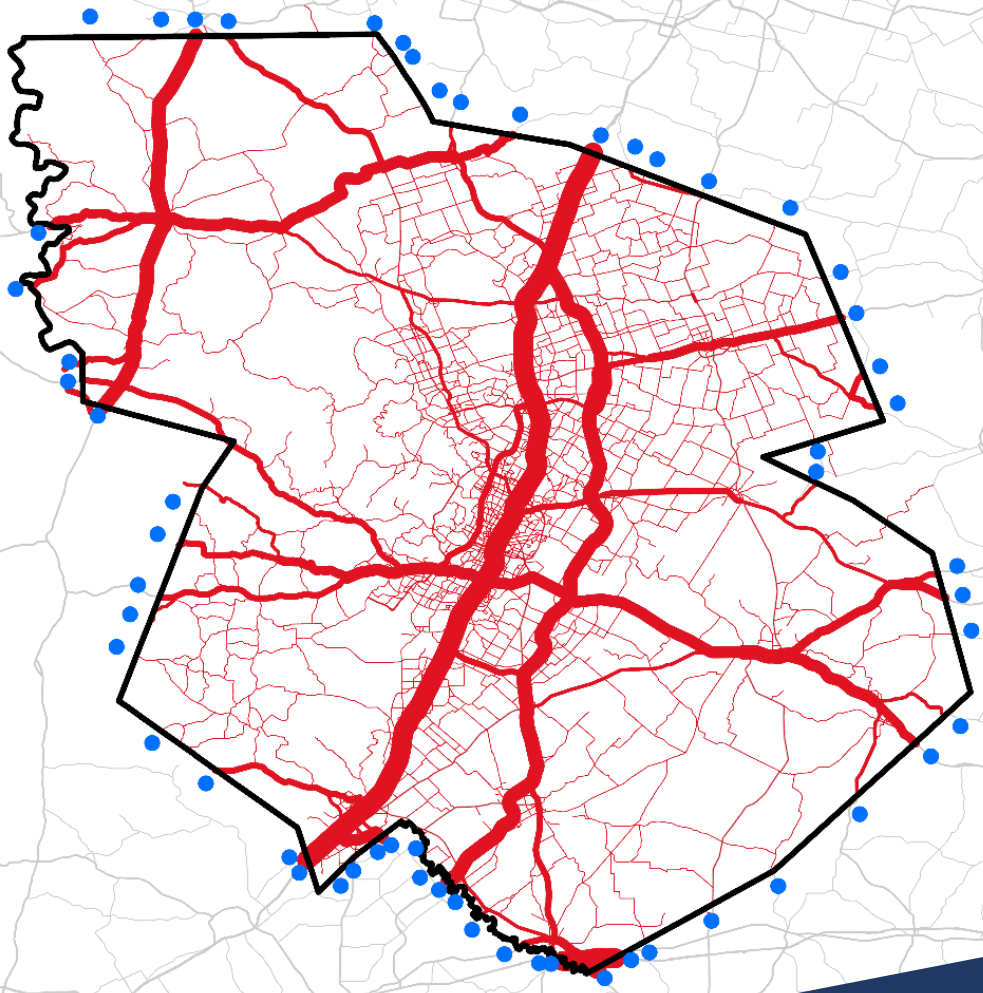
2015 CAPITAL AREA MPO EXTERNAL STUDY

Technical Summary

November 2018



Prepared by the
Texas A&M Transportation Institute

Texas A&M Transportation Institute

Texas Department of Transportation

2015 Capital Area MPO External Study Using Cellular and GPS Data

Technical Summary

Texas Department of Transportation Travel Survey Program

By

Byron Chigoy, Associate Research Scientist
L.D. White, Research Specialist IV
Michael Martin, Assistant Research Scientist
Gargi Singh, Assistant Transportation Researcher
and
Edwin N. Hard, Research Scientist

Performed in cooperation with the

Texas Department of Transportation

and the

Capital Area Metropolitan Planning Organization

and the

Federal Highway Administration

November 2018

TEXAS A&M TRANSPORTATION INSTITUTE

The Texas A&M University System

College Station, Texas 77843-3135

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration (FHWA) or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification or regulation.

ACKNOWLEDGEMENTS

The authors would like to thank Christeen Pusch, TxDOT Travel Survey Program Manager, and the Department for its continuing program to collect and analyze urban travel data to support travel demand modeling.

TABLE OF CONTENTS

List of Figures iv
List of Tables v
Study Background and Objectives 1
CAMPO Study Area 1
Identifying External Trips with GPS and Cellular Data 1
Data Collection 4
Traffic Counts 4
Cellular Data 5
GPS Data 6
Map Matching 8
Cellular Data 8
GPS Data 11
Data Expansion 13
Cellular Data 13
GPS Data 18
O-D External Trip Matrices 18
Results 18
Expanded External Trips and Recommended Estimations 19
SH-130 and IH-35 21

LIST OF FIGURES

Figure 1. CAMPO External Survey Geography and Zone Structure..... 3
Figure 2. Cellular External Trip (E-E) Assignment Flow Bands..... 9
Figure 3. Cellular External-Local (I-E/E-I) Assignment Flow Bands..... 10
Figure 4. Matching GPS Data to Road Segments..... 11
Figure 5. External Station Matching Example..... 12
Figure 6. Unmatched External Trips..... 13
Figure 7. Cellular External-Local Trips per Square Mile by Zone..... 15

LIST OF TABLES

Table 1. Cellular Area-Wide Weekday Averages. 5
Table 2. GPS Weekday Total Sample. 7
Table 3. GPS Weekday Total Sample Cars/Light Trucks. 7
Table 4. GPS Weekday Total Sample Medium and Heavy Trucks. 7
Table 5. Passive Data Sample Size Distributions. 16
Table 6. Cellular Area-Wide Weekday Expanded Trips. 18
Table 7. GPS Area-Wide Weekday Expanded Trips. 19
Table 8. Comparison of E-E Trips and Total External Traffic Percentages between GPS and Cellular Data. 19
Table 9. CAMPO Expanded External Trips. 20

STUDY BACKGROUND AND OBJECTIVES

This technical memorandum provides a summary of an external travel study performed for the Capital Area Metropolitan Planning Organization (CAMPO) in the spring of 2018 for origin-destination (O-D) data acquired for the spring of 2015. The Texas Department of Transportation (TxDOT) acquired the O-D for this time to coincide with traffic count data collected by TxDOT in the spring of 2015. The primary outcome of the study includes external-to-external (E-E), external-to-internal (E-I), and internal-to-external (I-E) travel data for passenger vehicles and trucks. TxDOT purchased cellular and Global Positioning System (GPS)-based O-D data for the study to provide estimates of external travel needed as input for the CAMPO travel demand model. The GPS data were acquired from INRIX and the cellular data from Airsage.

The use of passive data for external studies has become state-of-the-practice in recent years and provides significant benefits over prior methods such as roadside interviews and license plate matching techniques that garner safety and privacy concerns. The results of this study benefit many uses of the travel model such as evaluation of freeway improvements, new location corridors and tolling studies, as well as air quality/conformity assessments. The study's primary objective was the development of traffic matrices of external O-D interactions for the regional travel demand model. Since the use of passive technology to study travel is still evolving, the study also included a comparison of GPS and cellular data results. Researchers used data management techniques, algorithms, and analytical computations to address the technical and data processing challenges inherent in using large amounts of passive data.

Table 9 provides the expanded external trips for the CAMPO area. They represent summary results of the study and are explained with supporting information in later sections of this report.

CAMPO Study Area

The CAMPO study area includes Williamson, Travis, Hays, Bastrop, Burnet and Caldwell counties. Core elements of the study area and design include Traffic Analysis Zones (TAZs) and 'external' stations where major roadways and thoroughfares cross the study area boundary.

Figure 1 shows the study area, the study area boundary, the external stations, and the TAZs. The research team aggregated the CAMPO's TAZs into larger zones to compensate for cellular data's lower positional accuracy. These larger "cell capture" TAZs are based largely on census tract boundaries because they often align with the model zones and serve as the basis for cellular data expansion. In addition, the use of census tract geography coincides with American Community Survey (ACS) data products. The CAMPO's model TAZ structure contains 2102 TAZs and 59 external stations. The research team aggregated the TAZs into 975 larger TAZs to better capture cellular data.

