23-29]. Given these peculiarities, several studies used GPS traces to analyse vehicle emissions at different spatiotemporal scales[30,31], investigate the relationship between emissions and the urban environment[32], vehicle kilometres travelled and fuel consumption[33], or trip rates and travel mode choice[34]. Other studies concentrated on congestion-related emissions[35] or braking[36], emissions associated with ride-hailing[37] and bus stop positioning[38], the impact of urban policies[39], methods for emission modelling[40,41] and air quality monitoring[9].

Despite this variety of literature, it remains unclear what statistical patterns characterize the distribution of emissions per vehicle and road, how these distributions change in time and space, and how we can exploit this information to simulate emission reduction scenarios. For example, although it is reported that the distribution of emissions from on-road remote sensing sites across vehicles is skewed[42-44], this finding has been questioned given the inherent limitations of this type of measurement[45].

In this study, we analysed the estimated emissions of several air pollutants from thousands of private vehicles moving in different European cities. We used trajectories produced by onboard GPS devices to compute the vehicle emissions and matched the obtained emissions to the cities' road networks. We then studied how the emissions distribute across vehicles and roads to discover the statistical patterns that characterize emissions and investigated the relationships between emissions, human mobility and the road network's characteristics. Finally, we simulated two emission reduction scenarios in which a share of vehicles become zero-emission or limit their mobility, identifying strategies to drastically reduce emissions over a city while minimizing the share of vehicles targeted.

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy. [2]Institute of Information Science and Technologies, National Research Council, Pisa, Italy. ✉e-mail: bohm@diag.uniroma1.it; luca.pappalardo@isti.cnr.it

Table 1 | Summary statistics of the GPS data

Original GPS trajectories			
	Vehicles	Trips	Points	Average sampling rate (s)
London	2,745	117,930	2,978,989	72.9 (144)
Rome	9,188	254,088	2,221,206	207.7 (245.6)
Florence	4,782	51,000	291,598	261.3 (400.4)
GPS trajectories after filtering			
	Vehicles	Trips	Points	Average sampling rate (s)
London	2,721	233,627	2,936,512	58.3 (18)
Rome	9,069	216,083	1,033,487	76.8 (27.7)
Florence	4,471	35,145	86,187	75.8 (31.2)

The number of vehicles, trips and points, and their mean sampling rate (standard deviations are given in parentheses), are reported for Greater London, Rome and Florence before and after the filtering step. Note that after filtering, the number of trips of London is higher due to GPS data acquisition. See Supplementary Note 2 for details.

Our framework, which applies to any city provided the availability of vehicle GPS trajectories and road network data, may provide practical support for decision-makers to implement strategies to reduce emissions, improve citizens' well-being and design more sustainable cities[46–48].

Results
Computation of emissions. We used anonymous GPS trajectories describing 423,018 trips from 16,715 private light-duty vehicles moving in Greater London, Rome and Florence throughout January 2017 (Table 1). The spatiotemporal patterns of the vehicle trajectories are stable across the cities and the seasons of the year (Supplementary Note 1).

The trajectories were produced by onboard GPS devices, which automatically turn on when the vehicle starts, transmitting a point every minute to the server via a General Packet Radio Service connection[18–20]. When the vehicle stops, no points are logged or 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, it covers at least 2% of the total registered vehicles, and it is representative of the overall number of vehicles circulating in a city[18]. Figure 1a shows a sample of trajectories for 20 vehicles in Rome.

We defined a methodological framework to compute vehicle emissions from their raw GPS trajectories (Extended Data Fig. 1). We filtered the GPS trajectories so that the time between consecutive points is below a certain threshold (see Methods and Supplementary Note 2). For each vehicle, we estimated the instantaneous speed and acceleration at each point of its trajectory and filtered out points with unrealistic values (see Methods and Fig. 1b). We used a nearest-neighbour algorithm to assign the points to the cities' roads based on the road networks downloaded from OpenStreetMap[49] (see Methods).

The three cities are heterogeneous in their road networks: Rome is large, but with the sparsest network; London is huge, but with the densest network; Florence is small (~1/12 of Rome and ~1/15 of London in terms of land area), but with a dense road network (see Supplementary Note 3 for details).

We employed a microscopic emissions model[30] that uses speed, acceleration and fuel type to estimate the instantaneous vehicle emissions of CO_2, nitrogen oxides (NO_x), particulate matter (PM) and volatile organic compounds (VOC; see Methods). Finally, we computed each vehicle's overall emissions as the sum of all its

instantaneous emissions during the period of study. Analogously, we computed the overall amount of air pollutants on each road by summing all the instantaneous emissions from any vehicle passing along that road during the same period (Fig. 1c).

Patterns of emissions. We found, for all three cities, that the emissions were distributed across vehicles in a heterogeneous way: a few vehicles, which we call gross polluters, were responsible for a tremendous amount of the emissions. At the same time, most vehicles emitted considerably less (Fig. 2). The distribution of emissions per vehicle is associated with a Gini coefficient higher than 0.55, for all the cities and pollutants (Supplementary Note 4). In line with previous studies[42–44], we found that the top 10% of gross polluters in Florence, Rome and London were responsible for 47.5, 50.5 and 38.5% of the total CO_2 emitted during the month, respectively. The distributions of CO_2 emissions per vehicle of Rome and Florence are well approximated by a truncated power law with probability density function: $p(x) \propto x^{-\alpha} e^{-\lambda x}$, with parameters $\alpha = 1.13$ and $\lambda = 1.04 \times 10^{-3}$ for Rome (Fig. 2e), and parameters $\alpha = 2.12$ and $\lambda = 1.45 \times 10^{-3}$ for Florence (Fig. 2h), where e represents the natural logarithm. Similarly, London's distribution is well approximated by a stretched exponential with probability density function: $p(x) \propto x^{\beta-1} e^{-\lambda x^{\beta}}$, with parameters $\lambda = 5.7 \times 10^{-4}$ and $\beta = 1.26$ (Fig. 2b). These results are consistent with those we obtained for the other three pollutants (NO_x, PM and VOC): a truncated power law approximates well the distribution for Rome and Florence, and a stretched exponential approximates well the distribution for London (see Supplementary Notes 4 and 5 for details).

The picture is similar when considering the distribution of emissions per road: a few grossly polluted roads suffered from a substantial quantity of emissions, most of the roads suffered substantially fewer emissions. The distributions for all the cities and pollutants are associated with a Gini coefficient higher than 0.64 (Supplementary Note 4), and are well approximated by a truncated power law, with exponents $\alpha = 1.55$ and $\lambda = 1.08 \times 10^{-4}$ for Rome (Fig. 2f), $\alpha = 1.52$ and $\lambda = 1.30 \times 10^{-4}$ for Florence (Fig. 2i), and $\alpha = 2.59$ and $\lambda = 2.88 \times 10^{-4}$ for London (Fig. 2c). Both exponents α and λ are higher for London than for Rome and Florence, denoting a more even distribution of emissions per road (Supplementary Fig. 13). In Florence and Rome, the top 10% of 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 more than half of the city's total emissions of CO_2. Again, we found similar results for the other pollutants (Supplementary Note 4).

The above results held when changing the year's season (Supplementary Note 1). Also, the sample size of the dataset and the choice of filtering parameter θ did not affect the significance of our results: the shape of the distributions held even if we substantially reduced the sample size or changed θ (Supplementary Note 6).

