Supplementary Materials for

Scaling of Traffic Emissions by Converging Direct Measurements and Mobility Science

Ayşe Tuğba Öztürk[1], Helen Fitzmaurice[2], Olga Kavvada[3], Philippe Calvez[3], Ronald C. Cohen[4], Marta C. González[1,5,6*]

*To whom correspondence should be addressed. E-mail: martag@berkeley.edu.

This PDF file includes:

Notes S1 to S6

Figs. S1 to S6

Tables S1 to S2

Notes

S1 Predicting Missing Speed Values with K-Nearest Neighbors Algorithm

Uber Movement API (1) provides hourly speed profiles of street segments in several metropolitan areas starting from 2018. In this project, the time frame between 03/2018 and 12/2019 is analyzed. The data is provided monthly with average speed values and standard deviations of edges. However, speed profiles are not complete and only include road segments that attracted enough traffic in each hour to protect customer privacy. Figure **S1** illustrates the boundary of the road network analyzed and data availability for each segment. For each road segment, we averaged the speed values hourly for weekdays and weekends. Meaning for each road we can estimate speed values for at most 48 different time blocks. The blue and light blue road segments in Figure **S1** are where we know the majority of the speed values. While in gray roads we have no data on speeds for any of the time blocks.

Route assignment requires us to know all speed values for all time blocks therefore we utilized machine learning methods to predict the missing speed values. For each road segment, we created spatial, temporal and road type features such as highway type and speed limits. The features we generated for prediction are shown in Table **S1**. Then the missing speed values are predicted using the K-nearest neighbors (KNN) algorithm with a k value of 4. The final model had 7 kph mean error in the test set.

S2 Estimating Travel Demand After Shelter in Place Orders with SafeGraph Data

SafeGraph provides aggregated travel demand data between census block groups for before and after shelter in place (SIP) orders. We use SafeGraph data to estimate the change in the

