

# Modeling aggregate human mobility patterns in cities based on the spatial distribution of local infrastructure

Marco Miotti

Urban Informatics Lab, Department of Civil &  
Environmental Eng., Stanford University, USA  
[mmiotti@stanford.edu](mailto:mmiotti@stanford.edu)

Rishee Jain

Urban Informatics Lab, Department of Civil &  
Environmental Eng., Stanford University, USA  
[rishee.jain@stanford.edu](mailto:rishee.jain@stanford.edu)

## Abstract

*Understanding human mobility patterns in urban areas is key to solving a wide range of socio-technical problems at the human-infrastructure interface. Extending the intervening opportunities concept, we showcase a data-driven, network-based model that reproduces aggregate mobility patterns in cities. Using this model, we create a digital replication of daily travel across different trip purposes in 5 U.S. metropolitan areas and compare results against publicly available reference data. We find that our proposed model explains a large fraction of the variation in mean and median travel distance across the 5 cities. In particular, it accurately captures the effect of density on aggregate travel patterns. These findings add to evidence that human mobility patterns are strongly governed by the structure of the built environment. We discuss implications for the ongoing transformation of cities and for developing more sophisticated models that replicate human behavior based on crowd-sourced, spatio-temporal data streams.*

## 1. Introduction

Understanding human mobility patterns in urban areas is key to solving a wide range of socio-technical problems at the human-infrastructure interface. Applications of mobility models include the analysis of urban planning policies [1], trends in transportation technology [2], the simulation of epidemics [3], improving predictions of building energy consumption [4], and the general digital replication of human behavior [5, 6].

While individual mobility choices can be influenced by a large number of factors, aggregate mobility patterns have been found to correlate strongly with properties of the built environment, especially land use [7, 8, 9]. In particular, density, the degree of co-location between possible origins and destinations (through, for instance, mixed-use zoning), and the supply of and access to

different modes of transport influence travel demand and mobility patterns [10, 11]. These findings suggest that human activity is closely linked to properties of the built environment around them, and that modifications to that environment can affect travel activity. They also lead to the notion that travel is mostly a derived demand [9].

To model human mobility as a function of the built environment, many traditional travel demand models rely on four steps [12, 13]. The first two steps are used to estimate origin-destination matrices containing the number of trips or ‘flows’ from one place to another. Steps three and four are used to estimate the specific mode and route for each of those trips. While these four-step models are efficient to set up and run, their aggregate nature makes it difficult to forecast travel choices of individuals in relation to demographic properties.

More recently, activity-based models have been developed to model travel at the individual level by fitting utility functions to travel survey data [7, 14, 15]. While very powerful for certain applications, these bottom-up models require a substantial amount of location-specific data and effort to be set up, calibrated, and run [16, 17]. Enabled by the rise of information and communication technologies, data-driven approaches to travel activity modeling have therefore gained traction [13, 6]. These models tend to exhibit high spatial and temporal resolution and do not rely on expensive travel survey data and detailed information on the built infrastructure. However, they often lack the contextual information that would allow to explain, rather than just observe, human mobility patterns.

Here, we take an approach to replicate mobility patterns in metropolitan areas that combines advantages of the traditional models with opportunities to use novel, crowd-sourced datasets. Specifically, we focus on capturing human-infrastructure interactions and inter-dependencies by reproducing the generation and distribution of trips in the form of an origin-destination flow matrix. We design an algorithm that integrates aspects from the gravity model [18, 19], commonly

used in the four-step approach, with the intervening opportunities concept [20, 21] and the related radiation model [22, 23]. This algorithm allows us to reproduce flow matrices within a metropolitan area from pedestrian trips to long commutes and across different trip purposes. As inputs, it only relies on the features of local infrastructure (i.e., the location and approximate size of homes, jobs, and points of interest, as well as the travel distance between them) and a set of simple coefficients that are independent of the metropolitan area. We calibrate and validate the model using publicly available data, enabling our model to easily scale to a large number of cities across the United States and the world.

We apply this model to create mobility models for 5 U.S. cities across 5 different trip purposes. We analyze how closely the model is able to replicate real aggregate mobility patterns without knowing about congestion, travel times, demographic properties, and other aspects that are commonly included in activity-based travel demand models. Within a given metropolitan area, we evaluate how well the model is able to capture the impact of trip purpose and of key infrastructure features (i.e., density of the built environment) on aggregate travel behavior. We discuss implications of our findings for the ongoing transformation of cities and for building more sophisticated models that replicate behavior at the human-infrastructure interface, including those based on crowd-sourced, spatio-temporal data streams aiming to create ‘digital twins’ of mobility systems.

## 2. Methods

### 2.1. Modeling urban mobility patterns

We model the flows (or frequencies) of travel between each available origin and destination using an extended version of the intervening opportunities model [20]. This model is related to the recently introduced radiation model, both of which have been found to outperform the more common gravity model while requiring less location-specific parametrization [21, 22, 24].