Relationship with mobility and road features. To investigate the relationship between a vehicle's emissions and mobility patterns, we computed Spearman's correlation coefficient between the emissions and three mobility metrics (see Methods): the radius of gyration, indicating the characteristic distance travelled by an individual[50,51], the mobility entropy[52–54], characterizing the predictability of their visitation patterns, and the total travel time of the vehicles, a principal factor governing emissions. The travel time shows positive correlations, with the strength of the correlations varying from city to city. The radius correlates positively with a vehicle's emissions, whereas the entropy correlates negatively. In London, the travel time has a strong positive correlation (0.98) with the emissions (Table 2), the radius has an almost null correlation (0.09) and the entropy has a strong negative correlation (−0.72). In Rome, the strength of the correlation with travel time is high (0.8), and the radius (0.58) and entropy (−0.54) are positively and negatively correlated with the emissions, respectively. In Florence, the correlation coefficient of
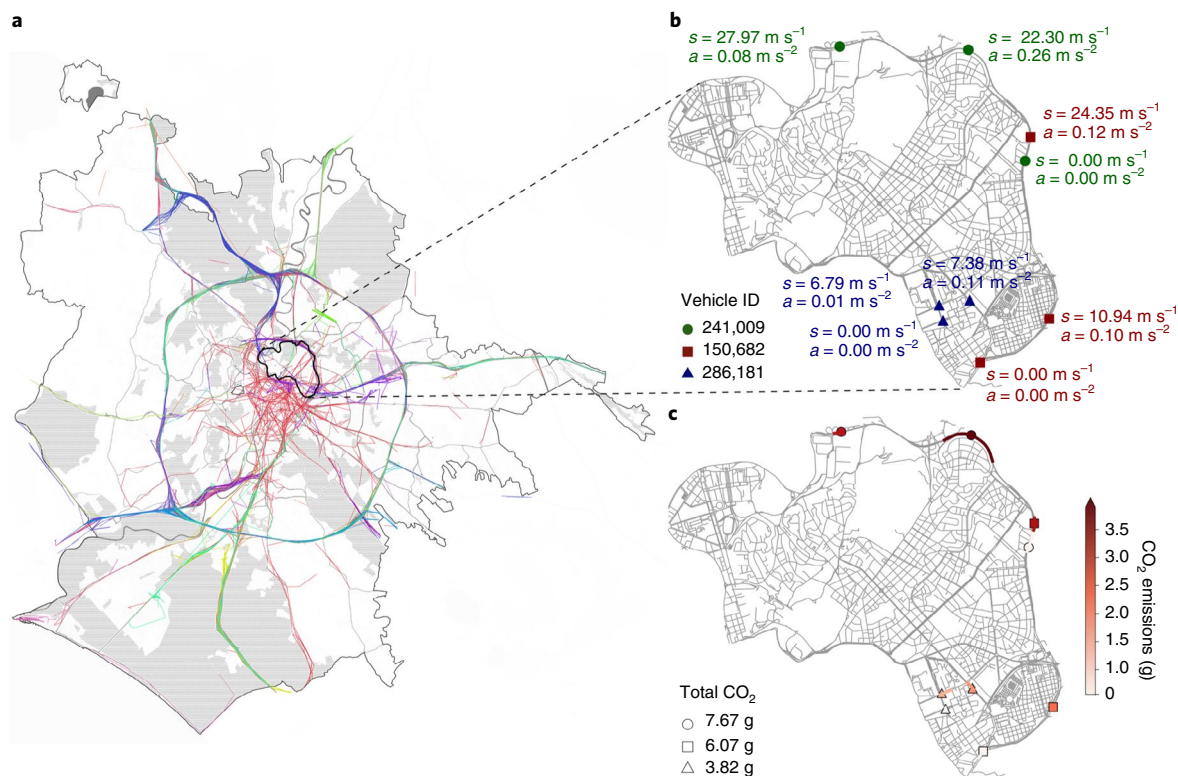
Fig. 1 | Computation of emissions from GPS trajectories. a, Visualization of the trajectories of 20 vehicles travelling in in Rome, Italy, during January 2017. Each colour represents a different trajectory. The grey area indicates the territory of the municipality of Rome. Plot generated with Python library scikit-mobility[51]. **b**, Visualization of the GPS points of three vehicles passing through a neighbourhood of Rome. Each symbol and colour indicates a different vehicle, with its own identity (ID) number; the corresponding instantaneous speed (s) and acceleration (a) are given for each point. The point with zero speed and acceleration is the first point of the trajectory sample. **c**, Instantaneous CO_2 emissions at each GPS point and for each road crossed. The points and the roads are coloured in a gradient ranging from white (low emission) to red (high emission). The legend shows the overall emissions for each vehicle. The road networks in **b** and **c** were plotted with the Python library OSMnx[49,63].

the travel time is reduced (0.48), and those of the radius and entropy show similar behaviour to Rome: the first is positive (0.30), the second is negative (−0.29).

As could be expected, the more a vehicle travels, the more emissions it produces. However, the vehicles with more regular and predictable behaviour generate the highest emissions, not those with more erratic behaviour. Indeed, mobility entropy is low when a vehicle performs a high number of recurring trips, indicating predictable travelling patterns. In contrast, it is high when the vehicle performs trips from various origins and destinations, denoting a more unpredictable travelling behaviour. The observed negative correlations suggest that gross polluters are more regular and predictable than low-emitting vehicles.

To deepen our understanding of these relationships, we used a generalized additive model[55] to express the emissions as a non-linear combination of the three mobility measures (Supplementary Note 7). We found that the radius and entropy contribute in an opposite way to determine a vehicle's emissions. On the one hand, for Rome and Florence, the greater a vehicle's radius of gyration, the greater its emissions (Supplementary Figs. 25 and 26). For London,

the radius of gyration's marginal contribution to the emissions is constant for radii of >7 km (Supplementary Fig. 24). On the other hand, for Rome and Florence, the greater a vehicle's entropy, the lower its emissions (Supplementary Figs. 25 and 26). For London, the negative marginal contribution of the entropy to the emissions only holds for an entropy value of >0.7 (Supplementary Fig. 24). We also performed a cluster analysis to group vehicles based on their radius of gyration, mobility entropy and travel time. We found two clusters, namely the predictable and erratic drivers, and found that the former emit typically more than the latter (Supplementary Note 7 and Supplementary Fig. 27).