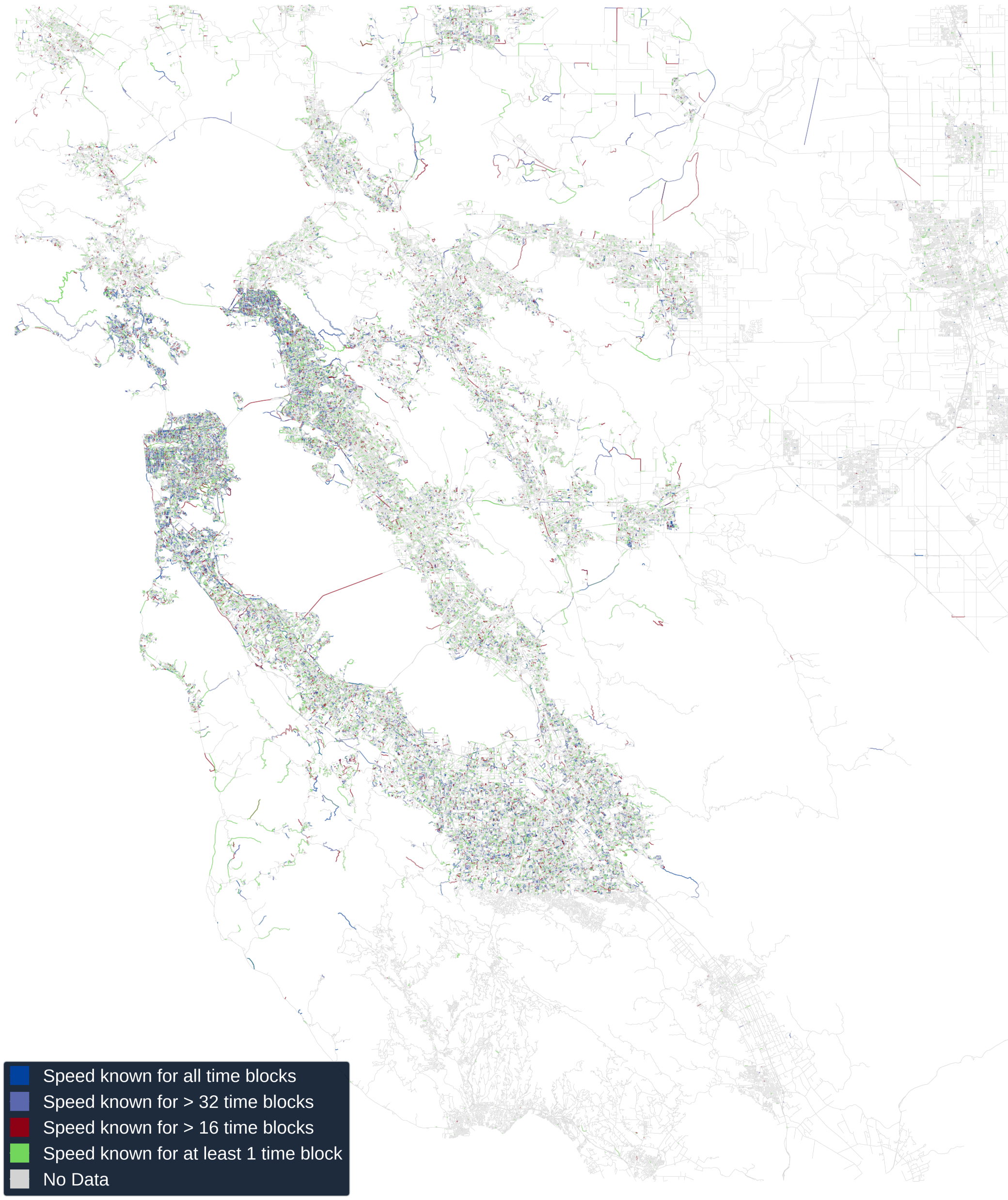

Fig. S1: Complete Bay Area driving road network and color coded road segments where Uber Movement provided speed data between 2018 and 2019. Observed speeds values are used to get the hourly averages for weekday and weekend speeds which makes 48 time blocks in total.

travel behavior due to SIP between origin and destinations. Then the calculated change is used to scale the TimeGeo travel demand data to get the after SIP travel demand (2). The main

Table S1: Prediction Model Features

Feature	Description	Type
speed_kph	Speed limit of the road segment	float
sin_hour	$\sin(2\pi hr/24)$ where *hr* is the local city hour between 0-23	float
cos_hour	$\cos(2\pi hr/24)$ where *hr* is the local city hour between 0-23	float
start_x	longitude of the start node of road segment	float
start_y	latitude of the start node of road segment	float
end_x	longitude of the end node of road segment	float
end_y	latitude of the end node of road segment	float
center_x	longitude of the centroid of the road segment	float
center_y	latitude of the centroid of the road segment	float
highway_motorway	Road type of the road segment	Binary, int
highway_motorway_link	Road type of the road segment	Binary, int
highway_primary	Road type of the road segment	Binary, int
highway_primary_link	Road type of the road segment	Binary, int
highway_secondary	Road type of the road segment	Binary, int
highway_secondary_link	Road type of the road segment	Binary, int
highway_trunk	Road type of the road segment	Binary, int
highway_trunk_link	Road type of the road segment	Binary, int
highway_tertiary	Road type of the road segment	Binary, int
highway_tertiary_link	Road type of the road segment	Binary, int
highway_residential	Road type of the road segment	Binary, int
highway_living_street	Road type of the road segment	Binary, int
highway_unclassified	Road type of the road segment	Binary, int

challenge at this step is the difference in granularity of the two datasets. TimeGeo data provides origin and destination coordinates for individual travelers, while SafeGraph data is aggregated at census block group level. To be able to perform scaling the two data sets need to be at the same granularity. We achieve this by rasterizing the two data sets. The steps followed are listed below:

Step 1: Rasterize the SafeGraph data and label each origin and destination census block group with a raster.

Step 2: Calculate the (%) change in travel demand between the rasterized origin and destinations, $c_{i,j}$, due to SIP orders.

Step 3: Rasterize the TimeGeo travel demand data and calculate the resulting travel demand between the rasterized origin and destinations.

Step 4: Use the $c_{i,j}$ calculated at step 2 to scale the TimeGeo travel demand and get the after shelter in place travel demand estimate between the rasters.

Step 5: Distribute the origin and destination coordinates within their rasters uniformly.

Due to the aggregation-disaggregation involved, we lost information on the home locations of the individual travelers.

S3 Vehicle Use Rates

At the census tract scale, we obtain the population and the vehicle usage rate of residents in that area. The American Community Survey provides commuting data on the level of census tracts (each containing roughly 5000 people) (3). Supplementary Figure **S2** exhibits the administrative boundaries and share of different commute modes within the regions.