In the original intervening opportunities model, given a set of locations  $U$ , the flow  $F$  of travel from origin  $i \in U$  to  $j \in U$  (or the probability of a trip taking place between  $i$  and  $j$ ) is expressed as [21]:

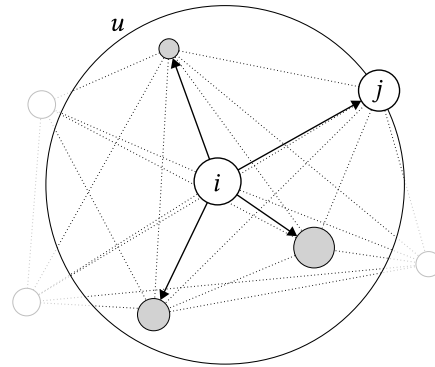
$$r_{i,j} = \left( \sum_{u: d(i,u) \leq d(i,j)} 1 \right)^{-\alpha} \quad (1)$$

$$F_{i,j} \propto \frac{r_{i,j}}{\sum_{j \in U} r_{i,j}}$$

where the  $r_{i,j}$  reflects the number of possible destinations  $u \in U$  that are at least as close to origin  $i$  as destination  $j$ , and  $\alpha$  is a coefficient. The higher  $\alpha$ , the higher the preference for closer destinations compared to those further away.

We model matrix  $F$  for each possible origin-destination pair at the census block group (CBG) level, meaning that origins and destination correspond to CBGs. Operating at the CBG level allows us to keep the number of possible origin-destination pairs under 20 million, and works well with our input data, much of which is also available at the CBG level. Contrary to previous implementations and to more accurately reflect short trips, we do include trips where  $i = j$  to allow for within-CBG travel.

Grouping origins and destinations by CBG also means that destinations can have vastly different levels of ‘significance’: a large shopping mall located in a given CBG will likely attract much more people than a single, small shop. Therefore, inspired by the gravity model [18, 19], we introduce weights  $w$  that reflect the significance of origins and destinations (see Figure 1 and Equation 2).



**Figure 1. Schematic of the extended intervening opportunities concept.** The daily number of trips from origin  $i$  to destination  $j$  depends on the number of possible destinations  $u$  that are at least as close to  $i$  as  $j$  is and their significance  $w$  (illustrated here by the size of the circle). The distance between two nodes is measured as travel distance on the road network.

**Table 1. Share of different types of trips in the 2017 National Household Travel Survey (NHTS, [25]).** The first five types (with a combined share of 75%) are included, modeled separately. POI = point of interest (retail, services, errands, recreational locations, health and child care, religious locations).

Trip purpose	Share	Modeled
Home to work (and reverse)	13%	Yes
Home to POI (and reverse)	38%	Yes
Work to POI (and reverse)	5%	Yes
Home to another home	6%	Yes
POI to another POI	13%	Yes
Work to another work location	2%	No
Other trips	23%	No

$$r_{i,j} = \left( \sum_{u:d(i,u) \leq d(i,j)} w_u \right)^{-\alpha} \quad (2)$$

$$F_{i,j} \propto \frac{r_{i,j}}{\sum_{j \in U} r_{i,j}} \times w_i$$

As another extension to previous implementations of the intervening opportunities model, we consider 5 different types of trips separately: home to work, home to shops and other points of interest, work to points of interest, home to other homes, and points of interest to other points of interest. Points of interest include retail, services, errands, recreational locations, health and child care, and religious locations. Implicitly, these 5 types of trips also reflect the corresponding return trips (such as work to home), and collectively cover 75% of all types of trips taking place within a metropolitan area (Table 1).

Because our model operates on an aggregate level, reproducing the total number of trips for a given purpose between a specific origin-destination pair, we do not need to ensure that a given individual returns to the same home location that they left at before heading to work. We also do not explicitly model tours (that is, a chain of trips containing more than 2 individual destinations).

A fourth and final change to the original equation is that we allow  $\alpha$  to change with the distance between origin  $i$  and possible destination  $u$ ,  $d_{i,u}$ . This modification allows the model to remain accurate for very short and very long trips simultaneously. Equation 2 therefore becomes:

**Table 2. Characteristics of the five metropolitan areas that are being modeled based on their Core-Based Statistical Areas (CBSAs, [26]).** San Diego and Rochester have been expanded to allow for longer trips. Cty = County, CBG = Census Block Group;  $n$  = average number of trips per day and person as measured by the National Household Travel Survey (NHTS, [25]).