The interpretation of these results implies that further analyses are required to garner additional data providing information about the motivations behind each vehicle's trip (for example, drivers' mobility diaries). For example, the erratic vehicles could emit less because they are primarily used for sporadic excursions towards unknown locations. In contrast, predictable drivers could be forced to use private vehicles because they live or work in neighbourhoods poorly served by public transportation. Also, the heterogeneity of emissions' distributions could be related to socioeconomic

Fig. 2 | Distributions of emissions. a,d,g, CO_2 emissions (expressed as grams per metre of road emitted during January 2017) on the road networks of Greater London (**a**), Rome (**d**) and Florence (**g**). The roads are coloured according to the level of emissions in a gradient ranging from yellow (low emission) to red (high emission). The road networks in **a,d,g** were plotted with the Python library OSMnx[49,63]. **b,e,h**, Plots, on the log–log scale, showing $P(X > x)$, i.e., 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 (**b**), Rome (**e**) and Florence (**h**). **c,f,i**, Plots, on the log–log scale, showing the CCDF (black dots) of the CO_2 emissions per road, together with the best fit (red curve), in Greater London (**c**), Rome (**f**) and Florence (**i**).
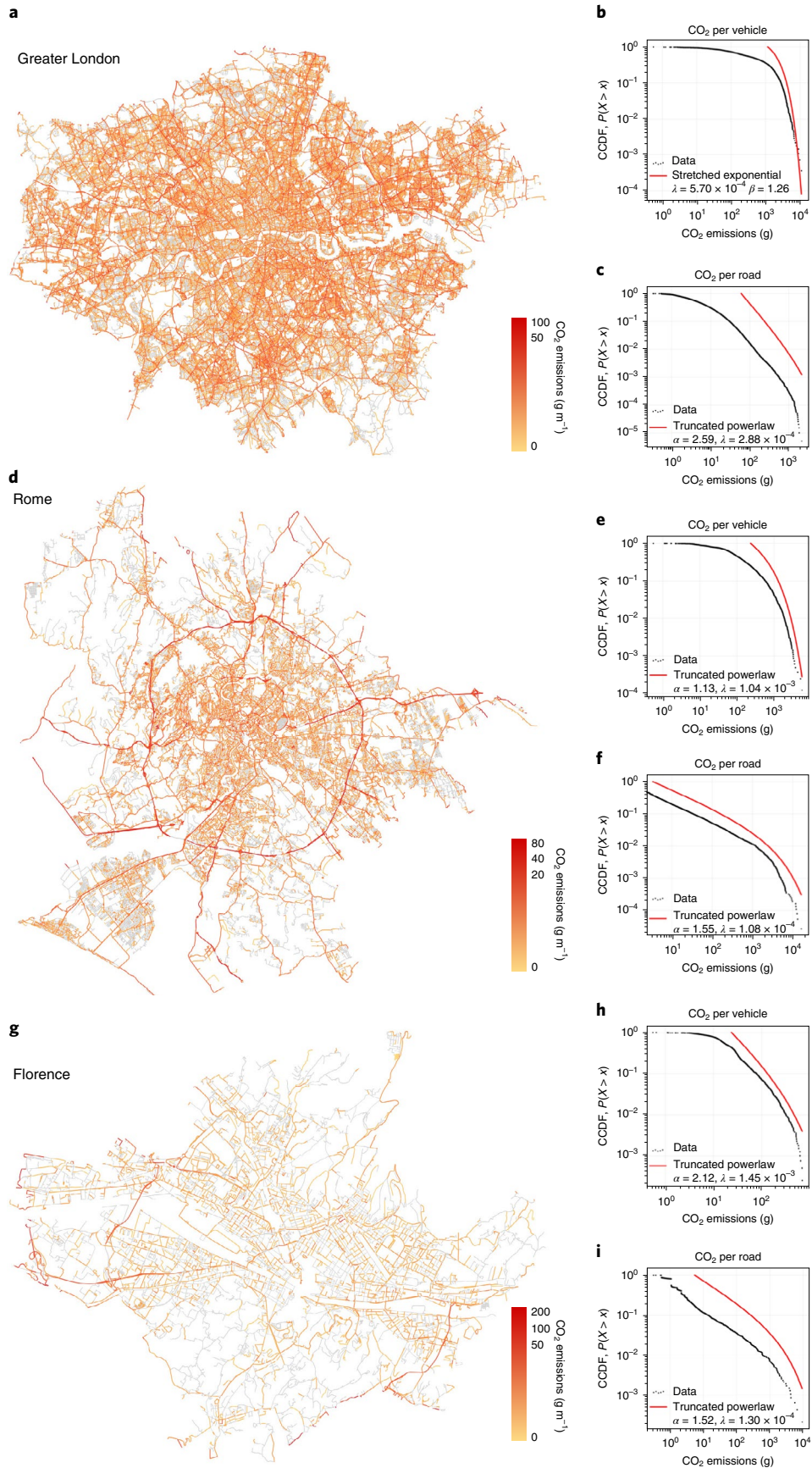
a

Greater London

b

CO$_2$ per vehicle

Data
Stretched exponential
$\lambda = 5.70 \times 10^{-4}$ $\beta = 1.26$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

c

CO$_2$ per road

Data
Truncated powerlaw
$\alpha = 2.59$, $\lambda = 2.88 \times 10^{-4}$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

CO$_2$ emissions (g m^{-1})
100
50
0

d

Rome

e

CO$_2$ per vehicle

Data
Truncated powerlaw
$\alpha = 1.13$, $\lambda = 1.04 \times 10^{-3}$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

f

CO$_2$ per road

Data
Truncated powerlaw
$\alpha = 1.55$, $\lambda = 1.08 \times 10^{-4}$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

CO$_2$ emissions (g m^{-1})
80
40
20
0

g

Florence

h

CO$_2$ per vehicle

Data
Truncated powerlaw
$\alpha = 2.12$, $\lambda = 1.45 \times 10^{-3}$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

i

CO$_2$ per road

Data
Truncated powerlaw
$\alpha = 1.52$, $\lambda = 1.30 \times 10^{-4}$

CCDF, $P(X > x)$

CO$_2$ emissions (g)

CO$_2$ emissions (g m^{-1})
200
100
50
0

Table 2 | Correlations of vehicle emissions with mobility metrics and road features

	Mobility metrics			Road features	
	Radius	Entropy	Travel time	Betweenness centrality	Length
London	0.09	−0.72	0.98	0.27	0.22
Rome	0.58	−0.54	0.80	0.30	0.35
Florence	0.30	−0.29	0.48	0.10	0.25

Spearman's correlation coefficients between CO_2 emissions per vehicle and vehicle mobility metrics (radius of gyration, uncorrelated entropy and travel time), and between CO_2 emissions per road and road features (betweenness centrality and length).

inequalities and the centre–periphery divide. For example, "new low-density settlements that often take on the character of sprawl and rely exclusively on private transportation" have been developed in Rome[56].

We also investigated the relationship between the emissions suffered by a road and network features such as road length and the betweenness centrality of the edge representing the road. The betweenness centrality is based on how frequently a road lies on the shortest paths connecting two crossroads (see Methods). Therefore, a road with a high centrality is more likely to be crossed. We found a positive correlation between the betweenness centrality and the emissions on roads for all three cities (Table 2). This result confirms that our emissions estimates are consistent with the roads' characteristics, as roads that are central in the network are more likely to host greater emissions. Similarly, we found a positive correlation with the road length (Table 2), as longer roads capture more points and thus more emissions. Similar results were obtained for the emissions of the other three pollutants in all three cities (Supplementary Note 7). Figure 2a,d,g shows the entire road networks of Greater London, Rome and Florence, respectively, with the emissions on each road normalized by the length of the road to highlight better the differences between the roads.

Simulation scenarios. Reducing emissions is becoming increasingly important for cities, and it is crucial to estimate the impact of policies targeting vehicles to reduce their footprint on a city's environment. We investigated the impact of vehicle electrification on the total amount of emissions and the distribution of those emissions across the roads. In particular, we studied how the electrification of a certain share of vehicles would change the emissions on the roads of the three cities. In this setting, even if a vehicle's electrification were to change its driver's mobility behaviour, the vehicle would not create any emissions.