Identifying External Trips with GPS and Cellular Data

Identifying the external station (or roadway) used by trips that cross the study area boundary is an essential part of an external travel study. For GPS data, the external station is determined using a process known as map matching, where a trip's GPS points are electronically associated with the roadway network. This process reveals the route used between the trip's origin and destination that allows the external station to be identified.

Cellular data is polygon-based rather than point-based (as with GPS data), which means that the analysis assigns trip O-D pairs to TAZs and the route used between the trip's origin and destination often must be estimated. Typical analysis methods for cellular O-D data assign the route or roadway by "association" and assume that trips occurring within a geographic travel shed drawn around a road to be on the road. This process works relatively well for isolated roadways but is challenging in developed and/or urban areas where other parallel or nearby roadways exist. Where this occurs, roadways in close proximity to each other would in fact serve distinct O-Ds that are prohibitive to distinguish empirically. To distinguish the distinct O-Ds served by all the region's external roadways, the study area for cellular data, as shown in Figure 1, includes 123 "halo" zones around the periphery of the study area in counties bordering the area. The development of halo zones involved a careful review of regional populations, roadway locations, and travel patterns. The halo zones provide a basis for a select link analysis (see the Map Matching section for details) that routes observed O-D flows through the appropriate external roadway.

Halo zones serve several additional purposes. First, they distinguish between resident and non-resident travel, with non-residents being those that live in the halo zones. Second, they capture cellular signals before they enter and after they exit the study area. Third, depending on forecasting needs, modelers can forecast each external zone independently, which allows for modeling different assumptions of growth in external traffic.

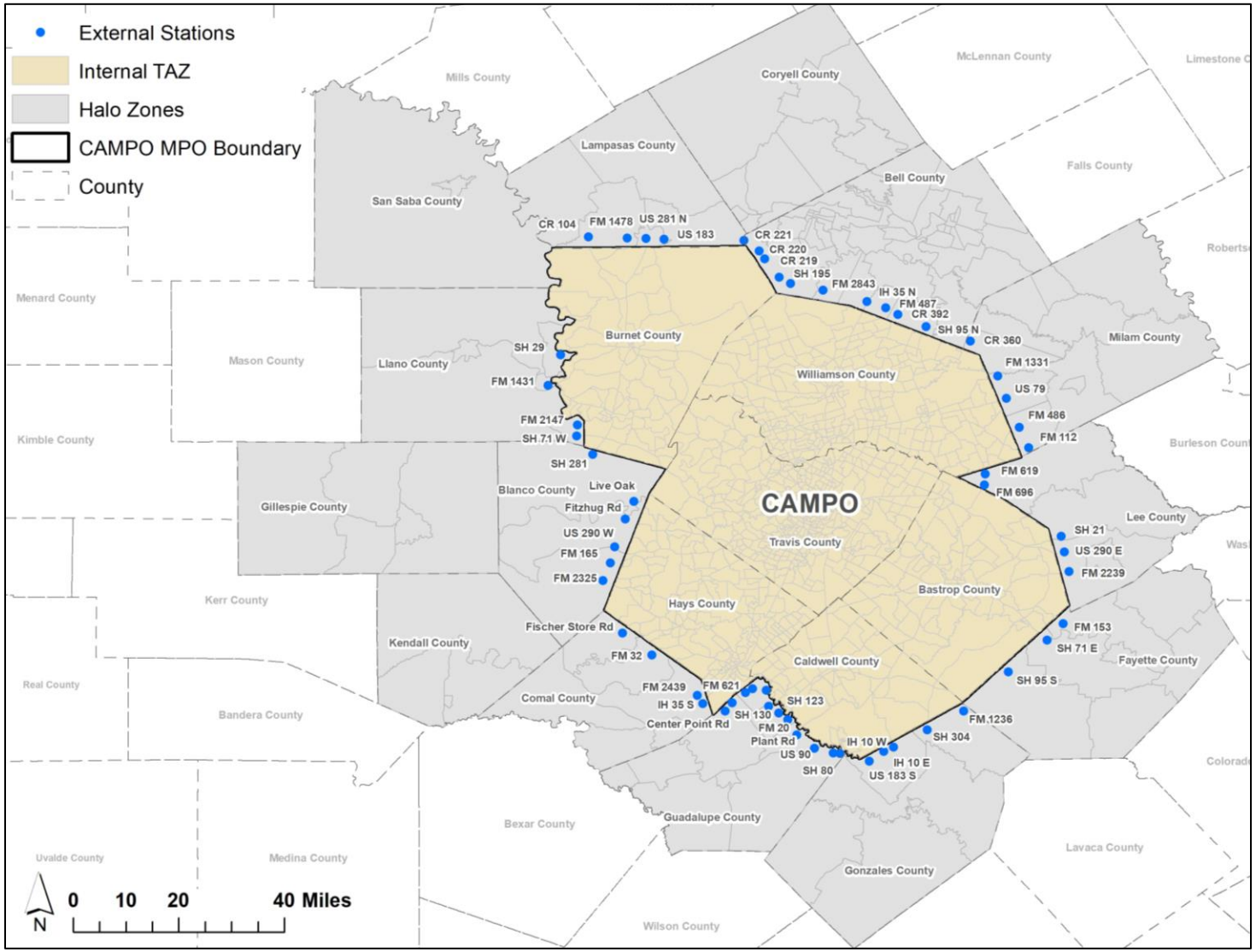


Figure 1. CAMPO External Survey Geography and Zone Structure.

DATA COLLECTION

Vendors of cellular and GPS data maintain large archives of data over many years and have extraction tools that efficiently pull data from their archives based on a range of dates and geographies provided for each study. The traffic counts are data collected in the "field" that require direct interaction with the roadway under study but provide relatively accurate ground truth data to baseline and expand the cellular and GPS data.

Traffic Counts

As with previous data collection methodologies, cellular and GPS data sources represent a sample of traveler information for a study area and therefore these data must be expanded to provide an estimate for the entire study area. The most common method to expand the data is the use of expansion factors derived from traffic counts. For external surveys, this equates to traffic counts at all external stations for a study area. Furthermore, to provide modelers the data needed to prepare a model for a particular study area, the survey results need to be disaggregated by non-commercial/passenger and commercial/truck vehicle types. To meet this need, the traffic data required for developing the expansion factors must be classification counts.

The passive O-D data purchased for the CAMPO area are from 2015. This is the same year TxDOT performed saturation counts in the Austin District. Additionally, TxDOT collects annual counts on on-system roads throughout the state every year. The combination of saturation and annual counts in the CAMPO area is useful for model calibration due to the abundance of counts for links within the model network. However, both saturation and annual counts are volume counts and do not provide the classification data needed for external station data expansion.

Due to the lack of current count data, the research team compiled historical classification count data from the 2005 CAMPO External Survey to develop 2015 classification count estimates for the study area. These data included 24-hour daily averages for non-commercial and commercial vehicles at each of the external locations for the 2005 survey. Additionally, the team extracted 2015 saturation and/or annual count data from the STARS II database for each 2015 external location. They also queried the STARS II database to identify permanent count locations that aligned or were near external locations and had classification count data. Upon review, researchers determined that none of the 2015 external sites had permanent count stations nearby that collected classification count data. Upon review of all sources, researchers estimated classification data for the stations using 2005 survey data.

Cellular Data

TxDOT coordinated with TTI to purchase one month of cellular data from AirSage for the period between April 15, 2015 and May 15, 2015 for an average weekday. This time period when data were acquired corresponds with dates of TxDOT's traffic saturation counts for the region. Analysis of the data show that there are an estimated 8,079,580 (expanded) average weekday trips in the CAMPO study area, including 17,581,197 internal trips and 498,383 external trips. TxDOT purchased the data for average weekday trips, as only weekday travel is modeled for the CAMPO area. Additional options that were purchased include trips for five time periods, for three trip purposes, and trip classification based on whether the trip was made by a resident of the study area or a or visitor (non-resident) from outside the study area. The time periods included early AM (12 a.m.-6 a.m.), the AM peak (6 a.m.-10 a.m.), Mid-Day (10 a.m.-3 p.m.), the PM peak (3 p.m.-7 p.m.), late PM (7 p.m.-12 a.m.), and a full 24-hour day.

