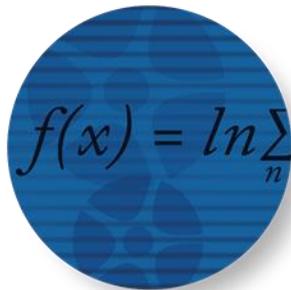


# Synopsis of New Methods and Technologies to Collect Origin-Destination (O-D) Data

May 2016



Better Methods. Better Outcomes.



---

Better Methods. Better Outcomes.

---

# **Synopsis of New/Emerging Methods and Technologies to Collect Origin-Destination (O-D) Data**

**Original:            January 2016**

**Final:                May 2016**

**Prepared for:**

**Federal Highway Administration**

<b>1. Report No.</b> FHWA-HEP—16-083	<b>2. Government Accession No.</b>	<b>3. Recipient's Catalog No.</b>	
<b>4. Title and Subtitle</b> Synopsis of New Methods and Technologies to Collect Origin-Destination (O-D) Data		<b>5. Report Date</b> May 13, 2016	
		<b>6. Performing Organization Code</b>	
<b>7. Authors</b> Ed Hard (TTI), Byron Chigoy (TTI), Praprut Songchitruksa, Ph.D., P.E. (TTI), Steve Farnsworth (TTI), Darrell Borchardt, P.E. (TTI), Lisa Green, Ph.D. (TTI)		<b>8. Performing Organization Report No.</b>	
<b>9. Performing Organization Name and Address</b> Texas A&M Transportation Institute (TTI) 2929 Research Parkway College Station, Texas 77843-3135		<b>10. Work Unit No. (TRAIS)</b>	
		<b>11. Contract or Grant No.</b> DTFH61-10-D-00004 Task Order 5001	
<b>12. Sponsoring Agency Name and Address</b> United States Department of Transportation Federal Highway Administration 1200 New Jersey Ave. SE Washington, DC 20590		<b>13. Type of Report and Period Covered</b> Apr 2015 to May 2016	
		<b>14. Sponsoring Agency Code</b> HEPP-30	
<b>15. Supplementary Notes</b> The project was managed by Federal Highway Administration COR, Sarah Sun, who provided the technical directions.			
<b>16. Abstract</b> This report provides an overview and detail on the use of cellular, GPS, and Bluetooth technologies for origin-destination (O-D) data. It discusses what each technology represents and its capabilities and limitations in relation to accuracy, sample saturation, and frequency. It includes takeaways and lessons learned from numerous studies in recent years that have used cell, GPS, and/or Bluetooth to collect O-D data. The report provides a comparison between the technologies in relation to the ability to provide O-D data by external trip types, by non-commercial and commercial vehicle categories, and by supplemental attributes such as residency status and routing. It discusses the suitability of each technology for planning versus operational O-D studies and for different geographic scales such as urban, regional, or statewide. The report provides potential users of O-D data sourced from cell, GPS, or Bluetooth general guidance on which technology or combinations of technologies is best suited for different O-D study types, sizes, and objectives.			
<b>17. Key Words</b> Origin-destination, cell, GPS, Bluetooth, technology, TAZ, O-D, E-E, external, corridor, passive		<b>18. Distribution Statement</b>	
<b>19. Security Classif. (of this report)</b> Unclassified	<b>20. Security Classif. (of this page)</b> Unclassified	<b>21. No. of Pages</b> 80	<b>22. Price</b> N/A

## Table of Contents

List of Figures .....	v
List of Tables.....	vi
EXECUTIVE SUMMARY .....	vii
<b>PART 1: SYNOPSIS.....</b>	<b>1</b>
<b>1.0 Synopsis of New Technology and Methods to Collect O-D Data.....</b>	<b>3</b>
1.1 Purpose and Objectives .....	3
1.2 Overview of New O-D Sources by Technology .....	4
1.3 Comparisons of Capabilities and Limitation of Each Technology .....	7
1.4 Takeaways from New Technology O-D Studies .....	7
1.5 Findings from Tyler, TX, Study Comparing O-D between Cell, GPS, and Bluetooth.....	9
1.6 Suitability by Study and Data Type .....	10
1.7 O-D by Technology: Advantages and Disadvantages .....	12
1.8 Concluding Remarks.....	12
<b>PART 2: BACKGROUND AND MORE DETAILED MATERIAL SUPPORTING THE SYNOPSIS.....</b>	<b>15</b>
<b>2.0 Background.....</b>	<b>17</b>
2.1 Evolution of New Technology for O-D Data.....	17
2.2 Importance of Understanding New Technology O-D Data.....	18
2.3 TTI’s Role and Experience.....	18
<b>3.0 Understanding Cellular O-D Data.....</b>	<b>20</b>
3.1 What They Are and Represent.....	20
3.2 Sample Size, Collection Timeframe, and Saturation .....	22
3.3 Location Accuracy of Cell Data .....	23
3.4 Processing and Analyzing Purchased Pre-Processed Cell Data .....	28
<b>4.0 Understanding GPS O-D Data .....</b>	<b>32</b>
4.1 What They Are and Represent.....	32
4.2 Sample Size, Collection Timeframe, and Saturation .....	34
4.3 Accuracy Margins .....	34
4.4 Considerations in Scoping an O-D Study Using GPS Data .....	35
4.5 Processing and Analyzing GPS Data.....	36
4.6 GPS-Based O-D Products .....	39
<b>5.0 Understanding Bluetooth O-D Data .....</b>	<b>41</b>
5.1 What They Are and Represent.....	41
5.2 Sample Size, Collection Timeframe, and Saturation .....	44

5.3	Accuracy of Bluetooth Reads .....	44
5.4	How Bluetooth Data Are Processed and Analyzed to Develop E-E Matrices.....	45
<b>6.0</b>	<b>Studies Using Cell, GPS, and Bluetooth O-D Data .....</b>	<b>46</b>
6.1	Cellular O-D Studies .....	46
6.2	GPS O-D Studies.....	50
6.3	Bluetooth O-D Studies .....	52
6.4	Studies Comparing Technologies .....	55
<b>Appendix A: Key Takeaways from Select Studies.....</b>		<b>60</b>
<b>Appendix B: Comparison of Characteristics by Technology.....</b>		<b>62</b>
<b>References .....</b>		<b>65</b>

## List of Figures

Figure 1. Illustration. On-call versus off-call sightings.....	21
Figure 2. Illustration. Handover sighting area. ....	21
Figure 3. Illustration. Clustering process to develop activity points. ....	22
Figure 4. Map. Example data capture areas around periphery of cell O-D study area. ....	26
Figure 5. Illustration. Cell E-E trips without E-I-E filter. ....	27
Figure 6. Equation. Proportionality of external TAZ total results. ....	28
Figure 7. Map. Smith County, TX, cell data proportion of total E-E trips by O-D pair. ....	31
Figure 8. Equation. Balanced O-D matrix. ....	37
Figure 9. Equation. O-D matrix factoring. ....	38
Figure 10. Equation. O-D matrix factoring optimization. ....	38
Figure 11. Illustration. Concept of how Bluetooth data are collected. ....	41
Figure 12. Photo. Installation of a TTI mobile Bluetooth reader. ....	43
Figure 13. Map. Travel routes of vehicles from segment of IH-494/TH 169 corridor. ....	51
Figure 14. Graph. Comparison of through trips on IH-35 in Austin by Bluetooth and ALPR.....	53
Figure 15. Map. Estimated Harbor Bridge O-D flows for major Corpus Christi destinations.....	54
Figure 16. Map. Study area design for collection of cell, GPS, and Bluetooth data in Tyler, TX.	56
Figure 17. Graph. Comparison of Bluetooth, GPS, and cell E-E results by station in Tyler, TX.	57
Figure 18. Illustration. Comparison of cell and GPS E-I/I-E trips in Tyler, TX. ....	58

## List of Tables

Table 1. Summary comparison of characteristics by technology. ....	7
Table 2. Suitability cell and GPS O-D data by study type and use.....	11
Table 3. External survey methods: advantages and disadvantages. ....	12
Table 4. Example output file of Airsage cell data.....	28
Table 5. Bluetooth reads and match percentages. ....	44
Table 6. Percentages of traffic surveyed from roadside interviews. ....	44
Table 7. Bluetooth, GPS, and Cell E-E results in Tyler, TX. ....	57
Table 8. Takeaways from select cell O-D studies.....	60
Table 9. Takeaways from select Bluetooth and/or ALPR O-D studies. ....	61

## EXECUTIVE SUMMARY

**The methods and practice of collecting origin-destination (O-D) data using cell, global positioning system (GPS), and Bluetooth® are still evolving and remain in a state of transition.** While the use of new technology sources for O-D data has indeed advanced and evolved, many questions and uncertainties still remain about the capabilities and limitations of each technology and how they compare, and which technology is best suited for different types of O-D studies.

The purpose of this report is to provide guidance on the use and application of new technology O-D technologies, as well as provide information on the types and scales of studies most suitable to each technology. It is important for agencies and practitioners in making informed decisions on what technology, or combination of technologies, to use for O-D studies.

The report is divided into two parts:

- **Part 1: The Synopsis**, which is Chapter 1 of the report.
- **Part 2: Background and More Detailed Material Supporting the Synopsis**, which includes Chapters 2, 3, 4, 5 and 6 of the report.

The Synopsis is intended to serve as a quick go-to reference for providing guidance on using cell, GPS, and Bluetooth for different types of O-D studies. Prior to purchasing (and investing) in new technology O-D data, it is recommended that potential users:

- Review the comparisons of the data elements related to each technology such as their accuracy, sample penetration, and how trips are defined. See Table 1 in the Synopsis.
- Review the suitability of cell and GPS O-D data by type of study and the various types of ways the data can be applied. See Table 2 in the Synopsis.
- Review the advantages and disadvantages of passive O-D data by technology. See Table 3 in the Synopsis.
- Review the key steps and considerations for scoping an O-D study using cell data. See section 3.3.2 in Chapter 3.
- Review the considerations for scoping an O-D study using GPS data. See section 4.4 in Chapter 4.
- Review the findings from the 2014 Tyler study that compare O-D results between cell, GPS, and Bluetooth O-D data. See section 6.4.1 in Chapter 6.

A review of the tables and sections referenced in the bullet points above will provide an overview of the best uses, capabilities, and limitations of each technology and provide guidance to potential users of the data. The above referenced tables and sections also address the key factors to consider in determining which technologies to use for passive O-D collection. These include:

- The size of the geographic area or corridor: Is it an urban corridor, a regional corridor, a small metropolitan planning organization (MPO) study area, a large metropolitan region, a statewide study?
- The type of study. Is it planning or operational?
- The spatial resolution needed in the data. Will the data need to be analyzed on specific highways or network links, or will it be analyzed in larger geographic areas such as Traffic Analysis Zones (TAZs), major activity centers, or census tracts?

- What is being measured or analyzed? Is it population flows, vehicles trips, or person trips?
- What temporal resolution is needed in the data? Are data needed by time of day, hourly or peak hour, peak period, daily, or for an activity or season?
- Will the study need to disaggregate travel by non-commercial and commercial vehicles?
- Will the study need to disaggregate travel by basic trip purposes?