We found that the electrification of just the top 1% of gross polluters would reduce emissions as much as electrifying a random 10% of vehicles. In Fig. 3, we show a case study for the entire city of Rome, as well as for a single neighbourhood, to investigate the impact of massive electrification on emissions. The results for London and Florence are provided in Supplementary Note 8. As the number of gross polluters that shift to electric engines grows, the impact on the roads of reducing emissions becomes more evident. In particular, if the top 10% of gross polluters shift to an electric engine, 107 roads would experience a notable reduction in the grams of CO_2 per metre (at least equal to $0.01\,\mathrm{g\,m^{-1}}$; Table 3 and Fig. 3d). In contrast, if 10% of the vehicles that shift to electric engines are chosen at random, only 18 roads would have a substantial reduction of emissions (Fig. 3b). These results hold for both single neighbourhoods and the entire city (see Fig. 3e,f for Rome and Supplementary Note 8 for London and Florence).

The percentage reduction of the overall emissions grows almost linearly when the share of electric vehicles is chosen at random.

In contrast, a generalized logistic function (GLF), also known as Richard's curve[57,58], approximates the growth rate when the gross polluters are electrified first. We used non-linear least-squares methods to fit the GLF, which describes the growth of a variable x as $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$, 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 (Fig. 3e) and for the whole city of Rome (Fig. 3f). The estimated growth rate r of the curve is 4.84×10^{-2} for the neighbourhood and 3.96×10^{-2} for the entire city of Rome. Its slope ν is almost the same for both the neighbourhood and the entire city (-1.55 and -1.56, respectively). The values of α and β are ~100 and -1, respectively, for both the neighbourhood and the city. Similar results hold for Florence (Supplementary Fig. 31b and Supplementary Table 11). In Greater London (Supplementary Fig. 31a), the growth starts slowly ($\nu = -0.86$): there are fewer vehicles with high emissions levels, and electrifying the most polluting vehicles is slightly less effective in reducing emissions than in the other two cities (Supplementary Table 11).

Given the increasing importance of remote working, especially since the COVID-19 pandemic[59,60], we simulated the impact of a massive shift to remote working on reducing vehicle emissions. This working style may affect individual mobility patterns, but we assumed that it eliminates commuting trips. Indeed, if an individual works from home, the most straightforward implication is the removal of commuting trips from their mobility habits. We identified vehicles' home and work locations (see Methods) and studied the emissions generated from their commuting patterns. We then performed a simulation in which a growing share of these commuters become home workers, that is, they no longer travel between their home and work locations.

We found that emissions reduction is more effective when the home workers are gross polluters. In this case, remote working for the top 1% gross polluters leads to the same reduction as if they were ~4% random vehicles (Supplementary Fig. 32). Again, a GLF fits well the emissions reduction when the gross polluters become home workers. In particular, we obtained 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 Note 8 and Supplementary Fig. 32 for details).

Overall, these results demonstrate that targeting specific profiles of vehicles can substantially improve emission reduction policies.

Discussion

Using GPS data to estimate the emissions from thousands of vehicles in three European cities of different sizes and characteristics, we have shown here the existence of gross polluters, that is, vehicles responsible for the greatest quantity of emissions. The existence of gross polluters has been reported in previous studies using measurements from on-road remote sensing sites[42–44]; however, these studies have been questioned because measurements from on-road sites cannot represent a vehicle's overall emission level[45]. Our study contributes to reshaping this discussion because our findings are based on a microscopic emission model that captures in great detail the instantaneous emissions of vehicles. We have added new elements to this debate, discovering that gross polluters exist in different cities and for different pollutants (CO_2, NO_x, PM and VOC) and reporting the existence of grossly polluted roads suffering the greatest amount of emissions.

The heterogeneous patterns governing the distribution of emissions across vehicles and roads are well approximated by heavy-tailed distributions, with exponents that vary from city to city and from pollutant to pollutant. These peculiar exponents may depend on the characteristics of the city's road network and people's commuting behaviour. For example, London has a vast and dense road network, and people use private vehicles less intensively[61] than
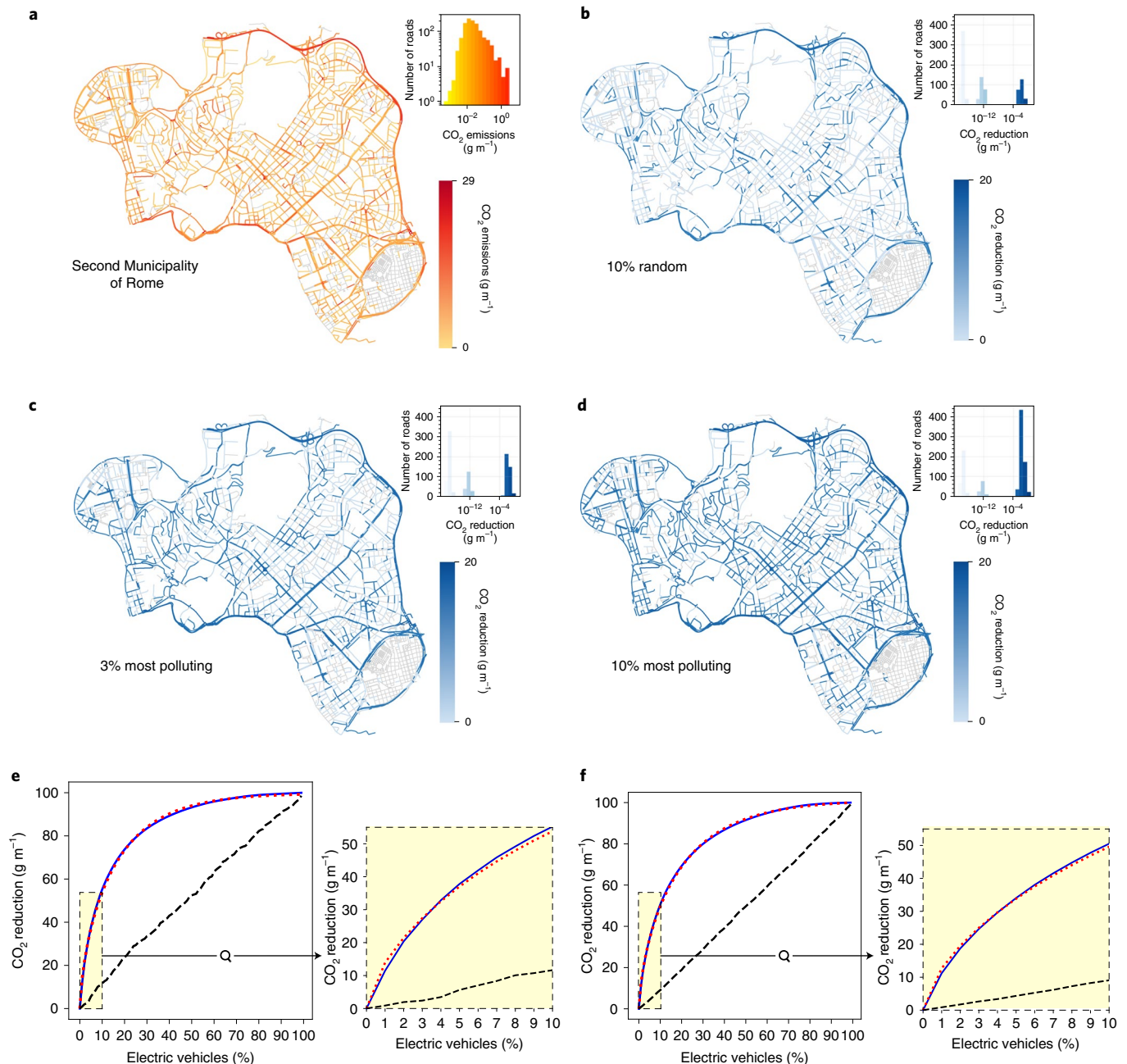
Fig. 3 | Simulation of vehicle electrification. Impact on CO_2 emissions of the electrification of the vehicles moving within a neighbourhood (the Second Municipality of Rome), as well as within the whole city of Rome, during January 2017. **a**, Distribution of emissions of CO_2 across the roads of the neighbourhood. The roads are coloured in a gradient ranging from yellow (low emission) to red (high emission), and the histograms show the distribution of the emissions across the roads, with the bars coloured according to the same gradient. **b**–**d**, Reduction of emissions experienced by the same roads if a random 10% of the vehicles (**b**), the top 3% gross polluters (**c**) and the top 10% gross polluters (**d**) shift to electric engines. The roads are coloured in a gradient ranging from light blue (low reduction) to dark blue (high reduction), and the histograms show the distribution of the reduction in emissions across the roads, with the bars coloured according to the same gradient. **e**,**f**, Percentage reduction of CO_2 emissions corresponding to the share (0–100%) of electric vehicles in the Second Municipality of Rome (**e**) and for the whole city of Rome (**f**). The insets zoom in on the first 10% share of electric vehicles. The solid blue lines correspond to the scenario when the vehicles to be electrified are chosen from the most polluting to the least polluting. The black dashed lines indicate the emission reduction when the vehicles to be electrified are chosen at random. The dotted red line is the GLF fit: $f(x) = \frac{\alpha}{(1+\beta e^{-rx})^{1/\nu}}$. The road networks in **a**–**d** were plotted with the Python library OSMnx[49,63].