A notable drawback of cellular data is the lack of clear data on commercial vehicle travel. Presumably, commercial vehicle travel is present in the data. However, since cellular data primarily tracks cellular phone usage ultimately tied back to a subscriber's home, it is not possible to determine whether a subscriber left home and started driving a truck when they got to work. Additionally, for cellular studies, regional trips made by non-residents and transients (i.e., briefly observed devices) are detected. For these trips, it is difficult to determine the residence, or primary activity place of the device, which precludes methods to determine if that place is associated with commercial vehicle activity. These trips are often external; however, their mode is unknown. For example, in the Tyler, Texas study TTI observed a significant number of interstate I-E and E-I trips stopped at a large gas station/truck stop. This study did not focus on making similar observations as in Tyler, but researchers would expect similar findings in the CAMPO study area. This indicates that trucks are present, but again the quantity of trucks cannot be determined via the data.

The E-E, I-E, and E-I trip-type designations use the trip's direction of travel relative to the study area and whether its origin or destination is inside or outside of the internal model region. Accordingly, a trip with an external zone origin and an internal zone destination is an E-I trip and vice versa for an I-E trip. E-E trips are those with both ends of the trip being external (i.e., located outside of the study area). It is important to note that for a trip to be E-E it must pass through the modeling region. Table 1 shows the percentage breakdown of each of these types of trips. Based on the cellular data, approximately 6.7 percent of all external traffic entering and exiting the CAMPO area are E-E trips.

Table 1. Cellular Area-Wide Weekday Averages.

	E-I	I-E	E-E	I-I	All Types
Total Trips	232,843	231,997	33,543	7,581,197	8,079,580
Percent Total Trips	2.88%	2.87%	0.42%	93.83%	100.00%
Percent Total External Trips	46.72%	46.55%	6.73%	NA	100.00%

The sample size calculation for cellular data is the penetration rate, which is the amount of cellular subscribers captured by AirSage data in the study area versus the total population of the study area.

Based on prior studies using Airsage data, researchers estimate that the penetration rate for this study was about 17 percent of the study area population, or about 42,500 persons based on the 2015 model population. This rate is largely dependent on the market penetration of cellular carriers (or carrier) with whom AirSage has an agreement. During the data collection period for this study, AirSage indicated an agreement with one major U.S. carrier in the region.

GPS Data

TxDOT and TTI acquired three months of pre-processed GPS trip data from INRIX for the CAMPO study area. This data is from INRIX's Insights™ Trips product and includes individual trip records with waypoints. The trip records contain time-stamped trip start and trip end locations. Waypoint data is included for each trip record and provides points between the trip ends to identify the trip's route. INRIX provided a combined dataset of GPS data sources from non-commercial passenger cars/vehicles, commercial/freight vehicles (including trucks), and mobile applications.

The collection period for GPS data extended from February 2015 to April 2015. As with the cellular data, the dates of data purchase correspond to TxDOT's traffic saturation counts for the region. The initial GPS dataset included 6,451,107 trip records. The study team processed each trip to determine if the trip was a weekday trip, whether or not it entered the study area, and if it entered the study area by way of a roadway with an external station. This process revealed that 3,344,771 trips from the dataset did not meet these criteria (1,124,911 were weekend trips), making the final sample for analysis 3,106,336 trips. Each record contained a unique device identification (ID) and the trip ID with time-stamped trip end locations. Trip ends were pre-processed by INRIX using a 10-minute dwell time (e.g., could be thought of as stop time) threshold and a 1-mile dwell distance. The actual trip end locations were anonymized by truncating the origin and destination coordinates to three digits of precision. The time between each GPS waypoint ranged between .016 and 10.0 minutes with an average of 0.83 minutes and median of 1.0 minutes.

There are several ways to consider the sample size of GPS data relative to daily traffic. One way is to consider it as the average sample per day, or the total sample divided by the number of days in the collection period. Another way to calculate the sample size is to calculate it as a pooled sample, which is the total of all trips for observation days. This analysis reports sample size as the average sample per day. Accordingly, sampled passenger vehicles, including I-I trips, accounted for 39.49 percent of the average sample per day and 15.72 percent of the total sample for all external stations. Sampled, trucks, including I-I trips, accounted for 60.51 percent of the average sample per day and 84.28 percent of the total sample for all external stations.

The sources of GPS data included in this sample are passenger vehicles, field service/local delivery fleets, and for hire/private trucking fleets. Within the sample, the distribution of these sources is 38.36 percent passenger vehicles, 47.27 percent field service/local delivery fleets, and 14.37 percent for hire/private trucking fleets.

Unlike cellular data, the GPS data require significant post-processing by the data user. This processing requires large computing resources and data management to extract and impute additional details of the data. These details include TAZ attribution, map matching to roadways, and study area determinations of trip ends for E-E, E-I, I-E attribution. Additional metrics that need to be calculated are speed and time between GPS waypoints. Despite the additional labor and computations involved in the analysis of GPS data, they offer a rich source of routing data and

flexibility as compared to cellular data. Table 2 through Table 4 provide unweighted distributions of total, passenger vehicles and truck trips.

Table 2. GPS Weekday Total Sample.

	E-I	I-E	E-E	I-I	All Types
Total Trips*	208,237	204,552	236,155	2,457,392	3,106,336
Percent Total Trips	6.7%	6.6%	7.6%	79.1%	100.0%
Percent Total External Trips	32.1%	31.5%	36.4%	NA	100.0%

*Unweighted results.

Table 3. GPS Weekday Total Sample Cars/Light Trucks.

	E-I	I-E	E-E	I-I	All Types
Total Trips*	41,443	40,158	20,464	1,128,463	1,230,528
Percent Total Trips**	3.4%	3.3%	1.7%	91.7%	100.0%
Percent Total External Trips**	40.6%	39.3%	20.0%	NA	100.0%

*Unweighted results.
**Total may not sum to 100 percent due to rounding.

Table 4. GPS Weekday Total Sample Medium and Heavy Trucks.

	E-I	I-E	E-E	I-I	All Types
Total Trips*	166,794	164,394	215,691	1,328,929	1,230,528
Percent Total Trips	8.9%	8.8%	11.5%	70.8%	100.0%
Percent Total External Trips	30.5%	30.1%	39.4%	NA	100.0%

*Unweighted results.

MAP MATCHING

Map matching is the process of matching coordinate data to a digital roadway network. Most mobile device users see this process when using routing directions on mapping apps. It is most commonly associated with GPS data, but in this study it also includes the process of matching cellular data to a roadway network. As with GPS data, there is no intrinsic association of the data to a given roadway.

Cellular Data

The map matching process for cellular data utilizes a combination of the CAMPO model network and the Texas Statewide Analysis Model (SAM) network. Stitching the two networks together allows for assigning trips outside of the study area in the external halo zones to the internal network based on the shortest path. The shortest path is typically a measure of distance or time between O-D pairs and the route to take that minimizes one or the other. The analysis for this study uses the shortest time path, which is calculated using the speed limits of the network.

TransCAD is software used in the development of travel demand models. One of its core abilities is to assign an O-D matrix of traffic to a network. Built into TransCAD is an analytical component, termed select link, to obtain the O-Ds of traffic that crosses a particular link during assignment. Accordingly, the analysis methodology uses a select-link analysis of roadway segments located at each of the 58 external stations. The select-link analysis process assigned E-E and I-E/E-I trips to the roadway network, while tracking the O-D of the trips that passed through each external link (see external stations in Figure 1).

Figure 2 shows that the E-E traffic appears logically assigned to the interstates and major roadways that bisect the study area. Figure 3 alternatively shows that I-E/E-I traffic also appears logically assigned to the major roadway network since more of the internal network is used to distribute the internal trips. Overall, the percentage breakdown of cellular derived external traffic per station is logical and consistent.