A generic guideline on which technology(ies) to use for an O-D study cannot be developed since what to use is dependent on study size, type, objectives, and the factors listed above. However, some general guidance on which source to consider can be gained from the following broad generalizations that can be made about the technologies:

- Spatial and Temporal Resolution
  - Cell O-D data are more accurate at aggregated levels; the more they are disaggregated, the less accurate they become.
  - In a typical urban scale, cell O-D data are best applied at an aggregated zone level since they generally do not have the positional accuracy for typical urban network assignment; GPS data are more accurate and can be applied at an urban zone or network levels.
  - Due to its coarse data collection frequency, a three-hour time period is the smallest time increment for which cell O-D data can be provided. GPS data can be provided as low as hourly time increments.
  - Cell and GPS O-D data are based on estimated trip ends; Bluetooth O-D data reflect points where vehicles were detected passing a sensor.
  - Cell data's coarse granularity in space and time makes them very difficult to ground-truth against a user's actual travel route in an urban scale.
- Sample Penetration and Frequency
  - Cell and GPS data are area wide, ubiquitous crowd sourced samples obtained from satellites and cell towers; Bluetooth is point sensor data collected at sensor locations along a highway.
  - Cell data have good sample penetration but low sample frequency; GPS has fair-to-poor sample penetration but has higher sample frequency than cell data.
- Vehicle Types and Bias
  - GPS data can be provided by various sources to distinguish between non-commercial and commercial trips, but cell data cannot.
  - GPS data (at the present time) over-represent commercial vehicles, but since the data are provided in non-commercial and commercial vehicle categories, the commercial bias can be addressed.
  - Researchers suspect that commercial vehicles are underrepresented in cell data and have a bias toward non-commercial vehicles.
  - Cell data can estimate trips by purpose, but GPS data cannot.
  - Third-party GPS data are a viable option for estimating O-D, especially for commercial vehicles. The quality and sample penetration of GPS data will improve over time.
- Other Attributes
  - Cell data can provide trips by basic trip purposes; GPS data cannot. However, TTI researchers believe more study is needed to understand the accuracy of cell-based trip purposes.

- Cell data can provide information on resident, visitor, and commuter travel; GPS cannot.

The experiences and lessons learned from previous research and new technology O-D studies have provided some clarity on the procedures and the best uses and applications of cell, GPS, and Bluetooth for O-D studies. However, more studies, trials, and research comparing the O-D results between these technologies, especially cell and GPS, are still needed. The 2014 Tyler external study comparing these technologies was for a small/mid-sized MPO in Texas with a one-county study area. A similar study is currently underway in the 13-county Dallas-Ft. Worth region of Texas to make the same technology O-D comparisons in a large, metropolitan MPO study area. Key elements and outcomes of this study from the Texas Department of Transportation's (TxDOT's) and the Texas A&M Transportation Institute's (TTI's) perspective include the following.

- The study assesses and compares external to external (E-E) trips between the three sources. Will cell data provide better results for longer distance E-E trips? Cell E-E results from Tyler were markedly low compared to GPS and Bluetooth.
- Will the results for GPS E-E trips compare well enough to Bluetooth E-E trips such that TxDOT can suspend its use of Bluetooth for E-E data collection?
- Will there be an improvement in the sample penetration of GPS data? Will the over-representation of commercial vehicles in GPS data be reduced?
- Will the use of census tracts as the internal TAZ structure for cell and GPS data capture prove to be a satisfactory size for the North Texas Council of Governments' (NCTCOG's) modeling? How will modeling results from this level of aggregation compare to the results of the models current TAZ structure?
- How will NCTCOG's model results compare to those developed using the passive data sources from this study?

Even with advances that have been realized over the past several years, a combination of technologies and providers is still the best or ideal approach for estimating all types and categories of O-D trips and movements. No one technology can collect all elements needed for a comprehensive O-D travel study. However, it is understood that budgetary constraints often preclude the ability to conduct ideal studies. If budget is a constraint, rather than just selecting one technology, potential purchasers of passive data should first explore the possibility of purchasing just the pieces or portions of data from each source that can be combined to best meet study objectives.

New technology data for O-D will continue to change and evolve and reports such as this will need to be updated frequently to stay current with changes and advances.

As previously noted Part 2 of the report provides more detailed information supporting the synopsis in this chapter. Chapters 2, 3, 4, and 5 of Part 2 of this report provide more background and detail about each technology—cell, GPS, and Bluetooth—on what the data truly are and represent, their sample size and penetration, their accuracy, and how each type of data can be processed and analyzed to develop O-D data. Chapter 6 includes examples of studies around the country that have used cell, GPS, and/or Bluetooth data for O-D data collection. These examples provide insights on how different types of O-D studies using passive data were designed and conducted and provide valuable insights on results and lessons learned.



# PART 1: SYNOPSIS



## 1.0 Synopsis of New Technology and Methods to Collect O-D Data

O-D data and information are critical in many transportation planning applications. The methods used to collect O-D data have been changing and evolving over the past decade, and they continue to evolve. Traditional methods to collect O-D data such as intercept surveys have, for the most part, been replaced with passively collected data. The primary technologies being tested and applied in the area of passive O-D data collection include cellular, GPS, and Bluetooth. Significant strides have been made in recent years in developing, refining, and comparing these new technologies for O-D. Cellular data have been used in O-D studies since about 2010, and to date more major O-D studies have been conducted using cell than GPS or Bluetooth. While cell data are a viable source for many types of O-D studies, some concerns remain due to the lack of information and clarity on how cell O-D results are produced.

In recent years, private sector data providers have worked (and continue to work) to develop methods and products using third-party GPS data to develop O-D information. The development of new GPS products to provide O-D data for a study area (similar to the current cell-based O-D product provided by Airsage) in 2015 and 2016 is the most recent advancement in O-D. While cell and GPS technology collect data by crowdsourcing from cell towers and satellites for an entire study area or region, Bluetooth (and Wi-Fi) technology can only collect data from vehicles as they pass by Bluetooth sensors located along a roadway.

The use of Bluetooth for O-D began before cell and GPS data. It was used for travel time and speed measurements before being used for O-D. Bluetooth O-D results have been validated by TTI with travel time and benchmarked with radio frequency identification (RFID) readers. However, the use of Bluetooth for O-D is limited (compared to cell and GPS) since it is based on single point detection equipment and the amount of detection equipment that would be needed for large scale studies. TTI has a long history of developing and using Bluetooth technology for travel time and O-D in Texas. TTI uses Bluetooth O-D data to compare and benchmark cell and GPS developed O-D data.

### 1.1 Purpose and Objectives

The methods and practice of collecting O-D data using cell, GPS, and Bluetooth are still evolving and remain in a state of transition. While the use of new technology sources for O-D data has indeed advanced and evolved, many questions and uncertainties remain about the capabilities and limitations of each technology and how they compare, and which technology is best suited for different types of O-D studies.

*While the use of new technology sources for O-D data has indeed advanced and evolved, many questions and uncertainties still remain about the capabilities and limitations of each technology and how they compare, and which technology is best suited for different types of O-D studies.*

This synopsis (along with background material in Chapters 2 through 6) provides potential users of these data, such as state Departments of Transportation (DOTs) and MPOs, state-of-the-practice information and guidance on the uses and applications of these new passive sources of O-D data. At the present time, which technology to use depends primarily on the level of spatial and temporal resolution needed to meet study objectives and desired level of accuracy. To this end, the objectives of this synopsis are to:

- Provide clarification on what the data collected or developed from each technology represent (e.g., are they trip traces, device movements, or population flows?).
- Provide a better understanding of the current capabilities and limitations of cell, GPS, and Bluetooth O-D data considering each technology's:
  - Data unit collected.
  - Positional accuracy.
  - Sample saturation/penetration.
  - Sample frequency.
  - Continuity of data stream.
  - How trips and trip ends are estimated and defined.
  - Processes used to anonymize data to retain confidentiality.
- Provide guidance on what new technology data or combinations of these data are best suited for different types and sizes of O-D studies.
- Overview the methods and mechanics of how the data are processed and analyzed to develop trip tables and other data elements needed for model input.
- Review key items and considerations in scoping a study using new technology data.
- Show the results of TTI comparisons between Bluetooth, GPS, and cellular O-D data by external trip types and characteristics.
- Provide highlights and lessons learned through TTI's and others' direct experience with cellular, GPS, and Bluetooth data.
- Provide guidance to potential purchasers and users of new technology data for O-D.

The following sections of the synopsis provide a summary overview of the following:

- O-D sources by technology.
- Capabilities and limitations of each technology.
- Takeaways and lessons learned from recent studies.
- Findings from the 2014 Tyler study comparing technologies.
- Suitability of technology by study type.

## 1.2 Overview of New O-D Sources by Technology

### 1.2.1 Cell Data

Cellular O-D data are a measure of estimated device movements or flows between pre-defined geographic areas/zones. The movements are developed based on analysis of mobile device sightings and activity locations over a set time period, ranging from weeks to months. The trips developed from cellular O-D data do not reflect actual trips, but rather the estimated trips derived from analysis of the device's movements and patterns over the subject time period.

Pre-processed cellular data to provide information on travel flows and movements for a prescribed area must be purchased from a private data provider/aggregator. Currently, Airsage, an Atlanta-based wireless information and data provider, is the only company in the United States that can provide large-scale information on device and population movements to study travel based on cellular data.

*The trips developed from cellular O-D data do not reflect actual trips, but rather the estimated trips derived from analysis of the device's movements and patterns over the subject time period.*

The sample size for cellular location data is one of their chief attributes due to the ubiquitous use and high market penetration of cellular devices. The sampling rate, or the percentage of the population that is represented in AirSage's data, is dependent on the market penetration rates of the carriers with whom AirSage has agreements. AirSage documentation says that on average their sighting locations are within about 300 meters in urban areas, but often they are within 50–100 meters.<sup>(1)</sup> However, location accuracy varies based on many factors such as by device

*The sampling rate, or the percentage of the population that is represented in AirSage's data, is dependent on the market penetration rates of the carriers with whom AirSage has agreements.*

type, device quality, and the size and density of cells in the cellular network. Trip estimates from cell data can be developed for average weekdays, weekends, or individual days. Unlike GPS data, cell data cannot be provided by peak hour. Due to its coarse data collection frequency, a three-hour time period is the smallest time increment for which cell data can be provided. In comparing cell O-D data to that of GPS and Bluetooth, it appears cell O-D data may underrepresent commercial vehicle trips.

Since its emergence in about 2010, there has been moderate consumption of pre-processed cell based O-D data in the United States by MPOs and DOTs, but few studies have been conducted to compare cell O-D results to other technologies and validate results. Cell data's coarse granularity in space and time makes it very difficult to ground-truth against a user's actual travel route in an urban scale. AirSage's data do not contain any location information about a subscriber's identity because all of their records have been encrypted to anonymize the specific individual or mobile device information.

Cell data can be used for external surveys, long-distance corridor studies, or studies of population flows between geographic areas/regions such as flows related to major events. Their use in the United States has primarily been for collection of external and internal trip data to develop O-D matrices to support and/or compare to regional modeling results. Cell data cannot discern vehicle type, but they can provide estimates on basic trip purpose, commuters, and residents versus visitors.

