Modelling the Trip Length Distribution of Shopping Center Trips from GPS Data

[1]Nicolaas J Jonker Pr Eng, [2]Christoffel J Venter Ph.D Pr Eng

NJ Jonker: Centre for Transport Development and Department of Civil Engineering, University of Pretoria, Pretoria, South Africa
CJ Venter: Centre for Transport Development and Department of Civil Engineering, University of Pretoria, Pretoria, South Africa

Abstract

Automobile trip lengths are increasingly used in the calculation of development impact fees or bulk services contributions, affecting the revenue collected by local authorities. It is however difficult to obtain accurate estimates of current or predicted trip distances, and the empirical evidence base is relatively thin. Global Positioning System (GPS) technology might provide a more accurate way of filling this data gap at a lower cost. The paper describes the use of mobile GPS loggers to collect and analyze trip length data for car-based trips to and from shopping centers, based on data collected from drivers in the Pretoria-Johannesburg area of South Africa. We verify the minimum stopped time criterion used to identify trip ends under local conditions. Significant variation in trip lengths is observed, but average trip lengths vary systematically by shopping center type and size. Average GPS-derived trip lengths were found to be significantly longer than those estimated using conventional surveys in Florida, especially for smaller centers, raising the possibility that conventional methods lead to underestimation of the traffic impacts of individual centers. Although the study confirms the feasibility of using mobile GPS loggers for measuring trip lengths, several methodological questions remain to be solved to improve the robustness of the findings.

Citation:

Jonker NJ, and Venter C (2019), Modelling the Trip Length Distribution of Shopping Center Trips from GPS Data. *ASCE Journal of Transportation Engineering, Part A: Systems.* Volume 145, Issue 1. https://doi.org/10.1061/JTEPBS.0000200.

[1] Corresponding Author. E-mail address: nico@itsglobal.co.za

[2] E-mail address: Christo.Venter@up.ac.za

Introduction

Trip lengths are important for several transportation planning activities, including the estimation of trip distribution models during traffic forecasting, and monitoring of aggregate travel activity in a region. In addition, a growing number of cities are using trip distances travelled to or from a development in the calculation of development impact fees or bulk services contributions payable by the developer (COTO, 2013; City of Albuquerque, 2004). It is therefore important to obtain accurate estimates of current or predicted trip distances. Unfortunately this is difficult to achieve. The most common source of trip length data is travel surveys, but serious questions have been raised about the accuracy of self-reported travel times (Kelly et al., 2013) and distances (Murakami and Wagner, 1999). Alternatively trip lengths can be estimated by simulating routes between user-reported origin and destination points, but these may or may not correspond to the actual routes travelled (Timmermans and Rasouli, 2013).

The advances made in Global Positioning System (GPS) technology over the last decade have created new opportunities for collecting more accurate trip data from larger samples at lower cost (Stopher et al., 2008). Whether collected via personal smartphone apps or dedicated in-vehicle GPS loggers, GPS data are finding many applications in transportation planning and traffic management (NCHRP, 2014). As GPS devices become more widely used for data collection, better and more robust procedures are needed for cleaning, validating, and analyzing these large data sets (Shen and Stopher, 2014a). This paper contributes to this effort by describing the use of mobile GPS loggers to collect and analyze trip length data for car-based trips to and from shopping centers. The data was collected from a sample of 726 drivers in the Pretoria-Johannesburg area of South Africa. The paper describes the methodology developed for identifying trip ends, for calculating trip lengths, and for aggregating trips by shopping center type, followed by the results of a statistical analysis to model trip length distributions for different shopping center types. These results might be useful for estimating the traffic impacts of new shopping developments, while the methodology described might serve as a basis for updating and transferring of the results to other areas and other trip purposes.

The paper starts with a brief description of the relationship between shopping developments and trip distances, and the use of GPS in transport research. Then follows the methodology used for collecting, cleaning, and extracting the data; the analysis and results; and lastly conclusions on the applicability of the method and results to future studies.

Background

This section discusses the existing literature on available trip length data for shopping centers, the use of GPS loggers in travel surveys and methods used to identify trips in a GPS data set.

Shopping Centers and Shopping Center Trip Lengths

The research focuses on car-based trips to shopping centers in urban and suburban locations, as opposed to single, isolated shops or downtown-type complexes without parking. Shopping centers are the predominant shopping venue in lower density cities such as those in most of the United States and South Africa (Putnam, 2000). For the purposes of this research, a shopping center or shopping mall is defined as a complex of shops located within one or more buildings in close vicinity to each other, with parking provided on the property for its customers. We explicitly exclude mixed use developments with office or residential components. However, as many shopping centers also host restaurant or entertainment sites, the trip purpose of trips in our sample is not limited to *shopping* trips but may include personal business and entertainment. Shopping, personal business and medical/dental trips are the single most important trip purpose, comprising 42% of all trips according to the 2009 National Household Travel Survey (Santos et al., 2011).

The classification of shopping centers is important as it is commonly used for land use development and traffic planning purposes (Prinsloo, 2010). Shopping centers can be classified based on characteristics as shown in Table 1, and range from Convenience Centers with the smallest sizes (in terms of Gross Leasable Area (GLA)) and the fewest shops, to Regional and Superregional Centers with more than 100,000 m^2 of GLA and more than 250 individual stores. There is reason to expect a relationship to exist between the size or type of center and the length of trips made to and from the center. The primary reason is that larger centers tend to attract more specialist stores with larger catchment areas, which increases trip lengths as shown in Table 1. A further reason is related to the density and location of centers: small convenience and neighborhood centers are densely scattered within neighborhoods and therefore attract short local trips, while larger centers are located on larger properties (often next to freeways) where more longer-distance trips are attracted. Stover and Koepke (2002) add that roadway level of service (LOS) and city size are additional factors that would affect shopping trip lengths.

Unfortunately the evidence on shopping center trip lengths and their relationship with shopping center type and size is scarce. When doing site impact studies traffic engineers can turn to the Institute of Transportation Engineers (ITE) Trip Generation Manual (ITE, 2012) or its equivalent for detailed information on trip generation rates for more than 40 types of retail developments. For guidance on trip lengths, the best source appears to be survey data from several sites in Florida where trip lengths ranging from 2.6 km (1.6 miles) for small convenience stores, 5.4km (3.39 miles) for other shopping centers, and 9.5 km (5.9 miles) for discount superstores have been recorded (e.g. Citrus County, 2006).

