Door-to-door Air Travel Time Analysis in the United States using Uber Data

Philippe Monmousseau, Daniel Delahaye, Aude Marzuoli, Eric Féron

Door-to-door Air Travel Time Analysis in the United States using Uber Data

Philippe Monmousseau, Daniel Delahaye
Optimization and Machine Learning Group
ENAC, Université de Toulouse
Toulouse, 31055, France
philippe.monmousseau@enac.fr

Aude Marzuoli, Eric Feron
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250, USA
amarzuoli3, feron@gatech.edu

Abstract—NextGen and ACARE Flightpath 2050 set some ambitious goals for air travel, including improving the passenger travel experience using door-to-door travel times as a possible metric. Using recently released Uber data along with other online databases, a reliable estimation of door-to-door travel times is possible, which then enables a comparison of cities performance regarding the good integration of their airports as well as a per segment analysis of the full trip. This model can also be used to better evaluate where progress should and can be made with respect to air passenger travel experience.

Keywords—Door-to-door travel times, Big Data, Air Transportation System, Multi-modal travel

I. Introduction

Seamless door-to-door travel and data sharing was deemed as needed by the European Commission's 2011 White Paper [1] and was reconfirmed by the Federal Aviation Administration (FAA) in 2017 [2]. Data sharing was already a main focus in the early 2000s and led at an air system level to the creation of the architecture SWIM - System Wide Information Management [3] - by Europe and later adopted by the FAA. NextGen [4] in the United States and ACARE Flightpath 2050 [5] aim to take a more passenger-centric approach, with ACARE Flightpath 2050 setting some ambitious goals, including some that are not measurable yet due to lack of available data. Regarding door-to-door travel times, it aims at having 90% of travellers within Europe being able to complete their door-to-door journey within 4 hours. In the US, the Joint Planning and Development Office has proposed and tested metrics regarding NextGen's goals, but there are still metrics missing from the passenger's veiwpoint, especially regarding door-to-door travel times [6]. The shift from flight-centric information to passenger-centric metrics was first explored by Cook et al. [7] within the project POEM - Passenger Oriented Enhanced Metrics, where they designed propagation-centric and passenger-centric performance metrics, and compared them with existing flight-centric metrics.

Door-to-door travel time estimation with a multi-modal approach has been previously studied but for travels contained within the same metropolitan area. Peer et al. [8] studied door-to-door travel times and schedule delays for daily commuters in a Dutch city, showing the importance of considering the correlation of travel times across different road links when estimating the overall travel time. Salonen and Toivonen [9] investigated the need of comparable models and measures for trips by car or public transport within Helsinki, introducing a multi-modal approach when considering the walking and waiting necessary to reach a station or a parking spot. Duran-Hormazabal and Tirachini [10] focused on travel time variability for multi-modal trips within Santiago, Chile, using both GPS data and surveyors to estimate the time spent in the different considered modes (walking, car, bus and metro). From an air travel perspective, Pels et al. [11] focused on the relative importance of the access time to airports in the passengers choice with a case study on the San Francisco Bay Area airports. These studies emphasized the importance of considering all relevant modes when estimating door-to-door travel times, but were limited in scope by the area considered and the data available.

Larger scale studies with a focus on air transportation was later possible thanks to the increasing use of mobile phone devices as datasources. In the United States, Marzuoli et al. [12] presented a method to detect domestic air passengers on a nationwide scale using mobile phone data, enabling a per leg analysis of the full door-to-door trip though the main focus was on passengers' behavior at airports. The passengers' experience in airports under major perturbations using this method and additional data from social media was further studied in [13]. In Europe, within the BigData4ATM project[1], García-Albertos et al. [14] presented a methodology for measuring the door-to-door travel time using mobile phone data and applied it between two Spanish cities, Madrid and Barcelona. Mobile phone data is however proprietary data and difficult to access for research.

Sun et al. [15] implemented a door-to-door minimum travel time estimation based on open source maps and datasets in order to study the possible competitiveness of air taxis. This study is also based on already available online data but aims in creating a method to measure the actual average door-to-door travel time once the trips are over enabling an analysis and comparison of the different modes. It adapts a method already used in a to-be-published work [16] applied to two intra-European multi-modal trips comparing air to rail. This

[1] www.bigdata4atm.eu

method is here improved by leveraging four different data sources (road data, flight data, phone data and census data) and is applied to compare air trips between five different cities in the United States, three on the West Coast and two on the East Coast.

This paper is organized as follows: Section II presents the model and data used to evaluate the full door-to-door journey time. Section III showcases several applications and comparisons enabled by this model and Section IV concludes this paper and discusses further research directions.