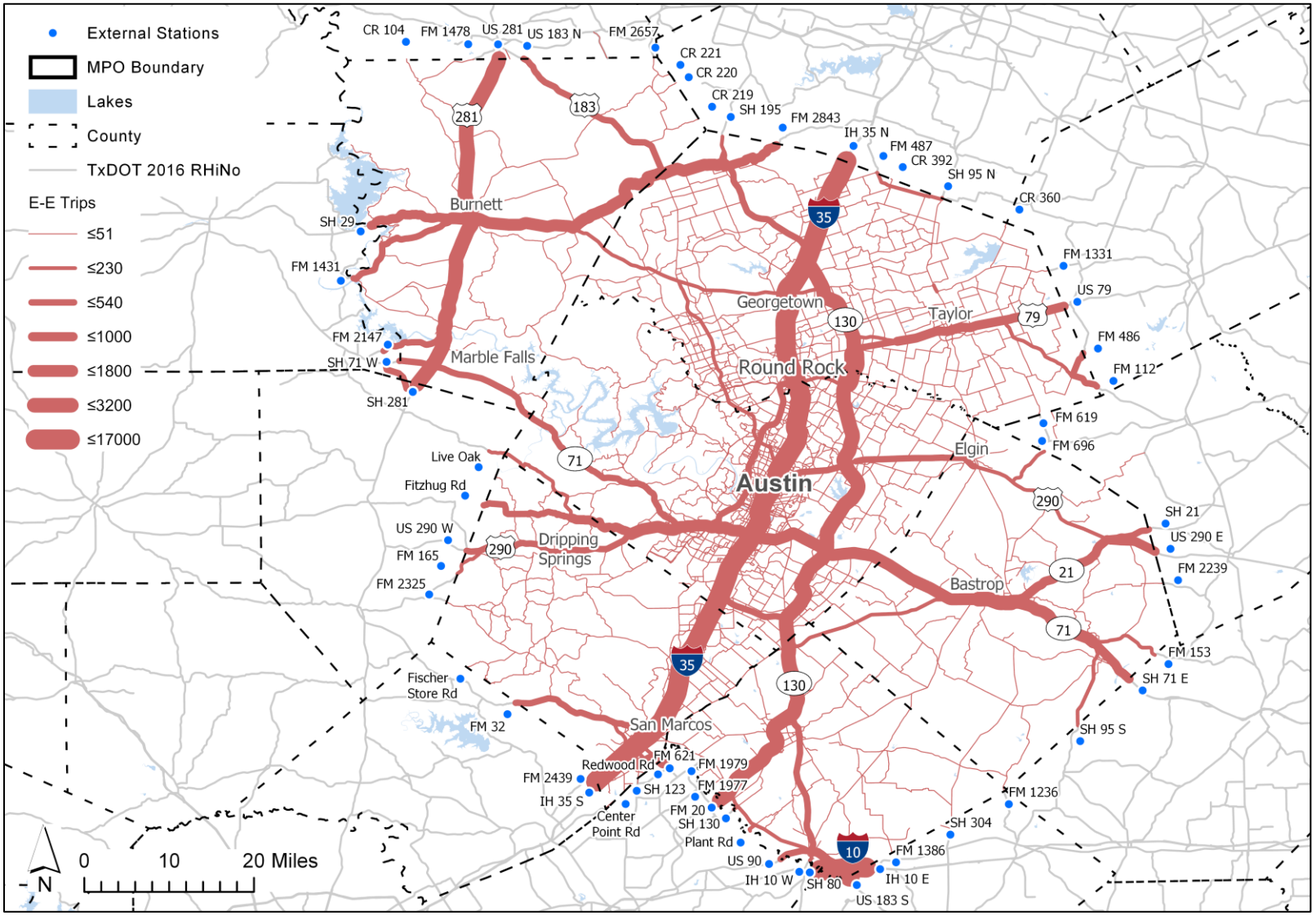


Figure 2. Cellular External Trip (E-E) Assignment Flow Bands.

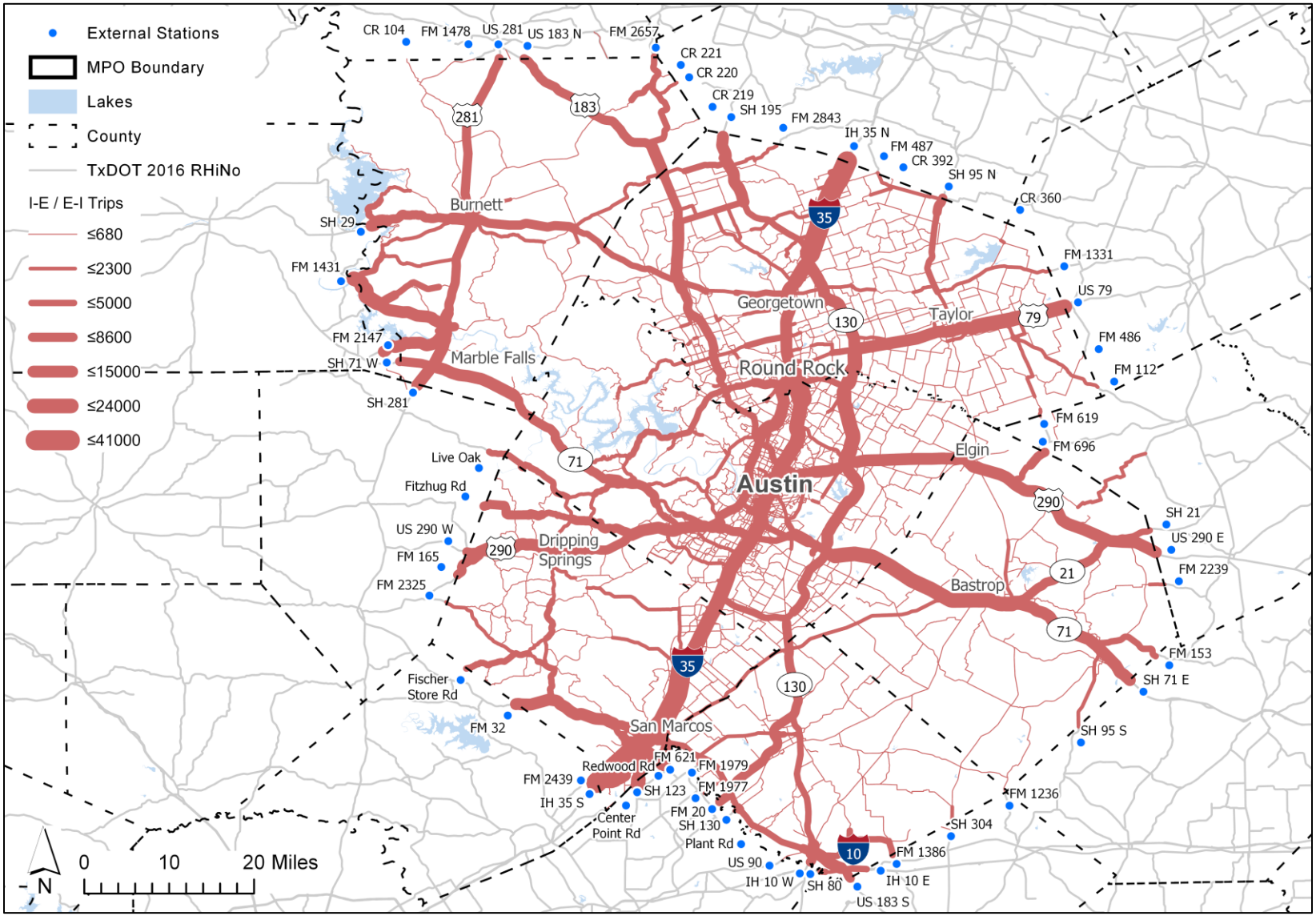


Figure 3. Cellular External-Local (I-E/E-I) Assignment Flow Bands.

GPS Data

Assigning GPS data to distinct road sections and determining ingress and egress external stations has significantly more challenges than those present with cellular data. The positional accuracy of GPS data is very good along with the time scale of the data (i.e., waypoints at 1-minute intervals or less). On the other hand, these data are large and difficult to work with using traditional GIS applications and methods. Data management strategies are key to working with large amounts of GPS data. These strategies include data chunking, efficient sorting of data in memory or on disk, and establishing efficient indexing routines. Figure 4 demonstrates a typical map matching problem using GPS data with waypoints superimposed on a highway system.