However the transferability to other parts of the world of United States (US) derived traffic data in general (Fan and Lam, 1997) and of trip distances in particular (City of Albuquerque, 2004) is questionable, especially for larger metropolitan areas. That trip distance results can vary widely across different regions is confirmed by travel survey results from South Africa suggesting average shopping trip lengths of more than 14 km (COJ, 2008), which is much longer than in the US. In the Netherlands, by contrast, simulation studies have produced average travel distances to multi-store shops of only 5.5 km (Timmermans and Rasouli, 2013). Thus the need to either estimate adjustment factors when transferring results to new areas, or to undertake new studies on local trip characteristics. What is of concern is both the lack of comprehensive data, and the absence of standard procedures for verifying and updating trip distance information in new areas.

This is a serious problem as trip lengths are needed as an input during the determination of development impact fees, a revenue source increasingly being used by local authorities to help finance infrastructure development (Tindale, 1991). In South Africa, the national guidelines on traffic impact assessments (COTO, 2012) specify that engineering service contributions for road infrastructure for new developments be calculated on the basis of not only trip generation but also the vehicle-kilometers of travel on external roads provided by a road authority. An average trip length ranging between 2.7 and 6.3 kilometers is recommended for shopping centers (depending on the size of the shopping center) (COTO, 2013), a figure based on some local surveys as well as the above mentioned Florida surveys (Citrus County, 2006). The concern is that insufficient local verification of such figures has been conducted – a problem that might be present in other parts of the world seeking to implement similar approaches to impact fee calculation. This paper helps to address the gap by developing and testing a novel data collection method using on-board GPS devices.

Use of GPS Devices in Travel Surveys

During the last decade GPS devices have been applied in a number of transport fields, including traffic studies (e.g. Quiroga, 1999), travel behavior surveys (see Shen and Stopher (2014a) for an overview), and physical activity studies (e.g. Rodriguez et al., 2005). GPS equipment has the capability of recording precise data on an individual's travel movements over a period of time. GPS data can thus provide accurate information on the number of trips made, trip durations and distances, trip timings, and routes used (Bricka et al., 2012). With accurate route information comes the ability to determine the use of different classes of roads, which is useful for pricing and mobility management purposes (Venter and Joubert, 2014). This makes the GPS survey method very attractive as compared to conventional travel survey or diary data, both in terms of accuracy and completeness (Stopher et al., 2008). Since 2002 more than 25 household travel surveys conducted in the US have used GPS supplements to assess and correct trip underreporting or misreporting (NCHRP 2014).

GPS Devices

Two types of GPS devices are currently in use, namely "passive" and "active" GPS systems. A "passive GPS" system is switched on by the surveyor, given to the participant, and requires no further input from the participant. The position of the vehicle or person is recorded at pre-set intervals (for example every 3 seconds), and in many cases also the speed and heading. The "active GPS" system requires the participant to type in information about the trip, such as trip purpose, and can thus produce richer data. However, studies have shown that participants often find this tedious and skip the task near the end of the survey (Stopher et al., 2008). Smartphone-based GPS surveys are being used by researchers in both active and passive modes, and have the possibility of further reducing survey costs (Shen and Stopher, 2014a).

Problems with GPS Data

The GPS survey has some problems as well, including signal loss, device malfunction, and battery limitations. These can cause inaccurate or incomplete data acquisition (Shen and Stopher, 2014a). A second source of error is the data processing method used, as post-processing is needed to identify trip ends and trip characteristics, and sometimes to detect the mode used. Although the performance of the algorithms for conducting these tasks is improving, they are not always accurate, prompting Bricka et al. (2012) to warn against the use of GPS data as the sole source of travel survey data. Other challenges with GPS data are: cost, integration into existing modelling paradigms, privacy, sample bias, and data management (NCHRP 2014).

Identifying Trip Ends

The correct identification of trip ends from a large GPS dataset is a problem that has not yet been adequately solved. In-vehicle devices can identify trips from when the ignition is turned on and off. For independent GPS devices, the primary variable used to identify possible trip ends is the time gap when the device was stationary, also referred to as the stopped time or dwell time (Du and Aultman-Hall, 2007). The challenge is to select a suitable time period: if it is too short, stops at traffic signals or in moving queues could incorrectly be identified as trip ends; if it is too long, short stops at destinations could be missed (Du and Aultman-Hall, 2007). While a range of stopped times of between 45 and 300 seconds have been tested, most studies recommend 120 seconds to be used (Srinivasan et al., 2009). Shen and Stopher (2014a) feel however that the 120-second rule lacks empirical and/or theoretical research to support it, and recommend a value of 60 seconds based on their research.

There are some additional patterns in GPS data that could assist in identifying possible trip ends, including (Du and Aultman-Hall, 2007):

- Heading changes of approximately 180˚;
- Parking patterns;
- Repeated use of road links;
- Distance from the road network; and
- Circuitous routes.

Research has also been done on determining trip ends and tours for commercial vehicles (truck and delivery vehicles) GPS data sets (e.g. Holguín-Veras and Thorson, 2000; Greaves and Figliozzi, 2008; Joubert and Meintjes, 2015). Ma et al. (2011; 2016) used the following criteria to determine commercial vehicle trip ends:

- Dwell time of 3 min (180 seconds); and
- Engine status, off or in park.

Trip ends can be clustered using Ward's hierarchical clustering method and aggregating clusters whose geometric medians are within the boundaries of a single property (Sharman and Roorda, 2011).

Since the correct identification of trips is the basis for all subsequent trip-level analysis, it is clear that more work is needed in this regard. The present research examines the use of stopped time in conjunction with road type and vehicle movement pattern to determine appropriate stopped times for South African conditions, as described later on.

Imputing Trip Purpose

A substantial literature exists around the imputation of trip purposes from passive GPS data (Liu et al., 2013). Algorithms typically use information on the land uses around the trip end, limited to a 200m to 300m radius, to impute the purpose of the trip (Wolf et al., 2001; Stopher et al., 2007). In addition, respondents may be requested to provide the addresses of locations they often visit, such as workplace, home, schools and shopping centers. If an end point is located near these addresses, then the trip purpose is more easily identifiable (Clifford et al. 2008). Trips made to mixed land use developments present a particular challenge to identifying the trip purpose. Other approaches have added activity duration extracted from the GPS trace as an additional predictive variable (e.g. Wolf et al., 2001). Chen et al. (2010) and Axhausen et al. (2003) have used probabilistic approaches to evaluate various factors that could provide some indication of trip purpose, including time of day, history dependence, distance to the various land uses, and land use characteristics. Machine learning and artificial intelligence algorithms seem to present a promising approach to identifying predictive relationships (e.g. Liu et al., 2013; Reumers et al., 2014). Nevertheless, results remain mixed: Shen and Stopher (2014a) report that success rates for trip purpose identification range from as low as 43% to about 75%.