### 1.2.2 GPS Data

GPS data points are obtained via satellite trilateration used to determine GPS device location and track movement via time-stamped coordinates. GPS data can be grouped into two categories: primary GPS data and third-party GPS data. Primary GPS data are unprocessed and collected first hand. Third-party GPS data are pre-processed data purchased through vendors such as INRIX, HERE (previously Nokia and NAVTEQ), or TomTom. For GPS data, a dwell time of about 10 minutes is generally considered a trip end candidate. However, ideally multiple thresholds would be best. For example, 10 minutes could be a solid trip end and between 2 and 10 minutes may require supplemental data such as geographic information system (GIS) networks, speed checks, and headings.

GPS-based data can be grouped into non-commercial and commercial categories since they can be provided by different sources such as mobile apps, in-vehicle navigation, and commercial fleet classes. However, unlike cellular data, GPS data cannot be disaggregated by resident and non-resident travel for a study area. Non-commercial categories primarily reflect consumer vehicles, while commercial categories mostly reflect fleet vehicles such as long haul trucking and local delivery or service fleets/vehicles.

Since early 2015, the use of third-party GPS data for O-D purposes has become available as new products have been (and are being) developed that can provide processed GPS O-D data

specific to a study type, scope, geography, and time period. Currently, O-D data from these products can be provided as:

- Trip records – time-stamped trip start and trip end locations.
- Trip records and waypoints or traffic message channels (TMCs) – waypoints or TMCs are provided in addition to the trip records.
- O-D matrices – the matrices are built based upon the consumer-provided zone structure.

The GPS data supporting these products have low sample penetration and a commercial vehicle bias. However, since the data can be provided in non-commercial and commercial categories, the bias can be mitigated. As data providers continue to add new GPS devices and sources, the sample penetration will improve and the commercial bias should be reduced. Waypoints from trip traces included in the GPS O-D data can be used to study routes used between trip ends. The addition of waypoints in the data allows for much greater use and multi-purposing of the data, such as the ability to perform individual or comparative studies of virtually all corridors within a regional data set.

GPS O-D data can be used for external surveys, corridor studies, travel pattern/routing studies, freight/commercial vehicle studies, select link analyses, and origin-destination matrix estimation (ODME). Due to its numerous sources of commercial fleet data, it is a good option for studying commercial vehicle and freight O-D and travel patterns. Trip data from GPS data providers can be developed and extracted for almost any time period since the data collection frequency for GPS data can be as low as a minute or less. The GPS technology is very accurate with recent studies reporting the accuracy of GPS devices on popular smartphones to be in the range of 5–8 meters. However, data purchasers may need to take into account anonymization measures applied to the raw data by the provider that can impact the accuracy of trip ends.

### 1.2.3 Bluetooth Data

Bluetooth is a wireless technology used for exchanging data over short distances. The technology is frequently embedded in mobile phones, GPS, and in-vehicle navigation systems, and each Bluetooth device has a unique alphanumeric identifier known as a Media Access Control (MAC) address. The MAC address of Bluetooth-equipped devices can be read as they pass by Bluetooth sensors located along a roadway. O-D data can be created using Bluetooth by matching MAC addresses between locations where Bluetooth sensor equipment has been installed.

Bluetooth data are collected via portable or permanent readers that are deployed at pre-selected locations. For an external survey, Bluetooth sensors are placed along highways around the periphery of a study area and MAC address readings are matched between external stations. The sample of Bluetooth reads and matches are expanded and balanced to classified traffic counts taken at each external station. The percentages of non-commercial and commercial E-E trips are estimated by applying the vehicle class results obtained from the counts.

While O-D data developed from cell and GPS data are based on estimated trip ends, O-D data from Bluetooth are based on where a device is detected at a sensor. Bluetooth data do not collect trips ends. In light of this, the primary purposes of Bluetooth O-D studies are to study the amount of through movements (E-E trips) for a study area or corridor or in estimating prevailing traffic patterns to assist on decisions related to route alignments and planning. The use of Bluetooth for O-D is generally suited for smaller scale studies, since data collection requires roadside sensors, unlike cell or GPS. For example, the use of Bluetooth to collect local external to internal/internal to external (E-I/I-E) trips within a study area would not be feasible due to the

exorbitant amount of sensors that would be needed and the fact the data do not reflect trip ends.

TTI has a long history of developing and using Bluetooth technology for travel time and O-D in Texas. TTI uses Bluetooth O-D data to compare and benchmark cell and GPS developed O-D data.

### 1.3 Comparisons of Capabilities and Limitation of Each Technology

O-D information can be developed using cell, GPS, and Bluetooth technology, but the O-D data from each source has unique characteristics that lend themselves to different applications. There is some overlap in these characteristics among the technologies, with no one technology possessing strengths in all characteristics. Table 1 provides a summary of the key characteristics by technology

**Table 1. Summary comparison of characteristics by technology.**

O-D Data Element	Technology/Method		
	Cellular	GPS Data Stream	Bluetooth (E-E Only)
Data unit	Cell sighting based on event: call, text, data use/exchange, or network handover	GPS ping; time-stamped coordinate	MAC address of device
Positional accuracy	300 meters (average)	1–10 meters	About a 100 meter range
Data saturation/penetration	Good, but varies	Relatively low	Varies by external station. Ranges from about 3–10 percent.
Sample frequency	Varies widely, in minutes	In seconds or minutes	In seconds
Continuous data stream?	No, random events	Sometimes, but typically pieces of trips captured	Yes, but only in about 100 meters range of reader
How trips and trip ends are estimated and defined	Based on activity points and clusters	Trip based on GPS data stream	MAC address matches between readers. Trip ends cannot be determined.
Anonymization	Encrypted to anonymize individual and device IDs through WISE technology	IDs scrambled and time/distance offsets applied. Actual trip ends may not be provided.	MAC address anonymized at field readers by removal of some digits of address, data aggregated prior to O-D table creation

### 1.4 Takeaways from New Technology O-D Studies

Airsage’s cell-based O-D product offering has been available since about 2009, while the first GPS O-D product offering was introduced in 2015. Because of this, more major O-D studies have been conducted using cell technology than GPS. The use of Bluetooth for O-D data began about the same time as cell, but since it is a point sensor data collection method it is generally limited to smaller scale studies. Takeaways from new technology O-D studies conducted between 2010 and 2016 that are detailed in Chapter 6 are provided in the following sections.

#### 1.4.1 Cell Studies

The majority of cell O-D studies have been for development of trip matrices and related trip characteristics for calibrating or validating model results or for comparing model results to those developed using cell O-D data. In all such studies, the model TAZ had to be aggregated to larger zones to capture cell trip ends due to the low positional accuracy of cell data. Table 8 in

Appendix A details the takeaways from cell O-D studies undertaken in the United States since about 2010. Common takeaways from these studies included the following:

- Cellular O-D data are an acceptable, lower cost option to traditional external O-D studies for use in modeling and long distance corridor studies.
- In comparing model results to cell results, the total number of trips was comparable, though differences become more apparent as the data are disaggregated.
- Important attributes of cell data are large sample penetration and the ability to distinguish between residents and visitors.
- There is uncertainty on how much to aggregate the internal TAZ zones to create new cell zones. One author suggested using the lowest level of resolution that is acceptable and aggregate TAZs accordingly.
- Establishing the size and boundaries for the external zones around the periphery of the study area that are needed to capture cell data is difficult. Check for erroneous trips between these external periphery zones to make sure they are not included in E-E trips.
- Cell data are better suited for estimating E-I/I-E trips than E-E trips. Study results for E-E trips were mixed.
- Cell data may under estimate home based work (HBW) trips, and over-estimate home based other (HBO), and non-home based (NHB) trips. More studies and comparisons are needed on cell-based trip purposes. One should be mindful of impacts on cell-derived trip purposes if there is a large population of students or shift workers in the region.
- Limitations include a lack of demographic and mode information and differences from model results across trip purpose.

#### 1.4.2 GPS Studies

Compared to cell, there are fewer GPS O-D studies because a GPS-based O-D product has only been available since 2015. The early GPS O-D studies performed by TTI were conducted prior to development of a private sector O-D product.

In the 2010–2011 timeframe, TTI studies found that GPS data could be used to develop O-D data needed for various types of transportation studies, but that—at the time—the GPS sample size and penetration levels were much too low for them to be a viable source. However, the 2014 study in Tyler showed that GPS was a viable source for O-D data and that the sample penetration of GPS data had improved since the early studies (an overview of the Tyler study is discussed in section 6.4 of this report, and summary findings and conclusions of this study are provided in section 1.5).

The introduction of INRIX’s Insights Trips O-D product in 2015 appears to be a significant advancement for the use of third-party GPS data for O-D. Takeaways from the Minnesota and Maryland studies, which have used this product are very positive. The Trips product’s ability to include waypoints between O-D trip ends increases the value and utility of the data. Takeaways from recent third-party GPS O-D studies include:

- The data can be used to incorporate real world trip data into a project at a scale not previously feasible.
- The data can be used to qualitatively assess travel markets and routing choices by travelers and employees O-D and routing patterns to calibrate a travel model.
- GPS data containing waypoints is a viable means of studying freight movements on a statewide basis.
- The data can be used to study commercial trips by vehicle weight class and provider profiles such as delivery fleets and private trucking fleets.

- The data can be used to study commercial trip flows between and within zip code zones.

### 1.4.3 Bluetooth Studies

Table 1 summarizes the takeaways from select O-D studies using Bluetooth and automated license plate recognition (ALPR). Bluetooth was found to be a useful technology in several O-D study settings, including external travel studies, corridor analyses with multiple routes, making tolling decisions about tolling along a corridor, determining the best bridge alignment, and assessing the need for a bypass route. As previously noted, the use of Bluetooth for large-scale studies is limited because they are point sensor data, which do not generally determine trip ends. Table 9 in Appendix A provides more detail on the takeaways from Bluetooth studies.

## 1.5 Findings from Tyler, TX, Study Comparing O-D between Cell, GPS, and Bluetooth

This section includes a summary of the findings and lessons learned for the 2014 Tyler study detailed in Chapter 6. It is the only study to date in the United States that was designed to compare the differences between cell, third-party GPS, and Bluetooth derived O-D data. The E-E results for cell and GPS data from the Tyler study were benchmarked against TTI's Bluetooth E-E results. The findings, conclusions, and takeaways from the Tyler study by various categories are provided in the following bullet lists.

### 1.5.1 General Findings

- The lack of positional accuracy in cell data is not well suited for smaller urban and rural TAZs.
- The 500 × 500 meter minimum zone size used to aggregate TAZs and capture cell data was too small. Researchers suggest future studies aggregate TAZs to a larger size. Due to this finding and the fact that Airsage expands its cell data to the census tract level, the 2016 Dallas-Ft. Worth study is using census tracts as the basis for the internal cell zone structure.
- The sample rate/penetration of GPS O-D data was low, especially when compared to cell data sample penetration.
- Accuracy of GPS trip ends and subsequent assignment to TAZs is impacted by anonymization of the data.
- The GPS data had a commercial vehicle bias. Fifty-seven percent of the GPS sample was from commercial/fleet vehicles. The bias was evident in areas and corridors with high commercial/freight activity and functionally higher thoroughfares.
- Cell data may have a non-commercial vehicle bias. Researchers suspect that commercial vehicles are underrepresented in cell data.
- The percentages of resident versus non-residents travel by external station obtained from the cell data compared well to those from Tyler's 2004 External Survey.