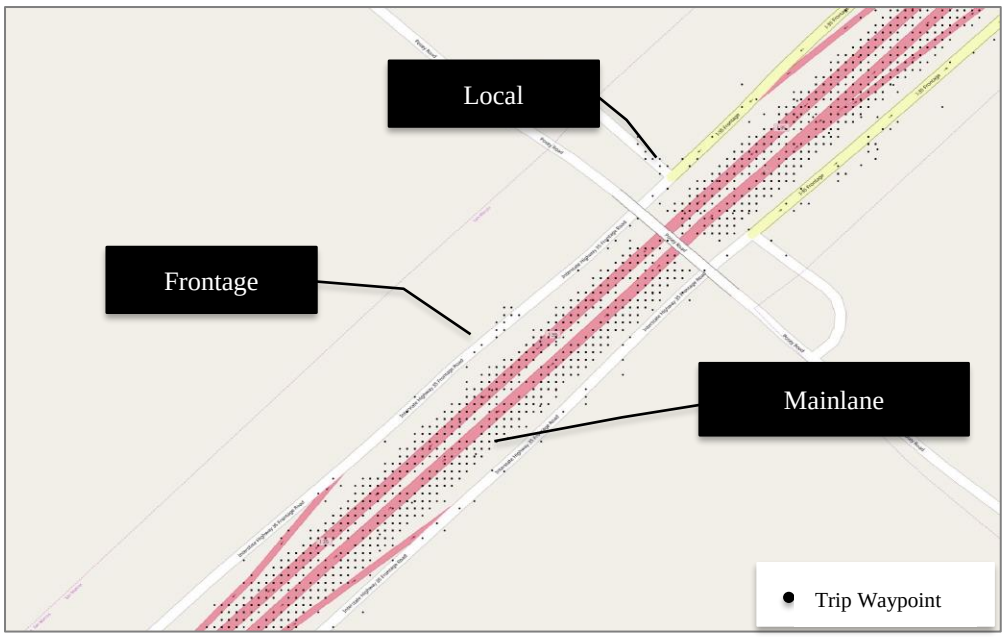


Figure 4. Matching GPS Data to Road Segments.

As illustrated, depending on the network and the precision of the coordinates, each GPS point matches to several different roadway sections. One way of solving this problem is to do a K-Nearest Neighbor (KNN) search for each point. KNN is an algorithm that attempts to locate a limited or unlimited set of results that are closest in space to a given point or object. However, this too has its drawbacks, as the return of the search can quickly become very large. To limit this problem, the algorithm imposes a limit of the three nearest roadway segments within a 100-meter radius.

The results of the KNN search determine the external roadway used when each trip crossed the study area boundary. Figure 5 illustrates an example of this process. The flag “insa” designates whether the point was in or out of the study area, and the flag “gaptest” indicates the first point that was not in the study area. In this illustration, each flagged waypoint cross-references to a table of external stations based on the highway names from the KNN search and highway names in the

external look-up table. The algorithm selects the highway name that matches the KNN search and is closest to the point.

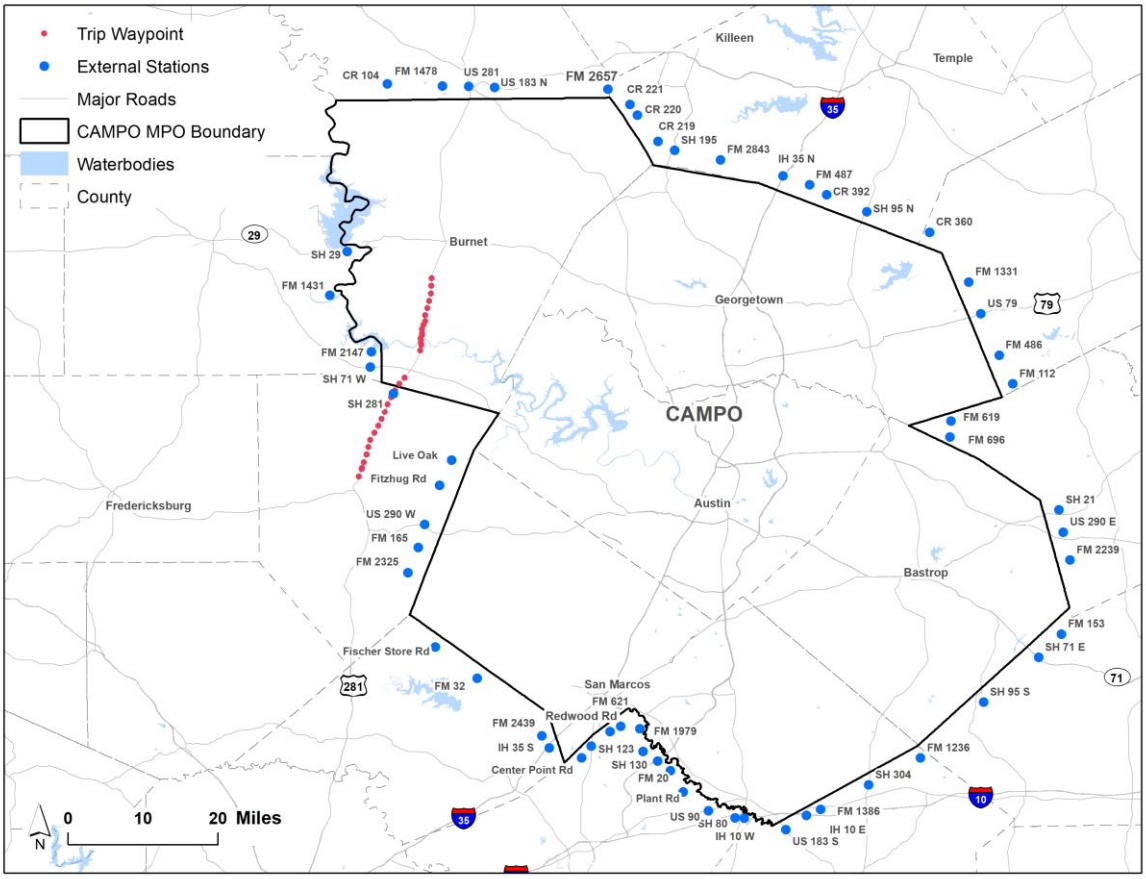


Figure 5. External Station Matching Example.

The results of the external station-matching algorithm indicate the following:

- Most matches are associated with the first KNN result.
- With no initial match, the algorithm searches the subsequent KNN results.
- When this process fails to find a match, the search broadens to include the immediate additional 10 points (+/-5 waypoint sequence) for each flagged waypoint and the results of the KNN for these points.

However, in some cases not all trips pass through one of the external stations in the study area, as illustrated in Figure 6. The reason for this is that travel demand model development does not account for all points of ingress/egress and focuses on major roads and local roads to major attractors. Although these roads are often the most logical routes, drivers often take "shortcuts."

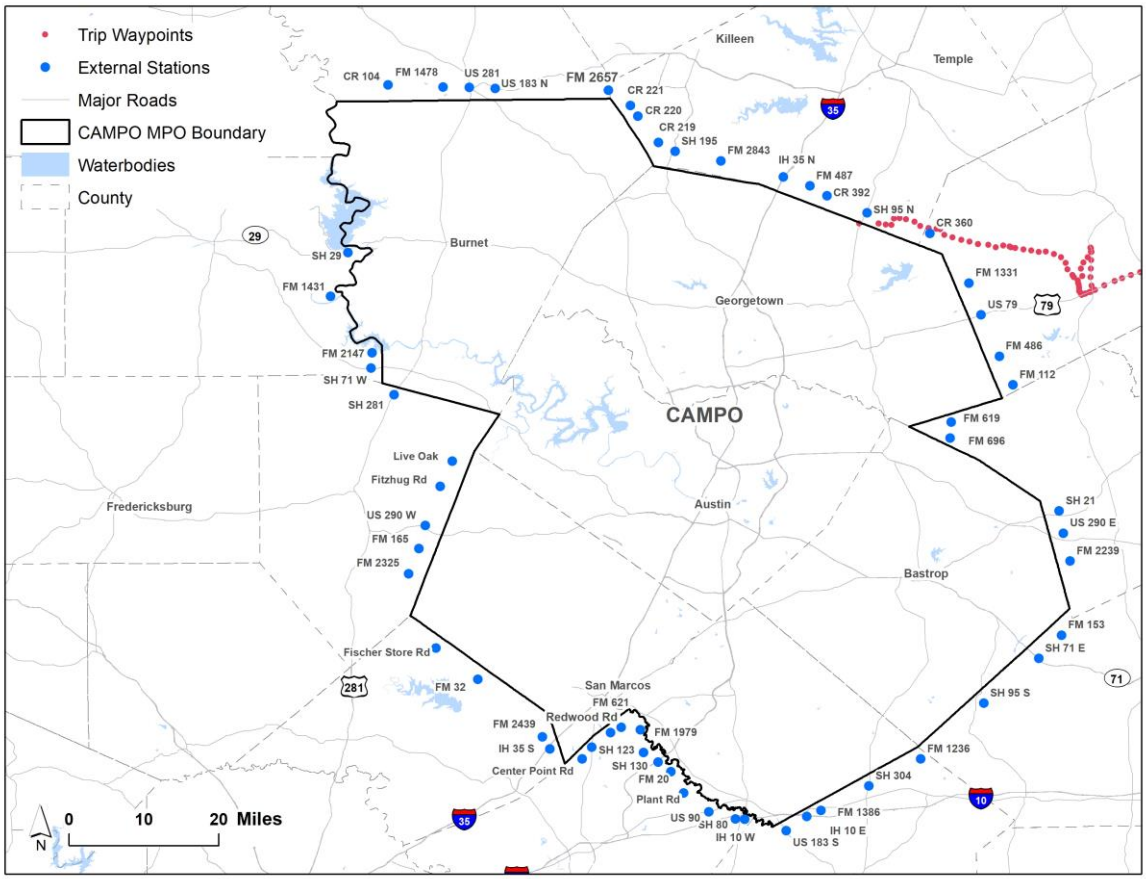


Figure 6. Unmatched External Trips.