II. MODEL

In the specific case of air travel and similarly to [14] and [15], the full door-to-door travel time can be decomposed into the five following times:

$$T = t_{to} + t_{dep} + t_{in} + t_{arr} + t_{from} \qquad (1)$$

where

- t_{to} is the time needed to go from the start of the journey to the departure airport
- t_{dep} is the time spent waiting and going through security processes at the departure airport
- t_{in} is the time actually spent in flight
- t_{arr} is the time spent going through security processes at the arrival airport
- t_{from} is the time needed to go from the arrival airport to the end of the journey

The measurement and estimation of these different times are described in the upcoming subsections. This study limits its scope, due to data availability, to the following five airports and associated cities: Boston's Logan International Airport (BOS), Ronald Reagan Washington National Airport (DCA), Los Angeles International Airport (LAX), Seattle-Tacoma International Airport (SEA) and San Francisco International Airport (SFO).

A. Travel time from origin location to departure airport and from arrival airport to final destination

Uber, a ride-sharing service implanted in major urban areas on six continents, recently released anonymized travel times data for certain of these urban areas, including the five considered US metropolitan areas. This data consist of the average travel time between zones (namely census tracts for US cities) within the serviced area from all Uber rides aggregated over each considered day. Depending on the availability of data, five additional different periods are considered:

- Early: from midnight to 7am
- AM: from 7am to 10am
- Midday: from 10am to 4pm
- PM: from 4pm to 7pm
- Evening: from 7pm to midnight

Before this data release, several studies were already conducted on the impact of Uber in urban transit, mainly in US metropolitan areas, Uber being initially introduced in the US. Li et al. [17] concluded that on average Uber tends to decrease

congestion in US urban areas where it was introduced. Hall et al. [18] studied the impact of Uber on the use of public transit system based on Uber's entry date in different cities and focused on whether Uber complemented or substituted public transit. Wang and Mu [19] studied Uber's accessibility in Atlanta, US by using the average wait time for a ride as a proxy and concluded that the use of Uber was not associated to a specific social category. Uber rides being part of the road traffic flow, this study considers Uber's travel times as accurate proxies of the actual travel time by car. The scope of this paper limits itself to road access and egress to the considered airports, though subway alternatives should be considered by using schedules and real time data for a more exhaustive analysis of these legs of multi-modal trips.

For this study, data was gathered using Uber's Movement API[2] over the period of January 1st 2018 to March 31st 2018. Each city was divided into their census tracts.

B. Dwell time at airports

The time spent at each airport depends on the airport, the specific flight and whether it is at boarding or unboarding process. It can be split into two component: a processing time t_{sec} necessary to get through security and through the airport to the desired gate and an extra wait time t_{wait} due to flight delays. The processing times are based on the average wait times at airports presented in the study of [12]. They are summarized in the following table:

TABLE I: TIME SPENT AT AIRPORTS

	BOS	DCA	LAX	SEA	SFO
Time at departure (min)	105	100	125	105	105
Time at arrival (min)	40	35	65	50	45

The extra wait times are based on the data published by the Bureau of Transportation Statistics (BTS) [20] and were calculated only for departure. They were calculated by subtracting the actual flight departure time with the scheduled departure time. No extra wait time was assumed at arrival in this study.

C. Time in flight

The actual flight time was also calculated using data from BTS using the actual departure and arriving times of all direct flights between each city pairs over the first three months of January. For this study, cancelled flights were not considered. Further studies should consider airline policies in order to estimate the impact of a cancelled flight on the full door-to-door time.

The daily average number of direct flights per day period are presented in Table II. This table does not count flights that were cancelled during the three month period considered. Over the chosen period, 25,937 flights were considered, corresponding to 2,405 early flights, 5,475 morning flights, 8,336 midday flights, 4,934 afternoon flights and 4,787 evening flights.

[2]movement.uber.com

| Segment	Early	AM	Midday	PM	Evening
BOS - DCA	3.71	13.55	22.19	11.19	8.74
BOS - LAX	2.07	9.87	8.13	12.42	2.35
BOS - SEA	1.08	6.55	0.00	3.16	2.42
BOS - SFO	5.35	11.29	7.94	5.65	5.00
DCA - BOS	3.55	9.16	24.13	12.03	10.65
DCA - LAX	0.00	5.68	0.00	5.74	0.00
DCA - SEA	0.00	2.84	0.00	2.84	0.00
DCA - SFO	0.00	2.90	0.00	2.84	0.00
LAX - BOS	2.94	10.68	8.42	2.15	10.74
LAX - DCA	0.00	5.74	5.74	0.00	0.00
LAX - SEA	6.65	10.39	24.55	13.97	14.32
LAX - SFO	7.68	18.65	44.19	23.65	22.61
SEA - BOS	2.52	1.33	4.48	0.00	4.94
SEA - DCA	0.00	2.84	2.87	0.00	0.00
SEA - LAX	12.74	11.32	20.42	11.77	13.48
SEA - SFO	9.39	11.35	25.00	9.23	14.55
SFO - BOS	1.04	8.16	13.55	4.81	7.94
SFO - LAX	10.45	26.61	35.52	23.68	20.87
SFO - SEA	8.94	8.00	21.77	14.32	15.81