### 1.5.2 E-E Trips

- TTI researchers believe that Bluetooth estimates for total (all vehicles) E-E trips were more accurate than those from cell and GPS data since they are based on an actual sample of trips and since the Bluetooth sample collects a more representative mix of commercial and non-commercial vehicles. The percentages of total E-E trips for Bluetooth, GPS, and cell were 27.7, 33.7, and 18.3, respectively.
- In developing estimates of E-E trips, travel time constraints between external zone pairs can be applied using Bluetooth and GPS data, but not for cell data.

- The estimates of total cell E-E trips were low, especially since they were developed based on a 24-hour period, unlike the Bluetooth and GPS results where travel time constraints were applied. The time constraint used was the time it takes to travel between each external station, plus about a 20 percent cushion. The results for E-E trips on the major interstate through the study area (IH-20) were 43 percent, 46 percent, and 23 percent, respectively, for Bluetooth, GPS, and cell.
- Bluetooth and GPS E-E trips compare well when only non-commercial vehicles are considered.
- Bluetooth, cell, and GPS E-E trips are most similar when only non-commercial vehicles are considered. The GPS data over-estimated commercial E-E trips due to its commercial bias.

### 1.5.3 E-I/I-E Trips

- Cell and GPS derived O-D data provide a better sample and distribution of E-I/I-E trips than those of prior traditional methods (e.g., intercept, license mail out, or postcard surveys).
- Due to its good sample penetration, cell data were the best source for total E-I/I-E trips, though GPS data may be better when considering only commercial E-I/I-E trips.
- It appeared cell data provided questionable trip ends in rural areas with poor cell coverage where estimated device locations are less accurate. Compared to the GPS, cell data had a greater number of trips coming from undeveloped areas.
- The similarity of the cell and GPS trip length frequency distributions (TLFDs) for E-I/I-E trips was not statistically similar, despite their appearing to be when charted. There is a greater variation in the TLFD of cell and GPS data when compared by station as opposed to the entire study area.
- The TLFDs need to be developed and compared between each external station to better reveal variations between the technologies.

New GPS O-D products have been introduced since the Tyler study was conducted in 2014. The products can provide waypoints or TMCs between trip ends to help determine routing. This feature was not available when the Tyler study was conducted. TTI researchers have been informed by GPS data aggregators that the sample size, especially for non-commercial sources of data, have improved since the Tyler study.

## 1.6 Suitability by Study and Data Type

Table 2 compares general suitability of cell and GPS technologies by study type and data use. This table was prepared considering the findings and lessons learned from the research and studies reviewed as part of this study. Bluetooth data were not included in this side-by-side comparison since they are point sensor data. However, Bluetooth is still a useful technology in O-D studies, despite not being conducive to large-scale regional studies. The other two O-D technologies—cellular and GPS—are more ubiquitous.

**Table 2. Suitability cell and GPS O-D data by study type and use.**

Suitability by Study and Data Use		Cell Data	Third-Party GPS Data	Comments
External Surveys	E-E trips	limited	✓✓	GPS more comparable to Bluetooth. Limited ability to apply E-E travel time constraints with cell. Several studies have found cell based E-E trips to be low.
	E-I/I-E trips	✓✓	✓	With appropriately-sized TAZs, cell data are best for total trips due to good sample penetration, though GPS may be better for commercial trips.
	Trip purpose	✓	✗	Cell estimates trip purposes based primarily on device's home and work locations. Purpose from GPS data could potentially be imputed based on land use.
	Commuter information	✓	✗	This is valuable information for many urban areas.
	Residency	✓	✗	Resident vs. non-resident splits needed for many models.
	Commercial/Freight	✗	✓	GPS splits O-D data into freight and non-freight sources (cars, apps, and freight categories).
	Demographic information	✓	✗	Based on census tract of the cell device's home location.
	Route information	✗	✓✓	GPS can determine route between O-Ds using waypoints or TMCs.
	Ability to apply travel time constraint	✗	✓	Typically needed to develop E-E trips/matrices.
Corridor Studies	Within urban areas (operational)	✗	✓✓	Cell not well suited due to low positional accuracy.
	Within urban areas (planning, select link analysis)	limited	✓✓	GPS has better positional accuracy, can provide directionality. GPS has ability to constrain data to corridors in urban settings.  Cell sample size makes it best for total traffic, plus it can inform on residents, visitors, commuters, etc. However, GPS needed for freight. Depends on study objectives.
	County to county	✓✓	✓	
	Multicounty metro regions	✓✓	✓	
	Between major metro areas	✓✓	✓	
Hourly or peak hour	✗	✓		
Time Period Options	Peak period	✓	✓✓	GPS better suited since it is collected in more frequent time increments.
	15 minute bins	✗	limited	Sample frequency and size probably too small to provide cell data in this time increment, same may be true for GPS
	Average weekday, weekend, etc.	✓	✓	Either is fine.
	Population/human activity movements	✓✓	✗	Good use of cell data.
Miscellaneous	Statewide O-D	✓✓	✓	Cell best due to sample size, but GPS needed for commercial/freight.
	ODME	✓	✓	Cell is best for regional estimations and GPS could work well for urban corridor and/or microsimulation studies.
	Travel time	✗	✓✓	GPS due to good accuracy and frequency of data points.
	Travel speed	✗	✓✓	GPS due to good accuracy and frequency of data points.
	Traffic operations studies	✗	✓✓	GPS due to good accuracy and frequency of data points.
	Freight studies	✗	✓	Current GPS data are biased toward freight. Appears to be promising source for freight planning studies.

### 1.7 O-D by Technology: Advantages and Disadvantages

The previous subsection highlighted the capabilities and limitations of cell and GPS data for key characteristics of O-D studies. Table 3 further summarizes the advantages and disadvantages—some of which have been previously discussed—associated with each technology. Despite some redundancy from the previous section, this table has been included to aid in drawing comparisons across the technologies and to fill in some gaps that were not discussed in previous sections. This information may be helpful to practitioners trying to decide which technology will best meet the needs of a given project.

**Table 3. External survey methods: advantages and disadvantages.**

8.5	Advantages	Disadvantages
<b>Cellular data</b>	<ul style="list-style-type: none"> <li>• Ease of implementation</li> <li>• Lower cost alternative to traditional O-D data collection</li> <li>• Good data saturation/penetration</li> <li>• Widespread geographic coverage</li> <li>• No limit on study time period (but time periods must be greater than three hours)</li> <li>• No equipment to purchase, deploy, or retrieve</li> <li>• Can estimate resident, non-resident commuter trips</li> <li>• Ability to estimate trip purpose</li> <li>• Good source for E-I/I-E trips with proper TAZ aggregation</li> </ul>	<ul style="list-style-type: none"> <li>• Concerns for accuracy of data at smaller geographic scales</li> <li>• Inability to provide route at smaller scales</li> <li>• Inability to distinguish between non-commercial and commercial vehicles</li> <li>• Unknowns about how results/outcomes are developed</li> <li>• More apt to collect trip chains and miss short trips</li> <li>• Low sampling frequency</li> <li>• Inability to provide time constrained E-E trips; difficulty in isolating E-E trips</li> </ul>
<b>GPS data</b>	<ul style="list-style-type: none"> <li>• Ease of implementation</li> <li>• Lower cost alternative to traditional O-D data collection</li> <li>• Good spatial and temporal resolution</li> <li>• No equipment to purchase, deploy, or retrieve.</li> <li>• Ability to identify routes between trip ends</li> <li>• No limit on study time period</li> <li>• Results can be applied to both urban TAZs and networks</li> <li>• Can distinguish between non-commercial and commercial vehicles</li> <li>• High sampling frequency</li> <li>• Viable source for E-I/I-E commercial trips</li> <li>• Good source for E-E trips, especially commercial</li> </ul>	<ul style="list-style-type: none"> <li>• Low data saturation/penetration in relation to traffic stream</li> <li>• Current bias toward commercial vehicles</li> <li>• Anonymization reduces accuracy of trip ends</li> <li>• More apt to collecting only portions of trips when navigation session in use; misses portion of trip when navigation turned off</li> </ul>
<b>Bluetooth data</b>	<ul style="list-style-type: none"> <li>• Collects samples of actual E-E trips</li> <li>• Good for quick smaller scale studies if equipment is available</li> <li>• Data available in real time and/or immediately after study</li> </ul>	<ul style="list-style-type: none"> <li>• Collects point sensor data</li> <li>• Data do not identify trip ends</li> <li>• Inability to collect, E-I/I-E trips; collects E-E data only</li> <li>• Cannot distinguish between non-commercial and commercial vehicles (though these splits can be estimated based on class counts)</li> <li>• Usually requires field work and equipment installation</li> </ul>

### 1.8 Concluding Remarks

This chapter represents a synopsis of the report. It is intended to serve as a quick go-to reference for providing guidance on using cell, GPS, and Bluetooth for different types of O-D studies. Prior to purchasing (and investing) in new technology O-D data, it is recommended that

potential users review this synopsis to gain an understanding of the capabilities, limitations, and suitability of cell, GPS, and Bluetooth in relation to the type of study under consideration.

Section 2 of this report provides more detailed information supporting the synopsis in this chapter. Chapters 2, 3, 4, and 5 of this report provide more background and detail about each technology—cell, GPS, and Bluetooth—on what the data truly are and represent, their sample size and penetration, their accuracy, and how each type of data can be processed and analyzed to develop O-D data. Chapter 6 includes examples of studies around the country that have used cell, GPS, and/or Bluetooth data for O-D data collection. These examples provide insights on how different types of O-D studies using passive data were designed and conducted as well as provide valuable insights on results and lessons learned.



## **PART 2: BACKGROUND AND MORE DETAILED MATERIAL SUPPORTING THE SYNOPSIS**



## 2.0 Background

### 2.1 Evolution of New Technology for O-D Data

Data and information on the origin and destination of travel and human activity is a core component in transportation planning and modeling. O-D data are needed and used in many transportation planning studies such as external surveys, household surveys, corridor studies, select link analyses, freight movement studies, and studies on long-distance travel and population flows. The need for O-D data for long-distance travel is especially important since these trips are typically not captured in urban or regional travel surveys.

The methods used to collect O-D data have been changing and evolving over the past decade. Traditional fielding methods such as roadside intercept surveys and video license capture methods are now dated and less viable (for many applications) due to concerns such as traffic safety and delay, privacy, respondent burden, and cost. Over the past 6–8 years, new technology methods such as Bluetooth, cellular data mining, and analysis of GPS data have emerged as new methods for collecting O-D data. During this time, a considerable amount of research and studies have been conducted using these new methods to collect and estimate O-D data for various types of studies. More recently, the use of Wi-Fi technology is being studied as a possible source to obtain O-D data.

The majority of the efforts using new technologies have been to obtain O-D data to develop base year trip tables for regional travel forecasting models, to provide seed matrices for corridor simulation analyses, or to provide O-D patterns for bypass and/or route alignment studies. Most of the uses have been to collect O-D data within urban or regional planning geographies. Fewer new technology uses have been to collect O-D for long-distance or statewide travel, though use of new technology for these purposes has been increasing in recent years.

While cellular data have been used for O-D studies (primarily external surveys) for many years, concerns exist about its accuracy when applied in small geographic areas and about the lack of detailed information and black box data used to develop cell-based O-D estimates. There is also a lack of clarity and readily available information on what *specifically* the data collected from cellular represent and how they compare to what is developed from other technologies.