	Ctys	CBGs	Population	$n$
San Francisco Bay	7	4013	6,654,837	4.27
Houston	10	3033	6,806,923	4.03
Sacramento	5	1655	2,730,268	4.07
San Diego	3	2920	5,866,335	4.24
Rochester NY	7	927	1,148,335	4.17

$$r_{i,j,k} = \left( \sum_{u:d(i,u) \leq d(i,j)} w_{u,k} \right)^{-\alpha_k(d_{i,u})} \quad (3)$$

$$F_{i,j,k} = \frac{r_{i,j,k}}{\sum_{j \in U} r_{i,j,k}} \times w_{i,k} \times \gamma_k$$

where  $k$  is one of the five modeled trip purposes (see Table 1), and  $\gamma_k$  is a scaling factor that is a function of the trip purpose. Specifically, we set  $\gamma_k$  such that the fraction of trips allocated to a given purposes matches the same fraction as measured in the NHTS, and the total number of trips per person (across all purposes) matches the corresponding number in NHTS as well:

$$\frac{\sum_{i,j} F_{i,j,k}}{\sum_{i,j,k} F_{i,j,k}} = f_k \quad (4)$$

$$\sum_{i,j,k} F_{i,j,k} = n \times \text{population}$$

where  $f_k$  is the share of trip purpose  $k$  (see Table 1), and  $n$  is the number of trips per person and day (see Table 2).

We apply Equations 3 and 4 to the processed input data for the five metropolitan areas shown in Table 2 to predict daily mobility flows for each of the five different trip purposes shown in Table 1. We then aggregate the five flow matrices for the different trip purposes to produce a flow matrix for each city that is intended to reflect the overall daily travel patterns across all types of trips. For each of the 5 metropolitan areas, we select all CBGs containing at least 1 resident, job, or point of interest that belongs to any of the counties

officially associated with the corresponding Core-Based Statistical Area (CBSA, [26]).

For an area with 4,000 CBGs, this results in  $4000^2 \times 5 = 8 \times 10^7$  flows that need to be calculated. We do not model trips that exceed a distance of 100 km, as these rarely occur as regular trips, and could not be properly modeled using the boundaries of the metropolitan areas. The equations are implemented in Python 3.7 using Numpy arrays and the Numba LLVM just-in-time compiler [27]. This implementation allows us to calculate  $8 \times 10^7$  values in under 5 seconds on a conventional computer. The short calculation time and efficient nature of the algorithm reflects an advantage of this modeling approach compared to bottom-up travel demand models.

## 2.2. Input data

Equation 3 requires four sets of input data on local infrastructure: (1) the number of people living in each node or CBG (weight  $w$  if the origin or destination is home), (2) the number of jobs located (weight  $w$  if the origin or destination is work), (3) the significance of points of interest (weight  $w$  if the origin or destination is a point of interest), and (4) the travel distance between each origin-destination pair ( $d_{i,j}$ ).

We obtain the number of people per CBG from the 2018 American Community Survey [28], and the number of jobs from the LEHD Origin-Destination Employment Statistics (LODES 7), which is based on the 2010 census [29]. For the significance of points of interest in each CBG, we aggregate the number of visitors per day from each point of interest contained in a CBG using a proprietary dataset called SafeGraph [30].

To calculate the distance between a given origin node and a given destination node, we use an Open Source Routing Machine (OSRM) server [31], fed with OpenStreetMap data [32]. For each origin-destination pair, we calculate the road travel distance using the corresponding CBG's center coordinates.

For CBGs that are far away from each other, this is likely accurate. For nodes that are close to each other, however, some trips between the those CBGs will be much shorter (with origins and destinations close to the border that's adjacent to the respective other CBG), and some trips will be longer. Therefore, using the distance between the centers to calculate the distance matrix may bias the resulting trip distance distribution. To circumvent this issue, we add or remove a random amount of distance from each origin-destination pair that is equivalent to the square root of the land area of the corresponding origin block group and destination block

group:

$$d_{i,j,adj} = d_{i,j} + \sqrt{A_i} \times r_i + \sqrt{A_j} \times r_j \quad (5)$$

where  $d_{i,j}$  is the road travel distance between the center of CBG  $i$  and the center of CBG  $j$ ,  $A_i / A_j$  are the land areas of the corresponding CBG, and  $r_i / r_j$  are independent random samples from a uniform distribution in the interval  $[-0.5, 0.5]$ .

Notably, this adjustment is only being done once for each origin-destination pair. At the level of an individual pair, the travel distance will therefore not reflect the average distance travel between the two CBG. On aggregate across a large number of pairs, however, the randomization of short trip distances is designed to reflect the real probability distribution of short trips across all of these pairs.