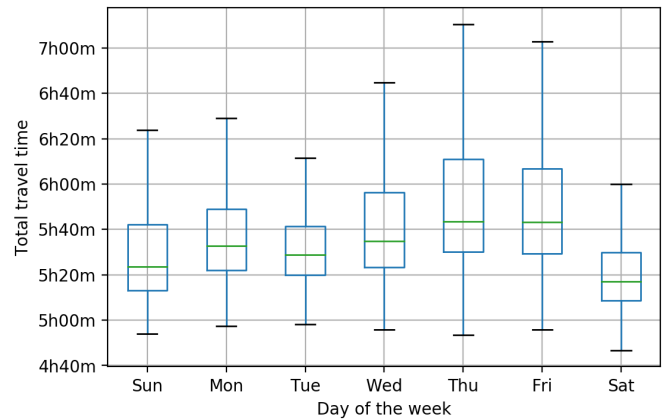Figure 1: Boxplots of door-to-door travel times from Los Angeles to San Francisco grouped by day of the week.

Figure 2: Boxplots of door-to-door travel times from San Francisco to Los Angeles grouped by day of the week.

D. Full door-to-door time

For this study, it was assumed that travellers planned their departure time to arrive at the departure airport exactly t_{sec} minutes before the scheduled departure time. This assumption is necessary to determine which period of the day to consider when retrieving the Uber average time between the initial zone and the departure airport to determine t_{to}. For days and zones where only daily aggregates were available in the Uber data, the daily aggregated times were used for each period of the day as a proxy. The same process was implemented to select the value of t_{from} based on the actual arrival time of the flight. The travel times were calculated for each scheduled flight from January 1st 2018 to March 31st 2018 if the flight was not cancelled.

III. APPLICATIONS

A. Overall trip comparison

Once the travel times for each flight are extracted for every census tract where there has been enough Uber trafic to or from the airport, it is possible to aggregate these times at a city level in order to obtain a single travel time per flight. A first way of aggregating would be using the actual distribution of passengers over the different census tracts, possibly using mobile phone data. Since this data is not available for this study, it was assumed a uniform distribution of passengers across the cities and the weight given to each census tract was proportional to its population. Information relative to each census tract were obtained via an online database[3], which is based on the 2010 census data gathered by the US government. This aggregation yields a single t_{to} per departing airport and a single t_{from} per arrival airport, which can then be used to obtain a single T total per flight.

From Table II, the busiest city pair is (Los Angeles, San Francisco), therefore this pair was used for the following analysis of full door-ot-door travel times. Figure 1 shows the
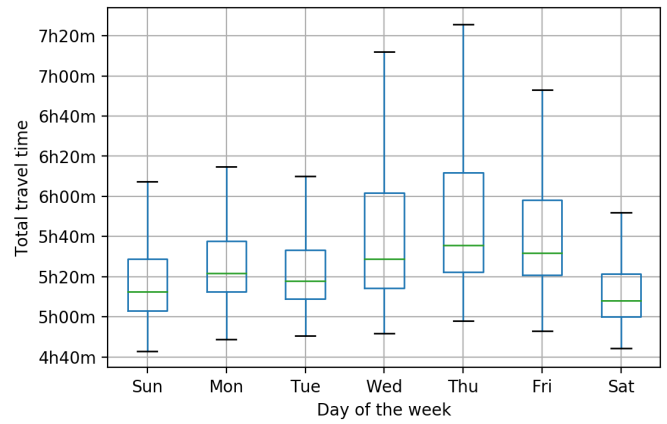
[3] www.usboundary.com

travel time distribution per day of the week under a boxplot form for trips from Los Angeles to San Francisco and Figure 2 for the return trips. The same pattern is found for these two trips: higher median traveling times and higher variability from Wednesday to Friday, and a week low on Saturday. The median traveling times are lower for the trip from Los Angeles to San Francisco than for the return trip, however there is a greater risk of longer travel times, shown by the greater size of the upper whiskers of the boxplots.