in Rome and Florence. Thus, one can argue that mobility behaviour in London leads to a distribution of emissions per vehicle that is more even than in Rome, which is characterized by a vast and sparse road network and intensive use of private vehicles[62].

Our study can be reproduced with any city provided the availability of vehicle GPS trajectories and road network data, and may

help to find more effective strategies to reduce emissions. For example, our study demonstrates that blocking the circulation based on an uninformed choice (for example, 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 have designed a precise model to estimate the overall reduction

Table 3 | CO$_2$ reduction experienced on roads in three different scenarios

CO$_2$ reduction (g m^{-1})	Number of roads with specific CO$_2$ reduction		
	10% random	3% most polluting	10% most polluting
≥10^{-4}	232	385	665
≥10^{-3}	137	258	469
≥10^{-2}	18	53	107
≥10^{-1}	0	6	10

The number of roads that experience certain levels of CO$_2$ reduction are presented according to three different scenarios of vehicle electrification: (1) when electrifying a random 10% of vehicles, (2) when electrifying the top 3% most polluting vehicles and (3) when electrifying the top 10% most polluting vehicles.

of emissions caused by the electrification of a particular share of vehicles or by reducing the number of commuting trips travelled by the vehicles (for example, caused by a transition to the home working of their drivers).

There are several directions in which this study can be extended. For example, because we focused on light-duty vehicles, all the results we have shown are valid for this fleet of vehicles only. Although they make up the vast majority of vehicles circulating in a city, we are aware that the absence of other vehicles, such as heavy-duty vehicles (for example, buses and trucks), generates an incomplete mosaic of the emissions within the urban environment. We hope, therefore, for a more comprehensive study that may include different types of vehicles.

Also, our analysis can be extended by investigating how the emission patterns vary between weekdays and weekends or weather conditions, and by considering more sophisticated simulation scenarios. For example, it would be interesting to investigate the impact of policies that aim to improve walking, transit or cycling on the distribution of emissions, the number of gross polluters and grossly polluted roads. Finally, the relationship between vehicle emissions and mobility patterns may be examined in more depth to investigate whether the observed heterogeneous distributions originate from other inequalities (for example, socioeconomic inequalities and the centre–periphery divide).

Meanwhile, our study may shape the discussion on measuring emissions with digital data and how to use such measurements to simulate emission reduction scenarios. If we learn how to use such a resource, we have the potential to monitor in real time the level of emissions in our urban environments and take immediate, informed actions when they overcome a certain tolerance threshold. This fact is crucial because the decisions that policymakers take depend on what we measure, how good our measurements are and how promptly we react to these measurements.

Methods

Data filtering. In our GPS dataset, each trajectory point is associated with a vehicle identifier, a trajectory identifier, a timestamp and a latitude and longitude pair. The sampling rate of the trajectory points may affect the estimate of instantaneous speed and acceleration. Because the mean time interval between trajectory points varies from city to city (it is about 1 min for London and 4 min for Rome and Florence, see Table 1), we performed a pre-processing step to align them. For each trajectory, we retained only those subtrajectories (that is, disjoint subsets of points) that satisfy two constraints: (1) there are at least two points and (2) the time interval between consecutive points is less than $\theta = 120$ s. The filtering step causes a drop in the number of points and, in consequence, of vehicles. We analysed the trends in the number of both vehicles and points resulting from the filtering step, varying the filtering parameter θ from 1 to 300 s (Supplementary Note 1). By choosing $\theta = 120$ s, we lost 53.5% of the points in Rome, 1.4% in London and 70.4% in Florence. Consequently, we discarded 1.3% of the vehicles in Rome, 0.9% in London and 6.5% in Florence (Table 1).

In previous work, different values of θ were used (for example, the time interval between the points was set to 1 s (ref. [36]), 3 s (ref. [37]), 5 s (ref. [30]) and 5–50 s (ref. [31])). Our choice derives from our data sampling rate and is a trade-off between the reliability of the results and the data coverage. As the last step, we computed, for each vehicle, the speed and acceleration at each point, and retained only those points with a speed of less than 300 km h^{-1} and an acceleration in the range -10 to $+10$ m s^{-2} (as suggested by Nyhan et al.[30]).

Road network and GPS point snapping. The road network of each city was extracted from OpenStreetMap (OSM)[49], 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 provided as a multigraph $G = (V, E)$, with V being the set of nodes v and E being the multiset of edges e, the edges representing public roads accessible to vehicles (including service roads). Each edge $e \in E$ comprises two identifiers that indicate the starting and ending OSM nodes, and a key that discriminates between parallel edges (if present). Moreover, it carries some information about the road it represents, such as its name, length and type (for example, whether it is a motorway or residential street). In this context, we call nodes $v \in V$ that have at least two roads (that is, edges $e \in E$) incident on them crossroads. To download, compute statistics and visualize the road networks, we conceived methods based on the Python library OSMnx[63]. Our matching step consisted of a ball tree nearest-neighbour algorithm that assigns each point of a GPS trajectory to the nearest edge in the road network. This point snapping step allowed us to assign vehicle emissions to the roads on which they were produced (Fig. 1c). Note that the nearest-neighbour algorithm used does not affect the quality of the emissions estimated with the microscopic model. Indeed, instantaneous emissions were computed on the basis of the speed and acceleration of the points computed from the GPS data (not matched yet with the road network). Moreover, our GPS data come with a quality index that ranges from 1 to 3 and indicates the precision of each GPS point's location. A point's maximum quality (that is, index = 3) indicates that at least four satellites' signals received by the GPS device are used for the trilateration of the point's location. In this study we used only points with the highest quality (that is, index = 3). Given this accuracy in the points' locations, the reconciliation problem is straightforward[64], and we assigned the location obtained from the GPS receiver to the nearest edge in the network.