To model trips that take place within a given CBG ( $i = j$ ), we take a similar approach, setting  $d(i, i)$  to a random number that is proportional to the square root of the area of that CBG:

$$d_{i,i} = \sqrt{A_i} \times q_i \quad (6)$$

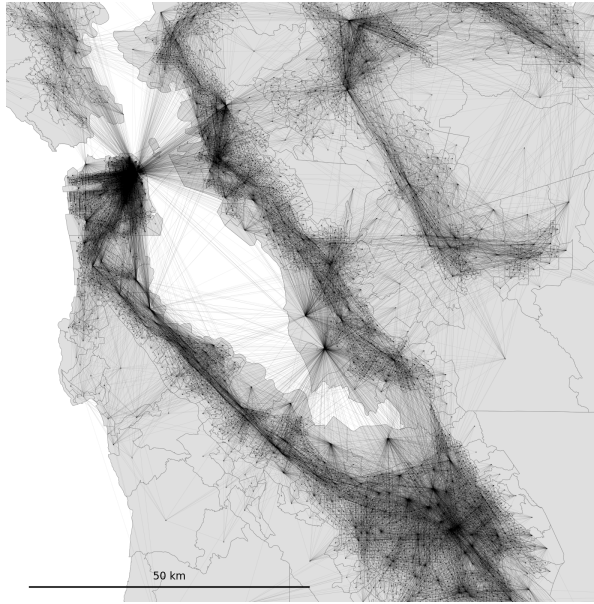
where  $q_i$  is a random sample from a uniform distribution in the interval  $[0.0, 1.0]$ .

## 2.3. Calibration of $\alpha$

We use two sets of metrics to calibrate the model coefficients,  $\alpha_k(d)$ . The first set of metrics consists of aggregate trip distance metrics compared against conventional travel survey data. The second set measures the similarity of the calculated flow matrices  $F$  for home to work trips against reference data from LODES. Rather than numerically fitting  $\alpha_k(d)$  to minimize the error compared to reference data, we calibrate the coefficients manually until a good fit across all metrics, cities, and trip purposes is achieved. Numerically fitting  $\alpha_k(d)$  could lead to over-fitting, and would still require subjective judgement about the weight of different metrics.

For the aggregate metrics, we compare aggregate statistics of the calculated flow matrix  $F$  to corresponding metrics in the NHTS. We compare the median trip distance, the mean trip distance, and the probability density function of trip distances for each trip purpose and for all trips combined. We predominantly use this method to calibrate  $\alpha$  as a function of trip purpose and trip distance, as NHTS is the more representative of the two reference data sets, and available for all trip purposes.

For comparing raw flow data, we compare the calculated flow matrix for commuter trips (home to work



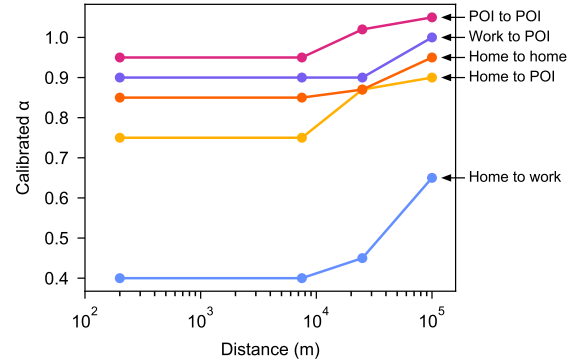
**Figure 2. Modeled daily mobility flows between census block groups (CBGs) across the San Francisco Bay Area.**

trips) to the corresponding matrix in LODES, which is also collected at the CBG level. In addition to comparing flows at a CBG level, we aggregate them to the zipcode level, summing over all trips that originate in any CBG belonging to a given origin zipcode and end in any CBG belonging to a given destination zipcode. This aggregation circumvents issues with the relatively sparse nature of the LODES flow matrix, and allows us to identify whether the accuracy of our model is increased at a lower spatial resolution. To calculate the similarity between the modeled flow matrix and the reference matrix either at the CBG or at the zipcode level, we use the common parts per commuter (CPC) metric, indicating the similarity between two origin-destination matrices  $F$  and  $G$  as a number between 0 and 1 [23]:

$$CPC_{F,G} = \frac{2 \times \sum_{i,j} \min(F_{i,j}, G_{i,j})}{\sum_{i,j} F_{i,j} + \sum_{i,j} G_{i,j}} \quad (7)$$

Before applying Equation 7, we normalize both matrices  $F$  and  $G$  to each sum to the same amount. We predominantly use this metric to validate our model, rather than for calibration purposes, meaning that we calibrate  $\alpha$  using aggregate metrics as described above, and then test its performance against disaggregate flow data.