Figures 3 & 4 show the hourly boxplot distribution of the full door-to-door travel times between these two same cities. The hour used for this distribution is based on the scheduled hour of the flight, not on the hour of the departure time of the full trip. Using this representation, it is still clear that the trip from Los Angeles has a lower median travel time but a greater upper tail distribution. This representation also enables to quickly see which hours to avoid if one wants to minimize the travel time. For instance, the worst moment to fly from Los Angeles or San Francisco is at 3pm due to a combination of high median travel times and a high variability. Flying out early does not necessarily imply avoiding rush hour: Travellers
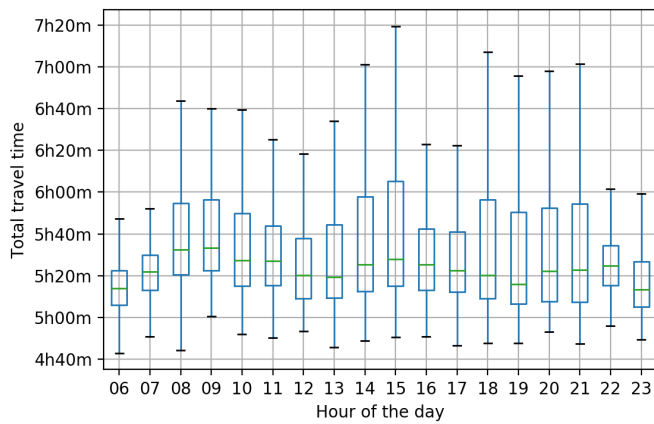
Figure 3: Hourly boxplots of door-to-door travel times from Los Angeles to San Francisco.
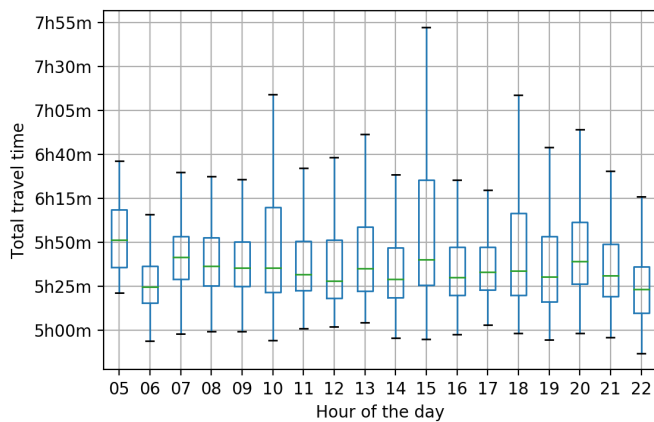

Figure 4: Hourly boxplots of door-to-door travel times from San Francisco to Los Angeles.

leaving at 5am from San Francisco have a median travelling time 30 minutes greater than those leaving at 6am.

B. Leg analysis

Another use of this model is to be able to analyze the differences between airports on every leg of the trip. Using the same aggregation as in Section III-A in order to have a single t_{to} and a single t_{from} per flight, it is possible to plot time histograms for each segment of the door-to-door trip. These histograms assume here a uniform passenger distribution within the census tracts, independent from their population density.

Regarding the city pair (Boston, Seattle), Figure 6 shows the histograms of the time spent going to and from the airports as well as the time spent in flight for both ways over the considered three month period. Figures 6a & 6c indicate that both airports are similarly well-integrated to their cities with the quasi-totality of the weighted egress or access time distributions under 30 minutes. For this city pair, the flight time is the major difference between travelling one way or another as shown in Figure 6b, assuming the processing times presented in Table I.

Focusing on a city pair within the same timezone, such as (Seattle, San Francisco), the time histograms can lead to a totally different conclusion. Figure 7 shows the histograms of the time spent going to and from the airports as well as the time spent in flight for both ways. In this case, the flight time distribution is similar for both ways of the trip as shown in Figure 7b. The access and egress time distributions show however a difference in the airport integration with their city. Figure 7a shows that most trips to SEA take less than 30 minutes while the majority of trips to SFO take more than 30 minutes. Similarly, Figure 7c shows that leaving SFO takes also more than 30 minutes while leaving SEA takes less than 30 minutes, though the 30 minute limit is less sharp than for the access time. For this city pair, the shift in the full door-to-door time distribution between the two directions shown in Figure 5 is essentially due to the better integration of SEA with Seattle and due to the slight time difference between access and egress to SFO.
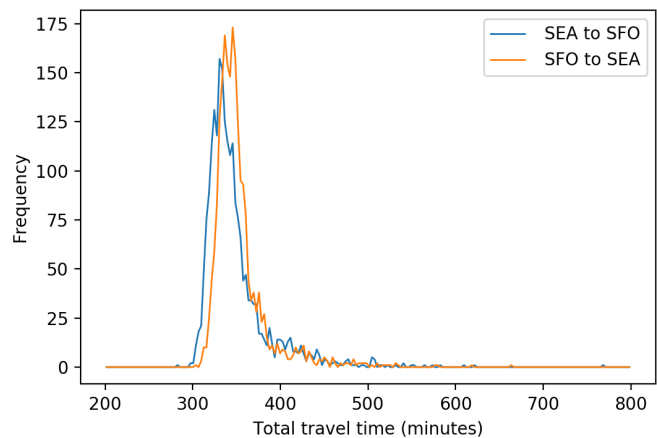