To get an estimate of the vehicle usage rates, we use the following relationship:

$$VUR(i) = P_{\text{drive alone}}(i) + \sum_{s=2}^{4} P^{s}_{\text{carpool}}(i)/s, \tag{1}$$

where $P_{\text{drive alone}}(i)$ and $P^{s}_{\text{carpool}}(i)$ are probabilities that residents in zone i drive alone or share a car with s people, respectively.

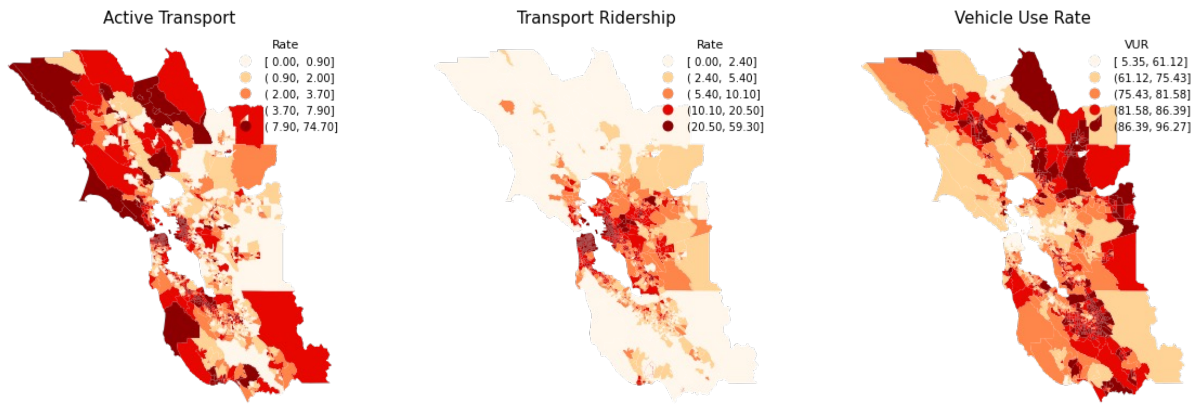

Fig. S2: Commute modes for the Bay Area census tracts derived from 2019 American Community Survey 1-year estimates

S4 Calibration of the StreetSmart Parameters

The StreetSmart model does not provide the energy indices for all fuel economies. In order to predict the fuel consumption of a range of vehicle fuel economies, the first step is to estimate the energy indices specific to each vehicle type.

We calibrated the model and obtained the energy index values using the fuel economies reported by the Environmental Protection Agency's (EPA) 2019 Fuel Economy report (4). The EPA uses the FTP-75, a standard speed profile to test for a car's urban fuel economy, and the HTP-75 to test the highway fuel economy (5). We calibrated the model separately for the urban and highway road types. This study used the idle time, travel time, acceleration and distance variables from the FTP-75 and HTP-75 speed profile with the reported fuel economy to calibrate the energy index ranges for each bin shown in Table **S2**.

We performed validation against Autonomie softwares vehicle fuel consumption simulations (6) for five different vehicle classes available in the software: *Conventional Pickup*, *Conventional Midsize SUV*, *Conventional Small SUV*, *Conventional Compact*, *Conventional midsize*. The resulting mean absolute error percentage is 4.98% and root mean squared error is 0.68

miles per gallon.

S5 Route Assignment and Speed Profiles

Predicting the fuel consumption for each trip in the road network requires a speed profile estimation. Since we don't have the exact routes taken by the travelers, we assign each trip a route in the network given the origin and destination coordinates. In this work we assumed that each traveler will take the route with the shortest travel time. Figure **S3**A illustrates the route assigned to the trip based on the shortest travel time between the origin and destination pairs.