### 3. Results

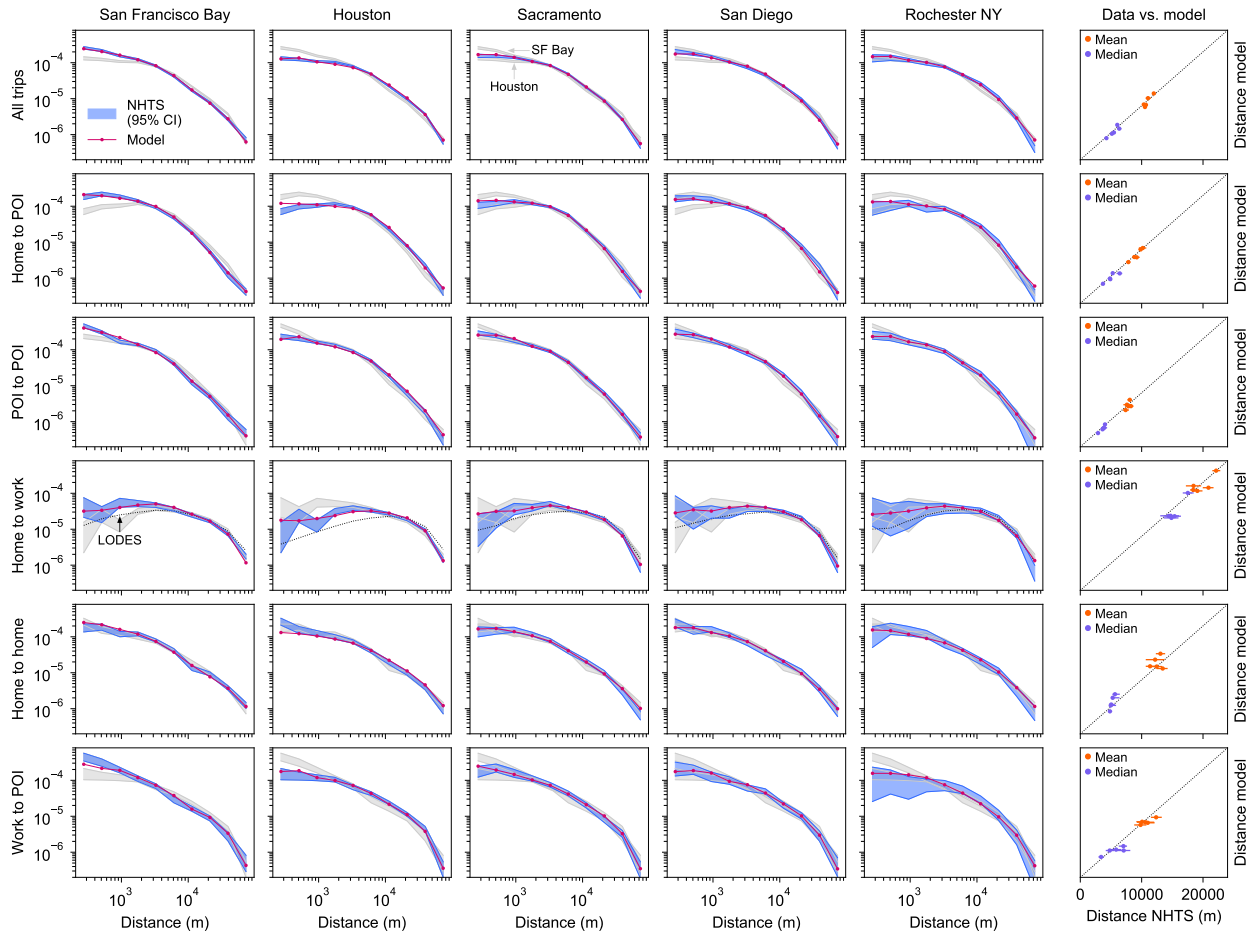


**Figure 3. Calibrated  $\alpha$  as a function of trip purpose and trip distance. The higher the value of  $\alpha$ , the higher the preference for destinations that are closer as opposed to destinations that are further away. POI = point of interest (retail, services, errands, recreational locations, health and child care, religious locations).**

We apply Equations 3 and 4 to the five selected metropolitan areas (Table 2) to reproduce daily mobility flows for the five different trip purposes (Table 1) after calibrating  $\alpha$ . A visual example of the resulting daily mobility flows is shown in Figure 2.

The calibrated  $\alpha$  for home to work trips is the lowest, meaning that people are willing to travel further (exhibit a lower preference for closeness of the destination) for trips to work than for the other trip types (Figure 3). For trips from home to points of interest such as retail, services, and recreational activities, the opposite is true.  $\alpha$  starts to increase for trips longer than 7.5 km across all types of trips. This indicates that above that threshold, the fixed relationship between the number of destination opportunities and the probability of traveling to a given destination suggested by the original intervening opportunities model starts to deviate.

With the calibrated  $\alpha$ , our model reproduces aggregate travel patterns in the 5 metropolitan areas to a high degree of accuracy (Figure 4). Across all trip purposes, the model is able to explain a substantial amount of the variance in median and average trip distance between the five cities. The explanatory power is highest for trips from home to points of interest (POIs), which represent the most common trip purpose (see Table 1) and exhibit the highest amount of variation between the five cities. The model can only explain little of the variation in mean distance from POI to POI and from homes to another home. Notably, however, those means exhibit less variation than for other trip types,



**Figure 4. Columns 1-5: Trip distance probability density functions (PDFs) across the five trip purposes (see Table 1) and five metropolitan areas (see Table 2). 95% confidence intervals for the reference data (2017 National Household Travel Survey [25], blue) are calculated using the bootstrap, resampling 200 new datasets of the same length from the original data. The dotted black line in the 4th row reflects the PDFs as inferred from the LEHD Origin-Destination Employment Statistics (LODES 7 [29]). Column 6: Predicted (y-axis) and measured (x-axis) mean and median trip distance for each of the 5 metropolitan areas.**