Within the past few years, several private sector data providers have been working (independently) on ways and methods to develop O-D data for transportation planning applications from GPS data. The source of these GPS data is from proprietary traffic or navigational apps and from GPS data purchased from third parties. In 2015, INRIX introduced its Insights Trips product developed to provide O-D data. In early 2016, HERE introduced its O-D product called Trip Data. TomTom has also developed capabilities for developing GPS-based O-D data for transportation planning purposes. As with cell O-D data, there are some questions and uncertainties about O-D data obtained from pre-processed GPS data, such as how robust are the data, what sources are used, to what extent are the source data between the providers the same, and where do they differ.

The use of Bluetooth for O-D began before cell and GPS data. It was used for travel time and speed measurements before being used for O-D. Bluetooth O-D results have been validated by TTI with travel time and benchmarked with RFID readers. However, the use of Bluetooth for O-D is limited (compared to cell and GPS) since it is based on single point detection equipment and the amount of detection equipment that would be needed for large-scale studies.

## 2.2 Importance of Understanding New Technology O-D Data

Today, new technology O-D data for transportation planning and operations purposes can be purchased from numerous private sector vendors that use cellular, GPS, or Bluetooth technology. It is important that users and purchasers of these O-D data understand that these technologies have different capabilities and limitations, which may make O-D data from one technology more suitable than that of another technology. Which type (or types) of O-D data to purchase depends on which source can best address the study's scope and objectives and yield the most accurate or desired results.

While the state-of-the-practice in new technology O-D data is still evolving, general guidance on which technology to use for different types of studies and applications can be provided from research, experience, and lessons learned from MPOs, DOTs, and consultants who have used the data, and from technology data comparisons completed to-date.

Before significant funds are expended to invest in these data, purchasers should examine prior studies and research to gain insights into which technology source (or combinations thereof) may be best suited for the scope and objectives of their study. Subsequent sections of this report will also provide guidance on which technology is best suited for different types of studies and applications.

## 2.3 TTI's Role and Experience

TTI's role in the evaluation and assessment of new technology methods for O-D data is that of an independent, impartial third party. TTI's primary objective has been, and remains to be, to provide clarification and guidance to MPOs, DOTs, and the transportation sector on what new technology data or combinations of these data are best suited for different types and sizes transportation studies that rely on O-D data. The agency has researched and performed numerous studies, trials, and/or field tests on new and emerging methods and technologies to collect and/or compare travel data using cellular data, primary (raw, unprocessed) GPS data, secondary (processed) GPS data, Wi-Fi data, Bluetooth data, and/or ALPR and web surveys. In 2014, TTI conducted a first-of-its-kind field study that compared cellular E-E and E-I/I-E movements to Bluetooth E-E data and private sector GPS data.

### 2.3.1 Work with Private Sector

Over the past six to eight years, TTI has worked with data from numerous private sector data aggregators including Airsage, INRIX, HERE, and TomTom. As part of these efforts, TTI has worked with technical experts from each of these vendors on criteria and evaluation methods for development O-D information from their data. TTI worked with Airsage to develop spatial and temporal criteria in developing O-D data and in troubleshooting anomalies arising from their data. Having worked with Airsage and analyzing its data for research purposes, the agency has a good understanding of what the data represent, generally how Airsage develops trips/flows, and how the data are processed and expanded.

Since 2011, TTI has had numerous meetings and discussions with INRIX, HERE, and TomTom related to potential GPS-based O-D product development, providing feedback on data review/assessment, detail on what consumers of O-D data need, and general specifications and criteria for use in processing O-D data. The exchanges included sharing ideas on development of methods, approaches, and criteria in development of GPS O-D data for transportation planning purposes. TTI first worked with TomTom in 2011 on a trial O-D project that provided new insights on the use GPS for O-D. In 2014, TTI's work with INRIX on development of external O-D data for a project in Tyler, TX, aided this firm in development of their new Insights Trips O-D product.

### 2.3.2 TTI Development of Bluetooth

TTI uses private sector sources for cellular and GPS data, but the agency collects its own Bluetooth and Wi-Fi data. It has been involved in the operation of an Automatic Vehicle Identification System (AVI) using electronic toll tags to provide real-time travel time and speed information since 1996. In 2008, the agency developed processes to effectively read Bluetooth signals and associated host software as a means to cost-effectively expand Houston's AVI system and provide more flexibility in collecting travel data. The unique process developed by TTI provides for more Bluetooth reads than standard processes and an application for a U.S. patent has been filed. The TTI Bluetooth product has been commercialized and more than 3,000 field installations exist worldwide. The accuracy of data from TTI's mobile and permanently installed Bluetooth readers has been vetted through ground-truthing and many years of tests and comparisons to AVI and Wi-Fi data. For this reason, where possible, TTI benchmarks O-D data developed from cellular or GPS sources against O-D data derived from TTI's Bluetooth devices as one means of evaluating and benchmarking O-D results.

Much of TTI's work in studying and using new/emerging technology for O-D has been sponsored by TxDOT as part of an effort to integrate state-of-the-practice methods into the agency's robust statewide travel survey program. However, no TxDOT funds were used in developing TTI's unique Bluetooth technology. The need for up-to-date external data for Texas MPOs has been TxDOT's primary impetus for sponsoring this work.

### 2.3.3 New Technology, Passive Data Are the Future

Despite the aforementioned concerns and limitations of new technology data, they are clearly the future of transportation data and they will ultimately replace or complement most traditional sources of data (and is already doing so in some cases). Purchased third-party data have many advantages such as sample size, geographic coverage, passive collection, no limit on study time periods, and many others.

## 3.0 Understanding Cellular O-D Data

### 3.1 What They Are and Represent

Cellular O-D data are a measure of estimated device movements between pre-defined geographic areas/zones. The movements are developed based on analysis of mobile device sightings and activity locations over a set time period, ranging from weeks to months. The trips developed from cellular O-D data do not reflect actual trips, but rather the estimated trips derived from analysis of the device's movements and patterns over the subject time period. Proprietary algorithms impute the estimated trips of a device based on its patterns between its home, work, and other activity locations. The movements of each sampled device in the study area are estimated individually and then expanded and aggregated to develop the total number of cell-based O-Ds for the study area's pre-defined geography.

*The movements of each sampled device in the study area are estimated individually and then expanded and aggregated to develop the total number of cell-based O-Ds for the study area's pre-defined geography.*

#### 3.1.1 Acquisition of Cellular Data

Pre-processed cellular data to provide information on travel flows and movements for a prescribed area must be purchased from a private data provider/aggregator. With the exception of research collaborations, raw cellular data are generally not made available from cellular/wireless carriers due to privacy concerns and confidentiality agreements they have with their subscribers.

Currently, Airsage, an Atlanta, GA-based wireless information and data provider, is the only company in the United States that can provide large-scale information on device and population movements to study travel based on cellular data. Airsage has an agreement with one or more nationwide cellular device carriers (e.g., Verizon, AT&T, Sprint, T-Mobile), which allows it to access wireless signaling data with the legal stipulation that the raw, unprocessed data be protected and not shared.

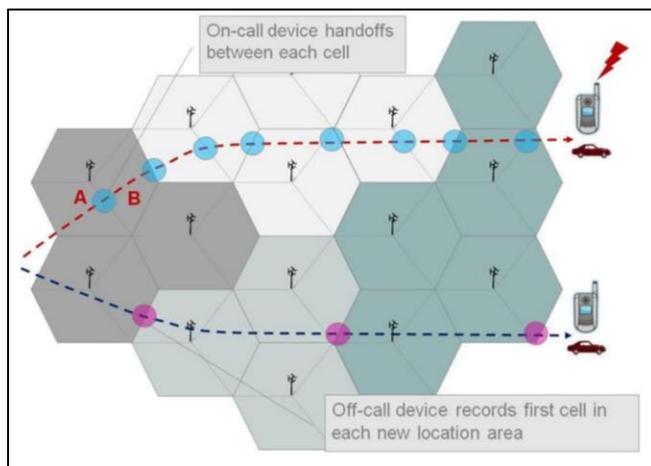
Airsage is the only source in the United States for large scale cell-based O-D data, because it is the only company that has an agreement with a mobile device carrier in the United States. There are other companies abroad that provide cell-based O-D information similar to Airsage because they have access to the cellular wireless carrier data. For example, INRIX provides cell-based O-D data and services similar to those of Airsage in the United Kingdom (UK) and many European countries.

Another source for collecting cell phone location data, but to a much smaller scale, is through active pinging of cell devices or tablets. This active approach, unlike the passive Airsage approach, must be used on a group of study participants who have agreed to allow their devices to be pinged to in order to identify their location. Cellint, headquartered in Tel Aviv, Israel, is one such provider of this active cell pinging service. According to their website, Cellint's patented technology uses pattern matching analysis on anonymous, real-time data extracted from the signaling links of mobile networks for all active mobile phones.<sup>(2)</sup> Their primary service is to provide real traffic monitoring of travel times and speeds using cellular data. The cell phone pinging approach has been used as an alternative means to evaluate accuracy of cell phone data. In 2013, NCTCOG used Cellint data in a study to compare the locational accuracy of trip data obtained from GPS loggers, a smartphone app, and cell triangulation.<sup>(3)</sup>

### 3.1.2 Cell Network Data, Activity Points, and Device Movements

Cellular network data are data derived from the interaction of mobile devices with the cellular network. They are generated each time the device interacts with the network. These data are either event driven or network driven. Mobile devices generate event-driven data each time the device interacts with the network, which can be when a call starts, a call ends, a text is sent or received, or during a data session/exchange. Event-driven interactions are also referred to as Call Detail Records. Network driven interactions include passive transmission of pilot signals between the mobile device and the Base Transceiver Station, which determines best connectivity and handovers when a device switches between two cell areas.<sup>(4)</sup> According to AirSage, as long a mobile phone is active on the cellular network, it receives wireless signals and uses them to anonymously determine locations. AirSage uses its Wireless Signal Extraction (WISE™) technology to aggregate and analyze signaling data. The WISE technology anonymizes the data stream and performs multiple layers of analysis to monitor the location and movement of mobile devices and the population of mobile users.<sup>(5)</sup>

Cell sightings, or signal interactions within the network, are opportunistic in nature since they occur randomly based on the activity of the user. Sightings can be generated when a cell device is active or on-call and interacting with the network or idle and off-call. When a device is active or on-call, its sightings are more numerous and useable in developing location-activity clusters. When a device is idle or off-call, its sightings are less frequent. A handover sighting occurs when a device is moving and changes from one cell coverage area to another. Handover sightings, by themselves, are not that useful in determining activity locations since they are transient points. Figure 1 and Figure 2 illustrate the concept of on-call versus off-call, and handover sightings. Figure 1 illustrates the frequency of on-call versus off-call sightings. The red rectangular box in Figure 2 shows a handover area, where a device in a moving vehicle (for example) would be switched from one cell coverage area to another.