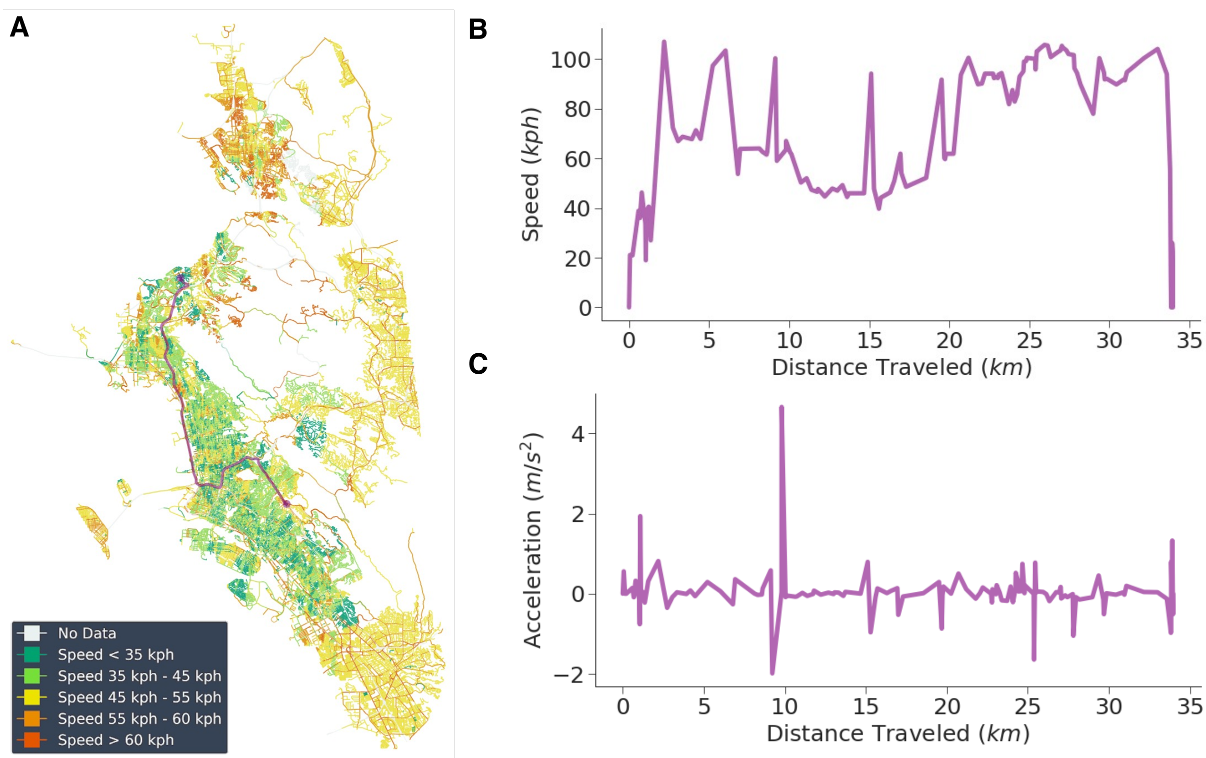

Fig. S3: An illustrative route assignment and speed profiles associated with the route for a trip given origin and destination coordinates

Once we find the path P traversed by the drivers, we extracted the idle time (T_{idle}), travel time (T_{move}), acceleration (a), and distance (L) variables by checking the properties of the edges

Table S2: Calibrated StreetSmart model parameters considering EPA's 2019 Fuel Economy Report. Fuel efficiency (FE) values are in the units of miler per gallon (mpg). The energy index values, k_i, are multiplied by 10^{-5}.

FE	Road Type	k1	k2	k3	k4
5	Highway	4123	117	8	2068
	Urban	94	35	11	2109
7	Highway	3070	87	6	1540
	Urban	68	27	7	1523
9	Highway	2526	72	5	1267
	Urban	55	34	3	1237
11	Highway	2076	59	4	1041
	Urban	46	17	5	1033
13	Highway	1772	50	3	889
	Urban	40	15	5	894
15	Highway	1541	44	3	773
	Urban	35	13	4	776
17	Highway	1379	39	3	691
	Urban	31	11	3	690
19	Highway	1241	35	2	623
	Urban	28	10	3	620
21	Highway	1126	32	2	565
	Urban	25	9	3	567
23	Highway	1026	29	2	514
	Urban	23	9	3	520
25	Highway	952	27	2	478
	Urban	21	8	2	479
27	Highway	883	25	2	443
	Urban	20	7	2	442
29	Highway	824	23	2	413
	Urban	19	7	2	418