Figure 5: Histogram comparison of the full door-to-door times for trips between San Francisco and Seattle

A complementary comparison of trip time distribution can be achieved by looking at the proportion of time spent per trip leg for each trip. Figure 8 presents a bar visualization of the average percentage of time spent per phase for the twenty considered trips. In this figure the trips are sorted according to the percentage of time spent in flight. A clear takeaway of this graph is that travellers spend on average less time going from the initial census tract to the departure airport and leaving from the arrival airport to the final census tract than dwelling at the departure airport. For short-haul flights such as between San Francisco and Los Angeles or between Boston and Washington D.C., travellers spend on average more time at the departure airport than in flight. This analysis could however be refined by considering less aggregated processing times, e.g. by calculating the processing times t_{sec} for travellers per origin-destination pairs.

Focusing on the access to the airport, it is possible to visualize the integration of airports with their city and the surrounding area by plotting the time needed to reach the
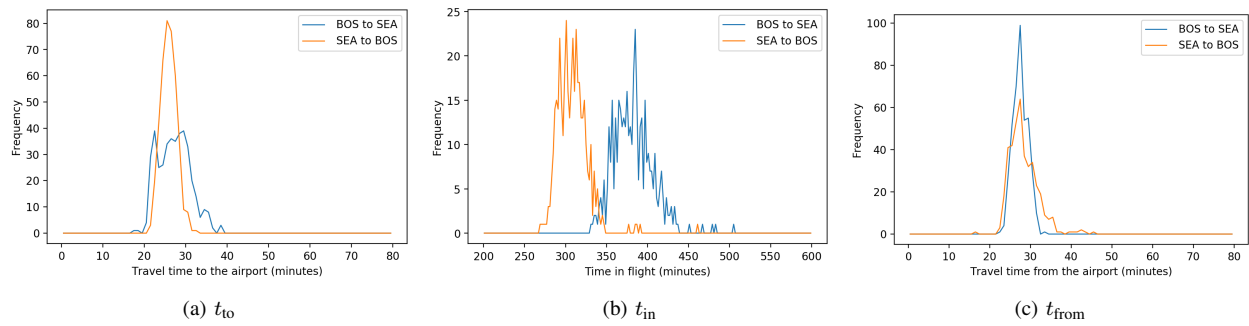
(a) t_{to} (b) t_{in} (c) t_{from}

Figure 6: Histogram comparison for different legs of the trips between Boston and Seattle

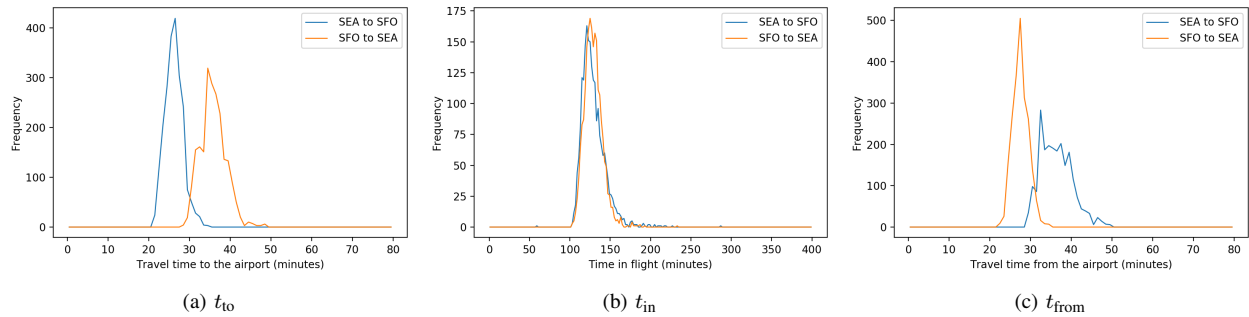(a) t_{to} (b) t_{in} (c) t_{from}

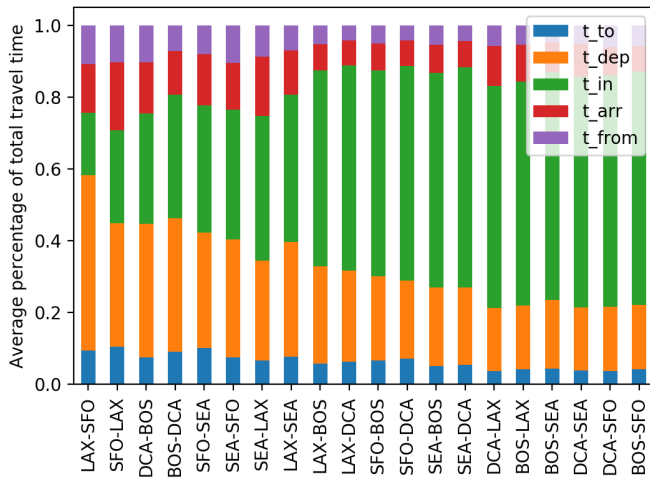Figure 7: Histogram comparison for different legs of the trips between San Francisco and Seattle