Computing emissions. We implemented a microscopic emissions model[30] to compute the instantaneous emissions associated with each trajectory point p. We denote the quantity of pollutant $j \in \{CO_2, NO_x, PM, VOC\}$ emitted at point p from vehicle u as $E_p^{j,u}$, and the instantaneous speed and acceleration of the vehicle at point p as s_p and a_p, respectively. Information about its engine type (whether it is petrol, diesel or liquefied petroleum gas) is available for each vehicle. This information determines, together with the type of pollutant, the emission factors f_i. We used the following equation to compute the instantaneous emissions $E_p^{j,u}$ of pollutant j from vehicle u at point p:

$$E_p^{j,u} = f_1^{j,u} + f_2^{j,u} s_p + f_3^{j,u} s_p^2 + f_4^{j,u} a_p + f_5^{j,u} a_p^2 + f_6^{j,u} s_p a_p \qquad (1)$$

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}). We show the variation of factors $f_1, ..., f_6$ with the vehicle's fuel type and acceleration in Supplementary Table 13.

Mobility measures and road centrality. We used three quantities to describe the mobility of a vehicle u:

- The radius of gyration[18,19,50] $r_g(u) = \sqrt{\frac{1}{n} \sum_{i \in P} \text{dist}[\mathbf{r}_i(u) - \mathbf{r}_{cm}(u)]^2}$, where P is the set of n points recorded for u, $\mathbf{r}_i(u)$ indicates the coordinates of trajectory point $i \in P$, $\mathbf{r}_{cm}(u)$ is the centre of mass of u and dist is the haversine distance between two points on earth
- The temporal-uncorrelated entropy[52–54] $S(u) = - \sum_{i=1}^{N_u} p_u(i) \log_2 p_u(i)$, where N_u is the number of distinct locations visited by u and $p_u(i)$ is the probability that u visits location i
- The travel time of u, computed as the sum of all the travel times of its trajectories
- We measured the centrality of a road, a proxy of its traffic volume in the city, as its betweenness centrality. In network science, the betweenness centrality of an edge e (that is, 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. The first step when identifying an individual's home and work locations is the selection of the starting and ending points of their trajectories[65]. The position of the starting (or ending) points of the trajectories that start from (or end at) the same semantic location may not coincide. This may happen (1) because a driver may park the vehicle within a certain radius of the location and (2) because the first point sent by the GPS device often lacks precision

and is discarded, the second point sent is taken as the starting point of the trajectory. For the above reasons, we spatially clustered these points within a radius of 250 m and took the centroid of each cluster as the vehicle's stop location.

To identify a vehicle's home and work locations, we used a principle commonly adopted in the literature[65]: the home location is the stop location corresponding to the most frequent cluster, and the work location is the stop location corresponding to the second most frequent cluster. We discarded the vehicles for which it is impossible to identify the most frequent stop location(s) (for example, the vehicle visited each location only once). We successfully identified home and work locations for 55, 31 and 16% of the vehicles moving in London, Rome and Florence, respectively. There are two main reasons we cannot identify the home and work locations of many vehicles in Florence. First, the average number of trajectories per vehicle (10.7) is much lower than in the other two cities (27.6 in Rome and 43 in London; see Table 1). The fewer trajectories a vehicle has, the more difficult it is to identify its home and work locations. Second, while a relatively small city, Florence is an essential hub for the surroundings. Thus, many users could live outside the city. Moreover, even if they live inside the city, given various restricted traffic areas in the city's historic centre, many could reach the workplace by public transportation or walking. This leads to a low number of commuting trajectories inside the city.

Reporting summary. Further information on research design is available in the Nature Research Reporting summary linked to this article.

Data availability
The data that support the findings of this study are not publicly available due to privacy restrictions and were used under licence for the current study. Aggregated source data for figures are available from the authors upon reasonable request.

Code availability
The Python code to reproduce the analyses in the study from public GPS data (taxi trips) is publicly available in GitHub (https://github.com/matteoboh/mobility_emissions) and the Zenodo repository[66].