FE	Road Type	k1	k2	k3	k4
31	Highway	773	22	1	388
	Urban	18	6	2	393
33	Highway	728	21	1	365
	Urban	16	6	2	368
35	Highway	691	20	1	346
	Urban	15	6	2	344
37	Highway	651	18	1	327
	Urban	15	5	2	326
39	Highway	619	18	1	311
	Urban	14	5	2	309
41	Highway	591	17	1	296
	Urban	13	5	1	295
43	Highway	564	16	1	283
	Urban	13	5	1	282
45	Highway	539	15	1	270
	Urban	12	4	1	270
47	Highway	517	15	1	259
	Urban	12	4	1	259
49	Highway	496	14	1	249
	Urban	11	4	1	251
51	Highway	477	14	1	239
	Urban	11	4	1	238
53	Highway	458	13	1	230
	Urban	10	4	1	230
55	Highway	443	13	1	222
	Urban	10	4	1	223

i and nodes at the end of edge i along the path P. We estimate the T_{idle} by checking the existence of the traffic lights or stop signs. The T_{move} is calculated using the length L_i of the edge i and the speed values v_i extracted from Uber Movement Speeds API. We included the acceleration if there is a stop (ie. trip start, trip end, traffic lights) or a speed change (i.e changing from an edge with speed value v_i to v_{i+1}). The braking time (T_{brake}) is acquired from UK highway code (7).

$$T_{\mathrm{i}}^{idle} = \begin{cases} \mathcal{N}\left(18, 5\right) & \text{if end node of edge i has a traffic light} \\ 3 & \text{if end node of edge i is stop sign} \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in P \qquad (2)$$

$$T_i^{move} = \frac{L_i}{v_{i,}} \qquad \forall i \in P \qquad (3)$$

$$a_i = \begin{cases} |\dfrac{v_i - 0}{T_{brake}}| & \text{if there is a stop} \\ |\dfrac{v_i - v_{i+1}}{\Delta t}| & \text{if there is a speed change} \end{cases} \qquad \forall i \in P \qquad (4)$$

Figure **S3** B and C illustrates the speed profile and acceleration profiles calculated using the Equations 2,3, and 4.

S6 Estimating the Emissions from Gallons Fuel Burned

The validation process requires converting the fuel gallons burned to CO_2 weight ($grams$) and CO_2 flux ($micromole/m^2/s$). The CO_2 weight is used in validation with PEMS-EMFAC model estimates and the CO_2 flux is used validation with $BEACO_2N$ estimate (8). We used the following equations in our conversions:

$$Fuel^g = Fuel^{gal} \cdot 3795 \frac{ml}{gal} \cdot 0.75 \frac{g}{ml} \tag{5}$$

$$C^g = Fuel^g \cdot 0.86 \tag{6}$$

$$CO_2^g = C^g \cdot \frac{44}{12} \tag{7}$$

$$CO_2^{flux} = CO_2^g \cdot \frac{1mole}{44g} \cdot 10^6 \cdot \frac{1}{A} \cdot \frac{1}{\Delta t}. \tag{8}$$

In the calculations, the density of fuel is taken as $0.75g/ml$ and $0.86g$ of C is emitted when $1\ gal$ of fuel is burnt. Note that $1\ mole$ of CO_2 weights $44g$ where $12g$ of it is C. The A in Equation 8 indicates the area.