Figure 8: Comparison of the average proportion of the time spent per trip phase for the twenty considered trips.

airport versus the distance of the census tracts with the airport. This distance is calculated using the internal point coordinates associated with each census tract from the usboundary.com database. Figure 9 shows a scatter plot of the average daily travel time to the airport versus the distance to the airport for all five considered airports. A linear regression of these average times with respect to distance is also plotted for a better comparison. A first observation is the disparity in range for the different airports: SFO has sufficient Uber rides coming from census tracts more than 120 km while DCA has a range

limited to 20 km. This difference could be explained by the difference in reach of the different airports but one should also keep in mind the possible limitation due to Uber data availability. When comparing the three airports with similar reach (namely BOS, LAX and SEA), the slopes of the linear regressions indicate that Seattle has the airport best integrated within its metropolitan area, followed by Boston and then Los Angeles.
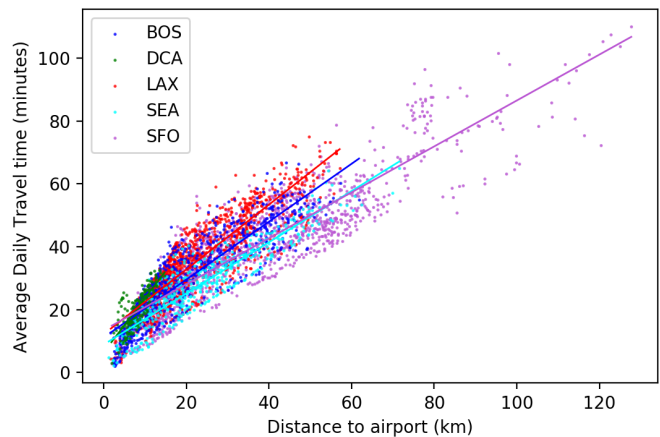

Figure 9: Scatter plot of the average daily travel time to the airport versus the distance to the airport. Straight lines indicate the linear regression fitted curves for each city.

C. Reach analysis

This model can also be used to visualize the accessible range of a city starting from a specific census tract. Once a starting point is chosen, it is possible to create a color map of the door-to-door times indicating the reach given by the air link. This map can be useful in better understanding the effects of urbanisation or severe weather.
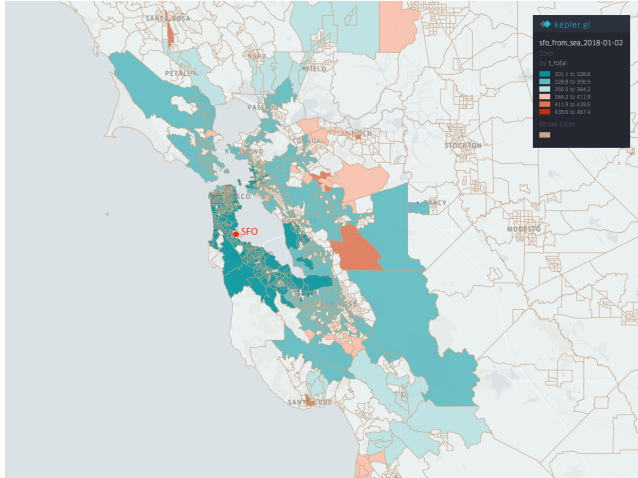
(a) From Seattle to San Francisco
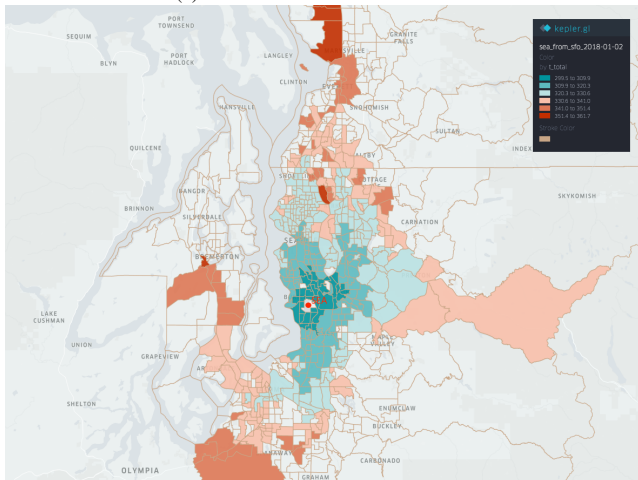
(b) From San Francisco to Seattle