**Figure 1. Illustration. On-call versus off-call sightings.**



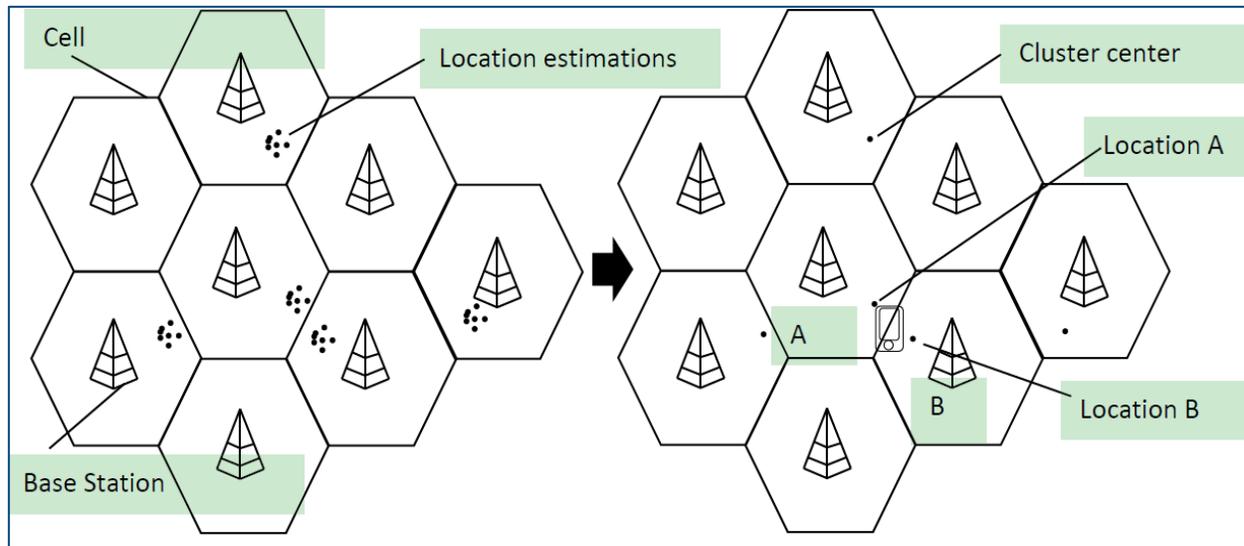
**Figure 2. Illustration. Handover sighting area.**

Source: (6).

### 3.1.3 Development of Cell O-D Data

To develop cellular O-D data, device movements are determined by identifying and analyzing device activity points developed based on the aggregation of individual cell sightings into

clusters. Activity points are created when a device remains in the same location for over five minutes. They represent the center of the clusters, which are based on a spatio-temporal analysis of the locations of all sightings in the cluster. Their locations are further refined and analyzed for arrival and departure times at the location and the duration of time the device remained at the location.<sup>(9)</sup> Figure 3 conceptually illustrates the clustering process and activity points in relation to cellular base stations. The home, work, and other main activity locations of the device are assumed based on day and night time clustering locations over a period of 2–6 weeks. The data are run through a series of pattern recognition and statistical clustering algorithms to identify activity points, and trip imputation algorithms are used to estimate trip ends based on stop dwell times, speed, and geographic proximity to other points.<sup>(5)</sup>



Source: (7).

**Figure 3. Illustration. Clustering process to develop activity points.**

Home locations are those locations where the mobile users spend the majority of their nights, with night time being defined between 9:01 p.m. and 6:00 a.m. Similarly, work locations are the locations where the user spends the majority of their days between 9:00 a.m. and 5:00 p.m.<sup>(5)</sup> TTI's O-D study in Tyler (Smith County), TX, using cell data defined the home and work locations based on an individual subscriber's day and night time clustering over a period of 14 days.<sup>(8)</sup>

Other non-home, non-work activity points for the device known as end points are determined based on a cluster duration that is stationary for five or more minutes. These end points in effect represent trip ends, which can be either an origin or a destination. Using the non-home/non-work activity trip ends, trip legs are developed around the device's home and work locations to estimate the device's trip patterns by predetermined time periods (e.g., average weekday, peak hour, weekend) based on the entire study time period.

### 3.2 Sample Size, Collection Timeframe, and Saturation

The sample size for cellular location data is one of its chief attributes due to the ubiquitous use and high market penetration of cellular devices. Airsage documentation claims they anonymously collect and analyze real-time mobile signals to produce over 15 billion anonymous locations every day in the United States from over 100 million mobile devices.<sup>(9)</sup> Similarly, INRIX

indicated it probes 1.5 billion events per day in the UK.<sup>(10)</sup> Any unique device/location combination may be represented numerous times, sometimes by many orders of magnitude. In actuality, only a small percentage of these sightings are used in development of O-D data. While it is not known exactly what percentage of the total anonymous locations is actually used, the sample size remains very large relative to traditional means of data collection.

The sampling rate, or the percentage of the population that is represented in AirSage's data, is dependent on the market penetration rates of the carriers with whom AirSage has agreements. For their *Nationwide Commute Report*, AirSage reports a sample size of the most reliable devices for the report ranging between 15–25 percent and that the sample size can vary across geographies, such as higher samples in urban areas and lower samples in rural areas.<sup>(11)</sup> In AirSage O-D studies in Moore County, NC, and Smith County, TX, about 17 percent of the residents within these respective study areas were sampled. INRIX indicates that in the UK their cell-based sample is 25–30 percent of the population.<sup>(10)</sup>

The penetration rate, simply put, is the ratio of the number of observed resident devices of a given area to the total population of the same area. For example, AirSage penetration rates use the 2010 U.S. census population of a census tract and the observed resident devices for the census tract. AirSage indicates that in the future they may expand their data to the census block group level.

According to their Trip Matrix Document, AirSage can provide processed cell O-D data for any contiguous window of time three or more hours in length defined by the customer. They can provide O-D data for a single day, average weekdays or weekends, and various peak periods and day parts. Three hours is the smallest time period for which processed cell O-D data can be provided since longer periods of time are needed to develop activity locations.

### 3.3 Location Accuracy of Cell Data

AirSage documentation states that on average their sighting locations are within about 300 meters in urban areas, but often they are within 50–100 meters.<sup>(1)</sup> However, location accuracy varies based on many factors such as by device type and quality and the size and density of cells in the cellular network. For example, one device may have an accuracy of 1,500 meters at 9 percent confidence while another has an accuracy of 250 meters at a 90 percent confidence level.<sup>(12)</sup> AirSage states they can limit the data points to only those with highly accurate locations. However, this could significantly reduce sample size and potentially bias the data toward a device type, carrier, or just areas with good cell infrastructure coverage.

AirSage reports that, on average, they receive 100 location signals (after filtering) from calls, texts, and data sessions for each device every day. However, there can be a lot of variation in the number and accuracy of sightings from one device compared to that of another. Some device types may be detected 50–60 times per day, while others are detected several thousands of times per day.<sup>(13)</sup> Factors influencing device sightings and accuracy include:

- Quality of device.
- Radio frequency characteristics.
- Device user.
- Cellular carrier coverage.
- Density of cellular network and towers.
- Density of the roadway network.
- Urban versus rural geography.

Since its development, there has been moderate consumption of pre-processed cell based O-D by MPOs and DOTs, but little scholarly work has been conducted on developing cell phone

location data mining algorithms and applications for comparison and further validation of results. Cell data's coarse granularity in space and time makes them very difficult to ground-truth against the user's actual travel route in an urban scale. There is vast uncertainty of a user's location when the user is not communicating with the network.<sup>(14)</sup>

*Cell data's coarse granularity in space and time makes them very difficult to ground-truth against the user's actual travel route in an urban scale. There is vast uncertainty of a user's location when the user is not communicating with the network.*

### 3.3.1 Privacy and Anonymization

Airsage's data do not contain any location information about a subscriber's identity because all of their records have been encrypted to anonymize the specific individual or mobile device information. Additionally, the encrypted ID for each mobile device is changed every 28 days to ensure an additional level of privacy protection.

Airsage states that their data are regularly tested by independent security auditors to ensure the data coming in and going out is fully anonymous. They work with carriers, partners, and neutral third-party auditors to strip out Customer Proprietary Network Information to remove all personally identifiable information. Airsage states that their data are anonymous and secure at all times.<sup>(15)</sup> Privacy protections listed on Airsage's website include:

- Using wireless aggregated carrier network data without any access to individual customer information.
- Being fully compliant with privacy laws and carrier privacy policies that prohibit third-party access.

### 3.3.2 Considerations in Scoping an O-D Study Using Cell Data

Travel studies collect data on the amount and characteristics of travel going within, into, out of, and through a study area. Trips going into the study area are termed E-I and those going out of the area are I-E. Similarly, trips staying within the study area are termed internal-internal (I-I) and those passing through the area are E-E. Today and in the past, transportation planning practice has relied on travel surveys to provide information to quantify trip generation and distribution for I-I trips. Data on external trip making are obtained from external surveys. These surveys are used to develop trip tables for E-I/I-E trips and E-E (external-through) trips and different subcategories such as resident, non-resident, non-commercial, and commercial (or truck). Today, these same products remain but the use of new passive technology is allowing greater flexibility in temporal aggregations along with similar options in aggregating resident versus non-resident classes of external travel.

Key steps and considerations for scoping an O-D study using cell data include the following:

1. Review how cell data are used to develop O-D data and understand its general spatial accuracy and trip end collection frequency and continuity.
2. Define goals and objectives for the study – is it for external travel and/or internal travel alone, travel model development, corridor study, or diurnal distribution of travel? These will help define the geographical scope, TAZ distribution/size for the study, and the audience.
3. Research market penetration of cell carriers in the study region. The cell data vendor may not have agreements with all providers. Those that it does not have agreements with may be more dominant in some regions. Additionally, variations in market penetration may affect rural areas more so than urban areas, especially rural areas with only one provider.

4. Review and become familiar with existing vendor products (e.g., Airsage's Trip Matrix product and Streetlight's web-based tools) and include both policy and technical/modeling personnel in the review to understand the various product options available for general planning and modeling uses.
5. Obtain general ball park pricing information from the vendor on the purchase of cell data for study alternatives with low, medium, and high numbers of cell zones; different time period analysis options; and different subscriber class and purpose attribute options. Use the ball park pricing information in development of the cell zone structure and deciding on which analysis options to purchase.
6. If applicable, identify the external stations to be included in the study. These are typically the major highways and thoroughfares that cross the study area boundary.
7. Develop the internal cell zone structure for the study area. The cell zone structure could be based on the study area's existing model or one that is under development. In the past, Airsage recommended a minimum cell TAZ size of 500 meters by 500 meters, but if possible, use larger cell TAZs than the minimum to improve the accuracy of O-D results.<sup>(16)</sup> Many studies have aggregated TAZs to enlarge their geographic size suitable for capturing cell trip ends. A 2016 study in the Dallas-Ft. Worth, TX, region used census tracts as the internal cell data zones. A key factor in the cost of cell O-D data is the number of analysis zones, so this should be kept in mind in developing the aggregated cell TAZ structure.
8. If applicable, develop external zones (e.g., districts or travel sheds) for the periphery around the study area. In coordination with the vendor, develop cell data capture areas that generally encompass the highway travel shed(s) related to one or more external stations for capturing cell data within these areas. These areas should be created around the periphery of the study area and consider the layout of major highways extending within, through, and nearby the area. Other factors such as small towns, major traffic generators, or other population centers within the external zones or near the study area boundary should also be considered. In general, a 30–45 minute travel-time buffer should be created around the study area to form the external periphery zones for cell data.<sup>(17)</sup> When developing external zones, it is important to keep in mind that cell data aggregations are based on activity points. For study areas with external roadways that are radially oriented, activity points collected within an arbitrarily drawn zone may be assumed to enter/exit the study area using its associated highway. However, more complex external networks pose a variety of challenges. These include:
  - Highways that exit the study area in very close proximity to each other.
  - Highways with conflicting (i.e., crossing) travel routes.
  - O-D activity points with multiple highway options for entering/exiting the internal study area.
9. These challenges effectively mean that E-I/I-E and E-E distributions among the various highways are subject to a degree of variability based on how the external zones are drawn. Figure 4 shows the external capture districts and external stations for an O-D study using cellular data in the Asheville, NC, area.<sup>(18)</sup>
10. Determine the study duration time period (beginning and end dates) for archived cell data to be aggregated and processed. The time periods to be used in the analysis of the data (e.g., peak periods, days, average weekdays) should be considered in determining the study duration time period. In general, the study time period for cell data could be from one to four weeks (or more), depending on study objectives and cost considerations.
11. Provide preliminary internal and external zone structures to the vendor for review and feedback in relation to suitability for cell data processing, accuracy margins, and more refined cost estimates.