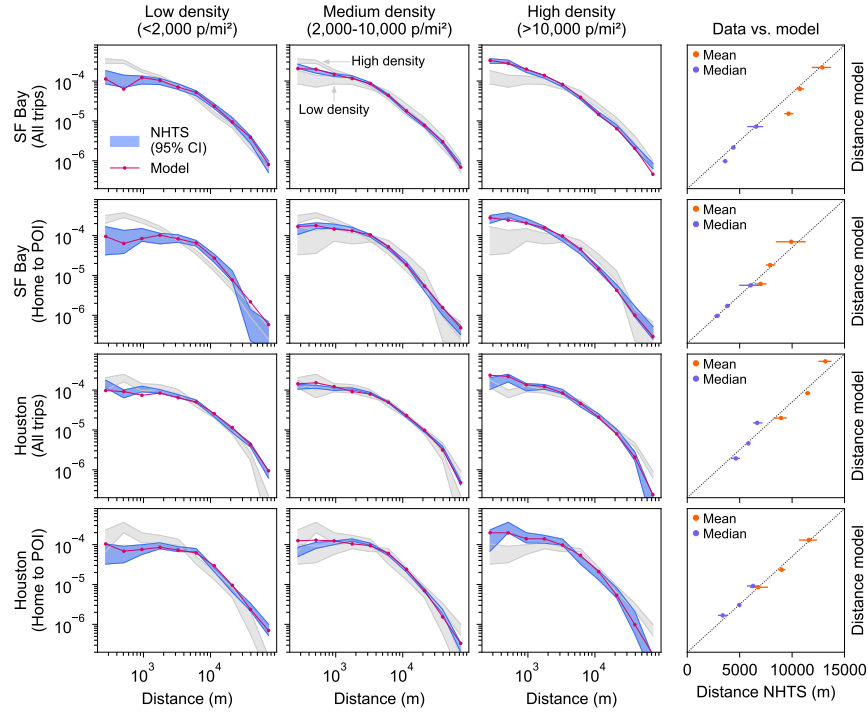
and are subject to larger confidence intervals due to the smaller amount of corresponding available data.

The detailed trip distance probability density functions (PDFs) generated by the model match the corresponding distributions as measured in the NHTS well, generally falling within the range of the 95% confidence intervals of the latter. The largest deviations can be observed at the extreme ends of the distribution, for very short trips (less than 1 km) and very long trips longer than 40 km). We also compare the PDF for home to work trips to the distribution in trip distances as predicted by LODES (Figure 5, 4th row). Compared to LODES, our model predicts more frequent short trips, and less frequent longer trips. Notably, however, LODES does not measure the trips that people actually

make; rather, LODES indicates the relationship between where people live and where people work, regardless of how often people actually travel between their work location and their home location, and whether they do so directly.

When comparing the disaggregate flow matrix for home to work trips predicted by the model to the same matrix from LODES, we observe a high level of similarity at the CBG level (0.52–0.59), and a very high level of similarity at the zipcode level (0.77–0.83, Table 3). The CPC value is similar for all five cities, indicating that the model—despite not having any city-specific calibration parameters—performs almost equally well for all modeled locations. The CPC could be further increased if  $\alpha$  was calibrated to maximize it. As noted





**Figure 5. Columns 1-3: Trip distance probability density functions (PDFs) for all trips as well as trips from home to points of interest (see Table 1) that take place within, leave from, or arrive at CBGs whose average population density falls into one of three specific brackets. Confidence intervals are calculated using the same approach as in Figure 4.**

previously, however, LODES does not measure the trips that people actually make, implying that calibrating  $\alpha$  to fit LODES data may not accurately reflect the daily travel choices that people make.

Finally, we classify all CBGs of the San Francisco Bay and Houston metropolitan areas into brackets of low ( $<2,000$  people/mi<sup>2</sup>), medium (2,000–10,000 people/mi<sup>2</sup>), and high ( $>10,000$  people/mi<sup>2</sup>) population

**Table 3. Similarity of the origin-destination flow matrix for home to work trips between our model and the LEHD Origin-Destination Employment Statistics (LODES 7 [29]). CPC = Common Part Per Commuter (see Equation 7); 0 = no similarity; 1 = perfect correlation.**

	CPC by spatial resolution	
	Census block group	Zipcode
San Francisco Bay	0.53	0.80
Houston	0.52	0.77
Sacramento	0.53	0.80
San Diego	0.54	0.77
Rochester	0.59	0.83

density. We then calculate aggregate trip statistics for each of the three population density brackets, including only trips whose origin and/or destination lies within the corresponding bracket. We compare those statistics to trips in NHTS where the household of the person making the trip is located in a CBG that falls within the same population density bracket. We compare statistics for all trips combined, as well as home to work trips, which represent the most common trip type (Table 1) and show the highest amount of variation between areas (see also Figure 4). For other trip types, the coefficient of variation of average travel distance relative to the number of available data points was too small to reliably assess the performance of our model.