Calculating Trip Lengths from GPS Data

After identifying trip ends, two main methods are used for determining trip lengths from GPS data. These are (Srinivasan et al. 2009):

- The point-to-point method - sum of distances between points over the entire trip (PP); and
- The link-to-link method - sum of link lengths over the entire trip after matching the GPS points to road network links (LL).

Each of these methods has advantages and disadvantages. The LL method requires accurate road network data to which the GPS data may be linked, but could be more accurate if an accurate GIS database is available. The PP method is more sensitive to inaccuracies due to signal loss (Srinivasan et al., 2009), but is much faster and easier to calculate. Caution should be taken to correct any errors or missing GPS points when using the PP method. Nevertheless Murakami and Wagner (1999) showed trip lengths estimated from GPS surveys to be more accurate than those obtained from recall surveys as the latter tend to be over-estimated.

Ma et al. (2011) list a set of criteria for identifying abnormal trips that should be excluded from the trip data set:

- Extremely short trip;
- Trip travel time equal to zero;
- Trip with unpractical high speed;
- Trip with origin or destination external to study area; and
- Trip in which the vehicle left the study area and returned.

This discussion does not consider travel mode identification, as this is applicable only to multimodal studies while the present study collected data only on automobile trips.

Methodology

The methodology consisted of data collection and cleaning, trip end identification, extraction of shopping center trips, and calculation of trip lengths as discussed below.

Collection of GPS Data

The GPS data used for this study were obtained from a previous study undertaken at the University of Pretoria. The purpose of the original study was to evaluate the impact of a new open-road tolling scheme on freeway driving patterns in Gauteng Province. While Gauteng is the smallest province in South Africa, with an area of 18 000 km^2, it is largely urban and has the largest population of 13 million residents. It includes the commercial and administrative hubs of Johannesburg and Pretoria.

Study participants were selected using a stratified random sampling strategy and recruited through face-to-face home visits. The sample was slightly biased in favor of lower-income drivers (Venter and Joubert, 2013), and subsequently weighted in order to ensure the sample is statistically representative of car owners, both spatially and demographically. The final sample consisted of 726 participants, observed across 2962 travel days. Thus the sample was sufficiently large and representative to be used for examining shopping center trips as in the present study. The survey took place between November 2011 and March 2012 (excluding holidays).

Each respondent kept the GPS logger in their vehicles for three days of normal driving activity, including weekdays and weekends. Respondents were given a small gift certificate as an incentive to participate. At the conclusion respondents completed a short questionnaire on household, demographical information, work location and vehicle details.

The GPS device used for this survey, a Tracking Key 3100-INT model manufactured by LandAirSea Systems, Inc., is a passive device which records its location every second while the vehicle is moving. It can record up to 300 hours of data and records the position with a horizontal accuracy of 2.5 m. The battery life is between 5 and 8 days. The device has no buttons and the fact that it requires no interaction with the user makes it extremely robust and simple to use.

Data Cleaning

The GPS data was cleaned in order to remove anomalies caused by signal loss or ineligible trips. Records with average speeds exceeding 140 km/h between data points were removed, as well as trips travelling outside the study area boundaries.

Identification of Trip Ends

Two alternative criteria were used to identify probable trip ends in the GPS data set, namely stopped time and repeated use of road links. The procedure included validating the critical stopped time to be used for local conditions.

Validation of Stopped Time Criterion

The stop time criterion makes use of the number of seconds for which the vehicle is stationary to determine whether the stop is a trip end or not. Rather than use the standard 120 second value from the literature (which has not been validated or South African conditions), a procedure was developed to determine the most appropriate value taking also the road type and location of the stop into account.

The approach essentially acknowledged that the likelihood of a stopped signal of a certain length corresponding to an actual trip end depends on the location of the stop. For instance, a stopped time of 60 seconds is less likely to indicate a true trip end if it occurred at a congested intersection than if it occurred outside the road reserve (e.g. in a parking lot). Unfortunately we could not apply a stopped time criterion differentiated by road type during automated data processing, as reliable road network data is not available for the study area. We therefore manually validated the stopped time criterion by randomly selecting 10% of trip logs, identifying trip ends based only on different stop time values ranging between 45 and 600 seconds, and inspecting trajectories on Google Earth™. As ground truth for trip ends were not known, a probabilistic approach was taken in which identified trip ends were evaluated as either *probably correct*, *probably incorrect*, or *uncertain* based on the location of the stop by visual inspection (see Table 2).

The reasoning behind the adopted criteria is as follows:

- On Freeways, On-Ramps or Off-Ramps
    - Under normal driving conditions trips do not end on freeways or on ramps.
    - Thus all trip ends identified on freeways were marked as incorrect and are believed to be caused by traffic congestion.
- On Major Roads (Class 2 main roads and arterials)
    - It is considered unlikely that a trip end would be located on a major road, as these are subject to access management and do not allow on-street parking. However, in some cases (e.g. overflow parking within the road reserve when off-street lots are full) stops of longer duration might be possible, but they will be marked as uncertain.
    - To determine this longer stopped time it is observed that under saturated traffic conditions, traffic may be stopped at signalized intersections for up to two cycles. At a

typical cycle length of 150 seconds, this would equate to a stopped time of 300 seconds. Thus stops within the road reserve of major arterials of shorter than 300 seconds would be considered as incorrectly identified stops.

- On Secondary Roads Near an Intersection
    - To be considered near an intersection, a trip end had to occur within 100m of an intersection on a class 3 (secondary) road.
    - As traffic congestion on secondary roads tends to occur mainly at intersections, there is a high probability that short stop times signify congestion rather than true trip ends. Thus stop times of up to one cycle length (150 seconds) was identified as incorrect trip ends.
    - Cycle failure is less likely on class 3 roads. Thus long stops (300 seconds plus) are more likely correctly identified trip ends than caused by oversaturated signals.
    - Stops of between 150 and 300 seconds could be due to either congestion or true stops, and are classified as uncertain.
- On Secondary Roads Far From an Intersection
    - Very short stops are not likely to be trip ends as there is not enough time. Given likely walking distances from class 3 roads to adjacent properties and back, it was estimated that at least 100 seconds would be required to complete even a short duration activity. Thus trip ends shorter than 100 seconds are labelled incorrect.
    - Stops of between 100 and 300 seconds are less likely but not impossible. On street parking is not common but occurs sometimes. Thus these are labelled as uncertain.
    - Long stops (300 seconds plus) are assumed to be correctly identified trip ends, as before.
- On Minor Road and Streets
    - The minimum stop time of 100 seconds still applies, but due to shorter walking distances between on-street parking spaces and properties, there is a larger possibility that shorter stops could be actual stops. Thus shorter stops were labelled uncertain.

