Louisiana Transportation Research Center

Final Report 646

ITE Trip Generation Modification Factors for Louisiana

by

Chester Wilmot, Ravindra Gudishala, Saba Doulabi, Mishuk Majumder*
Peter Stopher**
Angela Antipova***

****Louisiana State University***
*****Private Consultant***
******University of Memphis***

4101 Gourrier Avenue | Baton Rouge, Louisiana 70808
(225) 767-9131 | (225) 767-9108 fax | www.ltrc.lsu.edu

TECHNICAL REPORT STANDARD PAGE

1. Title and Subtitle
   ITE Trip Generation Modification Factors for Louisiana

2. Author(s)
   Chester Wilmot, Peter Stopher, Angela Antipova, Ravindra Gudishala, Saba Doulabi, and Mishuk Majumder

3. Performing Organization Name and Address
   Department of Civil and Environmental Engineering
   Louisiana State University
   Baton Rouge, Louisiana 70803

4. Sponsoring Agency Name and Address
   Louisiana Department of Transportation and Development
   P.O. Box 94245
   Baton Rouge, LA 70804-9245

5. Report No.
   FHWA/LA.17/646

6. Report Date
   April 2021

7. Performing Organization Code
   LTRC Project Number: 18-4SS
   SIO Number: DOTLT1000212

8. Type of Report and Period Covered
   Final Report
   July 2017 – December 2020

9. No. of Pages
   110

10. Supplementary Notes
    Conducted in Cooperation with the U.S. Department of Transportation, Federal Highway Administration

11. Distribution Statement

12. Key Words
    ITE trip generation rates; automated vehicle counting; built environment

13. Abstract
    The *Trip Generation Manual* published by the Institute of Traffic Engineers (ITE) is widely used to estimate trip generation of individual land uses. However, its trip rates are based on data collected over six decades and predominantly from sites in suburban areas. The result is vehicle trips tend to be overestimated in urban areas where transit is prevalent and pedestrian trips more frequent. In general, it is believed that the "built environment" surrounding a site has a significant impact on trip generation. To test this hypothesis, manual counts of trip generation were conducted at a sample of strip malls with varying surrounding population density, land use diversity, and traffic intensity, and regression analysis conducted to determine the impact of surrounding conditions on trip rates. It was found they reduced the root-mean-square-error of trip rates of strip malls in Louisiana by 36 percent compared to those estimated from the ITE trip equation. Different technologies were tested to automate counting of trip ends at land use sites; the use of image recognition from video recordings was found to be the most successful with 90 percent accuracy.

Project Review Committee

Each research project will have an advisory committee appointed by the LTRC Director. The Project Review Committee is responsible for assisting the LTRC Administrator or Manager in the development of acceptable research problem statements, requests for proposals, review of research proposals, oversight of approved research projects, and implementation of findings.

LTRC appreciates the dedication of the following Project Review Committee Members in guiding this research study to fruition.

LTRC Administrator/Manager
Julius Codjoe, Ph.D., P.E.
Special Studies Research Manager

Members
Jody Colvin
Ryan Hoyt
Cristine Gowland
Jared Chaumont
John Broemelsiek
Nick Ferlito

Directorate Implementation Sponsor
Christopher P. Knotts, P.E.
DOTD Chief Engineer

ITE Trip Generation Modification Factors for Louisiana

By
Chester Wilmot, Ravindra Gudishala, Saba Doulabi, Mishuk Majumder*
Peter Stopher**
Angela Antipova***

*Department of Civil Engineering
Louisiana State University
Baton Rouge, LA 70803

** Private Consultant, Phoenix, Arizona

*** Department of Earth Sciences, University of Memphis

LTRC Project No. 18-4SS
SIO No. DOTLT 1000212

conducted for
Louisiana Department of Transportation and Development
Louisiana Transportation Research Center

April 2021

Abstract

The *Trip Generation Manual* published by the Institute of Traffic Engineers (ITE) is widely used to estimate trip generation of individual land uses. However, its trip rates are based on data collected over six decades and predominantly from sites in suburban areas. The result is vehicle trips tend to be overestimated in urban areas where transit is prevalent and pedestrian trips more frequent. In general, it is believed that the “built environment” surrounding a site has a significant impact on trip generation. To test this hypothesis, manual counts of trip generation were conducted at a sample of strip malls with varying surrounding population density, land use diversity, and traffic intensity, and regression analysis conducted to determine the impact of surrounding conditions on trip rates. It was found they reduced the root-mean-square-error of trip rates of strip malls in Louisiana by 36 percent compared to those estimated from the ITE trip equation. Different technologies were tested to automate counting of trip ends at land use sites; the use of image recognition from video recordings was found to be the most successful with 90 percent accuracy.

Acknowledgments

The assistance of Amir Ghods, Ph.D., of SMATS in the use of the TrafficBox instrument, which detects Wi-Fi and Bluetooth signals, is gratefully acknowledged.

Implementation Statement

The research conducted in this study has led to identifying modification factors for the Institute of Transportation Engineer (ITE) trip rates for strip malls and has also identified an automated means of collecting trip data for individual land uses in general. The DOTD can use the automated data collection procedure to develop modification factors to ITE trip rates for different land uses in different settings in Louisiana. The study has also demonstrated the use of GIS as an interface between the user and the data needed to estimate trip rates for a particular land use in a particular location. While the GIS system only accommodates a single land use at the moment, as more land uses are surveyed and new modification factors identified, the convenience of the GIS interface will be an important feature of the process.

Table of Contents

List of Tables

List of Figures

Introduction

Background

The Institute of Transportation Engineers (ITE) has published trip generation rates of different land uses from studies conducted in the U.S. over the past 50-60 years. Over time, observations from new studies have been added and land uses have been disaggregated into finer land use categories. However, this has generally not resulted in more accurate estimates of trip generation rates at individual locations. The reason is that observed trip rates vary considerably from study to study. An example of this is shown in Figure 1. It is extracted from the ITE *Trip Generation Manual* (8th edition) and shows the results from 40 different sites where the number of trip ends during the peak hour between 4 and 6 pm per 1,000 square feet of supermarket floor area were measured [1]. As shown, observed trip rates at individual sites range from roughly 5 to 20 trip ends per hour per 1000 square feet of floor area. Observed values are shown as crosses in the diagram. Their dispersion around the fitted line (solid line in diagram) illustrates the variation in the data. A measure of the variation around the fitted line is the Coefficient of Determination, R-squared, which, in this case, shows 69 percent of the variation in the number of trip ends is captured by the fitted line.

The inability of the ITE trip generation rates to depict trip rates accurately at individual locations has long been recognized [1], [2], [3]. Even in the user guide of the ITE *Trip Generation Manual*, the suggestion is made that users must take care in using average values and may want to use local characteristics of a site to adjust average trip rates to a specific location [4]. In addition, since the trip rates published in the ITE *Trip Generation Manual* are mainly from data collected at suburban locations, the impact of public transportation, ridesharing, or pedestrian-friendly conditions often found in an urban environment, have largely been ignored [4]. To support this argument, a review of past ITE trip rate errors shown in Table 1 shows that overestimation is more common than underestimation, and errors in urban areas are generally larger than in suburban areas [2]. Mixed land use developments (i.e., where more than two land uses are present on the same site) are becoming increasingly popular among property developers, and they are also one of the land uses for which the DOTD has experienced inflated trip rate estimates in the past. There is a corresponding increase in research related to the trip generation rates of this land use [5], [6], [7], [8].

Figure 1. Observed trip generation at supermarkets

Specialty Retail Center
(814)

Average Vehicle Trip Ends vs: 1000 Sq. Feet Gross Leasable Area
On a: Weekday

Number of Studies: 4
Average 1000 Sq. Feet GLA: 25
Directional Distribution: 50% entering, 50% exiting

Trip Generation per 1000 Sq. Feet Gross Leasable Area

Average Rate	Range of Rates	Standard Deviation
44.32	21.30 - 64.21	15.52

Data Plot and Equation

Caution - Use Carefully - Small Sample Size

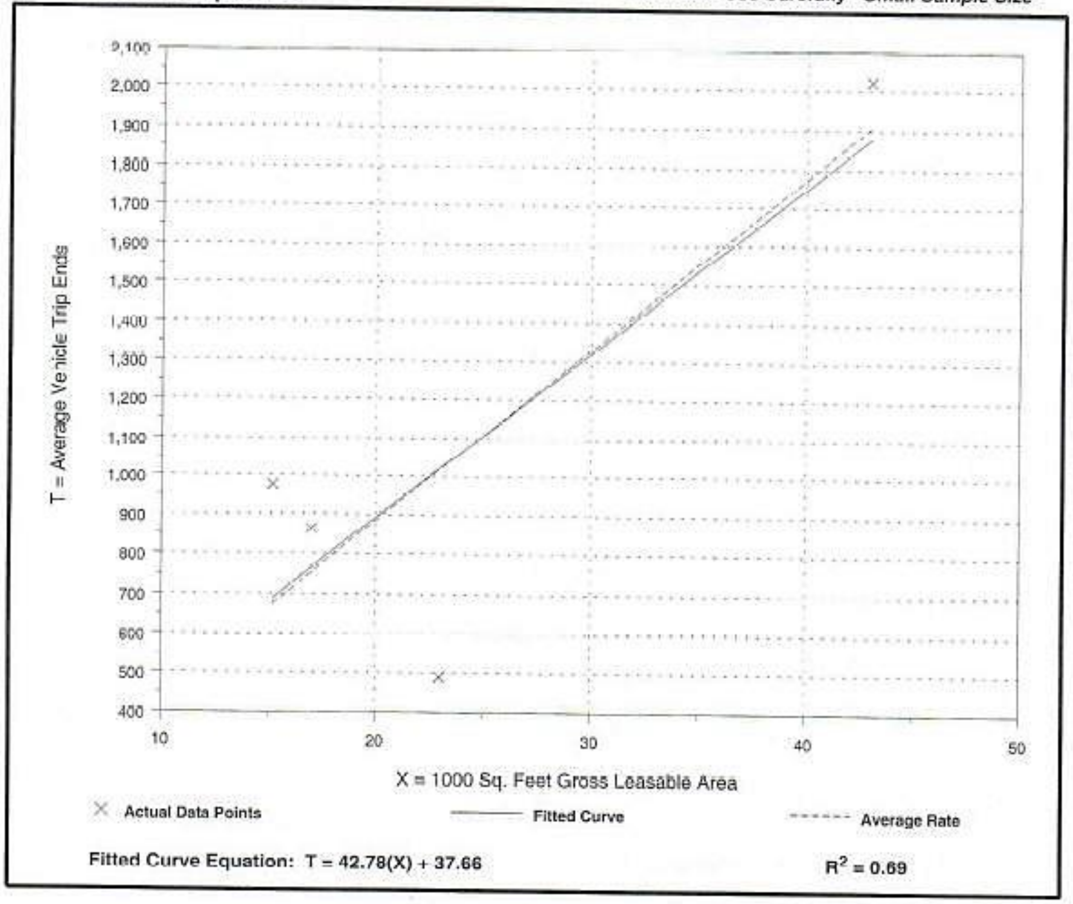


Table 1. ITE trip rate errors

	AM Peak			PM Peak			Automobile Mode Share		
Central Business District/Urban Core/Downtown	**-93%**	to	**1109%**	**-99%**	to	**11 %**	**8**	to	**100 %**
Eating / Restaurant	-93%	to	-57%	-99%	to	-70 %	17	to	57 %
Office	-80%	to	-22%	-62%	to	-21 %	56	to	95 %
Residential	-83%	to	15%	-80%	to	11 %	14	to	85 %
Restaurant		-35%			-26%		34	to	60 %
Retail	-17%	to	1109%*	-22%	to	8 %	8	to	100 %
Services		-14%			-66%				
Shopping		30%			3%				
Mixed-Use Development	**-109%**	to	**181%**	**-170**	to	**61 %**			
Mixed	-109%	to	38%	-80	to	61 %			
Town Center	-108%	to	181%	-170	to	-35 %			
Transit-Oriented Development	**-90%**	to	**20%**	**-92**	to	**35 %**	**50**	to	**96 %**
Office							50	to	96 %
Residential	-90%	to	20%	-92	to	35 %	53	to	93 %
Development near transit	**-58%**	to	**72%**	**-36**	to	**51 %**	**28**	to	**90 %**
Office							28	to	90 %
Residential	-58%	to	72%	-36	to	51 %	33	to	82 %
Suburban Activity Centers and Corridors	**-37%**	to	**-5%**				**54**	to	**98 %**
Office	-37%	to	-20%						
Residential		-5%							
Shopping							54	to	98 %

Literature Review

Factors Affecting ITE Trip Rates

There has been considerable research in the area of amending or adjusting ITE trip generation rates in the past 20 years. Two identifiable thrusts are the inclusion of the built environment (i.e., nature and intensity of surrounding land use and transportation service) in the adjustment of trip rates; and attention to mixed use developments and how they alter trip rates from single land uses as a result of multiple visits among the land uses at a mixed-use development.

One of the main impacts the built environment has on trip rates is that auto trips tend to be overestimated in urban/downtown areas in large cities where non-vehicle modes have a noticeable share of the trips [9]. This is because much of the data on which the ITE trip rates are based were collected in suburban areas where most travel is conducted by automobile [2]. Thus, while the trip generation of a site in terms of person trips may remain the same, vehicle trip generation using ITE rates are typically overestimated in areas such as urban centers, mixed use developments (MXD), and transit-oriented development (TOD) areas. To respond to this limitation, the 10th edition of the ITE *Trip Generation Manual* has provided new urban and person-based trip data for the first time [10].

One study that identified adjustments to the ITE trip rates was the study conducted under the supervision of the Planning Department of the City and County of San Francisco in 2002 [11]. It produced a look-up table of trip rates for different land uses. It was based, among other sources, on data collected from the 15,000 household MTC 2000 Bay Area Travel Survey in nine counties in the San Francisco Bay Area [11]. It recommended adjustments to ITE trip generation rates based on variables such as density and proximity to rail or ferry transit (within ½ mile, between ½ and 1 mile, beyond 1 mile) and urban context (urban area, high-suburban area, low-suburban area and rural area) [12].

The Trip Rate Information Computer System (TRICS) in the U.K. provides a similar trip estimation service as the ITE *Trip Generation Manual* does in the U.S. However, it uses a context-sensitive method to estimate both vehicular and multi-modal trip rates for a variety of land use types [13]. The system is sensitive to the location of the site (edge of town, town center, suburban area, etc.) and surrounding land use (retail zone, residential zone, high street, etc.) [13].

Regarding the study of mixed-use developments, the U.S. Environmental Protection Agency (EPA) has developed a new dataset and methodology in cooperation with ITE called MXD, which specifically addresses trip generation at mixed use developments [14]. The method was developed by studying the trip-generation impacts of mixed-use developments at 239 multiuse sites in six metropolitan areas. The required data were collected from various resources, such as household travel surveys and GIS databases, to create consistent land use and travel measures [11]. They tried to identify the design and other context characteristics of smart growth areas that engender walking, biking, and transit use as well as to identify the number of internally-generated trips. Knowing the geographic, demographic, and land use characteristics of an area surrounding a planned development, the EPA MXD model is able to estimate the share of walk and transit trips as well as daily vehicle miles of travel that will be generated by the development. In 2010, the EPA-SANDAG method was introduced as a result of the EPA MXD method being adapted for use in the San Diego region located in California [15].

TRB's National Cooperative Highway Research Program (NCHRP) launched a study, NCHRP Project 8-51, to estimate internal trips generated at mixed use development (MXD) sites. The method includes a classification system to identify internal trips at MXDs and, subsequently, the reduction in ITE rates based on internal capture levels [16]. Although the NCHRP 8-51 procedure is based on the ITE procedure, the proximity of interacting land uses, plus both A.M. and P.M. peak periods, are incorporated in the procedures. Estimation errors at MXDs reportedly decrease by approximately 50 percent from ITE-based methods when applying this method [16]. The final report of this project has been published as NCHRP Report 684 [16].

Several studies have been conducted on the impact of the built environment on trip rates at Portland State University by Clifton et al. [1]. They investigated the effect of the urban context on trip generation rates across three land uses (convenience stores, high-turnover restaurants, and drinking establishments) by surveying 78 establishments located in a variety of settings such as suburban and city centers. Vehicle trip rates were obtained manually using site visitor surveys as well as exiting and entering person and vehicle counts. The results were compared with the ITE trip rates and urban context adjustment models were produced incorporating built environment factors. They found the adjusted models improved vehicle trip rate estimates for convenience markets and drinking places in comparison with the ITE methodology, while no improvement was achieved for the high turnover restaurant land use category. Currans and Clifton (2015) also investigated the possibility of using Household Travel Surveys to adjust the ITE trip rates [3].

Schneider et.al (2015) developed two linear regression models for morning and afternoon peak hour to adjust ITE vehicle trip estimates across 50 smart growth areas in California [4]. Adjusted R squares of nearly 0.3 and 0.29 resulted for the morning and afternoon peak hour models. The models are appropriate to be used for “single land uses in several common categories, such as office, mid- to high-density residential, restaurant, and coffee/donut shop” and are suitable for planning level analysis.

The built environment is described in several terms. One description is that it includes the following Ds: development, density, diversity, design, destination accessibility, distance to transit, development scale, demand management, and demographics. Demographics, and some of the other factors listed above, are expected to capture the socioeconomic characteristics of individual sites [5]. Table 2 lists the built environment factors employed in different studies across the U.S. as reported in the literature.

Cervero et al (1997) introduced the original three Ds of density, diversity and design as the main categories of built environment (BE) characteristics [6], [7]. As a result of their survey on the influence of these three BE factors on travel behavior, it was revealed that the triune BE measures have a strong relationship with travel behavior by the negative influence they have on household VMT, but a positive impact on non-auto mode choice when non-commute trips are considered [6]. Moreover, the strong relationship of VMT and trip length with destination accessibility has been confirmed [5].

In urban studies, urban density is usually measured by population, employment, building floor area, housing units, and average lot building coverage per unit area. Also, activity density is defined as the aggregation of population and employment variables per areal unit [5] . The activity density affects mode choice considerably [8]. It is shown that an inverse relationship exists between density and vehicle trips [9]. The reason lies in the correlation among built environment measures. For example, an increase in population density also increases the diversity of land uses due to the economic opportunities dense regions offer in satisfying the various needs of the population. Because of the high value of land in dense regions, dense population and traffic congestion, environmental pollution

Table 2. Candidate measures affecting trip generation rates

Measures	Factors	Variables
Built environment	Density	Housing density
		Population and/or employment density
	Diversity	Land-use mix
		Non-residential land use
		Percent single-family housing units
		Retail employment
	Design (vehicle activity)	Total intersection density (per square mile)
		4-way intersection density
		Average block size including median block perimeter (miles), median block area (acres)
		Average street widths
		Number of lanes
		AADT
		Average building setback
		Pedestrian and bicycle facilities
	Destination accessibility	Regional (distance to CBD)
		Local (distance from home to the closest store)
		Job accessibility by auto
		Job accessibility by transit
	Distance to Transit	Transit route density or number of transit corridors
		Distance between transit stops
		Number of stations per unit area
	Demand management	Parking provided per service population
		Percent of site area covered by surface parking
Demographics	Socio-economic controls	Person/neighborhood paired by household size, occupation, income, race, gender, age of neighborhood, dwelling type (housing type), (rapid) transit service, roadway network, topography, regional location, number of children and some others [16]
		Lowest/ highest income household dropped from sample
		Neighborhood auto ownership levels

of automobiles and many other reasons, dense urban regions enjoy the greatest share of quality transit, public transit facilities, pedestrian walkways, and cycling routes.

According to Ewing and Cervero (2001), the aggregate elasticity of density and vehicle trips was calculated at about -0.05, meaning a one percent increment in the density of an area will result in approximately 0.05 percent reduction in vehicle trips [9].

Diversity is defined as the presence of different land uses in a specific area. This factor is usually measured as the percentage of commercial land use to total land area, or the percentage of employment to total population [10], [9]. With increasing diversity in an area, origins and destinations come closer together which results in a decrease in trip length and travel time for both work and shopping trips [11]. The literature also suggests that an increase in diversity increases walking trips, and mixed-use development can cause a reduction in single-occupant commuting [12], [11] . Transit developments are also supported by diverse areas. Consequently, vehicle trip generation studies have observed that diverse areas tend to have reductions in vehicle trip generation [10].

Design addresses street network features varying from a grid network located in dense urban areas to curved streets, loops, and cul-de-sacs in low density suburban areas. To characterize street networks and their impact on travel, the following factors are usually considered; street connectivity, directness of routing, block sizes, sidewalk continuity, and several other factors [9]. These factors have been classified into two main categories: macro-scale and micro-scale features reflecting neighborhood and immediate environmental characteristics. Macro-scale measures represent the street network connectivity through average block size, intersection density, and some other sub-factors, while micro-scale features reflect the walkability of neighborhoods through pedestrian amenities [10].

The impact of transportation network design on travel behavior has been demonstrated in several studies [6], [13], [14]. The network characteristics influence not only travel times by different modes, but also travel decisions. Moreover, the noticeable impact of urban design and transportation infrastructure on neighborhood auto ownership levels and distance driven for neighborhoods has been demonstrated in three case studies in Chicago, Los Angeles, and San Francisco [15]. According to Ewing et al. (2001), the aggregate elasticity of street network density and vehicle trips was calculated at about -0.05 [9].

According to Ewing and Cervero (2001), destination accessibility is defined as the ease of access to a destination at a local or regional scale [9]. Local accessibility is defined as the distance between home and the closest store, while regional accessibility is considered as the distance to the central business district or the number of jobs that can be accessed to them within a given travel time [5]. Based on analysis, trip length is clearly affected by destination accessibility [16].

As Ewing and Cervero (2010) have observed, transit accessibility not only affects mode share by increasing the possibility of using transit, but it also stimulates the non-motorized mode share [5]. Dense and diverse areas near a transit station usually provide better regional accessibility and more local opportunities. These factors provide a suitable context for the idea of trip chaining by walking to nearby stores and then using a transit station. Thus, vehicle trips and VMT would be decreased by being closer to transit [16]. Supporting their conclusion, a 15,000-household travel survey conducted by the Metropolitan Transportation Commission in San Francisco observed the travel behavior of residents living at different distances to rail stops and ferry terminals to be different. They classified the households into three discrete groups based on their distance to the transit stations (within a half-mile, between half and one mile, and beyond one mile) and found that using transit is four times higher for the residents living in a half-mile to the station in comparison with the others living further away. Additionally, within a half mile, walking was twice and cycling was three times as likely as for those living further away. Furthermore, 42 percent of daily trips of residents living and working within a half mile of transit or ferry stations were conducted by non-automobile mode share and also a third of the aforementioned households had no vehicles.

According to Handy (2015), the ITE methodology has caused overestimation of 48 percent for mixed-use sites and 94 percent overestimation for infill sites [17]. Another study found that well-supported transit-oriented developments (TODs) normally have between 30 and 50 percent less vehicle trip-generation than comparable suburban, vehicle-oriented areas [18]. According to Ewing et al. (2017) in a study of five US case studies, TODs create remarkably less demand for parking and driving in comparison with ITE estimates [19]. They conclude that vehicle trip generation rates and peak parking demand in TODs are approximately half of ITE estimates. Surprisingly, the majority of trips were found to be conducted with non-vehicle modes and only one quarter of all trips were conducted using vehicles [19].

Data Collection Procedures

Local trip data required for site-specific trip generation studies are typically collected manually. The most common manual approaches are either on-site observation or video recording, followed by analysis of the video data in the office, with the latter being much more popular due to safety and comfort concerns associated with on-site observation. However, low visibility due to poor lighting, low camera resolution, or adverse weather conditions can make recovering data from video images difficult. There is also the possibility that a vehicle passing through an entrance/exit can obscure other vehicles from view. Another issue is property managers do not usually allow on-site surveys and may even oppose the installation of traffic counters or video cameras in driveways leading to a facility [20]. One solution is to mount the camera in the road reserve.

Automated data collection for land use trip generation has been limited to using a handheld tablet to conduct intercept surveys and pneumatic tubes to count vehicles [8], [4]. Although a wide range of automatic devices are available for vehicle counting in different areas of transportation, they are not well developed for site-specific trip generation studies, their effectiveness has not been widely researched, and most devices are used for detecting rather than counting [21].

One alternative approach has been to use past household travel surveys. Currans and Clifton (2015) used data from the National Household Travel Surveys (NHTS) to collect travel-related information of a sample of households to adjust ITE trip generation rates for eight general land-use categories [3]. By disaggregating NHTS trips into trip ends and knowing the trip purpose of each trip, trips were translated into trips exiting and entering a particular land use by a specific mode, with a known number of occupants, at a stated time. These data, along with the built environment characteristics of the areas surrounding the trip ends, provide sufficient input to develop contextual mode share and vehicle occupancy adjustment models that can be applied directly to ITE trip generation estimates. Similarly, NCHRP Report 758 (2013) provides adjustment factors for ITE trip generation rates for infill developments through the development of mode share and vehicle occupancy factors extracted from NHTS [22].

Although using the NHTS releases analysts from the burdensome task of data collection, it is accompanied by several limitations. First, the HTS data do not include information on the size of the land use attracting trip ends. Given that the magnitude of land use (expressed as square feet of floor area, number of employees, number of rooms, etc.) is the main variable used in estimating ITE trip generation rates, this is a serious shortcoming. Second, trip purpose is not a definitive description of a land use. For example, “shopping” can encompass a wide array of land uses. This results in the method only being able to estimate the trip generation of broad land-use categories whereas the need is to estimate the trip generation of specific land uses.

Wi-Fi and Bluetooth Scanners

Wi-Fi and Bluetooth detectors are being used with increasing frequency in transportation. They are commonly used to measure travel time, vehicle delay, and occupancy [23]. Wi-Fi and Bluetooth data have also been used in demand estimation (i.e., predicting Origin-Destination (OD) matrices), obviating the need to conduct travel surveys and roadside traffic counts [6].

By using triangulation or locating the nearest cell tower, Call Detail Records (CDR) track trajectories of individual devices through space. Activities and travel modes are inferred by the amount of time spent at a location or the speed and interruptions of movement of individual trajectories of devices. Wi-Fi and Bluetooth trajectory data are also used to identify travel patterns and behavior. Travel patterns are usually identified by classification of homogeneous clusters based on the mode of travel, route choice, time of day, trip duration, as well as origins and destinations that can be estimated using Wi-Fi and Bluetooth data [24]. Wi-Fi and Bluetooth trajectory data have also been widely used to count, monitor, and track people to estimate pedestrian density, flow, and even wait time at transit stations [25]. Kalatian and Farooq distinguished different modes of walking, biking, and driving using Wi-Fi data by training decision tree-based and Deep Neural Network algorithms to achieve an 86 percent accuracy of correctly estimating observed behavior [26].

Video Camera

Video camera recording of activity on a facility provides an alternative to live on-site observation. It has the advantage of being able to review events in detail by rewinding the recording, using slow motion, and conducting the count in the comfort and safety of an office versus being out in the field. However, the labor-intensive task of manually counting vehicles still remains, unless a means of automating the counting process can be found. With video imaging, the possibility exists to automate counting by applying computer-based image detection.

The first step in image detection is to identify an object within the entire screen by using, for example, the background subtraction method which separates moving objects from the background of an image by noting pixel change. The next step is tracking detected objects from frame to frame using different trackers such as the Kalman box tracker. Unfortunately, most of these traditional object detection models are slow, cannot classify vehicles, and have low levels of accuracy. However, some neural network object detection models have been developed recently, are quicker and more accurate.

Currently, there are three popular modern object detection models available; Faster R-CNN, SSD, and YOLO (You Only Look Once) [27]. They are all neural-network based models that have been pre-trained to detect a few default objects. They can be operated on OpenCV (an open platform containing a library of computer vision algorithms). Any programming language can be used at the desired interface of the models, but C++ and Python are the most popular.

YOLO is the most effective and popular object detection model in current use. In the beginning of its application, YOLO divides an input image into a 13 x 13 grid of cells. Each cell then predicts a number of bounding boxes in the image, where each bounding box represents a potential object. An algorithm estimates the probability that each bounding box contains a valid object. The acceptance of a bounding box for further processing depends on whether the probability exceeds a threshold value. Threshold value ranges between 0 and 1 and are used to filter out bounding boxes that are unlikely to contain a valid object. For example, a threshold value of 0.25 means that bounding boxes with probability values less than 0.25 are eliminated.

An object detection program performs two activities: object detection and object recognition. Object detection involves determining whether objects are present in an image. The input of an object detector is the whole image, and the output is the class label and the probability a valid object is present. Object recognition involves identifying the type and location of an object in an image. Sub-regions of an image are used to establish the position of an object. The object recognition algorithm looks for the object in the image and identifies the boundaries of the object in a bounding box. The bounding box describes the height, width, and dimensions of a detected object.

Objective

The goal of this study was to obtain more accurate trip generation estimates of land uses in Louisiana than those obtained from the ITE *Trip Generation Manual*. This was achieved by pursuing two specific objectives:

1. Identify what factors, in addition to floor area, influence trip generation at individual land uses in Louisiana.
2. Investigate whether there are viable alternative ways of collecting trip generation data at land-use sites in place of the current method of manual counting.

Scope

The scope of this project is limited to determining the trip rate adjustment factors to the ITE trip generation rates for strip malls in the parishes of Lafayette, West Baton Rouge, East Baton Rouge, Livingston, Ascension, and Tangipahoa in Louisiana. The adjustment factors depend on the built environment characteristics of the area surrounding a strip mall, and these characteristics are captured from census data built into a GIS that is used to estimate the adjustment factors for any potential site in the area. The study also includes identifying, testing, and evaluating alternative means of automated trip generation detection with a view to measuring the trip generation rates of other land uses cost-effectively.

Methodology

Approach

Past studies have shown that the built environment, demographics, and transportation service all affect trip generation rates [3], [5]. The approach adopted in this research was to select variables available in secondary databases that reflect these features, incorporate these variables in a GIS, and then use the GIS as the medium of enquiry to identify adjustment factors to ITE trip generation rates for a particular site. This has been accomplished in this study for strip malls but could be repeated for any other land use provided trip rate observations are made. Since manual counting trip rates at individual sites is labor intensive, this study also investigated automated means of observing trip counts to individual locations. The method by which each of these objectives was accomplished is described below.

Identifying Trip Rate Adjustment Factors

One of the first tasks was to identify a particular land use to demonstrate the process of establishing adjustment factors to the ITE trip generation rates. After consultation with the Project Review Committee, strip malls were identified as a land use where trips appeared to be overestimated when using ITE trip generation rates in Louisiana. Strip malls were subsequently chosen for study in this project. Strip malls appear as Land Use 814, Specialty Retail Center, in the ITE *Trip Generation Manual*. The ITE trip generation rates for this land use are based on only 4 past studies. The average trip ends per day on a weekday are estimated by the linear equation: trips/day = 37.66 + 42.78(floor area), with floor areas being the total floor area of the mall measured in 1000s of square feet. A typical strip mall is shown in Figure 2. They are often located adjacent to arterial roads.

From the literature review, the factors in Table 2 were identified as contributing to the deviation of trip rates at individual sites from the values estimated in the ITE *Trip Generation Manual*. The three Ds of density, diversity, and design are considered the main contributors to trip rate deviation from ITE values [6]. To keep the process manageable by not including too many factors, and to ensure that the chosen factors could be meaningfully represented by variables commonly available in secondary data sets, the three factors of density, diversity, and design were chosen to demonstrate the process of establishing adjustment factors to the ITE trip generation rates in this study.

Figure 2. Typical strip mall

Identifying Secondary Data Sources

Considering the factors of density, diversity, and design, as built environment characteristics that affect trip generation, it was necessary to identify sources that can provide any of the variables in Table 2 that represent these factors. Variables in secondary data bases were sought that measure, or are closely associated with, the impact that the chosen factors of density, diversity, and design have on trip rates for that land use. For example, residential density of the catchment area surrounding the site could be used to measure the impact of density, while diversity could be measured by the amount of retail employment immediately surrounding the site. ADT of the adjacent street could be used to measure the impact of design. Different variables could be used depending on the strength of their association with the factor in question, their availability, and the land use being investigated.

Potential sources of data for use in this task are the U.S. Census and other federal and state agencies. Data sources used in this study are shown in Table 3. The data were used to obtain variables describing density, diversity, and design. Density can be measured either in terms of residential density (population per square mile) or residential and employment density combined. Land use diversity was measured by jobs to workers ratio (i.e., number of jobs to the resident workers of an area). Design was first measured by the road characteristics of the adjoining street with an entrance to the strip mall. Average Daily Trips (ADT), number of through lanes, urban/rural designation and shoulder availability were considered. However, road density within a half-mile radius of the strip was ultimately used.

Table 3. Secondary data sources used in study

Database Name	Database Address	Data Features
TIGER/Line® - Geography - U.S. Census Bureau	https://www.census.gov/geo/maps-data/data/tiger-line.html	Geographical data at different geographical levels
American FactFinder - Census Bureau	https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml	American Community Survey (ACS) 2015 (5-year estimate) is used for demographic information such as total population and median income. Urban/rural designation of roads
OnTheMap - U.S. Census Bureau	https://onthemap.ces.census.gov/	2015 Number of jobs and resident workers data for jobs to workers calculation
Louisiana Department of Transportation and Development	http://wwwapps.dotd.la.gov/engineering/tatv/	2016 ADT
Highway Performance Monitoring System (HPMS)	https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm	HPMS 2016 and 2017 ADT estimates
Google Maps	https://www.google.com/maps	Number of through lanes and shoulder availability

Preparing the GIS Database

After establishing the sample frame including the name and address of each strip mall in the study area, all the sites were geocoded into an ArcGIS program. Tiger-line shapefiles, demographic data, and job and traffic data (from the aforementioned databases) were loaded into ArcGIS software. The BE characteristics of all potential strip malls were calculated in a half-mile buffer zone around each site. A half-mile radius buffer zone was chosen as the area impacting a business as this is a common assumption in urban geography [4], [28], [29].

Demographic data were extracted from the 2015 ACS (American Community Survey) 5-year estimates at Block Group level across the study area. Using the common Geo-ID field of Tiger files and demographic data, demographic tables were joined to the geographic features in ArcMap. By having the total population of half-mile buffer zones and the area of the buffers, population density was calculated for all strip malls. Median income was similarly calculated in a half-mile buffer zone around strip malls.

In order to calculate land-use diversity around strip malls, the OnTheMap website was used. Although it was possible to use census data directly, the OnTheMap website provides an easy and user-friendly environment to download the number of resident workers (Home) as well as the number of available jobs (Work) in any area in the country. It is also possible to define different buffer zones around any location as well as selecting census data year, job types (all jobs, primary jobs, private jobs), labor market segment (by NAICS industry sector, race, ethnicity, educational attainment, sex, worker age, and monthly earning). All jobs, as well as the number of resident workers, were calculated in half-mile buffer zones around strip malls. The data were extracted in excel spreadsheets separately and were added to ArcMap. Then they were joined to the previous shapefile containing population and income data. By having the total jobs and total resident workers in the half-mile buffer zone around each strip mall, land-use diversity was expressed as the jobs to worker ratio (JWR). A value of JWR greater than 1 indicates an area where commercial, office, and industrial land uses are responsible for more jobs than those produced by local housing. In contrast, an area with low commercial, office, and industrial activity relative to residential land use will have a low JWR and a low diversity as it will be predominantly housing. Thus, the higher the JWR, the greater the mix of land uses and, therefore, the more diverse the land uses.

The traffic volume of the main road beside strip malls and road density in half-mile buffer zones around strip malls were two candidate variables to measure the impact of design of the street network on vehicle-trip generation rates of strip malls. Annual Average Daily Traffic (AADT) from different sources and various aspects of road density were investigated in this study. If insufficient data is available for the counts to be annual averages, the distinction is sometimes made by referring to counts of less than a year as Average Daily Traffic (ADT). Although DOTD provides ADT estimates for the entire state, such data were not available for all roads adjacent to the strip malls in the study. Therefore, the unknown ADT values of strip malls in the sample frame were imputed by establishing homogeneous road classes in the sample frame as well as the DOTD ADT database. The number of through lanes, shoulder availability, and the urban/rural designation of roads were the criteria to establish homogeneous groups. For all strip malls, these three variables were manually provided. Google Earth was used to identify the number of through lanes and shoulder availability, while the pre-calculated population density of half-mile buffer zones around strip malls allowed the identification of the urban/rural designation of the roads. Buffers with a population density of 1,000 people per square mile or less were considered rural roads, while those with more than 1,000 people per square mile were considered as urban roads. Likewise, all ADT points of the study area were characterized in terms of number of through lanes, shoulder availability, and urban/rural designation manually. In the next step, ADT values for the main roads of strip malls with unknown ADT values were imputed by obtaining the average value of all ADT points with similar characteristics (number of through lanes, shoulder availability and urban/rural designation) located in the same parish as the subject roads.

The Highway Performance Monitoring System (HPMS) also provides ADT estimates as a part of Vehicle Miles Traveled (VMT) estimates. By having permanent stations of traffic counts as well as temporary stations, the state's comprehensive traffic count program provides ADT estimates on almost all the links of the network. The latest versions of the HPMS ADT estimates, that is 2016 and 2017 estimates, were downloaded from the HPMS Geospatial Database and were added to the ArcMap workspace. Although very few links had no HPMS ADT estimates, for those cases with missing data, imputed ADT values from the above approach were assigned to such roads.

Road densities in half-mile buffer zones around strip malls were calculated using two road network databases, Tiger Lines and the DOTD transportation network database. Although it is known that the DOTD database provides more accurate and comprehensive network information compared to the Tiger Lines, calculating one measure (network density) from these two databases allows us to compare them and have a tangible understanding of their differences. To calculate road density, network shapefiles were intersected with the buffer layer of strip malls in ArcMap. After that, the sum of the length of roads inside the buffers were calculated and divided by the area of the buffers. Thus, network density around each strip mall was calculated. Due to the functional class availability for all links in the network in the DOTD database, it was also possible to evaluate the impact of road density of each road class within the buffer area. Hence, road densities of different functional classes were calculated for the sample frame separately. Eight different road functional classes can be found in this database: controlled access highway (code 1), secondary highway or major connecting road (code 2), local connecting road (code 3), local road (code 4), ramp (code 5), 4WD (code 6), ferry route (code 7), and tunnel (code 8).

Survey Site Selection

Values of variables describing the density, diversity, and design within a half-mile radius of each strip mall in the sample frame were calculated using the GIS system described above. The midpoint on each variable was determined and used to divide the sites into “high” and “low” categories on each variable. The sample frame was then stratified into eight (2^3) categories based on the high and low values of the variable values describing the density, diversity, and design of each site. Following this, 5 sites were randomly sampled from each category in the sample frame. Table 4 shows the number of selected strip malls in each parish by category. The first, second, and third digits of each category refers to the high (code 1) or low (code 2) values of residential density, JWR ratio, and the DOTD 2016 ADT counts on the adjoining road. Strip malls with more than 1,000 people per sq. mile residential density were considered as areas of high population density, while areas with less than 1,000 people per sq. mile were considered to have low population density. Buffer zones with a JWR value greater than 1.25 were considered areas of high land use diversity, and areas with lower values to be areas of low land use diversity. This value was used to ensure first, enough jobs for resident workers, and second, a job surplus of at least 25 percent, so that non-resident workers are entering the zone for the purpose of employment. The average DOTD 2016 ADT value of 19,412 vehicles per day among the sites in the sample frame, divided the strip malls into high and low traffic design categories if they had more or less ADT values than the reference value. Figure 3 shows the randomly selected strip malls in the Baton Rouge area by category. Figure 4 shows the spatial distribution of the 40 randomly selected strip malls across the entire study area. Figure 4 shows the study area as well as the strip malls surveyed in this study. Dots are used to show strip malls on the map.

Table 4. Randomly selected survey sites

Category	Parish						Total
	EBR	Lafayette	WBR	Ascension	Livingston	Tangipahoa	
111	2	2	0	0	0	1	5
212	2	0	1	0	1	1	5
211	2	0	1	1	1	0	5
222	1	0	0	2	1	1	5
121	3	1	1	0	0	0	5
221	1	0	1	2	1	0	5
122	2	1	0	2	0	0	5
112	1	0	0	2	0	2	5
Total	14	4	4	9	4	5	40

Figure 3. Strip malls in East Baton Rouge Parish by category

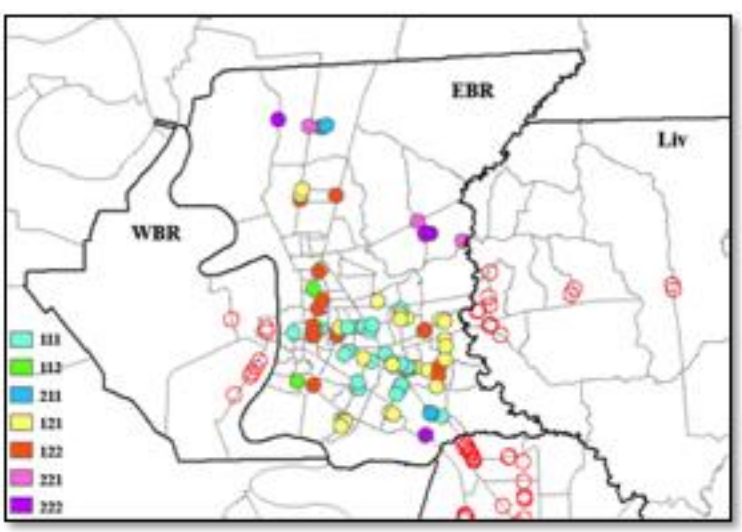
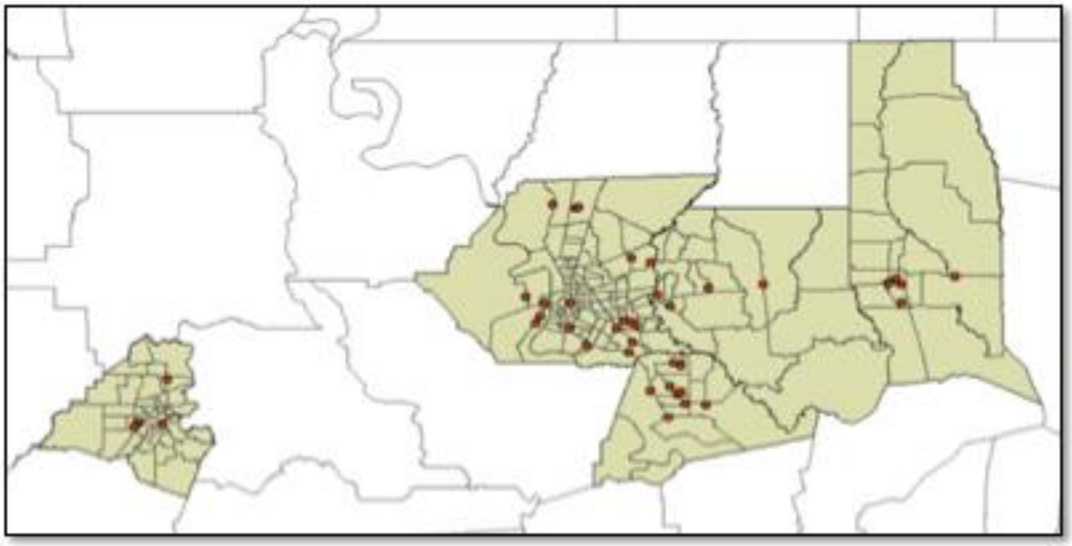


Figure 4. All selected sites by location



Conducting Manual Traffic Counts

Other studies that have investigated the difference between ITE trip generation rates and those observed at individual sites have typically relied on fieldworkers to count vehicle trip ends manually by time period at the individual locations. This is because it is often difficult to isolate trips to specific land uses unless it is a stand-alone land use, which has its own access and egress roads on which traffic counts can be conducted. In this study, the PRC was consulted regarding the time period for which ITE trip rates are most often used in assessing a site development application. They indicated they most often require trip rates during the morning and afternoon periods of peak hour traffic on the adjacent street. Since peak hour travel on adjacent streets can vary by location, survey times were adjusted to include both morning and evening peaks by conducting counts from 8 am on the first day, to 6 pm on the second day.

In conducting manual counting of vehicle trip ends, a fieldworker can either observe vehicles at the site or record vehicle movement on a video camera, or cameras, and then analyze the video footage in the office later. The latter procedure was adopted in this study for convenience and safety.

Office Preparation

In order for the fieldwork of a survey to be conducted efficiently, planning must be conducted in advance. This involves preparing a checklist of instruments, scheduling the fieldwork for each site, and assigning tasks among the fieldworkers. A checklist defines the list of instruments required to conduct the survey. A typical checklist is shown in Table 5.

Table 5. Checklist for fieldwork

Instrument Name	Number required (no's)	Check	
		yes	no
Camera			
Fully charged battery			
Traffic boxes			
Plastic pole			
Steel pole			
Steel angle			
Metal straps			
Wheel measurer			
Hammer			
Drill			
Safety vest			

The schedule of fieldwork defines a time plan for sites, which includes the survey day and departure time of a survey team from the office for each site. For a typical survey day, the number of sites for surveying was selected based on the required number of instruments and the distance of sites from the office. A single site was selected for a survey day when the site required the installation of all the available instruments, because the site had numerous entrances and exits. Similarly, a group of sites was selected for a survey day when the available instruments were sufficient to cover all entrances and exits. Selecting a group of sites for a survey day also took into account the distance between sites as well as the distance from the sites to the office, since all instruments had to be installed before 8 am. The departing time from the office was selected based on the distances of sites from the office and the number of sites served.

Work distribution is the assignment of survey work among the team members. Assignment of duties to individual workers in advance ensures smooth and efficient execution of the fieldwork. In this project, before each survey day, work was distributed among the team members. A team member was assigned to charging camera batteries and checking the availability of instruments before the survey day. On the survey day, the same person was responsible for ensuring all required instruments were loaded on the vehicle. Fieldworkers were assigned different duties at the site. For example, some members were responsible for finding suitable locations for camera installation, and some were responsible for installing and retrieving the cameras. After the survey, an individual was responsible for downloading video data from the cameras and uploading them to the server.

Execution of Fieldwork

The fieldwork was conducted according to the survey plan prepared at the office. On a survey day, the first task of the survey team was to load the instruments in the vehicle according to the checklist. It was found that it takes about 30 minutes to load the instruments in the vehicle. When the team reached a survey site, the first task was to find suitable locations for installation of the cameras. In selecting the actual location of a camera, the team preferred an existing pole, i.e., electricity or telephone pole, because the existing pole gives better support to the camera. When there was no existing pole, steel angles were used to support a 2-in. diameter camera pole. Two steel angles were driven into the ground on either side of the camera pole and then secured with clamps. The average mounting height of the cameras was 10 ft. because it was found this height generally prevents a vehicle in a closer lane obscuring the view of a vehicle in the next lane.

A few factors were checked during the camera installation. First, the charge level of the camera batteries was checked. Second, the clarity of the camera lenses was checked to ensure a clear video recording. Third, a real-time clock (for example, a smartphone clock) was shown in front of each video camera so that it could record the time. The purpose of this task was to find the difference between the camera clock and the real-time clock. This time difference was adjusted when manual and automated counting were conducted. Last, the installation angle of the camera was checked so that the camera covered a full view of the entrance or exit. An attempt was made to avoid including a view of adjacent roads as much as possible because vehicle movement on those roads confuse the manual and automated counting process. A typical view obtained by a correctly mounted camera is shown in Figure 5.

Figure 5. Typical view of a camera screen and clock

Retrieval of cameras was conducted after 6 pm on the day following installation. While retrieving the cameras, it was checked whether the cameras had successfully saved all the video data. A check was conducted to ensure that all the instruments were retrieved and loaded in the vehicle.

When the survey team reached the office, video data from the retrieved cameras were downloaded on the computer and then uploaded to the server. Batteries of all cameras were left to charge overnight. Last, a check was performed for the availability of instruments for the next survey day.

Data Storage

Accessibility, capacity of storage, and safety were considered for selecting data storage. In this study, a considerable volume of video data was collected from the survey, and multiple individuals were involved in manual counting. So, it was considered necessary that multiple individuals have access to the server simultaneously. At the same time, it was necessary to store the data safely. Thus, the server of the Intelligent Transportation System (ITS) at Louisiana Transportation Research Center (LTRC) was used for storing the data, which is secured, has a large capacity, and is accessible by multiple people at the same time. However, after downloading the video data from the cameras to the server, a copy of the data was also stored on an external hard drive for added security.

Counting Vehicles

A few rules were established regarding the manual count. First, it was decided to count vehicles in 5-minute intervals because it allows counts to be aggregated in any multiple of 5 minutes. Second, arrivals and departures were defined as the front of a vehicle passing a reference line on the access road. The numbers of arriving and departing vehicles were counted separately for each entrance or exit. Vehicles were classified into six classes: car, motorcycle, cycle, pedestrian, transit, and others. A spreadsheet template, as shown in Table 6, was used to record all manual counts.

Table 6. Sample manual counting sheet

Site name													
Date of survey:													
Camera details:													
No of Entrances:													
Time Start (hr:min:sec)	**Time End (hr:min:sec)**	**Entry details**						**Exit details**					
		Counts in every 5-min. interval						**Counts in every 5-min. interval**					
		Car	**Motorcycle**	**Cycle**	**Pedestrian**	**Transit**	**Others**	**Car**	**Motorcycle**	**Cycle**	**Pedestrian**	**Transit**	**Others**
8:00:00	8:04:59												
8:05:00	8:09:59												
8:10:00	8:14:59												
8:15:00	8:19:59												
--	--												
17:55:00	17:59:59												

Estimation of Error in Manual Counting

Different types of error may occur while conducting manual counts. These are classified as: total count error, classification error, and interval error. In manual counting, ground truth counts are not known, so it is difficult to estimate error. However, if repeated counting produces the same result, that value is considered the true number in this study. Thus, the error is considered to be the difference between the first count and the repeated counts. In this research, a few sites were randomly selected to conduct repeated counts to estimate error in manual counting.

Individual observations of count error are the difference between the repeated count and the first count in a time interval at a site. A measure of the average error in a time interval can be expressed by the root mean square error (RMSE) statistic. It will produce the average error experienced in the time interval in which the counts are reported (e.g., 15 minutes, 1 hour). However, because the count numbers are sometimes unknown it is sometimes difficult to interpret the results where error is reported in the number of vehicles. It is generally easier to interpret the results when they are expressed as a percentage (i.e., a relative measure rather than the absolute measure of RMSE), so the use of percent root mean square error (%RMSE) is recommended. Percent root mean square error (%RMSE) of the total count can be estimated from the following formula:

$$\%RMSE = \sqrt{\frac{\sum_{i,k}^{I,K}\left(\frac{N_{a,k,i}-N_{c,k,i}}{N_{a,k,i}}\right)^2}{I \times K}} \times 100 \qquad (1)$$

where,

i = a time interval.

I = total number of time intervals.

k = a site.

K = total number of sites.

$N_{a,k,i}$ = actual count of vehicles at site k in time interval i.

$N_{c,k,i}$ = counted vehicles at site k in time interval i.

Classification error: The classification error defines the difference between the actual classified counts and the counted vehicles for a particular vehicle class. A classification error occurs due to the placement of a count in a different vehicle class. Classification error will increase with an increase in vehicle classes because more vehicle classes result in lower counts in each class. The following formula can be used to estimate the percent RMSE of classification counts.

$$\%RMSE = \sqrt{\frac{\sum_{i,k}^{I,K}\left(\left(\frac{\sum_{v}^{V}\left(N_{a,k,i,v}-N_{c,k,i,v}\right)}{N_{a,k,i}}\right)\right)^{2}}{I \times K}} \times 100 \tag{2}$$

where,
i = a time interval.
I = total number of time intervals.
k = a site.
K = total number of sites.
v = a vehicle class;
V = total vehicle classes;
$N_{a,k,i,v}$ = actual count of vehicles at site k in time interval i for vehicle class v;
$N_{c,k,i,v}$ = counted vehicles at site k in time interval i for vehicle class v; and
$N_{a,k,i}$ = actual count of vehicles at site k in time interval i.

Interval error: Interval error occurs due to placement of a count in a different time interval than the one to which it belongs. Interval errors result in double counting because an error in one interval requires that there be an error in another. Equation (3) can be used to calculate the percent RMSE of interval counts.

$$\%RMSE = \sqrt{\frac{\sum_{i,k}^{I,K}\left(\frac{n}{N_{a,k,i}}\right)^{2}}{I \times K}} \times 100 \tag{3}$$

where,
i = a time interval.
I = total number of time intervals.
k = a site.
K = total number of sites.
n = the number of misreported vehicles at site k at time interval i which actually belong to time interval (i+1); and
$N_{a,k,i}$ = actual count of vehicles at site k in time interval i.

Automated Counting

Three alternative devices were tested to automate trip-end counting in this study: an infrared sensor, a video camera, and an instrument capable of detecting Bluetooth and Wi-Fi signals in its proximity.

The Pyro-Box Eco-counter

The infrared sensor, marketed as the Pyro-Box Eco-counter, uses changes in infrared radiation to detect passing objects. It is typically used to count pedestrians and cyclists with a human body providing a change in infrared radiation in the field of detection. Two models are available, one with a range of detection of 4 meters and one with a range of detection of 15 meters. Detection occurs in a directed beam. In tests on campus at LSU, the sensor was found to produce pedestrian counts with 91 percent accuracy amidst counts ranging from 10 to 53 pedestrians per 15-minute period. Because the walkway was 10 ft. wide at the point of observation, some pedestrians were hidden from view if they were walking perfectly abreast of someone closer to the device at the time of detection (this is called occlusion). Ideally, the sensor should be used where single-file movement occurs and a fixed object is on the far side of the path to block detection beyond the path. The test site on the LSU campus is shown in Figure 6. The figure shows a walking path with two light posts near to the path on LSU campus. The Pyro-Box is mounted at about 50 cm. height at the light post on the right side of the figure and is faced toward the light post on the left side of the figure so as to cover the walking path.

Figure 6. Pyro-Box test site

The Pyro-Box was considered when the observation of visitors to a strip mall were considered in contrast to vehicles. The thought was that parties of visitors could be identified and associated with a vehicle. However, it was found that it is difficult to observe pedestrians amidst cars and other obstructions, and that stationary pedestrians—as would occur if they stopped to chat or look in a shop window—were difficult to count accurately. In addition, the Pyro-Box operates best when there is a wall or some other fixed object the beam can terminate at. For this reason, the Pyro-Box was not investigated further as a means of automated counting of visitors to a strip mall.

Video Camera

Selection of Cameras

The cameras were selected based on the resolution, battery life, and weather factors. The resolution of a camera is an important factor that controls the success of automated and manual counting. One of the objectives of this research was to use pre-recorded videos for automated counting. Here, the success of automated counting depends on the quality of the video, as determined by the resolution of the camera. Good quality video increases the accuracy of manual counts because it enables individuals to recognize and report counts confidently. So, in this study, the minimum resolution of the cameras was considered as 480 pixels (frame size 480 x 640 pixels), which ensured a good quality video. Battery life was selected based on the duration of the survey. Since this project required to record videos continuously for two days, the minimum battery life of the cameras was selected to be a minimum of 40 hours. The weather factors were also considered for the selection of the camera. It was ensured that the cameras were able to operate in bad weather, for example, rain and fog.

Three types of cameras were selected in this research primarily for their properties but also due to their availability from other prior projects conducted by LTRC. The configurations of these cameras are shown in Table 7.

Table 7. Configuration of the selected cameras

Camera Name	Scout Video collection	Counting Camera	CountCam2 Traffic Recorder
Manufacturer	Miovision	CountCam	CountingCars
Resolution (pixels)	640	480 x 640	480 x 640
Storage (GB)	64 GB and extendable	Extendable	64 GB SDXC internal storage
Duration of recording (hrs)	64	Adjustable	50
Video format	.mp4	.mp4	.mp4
Display (inch)	5.5	6.5	Connectable to smart phone
Battery life (hrs)	72	48	50
Waterproof	yes	yes	yes
Operation temperature (^{0}F)	-40 to 140		Withstands summer heat and winter cold
Installation time (minutes)	5	5	5

Develop Algorithms

To detect objects in the video, it was necessary to select a deep-learning object detector, based on neural network artificial intelligence. After considering a number of candidate softwares the YOLO version 3 was selected for vehicle counting in this project. It was considered to be fast, accurate, and easier to work with than other versions of YOLO and also superior to other possible software. YOLO v3 was applied with the TensorFlow and Open CV libraries for vehicle detection and counting. Algorithms to apply YOLO were also required, which were programmed in Python.

The required packages for this conversion were TensorFlow 18.0, NumPy, OpenCV Python, and TQDM. TensorFlow is a deep learning library used for different applications, such as neural network applications. NumPy is a package of routines in Python which support many mathematical functions on multidimensional arrays. OpenCV (open-source computer vision library) is an open-source library used in computer vision applications. TQDM is a progress bar library which provides useful routines for nested loops.

Conversion of the YOLO Weight File to TensorFlow API

In this section, the YOLO weight file was converted to Tensorflow API using a few packages. The goal of the research was to use the Python programming language for the convenience of coding and implementation. Since the algorithms of the YOLO weight file are written in C/C++ programming language, it cannot be directly implemented using the Python programming language. One of the ways to implement it using python is applying Tensorflow architecture. In that case, the YOLO weight file has to be converted in Tensorflow architecture. So, in this research, the YOLO weight file was converted to Tensorflow architecture to implement it using Python programming language. The required packages for this conversion were TensorFlow 18.0, Numpy, OpenCV Python, and TQDM. First, the YOLO weight file was downloaded from the official YOLO website (pjreddie.com), which is opensource. Then, the path of the downloaded YOLO weight file was placed in the command prompt (a command-line interpreter in windows operating system) to convert the YOLO weight file to the TensorFlow format. This converted weight file was later used for vehicle detection**.**

Processing the Data

All video files need to be processed before uploading them to the program. First, a check is made to ensure that all the video files are in mp4 format; otherwise, they are converted to mp4 format. Second, if there are multiple video files, they are joined to make a single video file because the program cannot process multiple files at a time. Third, the unnecessary portions of the video files, i.e., the portions before 8 am and after 6 pm are trimmed. This pre-processing of video files can be conducted using any video editing software. In this research, Avidemux (a free video editing software) was used for processing video files.

Detection of Vehicles

In this section, the program draws reference lines, detects vehicles, and tracks the detected vehicles frame by frame. OpenCV captures and displays the first frame of the video file. A typical first frame of this study is shown in Figure 7. Note that with the camera in a fixed position, all frames of the video will have the same appearance, apart from objects that move through the frames.

Figure 7. The first frame of video

Next, the program allows the user to select two points on the displayed first frame, which are the starting and ending point of a reference line. These two points are drawn by the first and second left click of the mouse. The reference line drawn in this section is called the “midline.” A typical reference midline is shown as the yellow line in Figure 8. Two additional lines are then defined as the “right line” and the “left line.” These are used to define the direction of motion of an object that crosses the midline. These two lines are generated by the software parallel to the midline. They are the white lines shown in Figure 8.

Figure 8. Reference lines drawn in the program interface

Each video frame is converted to a common size of 128 x 128 pixels for processing. When YOLO detects an object, it draws a rectangular box around the detected object. The outputs of a processed image are rectangular bounding boxes and the score of those boxes. The score means a confidence level for a detected object, i.e., how confident YOLO is that the box contains a valid object. This value lies in the range of 0 to 1. The higher the score value, the higher the confidence that the object is indeed an object of interest.

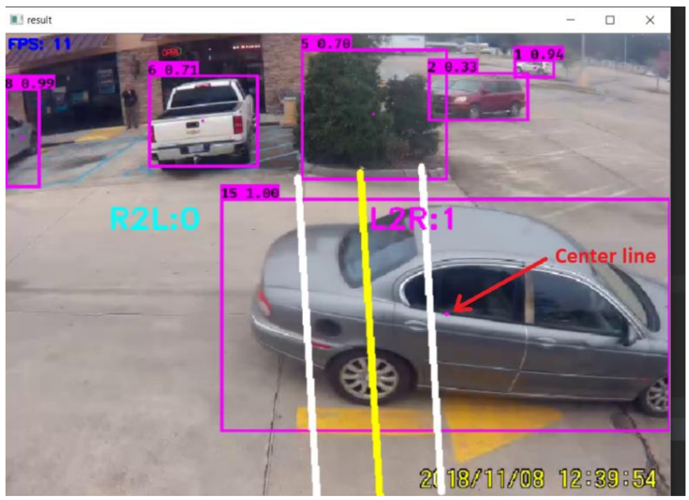
The bounding boxes are processed using a controlling factor that is a threshold value (a factor to screen out bounding boxes with low confidence levels). In this research, a threshold value of 0.2 was used, which means that bounding boxes having a score value higher than 0.2 are accepted for further processing. Those that have values below 0.2 are eliminated.

Tracking Vehicles

In this step, the sorted bounding boxes are tracked frame by frame. Since a video file consists of thousands of frames, an object has to be tracked from one frame to another to determine its direction of movement. In this study, KalmanBoxTracker was used for the tracking process. The tracker compares the current frame with the immediate previous frame using the pixel variance of the frame. When it finds a similarity in pixel values, it updates the object (i.e., bounding box) and memorizes it for consideration in the next frame. Then, it compares the updated frame to the next frame. The tracker titles each tracked bounding box by a numeric value such as 1, 2, 3, etc. In this way, the tracker tracks an object from frame to frame.

Rectangles are drawn around each tracked object on the computer screen as shown in Figure 9. Another purpose of drawing rectangles is to establish a small line in the center of each rectangle, which is used to count vehicles as described in greater detail below. A typical center line is shown in Figure 9 as marked by the red color arrow. The tracked rectangles are observed for vehicle counting.

Figure 9. Small line in the center of a rectangle



The program provides counts in spreadsheet, which contains the number of arriving and departing vehicles in a particular time interval. The counting speed of YOLO is different from the real-time clock. But this research requires counting vehicles in a time interval in real-time. So, at first, the program evaluates the number of frames in the provided time interval, for example, a 5-minute time interval. After that, the program calculates the time required to process a single frame. Then, it evaluates the time required to process all the frames in that time interval and considers the time as real-time interval.

Algorithm for Counting Vehicles

When a vehicle passes through an access road to a facility, the rectangles and the small line at the center of the rectangle (i.e., the center-line) passes the mid, left, and right lines. Depending on the sequence of the intersection of the "center-line" with the "midline" and the left and right lines, the direction of the vehicle is recorded. For example, when the centerline at the center of the rectangle intersects the left line first (i.e., the parallel line, drawn on the left side of the midline), the program deduces the vehicle has come from the left side. Afterward, when it intersects the midline, the program confirms that the vehicle has come from the left side and considers it as a count. The function counts the vehicles which come from the left side and right side, respectively, and are used to distinguish arrivals and departures.

Bluetooth and Wi-Fi Signal Detection Devices

Background

The rationale behind using a device that detects Bluetooth and Wi-Fi signals as a means of counting vehicle trips to a land use evolved from an earlier study conducted for LTRC involving the use of Bluetooth signals to estimate travel time [23]. It was found in that study that there was a close correlation between the number of devices sending out Bluetooth signals at a location and the volume of traffic on the road at that point. This seemed a logical result, and the question then arose whether the same device could be used to estimate visits to a particular site. If a close correlation could be found between Bluetooth and Wi-Fi signals and traffic to a land use, then deploying such devices to individual land use sites would provide automated trip data that heretofore required manual counting.

Selection of Devices

Several products were reviewed before deciding on a product called TrafficBox from the SMATS Company (https://www.smatstraffic.com). It is capable of detecting Bluetooth and Wi-Fi signals and recording the Media Access Control (MAC) address of each device whose signal reaches it. The TrafficBox records Wi-Fi and Bluetooth MAC addresses of Wi-Fi and Bluetooth signal sender devices located in its detection zone. The TrafficBox data can be exported in excel format (csv file) and include the fingerprint (MAC Address records), the detection type (Wi-Fi or Bluetooth), detection signal strength (RSSI value) and detection date and time.

The TrafficBox offers four different scanners, Bluetooth Classic Discovery Mode, Bluetooth Classic Paired Mode, Bluetooth Low-Energy (LE) Discovery Mode, and Wi-Fi signals. The company suggests different scanner combinations with the highest compatibility; Wi-Fi & Bluetooth Paired Mode, Bluetooth Paired Mode-Bluetooth

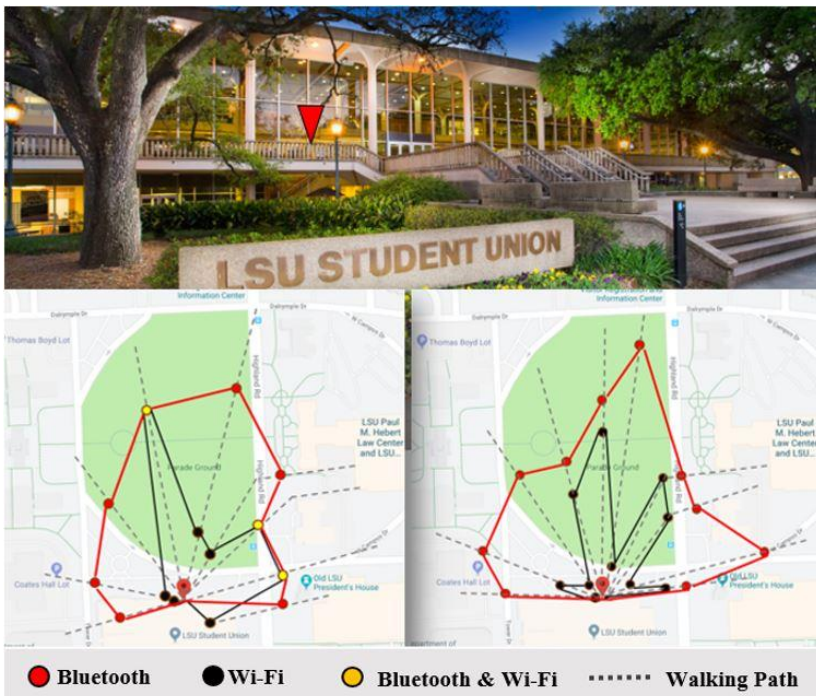
Classic, and Bluetooth Paired Mode-Bluetooth Low Energy. Table 8 shows some examples of the devices can be detected by these scanners. In this experiment, the combination of Wi-Fi and paired Bluetooth scanners was used because of the high compatibility of a Wi-Fi scanner with the Bluetooth scanner that detects Bluetooth signals from paired devices. However, two other Bluetooth modes – Bluetooth Low Energy (BTL) and Bluetooth Classic (BLC) Scanners – are available that are better when used in Bluetooth-only experiments.

Table 8. TrafficBox scanner

Bluetooth Classic Discovery Mode	**Bluetooth Low Energy Discovery Mode**	**Bluetooth Classic Paired Mode**	**Wi-Fi**
Car GPS units. Audio systems (if they are not paired with the driver's headset or any other devices) Very unlikely to detect cell phones as Bluetooth becomes discoverable only when the Bluetooth setting menu is open	Fitness gadgets Smart watches Can detect Bluetooth devices that are not already paired and does not need to have an open setting menu for detection	Detects if any two Bluetooth devices are paired and connected	Any type of Wi-Fi communication can be detected regardless of being connected to the internet or not.

Detection zones vary depending on antenna type, indoor vs outdoor environments, surrounding objects, and Received Signal Strength Indicator (RSSI value) of the signal-emitting device. Two antennas with different detection zone patterns are available: directional and omni-directional antennas. The omni-directional antennas have a circular detection zone while the directional antennas provide a stronger detection pattern in front of the device compared to the back. Directional antennas were used in this study. The slope, vegetation and surrounding environment all affect the detection zone when a device is set up in an outdoor environment. Our experiment with this device in different locations and at different heights confirms the dynamic detection zone in accordance with the surrounding environment. Figure 10 illustrates the results of a detection zone experiment of a TrafficBox with directional Wi-Fi and Bluetooth antennas set at a three-meter height above the street. The figure includes three images. The top image shows the LSU Student Union and the location of the TrafficBox when the experiment was conducted. The bottom right image shows the detection zone of one side of the TrafficBox, and the bottom left image shows the TrafficBox detection zone of the other side of the TrafficBox. In both images at the bottom, semi-circular detection zones can be seen for both sides of the device.

Figure 10. Detection zone experiment



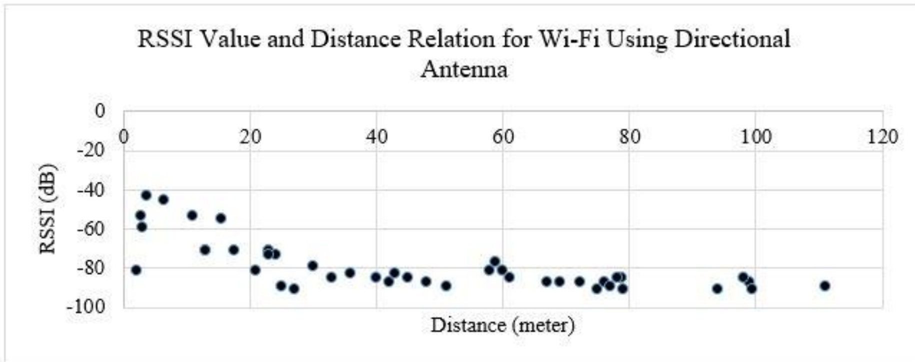
Gray dotted lines are walking paths from the farthest open space toward the device. Yellow points indicate the simultaneous detection of Bluetooth and Wi-Fi signals at the same location in less than 5 seconds. Red points represent the location of Bluetooth detections and black points represents the location of Wi-Fi detections. Both detection zones confirm that Bluetooth signals usually have a larger detection zone than Wi-Fi. In the lower left diagram in Figure 10, the furthest Bluetooth detection is at 223 m. and the closest is at 70 m. from the TrafficBox. The furthest Wi-Fi detection point occurred at 202 meters and the closest at 0 meters from the TrafficBox. The experiment shown in the diagram on the lower right-hand side of Figure 10 involved facing the other side of scanners toward the walking area. In this case, the furthest and closest Bluetooth were at 260 m. and 86 m., and the furthest and closest Wi-Fi detections were at 174 m. and 4 m. meters from the TrafficBox. This shows that there was little difference in the detection zones in front and behind the directional antennas in this case.

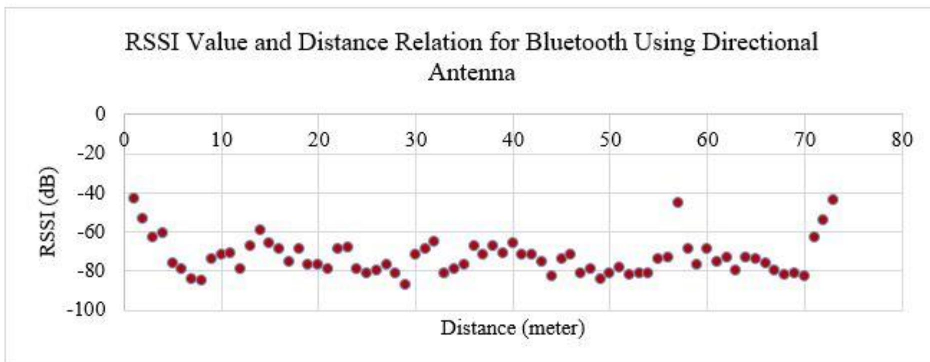
To adjust the detection zone of the TrafficBox, lead or aluminum can be used to block electromagnetic signals entering the sides of the TrafficBox covered by these metals. Because aluminum is lighter and cheaper than lead, we tested the use of multiple layers of aluminum foil behind the device. We found Wi-Fi signals could be blocked using thick layers of aluminum foil, but Bluetooth signals could not be blocked. However, the results of the test were not definitive because microwave signals such as Wi-Fi and Bluetooth signals can reflect off surfaces to move in different directions [30]. Thus, a device behind a TrafficBox with a shielded back, can still be detected if its signal is reflected off a surface in front of the TrafficBox. The result is that TrafficBoxes detect in roughly circular patterns, even with directional antennas.

Signal Strength

The Received Signal Strength Indicator (RSSI) is a measure of the strength of a signal from a device measured in decibels. RSSI values typically range between -30 and -100, with -30 being the strongest signal. Devices have vastly different signal strengths. RSSI values have been used to estimate the distance of a device from a scanner [31], [32], [33] However, we found RSSI values to be a poor indicator of distance given the wide range of signal strengths among devices and the relatively weak decline in signal strength with distance. As an illustration, Figure 11 shows the results of an experiment that was conducted testing the impact of distance on signal strength. A TrafficBox was located 3m above the ground and 2 mobile phones of different brands (Sony and Samsung) were used in detection mode for both Wi-Fi and Bluetooth. The phones were moved from beside the TrafficBox to more than 400 meters distance radially in different directions in a circle around the device. In each radial trajectory, fieldworkers halted at regular intervals for 10 seconds and noted their GPS location. The Sony phone was detected by the Wi-Fi scanner 68 times and by the Bluetooth scanner 119 times. In contrast, the Samsung phone was detected by the Wi-Fi scanner 70 times and by the Bluetooth scanner twice. This disparity between devices and between scanners, and even differences among TrafficBoxes with the same settings, was observed throughout the study.

Figure 11. RSSI as a function of distance





Data Storage

In a TrafficBox detections are recorded as MAC (Media Access Control) addresses stored in the "fingerprint" column of a csv file. Wi-Fi MAC addresses are recorded as a 12-digit hexadecimal number including all 6 bytes of the MAC address. Bluetooth MAC addresses are recorded as the last 3 bytes of a MAC address or the last 6-digit hexadecimal number. The "Detection Type" column in the file produced by the TrafficBox also differentiates between Wi-Fi and Bluetooth MAC addresses by giving them a value of 2 for Wi-Fi or 3 for Bluetooth. The GPS inside the TrafficBox allows the setting of local time automatically although it does not allow for Daylight Savings Time. The TrafficBox can be set to commence recording detections by date and time.

In the setting panel on the TrafficBox, MAC Hashing, Ignore Random MAC addresses, Ignore MAC Interval, and Probe Request affect data quality and eliminate privacy concerns. The MAC Hashing option stores detected MAC addresses in a modified form thereby eliminating the possibility of tracking a detection to a specific device. The Ignore Random MAC addresses option identifies devices using MAC Address Randomization

(MAR) and avoids recording a device multiple times. MAR replaces the unique MAC address of a device with randomly generated values in order to avoid being tracked by Wi-Fi detection technologies. Therefore, Wi-Fi devices using MAR will be detected and stored only by their first detection. The Ignore MAC Interval setting saves memory space and battery life by not recording repeated observations of the same device more frequently than a specified time period. In this study, the Ignore MAC Interval was set at 5 seconds so that a device was not recorded more frequently than at 5-second intervals. The Probe Request is a filtering tool to detect a particular type of Wi-Fi MAC address only. It can be set to capture cell phones with Wi-Fi turned on and searching for an access point and removes devices that are connected to the internet. Although it can remove computers and residents in the detection area, it will also exclude workers and visitors using free Wi-Fi at the mall. Therefore, Probe Request was not used in this study and undesired devices were filtered out using another filtering procedure described in the analysis section in detail.

Signal Frequency

Although there is no control over the time and frequency of communications between the TrafficBox and target devices, our experiments suggest a 1- to 2-minute time gap between repeated detections of eligible devices that are not currently in use (e.g., when a cell phone screen is off and Wi-Fi is disconnected but it is in searching mode). Eligible devices that are in use (e.g., when a person is listening to music using a Bluetooth-connected headphone) have a 1-30 second time gap in their communications with the TrafficBox. The time variation of communication is due to variations in the vendor. Although it is not possible to identify whether Wi-Fi and Bluetooth MAC addresses are coming from the same device, it is possible to identify the vendors of Wi-Fi and Bluetooth MAC addresses using freely available MAC address Look Up programs and websites such as Arul's Utilities (https://aruljohn.com/mac/3464A9639CFC). Although MAC address Look Up programs increase our understanding of the detection patterns of different devices, it is not practical when MAC Hashing is used. Since some vendors specifically produce non-portable hardware such as desktops, MAC address Look Up programs can be helpful in increasing an analyst’s understanding of the detection patterns of different device types.

Battery Life

The battery life of the TrafficBoxes was observed to be about 16 hours to 34 hours. Since the TrafficBoxes purchased in this study cannot be programmed to turn on and off at specific times and there are difficulties associated with commuting to survey sites multiple times to turn TrafficBoxes on and off, we allowed the TrafficBoxes to run from 8 am on the first survey day until the battery ran out on the second day. In general, TrafficBox data are available for the first day of survey at all survey sites and recorded data till 12 pm to 3 pm in different locations on the second day of the survey.

Identifying Relevant Detections

Since the RSSI value was found to be an unreliable indicator of the distance in this study, we tested the idea of identifying trips to an individual land use (e.g., strip mall) from Wi-Fi and Bluetooth data by putting two TrafficBoxes a certain distance from the farthest corners of the strip mall in a way that the detection areas of the TrafficBoxes would overlap the strip mall area. Thus, vehicle trips to the strip mall would be identified by being detected by both TrafficBoxes no more than a certain amount of time apart.

To implement this idea, two TrafficBoxes were located 200 meters from the two furthest corners of a strip mall as shown in Figure 12. The strip mall is shown by a black rectangle at the site. The two TBs are shown as red triangles in the diagram. Their detection zones, assumed to be 200 meters, are shown by yellow circles. As can be seen, the strip mall falls within the joint detection area. The 200-meter detection radius was chosen as the typical upper limit of Wi-Fi and Bluetooth devices so that any detected visit of the same device by both TrafficBoxes would likely represent a visit of that device to the joint detection area (i.e., the strip mall). Devices with smaller detection radii would not be jointly detected and thus would not be counted, while those with larger detection radii would allow joint detections in an area extending either side of the mall. Further qualifications on the use of this approach are addressed in the following sections.

Figure 12. Overlap identifying strip mall visits

Removing Static Devices

Because TrafficBoxes record all the Wi-Fi and Bluetooth devices in the detection zone, static devices at the site such as laptops and desktop computers paired with smart watches, headsets, and other devices will be recorded as well. These static devices need to be excluded from the data, since they do not belong to the visitors/customers. Subsequently, all devices that recorded continuously for 7 hours or more out of the 10 hours data collection time period on each day (from 8 am to 6 pm) were removed. This value was chosen as the result of a trial and error procedure to identify the most suitable time period for removing static devices at the site while not excluding devices carried by workers.

Identifying and Removing Through Traffic

Because TrafficBoxes were located beside the road, all Bluetooth and Wi-Fi MAC addresses of vehicles and their passengers passing by on the adjacent road, and even on surrounding roads are likely to be recorded. To remove these records, passing vehicles are defined as MAC addresses observed by both TrafficBoxes without a break of more than two minutes. Two minutes is assumed to be the shortest time required for a person to enter a store, conduct their business, and then emerge from the store again. As verified earlier, when the TrafficBoxes are mounted outside a building, they are able to record a MAC address of an emitting device in open space but observations cease whenever the device is carried into a building. As none of the passing vehicles are entering any building, they are recorded roughly every 5 seconds with the 5-second MAC ignore setting in operation. Therefore, their continuous observation identifies them as devices in passing vehicles and they are removed from the data.

Identifying Devices Visiting a Mall

In this step, all MAC addresses that have at least one visit (i.e., at least a two-minute break in a series of MAC address detections) are identified as ones visiting the mall. For such series of records, the very first record of the device is when the device enters the detection zone of an individual TrafficBox (coded as 1). After that, at least every 5 seconds the device is recorded (coded as 0), until the visit of a business starts as signified by at least a two-minute break in detections. The last record of these detections is coded as an entry to the mall. If only one visit happens, the person comes out after at least two minutes and the first detection after the visit is coded as a departure. After the departure record, the device is recorded (code 0) at least every 5 seconds till it gets out of the detection zone (code 3). Theoretically, in the case of visits to multiple stores in one strip mall, the "entry" will be the last detection before the first visit and "departure" will be the first detection after the last visit. However, it does not always happen precisely this way in practice. Sometimes, only one TrafficBox will pick up the signal when the person exits. Sometimes, only one TrafficBox will have detected the entry. So, there are times when we have a device detected by one TrafficBox only on entry, and both detect on exit, sometimes when both TrafficBoxes detect on entry, but only one on exit, and sometimes when both record both entry and exit.

Identifying Common Visits Recorded by Both TrafficBoxes

The next step identifies all the visits recorded by TrafficBoxes in terms of entries and departures and the time they happen. However, these visits might happen anywhere in the detection zone of the TrafficBoxes. In order to end up with visits to the overlapped area only, the first step is to select the visits that are recorded by both TrafficBoxes. In fact, if someone visits a land use in the detection zone of one of the TrafficBoxes (either of them) and passes the detection zones of both TrafficBoxes, this visit can be identified as a non-overlapped area visit, because the 2-minute gap or visit is recorded by only one of the TrafficBoxes. Figures 13 and 14 show all the possible visits at the site that are not recorded by both TrafficBoxes and therefore are removed from the data. Black arrows show the direction of movement. Figure 13 shows two TrafficBoxes and their detection zones. It shows that there was a visit in the detection zone of the TrafficBox located on the right side of the figure, but no visit in the detection zone of the TrafficBox on the left side. The converse situation is shown in Figure 14.

Figure 13. A visit to the right TB's detection zone

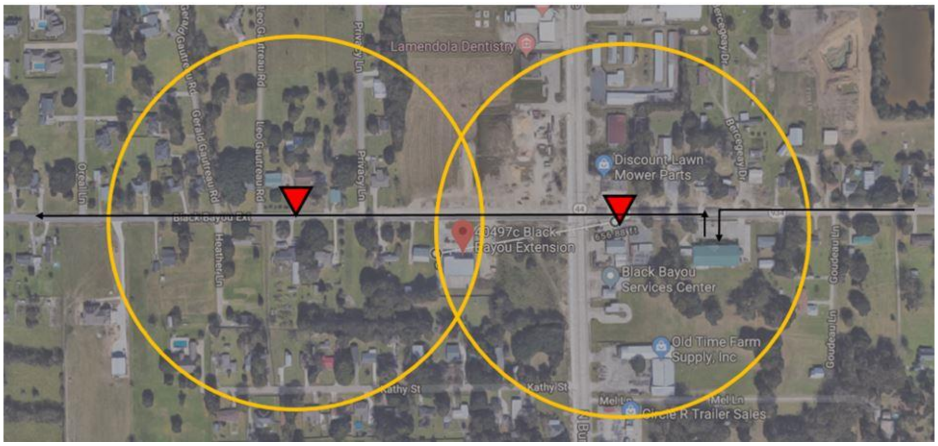
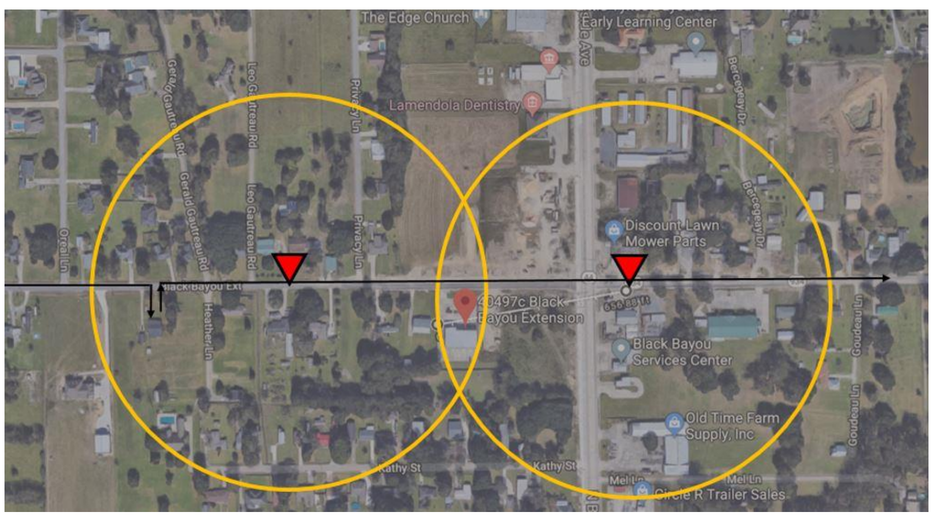
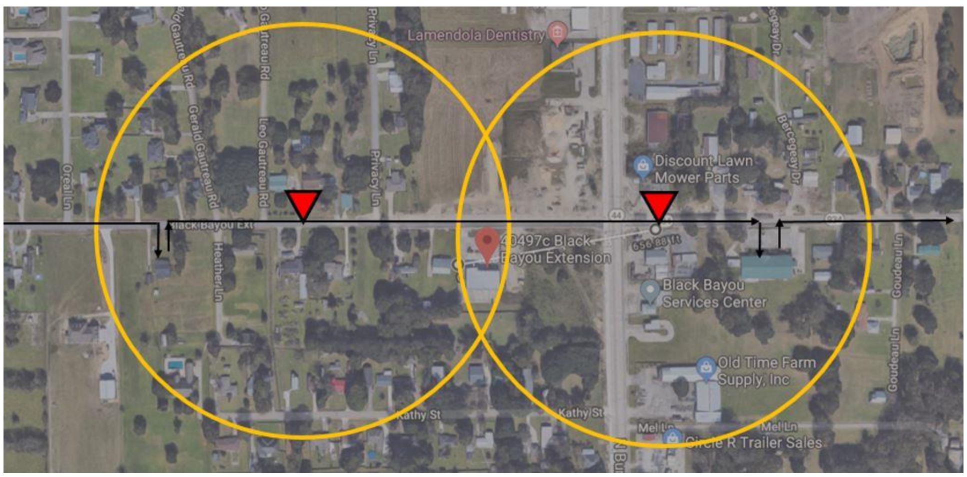


Figure 14. A visit to the left TB's detection zone



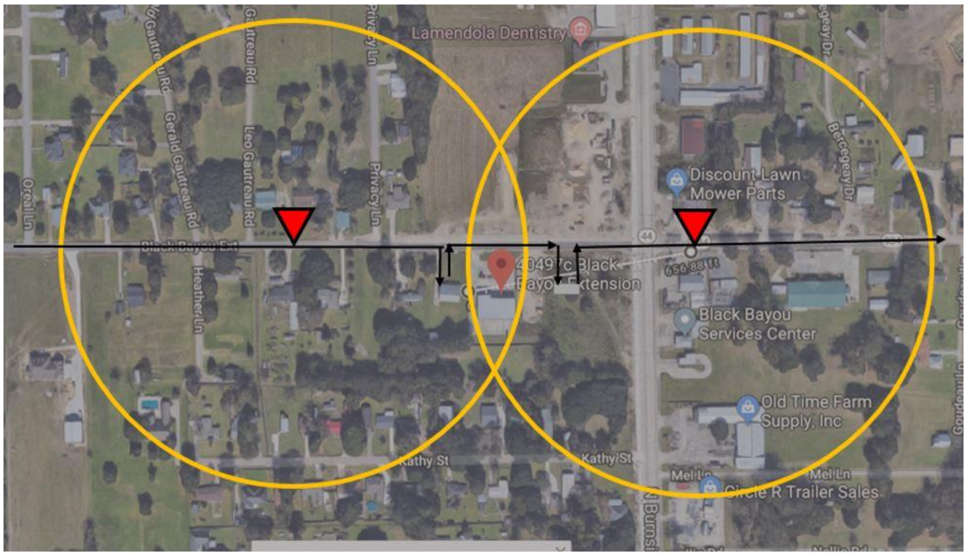
The other issue that might raise here is the possibility of both TrafficBoxes recording a visit, but the device is in fact visiting two separate land uses outside the common area as shown in Figure 15. The strip mall is shown by a black rectangle at the site. Two TrafficBoxes are located at 200 meters distant from the furthest corner of the mall and are shown with red triangles. Their detection zones are shown with yellow circles. The joint detection area covers the strip mall. A vehicle conducts a visit first in the detection zone of the left side TrafficBox, then passes by the target strip mall and then makes a second visit in the detection zone of the right side TrafficBox. Such trips are not frequent but they can be identified by checking the time of the visit recorded at each TrafficBox. Trips to the common area have more or less the same visit time recorded by both TrafficBoxes, while visits out of the common area, but recorded by both TrafficBoxes, have different visiting times.

Figure 15. Two visits out of overlapped area recorded by both TrafficBoxes



Although it is rare that a person visits two businesses that are closest right and left neighbors of the desired strip mall, we consider the most extreme case here to make sure all the undesired records are eliminated. Assuming Figure 16 shows a person visiting closest neighbors of the desired strip mall (but out of common area of the two TrafficBoxes) and is recorded by both TrafficBoxes. If a person starts a visit at the business A at time X, it takes at least 3 minutes and fourteen seconds after X to start the second visit at business B. It is the sum of 2 minutes visit at business A, 30 seconds walking time from business A to the person's car and exiting the parking area, at least 14 seconds driving between the two businesses (average speed of 30 mi/h and 200 meter distance between businesses) and another 30 seconds of parking inside of the second business parking area and walking inside business B. Therefore, only the visits recorded by two TrafficBoxes that are less than 194 seconds apart from each other are assumed to belong to the common area and desired strip mall.

Figure 16. Visits to businesses closest to either end of the joint detection area



Identifying Entries and Exits

After identifying entries and departures common to both TrafficBoxes of visits to a strip mall, final entry and exit times for each device/MAC Address are selected out of two entries and two departures recorded by the two TrafficBoxes. The two entries under consideration are the last records of a specific device right before the first visit recorded by each device, while the departures are the first detections after the last visit (in case of multiple visits at strip malls) recorded by the TrafficBoxes. Therefore, four records of each MAC address (entry and departure of TrafficBox no. 1 and entry and exit of TrafficBox no. 2) are produced. Final "entry" record of that MAC address is considered the latest record of entries by both TrafficBoxes, while its "exit" from the mall is considered as the earliest record of departures recorded by both TrafficBoxes.

Classifying Entries and Exits into 60-Minute Intervals

After identifying the entry and exit for each MAC address recorded by both TrafficBoxes in the overlapping area of the strip mall, they are sorted by recorded time. The number of combined Wi-Fi and Bluetooth entry, exit, and total detections by time interval are established. In this case a time interval of one hour was chosen since it provided a balance between the number of detections in each time interval and the number of intervals.

Establishing Adjustment Factors

Adjustment factors to be applied to the ITE trip generation trip rates in Louisiana were identified by first using the data collected at the 40 strip malls to identify the relationship between trip generation at strip malls in Louisiana and the expanded set of factors hypothesized to contribute to trip generation. That is, a regression model was estimated that related the observed trip generation at strip malls to the floor area of each mall, population density and land use diversity within a half-mile radius of the site, and the intensity of traffic on the adjacent road. Once this regression model was established, the difference between it and the equation in the ITE *Trip Generation Manual* was identified as the adjustment factor. To illustrate:

Say the new equation is: $New\ trip\ rate = a + b(floorarea) + c(density) + d(diversity) + e(traffic)$

Hypothetically, if the ITE trip rate equation for strip malls is: $ITE\ trip\ rate = f + g(floorarea)$

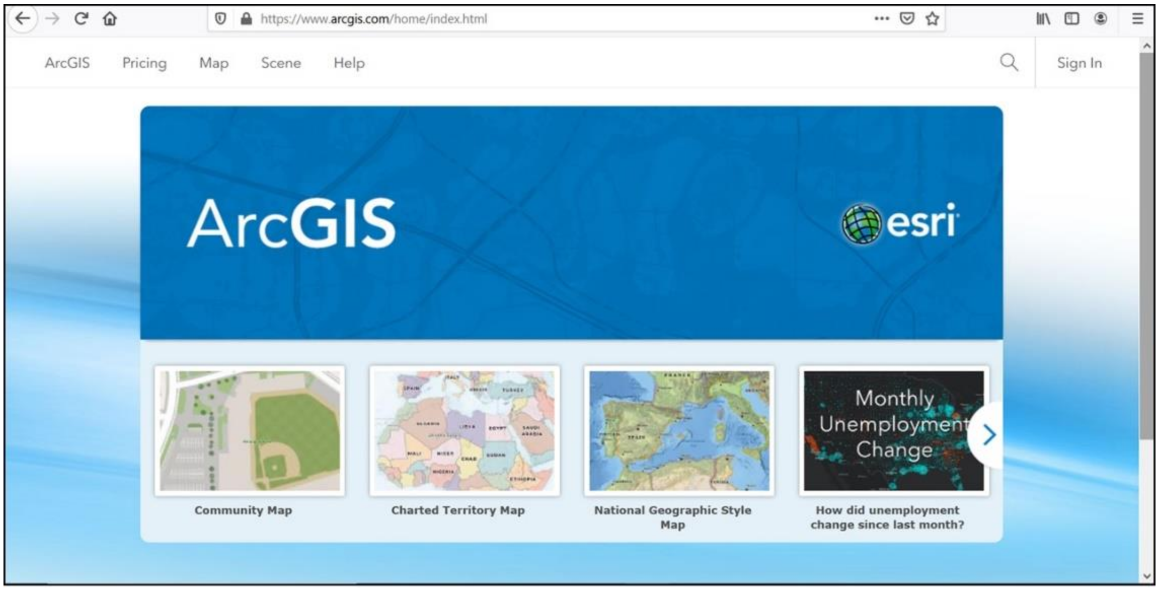
Then the adjustment factor is: $Adjustment\ factor = (a - f) + (b - g)(floorarea) + c(density) + d(diversity) + e(floorarea)$

GIS Interface

Creation of a Web Map for the ITE Trip Rates in Louisiana

Using the ArcGIS Online service from ESRI, a web map of the study area was created with the purpose of disseminating the modified trip generation estimates and allowing access to the data on which the estimates are based, in a convenient manner. The web map illustrates the results of the modified trip rates in the study area of this project specifically. Figure 17 shows the home page of ArcGIS accessed from www.arcgis.com.

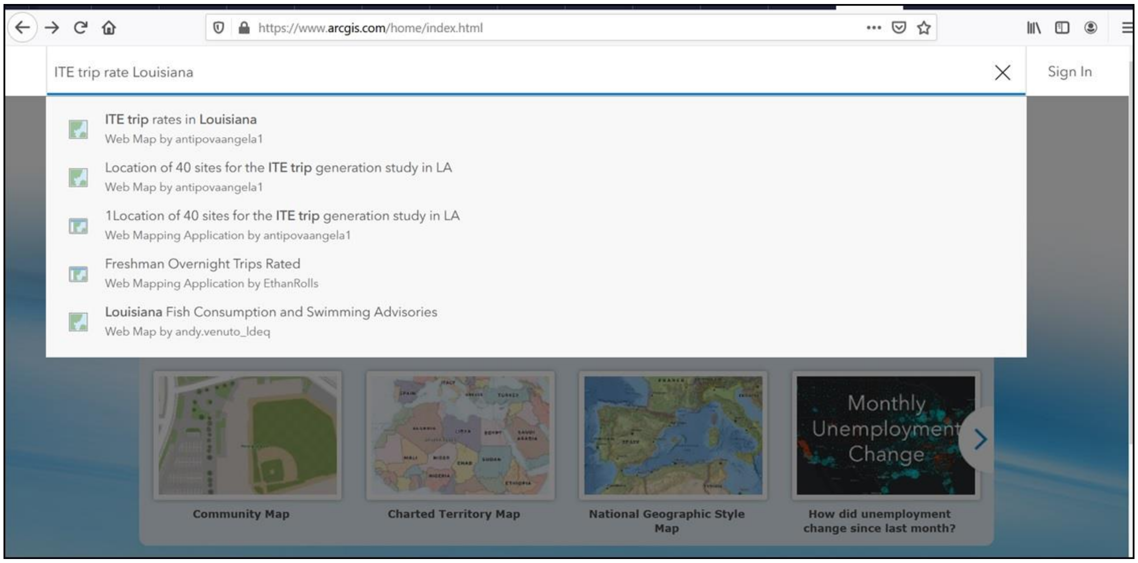
Figure 17. ArcGIS home page



Access

The web map titled the "ITE trip rates in Louisiana" can be accessed from the link: https://www.arcgis.com and by searching for: ITE, Trip rates, Louisiana (the keywords need to be typed in the top right corner to search ArcGIS Online). Below will appear the title: "ITE trip rates in Louisiana" as shown in Figure 18. By clicking on the title of the map, the map information page is seen. From there, by clicking on the map it can be opened in Map Viewer. Results from application of the program to the study area of this project are shown in the next chapter on Analysis and Results.

Figure 18. Accessing the modified trip generation rates



Analysis and Results

Data Analysis

Manual Counting

Forty strip malls were surveyed for two consecutive days from 8 am to 6 pm. Vehicle trips were counted by direction for every five-minute interval. After verifying the accuracy of the manual counts by conducting re-counts of random sites, seven out of 80 survey days with undesirable weather conditions (heavy rain) and incomplete video records (due to battery loss and similar issues) were identified and excluded from the whole day data. These seven days all belong to the second day of the site surveys. Data from the remaining 73 survey days of the 40 sites were incorporated into an Excel sheet along with the ground truth total trips for the whole day. However, three sites out of the seven omitted sites have complete data on the PM peak period (4 to 6 pm), so they were included in the data file used to conduct the afternoon peak period analysis.

According to the ITE *Trip Generation User's Guide* (8th edition), if traffic counts by time on the adjoining road are not available to allow the identification of the peak hour, it may be assumed that the hour in which the maximum number of trips to the sites occurs is the peak hour on the adjoining road. Because the number of trips were available in each 5-minute period between 4 pm and 6 pm in this study, the peak hour trips were calculated as the maximum number of trips to the site in 12 consecutive 5-minute counts between 4 pm and 6 pm for each day.

The results of the manual counts are shown in Table 9. Trip rates in the table are the number of trips per 1000 sq. ft. of gross floor area of the strip mall at each site. The average daily trip rate for all observations in the table (all sites, both days) is 40.82. This coincides closely with the fixed value of 42.78 in the 8th edition ITE *Trip Generation Manual*, but the observed variation in trip rates in the table show just how much is not captured by the ITE value. The standard deviation of the observed daily trip rates in Table 9. Total vehicle trip counts and trip rates of the survey sites is 26.08 trips per day, which is high, but it must be recalled that the sites were chosen to specifically provide a wide variation in three of the conditions thought to influence trip generation, namely residential density, land use diversity, and traffic intensity. Whether these three conditions are responsible for all the variation observed is the subject of investigation later in the report.

Table 9. Total vehicle trip counts and trip rates of the survey sites

Site No.	First Day Total Trips	First Day PM Peak Hour Trips	First Day Trip Rates	First Day PM Peak Hour Trip Rates	Second Day Total Trips	Second Day PM Peak Hour Trips	Second Day Trip Rates	Second Day PM Peak Hour Trip Rates
1	595	88	34.34	5.08	540	88	31.17	5.08
2	369	41	13.28	1.48	455	43	16.38	1.55
3	1054	105	102.51	10.21	1227	136	119.33	13.23
4	599	75	40.55	5.08	588	73	39.8	4.94
5	390	36	39.92	3.68	-	-	-	-
6	460	70	66.5	10.12	521	62	75.32	8.96
7	373	38	20.25	2.06	325	29	17.64	1.57
8	1137	188	59.06	9.76	1235	209	64.15	10.86
9	350	54	54.99	8.48	323	56	50.75	8.8
10	549	78	27.07	3.85	649	118	32	5.82
11	336	54	23.09	3.71	407	80	27.97	5.5
12	789	87	31.69	3.49	933	128	37.48	5.14
13	693	71	44.23	4.53	733	68	46.79	4.34
14	377	65	17.1	2.95	405	67	18.37	3.04
15	700	85	120.42	14.62	581	55	99.95	9.46
16	178	15	25.16	2.12	220	25	31.1	3.53
17	276	29	19.81	2.08	224	44	16.08	3.16
18	302	58	37.81	7.26	-	-	-	-
19	664	116	32.84	5.74	716	136	35.42	6.73
20	280	37	28	3.7	417	81	41.7	8.1
21	405	61	21.68	3.27	453	81	24.25	4.34
22	660	94	60.16	8.57	653	75	59.53	6.84
23	345	49	14.33	2.03	-	-	-	-
24	734	124	60.92	10.29	787	133	65.32	11.04
25	386	73	30.03	5.68	524	93	40.77	7.24
26	150	16	23.83	2.54	118	11	18.75	1.75
27	422	75	17.14	3.05	453	48	18.4	1.95
28	306	52	36.82	6.26	307	42	36.94	5.05
29	283	38	21.61	2.9	203	23	15.5	1.76

Site No.	First Day Total Trips	First Day PM Peak Hour Trips	First Day Trip Rates	First Day PM Peak Hour Trip Rates	Second Day Total Trips	Second Day PM Peak Hour Trips	Second Day Trip Rates	Second Day PM Peak Hour Trip Rates
30	621	93	88.65	13.28	653	103	93.22	14.7
31	143	13	18.77	1.71	-	-	-	-
32	300	66	34.34	7.55	-	35	-	4.01
33	235	40	22.42	3.82	904	28	86.24	2.67
34	943	141	24.27	3.63	-	105	-	2.7
35	286	69	25.77	6.22	904	47	81.44	4.23
36	67	8	6.62	0.79	115	7	11.36	0.69
37	1055	115	39.27	4.28	949	115	35.33	4.28
38	720	70	68.19	6.63	639	66	60.52	6.25
39	481	89	21.26	3.93	-	50	-	2.21
40	666	90	28.12	3.8	667	81	28.16	3.42

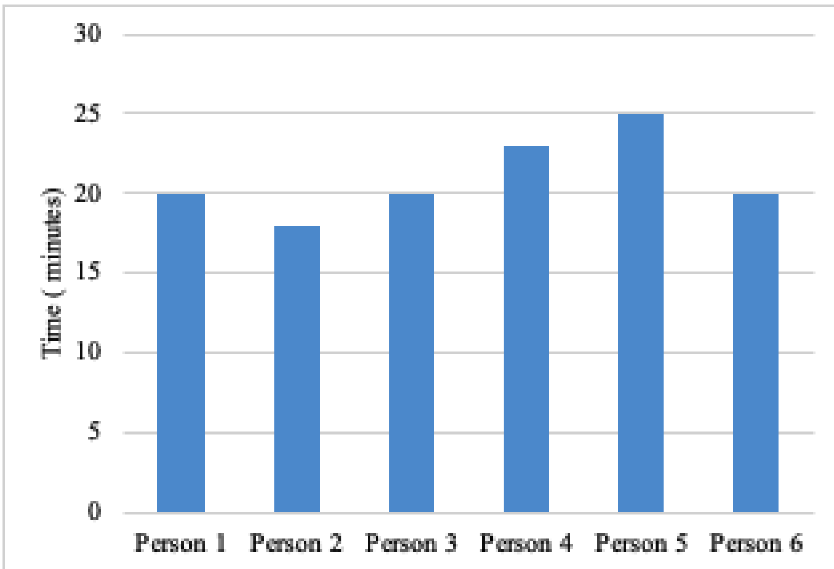
The vehicle counts, and subsequently the trip rates, are reasonably consistent between day 1 and day 2 in most cases, but some sites (e.g., sites 33 and 35) show a dramatic change between the days. This may be due to a special event on one of the days but considering such events are not likely to be known by the survey administrator in advance, the possibility of daily variations must be taken into consideration when designing the survey. One possibility may be to extend the counting period to identify outliers

Time Taken to Conduct Manual Counts

Conducting manual counts from pre-recorded videos in real-time is time-consuming. However, the ability of media players to speed up or slow down playback can save time. For example, the VLC media player has the ability to increase the playback speed up to 16 times.

In this study, six individuals conducted the manual counts. Their reported time for an hour video count is shown in Figure 19. From the figure, it can be observed that the average time taken to count one hour of recorded vehicle movement is 21 minutes. The individuals conducting the manual counting reported difficulty in maintaining concentration after looking at a particular point (i.e., reference line) on the computer screen for a long time. They reported that after counting continuously for about an hour, they had to take a rest. If this break is taken into consideration, then the average counting time for one hour of recording is probably on the order of 30 minutes. The quality of videos also controls the duration of manual counting. Low quality videos result in extended counting time because individuals cannot increase the playback speed without running the risk of failing to recognize all the vehicles.

Figure 19. Manual counting time



Error in Manual Counting

As mentioned in the Methodology chapter, there is no actual ground value to estimate errors of manual counting. But, errors can be estimated by conducting repeated counts and accepting them as actual counts. If the first repeated count is the same as the first count, it is considered confirmation that the first count was accurate. But, if the repeated count is different from the first count, conducting a second repeated counting is required. Since repeated counts cost double time and money, it is necessary to limit repeated counting to a few sites. In this study, five sites were randomly selected for repeated counting. Different people were selected to conduct the repeated counts from the first. Sites selected for repeated manual counting are shown in Table 10.

Table 10. Selected sites to measure error in manual counting

Site No	Site Name
1	6031 Siegen Ln, 70809
12	702 N Lobdell Hwy Suite 9, Port Allen, LA 70767
21	12240 Coursey Blvd, 70816
31	4404 Moss St, Lafayette, LA 70507
39	13091 Airline Hwy, Gonzales, LA 70737

Error in Daily Counts

First and repeated counts of daily entry, exit, and total vehicle trips at individual sites are shown in Table 11 for day 1 and Table 12 for day 2. Percent error is shown for each site for entry, exit, and total counts for day 1 and day 2 separately. Reviewing the errors in the tables shows under-counting is the dominant form of error. Over-counting occurs in only three cases: exit count of site 1 for day 1, exit count of site 21 for day 1, and entry count of site 31 for day 1. The magnitude of over-counting error also appears to be smaller than under-counting error from the results in Table 11 and Table 12. The potential reasons for under-counting from those who conducted the counting were reportedly poor visibility on the video screen, high speed of entering vehicles, and one vehicle obscuring the view of another at the point of observation.

Table 11. Daily count error for day 1

Site No.	Entry			Exit			Total		
	First	Repeat	Error (%)	First	Repeat	Error (%)	First	Repeat	Error (%)
1	306	309	1.0	289	287	-0.7	595	596	0.2
12	405	410	1.2	384	387	0.8	789	797	1.0
21	221	224	1.4	184	181	-1.7	405	405	0.0
31	76	75	-1.3	67	68	1.5	143	143	0.0
39	251	254	1.2	230	232	0.9	481	486	1.0

Table 12. Daily count error for day 2

Site No.	Entry			Exit			Total		
	First	Repeat	Error (%)	First	Repeat	Error (%)	First	Repeat	Error (%)
1	284	288	1.4	256	259	1.2	540	547	1.3
12	481	487	1.2	452	454	0.4	933	941	0.9
21	246	247	0.4	207	209	1.0	453	456	0.7
31	59	59	0.0	52	53	1.9	111	112	0.9
39	261	264	1.1	237	239	0.8	498	503	1.0

To get an overall estimate of the error in manual counts, percent root mean square errors (%RMSEs) of the total counts were calculated using formula (1) as shown in the methodology section. The percent RMSE for day 1 was found to be 0.65 percent and for day 2 was found to be 0.96 percent. The estimated RMSE of total counts error was found to be 0.82 percent.

Classification Error

The classified first counts were compared with the classified repeated counts to estimate classification error. The classification error in vehicle number, repeated counts, and error for day 1 is shown in Table 13 and for day 2 in Table 14. Classification errors of first, repeated, and total counts were estimated for day 1 and day 2 individually for each site.

Table 13. Daily classification count error for day 1

Site No	Entry			Exit			Total		
	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)
1	3	309	1.0	4	287	1.4	7	596	1.2
12	3	410	0.8	6	387	1.6	9	797	1.1
21	2	224	0.9	1	181	0.6	3	405	0.7
31	1	75	1.4	0	68	0.0	1	143	0.7
39	4	254	1.6	2	232	0.9	6	486	1.2

Table 14. Daily classification count error for day 2

Site No	Entry			Exit			Total		
	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)
1	2	288	0.7	4	259	1.5	6	547	1.1
12	9	487	1.8	3	454	0.7	12	941	1.3
21	3	247	1.2	4	209	1.9	7	456	1.5
31	0	59	0.0	1	53	1.9	1	112	0.9
39	2	264	0.8	3	239	1.3	5	503	1.0

The percent RMSE of classification counts was calculated according to equation (2) in the methodology section. The estimated RMSE of day 1 was found to be 1.02 percent and for day 2 was 1.18 percent. Total classification RMSE was found to be 1.10 percent.

Interval Error

Interval error occurs when a vehicle is counted in an incorrect time interval. In this research, a 5-minute interval was considered for counting. The probability of an interval error increases as the interval time decreases, i.e., the probability of an interval error when using 5-minute intervals is higher than when using 15-minute intervals. When interval time is decreased, counting has to be conducted in more subdivisions that increases the probability of errors.

In this study, interval error was calculated by comparing the first counts with the repeated counts. Table 15 shows the interval error in vehicle number, repeated counts, and error for day 1 and Table 16 shows the same data for day 2.

Table 15. Daily interval count error for day 1

Site No	Entry			Exit			Total		
	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)
1	3	309	1.0	4	287	1.4	7	596	1.2
12	3	410	0.7	5	387	1.3	8	797	1.0
21	2	224	0.9	4	181	2.2	6	405	1.5
31	2	75	2.7	0	68	0.0	2	143	1.4
39	4	254	1.6	1	232	0.4	5	486	1.0

Table 16. Daily interval count error for day 2

Site No	Entry			Exit			Total		
	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)	Error (Nos)	Repeated Counts	Error (%)
1	4	288	1.4	3	259	1.2	7	547	1.3
12	4	487	0.8	9	454	2.0	13	941	1.4
21	5	247	2.0	2	209	1.0	7	456	1.5
31	1	59	1.7	0	53	0.0	1	112	0.9
39	3	264	1.1	6	239	2.5	9	503	1.8

Equation (3), as shown in the methodology section, was used to estimate the percent RMSE of interval counts. RMSE of interval counts were found to be 1.23 percent for day 1 and 1.40 percent for day 2. The total interval RMSE was found to be 1.31 percent.

Fieldworker Feedback on Manual Count Error

In this study, the individuals who conducted manual counts reported some potential reasons for the errors in counts as follows:

- Due to a manual increase in frames per second in a media player, the chance of failure to recognize vehicles increases and is the main reason for total, classification, and interval error.
- It is hard to report vehicles for the videos which are recorded in the evening, heavy rain, or fog because, in these times, the quality of the video image is low.
- Raindrops obscure the camera lens, which makes a dark video frame, and individuals fail to report vehicles.
- Sometimes a queue of vehicles arrives and departs at the same time. In that case, the probability of error rises.

General Observations on Manual Counting

In the case of total error, it was observed that underestimation of counts is the most frequent scenario because individuals generally miss reporting vehicles. Overestimation of counts happens when a queue or a group of vehicles arrives or departs at high speed. Classification and interval errors do not have any effect on the total counts. Interval errors occur between the end and start of an interval causing an overcount in one interval and an undercount in the other, so they are not very important. Total and classification counts are the only counts generally considered in practice. Total counts are usually used to calculate daily trips. Interval counts are used to estimate peak hour volume or expanded traffic counts.

Automated Counting

Automated counting was investigated in this study as a possible way to avoid the labor-intensive practice of manual counting when establishing the trip generation of a land use. If automated counting can reduce manual effort sufficiently, local trip rates could more easily be established, sample size could be increased, and more land uses analyzed.

After reviewing different technologies potentially capable of automated counting, two were selected for further testing: video imaging and Bluetooth and Wi-Fi signal detection. The results of their application and evaluation is reported in the following sections.

Image Detection from Video Recordings

The analysis of images from video recordings was conducted in accordance with the method described in the Methodology section of this report. The effort involved in applying the method is the effort to install and retrieve video cameras at each site, and then to convert video images to counts. Automation only affects the second activity. In this study, each site took approximately 12 man-hours including recounts to validate the survey results. Processing the video by computer took approximately 2 man-hours per site including file preparation, program execution, and documenting the output using the system listed below:

Processor: Intel(R) Core (TM)i7-8750CPU @ 2.20 GHz 2.21 GHz

RAM: 2.7 GHz, 16.0 GB

GPU: NVIDIA GeForce GTX 1060, 6GB

On this platform, it was found that an hour of video takes 1.4 to 1.5 hours to process. It is recommended to use a high configuration computer for processing videos. A GPU of 2060 and graphics more than 6 GB are recommended for the processing.

Analyzing Video Files

In this study, all the videos were in mp4 format and did not require any format conversion. Videos were joined to form a ten-hour video. The unnecessary parts of videos were trimmed using the Avidmux software (an open-source video editing software). Then, the video file, file name, and time interval were provided to process the video.

In this section, analysis of pre-recorded videos by computer algorithms is reported on. They are used to automate the count of the entry and exit of vehicles to a site. Because computer processing of the video files took approximately 1.5 times the recording time, a sample of sites were randomly selected from the 40 sites to conduct automated counting.

The automated counting results from 10 randomly selected sites are shown in Table 17.

Table 17. Automated counting data in vehicles per day

Site No	Day 1		Day 2	
	Entry	Exit	Entry	Exit
1	295	269	268	235
3	481	485	571	564
5	172	168	164	162
11	154	145	182	181
15	335	331	281	264
18	114	124	112	131
21	212	164	229	184
23	162	145	121	112
28	145	134	132	124
32	134	124	118	109

Accuracy Evaluation

Automated counts were compared with manual counts to estimate the accuracy of the automated counting. The accuracy of daily entry, exit, and total counts of individual sites for day 1 are shown in Table 18 and day 2 in Table 19.

Table 18. Accuracy of automated counts for day 1

Site No	Entry			Exit			Total		
	Manual	Automated	Accuracy (%)	Manual	Automated	Accuracy (%)	Manual	Automated	Accuracy (%)
1	309	295	95.47	287	269	93.73	596	564	94.63
3	526	481	91.44	528	485	91.86	1054	966	91.65
5	196	172	87.76	194	168	86.6	390	340	87.18
11	172	154	89.53	164	145	88.41	336	299	88.99
15	354	335	94.63	346	331	95.66	700	666	95.14
18	149	114	76.51	153	124	81.05	302	238	78.81
21	224	212	94.64	181	164	90.61	405	376	92.84
23	180	162	90	165	145	87.88	345	307	88.99
28	162	145	89.51	144	134	93.06	306	279	91.18
32	159	134	84.28	141	124	87.94	300	258	86

Table 19. Accuracy of automated counting for day 2

Site No	Entry			Exit			Total		
	Manual	Automated	Accuracy (%)	Manual	Automated	Accuracy (%)	Manual	Automated	Accuracy (%)
1	288	268	93.06	259	235	90.73	547	503	91.96
3	617	571	92.54	610	564	92.46	1227	1135	92.5
5	194	164	84.54	186	162	87.1	380	326	85.79
11	207	182	87.92	200	181	90.5	407	363	89.19
15	297	281	94.61	284	264	92.96	581	545	93.8
18	142	112	78.87	145	131	90.34	287	243	84.67
21	247	229	92.71	209	184	88.04	456	413	90.57
23	132	121	91.67	124	112	90.32	256	233	91.02
28	158	132	83.54	149	124	83.22	307	256	83.39
32	129	118	91.47	114	109	95.61	243	227	93.42

Manual counts were considered as the true data to calculate the accuracy of the automated counting. The minimum and maximum accuracy of entry and exit counts were found to be 76.51 percent and 95.66 percent. The minimum and maximum total accuracy of individual sites were found to be 78.81 percent and 95.14 percent. The estimated accuracy of all sites was 89.57 percent.

A two-tailed paired t-test was performed to evaluate the similarity between manual and automated counts. When it is required to know the similarity between two variables of the same subject, a paired t-test is conducted. In this research, manual and automated counting was performed on 10 sites. So, manual and automated counts can be considered as two variables, and a site can be considered as the same subject of interest. It is required to know the difference in the observations of manual and automated counts which can be greater, smaller, or equal to zero. So, a two-tailed paired t-test was selected to perform in this study.

A few assumptions were considered for this test. It was assumed that the independent variable (i.e., a site) consists of two related groups (i.e., manual and automated counts), there are no significant outliers in the differences between manual and automated counts, and the distribution of differences between manual and automated counts shows an approximate normal distribution.

The following hypotheses were considered:

$$H_0: N_{ai} = N_{mi} \forall\, i$$
$$H_A: N_{ai} \neq N_{mi} \forall\, i$$
$$where,$$

$$N_{ai} = automated\ total\ daily\ count\ at\ site\ i$$

$$N_{mi} = manual\ total\ daily\ count\ at\ site\ i$$

A paired t-test was conducted for two cases for day 1 and day 2 individually. In the first case, the t-test was conducted for the difference between manual and automated counts. However, since there is a consistent trend to undercount, a second case was considered where the mean difference is subtracted from the difference. That is, in the second case, the test was conducted for the adjusted difference between manual and automated counts, where the adjustment was performed by the mean of the difference (manual – automated + mean). In effect, this is recognizing the consistent tendency for the automated counts to be undercounted and adjusting for it. The test results are shown in Table 20, Table 21, Table 22, and Table 23.

Table 20. Paired t-test for manual minus automated counts for day 1

Mean	44.10	
Standard deviation	18.91	
Standard Error of the Estimate (S.E.E.)	5.98	
t-statistic	7.37	
95% Confidence Interval	30.57	57.63

Table 21. Paired t-test adjusted for the mean for day 1

Mean	-0.60	
Standard deviation	18.91	
Standard Error of the Estimate (S.E.E.)	5.98	
t-statistic	-0.10	
95% Confidence Interval	-14.13	12.93

Table 22. Paired t-test for manual minus automated counts for day 2

Mean	44.70	
Standard deviation	20.38	
Standard Error of the Estimate (S.E.E.)	6.44	
t-statistic	6.94	
95% Confidence Interval	30.12	59.28

Table 23. Paired t-test adjusted for the mean for day 2

Mean	-2.84E-15	
Standard deviation	20.38	
Standard Error of the Estimate (S.E.E.)	6.44	
t-statistic	-4.41E-16	
95% Confidence Interval	-14.58	14.58

In the case of the difference between manual and automated counts as shown in Table 20 and Table 22, the null hypothesis is rejected for both day 1 and day 2. So, in this case, it is observed that the manual and automated counts are significantly different from each other. However, in the case of the adjusted difference between manual and automated counts, i.e., Table 21 and Table 23, the null hypothesis could not be rejected. So, when the tendency of automated counts to be undercounted is taken into effect, no discernible difference between manual counts and automated counts was observed in this study.

One of the assumptions of the two tailed paired t-test is that data shows a normal distribution so a normality test was conducted on the data. In this case, the manual and automated counting data for day 1 and day 2 were analyzed separately. It was found that only the data for manual counting from day 1 showed a normal distribution, while all other data did not. However, looking at the results above, evidence to reject the null hypothesis for the data in Table 20 and Table 22, and lack of evidence to reject it for data in Table 21 and Table 23, is so strong in each case that the failure of the data to display a normal distribution is unlikely to disqualify the finding that while a significant difference exists in the raw counts, if the consistent trend of video to undercount is taken into account, no significant difference exists.

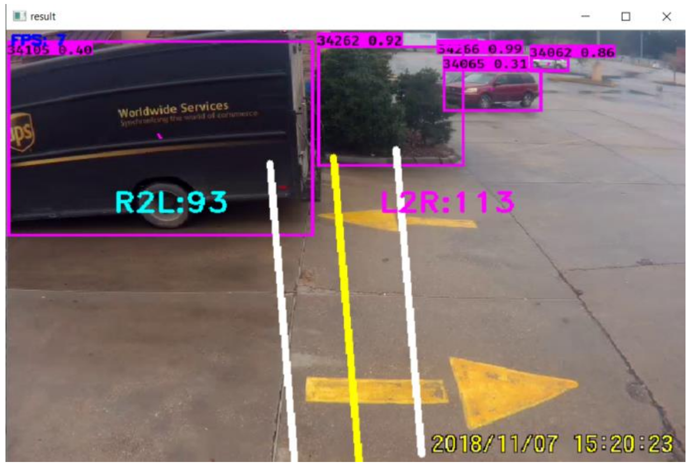
Potential Reasons for the Error

The potential reasons for error in automated counts are as follows. In the case of a parking lot, the arrival speed of a vehicle is generally higher than the departure speed. Figure 20 explains the scenario clearly. Here, vehicles leave the major road and take a right turn to enter the entrance of the parking lot, where they do not need to wait for any traffic signal or queue. But when vehicles depart from the parking lot, they have to wait for a gap in the major road, which causes a lower speed for departing vehicles. This program sometimes fails to count vehicles that have a high speed. The reason behind this is when vehicles arrive at high speed, the program does not get time to count because it appears in the frame for too short a time. For this reason, the error for arriving vehicles is higher than for departing vehicles.

Figure 20. Arrival and departure speed comparison

The camera angle is a major reason for reducing the accuracy of counts. When the camera is close to the entrance, i.e., the vehicle appears large and covers most of the frame, the program fails to count those vehicles. Moreover, when vehicles appear large, some parts of the vehicle are out of the frame, which makes it difficult for the program to detect the image as a vehicle. In Figure 21, the camera is very close to the entrance, and it does not cover the whole view of the entrance. As a result, the UPS vehicle appears large and some parts of the vehicle are out of the frame. Although most of the time the program can detect and count large vehicles, it is not the ideal view of the frame for automated counting.

Figure 21. Unsuitable camera angle and view



Visibility is an important factor that controls the quality of the video. Low visibility results in poor video quality. Rain, low light, evening recording, and cloudy weather causes low visibility. Raindrops obscure the camera lenses and result in a bad quality of video recording. In this case, the program cannot count accurately. When two vehicles arrive and depart at a time, one vehicle overlaps another. In this case, the program cannot detect the overlapped vehicle and counts only one vehicle.

Bluetooth and Wi-Fi Vehicle Detection

Field Experience

TrafficBox data can be exported in Excel format (CVS file) and include information on fingerprint (MAC Address records), detection type (Wi-Fi or Bluetooth), Received Signal Strength Indicator (RSSI) value, and detection date and time. An example of the output file from a TrafficBox is shown in Figure 22. Create-time is the recorded time in seconds from a reference date. According to the TrafficBox manual and our experiments with TrafficBoxes, the detection zone can vary depending on the antenna type, indoor versus outdoor environments, surrounding objects, and the signal strength of target devices.

Figure 22. TrafficBox recorded data

fingerprint	type	rssi	create_time
9.46269E+11	2	-91	1536757203
b827ebc27a10	2	-79	1536757205
d6ca8b	3	-76	1536757205
d428d556e932	2	-85	1536757207
9.46269E+11	2	-91	1536757208
9.46269E+11	2	-91	1536757211
d6ca8b	3	-79	1536757212
cc724d	3	-61	1536757212
cc724d	3	-79	1536757216
9.46269E+11	2	-91	1536757217
9.46269E+11	2	-91	1536757220

To expedite the data collection phase in a reasonable period of the project, four TrafficBoxes were purchased from the SMATS Company and were labeled TB#1 to TB#4. The first pair of TrafficBoxes (TB#1 and TB#2) were purchased a few months before the second pair. Five second Ignore MAC Interval was set on the TrafficBoxes to save memory space and battery life by not recording repeated MAC Addresses during the next five seconds after the first detection of a unique address (also referred to as a fingerprint). All the TrafficBoxes had a battery life of about 31 hours.

Correlation Between TrafficBox Detections and Vehicle Counts

A random sample of 6 strip malls were selected to compare the number of Bluetooth and Wi-Fi detections with manual counts of vehicles at each site. The count of detections and vehicle counts were made at the hourly level over the two days of observation at each site, resulting in up to 20 pairs of counts at each site if counts were conducted for the full 10 hours on each day. Table 24 shows the results of the comparison. As can be seen, correlation coefficients vary between 0.62 and 0.97 for the “combined” category representing total vehicle trip ends at each site. The detection ratio (i.e., the ratio of total Wi-Fi and Bluetooth detections divided by total vehicle trips) are the lowest at sites A (0.95) and E (0.81) causing 5.6 percent and 19.1 percent underestimation by the method. However, the detection ratio ranges between 1.5 and 2.87 in the remaining sites, leading to 49.67 percent to 187 percent overestimation. Also, the detection ratio for “in” and “out” categories are close, which suggests that almost the same number of entering devices remained detectable when leaving the site. Except in the first hour of the detection period, the detection ratio remained almost steady by time of day at each site. In addition, the correlation coefficients are always higher in the "out" category compared with the "in" category, suggesting the method is a very strong predictor of leaving vehicles (other than for site F). The correlation coefficients are above 0.81 compared to the entries (ranging from 0.43 to 0.89).

Table 24. Correlation coefficient and detection ratio of the random sample

	Correlation Coefficient			Detection Ratio (Automated/Manual)			Over/ underestimates
Sites	In	Out	Combined	In	Out	Combined	
A	0.89	0.94	0.97	0.91	0.98	0.95	-5.60%
B	0.8	0.93	0.94	2.67	3.11	2.87	187%
C	0.72	0.87	0.92	1.41	1.6	1.5	49.67
D	0.67	0.81	0.76	2.1	1.94	2.02	101.80%
E	0.49	0.81	0.75	0.76	0.86	0.81	-19.10%
F	0.43	0.59	0.62	1.97	1.63	1.79	79.33%

In Table 24, the correlation coefficients for combined in and out movements at the first three sites (i.e., A, B, and C) are high (above 0.92), but the other three sites (i.e., D, E, and F) have only moderate correlation coefficients (0.62 to 0.76). An attempt was made to identify which factors affect the quality of the predictions, by looking for factors that differed between the two groups of sites. Built-environment (BE) factors were investigated as a possible influence. The influence of variations in population density, land-use diversity, traffic volume on the adjacent road, and business types at each site were investigated. Residential density was determined from dividing the residential population by the size of catchment area around the strip malls (expressed in population within a 0.5-mile radius of the site). The jobs to resident workers ratio (JWR meaning the ratio of number of jobs to number of resident workers within a 0.5-mile radius of the site) was used as a measure of land use diversity. HPMS ADT 2016 counts were used as a measure of traffic volume on adjacent streets to the sites. Business types of the sites were collected form Google Popular Times data. The results are shown in Table 25.

Analyzing the results in Table 25, residential density is very high in the first group of sites (above 1400 people within a 0.5- mile radius of the sites) except site A. However, the second group has much lower residential density (below 720 people within a 0.5-mile radius of the sites). As expected, land use diversity is much lower in the first group compared to the second group (except site F). Because the detection zones are dynamic and may increase to larger radii for some devices with stronger signals, the joint detection area may include visits to surrounding land uses, which becomes problematic for sites with high land use diversity. It seems that the lower correlation coefficients of the second group may be due to higher land use diversity. However, there are two exceptions. Site A with low residential density and relatively high land use diversity had the best estimates in the first group. Site F with low residential density and low land use diversity had the least accuracy in the second group. These show that there must be other factors affecting the quality of the estimates beyond residential density and land use diversity around the sites.

Table 25. Built environment characteristics of the random sample

Sites	Manual			Automated			Combined R	Population Density	JWR	HPMS ADT 2016
	In	Out	combined	In	Out	combined				
A	358	304	662	282	261	543	0.97	666	1.65	15,200
B	154	132	286	411	410	821	0.94	1485.29	0.31	3,500
C	159	141	300	224	225	449	0.92	1752.52	0.23	24,992
D	354	346	700	628	501	1129	0.76	410.11	5.89	4,200
E	201	176	377	153	152	305	0.75	440.86	7.62	26,200
F	87	92	179	90	82	172	0.62	715.8	0.51	17,300

By increasing the traffic volume on the adjacent streets that are in view of the scanners, it may be possible to miss detecting some devices visiting the site because the capacity of the scanners may be exceeded due to a high volume of passing vehicles, which may affect the quality of the predictions. It was thought that the 5-second ignore MAC interval setting could worsen the situation at busy sites. However, this was found not to be true. As can be seen from Table 25, site F with the least vehicle trip ends (179) and relatively low traffic volume (17,300 vehicles per day) has the worst performance, while site A with the second highest vehicle trip ends (662) and almost the same level of traffic volume (15,200 vehicles per day) has the highest coefficient of correlation. Also, Table 25 shows the same level of traffic volumes in both groups, showing that traffic volume seems to have no effect on model performance by itself.

The type of businesses in the sites is another factor that may affect the results. If a site includes businesses that have more or less the same visiting time, the method performs the best. Here, we found that 75 minutes is the best threshold for a visit at the strip mall land-use category that is consistent with the conventional business types of this land-use category. For example, site A has eight businesses including two boutiques, two restaurants, two beauty salons, a coffee shop, and an eye care center. Site B includes a restaurant, two bars, a beauty salon, a coffee shop, and a vaporizer store. Site C has a real estate agency, beauty salon, and restaurants. Since the duration of visits of these businesses are close to each other, they could be captured well by the 75-minute threshold. However, if businesses with much higher and inconsistent durations of visit time are located at a strip mall, such as physical therapy at site F, and a healthcare clinic, art school, and dance academy at site E, the model performance drops. Site D includes a beauty salon, a postal store, a liquor store, and a cell phone store with less than 75 minutes visiting duration, so it seems the only reason for having 0.76 correlation coefficient value is due to the extension of the joint detection area and the inference from its surrounding land uses.

Another issue here is that the method identifies a MAC address with less than a 75-minute gap in its consecutive detection as a visit. It is unable to screen out vehicles passing by the desired site more than twice under the 75-minute threshold, unless we deploy two additional TBs beside the road before and after the land use. This way MAC addresses that passed the site more than once within less than the 75-minute interval will be screened out.

Regression Analysis

Regression analysis was conducted to determine local trip rates that include the impact of floor area, density, diversity, and traffic. Daily and peak-hour manual counts at the 40 survey sites serve as the dependent variable in this analysis, while floor area, density, diversity, and traffic, and floor area at each site are the independent variables. The purpose of the analysis is to determine whether the built environment factors such as density, diversity, and traffic have a significant impact on trip rates, and to develop locally-estimated trip rates that can be used to modify ITE trip rates.

Data

The data used in the regression analysis is shown in Table 26.

Table 26. Contextual data of survey sites

Site number	Floor area (1,000 sq.ft.)	Road Density	Worker + res.den (1000)	JWR
1	17.33	0.98	5.10	13.30
2	27.78	0.00	1.88	13.60
3	10.28	0.53	5.58	2.10
4	14.77	1.33	8.03	16.70
5	9.77	0.00	3.42	8.41
6	6.92	1.00	0.44	4.81
7	18.42	2.07	2.48	3.58
8	19.25	2.56	1.73	18.60
9	6.37	0.16	0.54	1.38
10	20.28	1.57	1.82	1.65
11	14.55	0.99	1.91	2.07
12	24.89	2.51	1.94	16.70
13	15.67	0.89	3.28	36.90
14	22.04	1.92	1.86	7.62
15	5.81	2.08	2.16	5.89
16	7.08	1.41	1.07	0.50
17	13.93	1.45	1.06	0.40
18	7.99	0.00	1.13	0.31
19	20.22	2.44	1.67	0.83
20	10.00	1.19	0.85	0.38
21	18.68	0.00	4.91	1.10
22	10.97	3.90	2.71	0.41
23	24.08	0.77	2.06	0.97
24	12.05	2.50	4.30	0.14
25	12.85	0.00	5.59	0.54
26	6.30	0.00	7.54	1.06
27	24.62	2.03	0.79	0.45
28	8.31	1.03	1.75	0.94
29	13.09	1.98	0.47	0.16
30	7.01	1.45	1.20	0.20
31	7.62	1.30	1.39	0.20
32	8.74	1.74	1.99	0.23
33	10.48	0.00	4.96	0.50
34	38.85	1.37	4.69	0.68
35	11.10	1.53	1.72	0.31

Site number	Floor area (1,000 sq.ft.)	Road Density	Worker + res.den (1000)	JWR
36	10.12	0.00	5.08	2.46
37	26.86	2.48	5.18	8.25
38	10.56	0.90	10.58	2.99
39	22.63	0.51	2.13	3.87
40	23.69	0.00	3.31	1.82

Correlation Analysis

Table 27 shows Pearson correlation between independent variables as well as the dependent variable. According to the table, whole day total trips have a moderately positive relationship with gross floor area (0.39), JWR (0.38), local connecting road density (0.38), and combined worker and residential density (0.22), where the numbers in parentheses are the correlation values. Also, these four variables are not strongly correlated with each other and, therefore, are good candidate variables for inclusion in the regression model. Median income is strongly correlated with residential density and is not strongly correlated with daily trips. Therefore, it was excluded from the regression equation estimated in the next section.

Table 27. Correlation of selected BE measures for whole day data

	Average Whole Day Trips	Area (Sq. ft.)	Residential Density	Combined Worker and Residential Density	JWR	2016 DOTD AADT	2016 HPMS AADT	Road Density	Controlled Access Highway Density	Major Connecting Road Density	Local Connecting Road Density	Local Road Density	Ramp Density
Average Whole Day Trips	1												
Area	0.39	1											
Residential Density	0.03	-0.07	1										
Combined Worker and Residential Density	0.22	0.01	0.79	1									
JWR	0.38	0.24	-0.3	0.11	1								
2016 DOTD AADT	0.09	0.17	0.04	0.16	0.3	1							

	Average Whole Day Trips	Area (Sq. ft.)	Residential Density	Combined Worker and Residential Density	JWR	2016 DOTD AADT	2016 HPMS AADT	Road Density	Controlled Access Highway Density	Major Connecting Road Density	Local Connecting Road Density	Local Road Density	Ramp Density
2016 HPMS AADT	-0.07	0.04	0.08	0.17	0.12	0.2	1						
Road Density	-0.02	-0.25	0.66	0.69	-0.06	0.05	-0.04	1					
Controlled Access Highway Density	-0.07	-0.04	-0.22	-0.07	0.37	0.03	-0.1	0.21	1				
Major Connecting Road Density	-0.06	0.2	0.06	0.19	0.07	0.08	0.24	0.19	-0.17	1			
Local connecting road density	0.38	0.14	-0.32	-0.27	0.05	0.07	-0.22	-0.13	0.01	-0.34	1		
Local Road Density	-0.11	-0.35	0.83	0.75	-0.32	-0.02	0.05	0.86	-0.19	0.14	-0.34	1	
Ramp Density	-0.01	-0.07	-0.18	-0.07	0.35	0.03	-0.12	0.31	0.89	-0.2	0.2	-0.1	1

A similar situation was found among the PM peak hour trips. Table 28 shows that local connecting road density, area, and JWR with 0.42, 0.38, and 0.25 correlation values with PM peak hour trips are good candidate variables for inclusion in a regression model estimating PM peak hour trips. The candidate variables also have low correlation with each other (< 0.23) thus minimizing multicollinearity in the regression model. All the variables are defined previously.

Table 28. Correlation of selected BE measures for PM peak hour data

	Average Whole Day Trips	Area (Sq. ft.)	Residential Density	Combined Worker and Residential Density	JWR	2016 DOTD AADT	2016 HPMS AADT	Road Density	Controlled Access Highway Density	Major Connecting Road Density	Local Connecting Road Density	Local Road Density	Ramp Density
Average Whole Day Trips	1.00												
Area	0.38	1.00											
Residential Density	-0.03	-0.07	1.00										
Combined Worker and Residential Density	0.05	0.02	0.78	1.00									
JWR	0.25	0.23	-0.30	0.11	1.00								
2016 DOTD AADT	0.07	0.17	0.05	0.18	0.30	1.00							
2016 HPMS AADT	0.20	0.16	-0.03	0.21	0.28	0.36	1.00						
Road Density	-0.16	-0.24	0.66	0.69	-0.06	0.07	-0.03	1.00					
Controlled Access Highway Density	-0.16	-0.04	-0.22	-0.07	0.37	0.02	0.08	0.20	1.00				
Major Connecting Road Density	-0.16	0.22	0.03	0.17	0.07	0.09	0.17	0.17	-0.17	1.00			
Local connecting road density	0.42	0.15	-0.34	-0.28	0.06	0.04	-0.03	-0.15	0.02	-0.34	1.00		
Local Road Density	-0.19	-0.34	0.84	0.75	-0.33	0.00	-0.07	0.86	-0.19	0.11	-0.35	1.00	
Ramp Density	-0.12	-0.07	-0.19	-0.07	0.36	0.02	0.05	0.31	0.89	-0.19	0.20	-0.11	1.00

Model Estimation

Multiple linear regression analysis was conducted on whole day (8 am-6 pm) and PM peak hour (the hour in which the maximum number of trips to the strip mall occurred) survey data. In the analysis, variables measuring area, density, diversity, and traffic are used as independent variables. Since different variables can be used in this role, the most important ones are evaluated here and the best combination identified. Two density measures (i.e., residential density and combined worker and residential density), a diversity measure (i.e., jobs to resident workers ratio), and five design measures (i.e., DOTD 2016 ADT, HPMS 2016 ADT, road density, road density by functional classes, and local connecting road density) resulted in ten different combinations of the BE factors. To compare the performance of different BE combinations, goodness of fit measures such as The Multiple Coefficient of Determination (R^2), standard error (SE), mean square error (MSE), model significance level (Significance F), and significance level of independent variables (t statistic and p-value) were used. From the estimation results of different combinations of independent variables shown in Table 29, the model that had the best R^2, standard error, mean square error, and significance level for the whole day analysis was the last model shown in the table using combined workers and residential density, JWR, and local connecting road density as the BE factors. Moreover, this combination of the BE factors has the highest significance level of the coefficients of the independent variables in a regression model. Table 30 shows the resulting whole day contextual trip generation model for the strip malls.

Table 29. Whole day model comparison of different BE combinations

Population Density		Land Use Diversity	Design (Network Characteristics)					R^2	SE	MSE	Significance F
Residential Density	Workers & Residential Density	JWR	LA DOTD 2016 AADT	HPMS 2016 AADT	Road Density	Road Density by Functional Class	Local Connecting Road Density				
■		■	■					0.28	238.8	57039.5	1.20E-04
■		■		■				0.30	235.4	55421.6	4.85E-05
■		■			■			0.30	235.4	55421.6	1.40E-04
■		■				■		0.28	239.6	57415.6	6.20E-06
■		■					■	0.42	212.8	45295.4	1.25E-07
	■	■	■					0.41	217.4	47282.1	7.58E-05
	■	■		■				0.30	237.1	56203.2	2.60E-05
	■	■			■			0.32	233.1	54351.4	8.40E-05
	■	■				■		0.29	237.4	56383.9	3.90E-06
	■	■					■	0.43	211.1	44584.9	0.74E-07

Table 30. Estimation of whole day regression equation

Regression Statistics	
Multiple R	0.66
R Square	0.43
Adjusted R Square	0.40
Standard Error	211.15
Observations	73

Table 31. Coefficients of whole day regression equation

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	67.77	73.47	0.92	0.36	-78.84	214.38
Area (1,000 sq. ft.)	10.25	3.40	3.01	0.00	3.47	17.04
Combined Worker & Residential Density (1,000 persons)	33.78	11.15	3.03	0.00	11.53	56.03
JWR Ratio	9.21	3.41	2.70	0.01	2.41	16.00
Local Connecting Road Density	115.25	27.61	4.17	0.00	60.15	170.35

The coefficients of the independent variables in the model are all statistically significant at the 95 percent level of significance (t-statistic above 2) with the exception of the constant. The model explains 43 percent of whole day trip variations in the data. Equation 4 shows the trip generation model for whole day analysis.

$$T = 67.77 + 10.25x_1 + 33.78x_2 + 9.21x_3 + 115.25x_4 \qquad (4)$$

where,

T = estimated trip ends per day at a strip mall with characteristics $x_1 - x_4$

x_1= gross floor area (in units of 1,000 square feet)

x_2 = combined worker and residential density (in 1000's of persons within 0.5-mile radius of the site)

x_3 = JWR ratio (ratio of number of jobs to number of resident workers within 0.5-mile radius of the site)

x_4 = local connecting road density (measured as miles of road within 0.5-mile radius of the site)

Likewise, Table 32 shows the results of different combinations of independent variables for the PM peak hour. Considering all the factors (R^2, SE, MSE, etc.), Table 34 reports the best fit for the PM peak hour data being a model specification that includes area, residential density, JWR, and local connecting road density as the independent variables.

Table 32. PM peak hour model comparison of different BE combinations

Population Density		Land Use Diversity	Design (Network Characteristics)								
Residential Density	Workers & Residential Density	JWR	LA DOTD 2016 AADT	HPMS 2016 AADT	Road Density	Road Density by Functional Class	Local Connecting Road Density	R^2	SE	MSE	Sig. F
■		■	■					0.15	71.20	5069.98	0.02
■		■		■				0.16	71.01	5041.79	0.02
■		■			■			0.17	70.70	4998.56	0.01
■		■				■		0.39	62.40	3894.29	6.4E-05
■		■					■	0.33	32.98	1087.63	6.6E-06
	■	■	■					0.15	71.30	5084.22	0.02
	■	■		■				0.16	71.08	5052.53	0.02
	■	■			■			0.16	70.88	5023.27	0.02
	■	■				■		0.38	62.80	3943.33	9.01 E-05
	■	■					■	0.32	33.26	1106.54	1.2E-05

Table 33. Estimation of PM peak hour regression equation

Regression Statistics	
Multiple R	0.58
R Square	0.33
Adjusted R Square	0.30
Standard Error	32.98
Observations	76

Table 34. Coefficients of PM peak hour regression equation

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	12.04	11.78	1.02	0.31	-11.45	35.53
Area (1,000 Sq. ft.)	1.50	0.53	2.85	0.01	0.45	2.54
Residential Density (1,000 in 0.5 mile radius)	5.57	3.04	1.83	0.07	-0.48	11.62
JWR	1.13	0.55	2.04	0.05	0.03	2.23
Local Connecting Road Density	17.84	4.32	4.13	0.00	9.22	26.46

According to the PM peak hour regression outcomes, the p-value is much smaller than the standard rejection region of 0.05 and the large values of the test statistics show the statistical significance of the coefficients of the independent variables at the 95 percent level of significance, with the exception of the constant and residential density. The model explains 33 percent of PM peak hour trip variations in the data. Equation 5 shows the estimated trip generation model for the PM peak hour.

$$T = 12.04 + 1.50x_1 + 5.57x_2 + 1.13x_3 + 17.84x_4 \qquad (5)$$

where,

T = estimated trip ends per hour during the peak at a strip mall with characteristics $x_1 - x_4$

x_1= gross floor area (in units of 1,000 square feet of the site)

x_2 = residential density (in 1000's of populations within 0.5-mile radius of the site)

x_3 = JWR ratio (ratio of number of jobs to number of resident workers within 0.5-mile radius of the site)

x_4 = local connecting road density (measured as miles of road within 0.5-mile radius of the site)

Testing for Autocorrelation, Multicollinearity, and Heteroscedasticity

To ensure that the models described in the previous paragraphs abide by the Gaussian assumptions on which linear regression is based, the models are tested for autocorrelation, multicollinearity, and heteroscedasticity.

Autocorrelation

Autocorrelation or serial correlation refers to the existence of correlation between error terms of observations in a regression model. The Durbin-Watson (DW) statistic is often used to check for autocorrelation in linear regression models. The DW statistic varies between zero and 4 representing positive and negative correlation among consecutive observations at either end. A DW value of 2 means no autocorrelation exists in the data. As shown in Table 35 and Table 36, DW values of 2.40 and 2.35 for whole day and PM peak hour analyses show very low levels of autocorrelation in this data set.

Table 35. Autocorrelation results for whole day analysis

Ordinary Least Squares Estimates			
SSE	3080098.26	DFE	68
MSE	45296	Root MSE	212.83
SBC	1006.07	AIC	994.61
MAE	160.60	AICC	995.51
MAPE	47.15	HQC	999.18
Durbin-Watson	2.40	Total R-Square	0.42
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.40	0.96	0.043

(Note: Pr < DW is the p-value for testing positive autocorrelation, and Pr > DW is the p-value for testing negative autocorrelation.)

Table 36. Autocorrelation results for PM peak hour analysis

Ordinary Least Squares Estimates			
SSE	80588.898	DFE	71
MSE	1135	Root MSE	33.69
SBC	766.77	AIC	755.12
MAE	25.97	AICC	755.98
MAPE	65.39	HQC	759.78
Durbin-Watson	2.35	Total R-Square	0.306
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.35	0.93	0.0705

Multicollinearity

Multicollinearity refers to high correlation between independent variables included in the model. The Variance Inflation Factor (VIF) is often used to measure multicollinearity in regression analysis. The VIF measures the number of times the variance of the estimated dependent variable is increased by the existence of collinearity (correlation) among independent variables in the model. The VIF varies between 1 (indicating no collinearity among the independent variables) to infinity when all independent variables are collinear. Generally, if the VIF is above 10, multicollinearity in a linear regression model is considered excessive and remedial action must be taken [34]. Table 37 shows the collinearity diagnostics for the whole day and Table 38 for the PM peak hour data. As can be seen from the table, the VIF values are well below 10 for all independent variables included in the models, which means there is not a significant level of multicollinearity among the independent variables in the model.

Table 37. Multicollinearity test for whole day data

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	67.77	73.47	0.92	0.36	0
Area	1	10.25	3.40	3.01	0.00	1.08
Residential & Workers Density	1	33.78	11.15	3.03	0.00	1.10
JWR	1	9.21	3.41	2.70	0.01	1.07
Local Connecting Road Density	1	115.25	27.61	4.17	<.0001	1.11

Table 38. Multicollinearity test for PM peak hour data

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	12.04	11.78	1.02	0.31	0.00
Area	1	1.50	0.53	2.85	0.01	1.08
Residential Density	1	5.57	3.04	1.83	0.07	1.25
JWR	1	1.13	0.55	2.04	0.05	1.17
Local Connecting Road Density	1	17.84	4.32	4.13	<.0001	1.16

Heteroscedasticity

Heteroscedasticity refers to inconsistency in the variation of residuals across values of the independent variables and should be treated in an OLS regression model as it decreases the precision of the estimates. Two common tests used in testing for heteroscedasticity, namely the Breusch-Pagan and White tests, were applied to the whole day and PM peak hour data. The Breusch-Pagan test checks for linear forms of heteroscedasticity, while the White test is used to test non-linear forms of heteroscedasticity in the data. They test the null hypothesis that the variance of the error terms is invariant across values of the independent variables. As shown in Table 39 and Table 40, the p-value (the probability of falsely rejecting the null hypothesis) for all the test-statistics are above the alpha level of 0.05. Therefore, whole day and PM peak hour data meet the homoscedasticity condition of linear regression models.

Table 39. Heteroscedasticity test on whole day data

Heteroscedasticity Test				
Equation	Test	Statistic	DF	Pr>ChiSq
Trips	White's test	15.83	14	0.32
	Breusch-Pagan	3.03	4	0.55

Table 40. Heteroscedasticity test on PM peak hour data

Heteroscedasticity Test				
Equation	Test	Statistic	DF	Pr>ChiSq
Trips	White's test	23.92	14	0.05
	Breusch-Pagan	4.44	4	0.35

Using the data in Table 26 and the ITE and Contextual models for daily trips in equations (6) and (4) respectively, the performance of the models can be assessed by comparing their predictions against the ground counts. The results of such a comparison is shown in Table 41 where the overall performance of each model is measured in terms of the percent root mean square error of the predicted counts from the ground counts. The percent root mean square error of the ITE model is 111 percent versus 71 percent for the model that includes the contextual factors. Thus, adding the density, diversity, and traffic intensity surrounding a site to estimation of the trip generation of a strip mall in this study, reduced the average error of predictions from 111 percent to 71 percent, a 36 percent reduction in error. While this is a substantial reduction in error, the magnitude of error still remaining shows that there are significant remaining factors influencing trip generation that have not been captured in this analysis.

Table 41. Comparison of ITE model and model including contextual factors

Site number	Ground count	ITE model	Contextual daily model
1	567.50	779.04	653.12
2	412.00	1226.09	541.28
3	1140.50	477.44	442.06
4	593.50	669.52	797.51
5	390.00	455.62	360.90
6	490.50	333.70	313.11
7	349.00	825.67	611.89
8	1186.00	861.18	789.87
9	336.50	310.17	182.45
10	599.00	905.24	533.26
11	371.50	660.11	414.59
12	861.00	1102.45	831.51
13	713.00	708.02	781.61
14	391.00	980.53	647.97
15	640.50	286.21	494.25
16	199.00	340.54	343.59
17	250.00	633.59	417.16
18	302.00	379.47	190.69
19	690.00	902.67	620.29
20	348.50	465.46	339.63
21	429.00	836.79	435.23

Site number	Ground count	ITE model	Contextual daily model
22	656.50	506.96	725.01
23	345.00	1067.80	481.85
24	760.50	553.16	625.95
25	455.00	587.38	393.29
26	134.00	307.17	396.81
27	437.50	1090.90	584.91
28	306.50	393.16	339.43
29	243.00	597.65	447.49
30	637.00	337.55	349.11
31	143.00	363.64	344.50
32	271.50	411.56	427.23
33	235.00	485.99	347.34
34	923.50	1699.66	788.57
35	286.00	512.52	418.83
36	91.00	470.59	365.76
37	1002.00	1186.73	879.87
38	679.50	489.42	664.67
39	481.00	1005.77	466.10
40	666.50	1051.12	439.17
%RMSE		111	71

Adjustment Factors

The eighth edition of the ITE *Trip Generation Manual* estimates whole day trip generation at a strip mall from the following formula:

$$T = 37.66 + 42.78x_1 \tag{6}$$

where,

T = estimated trip ends per day at a strip mall with a gross floor area of x_1

x_1= gross floor area of strip mall (in units of 1,000 square feet)

Under the assumption that the gross floor area used in this study is the same as the area used by ITE, subtracting equation (6) from the equation which included built environment factors in the formulation (i.e., equation (4)), produces an adjustment factor for daily ITE trips. That is, ITE trip estimates can be adjusted to reflect local conditions, including built environment characteristics of the area, by applying this adjustment factor to the trip generation estimates from ITE. The adjustment factor for whole day trips is shown in equation 7.

$$AF_{day} = 30.11 - 32.53x_1 + 33.78x_2 + 9.21x_3 + 115.25x_4 \tag{7}$$

where,

AF_{day} = Adjustment factor for whole day trip estimates from ITE

x_1= gross floor area (in units of 1,000 square feet of the site)

x_2= combined worker and residential density (in 1000's of population within 0.5-mile radius of the site)

x_3= JWR ratio (ratio of number of jobs to number of resident workers within 0.5-mile radius of the site)

x_4= local connecting road density (measured as miles of road within 0.5-mile radius of the site)

The eighth edition of the ITE *Trip Generation Handbook* estimates afternoon peak hour trip generation equation from the following formula:

$$T = 21.48 + 2.40x_1 \tag{8}$$

where,

T = estimated peak hour trip ends at a strip mall with a gross floor area of x_1

x_1= gross floor area of strip mall (in units of 1000 square feet)

By subtracting equation 8 from the equation 5 for peak-hour trips, an adjustment factor for ITE peak-hour trips can be established in the same manner as established for whole day trips above. ITE trip estimates can be adjusted to reflect local conditions by applying the adjustment factor shown in equation 9.

$$AF_{peak\ hour} = -9.44 - 0.9x_1 + 5.57x_2 + 1.13x_3 + 17.84x_4 \tag{9}$$

where,

$AF_{peak\ hour}$ = Adjustment factor for peak hour trip estimates from ITE

x_1 = gross floor area (in units of 1,000 square feet of the site)

x_2 = residential density (in $1000's$ of population within 0.5-mile radius of the site)

x_3 = JWR ratio (ratio of number of jobs to number of resident workers within 0.5-mile radius of the site)

x_4 = local connecting road density (measured as miles of road within 0.5-mile radius of the site)

To get an idea of the magnitude of the adjustment factors above, consider the strip mall at site 1 which has a floor area of 17,330 square ft., a residential density of 1,500, jobs to worker ratio (JWR) of 13.30, and 2 miles of connecting road:

From equation (7): $AF_{day} = 30.11 - 32.53(17.33) + 33.78(1.5) + 9.21(13.3) + 115.25(2) = -130$

That is, the ITE trip rate must be reduced by 130 vehicles per day from its estimate of (37.66+42.78(17.33)) = 779 vehicles per day, according to the adjustment factor equation developed in this study. Considering the manual count for site 1 was 567.5 vehicles per day (see Table 41), reducing the ITE estimate of 779 to 649 is an improvement but it does not compensate entirely.

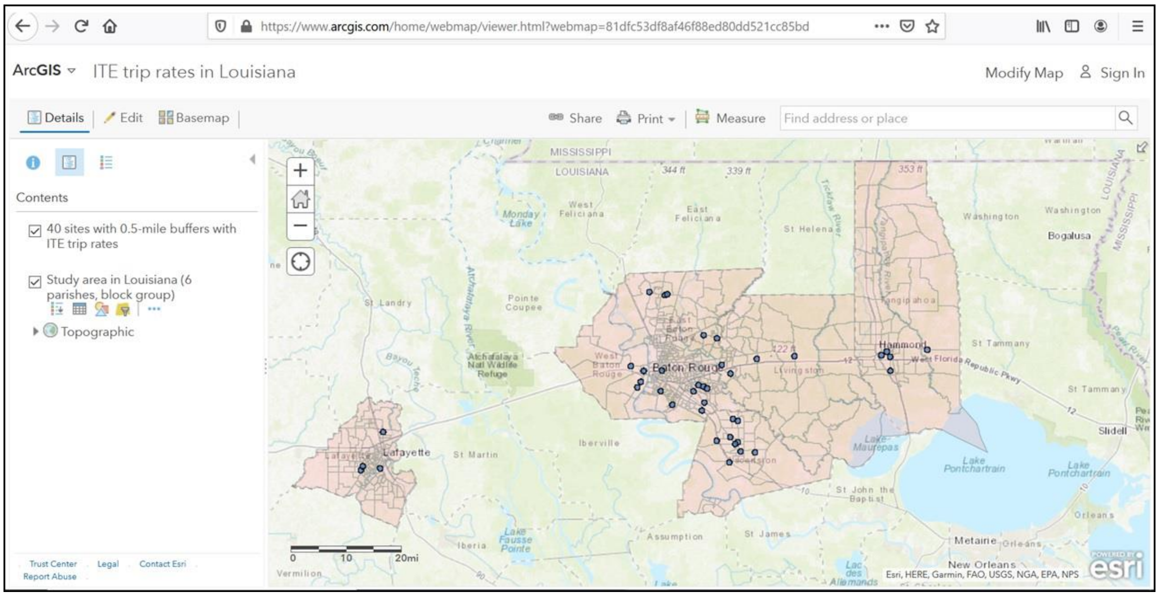
GIS System

The GIS system employed in this study is ArcGIS Online as described in the Methodology section. It allows access to a GIS application via the web to any user authorized by the developer of an application. Access is via the ESRI website as described in the Methodology.

Content

The following geospatial data have been added to the specific ArcGIS Online map prepared for this project: a data layer presenting a study area of six parishes in Louisiana and 40 sites with the 0.5-mile buffers with observed and estimated trip rates. The online map is shown in Figure 23. The side panel is the table of contents for the map. There are three entries in the Contents: the basemap ("Topographic"), 40 sites used in the project, and the study area of 6 parishes in the Louisiana data layer.

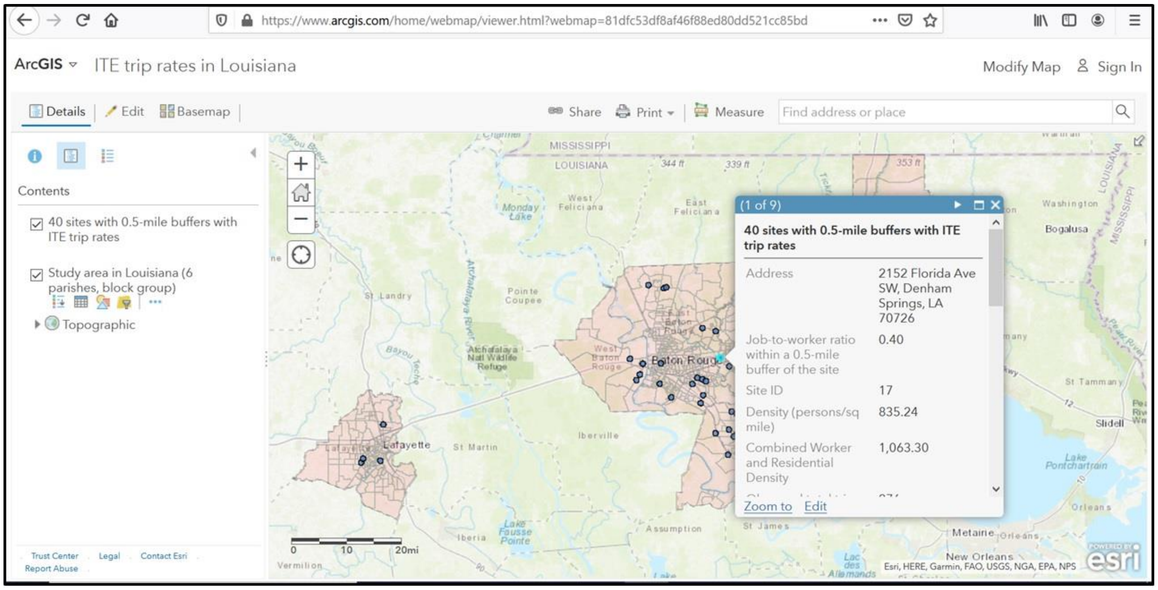
Figure 23. Map of study area



Pop-up Box

A pop-up window has been configured for the 40 sites with estimated and observed trip rates" data layer so that clicking on any site brings in the pop-up attributes associated with this site. Site attributes such as the address of the site, job-to-worker ratio, site ID, and population density, are shown in the pop-up window (see Figure 24).

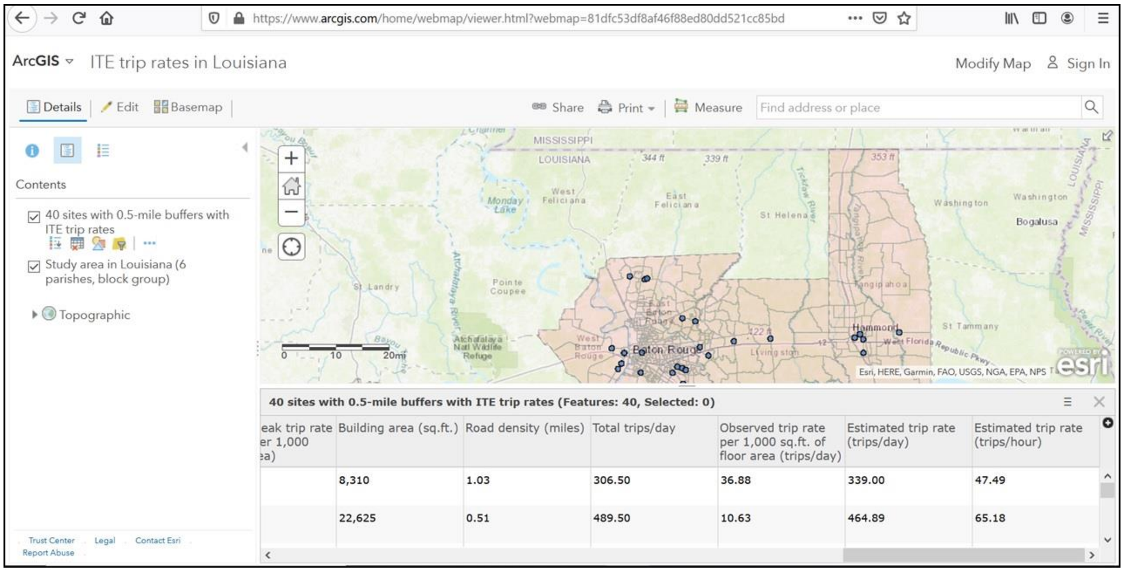
Figure 24. Pop-up box showing individual site features



Attributes

Attributes have been configured by selecting the fields to be displayed and giving more descriptive names. When a user clicks on the Show Table icon in the Contents, a table appears on the screen showing the attributes of all sites in detail. The table is closed by clicking on the X in the top right-hand corner of the table. The attribute table is shown in Figure 25.

Figure 25. Attribute table

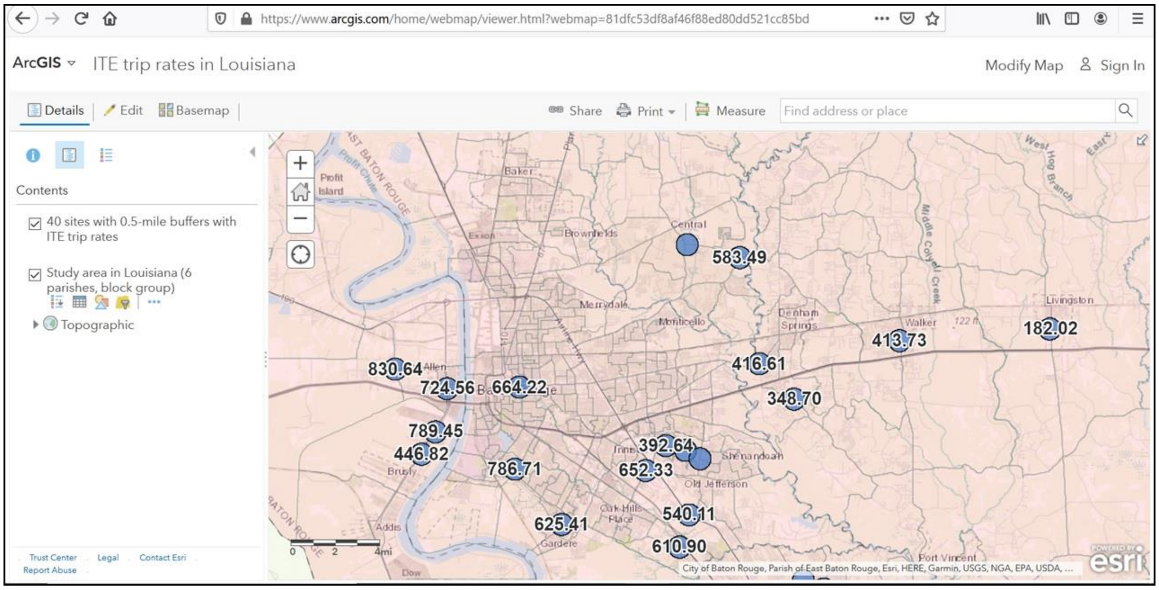


Labels

For this web map, we set a visible range for labels so the labels will not be visible unless a user zooms in to a certain point. A label showing estimated trip rates (trips/day) of each site appears on the map; the labels only show when a user/viewer zooms in further. There are several ways a user can zoom in: using the + button; by double-clicking, or by using the mouse scroll button. Labels are shown in Figure 26.

Zooming in to a metropolitan area (1:320,000 scale) brings in labels showing estimated trip rates (trips/day) of each site appearing on the map. In order to keep the map uncluttered and readable, the map presents only the data that is relevant to the user at a certain scale. The labels (showing estimated trip rates measured as trips/day) disappear when zooming further out.

Figure 26. Site labels



Conclusions

The objectives of this study were, first, to identify what factors beyond floor area influence trip generation at land use sites in Louisiana, and, second, to investigate alternative ways of collecting trip generation data in place of the current labor-intensive method of manual counting.

In addressing the first objective of the study, it was decided to count the trips to a sample of strip malls in Louisiana and then estimate the impact the contextual characteristics of these sites had on trip making. Contextual factors considered included residential density, land use diversity, traffic intensity, and household income in the area surrounding the sites. Household income was highly correlated with residential density and was omitted from further consideration. The results show that models of trip generation containing contextual factors reduce the percent root mean square error by 36 percent from that obtained by using the trip generation estimates from the ITE *Trip Generation Manual*. However, although the model containing contextual factors improves the estimates, it only explains about one half of the trip variation in the data. Determining other factors that influence trip generation rates should be the subject of future research.

The survey found that manual counts can have errors of approximately one percent. It was also found that the average trip rate observed in the survey was 40.82 trips/day/1,000 square ft. of gross leasable area compared to 42.78 in the ITE *Trip Generation Manual*. However, the observed trip rates at the 40 sites in the survey showed a standard deviation of 26.08 trips/day/1,000 square ft. of gross leasable area, indicating there are large variations in trip rates from site to site. As indicated in the previous paragraph, only some of that variation is captured by the contextual factors of density, diversity, and traffic intensity.

The possibility that the number of Wi-Fi and Bluetooth detections at a site may reflect the number of visits to a site was investigated in this study. The notion was that if there is a close correlation between trip ends and Wi-Fi and Bluetooth detections, then trip ends could be inferred by observing the number of Wi-Fi and Bluetooth detections. Because detections could take place automatically, the potential was there to automate the counting of visits to a site.

Analyzing Wi-Fi and Bluetooth data in six strip malls showed correlations of combined Wi-Fi and Bluetooth counts with manual vehicle counts, which varied from 0.92 to 0.97 in areas with low land use diversity to 0.62 to 0.76 in highly diverse areas. There are two potential reasons for this result. First, in highly diverse areas the joint detection area may increase due to the presence of devices with strong signals at surrounding land uses. Therefore, interference from other land uses will decrease the correlation between automated and actual counts. Second, the inability of the method to screen out vehicles passing by the desired site more than twice within the 75-minute threshold is more likely to occur in diverse areas because of the multiple opportunities to satisfy the purpose of a trip in that environment. This situation can be remedied by deploying two additional TrafficBoxes beside the road before and after the land use to identify such vehicles but was not implemented in this study. These two possible reasons for a change in the relationship between trip ends and Wi-Fi and Bluetooth detections should be the subject of future research.

The investigation of image processing of video recorded traffic movement on access to strip malls, produced promising results. Overall, the processing of video imagery to estimate the entry and exits of vehicles at land use sites produces estimates that are roughly 90 percent of the actual vehicle counts. Of great significance in this finding, however, is that this error is exclusively the result of undercounting. In fact, among all the observations of daily trip generation estimates at the 40 sites over two days, the range of estimates from video image detections ranged from 78.8 to 95.4 percent of the ground counts. Undercounting is due to poor quality images (e.g. adverse weather, poor light), vehicles passing through the observation area too quickly, or a vehicle in the closer lane obscuring the presence of a vehicle in the next lane. If observations in bad weather or poor light are excluded and an effort is made to limit the source of undercounting, much better accuracy could be achieved. It should also be noted that the Wi-Fi and Bluetooth detection method is not subject to these problems.

Overall, image processing of video data is the best way to replace manual counting with an automated process. However, it still requires that the camera be installed in the field and retrieved when observation is complete. With cameras with long battery life, or camera batteries supplemented with solar panels, observation periods could be extended making installation and retrieval a smaller portion of the total cost.

To be able to use secondary data to estimate trip generation, thereby eliminating the need for any fieldwork, would be a great improvement on current methods. One of the data sources encountered in this study that possibly has the potential to achieve that is the business profile data in Google Maps. It shows the number of customers by hour of the day seven days a week at individual businesses. Information is not available for all businesses and the diagram showing the customer numbers is not annotated so it is difficult to identify actual numbers. However, there is a rough scale and by looking at different businesses, the scale can be roughly interpreted. The type of each business is specified along with other information such as opening and closing times. Seeing this is a large dataset, the question is whether it could be used to estimate the trip generation of individual businesses together with the interaction they may impose on each other due to being in close proximity to each other. This topic is suggested for future research.

Adjustment factors to the ITE trip generation rates have been developed for daily and afternoon peak hour periods. They provide estimates that are, overall, 36 percent more accurate than the ITE estimate. The adjustment factors have been incorporated into a GIS system that uses census data to estimate the values of the contextual factors at a site, and then uses this information to adjust the ITE estimates to improved values.

Recommendations

The following recommendations are made based on the conclusions of this study:

1. Use video camera footage and image detection technology developed in this study to estimate trip generation rates of land uses for which ITE estimates appear suspect for conditions in Louisiana. Extend the period of observation to at least one week to maximize data collection, minimize installation and retrieval time, and observe fluctuation in traffic over the week. Keep a record of time spent in conducting this activity so that a cost comparison can be made between this method and the conventional method of estimating trip rates through manual counts.

2. Investigate whether secondary data sources exist which can serve as input data to a process that can be used to estimate trip production at any individual land use.

Acronyms, Abbreviations, and Symbols

Term	Description
AASHTO	American Association of State Highway and Transportation Officials
BE	Built environment
cm	centimeter(s)
DOTD	Louisiana Department of Transportation and Development
FHWA	Federal Highway Administration
ft.	foot (feet)
in.	inch(es)
ITE	Institute of Traffic Engineers
LTRC	Louisiana Transportation Research Center
lb.	pound(s)
M.	meter(s)
TB	TrafficBox
YOLO	You Only Look Once

References

[1] ITE, "Trip Generation, 8th edition," Institute of Transportation Engineers, Washington D.C., 2008.

[2] K. M. Currans, "Improving Vehicle Trip Generation Estimations for Urban Contexts: A Method Using Household Travel Surveys to Adjust ITE Trip Generation Rates," in *93rd Annual Meeting of the Transportation Research Board*, Washington, DC, 2013.

[3] K. J. Clifton, K. M. Currans and C. D. Muhs, "Adjusting ITE's Trip Generation Handbook For Urban Context," *Journal of Transport and Land Use,* vol. 8, no. 1, pp. 5-29, 2015.

[4] R. J. Schneider, K. Shafizadeh and S. L. Handy, "Method to Adjust Institute of Transportation Engineers Vehicle Trip-Generation Estimates in Smart-Growth Areas," *Journal of Transport and Land Use,* vol. 8(1), pp. 69-83, 2015.

[5] R. Ewing and R. Cervero, "Travel and the Built Environment: A Meta-Analysis," *Journal of the American planning association,* vol. 76(3), pp. 265-294, 2010.

[6] R. Cervero and K. Kockelman, "Travel Demand and the 3Ds: Density, Diversity, and Design," *Transportation Research Part D: Transport and Environment,* vol. 2(3), pp. 199-219, 1997.

[7] G. Tian, R. Ewing, A. White, S. Hamidi, J. Walters, J. P. Goates and A. Joyce, "Traffic Generated by Mixed-Use Developments: Thirteen-Region Study Using Consistent Measures of Built Environment," *Transportation Research Record,* vol. 2500(1), pp. 116-124, 2015.

[8] K. J. Clifton, K. M. Currans, A. C. Cutter and R. Schneider, "Household Travel Surveys in Context-Based Approach for Adjusting ITE Trip Generation Rates in Urban Contexts," *Transportation Research Record,* vol. 2307(1), pp. 108-119, 2012.

[9] R. Ewing and R. Cervero, "Travel and the Built Environment: A Synthesis," *Transportation research record,* vol. 1780(1), pp. 87-114, 2001.

[10] K. J. Clifton, K. M. Currans and C. D. Muhs., "Contextual Influences on Trip Generation," Transport Research and Education Center, Portland, Orlando, 2012.

[11] L. D. Frank, "Land Use and Transportation Interaction: Implications on Public Health and Quality of Life," *Journal of Planning Education and Research,* vol. 20(1), pp. 6-22, 2000.

[12] B. B. Brown, I. Yamada, K. R. Smith, C. D. Zick, L. Kowaleski-Jones and J. X. Fan, "Mixed Land Use and Walkability: Variations in Land Use Measures and Relationships with BMI, Overweight, and Obesity," *Health & place,* vol. 15(4), pp. 1130-1141.

[13] E. Avineri and J. N. Prashker, "The Impact of Travel Time Information on Travelers' Learning Under Uncertainty," *Transportation,* vol. 33(4), pp. 393-408, 2006.

[14] E. Leck, "The Impact of Urban Form on Travel Behavior: A Meta-Analysis," *Berkeley Planning Journal,* vol. 19, no. 1, pp. 37-58, 2011.

[15] J. Holtzclaw, R. Clear, H. Dittmar, D. Goldstein and P. Haas, "Location Efficiency: Neighborhood and Socio-Economic Characteristics Determine Auto Ownership and Use-Studies in Chicago, Los Angeles and San Francisco," *Transportation planning and technology,* vol. 25(1), pp. 1-27, 2002.

[16] S. Handy, K. R. Shafizadeh and R. Schneider, *California Smart-Growth Trip Generation Rates Study (No. CA13-1940),* Davis, California: Davis Urban Land Institute and Transportation Center, 2013.

[17] S. Handy, "Trip Generation: Introduction to the Special Section," *Journal of Transport and Land Use,* vol. 8(1), pp. 1-4, 2015.

[18] B. S. Bochner, K. M. Currans, S. P. Dock, K. J. Clifton, P. A. Gibson, D. K. Hardy, K. G. Hooper, L.-J. Kim, R. S. McCourt, D. R. Samdahl, G. H. Sokolow and L. F.

Tierney, "Advances in Urban Trip Generation Estimation," *Institute of Transportation Engineers. ITE Journal,* vol. 86(7), p. 17, 2016.

[19] R. Ewing, G. Tian, T. Lyons and K. Terzano, "Trip and Parking Generation at Transitoriented Developments: Five US case studies," *Landscape and Urban Planning,* vol. 160, pp. 69-78, 2017.

[20] M. Givechi and G. Sarkodee, "Establishment of Local Trip Generation Rates or Equations for Mixed-Use Developments in Kansas," Kansas Department of Transportation, Topeka, Kansas, 2012.

[21] R. Greene-Roesel, M. C. Diogenes, D. R. Ragland and L. A. Lindau, "Effectiveness of a Commercially Available Automated Pedestrian Counting Device in Urban Environments: Comparison with Manual Counts.," Safe Transportation Research & Education Center, Institute of Transportation Studies, Berkeley, California, 2008.

[22] "Trip Generation Rates for Transportation Impact Analyses of Infill Developments," Transportation Research Board of the National Academics, Washington, DC, 2013.

[23] R. Gudishala, C. Wilmot and A. Mokkapati, "Travel Time Estimation Using Bluetooth," Louisiana Transportation Research Center, Baton Rouge, Louisiana, 2016.

[24] Y. Malinovskiy, N. Saunier and Y. Wang, "Analysis of Pedestrian Travel with Static Bluetooth Sensors," *Transportation Research Record,* vol. 2299, no. 1, pp. 134-147, 2012.

[25] A. Kurkcu and K. Ozbay, "Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors," *Transportation Research Record,* vol. 2644, no. 1, pp. 72-82, 2017.

[26] A. Kalatian and B. Farooq, "Mobility Mode Detection Using WiFi Signals," in *2018 IEEE International Smart Cities Conference*, Kansas City, 2018.

[27] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[28] J. M. Daisa and T. Parker, "Trip Generation Rates for Urban Infill land uses in California," *ITE Journal,* vol. 79, no. 6, pp. 30-39, 2009.

[29] R. Cervero and G. B. Arrington, "Vehicle Trip Reduction Impacts of Transit-Oriented Housing," *Journal of Public Transportation,* vol. 11(3), p. 1, 2008.

[30] A. Burgi, P. Frei, G. Theis, E. Mohler and et.al., "A Model for Radiofrequency Electromagnetic Field Predictions at Outdoor and Indoor Locations in the Context of Epidemiological Research," *Bioelectronmagnetics,* vol. 31, no. 3, pp. 226-236, 2009.

[31] U. Bandara, M. Hasegawa, M. Inoue, H. Morikawa and T. Aoyama, "Design and Implementation of a Bluetooth Signal Strength Based Location Sensing System," *In Proceedings. 2004 IEEE Radio and Wireless Conference,* vol. IEEE Cat. No. 04TH8746, pp. 319-322, September 2004.

[32] Y. Wang, X. Yang, Y. Zhao, Y. Liu and L. Cuthbert, "Bluetooth Positioning Using RSSI and Triangulation Methods," *In 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC),* pp. 837-842, January 2013.

[33] M. E. Rida, F. Liu, Y. Jadi, A. A. A. Algawhari and A. Askourih, "Indoor Location Position Based on Bluetooth Signal Strength," in *2015 2nd International Conference on Information Science and Control Engineering, IEEE*, Shanghai, China, 2015.

[34] A. Alin, "Multicollinearity," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 2(3), p. 370–374, 2010.

This public document is published at a total cost of $200. 29 copies of this public document were published in this first printing at a cost of $200. The total cost of all printings of this document including reprints is $200. This document was published by Louisiana Transportation Research Center to report and publish research findings as required in R.S. 48:105. This material was duplicated in accordance with standards for printing by state agencies established pursuant to R.S. 43:31. Printing of this material was purchased in accordance with the provisions of Title 43 of the Louisiana Revised Statutes.