12. Select period aggregations for O-D data analyses needed to address the study area's planning and modeling needs. Time-of-day options include various day parts and peak periods. Further, aggregation options are based on the total study time or averages by day, week, average weekday or weekend, etc. Select time periods for analysis weighing study needs and costs.

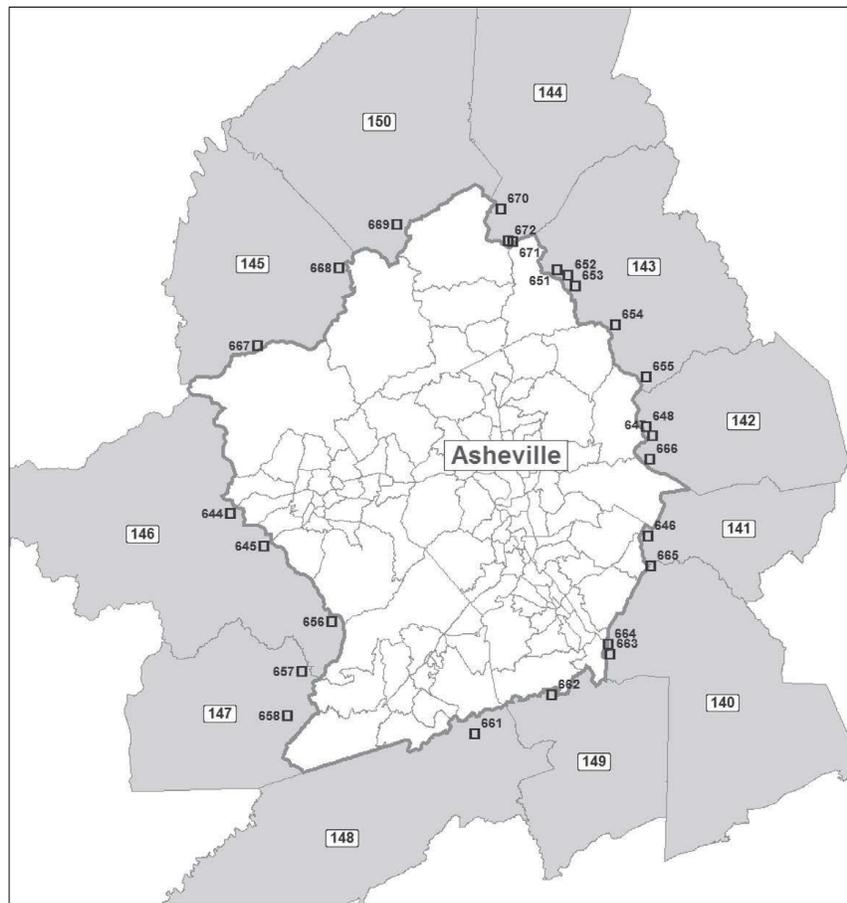


Figure 4. Map. Example data capture areas around periphery of cell O-D study area.

13. If applicable, select an E-I-E filter option to correctly identify E-E trips that pass through the study area. This filter ensures that the results for E-E trips begin outside of the study, enter the study area, and then exit the study area. Without this filter, the E-E totals will also include trips between the travel shed zones as E-E trips, which are not accurate for modeling purposes. Figure 5 illustrates E-E trips in a study area where the E-I-E filter is not included. The red lines show E-E trips between the travel sheds, which are typically not considered true E-E trips.
14. If desired, select trip purpose options. The current cell data vendor in the United States has two classification schemes for trip purpose options. One includes HBW, HBO, and NHB. A second option includes many combinations of home, work, and other. Selection of trip purposes is not needed for analysis of external travel, but selecting at least the three class-purpose options allows for a basic analysis of trips by purpose, especially if the I-I trips are to be analyzed.

15. For external studies, bi-directional traffic counts should be collected at all roads that cross the internal study area or at minimum those roads with 500 or greater average annual daily traffic. Counts should be taken for three consecutive weekdays occurring during the cell data analysis period. If cell data are being analyzed for typical weekend traffic, counts should be collected for at least one weekend occurring during the cell data analysis period.

The vendor's options platform should allow for varying degrees of level of detail for the overall study and analyses based on the study size and design and options selected by the purchaser. The cost of data may vary widely depending on the number of zones, day and time-of-day aggregation options (e.g., average weekday or weekend day, peak periods), trip purpose attribute options selected, resident class attributes, and various other filters. Table 4 shows an example output file of Airsage data.

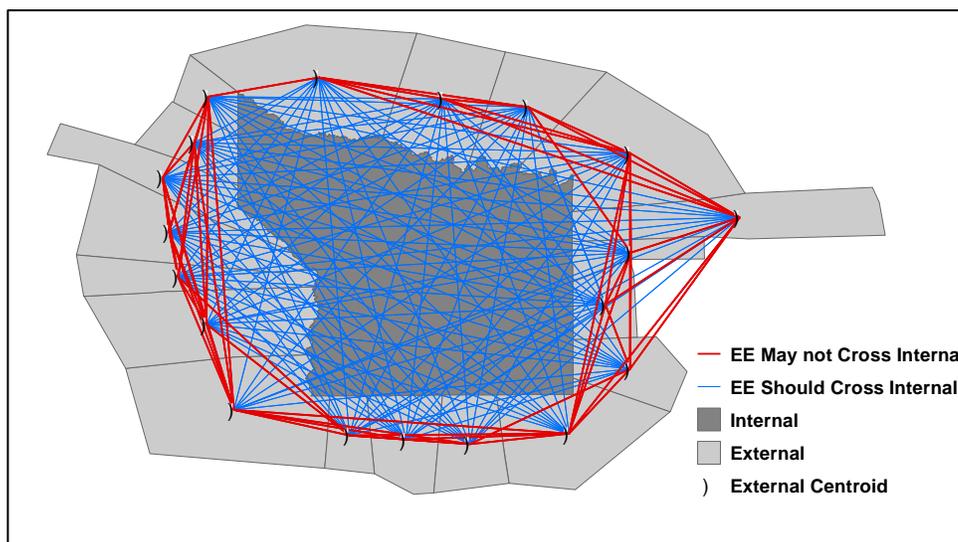


Figure 5. Illustration. Cell E-E trips without E-I-E filter.

**Table 4. Example output file of Airsage cell data.**

Origin Zone	Destination Zone	Start Date	End Date	Aggregation (workday)	Subscriber Class	Purpose	Time of Day	Count
239	188	20120920	20121018	WD	Resident	HBO	H19:H24	2.43
180	507	20120920	20121018	WD	Resident	NHB	H6:H10	3.88
170	105	20120920	20121018	WD	Non Resident	NHB	H6:H10	1.09
244	254	20120920	20121018	WD	Resident	HBO	H6:H10	0.97
502	192	20120920	20121018	WD	Resident	HBO	H10:H15	0.37
161	248	20120920	20121018	WD	Through	NHB	H15:H19	0.55
506	130	20120920	20121018	WD	Resident	HBO	H0:H6	0.99

Source: (19).

### 3.4 Processing and Analyzing Purchased Pre-Processed Cell Data

This section describes key steps for the quality control and post-processing of cell data including troubleshooting and checking techniques. Obtaining the most useful results from cell data can be an iterative process. Typically the sponsoring agency will work with a vendor such as Airsage to address data issues and reconfigure capture areas and/or cell TAZ zones, if needed. The cell data will need to be re-processed and re-sent if any changes to the internal or external zone structures are made and to address any other issues identified. This process of reconfiguring zone structures and reprocessing the data may need to be repeated a few times until final cell data set is established.

*When comparing the cell data to existing data sources and/or models, the analyst should keep in mind that all trip data are in the form of device trips rather than vehicle trips.*

When comparing the cell data to existing data sources and/or models, the analyst should keep in mind that all trip data are in the form of device trips rather than vehicle trips. This should be most pronounced in urban areas and less pronounced with interactions between TAZs separated by some degree of distance and/or between external TAZs to other external and internal TAZs. For all practical purposes, it is assumed that device trips are the same as person trips.

#### 3.4.1 Initial Review of Cell Data

Results of an O-D study using cell data are typically delivered in a comma-delimited text file. Initial review of the results should include the following:

- Summarize results for each external TAZ and compare the summaries to count data. Results should be compared to traffic counts for all roads which cross the boundary of the internal study area.
- Typically for external TAZs, the cell data results will not match the traffic counts. This can be corrected at later phases of the study. However, it is beneficial to review the proportionality of each external TAZ total results to the total for all external TAZs:

$$\frac{\sum EI + \sum IE + 2 \cdot \sum EE}{\sum (EI + IE + 2 \cdot EE)}$$

**Figure 6. Equation. Proportionality of external TAZ total results.**

- Review TAZs to identify those where E-I+I-E trips are greater than I-I trips.
- Identify cross tabulations of external TAZs which share a border or those where the straight line drawn between them does not cross the internal study area. Any trips between these pairs may be anomalous and warrant further investigation. For small areas this is best done visually; for larger areas, a geospatial analysis should be used.

### 3.4.2 Detailed Review of Cell Data

The initial review of the data is beneficial for quickly identifying potential errors in the cell data and addressing them with the vendor. However, it is recommended that the cell data also be subject to a more comprehensive review to include the following:

- Cross tabulating total trips for each zone against existing data about that zone. These data can include total population and/or total employment, acres of urbanized area, and/or point level data for special generators. This comparison may identify anomalies in the cell data related to the proximity of a TAZ boundary to a population and/or employment center within another TAZ. These comparisons are best accomplished using database and geospatial analysis software.
- Review of major HBW origins. Current practice is to assume that where a device spends the majority of its evening is its home location and where it spends the majority of its day is its work location. However, reviewers of cell data have observed that for some major generators this assumption may produce errors for such generators as universities and major employment areas that have 24-hour shifts.<sup>(20)</sup>
- Comparison of cell data results to those from an existing model such as comparing:
  - Trip flows by purpose by zone and/or by district (and aggregation of many zones).
  - TLFID by purpose.
  - Traffic assignment (keeping in mind that cell data represent person trips).
- Comparison of cell data results for HBW total trips and TLFIDs to U.S. Census Journey-to-Work data, Census Transportation Planning Package, and Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics.<sup>(21)</sup>
- Identifying desire lines of major O-D patterns.
- Review of trip purpose totals for the study area against national data sources such as the National Household Travel Survey and previous regional travel surveys, where available. It would not be expected for totals from these surveys to match, although the proportion of total trips by trip purpose may yield useful baseline information.
- Existing studies have found cell data to compare well with survey data on internal HBW and home-based non-work trips. For NHB trips, significant differences have been observed and may be due to:
  - Missed short trips (e.g., work based trips).
  - Presence of commercial vehicle trips.<sup>(21)</sup>

### 3.4.3 Development of Cell Data O-D Trip Tables

Current cell data acquisitions will include total trips and stratification of trips by trip purpose (if applicable), and visitor versus resident, depending on the level of detail indicated at purchase. At this time, cell data cannot distinguish between non-commercial versus commercial, so trip tables based on these vehicle classes cannot be developed.