    - Low traffic congestion levels and low control delay occurs at junctions, thus medium and longer stop times are more likely to indicate true trip ends (as compared to higher order roads). Thus any stopped time longer than 100 seconds was labelled as correct.
- Within Property
    - All trip ends within a property was marked as correct.

The results of the analysis are shown in Figure 1. The optimal stop time is the stop time value which minimizes both the correct trip ends missed and the sum of the uncertain and incorrect trip ends. The analysis shows that the minimum occurs at a stop time of 110 seconds, which compares well with the 120 second rule commonly found in the literature. The 110 seconds was subsequently used as the stop time value for this study.

Repeated Use of Road Links

Repeated use of road links occurs when a vehicle travels along a road and then turns around, returning along the same road. This was used as an alternative criterion to identify trip ends within shopping centers, where stops of very short duration may occur which may cause them to be missed. In such cases trip ends will more likely be identified by searching for the traces that a vehicle makes when being parked. The method used to detect these movements was as follows:

- Calculate the distance between the 50th preceding and 50th next point;
- Calculate the distance between the 40th preceding and 40th next point;
- Calculate the distance between the 30th preceding and 30th next point;
- Calculate the average of the three distances;
- If the average distance is less than 20m, the point is marked as a trip end.

The 20m distance is based on the width of a road reserve including two lanes per direction and a 6m wide median as shown in Figure 2. The width of the two center lanes plus the median equals 13m. The accuracy level of the GPS points of 2.5m was also taken into account by adding 2.5m on both sides, which equals 18m, and then rounding off to 20m.

Cleaning Trip Ends

In some cases the same trip end was identified by both of the preceding criteria, producing consecutive trip ends located in close vicinity to each other. During cleaning of trip ends, consecutive stops that were within 300m of each other were collapsed into a single stop, on the basis that 300m is a comfortable walking distance that should not need two vehicular stops to be made. A second cleaning criterion was added to remove identified stops resulting from heavy congestion rather than true trip ends. The average speed from the previous 30th point to the next 30th point around an identified stop was calculated, and if this was less than 10km/h the stop was assumed to be indicative of slow movement due to congestion rather than a true trip end, and removed from the database.

Identifying Trip Ends at Shopping Centers and Home

The coordinates of trip ends were compared to a database (SACSC, 2012) containing all shopping centers in Gauteng Province. It was assumed that all trip ends located within a specified radius of a shopping center's centroid were relevant. This radius can be expected to vary with the size of the center, and should be long enough to cover the parking area of the shopping center but not so long as to pick up other trip ends on adjacent properties.

Shopping centers were classified by size (Gross Leasable Area), using the cut-offs shown in Table 1. In order to determine an appropriate radius per shopping center type, ten shopping centers were randomly selected from each of the types, except for the super regional centers of which all five was selected. The distance between the centroid and the furthest boundaries/parking spaces of each center was measured from satellite imagery, and the average was used as the proximity criterion in the further identification of shopping trip ends. Table 1 shows the recommended radii. As was expected, the radius increases with shopping center size.

In order to differentiate between primary (Home to Shop and Shop to Home) and secondary trips, the participant's home coordinates had to be determined. This was done with reference to the location where the vehicle spent most of the evening and early morning hours (night hours), in conjunction with the home suburb provided in the GPS questionnaire.

Calculating Trip Lengths for Shopping Center Trips

For each trip log, a cumulative distance was calculated using the point-to-point method by adding the distances between the points recorded every second. With shopping center trips already identified, the trip length *to* a shopping center was calculated by subtracting the shopping center trip end's cumulative distance from the previous trip end's cumulative distance value. For trips *from* a shopping center, the same method is followed except the next trip end after the shopping center trip end's cumulative distance is used. The results are discussed and analyzed in the next section.

Analysis and Results

Results of interest include not only the average trip lengths estimated from the GPS sample, but also how they differ across shopping center types and sizes, and across trip direction (to versus from shops). We also examine the distribution of trip lengths, and derive appropriate trip length distribution models. Finally, we consider sample size issues for future applications of this method.

Shopping Center Trip Lengths

From the sample, trip lengths for the following trip types were calculated:

- Home to Shopping Center;
- Non-home to Shopping Center;
- Shopping Center to Home; and
- Shopping Center to Non-home.

The number of trips, average trip lengths, and standard deviations of each trip type are shown in Table 3. A total of 1,767 shopping center trips were retrieved, of which the majority had the other trip end at a non-home location. The average trip lengths range from 7.0 to 7.8km and the standard deviations range from 9.9 to 11.4km, and are similar across all trip types. An ANOVA test showed that there was no statistically significant difference in the average trip length between home-based and non-home-based shopping center trips (F value = 0.43, 95% confidence). For the rest of the analysis the trip types were combined into the following three trip types:

- To Shopping Centers;
- From Shopping Center; and
- To and From Shopping Centers (combination of first two).

Average Trip Lengths for Different Shopping Center Types

The number and percentage of trips per shopping center type are shown in Table 4. The Neighborhood Centers have the largest percentages of trips in this data set while the Super Regional Centers attract less than 5% of the total trips. The rest of the trips are spread in similar percentages from 10% to 15%. This follows the distribution of shopping centers per type in Gauteng.

The average trip lengths per shopping center type are shown in Table 4. The relatively large standard deviations (exceeding the average values) suggest high variability in the data, i.e. that all shopping centers tend to attract trips from a wide catchment area. However, trip lengths do on average increase with the size of the shopping center, in line with expectations: Neighborhood Centers attract the shortest trips (6.3km) and Super Regional Centers the longest (11.8km). In order to test whether the differences in trip lengths are statistically significant, a non-parametric Kruskal-Wallis H test was performed between successive center types. Due to the skewness of the data a simple t-test is not appropriate. The H test showed that, at the 95% confidence level, trip lengths differ between some of the groups only (Table 4). Centers can be clustered into three groups with statistically similar trip lengths, namely Convenience Centers in a first cluster; Neighborhood Centers, Community Centers and Small Regional Centers in a second cluster; and Regional Centers and Super Regional Centers in a third cluster.