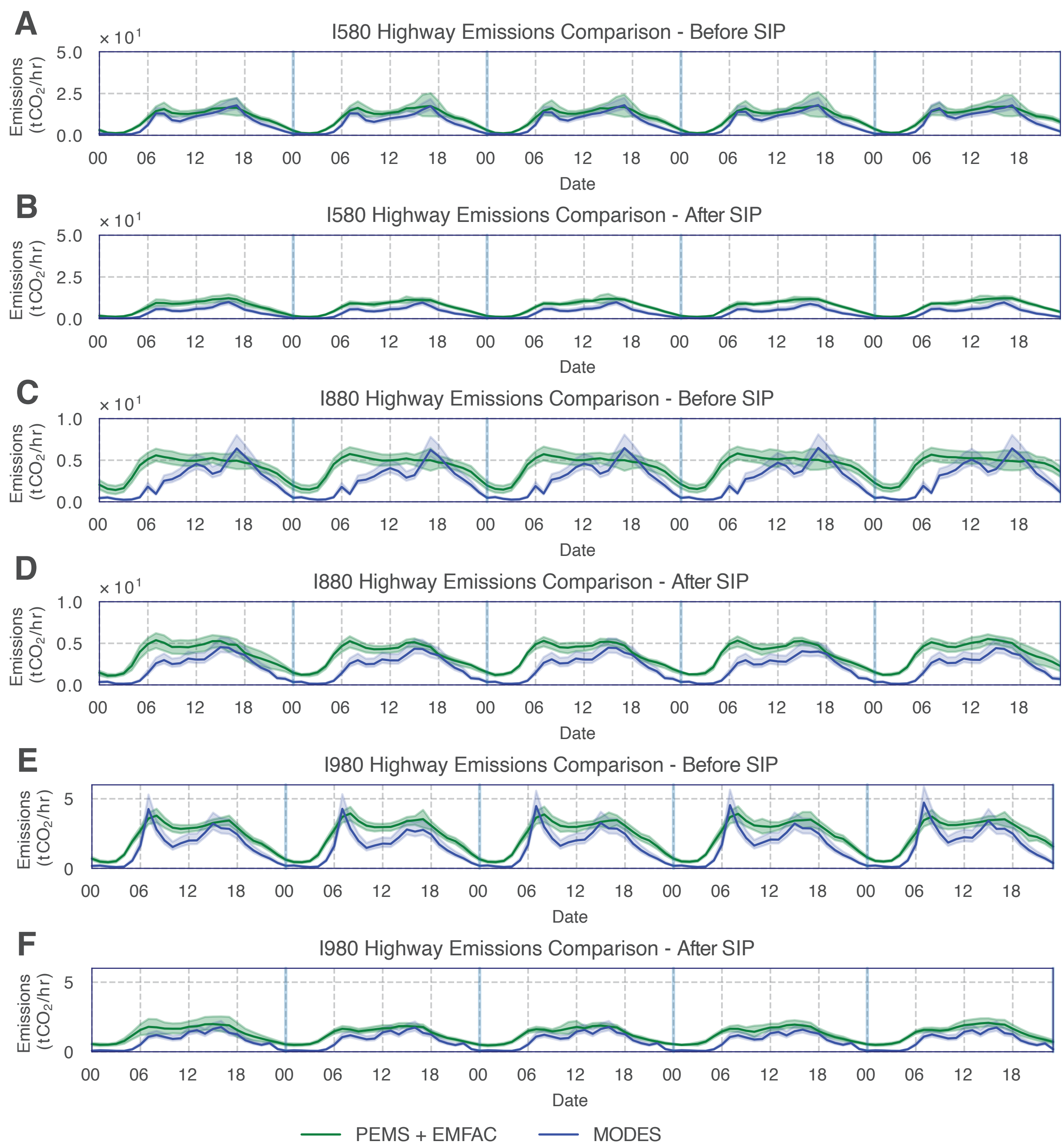

Fig. S4: Comparison of emissions estimates between PEMS-EMFAC model and MODES for I880, I580 and I980. I880 highway emissions before (A) and during (B) the shelter in place decisions. I580 highway emissions before (C) during (D) the shelter in place decisions. I980 highway emissions before (E) during (F) the shelter in place decisions.