Figure 10: Door-to-door travel times between the city pair (Seattle, San Francisco) starting from their city centers

In order to better understand the differences in egress times between the city pair (Seattle, San Francisco), the door-to-door map are plotted in Figure 10. For these maps, the starting census tract was the one containing the city hall, associated with the city center of each city. The full door-to-door time was then averaged per census over one day, January 2nd 2018. The color scaling is different in the maps and that realization is already an indication of the differences in the car reach of each airport. In Figure 10b, the full color scale (from 299 to 362 minutes) is almost totally encompassed within the first two color levels of Figure 10a (from 300 to 356 minutes). Both maps have a time expansion with a quasi

linear form, due to the main roads servicing these cities. For Seattle (Figure 10b) there is a clear difference between the South-North propagation and the East-West propagation due to the highway I-5 following a North-South direction. For San Francisco (Figure 10a), the reach of SFO is almost rectangular and follows the orientation of the Bay and the many highways (e.g. I-280 and I-880) on both side of the Bay. The zones in orange and red surrounded by green zones correspond to areas close to parks and hence with less housing and fewer roads.

This visualisation method can also be used to study the evolution of the reach of a city over time, which can be useful for analyzing the effect and reach of natural disasters. The same severe weather perturbation as in [13] was investigated here. On January 4th 2018, a winter storm nicknamed "Bomb Cyclone" hit the East Coast of the United States leading to the closure of some of the main Northeastern airports and the cancellation of a majority of flights servicing the area. Figure 11 shows the average door-to-door times to reach different census tracts in the Boston area starting from the city center in Washington D.C. both before and after the Bomb Cyclone.

Note that the color scale is different in the two maps, indicating that on the day after the Bomb Cyclone, the minimum time to reach Boston increased of twenty minutes compared to before the Bomb Cyclone and the maximum time increased of ten minutes. Not only does the map shift towards the red, indicating longer travel times, but some zones did not have enough Uber data, which means that they were not well or not at all serviced, especially in the South and the North West. Setting the Bomb Cyclone analysis aside, one can notice that Boston has a road structure different than Seattle and San Francisco since the reach of its airport is star-shaped and not linear.

IV. CONCLUSION

This paper leveraged Uber's recently released data and integrated it with three other data sources in order to create a model of the full door-to-door travel time for flights between five different cities of the United States. It could however be implemented for any world city pairs with available ride-sharing or taxi data as well as sufficient flight time information. By aggregating the full door-to-door travel times at a city level, it enables both the pairwise comparison of the different travel times per trip segment between two cities as well as an analysis over time of the time necessary to join two specific cities. This model can also be used to evaluate on a national level some passenger-centric objectives within NextGen in the US and ACARE in Europe regarding the good integration of airports within their cities.

Further studies should consider using alternative modes to reach the airport such as the subway as well as different modes linking cities (e.g. rail) when such modes exist. Additionally, knowing the actual daily proportion of travellers using the different approaches (road or rail) would enable a better daily evaluation of the full door-to-door travel time. A possible
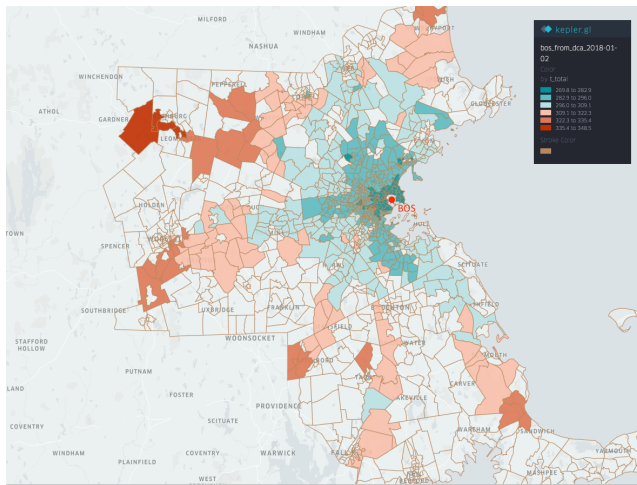
(a) On January 2nd, 2018
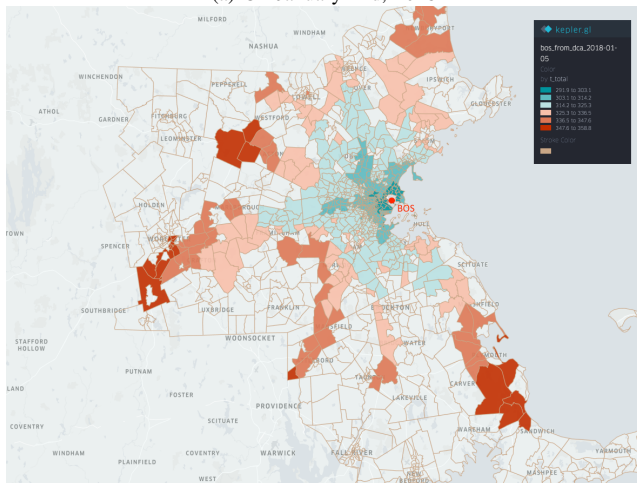
(b) On January 5th, 2018