Our model reproduces the impact of population density on travel behavior to a high degree of accuracy (Figure 5). As before, the predicted PDFs fall within the range of the confidence intervals generated from NHTS data. Deviations occur only for short trips ( $<1$  km) and long trips ( $>40$  km). In particular, our model is able to explain almost all of the variation in average and median trip distance between low, medium, and high density areas in both cities, for all trips combined as well as for home to point of interest trips specifically.

## 4. Discussion and Implications

We find that a large part of the heterogeneity in human mobility patterns between different metropolitan areas can be explained solely based on information regarding the local infrastructure (i.e., the location of different origins and destinations, and the travel distance between them). These findings add to evidence that human mobility patterns are strongly governed by the location of origins and destinations, and illustrates the inter-dependency between human behavior and infrastructure development. Our findings reaffirm that the overall demand for mobility can be lowered by increasing density and by fostering mixed-use zoning and development. They also confirm earlier findings [21] that suggest that models of the intervening opportunities class work across different metropolitan areas without the need for recalibration, implying that there is a certain universality to human mobility patterns.

We observe that our model is particularly accurate at reproducing the impact of population density on travel behavior. In the model, density is encoded as the general urban structure around the locations and between them. As result, our approach has advantages over previously proposed regression-type models that include one or more metrics for density and/or diversity in each location as predictors (e.g. [10, 33, 11]). These models are often not able to distinguish between a case where the density right around the origin is high, but falls off quickly, and a case where the density around the origin is generally high. Some statistical models use more sophisticated predictors representing urban form to mitigate this issue (e.g., [34]), but such predictors still reflect a substantially more indirect way of measuring the relationship between density or diversity and mobility patterns. In addition, the mechanistic nature of our model may make it less prone to self-selection effects (see e.g. [33]).

As such, our findings reiterate the need to capture interactions and inter-dependencies between human mobility and physical infrastructure. They may imply, for instance, that increasing density substantially in a small area of a city is less effective at reducing travel demand than increasing it less substantially, but across a larger area. The specific relationships between urban form and travel demand as implied by the model presented here, however, require further investigation and verification across more cities both in the United States and across the world.

The overall accuracy with which our model is able to reproduce aggregate travel patterns illustrates how little information is needed to explain a large

amount of the variation in different travel patterns across different types of trips, cities, and areas within these cities. Congestion, travel times, travel costs, access to different modes of transport, and demographic properties are important for the precise route or modal split in relation to demographic properties, but may only play a subordinate role in shaping aggregate travel patterns, where they can largely be captured by simple coefficients ( $\alpha$  in our case). While there is evidence that the intervening opportunities concept works well in different cities across the planet [21], the accuracy of our model may fall off if it was applied to other countries without re-calibrating  $\alpha$ . We expect this to be particularly true in places where the average travel speed for trips of a given distance and/or the (financial or geographical) access to different modes of transportation is substantially different from U.S. locations.

The increasing amount of available crowd-sourced data may further expand the potential of high-level, infrastructure-based travel models such as the one proposed here. These types of data could serve both to increase the model sophistication without sacrificing its key advantages and to allow for more refined and accurate calibration and validation. Such efforts may be able to complement or even partially replace detailed, complex, and expensive utility-based bottom-up travel models to assess the impact of infrastructure and technology transformations on mobility behavior, from the city to the local street level.

While travel times and congestion levels may not have been necessary to reproduce aggregate travel patterns shown here, such information could be used to add a temporal dimension to the model. As one possible application, this would allow for the combination with building models to add to a growing body of literature that investigated the relationship between building energy use and mobility patterns [4, 35].

A spatio-temporal map of mobility patterns, generated by a version of our model that has been expanded by a temporal dimension, could be compared to and updated by real-time travel information. This would enable us to detect deviations from the typical travel demand and could allow for proactive rather than reactive traffic management and congestion pricing. More generally, through the integration with other urban models and novel data streams, frameworks such as the one proposed here could form a comprehensive ‘digital twin’ of human mobility, capable of evaluating various urban development and emission reduction scenarios across different scales and sectors in cities around the globe.



## 5. Acknowledgements

The research presented in this paper was supported in part by a TomKat Center Postdoctoral Fellowship in Sustainable Energy, a Swiss National Science Foundation (SNSF) Postdoctoral Fellowship, and the U.S. National Science Foundation (NSF) under Grant No. 1941695. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. NSF.