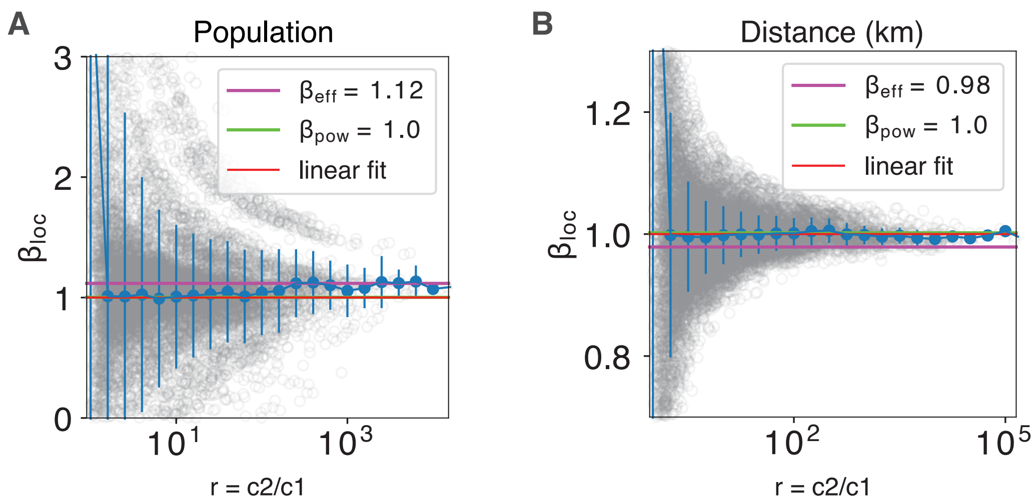

Fig. S5: Tomography of scaling plots for the equation $\beta_{loc} = \frac{\log(Y_2/Y_1)}{\log(X_2/X_1)}$ as defined by Barthelemy et. al (9). where Y_i is the emissions produced by city i and X_i is the independent variable for city i. β_{eff} is the mean of the β_{loc} values for the city with the lowest variance in β_{loc} (9). β_{pow} is the scaling coefficient calculated with regular least squares fitting. (A) Tomography of scaling plots where independent variable is *population* and dependent variable is the emissions. local exponent, β_{loc} converges to $\beta_{eff} = 1.12$ as the ratio $r = \frac{X_2}{X_1}$ increases. (B) Tomography of scaling plots where independent variable is *vehicle kilometers traveled* and dependent variable is the emissions. local exponent, β_{loc} converges to $\beta_{eff} = 0.98$ as the ratio $r = \frac{X_2}{X_1}$ increases.

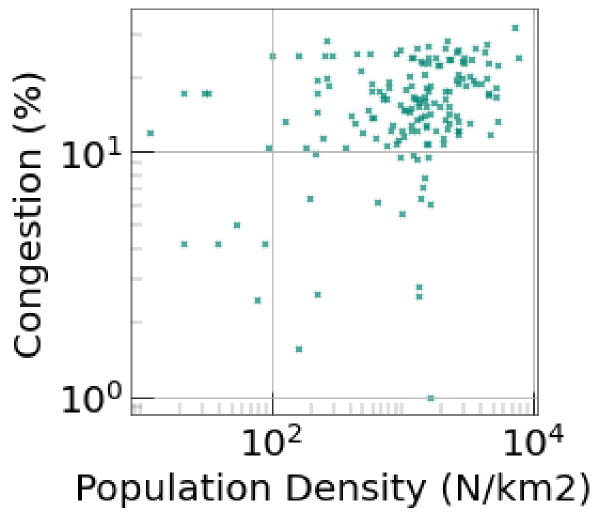

Fig. S6: Relationship between population density of a city and the congestion levels experienced by the city residents. High population density results in higher congestion levels experienced while at low population density the congestion levels are noisy.

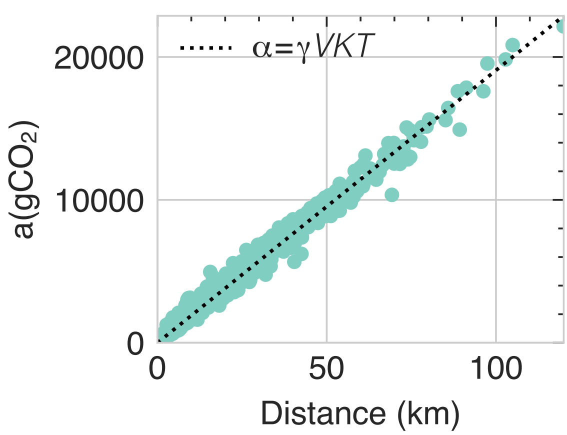

Fig. S7: Relationship between α and VKT for 25 mpg vehicle. The best-fit line is given with the equation $\alpha = \gamma VKT$ where $\gamma = 190.6 \pm 33.3$.