GPS-derived trip lengths are longer than those obtained from conventional surveys reported for the United States. The weighted average GPS-based trip length of 7.7km is about twice as long as the value of 3.39 km reported in the Florida Studies Database (City of Albuquerque, 2004). The difference is more marked for smaller stores: compare values for small convenience stores (8.2km vs. 2.6km), with those for regional centers (termed superstores in the US) (10.2km vs. 9.5km) (Citrus County, 2006).

Whether this result reflects true underlying differences between Florida and Gauteng, or an instrumental difference between the two methods (survey-based versus GPS-based) is impossible to say. Firstly, land use patterns could differ between Florida and Gauteng, which could influence the distribution of origins and destinations and lead to different trip distance results. Similarly, shopping behavior might vary by location, with regard to the preferences of shoppers in choosing between closer convenience-type stores and more distant regional centers. Alternatively, different results may be attributable to differences in the sample composition of the GPS-based and the conventional survey-based methods, particularly related to the treatment of pass-by trips. As compared to single-stop trips (e.g. home-shop-home), the presence of pass-by trips (trips made as intermediate stops on the way from an origin to a primary trip destination) would tend to increase trip distances associated with individual shopping centers, especially if the primary trip is a longer-distance work trip. The GPS-based methodology would tend to include more pass-by trips in the sample than travel survey or interview-based methods, as respondents are more likely to omit intermediate stops during recall surveys (Bricka and Bhat, 2006). This explanation is supported by the fact that the difference between GPS-based and survey-based trip lengths is largest for the smaller types of centers, which are more likely to be visited as pass-by trips than larger regional centers (Smith, 1986). This is especially true for Convenience Centers, the smallest category of centers, which have an unexpectedly long average GPS trip length of 8.2km. Furthermore, trip chaining is very common in our database as the number of trip segments linking non-home and shopping center destinations far outnumber the number of segments between home and shopping centers (Table 3).

Clearly further research is needed into this issue, as it goes to the heart of whether GPS-based data can provide more accurate estimates of trip lengths than traditional methods. A fair amount of research has been done on the *number* of pass-by trips to retail sites (e.g. Faghri et al., 1999; ITE, 2012; Smith, 1986), while GPS data have proven to be a very useful data source in the already substantial literature on trip chaining behavior (e.g. Huang and Levinson, 2016; Ma et al., 2016). Work of this nature can be usefully extended to clarify the impacts of trip chaining on trip distances as measured by GPS data.

Trip Length Frequency Distributions

The purpose of this section is to estimate best-fitting functions to the trip length distributions for each type of shopping center. This would be valuable during updating of the findings for the purposes of transferring the results to other areas, as smaller trip samples obtained locally could be used to estimate the parameters under the assumption that the distributional form remains constant. Trip length frequency distributions were calculated with intervals of 1 km. A frequency distribution function was estimated by fitting Gamma, Weibull and Exponential distributions through the trip length frequency points for each trip type (To Shop, From Shop and To & From Shop). Pearson et al. (1974) recommend that the Gamma and Weibull distributions should be used for modelling trip length distributions. The least squares method was used to estimate the function parameters.

The least squares method does not force the function mean to equal the mean of the sample. As the mean trip length is important for the purposes of calculating impact fees, it was deemed desirable to use an adjusted least squares method that minimizes the sum of the squared errors while forcing the mean to equal the sample mean. The results of both methods are similar, as shown in Table 5, although the latter method decreases the goodness of fit slightly.

Table 5 suggests that the Weibull distribution produces the best fit for all three trip types in terms of the R^2 and RMS errors. The best fitting parameters for the Weibull function are given in Table 6, together with the regression statistics. The final functions were estimated using the combined trip type (To & From Shopping). The parameters α and β refer to the standard shape and scale parameters of the Weibull distribution, while the mean is given as:

$$\mu = \alpha * \Gamma(1 + \frac{1}{\beta})$$

Where:

α = Shape parameter

β = Scale parameter

μ = Mean

The goodness of fit values are reasonably high for most shopping center types, suggesting that the adjusted Weibull distribution fits the data quite well. The exception is the Super Regional Centers which have a low R^2 value, although this is likely due to the small sample size rather than a more fundamental reason. We suggest the results for Super Regional Centers should be used with caution. Figure 3 plots the final distributions, confirming that the larger shopping center types draw fewer shorter distance trips (i.e. from the immediate vicinity), but more trips from further afield.

Sample Size Considerations

The high standard deviations observed in the trip lengths raise the question of the size of the samples required for accurate estimation of average trip lengths in future applications of GPS methods. We calculated the level of confidence attached to this sample for every shopping center type, assuming the population variance to equal the sample variance, for a maximum error of 1 kilometer in the mean. The results (Table 7) show that a confidence level of at least 95% was achieved for Neighborhood, Community, and Small Regional Centers; about 80% for Convenience and Regional Centers; and about 50% for Super Regional Centers. This suggests that future applications of the technique might need to expend extra effort to collect more data from the smallest and largest types of centers.

Should a 95% confidence level be required for each individual shopping center type (and still assuming a 1-kilometer maximum error), approximately 2,840 data points should be available (Table 7, last row). Regional Shopping Centers, having the largest variance in trip lengths, require the largest individual sample of 650 trips.

The challenge is to determine how many participants and what duration is required to collect the required data using GPS surveys. The answer depends on how respondents are recruited. If recruitment takes place at representative shopping centers, or results are required for only a subset of center types, the final column of Table 7 can be used as a guideline for the sample required. Should a shopping center trip rate of 0.84 shopping center trips per person per day (calculated from the present GPS data), and a loss of 15% of participants' data due to signal loss or post-processing error be assumed, the total number of participant-days to be collected can be calculated as 1/(0.84*0.85), or 1.4, times the number in the final column of Table 7.

Should the sample be drawn from a random group of drivers in an area, it is not possible to control the sample size for each center type very accurately. In this case, the total sample size needs to be estimated using the expected distribution of trips per center type. For instance, in our data, Regional Shopping Centers (requiring the largest amount of trips for a 95% confidence interval) constitute 15% of the total sample. In order to end up with 650 Regional Shopping Center trips, a total of 1.4*650/0.15, or 6,067 participant-days, needs to be collected. With this sample, all other shopping center types will have enough data to reach a 95% confidence level, with the exception of Super Regional Centers which, as mentioned above, needs to be treated with caution and for which additional targeted data may need to be collected.