DATA EXPANSION

Data expansion is the process of growing sample data to match a known population. Part of data expansion is calculating different weights representing each category of the population relative to the sample, and vice-versa. This process controls for biases in the sample data that may be over-represented. For external analysis, the basis for expansion is traffic counts, which represent the known population. There are many techniques for weighting data. This study uses a process known as Iterative Proportional Fitting (IPF). IPF is a procedure to fix values in a matrix to a known total. The process is iterative because it estimates initial weights. From the initial estimate, the resultant error is calculated, and the process then recalculates the weight using the error until it minimizes a specified minimum error. The IPF procedure seeks to minimize the weighted mean absolute percentage error (MAPE). This objective best controls for errors in higher volume roadways rather than large errors in smaller volume roadways.

Cellular Data

Cellular data from AirSage represents expanded results. AirSage expands cellular data by aggregating households in the cellular data, also known as household clusters, to census tracts. The ratio of household clusters to the total households in the census tract is the penetration rate of the

cellular data. AirSage applies this ratio to each trip made by each household cluster. AirSage provides only expanded results, which the study team post-processes as part of the external analysis. Figure 7 shows the locations that produce and attract the highest number of I-E/E-I trips per square mile within the CAMPO area.

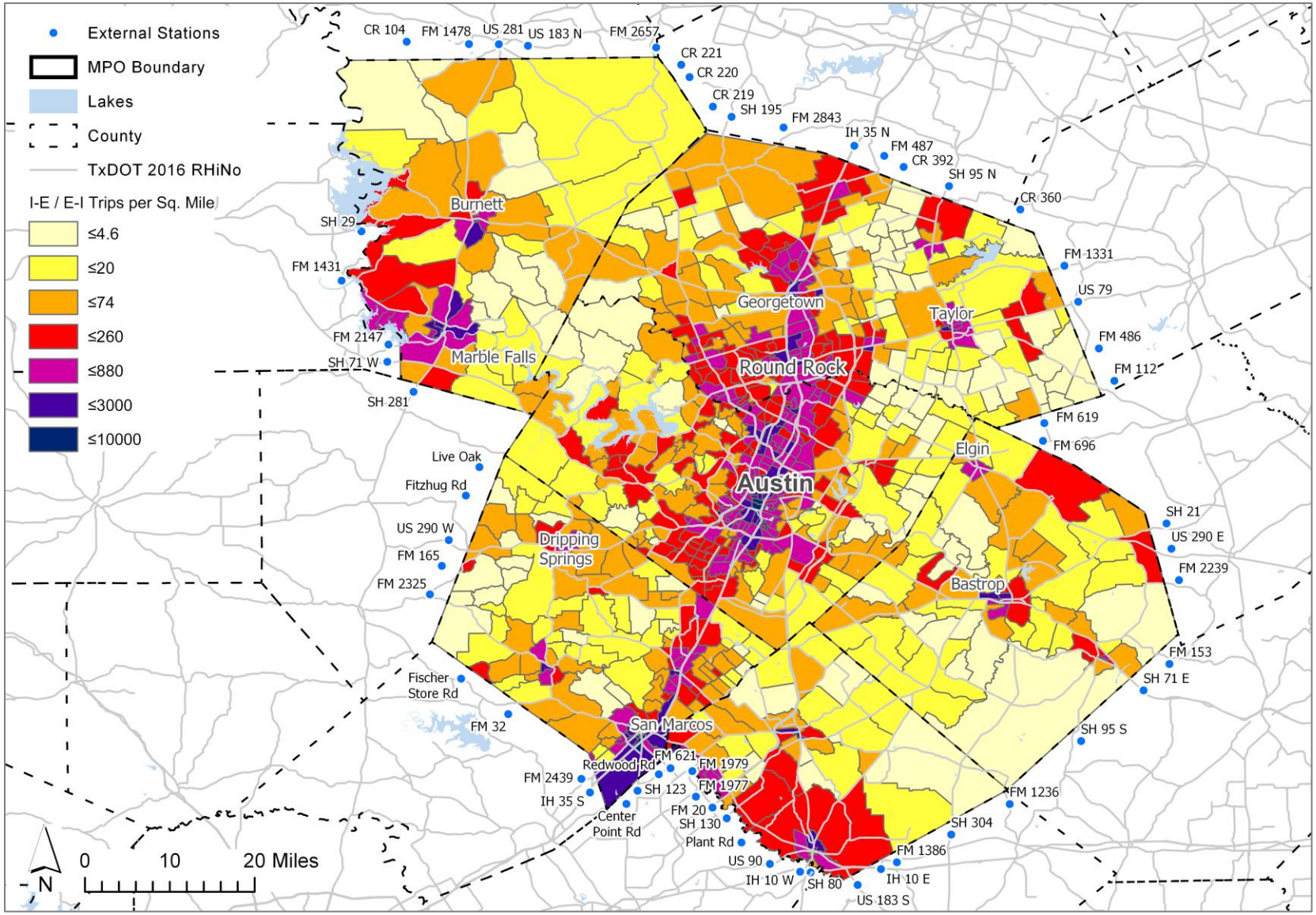


Figure 7. Cellular External-Local Trips per Square Mile by Zone.

Table 5 provides sample size distributions for each of the external roadways for cellular and GPS data, sorted by roadway total volume. In reviewing Table 5, it is important to keep in mind that cellular data represents an expanded sample, which is subject to error. Ultimately, what the cellular data provides is an approximate distribution of trips that would utilize the roadways. In later analysis, this distribution and overages are corrected for by fitting the data to the traffic counts. Presuming that cellular data is mostly passenger vehicles, it is approximately 210 percent of traffic counts. Alternatively, presuming that cellular data is representative of trucks and passenger vehicles, it is approximately 142 percent of traffic counts. AirSage and TTI researchers believe that cellular data has a significant passenger vehicle bias and the proportion of passenger to commercial vehicles in the data is unknown.

The sample size overages contained in the cellular data reflect that the cellular data is device trips and not vehicular trips. In light of this, this study team computes correction factors that negatively expanded the cellular results relative to traffic counts of cars. These cellular correction factors apply only to the external trips in the cellular data.

In Table 5, the average GPS data is based on the total workdays during the study dates, which is 65 of a total of 88 days during the study period from February 2, 2015 to May 2, 2015.

Table 5. Passive Data Sample Size Distributions.