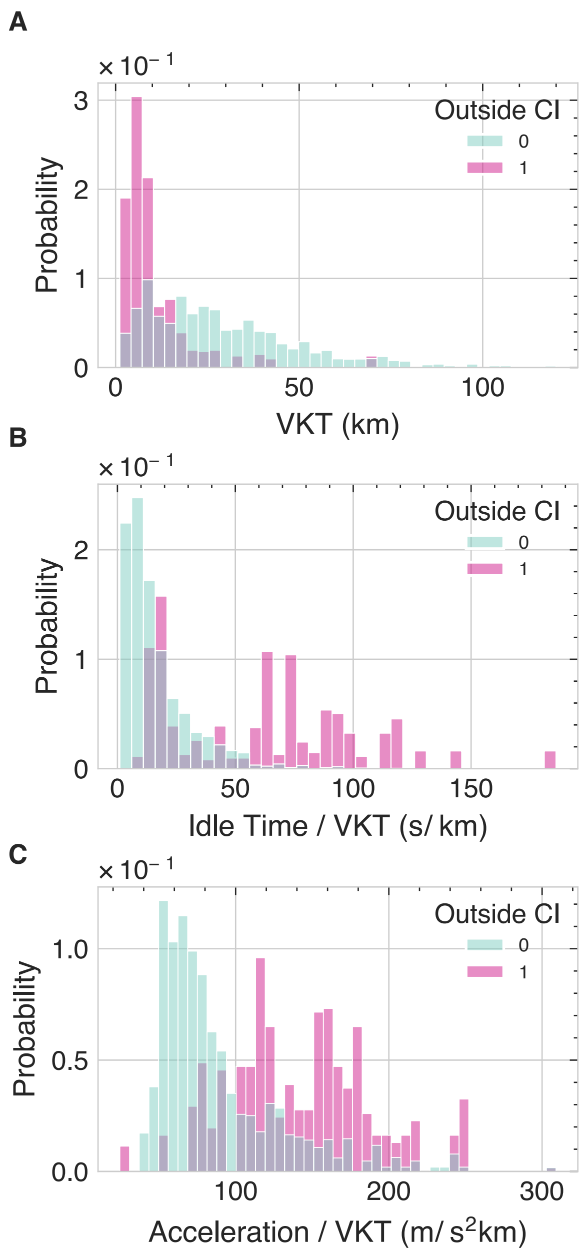

Fig. S8: Characteristics of the trips whose predicted emissions are inside (0) and outside (1) the confidence interval. (A) Distribution of the VKT. Majority of the trips' emissions predicted outside the confidence interval are shorter distance trips. (B) Distribution of the idle time per VKT. Trips whose emissions are predicted outside the confidence interval have a larger idle time to VKT ratio. (C) Distribution of the acceleration per VKT. Trips whose emissions are predicted outside the confidence interval have a larger acceleration to VKT ratio.

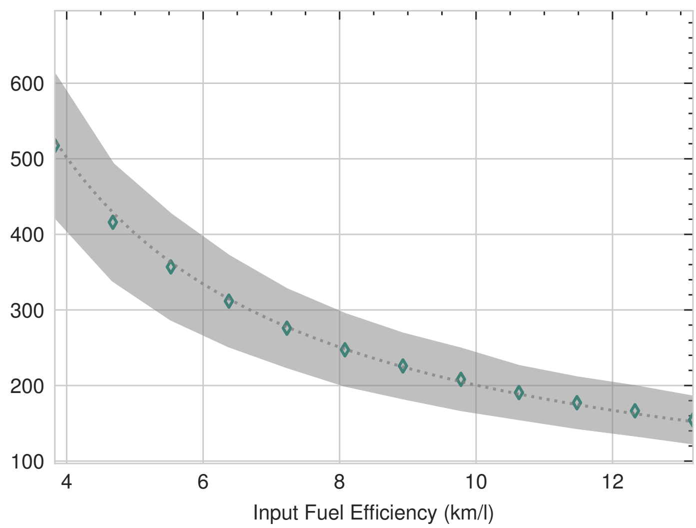

Fig. S9: Power-law relationship between the VKT multiplier, γ, and input fuel efficiency. The best fit line is $\gamma = \frac{2006.53 \pm 312.32}{FE}$. Then the full equation to predict the emissions is as follows: $Y = \frac{2006.53}{FE} VKT_i exp(\beta_c X)$

References

1. Uber movement speeds. https://movement.uber.com/?lang=en-US. Accessed: 2020-03-01.

2. Xu, Y., Çolak, S., Kara, E. C., Moura, S. J. & González, M. C. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nature Energy* **3**, 484–493 (2018).

3. Bureau., U. C. American community survey 1-year estimates. https://data.census.gov/cedsci/table?q=commute&tid=ACSST1Y2019.S0801&moe=false (2019).

4. Fuel economy guide 2019. https://www.fueleconomy.gov/feg/pdfs/guides/feg2019.pdf. Accessed: 2020-03-01.

5. Agency, E. P. Dynamometer drive schedules. {https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules}. Accessed: 2020-03-01.

6. Laboratory, A. N. Autonomie software. https://www.autonomie.net. Accessed: 2020-03-01.

7. Uk highhway code. https://www.highwaycodeuk.co.uk/answers/what-is-the-stopping-and-braking-distance-of-a-car.

8. Fitzmaurice, H. *et al.* Assessing vehicle fuel efficiency using a dense network of co_2 observations. *Atmospheric Chemistry and Physics Discussions* **2021**, 1–15 (2021). URL https://acp.copernicus.org/preprints/acp-2021-808/.

9. Barthelemy, M. Tomography of scaling. *Journal of the Royal Society Interface* **16**, 20190602 (2019).