The cost of such an exercise might be significant, but two issues need to be considered. Firstly, GPS surveys are very suited to collecting data from the same respondent over multiple days, which lowers the total cost significantly due to efficiencies in recruitment, administrative, and data manipulation costs. For instance, we could have reached the above sample size of 6,067 participant-days by extending our current survey from three to seven days with the same set of drivers, at minimal additional cost. Secondly, with many cities using bulk services contributions as a source of revenue to fund infrastructure development, the cost and effort might easily be justified.

Conclusions and Recommendations

With many cities adopting development impact fees and bulk services contributions as a source of revenue to fund infrastructure development, it is important that the data underlying the estimation of traffic impact assessments be relevant and accurate. Trip lengths is one such data element increasingly used in several jurisdictions in the USA and South Africa. The paper develops and tests the use of a novel approach using mobile GPS loggers for collecting trip length data, with a view to assessing its potential for helping to improve the available evidence base on trip characteristics. It is applied specifically to trips to or from shopping centers, based on a large sample GPS study in Gauteng Province in South Africa.

The main finding is that the GPS-based method is feasible and capable of delivering trip length data across a wide range of shopping center trip types and areas. The accuracy of the data was such that relationships between trip lengths and shopping center types could clearly be identified. A more refined method for identifying trip ends from the large GPS dataset was tested. It was found that a stop time of 110 seconds, together with repeated use of road links, were the most suitable criteria for identifying trip ends in this data. This method is especially useful where no reliable GIS layer indicating road classification is available, as often occurs in developing countries. Several criteria for cleaning trip data had to be applied, reflecting the fact that GPS data still requires significant amounts of additional manipulation before it is usable for analysis.

Significant variation in trip lengths was observed, with standard deviations exceeding mean values for each category of shopping center. As expected the average trip length increases as the shopping center size increases and as one moves up the shopping center hierarchy. A similarity test showed that shopping center types can be collapsed into three clusters with similar average trip lengths, with Neighborhood Centers, Community Centers and Small Regional Centers in a cluster with the lowest trip lengths; and Regional Centers and Super Regional Centers in a cluster with the highest.

Average GPS-derived trip lengths were found to be significantly longer than those reported in previous conventional travel survey or interview-based recall studies in the United States. The data does not allow us to tell whether this is due to true underlying differences in trip lengths due to land use and behavioral factors, or to instrumental differences between the GPS and survey-based methods. It is possible, although impossible to prove at present, that GPS-derived samples include more pass-by trips, as respondents are more likely to omit intermediate stops during interview-based recall surveys. If this were true, it would imply that GPS-based methods are less biased and less likely to underestimate trip distance distributions to specific land uses than survey-based methods. It is an important point as current methods used by some jurisdictions to estimate development impact fees and bulk services contributions in both South Africa and the US depend on trip distance information derived from survey-based data. It was estimated that these data might underestimate the quantum of impact and therefore the impact fee payable by individual developers by as much as 3 times under certain assumptions (Jonker, 2016). Given the potential importance of this result for the revenues collected by local authorities, it would be beneficial to confirm and explore it with further GPS-based studies specifically designed to differentiate between primary and secondary trips made as part of trip chains, and in other locales. Further research explicitly comparing GPS-based and survey-based data obtained from the same sample would also be highly useful.

Overall, the use of vehicle-based GPS technology to extract trip length data holds potential for more accurate distance estimation, and (although not reported here) for identifying the amount of travel occurring on different classes of roads. However, further methodological refinement and validation is needed before it can be considered as a robust data source, specifically with regard to the following:

- Differences in errors need to be investigated between the point to point methods and the link to link method when calculating trip distances. Clearer guidelines need to be developed around which are more appropriate to use under different circumstances, depending on the magnitude of errors in both the available GIS and GPS data.
- Trip identification algorithms and GPS data cleaning methods need further work, specifically with regard to balancing the additional complexity of using additional data sources (like road networks) when identifying trip ends, with the marginal benefit in terms of better trip identification. There is great need for studies relating derived trip ends to ground truth obtained by other means from the same sample (e.g. Shen and Stopher, 2014b).
- As with all data collection, sampling issues remain critical. If, as in the present study, the GPS data was collected for a different purpose, the potential for (known or unknown) bias in the data is a potential problem that can only partially be corrected by weighting. To get fully

representative trip length data, properly designed GPS sampling strategies and quality control procedures still need to be developed.

The work further underlines the importance of doing local studies before data are transferred from one country to another, particularly if they are to be the basis for decisions as important as development impact fees. Updating and transfer of trip lengths might be simplified by looking at the distribution of trip lengths rather than the average value. This research found that the Weibull distribution fits the GPS-derived trip length data very well, and might be used in further studies.

Acknowledgements

The authors which to acknowledge and thank:

Dr Adrian Ellison from the University of Sydney for his useful comments; and

The South African National Road Agency for sponsoring the original data collection.

References

Axhausen, KW, Schonfelder, S, Wolf, J, Oliveira, M and Samaga, U (2003). ’80 weeks of GPS traces: Approaches to enrich the trip information’, *Arbeitsbericht Verkehrs- und Raumplanung, 178*, *Institut für Verkehrsplanung und Transportsysteme*, ETH Zürich, Zürich.

Bricka, S and Bhat, C (2006). “Comparative analysis of global positioning system-based and travel survey-based data”. *Transportation Research Record: Journal of the Transportation Research Board*, 1972, 9-20.

Bricka, SG, Sen, S, Paleti, R and Bhat, CR (2012). ‘An analysis of the factors influencing differences in surveys-reported and GPS-recorded trip’, *Transportation Research Part C,* 21(2012), 67-88.

Chen, C, Gong, H, Lawson, C and Bailostozky, E (2010). ‘Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study’, *Transportation Research Part A,* 44(2010), 830-840.

Citrus County (2006). *Impact Fee Study Draft Report*, Tindale-Oliver & Associates, Inc. Lecanto, Florida.

City of Albuquerque (2004). *Roadway facilities impact cost study, Summary Report*, Tindale-Oliver & Associates, Inc., Tampa, Florida.

City of Johannesburg (COJ) (2008). *Policy for Engineering Service Contributions for Roads and Stormwater*, Johannesburg Road Agency (JRA), Final Report. Johannesburg.

Clifford, E, Zhang, J and Stopher, P (2008). *Determining trip information using GPS data*. Institute of transport and logistics studies, The University of Sydney, ITS-WP-08-01.

Committee of Transport Officials (COTO) (2012), *South African Traffic Impact and Site Traffic Assessment Manual,* TMH 16, Volumes 1 and 2. Pretoria.

Committee of Transport Officials (COTO) (2013), *South African Trip Data Manual*, Version 1.01. TMH 17. Pretoria.

Department of Transport (DOT) (2005). *The First South African National Household Travel Survey 2003 – Key results of the National Household Travel Survey*, Final Report. Pretoria.

Du, J, and Aultman-Hall, L (2007). 'Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification', *Transportation Research Part A*, 41(3), 220-232.

Faghri, A, Aneja, S and Vaziri, M (1999). 'Estimation of percentage of pass-by trips generated by a shopping center using artificial neural networks'. *Transportation Planning and Technology*, 22(4), 271-286.

Fan, HS and Lam, SH (1997). 'Parking generation of commercial developments in Singapore'. *Journal of Transportation Engineering*, 123(3), 238-242.

Greaves, SP and Figliozzi, MA (2008). 'Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology: Issues and Potential Applications', *Transportation Research Record: Journal of the Transportation Research Board*, 2049, 158-166.

Holguín-Veras, J and Thorson, E (2000). 'Trip length distributions in commodity-based and trip-based freight demand modeling: investigation of relationships'. *Transportation Research Record: Journal of the Transportation Research Board*, 1707, 37-48.

Huang, A and Levinson, D (2016). ‘A model of two-destination choice in trip chains with GPS data’. *Journal of Choice Modelling*, in press.

Institute of Transportation Engineers (ITE) (2012), *Trip Generation Manual*, 9th edn, Washington, D.C.

Joubert, JW and Meintjes, S (2015). ‘Repeatability & reproducibility: Implications of using GPS data for freight activity chains’. *Transportation Research Part B: Methodological*, 76, 81-92.

Kelly, P, Krenn, P, Titze, S, Stopher, P and Foster, C (2013). ‘Quantifying the Difference Between Self-Reported and Global Positioning Systems-Measured Journey Durations: A Systematic Review’, *Transport Reviews*, 33(4), 2013, 443-459.

Lu, Y, Zhu, S and Zhang, L (2013). ‘Imputing trip purpose based on GPS travel survey data and machine learning methods’. In *Proceedings of the 92nd Annual Meeting of the Transportation Research Board.* Washington, D.C.

Ma, X, McCormack, E and Wang, Y (2011). ‘Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program’, *Transportation Research Record: Journal of the Transportation Research Board*, 2246, 92-100.

Ma, X, Wang, Y, McCormack, E, and Wang, Y (2016). ‘Understanding Freight Trip Chaining Behavior Using Spatial Data Mining Approach with GPS Data’, *Transportation Research Record: Journal of the Transportation Research Board*, 2596, 44-54.

Murakami, E and Wagner, DP (1999). ‘Can using global positioning systems (GPS) improve trip reporting?’, *Transportation Research Part C,* 7(1999), 149-165.

National Cooperative Highway Research Program (NCHRP) (2014), *Applying GPS Data to Understand Travel Behaviour, Volume 1: Background, Methods and Testing*, Report 775. Washington, D.C.

Pearson, DF, Stover, VG and Benson, JD (1974), *A Procedure for Estimation of Trip Length Frequency Distributions*, Texas Transport Institute, Report No.: TTI-2-10-74-17-1. Austin, Texas.

Prinsloo, DA (2010). ‘Classification and hierarchy of retail facilities in South Africa’, *Urban Studies*, South Africa.

Putnam, R (2000). *Bowling alone: the collapse and revival of American community*, Simon and Schuster, New York.

Quiroga, CA and Bullock, D (1999), ‘Measuring control delay at signalized intersections’, *Journal of Transportation Engineering*, 125(4), 271-280.

Reumers, S, Liu, F, Janssens, D. and Wets, G (2014). ‘The Annotation of Global Positioning System (GPS) Data with Activity Purposes Using Multiple Machine Learning Algorithms’. *Mobile Technologies for Activity-Travel Data Collection and Analysis*, Rasouli S. and Timmermans H (eds), IGI Global, 119-133.

Rodriguez, D, Brown, AL and Troped, PJ (2005). ‘Portable global positioning units to complement accelerometry-based physical activity monitors’, *Medicine & Science in Sports & Exercise*, 37(11 Suppl), S572–S581.

Santos, A, McGuckin, N, Nakamoto, HY, Gray, D, and Liss, S (2011). *Summary of Travel Trends: 2009 National Household Travel Survey*. Federal Highway Administration, Washington, DC.

Sharman, BW and Roorda, MJ (2011). ‘Analysis of freight global positing system data: clustering approach for identifying trip destinations’, *Transportation Research Record: Journal of the Transportation Research Board*, 2246, 83-91.

Shen, L and Stopher, PR (2014a), ‘Review of GPS travel survey and GPS data-processing methods’, *Transport Reviews*, 34(3), 316-334.

Shen, L and Stopher, PR (2014b), ‘Using SenseCam to pursue “ground truth” for global positioning system travel surveys’. *Transportation Research Part C: Emerging Technologies*, 42, 76-81.

Smith, S.A., (1986). ‘A Methodology for consideration of pass-by trips in traffic impact analyses for shopping centers’. *ITE Journal*, August, 37-40.

South African Council of Shopping Centers (SACSC) (2012), *South African Shopping Center Directory*. Johannesburg, South Africa.

Srinivasan, S, Bricka, S, and Bhat, C (2009), *Methodology for converting GPS navigational streams to the travel-diary data format*. Department of Civil and Coastal Engineering, University of Florida.

Stopher, PR, FitzGerald, C and Zhang, J (2008), 'Search for a global positioning system device to measure person travel', *Transportation Research Part C*, 16(3), 350-369.

Stopher, PR, Clifford, E, Zhang, J and FitzGerald, C (2007). 'Deducing mode and purpose from GPS data'. In *Proceedings of the Transportation Planning Applications Conference of the Transportation Research Board*. Daytona Beach, FL: Academic Press.

Stover, VG and Koepke, FJ (2002), *Transportation and Land Development*, 2nd ed, Institute of Transport Engineers.

Tindale, SA (1991). 'Impact Fees – Issues, Concepts and Approaches', *ITE Journal*, May, 33-40.

Timmermans, H and Rasouli, S (2013). 'Assessment of model uncertainty in destinations and travel forecasts of models of complex spatial shopping behaviour', *Journal of Retailing and Consumer Services*, 20(2), 139-146.

Venter, CJ and Joubert, JW (2013). 'Use of Multisource GPS Data to Characterized Multiday Driving Patterns and Fuel Usage in a Large Urban Region', *Transportation Research Record: Journal of the Transportation Research Board*, 2338, 1-10.