References
1. Venter, Z. S., Aunan, K., Chowdhury, S. & Lelieveld, J. COVID-19 lockdowns cause global air pollution declines. *Proc. Natl Acad. Sci. USA* **117**, 18984–18990 (2020).
2. Le Quéré, C. et al. Temporary reduction in daily global CO₂ emissions during the COVID-19 forced confinement. *Nat. Clim. Change* **10**, 647–653 (2020).
3. He, G., Pan, Y. & Tanaka, T. The short-term impacts of COVID-19 lockdown on urban air pollution in China. *Nat. Sustain.* **3**, 1005–1011 (2020).
4. Liu, Z. et al. Near-real-time monitoring of global CO₂ emissions reveals the effects of the COVID-19 pandemic. *Nat. Commun.* **11**, 5172 (2020).
5. Forster, P. et al. Current and future global climate impacts resulting from COVID-19. *Nat. Clim. Change* **10**, 913–919 (2020).
6. IPCC *Climate Change 2014: Mitigation of Climate Change* (eds Edenhofer et al.) 599–670 (Cambridge Univ. Press, 2014).
7. Ritchie, H. *Sector by Sector: Where Do Global Greenhouse Gas Emissions Come From?* (Our World in Data, 2020); https://ourworldindata.org/ghg-emissions-by-sector
8. United Nations General Assembly *Transforming Our World: The 2030 Agenda for Sustainable Development* (United Nations, 2015).
9. deSouza, P. et al. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: methods development and lessons learned. *Sustaina. Cities Soc.* **60**, 102239 (2020).
10. Chong, H. S., Kwon, S., Lim, Y. & Lee, J. Real-world fuel consumption, gaseous pollutants, and CO₂ emission of light-duty diesel vehicles. *Sustain. Cities Soc.* **53**, 101925 (2020).
11. Luján, J. M., Bermúdez, V., Dolz, V. & Monsalve-Serrano, J. An assessment of the real-world driving gaseous emissions from a Euro 6 light-duty diesel vehicle using a portable emissions measurement system (PEMS). *Atmos. Environ.* **174**, 112–121 (2018).
12. Chatterton, T., Barnes, J., Wilson, R. E., Anable, J. & Cairns, S. Use of a novel dataset to explore spatial and social variations in car type, size, usage and emissions. *Transp. Res. D* **39**, 151–164 (2015).
13. i Diao, M. & Ferreira, J. Jr. Vehicle miles traveled and the built environment: evidence from vehicle safety inspection data. *Environ. Plan. A.* **46**, 2991–3009 (2014).
14. Kancharla, S. R. & Ramadurai, G. Incorporating driving cycle based fuel consumption estimation in green vehicle routing problems. *Sustain. Cities Soc.* **40**, 214–221 (2018).
15. Choudhary, A. & Gokhale, S. Urban real-world driving traffic emissions during interruption and congestion. *Transp. Res. D* **43**, 59–70 (2016).
16. Ferreira, J. C., de Almeida, J. & da Silva, A. R. The impact of driving styles on fuel consumption: a data-warehouse-and-data-mining-based discovery process. *IEEE Trans. Intell. Transp. Syst.* **16**, 2653–2662 (2015).
17. Zheng, F., Li, J., van Zuylen, H. & Lu, C. Influence of driver characteristics on emissions and fuel consumption. In *20th EURO Working Group on Transportation Meeting* (eds Esztergár-Kiss, D. et al.) 624–631 (Elsevier, 2017).
18. Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D. & Giannotti, F. Understanding the patterns of car travel. *Eur. Phys. J. Spec. Top.* **215**, 61–73 (2013).
19. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
20. Gallotti, R., Bazzani, A. & Rambaldi, S. Towards a statistical physics of human mobility. *Int. J. Mod. Phys. C* **23**, 1250061 (2012).
21. Luca, M., Barlacchi, G., Lepri, B. & Pappalardo, L. A survey on deep learning for human mobility. *ACM Comput. Surv.* **55**, 1–44 (2021).
22. Barbosa, H. et al. Human mobility: models and applications. *Phys. Rep.* **734**, 1–74 (2018).
23. Çolak, S., Lima, A. & González, M. C. Understanding congested travel in urban areas. *Nat. Commun.* **7**, 10793 (2016).
24. Lwin, H. & Naing, T. Estimation of road traffic congestion using GPS data. *Int. J. Adv. Res. Comput. Commun. Eng.* **4**, 1–5 (2015).
25. Stipancic, J., Miranda-Moreno, L., Labbe, A. & Saunier, N. Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected GPS data. *Transp. Lett.* **11**, 391–401 (2019).
26. Camargo, C. Q., Bright, J., McNeill, G., Raman, S. & Hale, S. A. Estimating traffic disruption patterns with volunteered geographic information. *Sci. Rep.* **10**, 1271 (2020).
27. Jenn, A. Emissions benefits of electric vehicles in Uber and Lyft ride-hailing services. *Nat. Energy* **5**, 520–525 (2020).
28. Liang, X. et al. Air quality and health benefits from fleet electrification in China. *Nat. Sustain.* **2**, 962–971 (2019).
29. Rolnick, D. et al. Tackling climate change with machine learning. *ACM Comput. Surv.* **55**, 1–96 (2019).
30. Nyhan, M. et al. Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmos. Environ.* **140**, 352–363 (2016).
31. Liu, J., Han, K., Chen, X. M. & Ong, G. P. Spatial-temporal inference of urban traffic emissions based on taxi trajectories and multi-source urban data. *Transp. Res. C* **106**, 145–165 (2019).
32. Reznik, A., Kissinger, M. & Alfasi, N. Real-data-based high-resolution GHG emissions accounting of urban residents private transportation. *Int. J. Sustain. Transp.* **13**, 235–244 (2019).
33. Wang, X., Grengs, J. & Kostyniuk, L. Using a GPS data set to examine the effects of the built environment along commuting routes on travel outcomes. *J. Urban Plan. Dev.* **140**, 04014009 (2014).
34. Cervero, R. & Kockelman, K. Travel demand and the 3Ds: density, diversity, and design. *Transp. Res. D* **2**, 199–219 (1997).
35. Gately, C. K., Hutyra, L. R., Peterson, S. & Sue Wing, I. Urban emissions hotspots: quantifying vehicle congestion and air pollution using mobile phone GPS data. *Environ. Pollut.* **229**, 496–504 (2017).
36. Chen, J. et al. Mining urban sustainable performance: GPS data-based spatio-temporal analysis on on-road braking emission. *J. Clean. Prod.* **270**, 122489 (2020).
37. Sui, Y. et al. GPS data in urban online ride-hailing: a comparative analysis on fuel consumption and emissions. *J. Clean. Prod.* **227**, 495–505 (2019).
38. Yu, Q. et al. Mobile phone GPS data in urban customized bus: dynamic line design and emission reduction potentials analysis. *J. Clean. Prod.* **272**, 122471 (2020).
39. Rahman, M. N. & Idris, A. O. Tribute: trip-based urban transportation emissions model for municipalities. *Int. J. Sustain. Transp.* **11**, 540–552 (2017).
40. Zhu, S., Kim, I. & Choi, K. High-resolution simulation-based analysis of leading vehicle acceleration profiles at signalized intersections for emission modeling. *Int. J. Sustain. Transp.* **15**, 375–385 (2020).
41. Aziz, H. M. A. & Ukkusuri, S. V. A novel approach to estimate emissions from large transportation networks: hierarchical clustering-based link-driving-schedules for EPA-MOVES using dynamic time warping measures. *Int. J. Sustain. Transp.* **12**, 192–204 (2018).
42. Guenther, P., Bishop, G., Peterson, J. & Stedman, D. Emissions from 200 000 vehicles: a remote sensing study. *Sci. Total Environ.* **146–147**, 297–302 (1994).
43. Brand, C. & Boardman, B. Taming of the few—the unequal distribution of greenhouse gas emissions from personal travel in the UK. *Energy Policy* **36**, 224–238 (2008).
44. Huang, Y. et al. Remote sensing of on-road vehicle emissions: mechanism, applications and a case study from Hong Kong. *Atmos. Environ.* **182**, 58–74 (2018).
45. Huang, Y. et al. Re-evaluating effectiveness of vehicle emission control programmes targeting high-emitters. *Nat. Sustain.* **3**, 904–907 (2020).

46. Batty, M. et al. Smart cities of the future. *Eur. Phys. J. Spec. Top.* **214**, 481–518 (2012).

47. Kitchin, R. The real-time city? Big data and smart urbanism. *GeoJournal* **79**, 1–14 (2014).

48. Voukelatou, V. et al. Measuring objective and subjective well-being: dimensions and data sources. *Int. J. Data Sci. Anal.* **11**, 279–309 (2020).

49. OpenStreetMap contributors https://www.openstreetmap.org (2017); planet dump, retrieved from https://planet.osm.org

50. González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).

51. Pappalardo, L., Simini, F., Barlacchi, G. & Pellungrini, R. scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data. Preprint at http://arxiv.org/abs/1907.07062 (2019).

52. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).

53. Eagle, N. & Pentland, A. S. Eigenbehaviors: identifying structure in routine. *Behav. Ecol. Sociobiol.* **63**, 1057–1066 (2009).

54. Pappalardo, L. et al. An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2**, 75–92 (2016).

55. Hastie, T. & Tibshirani, R. Generalized additive models. *Stat. Sci.* **1**, 297–310 (1986).

56. Lelo, K., Monni, S. & Tomassi, F. Socio-spatial inequalities and urban transformation. The case of Rome districts. *Socioecon. Plan. Sci.* **68**, 100696 (2019).

57. Richards, F. J. A flexible growth function for empirical use. *J. Exp. Bot.* **10**, 290–300 (1959).

58. Fekedulegn, D. & Colbert, J. Parameter estimation of nonlinear growth models in forestry. *Silva Fenn.* **33**, 653 (1999).

59. Vyas, L. & Butakhieo, N. The impact of working from home during COVID-19 on work and life domains: an exploratory study on Hong Kong. *Policy Des. Pract.* **4**, 59–76 (2021).

60. Nagel, L. The influence of the COVID-19 pandemic on the digital transformation of work. *Int. J. Sociol. Soc. Policy* **40**, 861–875 (2020).

