Trip length distribution of TNC trips: based on empirical data in Chicago

Irene Martínez* and Wen-Long Jin*

January 5, 2020

Keywords: *bathtub model, trip length distribution, TNC empirical data*

1 Introduction

Understanding and preventing traffic congestion in urban areas is a problem of growing importance. Recently, the potential time savings of modeling the traffic dynamics based on the relation between average density and average flow at the network level has been established. This relation is referred as Network (or Macroscopic) Fundamental Diagram (NFD), and empirical findings support the existence of this flow-density relation (Godfrey, 1969; Geroliminis and Daganzo, 2008; Ambühl and Menendez, 2016). Based on it, the network dynamics can be modeled with the so-called "bathtub model", that captures the inflow, outflow and current accumulation of vehicles in the network. This term was proposed by Vickrey (2019) and has been used in the literature (Small and Chu, 2003; Fosgerau, 2015; Arnott, 2013). The model describes an aggregated relation between the finishing trips in the city and the total accumulation of cars. It is the simplest way to describe the traffic network dynamics, by tracking the number of vehicles, $n(t)$, as the difference between the cumulative arrival to the network, $I(t)$, and the cumulative departure from the network, $O(t)$. It explicitly tracks the interaction between the supply and demand, since $I(t) = O(t + T(t))$, where $T(t)$ denotes the time duration of a trip starting at time t. The model was claimed to be mathematically intractable (Arnott, 2013), because the delay is endogenous. However, assuming that the trip distance of the users in the network follows a negative exponential distribution (NED), Vickrey (1991) derived an outflow expression that does not have a endogenous delay

$$O(t) = n(t)V(n(t))\frac{1}{B\bar{L}}, \tag{1}$$

where $V(n)$ is the speed-accumulation relation determined by the NFD, B represents the total network length and $\bar{L}$ is the average trip length. We refer to this version as "Vickrey's bathtub

*Department of Civil and Environmental Engineering, Institute of Transportation Studies, 4000 Anteater Instruction and Research Bldg, University of California, Irvine, CA 92697-3600.

model". More recently, the same outflow function was independently derived by Small and Chu (2003); Daganzo (2007) and is sometimes referred as 'accumulation-based' model. If the trip distance distribution does not follow a NED, the linear relation between the finishing trip rates and the circulating flow (Eq. 1) only holds for steady state traffic (Gonzales and Daganzo, 2012; Lamotte et al., 2018). In the past, to solve the generic bathtub model (i.e. without assumption of NED), researchers have employed numerical methods e.g. Daganzo and Lehe (2015); Lamotte et al. (2018); Mariotte et al. (2017), and have referred to this approach as 'trip-based' model. This model is able to account for the trip length distribution (TLD) among users, and track the trip completion of individual trips. Recently, Jin (2019) proposed an approach to solve the bathtub model for any TLD by tracking the changes in the distribution of remaining trip distance.

In summary, "Vickrey's bathtub model" is based on the assumption that the TLD follows a negative exponential, but this has never been validated with real data. Moreover, trip lengths have been shown to play an important role on the regional traffic dynamics (Mariotte et al., 2017; Batista et al., 2019). Furthermore, the trip length of the users is a fundamental variable that is needed for studying other aspects, e.g. it is an important input for the 'trip distribution' of the four-step model. It is also important for mileage tax and traffic management and control (e.g. transit scheduling). During many years data about the trip length has been sparse, mainly because it has been collected through surveys. Other ways to collect the trip distance information is trough mobile phone data (Colak et al., 2016) or even GPS traces (Cich et al., 2015). In light of the importance of the TLD in a city, we present a systematic analysis of the trip distances of Transportation Network Companies (TNC) trip data based on empirical measurements in the city of Chicago to establish a TLD.

2 Trip length distributions

2.1 Models

Based on travel surveys from 1995 and 2006, Thomas and Tutert (2013) studied the transferability of the trip distribution models in time and space. To do so, the authors estimate the trip distribution functions for different education and urbanization levels and establish that the trip distribution can be explained by a "negative exponential-to-the-power law" distribution, i.e. $f(l) = e^{a+bl^{0.4}}$. Its slope was reported to vary for different education levels (Thomas and Tutert, 2013). Therefore, they argue that a single stable distribution function is not enough to describe the aggregated travel behavior. More recently, Colak et al. (2016) estimated the distribution of the straight-line distance for 5 different cities and concluded that the trip distribution follows a log-normal distribution, i.e.,

$$f(l;\mu,\sigma) = \frac{1}{l\sigma\sqrt{2\pi}} e^{-\frac{(ln(l)-\mu)^2}{2\sigma^2}}. \tag{2}$$

On the other hand, a trip NED function could be assumed (Vickrey, 1991) with a generic expected trip length that is not necessarily time independent, $L(x,t)$, the probability density function (pdf) would follow

$$f(l;L(x,t)) = \frac{1}{L(x,t)} e^{-\frac{l}{L(x,t)}}. \tag{3}$$

2.2 Methodology

The interest of this paper is to determine the TLD based on empirical data and check whether the necessary hypothesis (TLD is a time-independent NED) to formulate the Vickrey's bathtub model holds. This assumption has never been validated before and exiting studies suggest otherwise (Thomas and Tutert, 2013; Colak et al., 2016).

To test whether the probability distribution of a sample follows a reference probability distribution (RD), the Kolmogorov–Smirnov (KS) test is used. This statistic quantifies the distance between the cumulative distribution function of the RD and the empirical one. First, the maximum likelihood method is used to estimate the parameters of a given distribution, e.g. L for the NED (Eq. 3), and μ and σ for the log-normal distribution (Eq. 2), from a sample of trips. Then, the null hypothesis, H_0, is defined as "the calibrated RD is the under-laying probability distribution of a different sample of trips", and the KS statistic is used to determine whether H_0 can be rejected or not. Both samples need to have the same characteristics, e.g. correspond to the same zone, time of day, day of week, etc. The paper calibrates and validates for: (i) The weekday trips (Mon-Fri) during the whole day in all zones, (ii) the Tuesday trips from 7-9am (morning peak) and from 1-3pm (off peak) in all zones; and, (iii) all weekday trips (Mon-Fri) during the whole period for three different community areas (with highest number of internal trips). These analyses will help to determine whether the TLD depends on the day-of-the-week and whether it is time- and/or location-dependent.

3 Empirical (preliminary) results

The data used[1] corresponds to the information that TNC in the city of Chicago have collected over more than one year, starting on November 2008. The trips are reported in 15 minute intervals, by rounding the starting time to the nearest interval. More than 73M trips are recorded with the length (in miles) and duration (in seconds). Moreover, the trips origins and destinations are zone-based, corresponding to the 77 community areas in Chicago. The average trip length in each zone is not constant (**Figure** 1). Therefore, the average trip distance should be location-dependent.

The calibration of different TLD and validation with KS test will be presented in the full paper. Here, an analysis of the average trip length is presented. **Figure** 2a shows the frequency of trip lengths aggregated in an hourly way. The data corresponds to all trips recorded in Chicago during the year that took place on a Tuesday. The TLD changes over time. Thus, the time-independent NED cannot hold for TNC trips. This is also confirmed by checking the average trip length over

[1] Publicly available at https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p.

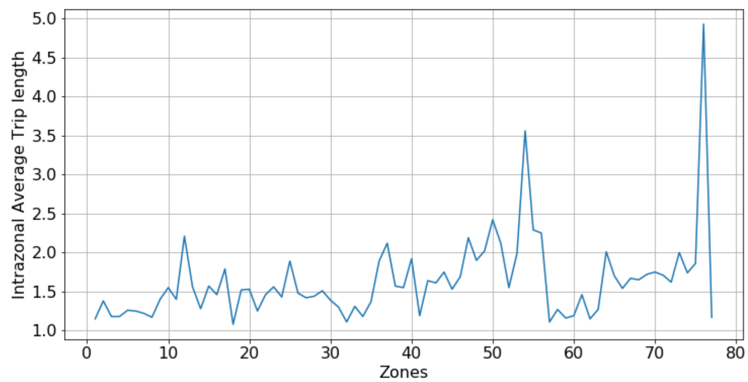
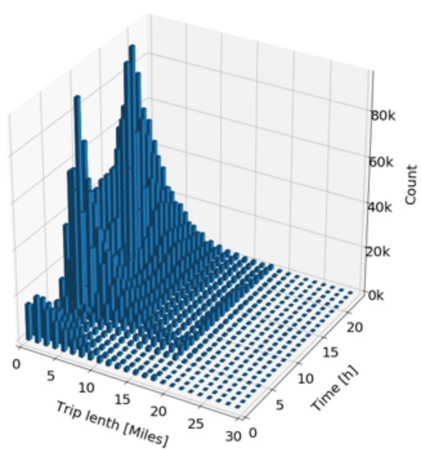
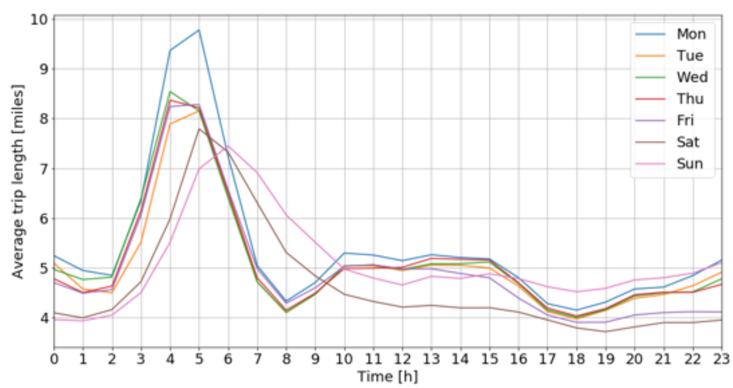


Figure 1: Average trip length over the whole year for intra-zonal trips. Note that the average trip length for the whole data set is 3.7 miles.



(a)



(b)

Figure 2: (a) Trip length frequency distribution over time on Tuesdays. (b) Average trip length for each 1 hour interval for different days of the week.

time for different days of the week (**Figure** 2b). Whether the TLD follows a NED with time and location dependent average trip distance (Eq. 3) or rather another time-dependent pdf will be established in the full paper.

4 Discussion

This work points out the importance of the trip length distribution, both from the practical and modeling perspective. The TLD, jointly with the trip rate, defines the travel demand and dictates the congestion dynamics in a road network. Moreover, the assumption of time-independent NED for trip lengths does not hold for the TNC trips, since the average trip length is time-dependent. Thereafter, the "Vickrey's bathtub model" cannot be used as a simplified analysis of the network traffic dynamics. Rather, other more complete models that account for the trip progression and trip completion rate with a given TLD need to be used.

References

Ambühl, L. and Menendez, M. (2016). Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C: Emerging Technologies*, 71:184–197.

Arnott, R. (2013). A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, 76:110 – 121.

Batista, S., Leclercq, L., and Geroliminis, N. (2019). Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. *Transportation Research Part B: Methodological*, 122:192 – 217.

Cich, G., Knapen, L., Bellemans, T., Janssens, D., and Wets, G. (2015). Trip/stop detection in gps traces to feed prompted recall survey. *Procedia Computer Science*, 52:262 – 269. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).

Colak, S., Lima, A., and Gonzalez, M. C. (2016). Understanding congested travel in urban areas. *Nature communications*, 7:10793.

Daganzo, C. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.

Daganzo, C. F. and Lehe, L. J. (2015). Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological*, 75:89 – 99.

Fosgerau, M. (2015). Congestion in the bathtub. *Economics of Transportation*, 4(4):241–255.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759 – 770.

Godfrey, J. (1969). The mechanism of a road network. *Traffic Eng. Control*, 11:323–327.

Gonzales, E. J. and Daganzo, C. F. (2012). Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, 46(10):1519 – 1534.

Jin, W.-L. (2019). Generalized bathtub model of network trip flows.

Lamotte, R., Murashkin, M., Kouvelas, A., and Geroliminis, N. (2018). Dynamic modeling of trip completion rate in urban areas with mfd representations. Washington, D. C. Presented in 97th Annual Meeting of the Transportation Research Board.

Mariotte, G., Leclercq, L., and Laval, J. A. (2017). Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological*, 101:245 – 267.

Small, K. A. and Chu, X. (2003). Hypercongestion. *Journal of Transport Economics and Policy*, 47:19–52.

Thomas, T. and Tutert, S. (2013). An empirical model for trip distribution of commuters in the netherlands: transferability in time and space reconsidered. *Journal of Transport Geography*, 26:158 – 165.

Vickrey, W. (1991). Congestion in midtown manhattan in relation to marginal cost pricing. *Unpublished notes, Columbia University*, 13.

Vickrey, W. (2019). Types of congestion pricing models. *Economics of Transportation*, 20:100140.