Venter, CJ and Joubert JW (2014). 'Tax or toll? GPS-based assessment of equity impacts of large scale electronic freeway tolling in Gauteng, South Africa', *Transportation Research Record: Journal of the Transportation Research Board*, 2450, 62-70.

Wolf, J, Guensler, R and Bachman, W (2001). 'Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data'. *Transportation Research Record: Journal of the Transportation Research Board*, 1768, 125-134.

Tables

Table 1. Shopping Center Classification

Classification	Size of Center (m2) (GLA)	Number of stores	Size of land (ha)	Average radius of primary trade area	Median travel time to the center	*Recommended Radius (m)
Convenience Centers	500 to 5000	5 to 25	0.15 to 1.5	1 to 1.5 km	2 to 3 minutes	70
Neighbourhood Centers	5000 to 12000	25 to 50	1.5 to 3.6	1.5 to 2.0 km	4 to 9 minutes	100
Community Centers	12000 to 25000	50 to 100	3.6 to 7.5	2.5 to 3.0 km	6 to 14 minutes	150
Small Regional Centers	25 000 to 50 000	75 to 150	7.5 to 15.0	3.0 to 5.0 km	10 to 16 minutes	220
Regional Centers	50 000 to 100 000	150 to 250	More than 15.0	5.0 to 8.0 km	14 to 20 minutes	280
Super Regional Centers	More than 100 000	More than 250	-	More than 10.0 km	24 to 30 minutes	300

Source: Prinsloo, 2010

* This column was calculated by analysis of a sample of shopping centers

Table 2. Criteria Used to Evaluate Trip Ends

Maximum Stop Time (seconds/minutes)	Location of Trip End					
	On Freeways, On-Ramp or Off-Ramp (Class 1)	On Major Roads (Class 2)	On Road Near Intersection (Class 3)	On Road Far From Intersection (Class 3)	On Minor Road (Class 4 & 5)	Within Property
45 / 0:45	Incorrect	Incorrect	Incorrect	Incorrect	Uncertain	Correct
60 / 1:00	Incorrect	Incorrect	Incorrect	Incorrect	Uncertain	Correct
80 / 1:20	Incorrect	Incorrect	Incorrect	Incorrect	Uncertain	Correct
100 / 1:40	Incorrect	Incorrect	Incorrect	Uncertain	Correct	Correct
120 / 2:00	Incorrect	Incorrect	Incorrect	Uncertain	Correct	Correct
150 / 2:30	Incorrect	Incorrect	Incorrect	Uncertain	Correct	Correct
180 / 3:00	Incorrect	Incorrect	Uncertain	Uncertain	Correct	Correct
240 / 4:00	Incorrect	Incorrect	Uncertain	Uncertain	Correct	Correct
300 / 5:00	Incorrect	Uncertain	Correct	Correct	Correct	Correct
600 / 10:00	Incorrect	Uncertain	Correct	Correct	Correct	Correct

Table 3. Average Trip Lengths per Trip Type

Trip Type	Trips Identified in Data Set	Average Trip Length (km)	Standard Deviation (km)
Home to Shopping Center	134	7.3	11.3
Non-home to Shopping Center	725	7.8	10.9
ALL To Shopping Center	859	7.8	11.0
Shopping Center to Non-home	722	7.8	9.9
Shopping Center to Home	186	7.0	11.4
ALL From Shopping Center	908	7.6	10.3
ALL Shopping Center Trips	1,767	7.7	10.6

Table 4. Average Trip Lengths and Sample Properties per Shopping Center Type

Shopping Center Type	To Shop	From Shop	To & From Shop			
	Average Trip Length (km)	Average Trip Length (km)	Average Trip Length (km)	Standard Deviation	Number (%) of Shopping Center Trips in Sample	p-value for difference between groups (Kruskal-Wallis H test)*
Convenience Centers	8.5	7.8	8.2	13.0	284 (16%)	--
Neighborhood Centers	6.5	6.1	6.3	9.1	507 (29%)	*0.02**
Community Centers	6.6	7.6	7.1	8.7	332 (19%)	0.20
Small Regional Centers	8.2	6.0	7.1	9.0	311 (17%)	0.50
Regional Centers	9.7	10.7	10.2	13.0	268 (15%)	*0.001**
Super Regional Centers	10.3	13.1	11.8	12.7	65 (4%)	0.40
All Shopping Centers	7.8	7.6	7.7	10.6	1,757 (100%)	--

**Italic indicates significant difference between preceding and current groups, 95% significance*

Table 5. Goodness of Fit Evaluation for Trip Length Distributions To and From Shopping Centers

Variable	Function Mean Not Equal to Sample Mean			Function Mean Equal to Sample Mean		
	Gamma	Weibull	Exponential	Gamma	Weibull	Exponential
R^2	0.912	*0.920*	0.872	0.839	*0.868*	0.836
*RMS Error	*0.010*	0.011	0.012	*0.014*	*0.014*	*0.014*
Error Squared	*0.006*	0.007	0.009	*0.011*	*0.011*	*0.011*
**Mean Diff (km)	-2.6	-2.7	*-1.8*	*0.0*	*0.0*	*0.0*

* RMS: Root Mean Square

** Mean Diff: The difference between function mean and sample mean

*** Italic indicates best fit value

Table 6. Weibull Distribution Results for Shopping Center Types (equal means)

Shopping Center Type	Alpha (α)	Beta (β)	R^2	RMS Error	Error Squared
Convenience Center	0.83	7.41	0.702	0.025	0.022
Neighbourhood Center	0.99	6.26	0.795	0.021	0.018
Community Center	1.13	7.45	0.823	0.017	0.010
Small Regional Center	1.16	7.48	0.724	0.023	0.018
Regional Center	1.07	10.48	0.789	0.014	0.008
Super Regional Center	1.22	12.56	0.396	0.021	0.012

Table 7. Confidence levels and required samples sizes

Shopping Center Type	Number (%) of Shopping Center Trips in Sample	Confidence Level with this Sample and Error of 1 km	No of Trips Required for 95% Confidence Level and Error of 1 km
Convenience Center	284 (16%)	81%	646
Neighbourhood Center	507 (29%)	99%	318
Community Center	332 (19%)	96%	289
Small Regional Center	311 (17%)	95%	312
Regional Center	268 (15%)	79%	650
Super Regional Center	65 (4%)	47%	622
All Shopping Centers	1,757 (100%)	--	2,837