## References

- [1] P. W. G. Newman and J. R. Kenworthy, "The transport energy trade-off: Fuel-efficient traffic versus fuel-efficient cities," *Transportation Research Part A: General*, vol. 22, no. 3, pp. 163–174, 1988.
- [2] M. Salazar, F. Rossi, M. Schiffer, C. H. Onder, and M. Pavone, "On the Interaction between Autonomous Mobility-on-Demand and Public Transportation Systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2262–2269, IEEE, nov 2018.
- [3] B. D. Dalziel, B. Pourbohloul, and S. P. Ellner, "Human mobility patterns predict divergent epidemic dynamics among cities," *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1766, pp. 1–6, 2013.
- [4] N. Mohammadi and J. E. Taylor, "Urban energy flux: Spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction," *Applied Energy*, vol. 195, pp. 810–818, 2017.
- [5] M. C. González, C. A. Hidalgo, and A. L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [6] L. Pappalardo and F. Simini, *Data-driven generation of spatio-temporal routines in human mobility*, vol. 32. Springer US, 2018.
- [7] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 1, pp. 1–28, 2000.
- [8] S. Handy, "Smart growth and the transportation-land use connection: What does the research tell us?," *International Regional Science Review*, vol. 28, no. 2, pp. 146–167, 2005.
- [9] B. van Wee and S. Handy, "Key research themes on urban space, scale, and sustainable urban mobility," *International Journal of Sustainable Transportation*, vol. 10, no. 1, pp. 18–24, 2016.
- [10] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: Density, diversity, and design," *Transportation Research Part D: Transport and Environment*, vol. 2, pp. 199–219, sep 1997.
- [11] S. Hankey and J. D. Marshall, "Impacts of urban form on future US passenger-vehicle greenhouse gas emissions," *Energy Policy*, vol. 38, no. 9, pp. 4880–4887, 2010.
- [12] M. G. McNally, "The Four Step Model," in *Handbook of Transport Modeling* (D. A. Hensher and K. J. Button, eds.), ch. 3, Irvine, CA, USA: Pergamon, 2nd ed., 2008.
- [13] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 162–177, 2015.
- [14] M. Balmer, K. W. Axhausen, and K. Nagel, "Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1985, pp. 125–134, 2006.
- [15] M. Adnan, C. Author, F. C. Pereira, C. Miguel, L. Azevedo, K. Basak, M. Lovric, S. Raveau, Y. Zhu, J. Ferreira, C. Zegras, and M. E. Ben-akiva, "SimMobility : A Multi-Scale Integrated Agent-based Simulation Platform," no. January, pp. 1–18, 2015.
- [16] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, and M. C. González, "The TimeGeo modeling framework for urban motility without travel surveys," *Proceedings of the National Academy of Sciences*, vol. 113, pp. E5370—E5378, sep 2016.
- [17] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.
- [18] S. Erlander and N. F. Stewart, *The Gravity Model in Transportation: Theory and Extensions*. VSP, 1990.
- [19] L. Pappalardo, S. Rinzivillo, and F. Simini, "Human Mobility Modelling: Exploration and Preferential Return Meet the Gravity Model," *Procedia Computer Science*, vol. 83, pp. 934–939, 2016.
- [20] S. A. Stouffer, "Intervening Opportunities: A Theory Relating Mobility and Distance," *American Sociological Review*, vol. 5, p. 845, dec 1940.
- [21] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: Universal patterns in human urban mobility," *PLoS ONE*, vol. 7, no. 5, 2012.
- [22] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, pp. 96–100, feb 2012.
- [23] Y. Yang, C. Herrera, N. Eagle, and M. C. González, "Limits of Predictability in Commuting Flows in the Absence of Data for Calibration," *Scientific Reports*, vol. 4, pp. 1–9, may 2014.
- [24] J. P. Bagrow and Y. R. Lin, "Mesoscopic structure and social aspects of human mobility," *PLoS ONE*, vol. 7, no. 5, pp. 3–8, 2012.
- [25] FHWA, "2017 National Household Travel Survey," 2018.
- [26] U.S. Patent and Trademark Office, "Patenting In U.S. Metropolitan and Micropolitan Areas Regional Components January 2000 – December 2015 Reports," 2020.
- [27] S. K. Lam, A. Pitrou, and S. Seibert, "Numba," in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15*, (New York, New York, USA), pp. 1–6, ACM Press, 2015.
- [28] United States Census Bureau, "2018 American Community Survey (ACS)," 2020.
- [29] United States Census Bureau, "Longitudinal Employer-Household Dynamics Data," 2020.
- [30] SafeGraph, "POI Data, Business Listings, & Foot-Traffic Data," 2020.

- [31] “Project OSRM,” 2020.
- [32] Open Street Map Foundation (OSMF), “OpenStreetMap,” 2020.
- [33] B. B. Zhou and K. M. Kockelman, “Self-Selection in Home Choice,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2077, pp. 54–61, jan 2008.
- [34] J. Rajamani, C. R. Bhat, S. Handy, G. Knaap, and Y. Song, “Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1831, pp. 158–165, jan 2003.
- [35] E. Barbour, C. C. Davila, S. Gupta, C. Reinhart, J. Kaur, and M. C. González, “Planning for sustainable cities by estimating building occupancy with mobile phones,” *Nature Communications*, vol. 10, no. 1, 2019.