Figure 11: Door-to-door travel times from the city center in Washington D.C. to Boston before and after the Bomb Cyclone of January 2018

method to determine this proportion would be by using aggregated information from GPS or mobile phone sources.

ACKNOWLEDGMENT

REFERENCES

[1] *White Paper: Roadmap to a Single European Transport Area - Towards a Competitive and Resource Efficient Transport System.* Brussels: European Commission, Mar. 2011.

[2] *NextGen Priorities - Joint Implementation Plan Update Including the Northeast Corridor.* Federal Aviation Administration, Oct. 2017.

[3] J. Meserole and J. Moore, "What is System Wide Information Management (SWIM)?" in *2006 IEEE/AIAA 25TH Digital Avionics Systems Conference.* Portland, OR, USA: IEEE, Oct. 2006, pp. 1–8.

[4] NextGen Integration and Implementation Office, "NextGen Implementation Plan," in *Federal Aviation Administration*, 2009.

[5] E. Commission, Ed., *Flightpath 2050: Europe's Vision for Aviation ; Maintaining Global Leadership and Serving Society's Needs ; Report of the High-Level Group on Aviation Research*, ser. Policy / European Commission. Luxembourg: Publ. Off. of the Europ. Union, 2011, oCLC: 930887434.

[6] Y. O. Gawdiak and T. Diana, "NextGen Metrics for the Joint Planning and Development Office," p. 10, 2011.

[7] A. Cook, G. Tanner, S. Cristóbal, and M. Zanin, "Passenger-Oriented Enhanced Metrics," p. 8, 2012.

[8] S. Peer, "Door-to-door travel times in RP departure time choice models: An approximation method using GPS data," p. 17, 2013.

[9] M. Salonen and T. Toivonen, "Modelling travel time in urban networks: Comparable measures for private car and public transport," *Journal of Transport Geography*, vol. 31, pp. 143–153, Jul. 2013.

[10] E. Durán-Hormazábal and A. Tirachini, "Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in Santiago, Chile," *Research in Transportation Economics*, vol. 59, pp. 26–39, Nov. 2016.

[11] E. Pels, P. Nijkamp, and P. Rietveld, "Access to and competition between airports: A case study for the San Francisco Bay area," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 1, pp. 71–83, Jan. 2003.

[12] A. Marzuoli, E. Boidot, E. Feron, and A. Srivastava, "Implementing and validating air passenger-centric metrics using mobile phone data," *Journal of Aerospace Information Systems*, 2018.

[13] A. Marzuoli, P. Monmousseau, and E. Feron, "Passenger-centric metrics for Air Transportation leveraging mobile phone and Twitter data," in *Data-Driven Intelligent Transportation Workshop - IEEE International Conference on Data Mining 2018*, Singapore, Nov. 2018.

[14] P. García-Albertos, O. G. C. Ros, and C. Ciruelos, "Understanding Door-to-Door Travel Times from Opportunistically Collected Mobile Phone Records," p. 8, 2017.

[15] X. Sun, S. Wandelt, and E. Stumpf, "Competitiveness of on-demand air taxis regarding door-to-door travel time: A race through Europe," *Transportation Research Part E: Logistics and Transportation Review*, vol. 119, pp. 1–18, Nov. 2018.

[16] P. Monmousseau, D. Delahaye, A. Marzuoli, and E. Feron, "Door-to-door travel time analysis from Paris to London and Amsterdam using Uber data," in *Ninth SESAR Innovation Days*, Athens, Greece, 2019.

[17] Z. Li, Y. Hong, and Z. Zhang, "Do On-demand Ride-sharing Services Affect Traffic Congestion? Evidence from Uber Entry," p. 35, 2016.

[18] J. D. Hall, C. Palsson, and J. Price, "Is Uber a substitute or complement for public transit?" *Journal of Urban Economics*, vol. 108, pp. 36–50, Nov. 2018.

[19] M. Wang and L. Mu, "Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA," *Computers, Environment and Urban Systems*, vol. 67, pp. 169–175, Jan. 2018.

[20] Bureau of Transportation Statistics, "Bureau of Transportation Statistics, About BTS." [Online]. Available: http://www.rita.dot.gov/bts/about