Highway	Count of Cars	Count of Trucks	Total Count	Cellular	GPS Cars	GPS Trucks	% Cellular Sample to Cars	% Cellular Sample to Counts	% GPS Car to Counts (Avg)*	% GPS Trucks to Counts (Avg)*#
IH 35 S	67,729	16,274	84,003	86,805	28,143	208,560	128%	103%	0.60%	20%
IH 35 N	47,960	13,717	61,677	110,827	21,941	201,122	231%	180%	0.70%	23%
IH 10 E	23,065	7,979	31,044	18,965	11,875	103,996	82%	61%	0.80%	20%
IH 10 W	19,693	6,794	26,487	18,746	13,319	107,542	95%	71%	1.00%	24%
US 290 E	10,910	2,057	12,967	15,120	4,849	12,908	139%	117%	0.70%	10%
SH 71 E	10,465	2,029	12,494	12,017	5,097	8,742	115%	96%	0.70%	7%
RM 1431	11,712	616	12,328	19,859	855	1,710	170%	161%	0.10%	4%
SH 123	9,186	1,110	10,296	14,496	1,737	7,936	158%	141%	0.30%	11%
SH 21 E	7,047	1,777	8,824	3,517	2,693	10,423	50%	40%	0.60%	9%
SH 195 N	8,039	756	8,795	12,214	3,865	5,628	152%	139%	0.70%	11%
RM 2147	8,352	440	8,792	14,174	849	1,225	170%	161%	0.20%	4%
US 281 S	8,022	422	8,444	13,384	2,110	6,775	167%	159%	0.40%	25%
US 183 S	6,691	1,187	7,878	11,178	3,490	13,838	167%	142%	0.80%	18%
US 79 E	5,253	1,288	6,541	20,791	1,677	6,248	396%	318%	0.50%	7%
US 290 W	5,893	505	6,398	5,628	2,211	4,144	96%	88%	0.60%	13%
US 281 N	5,373	283	5,656	10,691	1,403	4,777	199%	189%	0.40%	26%
SH 95 N	4,993	646	5,639	7,017	829	5,069	141%	124%	0.30%	12%
SH 29 W	5,287	278	5,565	14,879	905	3,071	281%	267%	0.30%	17%
FM 621	5,239	255	5,494	18,259	289	733	349%	332%	0.10%	4%
SH 130 S	5,064	267	5,331	12,295	2,814	18,101	243%	231%	0.90%	104%
SH 80 S	4,074	802	4,876	7,480	1,404	6,685	184%	153%	0.50%	13%

Highway	Count of Cars	Count of Trucks	Total Count	Cellular	GPS Cars	GPS Trucks	% Cellular Sample to Cars	% Cellular Sample to Counts	% GPS Car to Counts (Avg)*	% GPS Trucks to Counts (Avg)*#
US 183 N	4,362	230	4,592	8,073	1,468	3,561	185%	176%	0.50%	24%
SH 71 W	4,170	219	4,389	6,842	1,361	2,633	164%	156%	0.50%	18%
Redwood Rd	3,852	203	4,055	0	289	505	0%	0%	0.10%	4%
RM 32	3,226	352	3,578	13,138	759	1,175	407%	367%	0.40%	5%
FM 2439	2,659	645	3,304	448	467	1,177	17%	14%	0.30%	3%
FM 696 E	2,074	265	2,339	6,014	130	499	290%	257%	0.10%	3%
FM 2657	1,930	102	2,032	3,638	488	635	188%	179%	0.40%	10%
US 90	1,650	318	1,968	1,093	541	2,325	66%	56%	0.50%	11%
Center Point Rd	1,762	93	1,855	0	305	1,387	0%	0%	0.30%	23%
RM 165	1,675	163	1,838	5,765	621	693	344%	314%	0.60%	7%
FM 1979	1,346	115	1,461	3,108	123	322	231%	213%	0.10%	4%
SH 304 S	1,258	176	1,434	825	638	1,174	66%	58%	0.80%	10%
SH 95 E	962	315	1,277	1,576	225	891	164%	123%	0.40%	4%
FM 20	711	258	969	2,135	146	251	300%	220%	0.30%	1%
RM 2325	837	97	934	1,039	185	224	124%	111%	0.30%	4%
FM 2239 E	740	148	888	1,140	255	674	154%	128%	0.50%	7%
FM 1331	572	89	661	1,245	88	132	218%	188%	0.20%	2%
FM 619 E	552	91	643	180	40	69	33%	28%	0.10%	1%
FM 2843	532	98	630	3,205	118	256	602%	509%	0.30%	4%
Fischer Store Rd	596	31	627	3,194	315	680	536%	509%	0.80%	34%
FM 112 E	512	87	599	2,187	46	235	427%	365%	0.10%	4%
FM 487 N	527	50	577	836	0	0	159%	145%	0.00%	0%
FM 153 E	416	76	492	3,132	62	234	753%	637%	0.20%	5%
FM 486 E	379	93	472	322	71	311	85%	68%	0.30%	5%
FM 1977	363	40	403	1,128	7	17	311%	280%	0.00%	1%
FM 1386 E	332	17	349	3,957	0	0	1192%	1134%	0.00%	0%
CR 221	267	14	281	0	27	77	0%	0%	0.20%	8%
FM 1478	217	11	228	3	7	20	1%	1%	0.00%	3%
Fitzhugh Rd	210	11	221	6,212	307	266	2958%	2811%	2.20%	37%
FM 1326 E	139	38	177	289	0	0	208%	163%	0.00%	0%
CR 220	140	7	147	1,527	5	3	1091%	1039%	0.10%	1%
RM 962	111	6	117	0	43	151	0%	0%	0.60%	39%
Plant Rd	108	6	114	676	324	1,154	626%	593%	4.60%	296%
CR 104	97	5	102	47	173	520	48%	46%	2.70%	160%
CR 219	70	4	74	5	153	146	7%	7%	3.40%	56%
CR 392 N	65	3	68	0	122	268	0%	0%	2.90%	137%
CR 360	41	2	43	603	265	672	1471%	1402%	9.90%	517%
Total	**319,507**	**63,960**	**383,467**	**531,924**	**122,529**	**762,570**	**166%**	**139%**	**0.60%**	**18%**

*The percentage of GPS trips being greater than 100 percent is likely a result of potential inaccuracies in the traffic count.

GPS Data

Table 5 shows that commercial vehicles dominate GPS data. This makes it ideal for developing external truck O-D matrices. Overall, the average sample for trucks was 18 percent, which generally increase as the roadway volume increases. This matches expectations as commercial vehicles make up a greater portion of freeway traffic than non-freeway traffic. The results from cars was not as promising, with a very low average sample size of 0.6 percent. As with trucks, the average sample size increases with volume.

Expansion of GPS results uses the total of all weekday trips observed in the data. This total compensates for empty cells within a trip O-D matrix. This phenomenon occurs due to the lack of observations of trips between an O-D pair on any given day. Using the total trips, rather than a daily average, allows for the capture of trips within these empty cells over time, rather than using a snapshot. As with cellular data, the expansion process uses IPF to calculate correction factors. The procedure expands trucks negatively and cars positively.

O-D External Trip Matrices

The travel demand model directly applies the final O-D trip matrices. These matrices are calculated by developing a ratio of the sample trips to the expanded trips for each external roadway. This ratio is then applied to all O-D pairs associated with that external roadway. For trucks, the O-D matrix only uses the GPS truck results. For passenger vehicles, the resulting trip table is an average of the cellular data. However, the study team's goal is to refine this process using revealed biases and other results.

RESULTS

Table 6 and Table 7 show the expanded results for cellular and GPS data. Overall, the expansion process altered the distribution of external trips, to correct for the differences in the sample and the traffic counts. Additionally, the E-E as a percent of all trips did increase, which is a reflection of the unexpanded data's bias toward trucks.

Table 6. Cellular Area-Wide Weekday Expanded Trips.

	E-I	I-E	E-E	Total	Unexpanded E-E
Total Trips	130,399	130,399	26,621	287,419	33,543
Percent Total Trips*	45.4%	45.4%	9.3%	100.0%	6.7%

*May not sum to 100 percent due to rounding.

Table 7. GPS Area-Wide Weekday Expanded Trips.

	E-I	I-E	E-E	Total*	Unexpanded E-E
GPS Car Trips	116,693	116,693	43,223	276,609	20,464
GPS Truck	15,879	15,879	16,123	47,881	215,691
Total Trips	132,572	132,572	59,346	324,490	236,155
Percent Car Trips*	42.2%	42.2%	15.6%	100.0%	20.1%
Percent Truck Trips*	33.2%	33.2%	33.7%	100.0%	39.4%
Percent Total Trips*	40.9%	40.9%	18.3%	100.0%	36.4%

*May not sum to 100 percent due to rounding.

Table 8 shows a summary comparison of the total and percentage of E-E trips between the GPS and cellular data. The percentage of E-E trips reflects the percent of trips in relation to E-I/I-E trips for the entire study area.

Table 8. Comparison of E-E Trips and Total External Traffic Percentages between GPS and Cellular Data.

Vehicle Category	GPS-Expanded		Cellular Expanded	
	E-E Trips	% E-E	E-E Trips	% E-E
Car	43,223	15.63%	26,621	9.26%
Truck	16,123	33.67%	NA	NA
Total	59,346	18.29%	26,621	9.26%

The percent of car E-E trips for the study area is 15.63 percent and 9.26 percent from cellular data. One potential reason for this difference may be the sparse sample provided by the GPS data.