61. *Travel in London: Report 12* (Transport for London, 2019); http://content.tfl.gov.uk/travel-in-london-report-12.pdf

62. Fuschiotto, A. et al. *Rapporto Mobilità 2019* (Dipartimento Mobilità e Trasporti Roma Capitale, 2019); https://romamobilita.it/it/media/pubblicazioni/rapporto-mobilita-2019

63. Boeing, G. OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).

64. White, C. E., Bernstein, D. & Kornhauser, A. L. Some map matching algorithms for personal navigation assistants. *Transp. Res. C* **8**, 91–108 (2000).

65. Pappalardo, L., Ferres, L., Sacasa, M., Cattuto, C. & Bravo, L. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Sci.* **10**, 29 (2021).

66. Bohm, M., Nanni, M. & Pappalardo, L. matteoboh/mobility emissions: code release for *Nature Sustainability* paper https://doi.org/10.5281/zenodo.6124225 (2022).

Author contributions
M.B. designed the study, performed the data pre-processing and cleaning, developed the code for and performed the statistical analyses, experiments and simulations, created the plots, contributed to the interpretation of the results and wrote the paper. M.N. provided the data, contributed to the work methodology and interpretation of the results, and revised the paper. L.P. developed the concept and designed the study, contributed to the work methodology and interpretation of the results, wrote the paper and coordinated the study. All authors read, edited and approved the final version of the paper.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41893-022-00903-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-022-00903-x.

Correspondence and requests for materials should be addressed to Matteo Böhm or Luca Pappalardo.

Peer review information *Nature Sustainability* thanks the anonymous reviewers for their contribution to the peer review of this work.
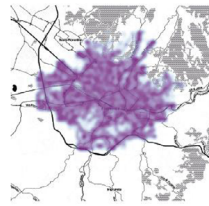
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. **Trajectory selection**

Selection of the trajectories inside the region of interest (e.g. Florence).
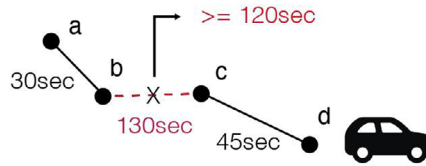
Running example in Florence

4,782 vehicles
51,000 trips
291,598 points

2. **Time filtering step**

Select sub-trajectories (i) composed by at least two points, (ii) whose time distance between consecutive points is smaller than 120 seconds.

a
>= 120sec
b   c
30sec
130sec
d
45sec

4,471 vehicles
35,145 trips
86,187 points

3. **Speed and acceleration estimates**

Compute instantaneous speed and acceleration at each point.

a S: 0.00 m/s
A: 0.00 m/s²
c S: 0.00 m/s
A: 0.00 m/s²
b
d S: 24.35 m/s
A: 0.12 m/s²
S: 22.30 m/s
A: 0.26 m/s²
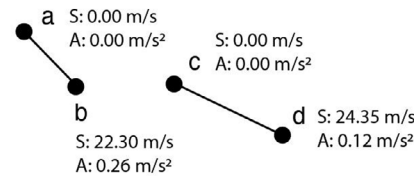
S = speed
A = acceleration

4. **Emissions estimates**

Compute instantaneous emissions of CO_2, NO_x, PM, VOC at each point.
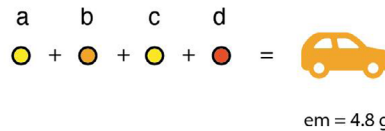
a em= 0.5 g
b
c em=0.5 g
d em=2.3 g
em=1.5 g
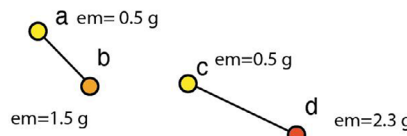
Types of emissions (em)

CO_2 - carbon dioxide
NO_x - nitrogen oxides
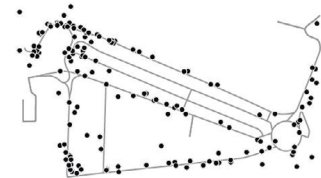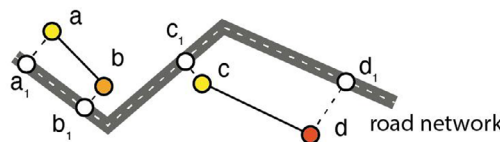PM - particulate matter
VOC - volatile organic compounds

5. **Vehicle aggregation**

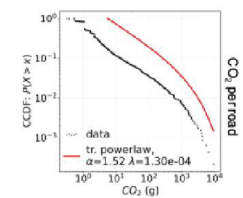Aggregate the emissions for each vehicle, and study the distributions.

a b c d
○ + ○ + ○ + ○ =

em = 4.8 g
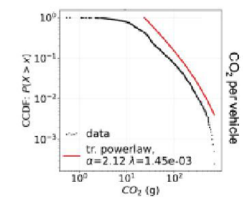
6. **GPS points snapping**

Assign each point to its nearest road in the network.

a c_1
b c
a_1 d_1
b_1 d road network

7. **Road aggregation**

Aggregate the emissions for each road, and study the distributions.

a_1
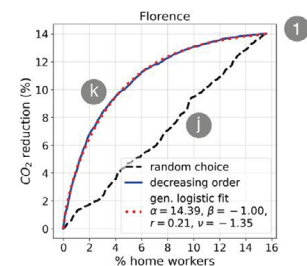b_1 c_1
d_1
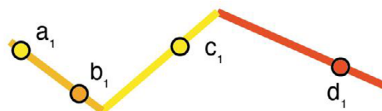
8. **Simulations**

Analyse the impact on emissions' reduction of two scenarios: home working and vehicles' electrification.

① ②

① x% of the drivers stop commuting
② x% of the vehicles become electric

ⓙ x% chosen at random
ⓚ x% chosen starting from the most polluting ones

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Step-by-step procedure for the computation of emissions and results analyses. The left column describes the steps followed starting from the data, passing through the data processing, and ending with the analyses performed. The central column shows a schema of what happens in each step. The right column shows some numbers and results in support of the central column. The heatmap in step 1 is plotted with the Python library scikit-mobility[51]. The small road network in step 7 is plotted with the Python library OSMnx[63]. Car icons from the Noun Project (thenounproject.com).

nature research

Corresponding author(s): DBPR NATSUSTAIN-21049605

Last updated by author(s): May 11, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected by a company and provided to us (anonymized) for research purposes
Data analysis	We used python programming language for our analyses, creating our custom code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets analysed during the current study are not publicly available due to privacy reasons. An aggregated version of the data is available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☐ Behavioural & social sciences   ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Estimation of emissions from private vehicles based on their spatio-temporal trajectories; analysis of emission patterns for vehicles and roads; simulation of emission reduction scenarios.
Research sample	Spatio-temporal trajectories of 14,429 vehicles travelling in Rome, Greater London, and Florence during January 2017.
Sampling strategy	All those vehicles carrying on-board GPS for insurance purposes managed by the company that provided us the data.
Data collection	The vehicles' spatio-temporal trajectories are produced by on-board GPS devices that automatically turns on when the vehicle starts, sending data to a GPRS server every 1 minute approximately.
Timing and spatial scale	The dataset covers the entire month of January 2017.
Data exclusions	No data were excluded from the analysis.
Reproducibility	The study is reproducible starting from the dataset because the computations are deterministic.
Randomization	Not relevant for this study, no randomizations were performed.
Blinding	Not relevant for this study, no blinding was needed.

Did the study involve field work?   ☐ Yes   ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging