On Time-dependent Trip Distance Distribution with For-hire Vehicle Trips in Chicago

Irene Martínez
Institute of Transportation Studies - UC Irvine
4000 Anteater Instruction and Research Bldg (AIRB)
Irvine, CA 92697-3600
Tel: 949 351 7098; Email: irenem3@uci.com
Orcid number: 0000-0002-3052-4791

Wen-Long Jin
Institute of Transportation Studies - UC Irvine
4000 Anteater Instruction and Research Bldg (AIRB)
Irvine, CA 92697-3600
Tel: 949 824 1672; Email: wjin@uci.edu
Orcid number: 0000-0002-5413-8377

ABSTRACT

For transportation system analysis in a new space dimension with respect to individual trips' remaining distances, vehicle-trips demand has two main components: the departure time and the trip distance. In particular, the trip distance distribution (TDD) is a direct input to the bathtub model in the new space dimension, and a very important variable to consider in many applications, such as the development of distance-based congestion pricing strategies or mileage tax.

To have a good understanding of the demand pattern, both the distribution of trip initiation and trip distance should be calibrated from real data. In this paper, we assume that the demand pattern can be described by the joint distribution of trip distance and departure time. In other words, the TDD is assumed to be time-dependent, and a calibration and validation methodology of the joint probability is proposed, based on log-likelihood maximization and the Kolmogorov-Smirnov test. The calibration method is applied to empirical for-hire vehicle trips in Chicago, and we conclude that TDD varies more within a day than across weekdays. We also reject the hypothesis that the TDD follows a negative exponential, a log-normal, or Gamma distribution. However, the best fit is systematically observed for the time-dependent log-normal probability density function. In the future, other trip distributions should be considered and also non-parametric probability density estimation should be explored for a better understanding of the demand pattern.

Abstract Word count: 249

Keywords: trip distance, trip distribution, empirical data, for-hire vehicle trips

INTRODUCTION

In order to improve mobility, researchers study ways of “shaping travel demand". Many different strategies can be considered, and most involve the reduction of motorized vehicles traveled distances. For example, reducing the number of motorized trips, or increasing vehicle occupancy level for the motorized trips (*1*). Transportation demand in a road network is traditionally stored in origin-destination (OD) matrices, where each cell represents the number of vehicle trips between each OD pair[1]. However, the estimation of OD matrices is generally not very accurate due to the limited data available. In fact, the problem is known to be under-determined when one tries to estimate OD matrices from link flows. Further, the estimation of dynamic OD matrices is very computationally expensive. However, the demand calibration or modeling is a key component in any traffic flow model in order to accurately model traffic congestion and propose adequate operational strategies to alleviate it.

Recently, a new paradigm for transportation system analysis has been introduced where the spatial dimension is relative. The idea is to disregard the network topology and use space as a relative distance to the destination (*2, 3*). In this new paradigm, the travel demand is described by the number of trips initiated at any given time and the trip distance distribution (TDD) of these trips (*4*). This demand definition is in fact a direct input (or assumption) of the so-called bathtub models (*2–5*), which have been gaining interest in recent years among the research community. While the supply side of the bathtub models has had a lot of attention in the literature, the demand calibration has been overlooked. Several studies have highlighted the important role of trip distance (in this case at a regional level) on the accurate prediction of traffic dynamics (*6–8*). More recently, it has been observed that the mean trip distance changes over the day based on mobile phone data (*9*), which contradicts the common assumption of time-independent TDD for the bathtub model (*2, 3, 10, 11*).

In transportation science, the trip distance of the users is a fundamental variable that is needed for studying different aspects. For example, it is an important input in the ‘trip distribution’ step of the four-step model, since it represents a measure of travel impedance. In the past, the data available to estimate the trip distance has been sparse, because it was collected through surveys. However, in recent years, other ways to collect the trip distance information have become available thanks to the mobile phone data (*12–14*), and GPS’ traces (*15, 16*). Several empirical studies have reported different TDD, such as log-normal distribution (*14, 17, 18*). These new technologies for data collection will allow to obtain larger data sets and a more accurate estimation of demand. However, the collection of such detailed information might lead to privacy concerns of users. When considering the new relative space paradigm for transportation, the data requirements are lower and privacy might be easier to be guaranteed, because only two variables should be collected. Thus, GPS’ traces and mobile phone data present a potential data source to estimate the demand for the new transportation paradigm. However, there is no systematic calibration procedure in the literature for this demand definition.

In summary, the study of the TDD is gaining interest in the research community, both from the practical and modeling perspective. The TDD jointly with the trip initiation rate defines the travel demand and dictates the congestion dynamics in a road network. It is natural to think of the demand as a joint distribution of trip distances and departure time, although this concept is

[1]Depending on the duration of the time interval, the OD matrix is said to be static (long period of time) or dynamic (short period).

rather novel (*4*). In this paper, we propose a trip demand estimation methodology for such joint distribution. Then, the method is used to calibrate the TDD of for-hire vehicle trips reported by the Transportation Network Providers in Chicago, for two common TDD assumed in the literature and a generalized function. Through statistical testing, we reject the hypothesis that the TDD is time-independent and also reject the hypothesis that it follows any of the distributions considered in this paper, which reveals the necessity of further studying the joint distribution of trip distances and departure time since the common assumptions in the literature are not valid , at least for the data set considered in this paper.

The rest of the paper is organized as follows. Section 2 presents a literature review of TDD assumptions and models. Then, Section 3 reviews the definition of the joint trip distance and departure time distribution and proposes the methodology for its calibration. In Section 4 an overview of the data used in this paper is presented and we establish the time dependency of the TDD through hypothesis testing. Then, both calibration and validation of three different time-dependent probability density functions are performed, and the results are discussed in Section 5. Section 6 discusses the contributions, limitations, and practical implications of this study and we conclude the paper with a short summary in Section 7.

LITERATURE REVIEW

In this section we will first summarize the assumptions on TDD for bathtub models. Then, we review some existing models and empirical calibrations of TDD.

Trip distance assumptions in bathtub models

The "bathtub model" was a term proposed by Vickrey (*2, 3*), which described an aggregated model capturing the completion of trips in the city as a function of the number of vehicles in it. The dynamics are described by a conservation equation of the active vehicles in the system. The supply side of the bathtub models has been studied extensively in the literature, e.g. through the network fundamental diagram calibration, a.k.a. macroscopic fundamental diagram. The demand is another important input of the bathtub model. However, the estimation of the TDD has been largely overlooked in the bathtub model literature. This paper aims to fill this gap by defining and estimating the time-dependent TDD. This will also help to have a better understanding of the demand.

There is a special case where the traffic dynamics can be modeled with a simple ordinary differential equation. This was first derived assuming that the trip distance of the users in the network follows a (time-independent) negative exponential (NE) distribution (*2, 3*), such as

$$\varphi(x)_{NE} = \frac{1}{B}\exp\left(\frac{-x}{B}\right), \tag{1}$$

where B represents the average trip distance and $\varphi(x)$ is the probability density function of trip distances x. We refer to this bathtub model as *Vickrey's bathtub model*. Later, the same dynamics were independently derived by other authors (*10, 11*). However, the assumption of NE distribution was not explicitly stated in these later works. Empirical findings in Yokohama (*19*) suggest that the average distance of trips, B, does not change much over time, and is a homogenous parameter for the whole city, which relates the flow in the network with the completion rate of trips. This would be consistent with the NE assumption (*20*). However, more recent studies contradict these observations and have reported that the average trip length in an area does change over time (*9, 21*).

Other TDD have been assumed in the literature for bathtub models. For example, the same trip distance for all travelers (*5*). This model with homogeneous users is often referred as *the basic bathtub model*. Recently, the so-called *generalized bathtub model* was derived for any given TDD (*4*). The completion rate of trips for the generalized bathtub model depends on the distribution of remaining trip distances (*4*).

The so-called "trip-based model" (*5, 22–24*), is developed as a reformulation of the bathtub model, where the objective is to track individual users' trip progression. Most of these studies consider constant trip distances and thus are equivalent to the basic bathtub model. A framework to determine explicit distributions of travel distances has been proposed based on information of vehicle-trips in the city network through sampling of a set of (virtual) trips (*7*). These distributions were assumed to be time-independent but the probability density function was not calibrated. Later, the framework was extended to assume that a single trip can change its route depending on the traffic conditions. However, in this paper, we will assume that the trip distances are fixed for individual vehicle-trips.

In summary, there are several assumptions in the literature regarding the TDD for bathtub models. Many researchers have assumed constant TDD (*5, 24, 25*), and others have assumed NE distribution either explicitly (*2, 3*) or implicitly (*10, 11*). In this paper, we will test whether the trip distance is time-independent and whether the trip distance follows a negative exponential distribution.

Existing trip distance distribution models

In general, trip distance (or trip length) is studied through regression models based on population density and other demographic characteristics (*13*). The distribution step of the four-step model has been traditionally done based on Gravity models, using a functional distribution of the travel impedance, which parameters are calibrated subsequently. The travel impedance is in most of the cases defined by the trip distance. Thus, different functional forms of TDD have been considered in the literature. The most popular functions are negative exponential, power-law or a combination as $\varphi(x) = x^a \exp(-bx)$, (*26*). Other TDD have been used too, e.g. log-normal, or logistic distribution (*27*). For example, (*17*) established that the trip distribution can be well explained by a "negative exponential-to-the-power law" distribution. The slopes of this function were reported to vary for different education levels and the average trip distance was shown to increase over time (*17*). However, considering an "negative-exponential-to-the-power-law" expression $\varphi(x) = e^{a+bx^{0.4}}$, does not necessarily satisfy $\int_0^\infty \varphi(x)dx = 1$ for all values of a and b. This property is one of the most important characteristics of a probability density function, since the area under the curve should be 1. Particularly, one can derive the required parameter $a = -0.187$, given the estimated value of $b = 1.5$ by (*17*), which is not consistent with the empirically estimated value $a = 6$. A more generic relation can be established between a and b, solving $\int_0^\infty e^a e^{-bx^{0.4}} dx = 1$, i.e. $a = \ln\left(\frac{8b^{\frac{5}{2}}}{15\sqrt{\pi}}\right)$.

The calibration in TDD for Gravity models is traditionally done with limited survey data, thus the results were not very accurate. More recent findings based on mobile phone data (*14*) suggest very interesting similarities between TDDs across the five cities analyzed. In particular, they concluded that the straight line distance between origins and destinations for commuting trips follows a log-normal (LN) distribution, i.e.

$$\varphi_{LN}(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right), \tag{2}$$

where μ is the mean of the variable's natural logarithm and the σ is the standard deviation of the variable's natural logarithm. The values obtained from the fitted log-normal distribution ranged from $\mu \in [1.6; 2.1]$ and $\sigma \in [0.7; 1.2]$, (*14*). Later, another study based on 3 months of empirical data of Didi trips in 10 different cities in China also considered log-normal distribution of trip distance for the parameter estimation (*28*). The log-normal TDD has been assumed for other modeling purposes (*29*) as well.

All these empirical studies aggregated the trip distances across hours (or days) and calibrated a single distribution. This means that their studies (indirectly) assumed that the TDD is time-independent. However, it is important to consider trip distance variation for micro-simulations (*6*). For this reason, the present paper defines the joint distribution of departure time and trip distance (*4*) as the demand and proposes a method to study and calibrate the time-dependent TDD. Very recently the mean trip distance (MTD) variations across on- and off-peak hours have been studied (*9*). They observe with empirical data that the MTD changes over time. However, the trip distance analysis was based on regional paths, i.e. single trips were "cut" into regions and for each region the TDD was obtained. Therefore the analyzed TDD were dictated by the topological features of the city network and the network partitioning. There are three main differences between this paper and (*9*): First, in this study we consider the trip distance analysis of the whole trips, which allows to study the demand, instead of studying the distribution of partial trip distances in regions that have been defined for the purpose of modeling. Thus, our approach is more generic and the study of the TDD can serve other purposes than being the input of a bathtub model. Second, in this study we assume a continuous TDD that is time-dependent and we calibrate the parameters of three assumed distributions, based on data. On the other hand, (*9*) do not present any calibration of TDD, and the time variation is only studied and discussed for the MTD. Finally, the present paper studies the TDD through standard statistical hypothesis testing, which allows to draw conclusions by rejection of certain hypothesis, rather than only by looking at graphical representation of the data.

METHODOLOGY

We assume that the TDD is time-dependent, as discussed in (*17*). However, we argue that this time-dependency might be at a shorter time-scale, i.e. within day or across days, instead of considering a change over years. In the following, we define the time-dependent TDD as a mixed continuous-discrete joint distribution.

Definition of joint probability function

The concept of joint probability density function for trip lengths and time is a very natural concept, but very novel too (*4*). A joint probability function defines the likelihood of two events occurring together at the same instant. This joint distribution can be discrete, continuous or a mix. If we define the joint distribution of trip distance and departure time, $\varphi_{T,X}(t,x)$, as a continuous function the probability that between t_0 and $t_0 + \Delta t$ a trip of distance $x \in [x_0, x_0 + \Delta x]$ is initiated is $\int_{t_0}^{t_0+\Delta t} \int_{x_0}^{x_0+\Delta x} \varphi_{T,X}(t,x)dxdt$.

There is no empirical study that has tried to calibrate this joint distribution $\varphi_{T,X}(t,x)$. This paper proposes a systematic approach to fill this gap. Notice that this joint distribution can be continuous or discrete, both in time and space. In this paper, we assume that $\varphi_{T,X}(t,x)$ is discrete in time and continuous in space. The reasons to do so are twofold: First, this allows us to assume that at a given time interval, the trip distances follow a well established (continuous) probability density function, as the ones suggested in the literature, e.g. NE distribution or LN distribution. Second, the data that we will use for the estimation is not continuous in time, but trip's start time is rounded to the nearest 15 minutes. This means that we cannot differentiate between trips that started at 7:55 AM and trips that started at 8:05 AM, they are all reported to have started at 8:00 AM.

From the axiom of probability the joint probability density function can be defined mathematically as

$$\varphi_{T,X}(t,x) = \varphi_T(t) \cdot \varphi_{X|T}(x|t), \tag{3}$$

where $\varphi_T(t)$ is the marginal probability function and $\varphi_{X|T}(x|t)$ is the conditional probability of trip distance given a time t. This conditional distribution is in fact one input of the generalized bathtub model (*4*). Then, we can estimate $\varphi_{T,X}(t,x)$ by estimating both the trip generation rate mass function, and the conditional probability of trip distance for any given time. In the following, the subscripts are omitted from the probability functions to simplify the notation.

Being the time a discrete variable, the marginal probability of trip generation is defined as

$$\varphi(t) = \frac{e(t)}{E(T)}, \tag{4}$$

where $E(t) = \sum_{i=0}^{t} e(i)$ is the cumulative initiation of trips , and T indicates the end of the day. The number of trips starting between time t and $t+1$ is $e(t)$, where t is the interval. Thus, $\varphi(t)$ is the marginal probability of a trip occurring during time interval t and $t+1$ and (**Equation 4**) is discrete. We define the conditional probability of trip distance for a given time t as $\varphi(x|t) = \varphi(x;\bar{z}(t))$, where $\bar{z}(t)$ are the parameters that define the trip distance probability density function and are time-dependent. Thus, we have

$$\varphi(t,x) = \frac{e(t)}{E(T)} \varphi(x;\bar{z}(t)). \tag{5}$$

This mixed joint distribution based on discrete time intervals and continuous trip distance x is a well-defined joint probability function since it integrates to 1. The volume below the curve is

$$\sum_{t=0}^{T} \left[(t+1-t) \cdot \int_0^\infty \frac{e(t)}{E(T)} \varphi(x;\bar{z}(t)) dx \right] = \sum_{t=0}^{T} \left[\frac{e(t)}{E(T)} \int_0^\infty \varphi(x;\bar{z}(t)) dx \right].$$

Clearly, the volume is 1 since $\int_0^\infty \varphi(x;\bar{z}(t))dx = 1$ by definition of conditional distribution , and $\sum_{t=0}^{T} \frac{e(t)}{E(T)} = 1$, by definition of marginal mass function.

3.2. Hypothesis considered

In this paper, we will study the time-dependency through **Hypothesis 1**.

Hypothesis 1: Null hypothesis: The TDD is time-independent.

Then, we will study the joint probability distribution by considering three different possible functions. We consider the NE distribution **Equation 1** and the log-normal distribution **Equation 2**, since other empirical data was calibrated under that assumption (*14, 28*). Moreover, we will also consider a Gamma distribution,

$$\varphi_{Ga}(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}, \tag{6}$$

where $\Gamma(\cdot)$ is the Gamma function. This distribution has not been considered in the literature as a possible underlying function of the TDD. However, it will be considered in this paper, since $\varphi_{Ga}(x;\alpha,\beta)$ is a generalization of the NE distribution and for some values of α it has a similar " form" than the log-normal distribution. Furthermore, the Gamma distribution can be considered a particular case of a combination between exponential and power-law function, which as also been assumed to describe TDD (*26*).

If Hypothesis 1 is rejected, the joint probability functions defined above will be considered to be time-dependent. Then, we will test **Hypothesis 2a, 2b** and **2c**, with the following null hypothesis (H_0):

Hypothesis 2a: H_0: The hourly TDD follows a NE distribution.
Hypothesis 2b: H_0: The hourly TDD follows a log-normal distribution.
Hypothesis 2c: H_0: The hourly TDD follows a Gamma distribution.

These functions are presented here:

$$\varphi_{NE}(t,x) = \frac{e(t)}{E(T)}\frac{1}{B(t)}\exp\left(\frac{-x}{B(t)}\right), \tag{7}$$

which has a single time-dependent parameter: the average trip length $B(t)$.

$$\varphi_{LN}(t,x) = \frac{e(t)}{E(T)}\frac{1}{x\sigma(t)\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-\mu(t))^2}{2\sigma^2(t)}\right). \tag{8}$$

which has two time-dependent parameters $\mu(t)$ and $\sigma(t)$.

$$\varphi_{Ga}(t,x) = \frac{e(t)}{E(T)}\frac{[\beta(t)]^{\alpha(t)}x^{\alpha(t)-1}\exp(-\beta(t)x)}{\Gamma(\alpha(t))}, \tag{9}$$

which also has two time-dependent parameters $\alpha(t)$ and $\beta(t)$.

In this paper, trips are aggregated in an hourly basis, creating the discrete time variable $t = \{0,...,23\}$. As an example, all trips that started between 7:52.5 AM and 8:52.5 AM will be sampled and labeled as $t = 8$. However, the presented methodology could be applied to any time interval aggregation, e.g. 15 minutes, 5 minutes, etc. if the data is detailed enough.

The rest of this section will explain the procedure of the calibration-validation method. For each hour of the day, the trips will be randomly divided in two samples: one calibration sample used for the maximum likelihood estimation of parameters; and one validation sample to perform statistical tests.

3.3. Calibration: Maximum likelihood estimation

The calibration will be done based on the maximum likelihood parameter estimation method. This method estimates the parameters of a given distribution from a sample of trips, e.g. $B(t)$ for the NE distribution **Equation 7** at time t, etc. For simplicity, we omit the time index t in the following. This parameter estimation method is a typical calibration technique that maximizes the log-likelihood, $\mathscr{L}$, that a set of observations are drawn from a distribution with one or more parameters. The likelihood of a given parameter z given $x_1,...,x_n$ observations is defined as $L(z|x_1,...,x_n) = \Pi_{i=1}^{n} L(z|x_i)$, where the likelihood $L(z|x_i) = \varphi(x_i;z)$. Then, the log-likelihood is defined as

$$\mathscr{L} = \ln\left(\Pi_{i=1}^{n}\varphi(x_i;z)\right), \tag{10}$$

which is maximized by solving $\frac{d\mathscr{L}}{dz} = 0$.

For the NE function **Equation 7** the parameter B can be estimated given a certain number n of measurements x_i. The likelihood is

$$L(B|x_1,...,x_n) = \left(\frac{1}{B}\right)^n \exp\left(-\frac{1}{B}\sum_{i=1}^{n} x_i\right). \tag{11}$$

And maximizing **Equation 10** with the likelihood in **Equation 11** leads to to $B^{ML} = \frac{\sum_{i=1}^{n} x_i}{n}$, where the superscript ML indicates that it has been estimated through the maximum log-likelihood method. Similarly, this can be done for distributions with more than one parameter. For the log-normal distribution **Equation 2**, the log-likelihood is:

$$\mathscr{L}(\mu,\sigma|x_1,...x_n) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \sum_{i}^{n}\ln(x_i) - \frac{\sum_i^n \ln(x_i)^2}{2\sigma^2} + \frac{\sum_i^n \ln(x_i)\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}.$$

Thus, the maximum likelihood estimators are

$$\mu^{ML} = \frac{\sum_{i=1}^{n}\ln(x_i)}{n} \text{ and } \sigma^{ML} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\ln x_i - \sum_{i=1}^{n}\frac{\ln x_i}{n}\right)^2}. \tag{12}$$

The same procedure could be applied to find the Gamma maximum likelihood estimators, but to solve $\frac{\partial\mathscr{L}}{\partial\alpha} = 0$ and $\frac{\partial\mathscr{L}}{\partial\beta} = 0$, leads to a fixed point problem. Thus, there is no closed form expression for these estimators. To obtain the solution a numerical scheme should be used, e.g., using a fixed point iteration scheme. However, there are many software that can do the maximum likelihood estimation for Gamma distribution without requiring to code or solve the fixed point problem. For this paper we will use Python's library *stats* to obtain the Gamma maximum likelihood estimators.

3.4. Validation: Kolmogorov-Smirnov tests

To test whether the probability distribution of a sample follows a reference probability distribution (RD), the Kolmogorov-Smirnov (KS) test can be used (*30, 31*). There are two types of KS test: one sample test and two sample test. In the first one the null hypothesis, H_0, is defined as: "the validation data set $x_1,...,x_n$ follows the estimated distribution from the calibration data set", which is how we define **Hypothesis 2a, 2b** and **2c**. The KS statistic for the one-sample test is defined as

$$D_n = \sup_x |\Omega_n(x) - \Omega(x)|, \tag{13}$$

where $\Omega_n(x)$ is the empirical cumulative distribution function, and $\Omega(x) = \int_0^x \varphi(s)ds$ is the reference cumulative distribution, which can be obtained from the calibration step in the previous section. The KS statistic quantifies the distance between the cumulative distribution function of the theoretical distribution and the empirical one (*31*). This test can only be used for continuous probability functions $\varphi(x)$. To determine whether the null hypothesis can be rejected or not, the KS statistic should be compared to the critical value. The null hypothesis, H_0, is rejected if the test statistic, D_n, is greater than the critical value, C_n, obtained from a table. For large sample sizes and for significance level $\alpha = 0.05$, the critical value can be approximated as

$$C_n(\alpha = 0.05) = \frac{1.36}{\sqrt{n}}, \tag{14}$$

where n is the sample size, for the one sample KS test.

For the two-sample KS test, the null hypothesis, H_0, is that both empirical data sets were sampled from populations with identical distributions. This test will be done to compare the empirical subsets (e.g. trip distance from different time of day or different days) and test whether they can be assumed to be a sample of the same TDD. If the hypothesis is rejected, the TDD of different times of day are not samples from the same distribution and it can be concluded that the TDD is time-dependent. The KS statistic for the two-sample test is

$$D_{n,m} = \sup_x |\Omega_n(x) - \Upsilon_m(x)|, \tag{15}$$

where $\Omega_n(x)$ is the cumulative distribution of the first sample and n is its sample size; and $\Upsilon_m(x)$ is the the second cumulative distribution with sample size m. Again, if the critical value $C_{n,m} < D_{n,m}$, the null hypothesis is rejected. The critical value can be approximated for large sample sizes as

$$C_{n,m}(\alpha = 0.05) = 1.36\sqrt{\frac{n+m}{nm}}. \tag{16}$$

4. EMPIRICAL DATA

4.1. Data Overview

The data used[2] corresponds to the information that Transportation Network Providers in the city of Chicago have collected since late 2018. Each trip recorded has unique identifier and information about the trip distance (in miles), duration (in seconds), starting and ending times (that are reported in 15 min intervals, by rounding the actual time to the nearest interval), trip fare and tip, and information whether other trips where pooled. The trips origins and destinations are zone-based, corresponding to the 77 community areas in Chicago, depicted in Figure 1. We are not interested in including the trips with origin or destination in the airport or other peripheral areas. For this reason, we consider a limited (more or less convex) area for the analysis, represented in yellow in Figure 1.

The data is cleaned to ensure that the trips recorded are meaningful, i.e., trips that lasted for less than 10 seconds or have average speeds higher than 80 mph or lower than 1mph are discarded from the data set. The data selected for this paper are trips that took place in 2019 that have both pick-up and drop-off points in the yellow community areas marked, which corresponds 45 community areas and 72.7M trips. The millions of trips initiated and ended in each community area and its associated mean distance is depicted in Figure 2. Clearly, the average trip distance is location-dependent. However, the analysis of location dependent trip distance is out of the scope of this paper. Recently, the spatial variation of ride-hail trip demand of this data set has been analyzed (*32*). Notice that the areas with higher number of trips correspond to the areas with lower average trip distance. These community areas (6, 7, 8 ,22, 24,28 and 32) also correspond to the downtown of Chicago, where many trips are likely to originate and end in the downtown area itself, and are

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

short trips.

The data analysis in this paper will be based on two different samples of trips from the data described. The first set of data analyzed will be all the trips over a given week. In particular, the trips that took place during week 11, since it corresponds to the week of 2019 with most trips recorded. This sample has a 1.63M trips with average trip distance of 3.47 miles and standard deviation of 2.80 miles. For this set, the trip generation rate and the mean trip distance (MTD) evolution over time will be analyzed in the following subsection. This will be done to study **Hypothesis 1**. Then, the time-dependent joint distribution will be estimated for 13th March 2019 (i.e. Wednesday of Week 11) in Section 5.1. This day had a total of 195119 trips with average distance of 3.53 miles and 2.85 miles standard deviation. As explained in Section 3, the estimation of the density function of trip distances will be done per hour. Each hour data will be split in two random equally large data samples, one for calibration and the other for validation. Finally, a larger data sample will be considered (∼3.64 M trips) to perform further calibration-validation analysis in Section 5.2. to study the variation of TDD over a single day and across days.

4.2. Time-dependent trip distance

In this section a detailed analysis of the demand over time is presented. The data sample used corresponds to the trips recorded on the Week 11 of 2019. As expected from common knowledge, the trip generation at different times of day is not the same, see Figure 3(a,b) where the trip initiation for the different days is depicted with respect to time. However, the underlying TDD could potentially be time-independent. A necessary condition for the trip distribution to be time-independent is that the average trip distance is time-independent. Therefore, we will do hypothesis testing on the MTD for each hour of data. A first look into the average trip distance over time in Figure 3(c,d) suggests that the average trip distance might be time-dependent. To test this, a "one-way ANOVA" test will be used, considering the following null hypothesis:

Hypothesis 3: H_0: The mean trip distance (MTD) is the same across hours, i.e. $\bar{x}_{t=0} = \bar{x}_{t=1} = ... = \bar{x}_{t=23}$.

Each day's MTD represents a data point for the groups "hour-of-day". This leads to 24 sets of data and 5 data point in each group. We verify that the assumption of normal distribution and homogeneity of variances are met for all 24 sets of data. Considering **Hypothesis 3**, the statistic for the ANOVA test is 37.9 and p-value of $8.06 \cdot 10^{-49}$. Thus, we reject the hypothesis that the MTD is time-independent (over day). Note that the MTD for Mon-Fri follows same trend. This suggests that the TDD may very within day, with a fixed pattern from day-to-day. This will be studied in Section 5.2.1. Further, we excluded from the analysis the MTD on the weekends, because it is reasonable to assume that the nature of these trips is different.

Another observation to highlight from Figure 3 is that the average trip distance is higher for the hours of day that the trip generation is lower. This is so for all the days of the week. Notice that this relation is consistent with the observations in the previous subsection (i.e. Figure 2) where regions with lower production and attraction of trips had longer average trip distances. It seems that during the early morning, where the demand is low (especially between 4 AM and 7 AM), longer trips are more likely to happen than during other times of the day, increasing the mean value. The reasons for this relation may be due to the nature of the demand. For example, it is natural that people with long trip distances choose another transportation mode than ride-hailing, especially during peak congestion time. Another possible explanation is that riders with larger trip distances try to avoid the early morning congestion by requesting a for-hire trip earlier. Alternatively, another reason to not observe these longer trip distances in the afternoon is that these commuters might use other modes of transportation that were unsafe (or not available) in the early morning. Finally, this relation might be explained with spatial variations of trip distances, as reported in Figure 2. All these assumptions should be tested in future studies.

The MTD and standard deviation for each hour are compared pairwise between weekdays and weekends in Figure 4. The figure also shows that the standard deviation is systematically lower than the MTD, which indicates that the assumption of NE distribution (*2, 3*) might not be adequate, since the expected value and standard deviation are equal in a negative exponential distribution. This will be tested in Section 5.1. and Section 5.2.2.

The aggregated trip distance probability mass function (PMF) across all week days is presented in Figure 5a, while individual days PMF are in Figure 5b for completeness purposes. These aggregated mass functions for the whole day have a similar form than the Gamma or log-normal probability density functions. However, the estimation of these aggregated distributions is not the purpose of the paper, because we have rejected **Hypothesis 1** and concluded that the TDD is time-dependent. In the next section we will perform a time-dependent calibration-validation analysis.

5. EMPIRICAL CALIBRATION AND VALIDATION

5.1. Joint distribution for single day

In this section a maximum-likelihood estimation from (**Equations 7 - 9**) is done for the 13th March 2019 trip data. First, we present the joint empirical probability mass function for that day in Figure 6a, where trips have been aggregated in an hourly way and the trip distance x is also discrete, in intervals of ~ 0.42 miles. First, the marginal distribution $\varphi(t)$ can be calibrated from **Equation 4**. This marginal mass function is non-convex due to the two peak periods. The conditional TDD $\varphi(x|t)$ is calibrated and validated for each hour of the day, $t \in [0,23]$, and we test **Hypothesis 2a, 2b, 2c** for each hourly sample data. The sample size for both calibration and validation is presented in Figure 6b.

The parameters from the estimation for the three distributions considered are omitted here for the sake of brevity. The KS statistics **Equation 13** are presented in Figure 7 for validation purposes. The subscript indicates what is the underlying distribution, i.e. $D_{n,NE}$ for the NE distribution, $D_{n,LN}$ for the log-normal distribution and $D_{n,Ga}$ for the Gamma distribution, where n is the sample size shown in Figure 6b. Figure 7 also presents the hourly critical values C_n, based on **Equation 14**. Comparing $D_{n,NE}$ and C_n, we reject **Hypothesis 2a** for all hours. By comparison of $D_{n,LN}$ and C_n, we reject **Hypothesis 2b** for all hours, except 4 AM. Thus, we fail to reject the null hypothesis that the samples comes from a log-normal distribution at 4 AM. Finally, we reject **Hypothesis 2c** for most hours by comparing $D_{n,Ga}$ and C_n, but from 2 AM to 4 AM the results of the KS statistic are inconclusive for the Gamma distribution.

Between 4 AM and 9 PM, we can see that $D_{n,LN}$ is much lower than the other two KS statistics. However, from 9-11 PM and at 12AM, 2AM and 3 AM the Gamma distribution has the lowest KS statistic. Moreover, notice $D_{n,LN}$ is remarkably lower than the other two statistics for larger sample sizes, and the statistic increases for lower sample sizes. On the other hand, the KS statistic for the other two distributions is lower for smaller sample sizes. The hypothesis of the KS test is that "the maximum difference between the empirical and the theoretical cumulative density functions tends to zero for increasingly sample sizes". Thus, for most of the hours the log-normal distribution it is the best assumption from the three considered. Therefore, the "best" estimated joint distribution for this day considered in this paper is

$$\varphi_{LN}(t,x;\mu(t),\sigma(t)) = \frac{e(t)}{195119}\frac{1}{x\sigma(t)\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-\mu(t))^2}{2\sigma^2(t)}\right),$$

presented in Figure 8a. For a better representation Figure 8b-e presents the cuts of the joint probability $\varphi(t,x)$ at different times. The time-dependent parameters are depicted in Figure 9a.

For the sake of comparison with the aggregated TDD for the whole day, the calibration and validation of the log-normal distribution is also done, see grey horizontal lines in Figure 9a. Compared to the parameters obtained from other cities, σ^{ML} lies between the range observed in other cities (*14*), i.e. $\sigma^{ML} \in [0.71.2]$, while μ^{ML} is about half of the reported lower bound. The estimation for the 24h data set underestimates μ^{ML} for early morning and over-estimates it for the peak periods and during midday, compared to the proposed time-dependent calibration. On the other hand, the estimation of σ^{ML} is generally underestimated and only overestimated during the rush hour periods. Figure 9b presents again the hourly $D_{n,LN}$ and C_n and the whole day KS statistic and critical value for comparison. Notice that the difference (both absolute and relative) between the D_n and C_n is much larger than when estimating a TDD for an hourly base. Thus, a better estimation of the TDD distribution can be achieved if it is done considering time-dependent parameters.

5.2. Time variation analysis

In this section we consider a larger data sample (from multiple days), to ensure the robustness of the results reported for a single day. Moreover, we are interested in studying whether time variation in TDD is larger across days or within days. To do so, we consider the data collected during four different hours (2 off-peak and 2 peak hours) for all weekdays (except Friday) in 2019. Therefore, there is a total of 16 data sets, which are labeled as in Table 1. We will focus only on the conditional distribution estimation, $\varphi(x|t)$, since the data is sampled for certain hours of the day.

TABLE 1: Summary of Data Sets for analysis in Section 5.2.

Time of day	Mondays	Tuesdays	Wednesdays	Thursdays
4-5 AM	Set 1	Set 5	Set 9	Set 13
8-9 AM	Set 2	Set 6	Set 10	Set 14
1-2 PM	Set 3	Set 7	Set 11	Set 15
7-8 PM	Set 4	Set 8	Set 12	Set 16

The cumulative mass function and the cross-comparison across days and hours of days is presented in Figure 10. A first look into the cumulative mass function highlights the similarity between days for each given hour. In the following we analyze only the Sets 5-12 (i.e. Tuesdays and Wednesdays), because they correspond to the days with higher reported difference in the cumulative mass functions, see Figure 10(f,i). The pair-wise comparison for all 4 hours is depicted in Figure 11(a). The similitude between the early morning hours (4-5 AM) is not as evident as the other hours analyzed. Figure 11(b) presents the differences between the cumulative mass functions on Tuesdays for different hours of day compared to the morning peak. This figure reinforces the hypothesis in Section 4.2 that the TDD time-dependency changes more over a single day than across days of the week.

5.2.1. Two sided KS test

In this section we consider the following hypothesis:

Hypothesis 4: H_0: The empirical samples from Set x and Set y come from the same underlying distribution.

To test it, we will use the two-sample KS test. The KS statistic **Equation 15** are presented in Figure 12. In particular the data sets of Wednesday samples are tested among each-other in Figure 12a; and against the data sets in Tuesday in Figure 12b. The pairs with darker color present more similitude among their cumulative distribution. The diagonal comparison in Figure 12a is ignored, since it compares the same sample. The $C_{n,m}$ from **Equation 16** is lower than 0.0089 in all cases. All the null hypothesis are rejected, except one. The test to compare Set 7 with Set 11 leads to a KS statistic of $D_{n,m} = 0.0019 < 0.0031 = C_{n,m}$ and we fail to reject the null hypothesis in this case.

The diagonal terms in Figure 12b, are darker than the non-diagonal squares in 12a, which indicate that the distributions from the same hour are more similar across days than the distributions of a given day for different hours. The peak hour TDD have higher similitude than the off-peak TDD. Furthermore, the lightest pairs, which correspond to higher differences in the cumulative distributions are always for data Sets 5 and 9, which corresponds to trips sampled at 4 AM. This suggests that TDD during the early morning is notably different than the rest, as suggested by Figure 11.

5.2.2. Calibration-validation of Tuesday and Wednesday subsets

In this subsection we will perform the validation and calibration as presented in Section 3 of different samples of the data sets corresponding to Tuesday and Wednesday (Set 5 to Set 12). The results are shown in the first part of Table 2. The log-likelihood is calculated from the calibrated data as **Equation 10**. From the results in the Table 2 we see that for each data set the maximum log-likelihood is obtained for the calibration of log-normal distribution, except for the Sets 5 and 9. For these sets the log-likelihood from Gamma distribution fit is either larger or equal to the log-likelihood from the log-normal calibration. Notice that the $\mathscr{L}$ for the NE distribution is consistently lower than the other two, indicating a worse fit to the data.

The KS statistic, D_n, is presented in the third part of Table 2, for each TDD. For the given sample sizes of all sets (Figure 13a black color), the KS critical value is much lower (≤ 0.0065) than any D_n. Thus, based on the sample trips from all 2019, the null hypothesis should be rejected for all distributions considered. However, the KS statistic for the log-normal distribution is systematically lower than the other for all sets of data considered. As a graphical representation, the empirical validation data histograms are presented in Figure 14, with the superposition of the theoretical NE distribution, LN and Gamma conditional distributions estimated from the calibration sample. From visual inspection it is clear that the log-normal distribution captures better the mode of the trip distance.

Recall that KS critical value decays inversely proportional to square root of n with increasing sample size, as per **Equation 14**. In general, KS tests are performed with samples sizes of less than 10,000 data points, i.e. critical $C_n(\alpha = 0.05) > 0.0136$. Thus, the same calibration and validation procedure is done for different samples. Five samples are drawn and tested for each set, presented in Figure 13a with their corresponding sample sizes. Using different samples (with different sizes) multiple one-sample KS tests can be performed to ensure that the conclusion of the

TABLE 2: Calibration and Validation results for the full sets (all year data).

	NE Distribution	**LN Distribution**	**Gamma Distribution**
Parameters			
Set 5	B^{ML}=4.807	$\mu^{ML} = 1.237$; $\sigma^{ML} = 0.862$	$\alpha^{ML} = 1.650$; $\beta^{ML} = 0.343$
Set 6	B^{ML}=3.346	$\mu^{ML} = 0.932$; $\sigma^{ML} = 0.750$	$\alpha^{ML} = 1.963$; $\beta^{ML} = 0.587$
Set 7	B^{ML}=3.431	$\mu^{ML} = 0.912$; $\sigma^{ML} = 0.809$	$\alpha^{ML} = 1.708$; $\beta^{ML} = 0.498$
Set 8	B^{ML}=3.254	$\mu^{ML} = 0.902$; $\sigma^{ML} = 0.757$	$\alpha^{ML} = 1.949$; $\beta^{ML} = 0.599$
Set 9	B^{ML}=4.788	$\mu^{ML} = 1.237$; $\sigma^{ML} = 0.858$	$\alpha^{ML} = 1.667$; $\beta^{ML} = 0.348$
Set 10	B^{ML}=3.304	$\mu^{ML} = 0.918$; $\sigma^{ML} = 0.751$	$\alpha^{ML} = 1.953$; $\beta^{ML} = 0.591$
Set 11	B^{ML}=3.397	$\mu^{ML} = 0.900$; $\sigma^{ML} = 0.812$	$\alpha^{ML} = 1.696$; $\beta^{ML} = 0.499$
Set 12	B^{ML}=3.241	$\mu^{ML} = 0.903$; $\sigma^{ML} = 0.749$	$\alpha^{ML} = 1.981$; $\beta^{ML} = 0.611$
Log-likelihood, $\mathscr{L}$			
Set 5	$-6.31 \cdot 10^{-4}$	$-6.16 \cdot 10^{-4}$	$-6.15 \cdot 10^{-4}$
Set 6	$-7.31 \cdot 10^{-5}$	$-6.83 \cdot 10^{-5}$	$-6.94 \cdot 10^{-5}$
Set 7	$-4.33 \cdot 10^{-5}$	$-4.11 \cdot 10^{-5}$	$-4.19 \cdot 10^{-5}$
Set 8	$-7.79 \cdot 10^{-5}$	$-7.29 \cdot 10^{-5}$	$-7.40 \cdot 10^{-5}$
Set 9	$-5.75 \cdot 10^{-4}$	$-5.60 \cdot 10^{-4}$	$-5.60 \cdot 10^{-4}$
Set 10	$-7.37 \cdot 10^{-5}$	$-6.88 \cdot 10^{-5}$	$-7.00 \cdot 10^{-5}$
Set 11	$-4.31 \cdot 10^{-5}$	$-4.09 \cdot 10^{-5}$	$-4.17 \cdot 10^{-5}$
Set 12	$-7.91 \cdot 10^{-5}$	$-7.39 \cdot 10^{-5}$	$-7.50 \cdot 10^{-5}$
KS statistic, D_n			
Set 5	0.109	0.0333	0.0509
Set 6	0.169	0.0474	0.0784
Set 7	0.144	0.0539	0.0875
Set 8	0.165	0.0287	0.0590
Set 9	0.114	0.0419	0.0452
Set 10	0.170	0.0433	0.0762
Set 11	0.148	0.0433	0.0787
Set 12	0.167	0.0296	0.0581

test is not affected by the sample size. The D_n for all these data samples are presented in Figure 13b. For all, the null hypotheses from **Hypothesis 2a, 2b, 2c** are rejected. Therefore, we conclude that the trip distance distribution does not follow neither a NE, a log-normal nor a Gamma distribution. We have also the following observations from Figure 13b:

- Increasing the sample size increases the KS statistic for the NE distribution. In other words, the larger the sample size, the larger the difference is between the sample and NE distribution.
- For the log-normal and Gamma distributions a larger sample size does not affect much D_n. However, for lower sample sizes the Gamma and LN validations lead to similar KS statistics, while for larger sample sizes the log-normal leads to consistently lower D_n than the Gamma distribution.
- For very large samples, the $D_{n,LN}$ is systematically lower than D_n from the other two distributions considered. This result corroborates the conclusion in Section 5.1., that the best approximation would be to assume that TDD follows a time-dependent log-normal distribution.

6. PRACTICAL IMPLICATIONS, CONTRIBUTIONS AND LIMITATIONS

Travel demand can be obtained from a travel demand model (four-step model, activity-based model, etc.) or calibrated from empirical data. In this paper, we focus on the calibration of travel demand, especially the TDD, for transportation system analysis in a relative space dimension (*2–4*). TDD is not only important to model traffic congestion, but also for designing and evaluating mileage tax and other real-time operation and management schemes, including transit scheduling and distance-based congestion pricing in high-occupancy-toll lanes. Further, the TDD can also be used to calibrate and validate the resulting trips from activity-based models.

In this paper, we propose a calibration procedure to find the demand pattern empirically through the definition of the joint distribution of departure time and trip distance and maximum likelihood estimation. We used the Chicago data set to do some statistical tests on the TDD. We have rejected most of the hypotheses considered in this paper. From **Hypotheses 1, 3 and 4**, the TDD should be considered time-dependent, which is consistent with another recent study (*9*). Testing **Hypothesis 2** showed that the NE distribution is a very bad approximation for TDD, compared to the other two distributions considered (Gamma, as a generalization of NE distribution, and log-normal distribution, assumed by some researchers). Therefore, the trip dynamics of Vickrey's bathtub model are only an imprecise approximation of the actual bathtub dynamics.

However, none of the distributions considered in this paper present a good fit for TDD. From the behavioral point of view, the distribution will depend on land use, mode choice, and departure time choice of travelers. Whether a generic TDD can be found to describe any demand pattern in any region is an open and challenging question. It is likely that a site and mode-specific calibration is needed for each case. This highlights the importance of using the generalized bathtub model (*4*) over the most common bathtub models used in the literature that assume directly (*2, 3*) or indirectly (*10, 11*) a time-independent negative exponential distribution of trip distances, or constant trip distances (*5, 25*).

In summary, the contributions of this paper are threefold. First, it provides a systematic way to define and estimate the demand in the space-time domain, where the space dimension is with respect to the remaining trip distances. This is achieved defining the joint distribution of departure time and trip distance as the product of the marginal and conditional distribution. Second, based on empirical data of Chicago for-hire trips, it proves through statistical methods that the TDD is time-dependent, and that the variation is larger within the day than across days. Third, it rejects the hypothesis that the trip distribution follows a negative exponential distribution (as assumed in Vickrey's bathtub model), a log-normal distribution (considered by several authors), or the Gamma distribution. Other observations from the empirical data analyzed are that the average trip distance changes both spatially and in a temporal manner. Also, under peak periods the average trip distance is lower, while in the early morning a higher percentage of longer trips take place. In all, a case-by-case data-based demand calibration might be needed for real-time control and traffic modeling. Even if a site-specific calibration of demand is needed, the relative space dimension paradigm allows for easier calibration and definition of demand than the traditional OD demand estimation problem in transportation analysis.

Finally, it is important to highlight that the data set used is from for-hire vehicles. For this reason, the results of this paper might not be representative of other types of trips. We are only analyzing data from a sample of the total demand, i.e., considering only a mode. Thus, to have a better understanding of the total demand pattern, further calibration of other empirical data sources is required, ideally based on privately owned vehicle-trips. Since we conclude that the most common assumptions in the literature on TDD are not adequate for for-hire trips, we want to highlight the importance of carefully revising and selecting the demand assumptions for other modes. For this reason, studying empirical trip distances of other modes is important to define more adequate assumptions on the demand in the future.

7. CONCLUSIONS

The study of TDD has been gaining interest in the research community due to the newly available data collection technologies. In this paper, we consider a different type of demand definition, where the geographical location of the origins and destinations is ignored. We have presented a calibration and validation procedure to study the demand as a joint probability density function of trip distance and departure time. In the past, the TDD has been studied and calibrated, but always based on daily (or larger temporal-scale) aggregation of trips. However, the TDD could depend on the day in the same way that the trip initiation rate changes during the day. In this paper, we consider a time-dependent TDD and propose a methodology to calibrate the joint distribution of trip distance and trip initiation rate.

Further, using the for-hire trips data set from Chicago we prove through hypothesis testing that TDD is indeed time-dependent. For a day in 2019, we have performed both calibration and validation of three different joint distributions (time-dependent NE, log-normal, and Gamma distributions). With statistic tests, such as the Kolmogorov-Smirnov test we have rejected the hypothesis that samples of trips at different times of day follow any of the considered distributions. Among the three distributions tested, the log-normal is the most promising, because it consistently leads to lower KS statistics and higher log-likelihood. One of the time-dependent parameters estimated for the log-normal distribution matches well with other calibration in other cities (*14*), while the other parameter is consistently lower.

In summary, this paper highlights the importance of considering other distributions in the future to study the demand. Future research should consider the calibration and validation of Weibull distribution, Generalized Extreme Value distribution, and even consider non-parametric calibration (*33*). In the future, we are also interested in performing a location-dependent TDD analysis. The proposed methodology can be applied to smaller regions, even considering each individual community area. This type of study might bring some light into mobility patterns and also regional trip length analysis.

ACKNOWLEDGMENTS

This research has been partially funded by the Graduate Balsells Fellowship program. The second author would like to acknowledge the support of NSF-SCC- CMMI#1952241, entitled "SCC-PG: Addressing Unprecedented Community-Centered Transportation Infrastructure Needs and Policies for the Mobility Revolution".

The authors confirm their contribution to the paper as follows: study conception and design: I.M. and W-L.J.; statistical analysis: I.M.; interpretation of results: I.M. and W-L.J.; draft manuscript preparation: I.M. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Cervero, R. and K. Kockelman, Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, Vol. 2, No. 3, 1997, pp. 199–219.

[2] Vickrey, W., Congestion in midtown Manhattan in relation to marginal cost pricing. *Unpublished notes, Columbia University*, 1991.

[3] Vickrey, W., Congestion in midtown Manhattan in relation to marginal cost pricing. *Economics of Transportation*, Vol. 21, No. January, 2020, p. 100152.

[4] Jin, W.-L., Generalized bathtub model of network trip flows. *Transportation Research Part B*, Vol. 136, 2020, pp. 138 – 157.

[5] Arnott, R., A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, Vol. 76, 2013, pp. 110 – 121.

[6] Yildirimoglu, M. and N. Geroliminis, Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. *Transportation Research Part B: Methodological*, Vol. 70, 2014, pp. 186–200.

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

[8] Mariotte, G., L. Leclercq, S. F. A. Batista, J. Krug, and M. Paipuri, Calibration and validation of multi-reservoir MFD models: A case study in Lyon. *Transportation Research Part B: Methodological*, Vol. 136, 2020, pp. 62–86.

[9] Paipuri, M., Y. Xu, M. C. González, and L. Leclercq, Estimating MFDs, trip lengths and path flow distributions in a multi-region setting using mobile phone data. *Transportation Research Part C: Emerging Technologies*, Vol. 118, 2020, p. 102709.

[10] Small, K. A. and X. Chu, Hypercongestion. *Journal of Transport Economics and Policy*, Vol. 47, 2003, pp. 19–52.

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

[12] González, M. C., C. A. Hidalgo, and A. L. Barabási, Understanding individual human mobility patterns. *Nature*, Vol. 453, No. 7196, 2008, pp. 779–782.

[13] Calabrese, F., M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti, Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, Vol. 26, 2013, pp. 301–313.

[14] Colak, S., A. Lima, and M. C. Gonzalez, Understanding congested travel in urban areas. *Nature communications*, Vol. 7, 2016, p. 10793.

[15] Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira, and K. W. Axhausen, Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information. *Transportation Research Record*, Vol. 1870, No. 1, 2004, pp. 46–54.

[16] Gong, L., T. Morikawa, T. Yamamoto, and H. Sato, Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences*, Vol. 138, 2014, pp. 557–565.

[17] Thomas, T. and S. Tutert, An empirical model for trip distribution of commuters in The Netherlands: transferability in time and space reconsidered. *Journal of Transport Geography*, Vol. 26, 2013, pp. 158 – 165.

[18] Cich, G., L. Knapen, T. Bellemans, D. Janssens, and G. Wets, TRIP/STOP Detection in GPS Traces to Feed Prompted Recall Survey. *Procedia Computer Science*, Vol. 52, 2015, pp. 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).

[19] Geroliminis, N. and C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 759 – 770.

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

[21] Ramezani, M., J. Haddad, and N. Geroliminis, Dynamics of heterogeneity in urban networks: Aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological*, Vol. 74, 2015, pp. 1–19.

[22] Fosgerau, M., Congestion in the bathtub. *Economics of Transportation*, Vol. 4, No. 4, 2015, pp. 241–255.

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

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

[25] Arnott, R. and J. Buli, Solving for equilibrium in the basic bathtub model. *Transportation Research Part B: Methodological*, Vol. 109, 2018, pp. 150–175.

[26] de Dios Ortúzar, J. and L. G. Willumsen, *Trip Distribution Modelling*, John Wiley & Sons, Ltd, chap. 5, pp. 175–206, 2011.

[27] de Vries, J. J., P. Nijkamp, and P. Rietveld, *Exponential or power distance decay for commuting? An alternative specification.* Tinbergen Institute Discussion Paper, No. 04-097/3, Amsterdam and Rotterdam, 2004.

[28] Yang, H., J. Ke, and J. Ye, A universal distribution law of network detour ratios. *Transportation Research Part C: Emerging Technologies*, 2018.

[29] Lehe, L. J., Downtown tolls and the distribution of trip lengths. *Economics of Transportation*, Vol. 11-12, No. October, 2017, pp. 23–32.

[30] Conover, W., *Practical Nonparametric Statistics*, Vol. II, chap. Chapter 6. 3rd ed., 1999.

[31] Chakravarti, I. M., R. G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*. John Wiley Sons, 1967.

[32] Dean, M. D. and K. M. Kockelman, Spatial variation in shared ride-hail trip demand and factors contributing to sharing: Lessons from Chicago. *Journal of Transport Geography*, Vol. 91, 2021, p. 102944.

[33] Kvam, P. H. and B. Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering*, 2007.

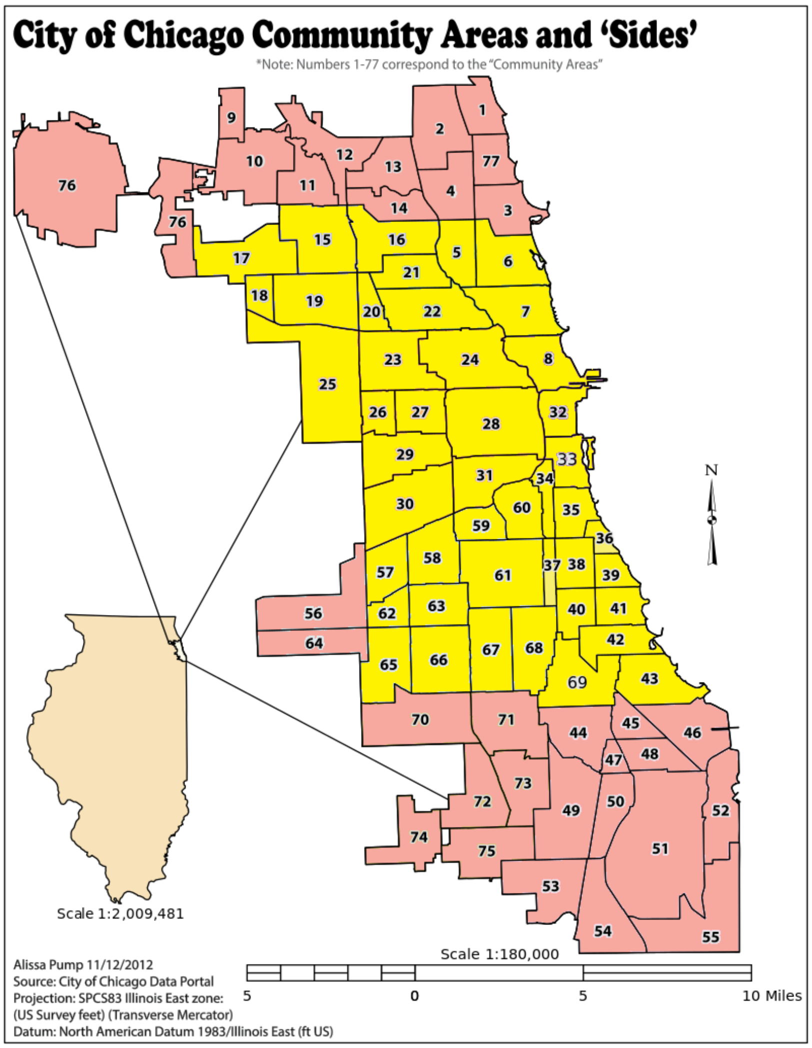


FIGURE 1: Chicago Community Areas. Zone of interest in yellow.

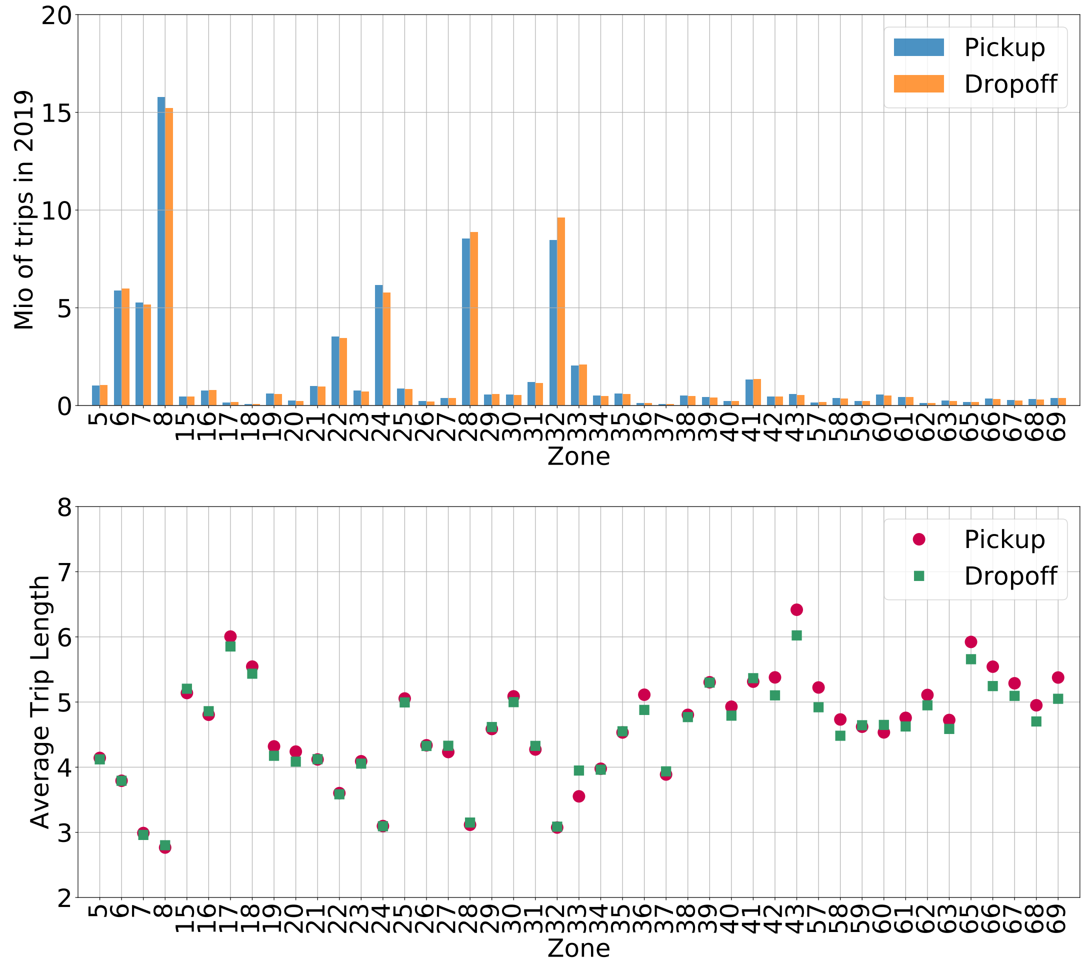


FIGURE 2: Millions of strips started/completed in each one the community zone of interest and associated average trip distance.

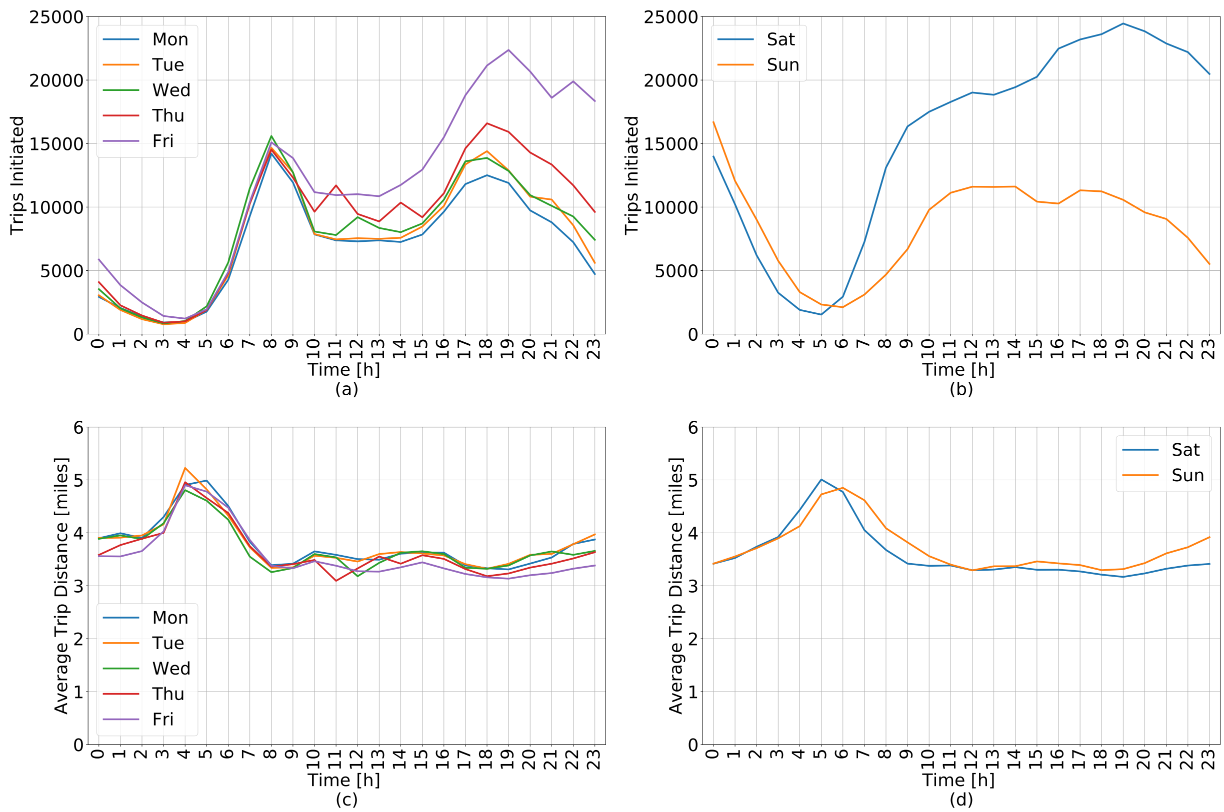


FIGURE 3: Trip initiation and distance analysis in zone of interest for week 11 of 2019. (a) Trip generation at different time of day in weekdays. (b) Trip generation at different hours on the weekend. (c) Average trip distance of trips initiated in each hour for weekdays. (d) Average trip distance of trips initiated in each hour for Saturday and Sunday.

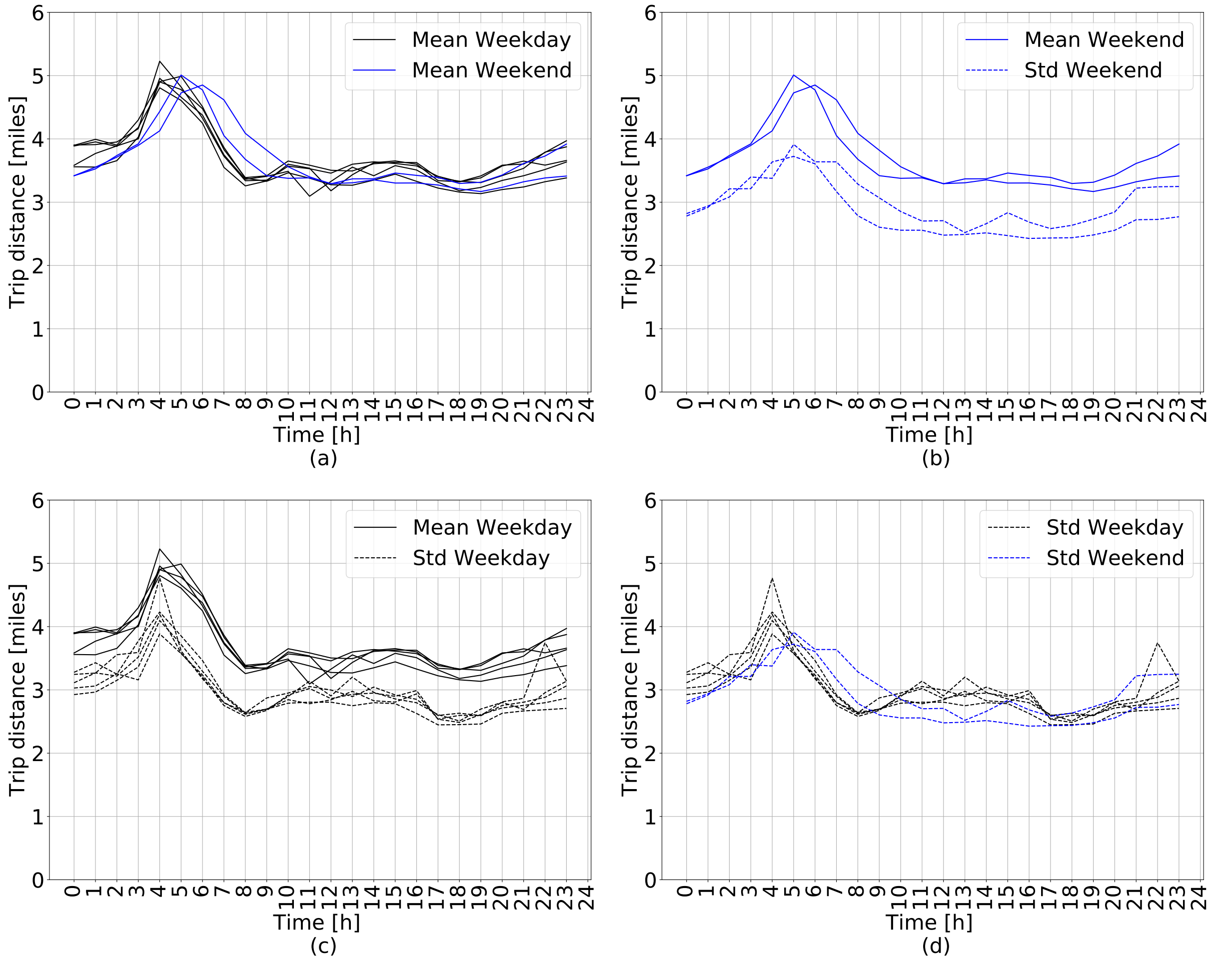
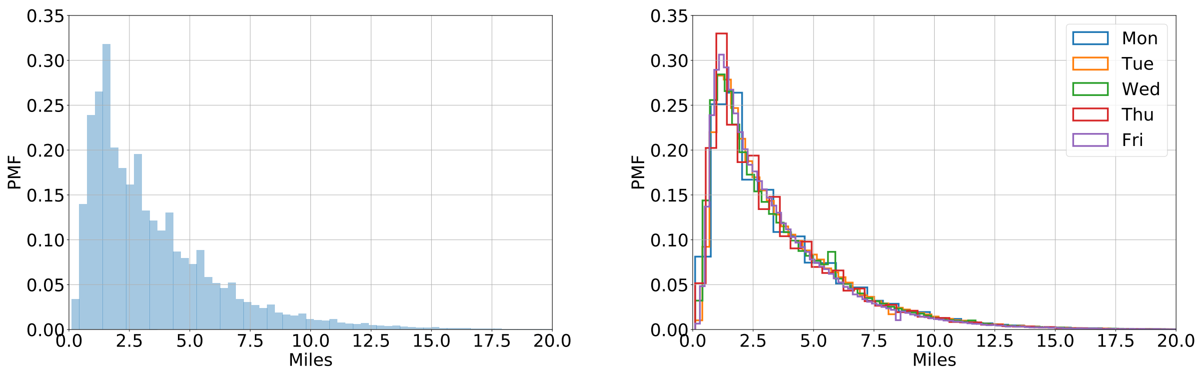
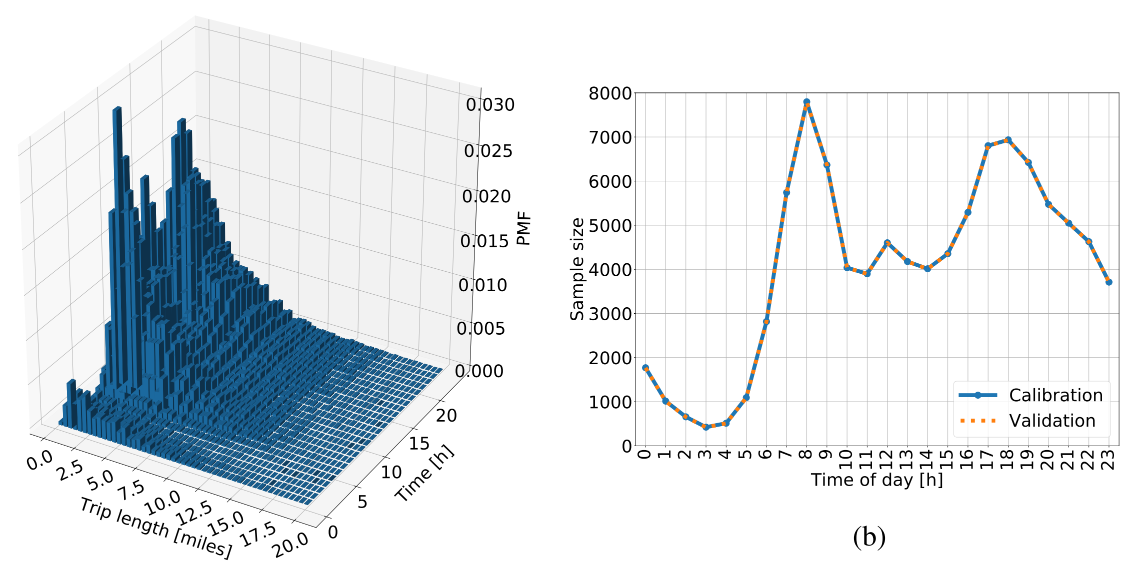


FIGURE 4: Trip distance mean and standard deviations over time for weekdays (black) and weekends (blue). (a) Compares MTD between weekdays and weekends, (b) Compares mean and standard deviation of trip distance for weekends, (c) compares mean and standard deviation of trip distance for weekdays and (d) compares standard deviation of trip distance between weekdays and weekends.



(a) (b)

FIGURE 5: Trip length empirical probability mass functions for (a) All days of the week 11 of 2019 (b) Individual work days, i.e. March 11th to March 15th.



(a) (b)

FIGURE 6: Joint distribution for trip distance and trip initiation for Wednesday 13th Match 2019 (a) Empirical joint probability mass function; (b) Sample sizes of calibration and validation subsets for each hour.

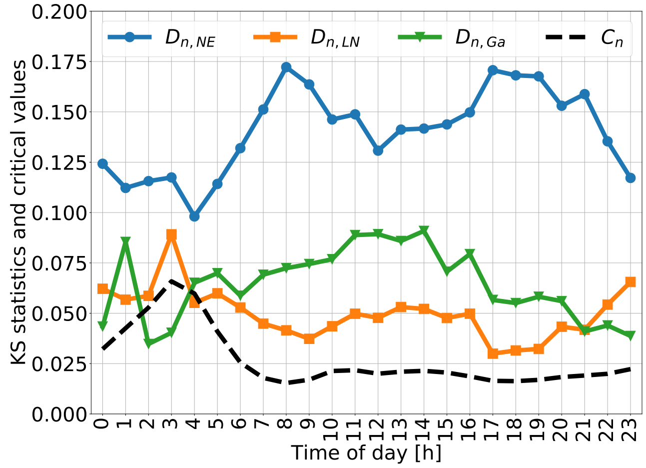


FIGURE 7: KS statistic for each hourly estimation for the three different TDD considered. Blue $D_{n,NE}$ for the NE distribution, orange $D_{n,LN}$ for the log-normal distribution and green $D_{n,Ga}$ for the Gamma distribution. The critical values C_n depend on the sample size of each hourly data and are depicted in a dashed black line.

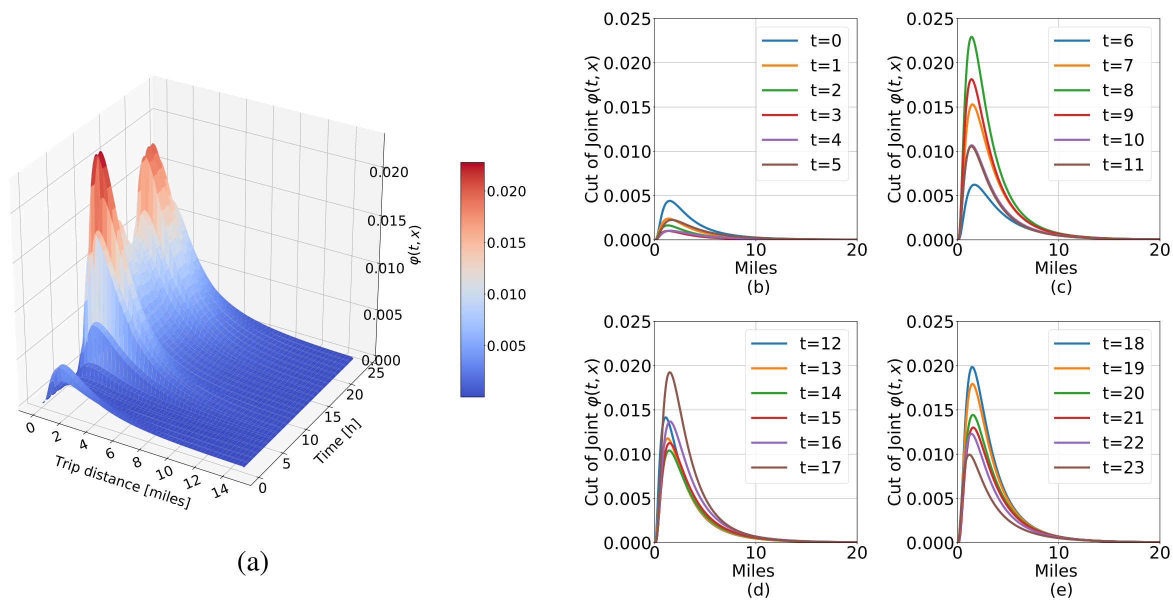
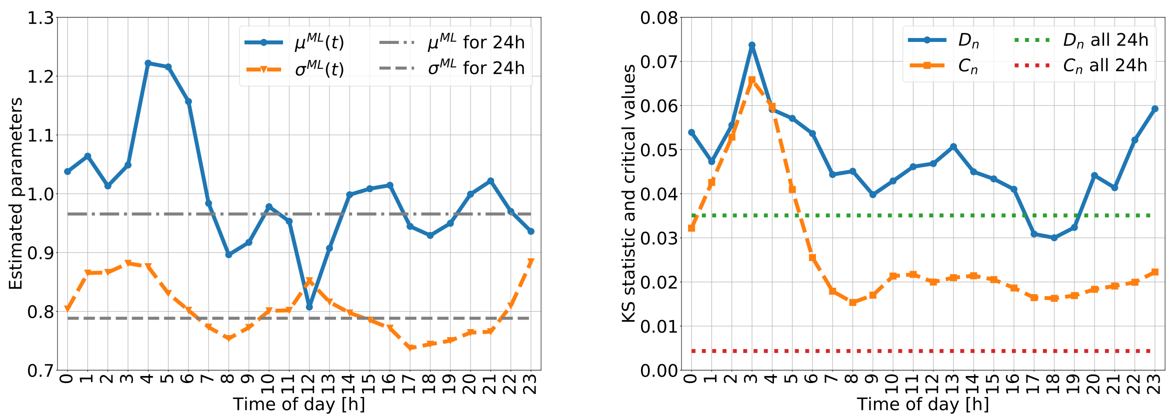
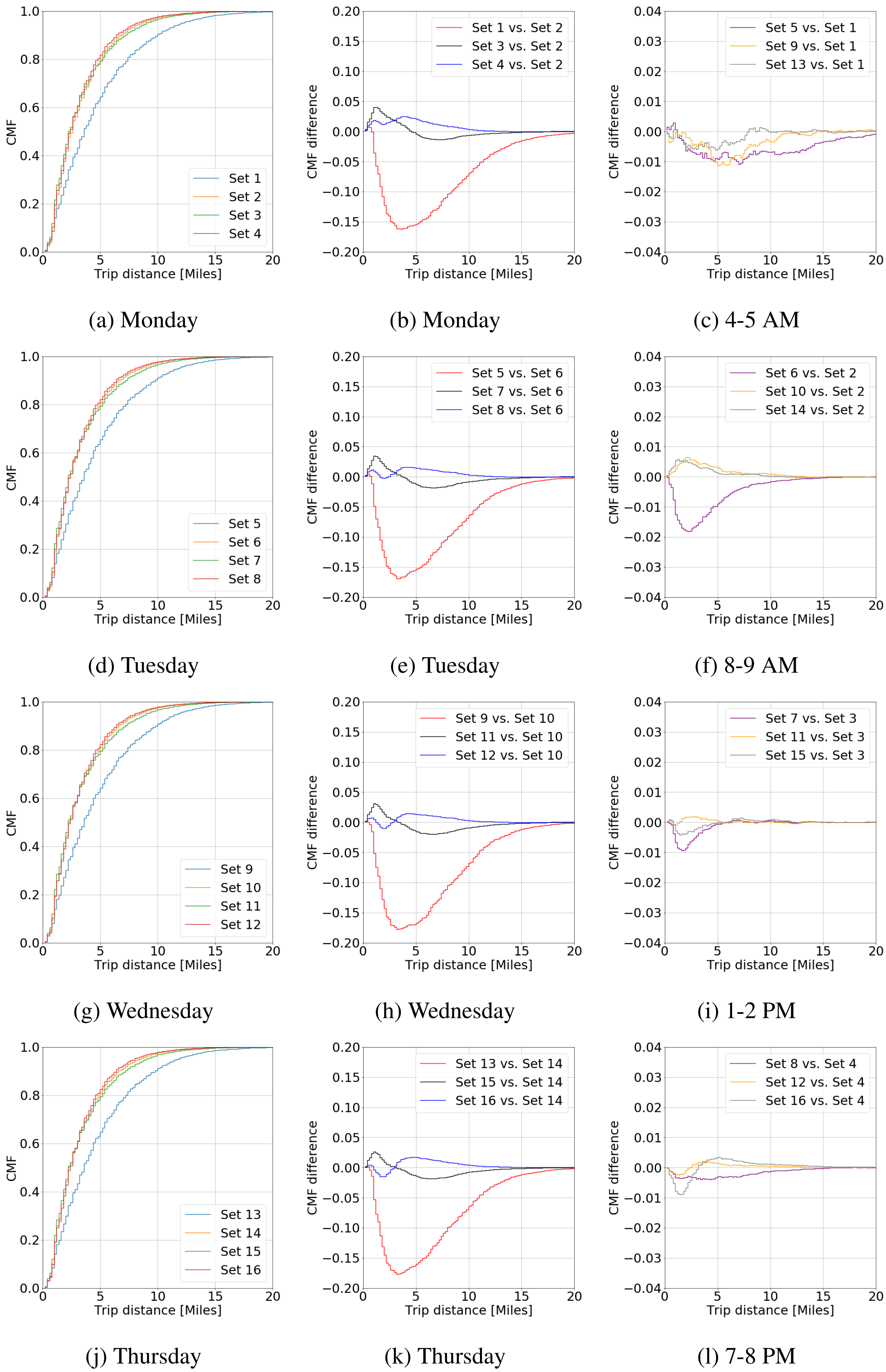


FIGURE 8: (a) Calibrated joint distribution, (b-e) Cut of joint distribution represented as a curve for each hour $t \in [0, 23]$, i.e. the conditional probability of trip distance scaled down by the probability of a trip starting at that time interval.



(a) (b)

FIGURE 9: (a) time-dependent parameters of log-normal estimated parameters and the estimated parameters for the whole day. (b) The KS statistic D_n and the critical KS value C_n for each hour and for the 24h analysis.



(a) Monday (b) Monday (c) 4-5 AM

(d) Tuesday (e) Tuesday (f) 8-9 AM

(g) Wednesday (h) Wednesday (i) 1-2 PM

(j) Thursday (k) Thursday (l) 7-8 PM

FIGURE 10: (a,d,g,j) Cumulative mass functions for the different sets of data described in Table 1. (b,e,h,k) Difference in cumulative mass functions for a given day across time of day (d,f,i,l) Difference in cumulative mass functions across days for a given time.

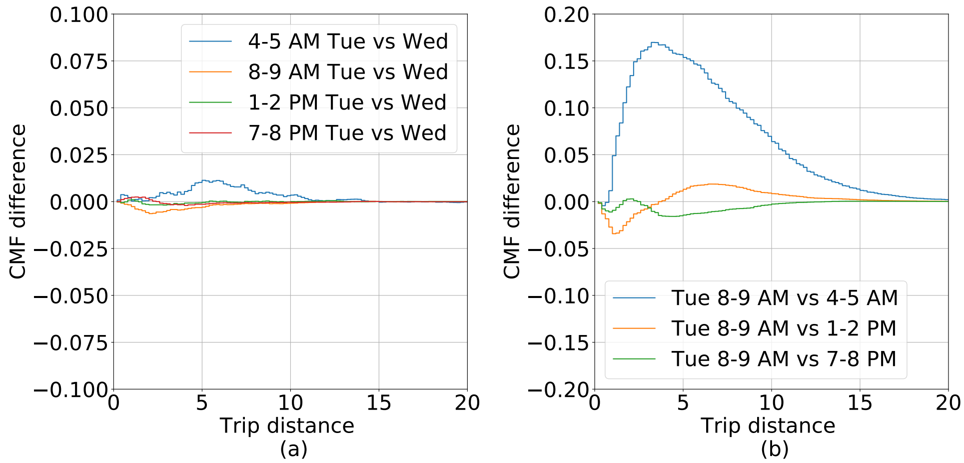


FIGURE 11: Trip distance cumulative mass function difference across subsets. (a) Comparison between Tuesday and Wednesday (b) Comparison within day for Tuesday data.

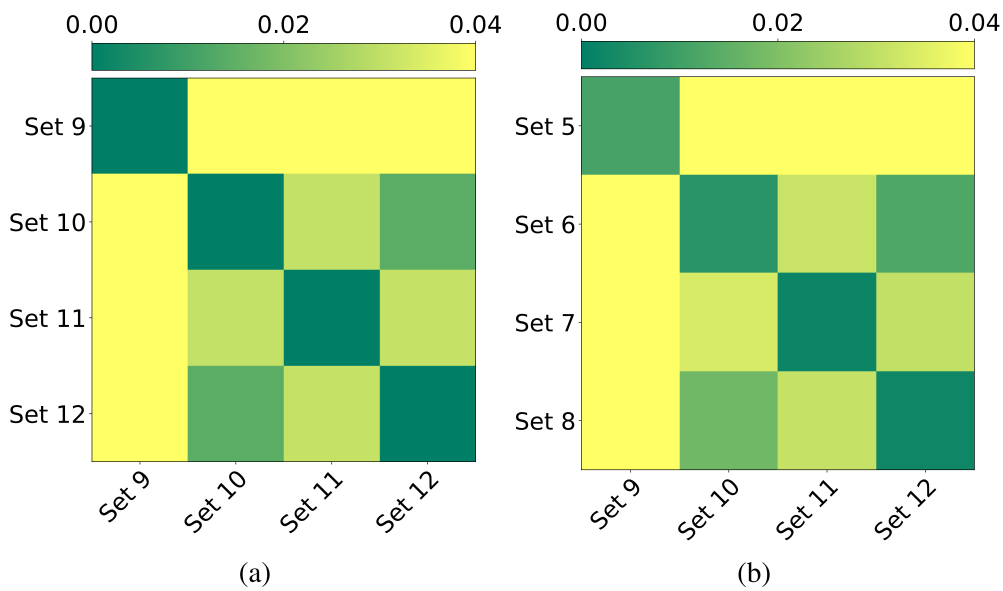
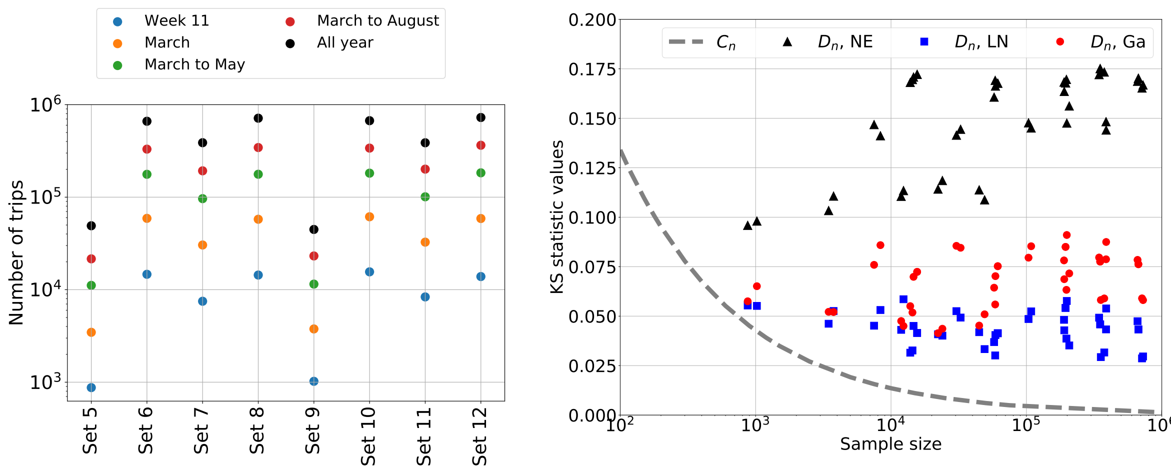
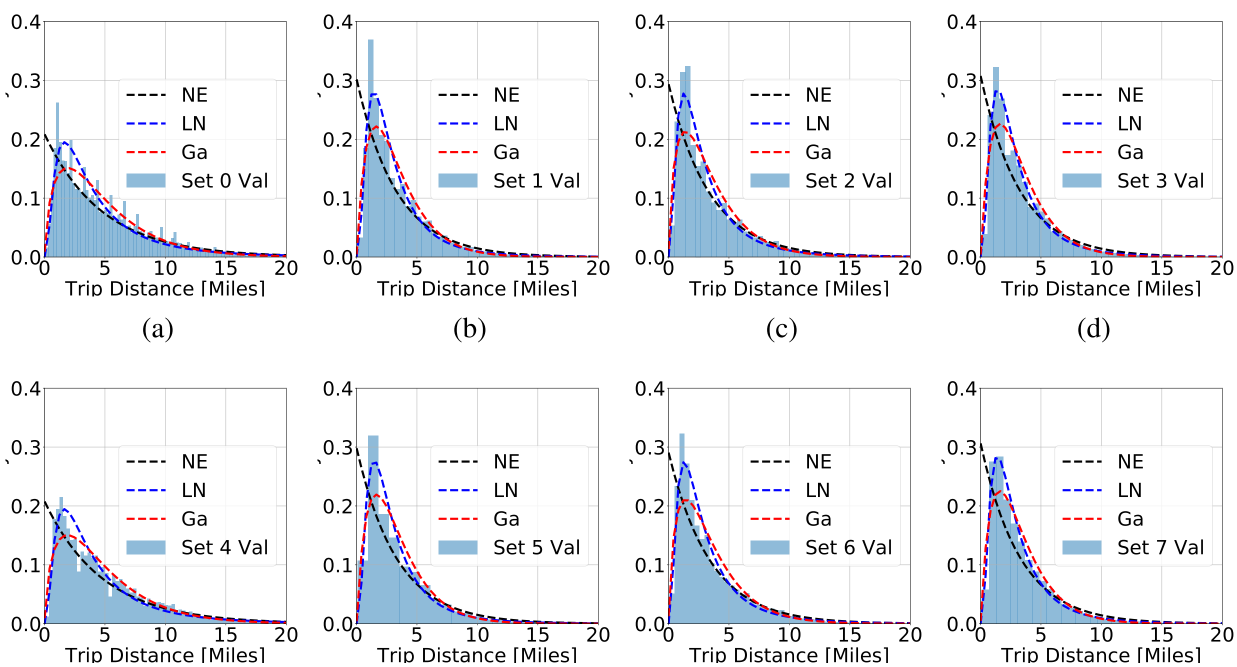


FIGURE 12: Two sample KS statistic to compare data sets. The color of each square represents the value of the KS statistic, $D_{n,m}$. (a) Compares the 4 hourly samples for Wednesday among each other. (b) Compares the hourly data sets from Tuesdays to the ones from Wednesdays.



(a) (b)

FIGURE 13: (a) Number of trips for each sample considered. Each color represents a sample of the data set during a different period of time; (b) Analysis of KS statistic, D_n, from each TDD considered for different sample sizes against the KS critical value, C_n (dashed grey line).



(a) (b) (c) (d)

(e) (f) (g) (h)

FIGURE 14: Empirical conditional TDD of validation subsets and the three maximum likelihood calibration distributions, $\varphi(x|t)$, obtained from calibration subsets with parameters in Table 2.