Expanded External Trips and Recommended Estimations

Table 9 provides expanded, aggregate external estimations as a percent of total trips by station. The recommended estimations are those of the greatest confidence. The external estimations for cars are derived from the cellular results and where the cellular results were inconclusive, the GPS results were used. The external estimations for trucks use the GPS data, as this data is predominantly truck data.

Table 9. CAMPO Expanded External Trips.

Highway	Count of Cars*	Count of Trucks*	Total Count*	% IE+EI Cars (Recommended)	% EE Cars (Recommended)	% IE+EI GPS Cars (For Reference)	% EE GPS Cars (For Reference)	% IE+EI GPS Trucks (Recommended)	% E-E GPS Trucks (Recommended)
IH 35 S#	67,729	16,274	84,003	94.5%	5.5%	79.8%	20.2%	54.2%	45.8%
IH 35 N	47,960	13,717	61,677	92.6%	7.4%	74.4%	25.6%	51.4%	48.6%
IH 10 E	23,065	7,979	31,044	29.3%	70.7%	38.7%	61.3%	25.9%	74.1%
IH 10 W	19,693	6,794	26,487	7.4%	92.6%	11.2%	88.8%	NA	NA
US 290 E	10,910	2,057	12,967	95.9%	4.1%	94.3%	5.7%	89.9%	10.1%
SH 71 E	10,465	2,029	12,494	92.8%	7.2%	95.8%	4.2%	88.9%	11.1%
RM 1431	11,712	616	12,328	98.3%	1.7%	95.1%	4.9%	92.5%	7.5%
SH 123	9,186	1,110	10,296	99.9%	0.1%	88.8%	11.3%	74.6%	25.4%
SH 21 E	7,047	1,777	8,824	90.6%	9.4%	64.4%	35.6%	49.2%	50.8%
SH 195 N	8,039	756	8,795	99.1%	0.9%	95.0%	5.0%	84.1%	15.9%
RM 2147	8,352	440	8,792	97.1%	2.9%	89.5%	10.5%	94.0%	6.0%
US 281 S	8,022	422	8,444	89.3%	10.7%	59.0%	41.0%	60.0%	40.0%
US 183 S	6,691	1,187	7,878	59.7%	40.3%	42.6%	57.4%	39.4%	60.6%
US 79 E	5,253	1,288	6,541	97.1%	2.9%	87.8%	12.2%	73.8%	26.2%
US 290 W	5,893	505	6,398	95.4%	4.6%	93.1%	6.9%	85.4%	14.6%
US 281 N	5,373	283	5,656	77.2%	22.8%	60.6%	39.4%	53.9%	46.1%
SH 95 N	4,993	646	5,639	98.1%	1.9%	89.3%	10.7%	77.4%	22.6%
SH 29 W	5,287	278	5,565	93.8%	6.2%	81.8%	18.2%	72.3%	27.7%
FM 621	5,239	255	5,494	91.0%	9.1%	80.6%	19.4%	69.0%	31.0%
SH 130 S#	5,064	267	5,331	83.3%	16.7%	62.5%	37.5%	45.1%	54.9%
SH 80 S	4,074	802	4,876	99.0%	1.0%	75.6%	24.4%	51.5%	48.5%
US 183 N	4,362	230	4,592	96.8%	3.2%	92.1%	7.9%	88.6%	11.4%
SH 71 W	4,170	219	4,389	95.6%	4.4%	64.8%	35.2%	52.3%	47.7%
Redwood Rd**	3,852	203	4,055	91.3%	8.8%	91.3%	8.8%	73.5%	26.5%
RM 32	3,226	352	3,578	96.3%	3.7%	85.8%	14.2%	68.0%	32.0%
FM 2439	2,659	645	3,304	100.0%	0.0%	82.2%	17.8%	77.6%	22.4%
FM 696 E	2,074	265	2,339	98.6%	1.4%	95.3%	4.7%	88.6%	11.4%
FM 2657	1,930	102	2,032	100.0%	0.0%	95.5%	4.5%	92.2%	7.8%
US 90	1,650	318	1,968	62.2%	37.8%	65.5%	34.5%	74.5%	25.5%
Center Point Rd**	1,762	93	1,855	72.5%	27.5%	72.5%	27.5%	82.6%	17.4%
RM 165	1,675	163	1,838	97.4%	2.6%	98.1%	1.9%	92.7%	7.3%
FM 1979	1,346	115	1,461	100.0%	0.0%	92.6%	7.4%	94.8%	5.2%
SH 304 S	1,258	176	1,434	100.0%	0.0%	49.1%	50.9%	55.8%	44.2%

Highway	Count of Cars*	Count of Trucks*	Total Count*	% IE+EI Cars (Recommended)	% EE Cars (Recommended)	% IE+EI GPS Cars (For Reference)	% EE GPS Cars (For Reference)	% IE+EI GPS Trucks (Recommended)	% E-E GPS Trucks (Recommended)
SH 95 E	962	315	1,277	96.3%	3.7%	81.0%	19.0%	65.0%	35.0%
FM 20	711	258	969	96.3%	3.7%	92.7%	7.3%	87.0%	13.0%
RM 2325	837	97	934	100.0%	0.0%	96.7%	3.3%	95.8%	4.2%
FM 2239 E	740	148	888	100.0%	0.0%	79.1%	20.9%	73.4%	26.6%
FM 1331	572	89	661	95.8%	4.2%	99.0%	1.1%	88.4%	11.6%
FM 619 E	552	91	643	100.0%	0.0%	89.5%	10.5%	75.4%	24.6%
FM 2843	532	98	630	69.5%	30.5%	96.6%	3.4%	93.6%	6.4%
Fischer Store Rd	596	31	627	99.0%	1.0%	91.3%	8.7%	93.3%	6.7%
FM 112 E	512	87	599	42.2%	57.8%	64.7%	35.3%	72.7%	27.3%
FM 487 N	527	50	577	92.0%	8.0%	0.0%	0.0%	0.0%	0.0%
FM 153 E	416	76	492	94.7%	5.3%	85.2%	14.8%	45.5%	54.6%
FM 486 E	379	93	472	1.2%	98.8%	60.2%	39.8%	35.9%	64.1%
FM 1977	363	40	403	100.0%	0.0%	42.9%	57.1%	85.7%	14.3%
FM 1386 E	332	17	349	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
CR 221**	267	14	281	63.2%	36.8%	63.2%	36.8%	100.0%	0.0%
FM 1478	217	11	228	100.0%	0.0%	86.1%	13.9%	100.0%	0.0%
Fitzhugh Rd	210	11	221	92.3%	7.7%	89.4%	10.6%	100.0%	0.0%
FM 1326 E	139	38	177	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
CR 220	140	7	147	100.0%	0.0%	100.0%	0.0%	25.0%	75.0%
RM 962**	111	6	117	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%
Plant Rd	108	6	114	88.7%	11.3%	78.4%	21.6%	100.0%	0.0%
CR 104	97	5	102	100.0%	0.0%	87.5%	12.5%	100.0%	0.0%
CR 219	70	4	74	100.0%	0.0%	94.3%	5.7%	100.0%	0.0%
CR 392 N**	65	3	68	85.3%	14.7%	85.3%	14.7%	100.0%	0.0%
CR 360	41	2	43	100.0%	0.0%	90.0%	10.0%	100.0%	0.0%
Total	**292,375**	**47,697**	**340,072**	**90.7%**	**9.3%**	**84.4%**	**15.6%**	**66.3%**	**33.7%**

*Unique trips is differentiated from total counts (e.g., the total count of cars is not the sum of the count of cars) because each E-E trip crosses two external count stations and is thus double counted. Accordingly, to derive the total unique vehicle count, the E-E trips at each station are divided by a factor of two.

**External estimation uses GPS results.

SH-130 and IH-35

SH-130 is a north-to-south toll road that parallels IH-35 through the CAMPO region. To the south SH-130 intersects IH-10 and connects to IH-35 just north of Georgetown. SH-130 provides congestion relief and travel reliability for the IH-35 corridor. The process used to assign cellular data trips is not designed to control for the impedances associate with toll roads. These include

additional costs it takes to utilize the toll road, nor does it control for the congestion encountered along IH-35 through the CAMPO region. Accordingly, it is recommended that both the cellular and GPS results be considered when developing E-E estimates for the two roadways. It is also recommended that additional research be conducted to better utilize cellular or location based services data for transportation networks with significant toll facilities.