Processing cell data into trip tables should consider the goals and objective of the study. Considering: 1) if the study is just a comparison; 2) providing external data for input to the model; and/or 3) being directly applied to a highway network in lieu of model development.

For comparison studies, the data delivered by the vendor can be processed directly from the data files provided. For external studies, cellular results will not typically match known traffic counts, in which case the cellular data may need to be balanced and fitted to traffic counts, using techniques such as Fratar (see section 4.5.2).

For assigning the data directly to a highway network, several processes are available. The existing model network for the study area can either be adapted to the cell data TAZ structure or the cell data can be disaggregated back to the existing TAZ structure using population and employment values for each TAZ. To obtain results that more closely match traffic counts for a region, the cell data can be adjusted using ODME processes. ODME is a short-term technique usually used for short-term forecasting; traffic flows produced as result are based on statistical fitting and as result can be over-fit.<sup>(21)</sup>

### 3.4.4 Visualizing of Cell Data O-D Results

Besides benefits to traffic modeling applications, cell data can also benefit urban, regional, and statewide planning efforts. The data can be easily adapted to and incorporated into a number of software applications including Microsoft<sup>®</sup> Excel/Access, R, Relation Database Management Systems (e.g., PostgreSQL), GIS applications, and travel demand model applications. Each of these packages has the ability to produce graphics to visualize the cell phone data. Visualization of cell data helps simplify and communicate to the public and stakeholders the complex nature of travel patterns. For example, the “Origin Destination Analysis for Moore County, NC” helped communicate and quantify to residents the composition of traffic along the region’s primary north-south corridor relative to resident and non-resident travel. In doing so, it helped establish the corridor as a one of regional and statewide importance rather than just local. In this regard, the perceived objective nature of cell phone data allayed resident concerns of biases in other sources of data.<sup>(22)</sup> Other examples include Figure 7, which illustrates the relative proportion of external travel between external stations for a region (Smith County, TX) using cell data.

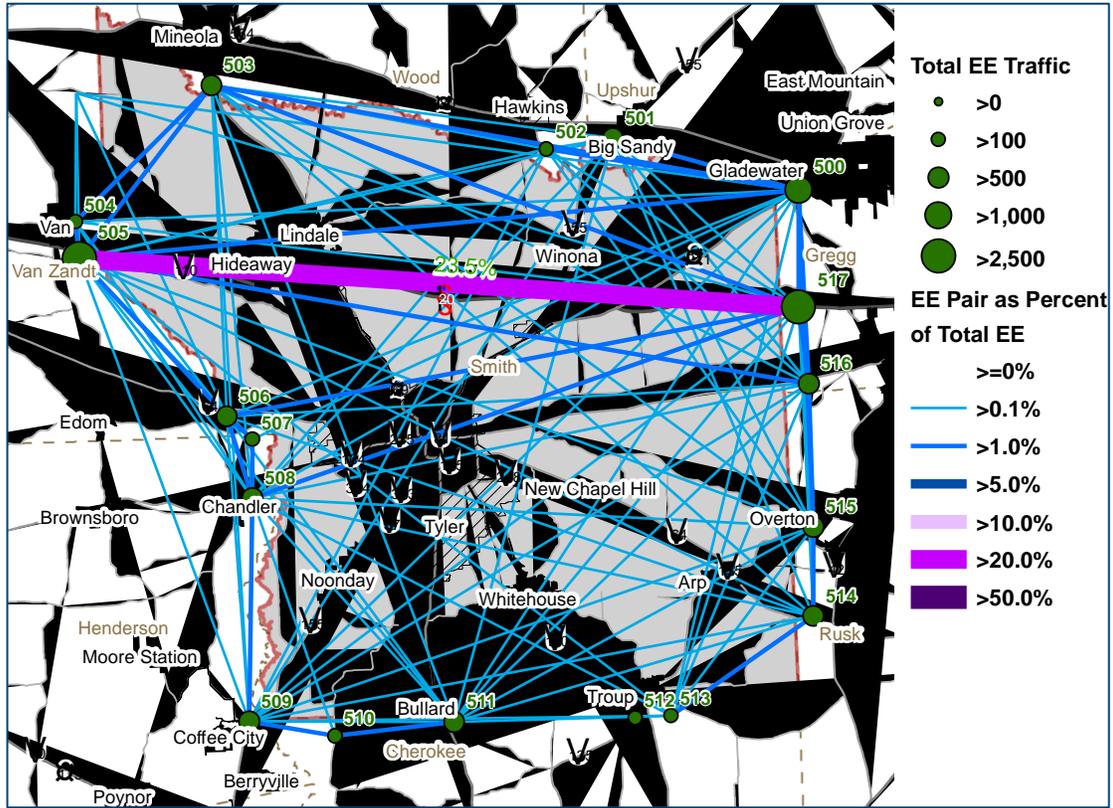


Figure 7. Map. Smith County, TX, cell data proportion of total E-E trips by O-D pair.

## 4.0 Understanding GPS O-D Data

### 4.1 What They Are and Represent

#### 4.1.1 GPS Data, Acquisition, Sources, and Privacy

GPS receivers take information from the GPS satellites and use trilateration to calculate the device locations. A GPS receiver must be locked on to the signal of at least three satellites to calculate a coordinate (latitude and longitude) and track the device movement. After the time-stamped positions are determined, the other information such as speed and bearing can be calculated. The GPS device movement creates a series of time-stamped coordinates, which can be used to derive various attributes, infer trip ends, and finally construct O-D information.

There are two categories of GPS data:

1. Primary GPS Data. Primary GPS data are raw, unprocessed GPS data collected through first-hand means such as using GPS tracking devices in vehicles to obtain travel-time data for a floating car study or to obtain trip data as part of a household travel survey.
2. Third-Party GPS Data. Third-Party GPS data are data obtained from data providers such as HERE, TomTom, or INRIX who continuously collect, purchase, and compile GPS data from a variety of sources for eventual sale to businesses or government agencies. GPS data acquired from third-party providers will be pre-processed to anonymize it and/or to provide it in pre-established formats and outputs.

In acquiring private sector GPS data for a study, specifications and criteria should be discussed with the data provider to ensure that the data provided are suited for the study's objectives. Key specifications include data collection time period, definition/criteria for a trip and trip end, and variables to be included for trip data. Currently, third-party GPS data providers can provide anonymized trip data for a pre-defined study area in the following forms:

- An aggregated data set of trip origins and destinations.
- An aggregated data set of trip origins and destinations with waypoints or TMCs to reveal general routing.
- Trip matrices based on a TAZs provided by the purchaser.

To ensure the anonymity of the data set, the data provider will apply anonymization techniques either in time and/or space to ensure user privacy. For example, the first few minutes of driving and last few minutes of driving may be removed in order not to reveal the exact location of the trip beginning. Additionally, a random offset from the actual location of the beginning and end points could be imposed. The anonymization rules should be communicated by the data provider so the analyst knows the limitations and potential pitfalls associated with the use of the data. In addition, the GPS data sources may also scramble unique device IDs periodically to ensure that the devices are not traceable on a long-term basis.

*The anonymization rules should be communicated by the data provider so the analyst knows the limitations and potential pitfalls associated with the use of the data.*

#### 4.1.2 How Trips Are Developed Using GPS Point Data

Trips are developed from GPS points by identifying the trip ends. The trip ends are generally identified by computing the dwell time from the GPS data points. The dwell time is the time duration that a device remains at the same location. For GPS data, a rule of thumb is at least two minutes of dwell time to be a candidate for a trip end. More sophisticated rules may be

considered in identifying legitimate trip ends. These rules include the consideration of point of interest locations and the street maps in conjunction with the trip ends to reduce the chance of misidentifying traffic stops as trip end locations. A pair of trip ends is considered a trip. A series of trip records can be used to construct an O-D table using any well-known analytical tools (e.g., pivot table in Microsoft® Excel).

#### 4.1.3 Types of Trips Collected with GPS-Based O-D Data

Types of trips collected with GPS-based O-D data depend on the data sources. In the case of INRIX and HERE data, the provider knows whether the data sources are coming from in-vehicle navigation systems, after market GPS units, fleet/freight vehicles, or mobile applications. From these sources, the O-D data can be constructed or produced separately from each source or into the general categories of commercial (fleet/delivery) and non-commercial (consumer) vehicles. The time stamp information associated with each trip can be used to build O-D by time of day, day of week, or weekday/weekend basis. Unlike cell data, GPS data can be used to develop results in hourly increments.

#### 4.1.4 Chief Purposes and Uses of GPS O-D Data

Since the early 2000s, the chief uses of GPS O-D data have been for collection of trip data as part of household surveys primarily as a means to estimate the amount of trip under-reporting in CATI or paper surveys. This application used primary GPS data collected first hand via GPS tracking devices provided to a subsample of participants of a household survey. GPS data have been commonly used in floating car travel time studies, but here it was to collect travel time and speed data not O-D.

Over the past year, the use of third-party GPS data for O-D purposes has become available as new products have been (and are being) developed that can provide processed GPS O-D data specific to a study type, scope, geography, and time period. More detail on new GPS O-D products is provided in section 4.6. Currently, the GPS data supporting these products have low sample penetration and a commercial bias. However, as data providers continue to add new GPS devices and sources, the sample penetration will improve and the commercial bias should be reduced. Since the Tyler study was conducted in early 2014, INRIX has indicated that the saturation of the non-freight GPS data has improved relative to the freight data.

GPS O-D data can be used for external surveys, corridor studies, travel pattern/routing studies, and freight/commercial vehicle studies, select link analyses, and ODME. Due to its (current) commercial bias, it is a good source for studying commercial vehicle and freight O-D and travel patterns. In the 2014 Tyler study, INRIX O-D data were used to develop E-E and E-I/I-E matrices and results were considered reasonable. For E-E trips, GPS O-D may be a better option than cell data since trips can be identified at the exact highway locations where they cross the study areas boundary and broken down by commercial and non-commercial categories and expanded to traffic counts. On the other hand, cell data are needed to provide estimates of resident versus non-resident travel coming into and out of a study area. The GPS O-D data provided by INRIX for the Tyler study was developed using a preliminary beta version of its current Insights Trips product.

Studies by the Minnesota and Maryland DOTs are currently underway that are using INRIX's Insights Trips product for GPS O-D data. The Minnesota project is using GPS data to study trip O-D and routing patterns to support a congestion relief study along IH-494/TH62 in the Twin Cities Metropolitan Area. The Maryland project is using GPS O-D data for a comprehensive statewide freight fluidity analysis. Section 6.2 provides more information on these projects.

## 4.2 Sample Size, Collection Timeframe, and Saturation

Since the use of GPS data for O-D is still new and evolving, there are only a few recent studies from which an estimate of sample penetration can be obtained. In 2011, TTI partnered with a national data provider to assess the viability of its GPS data for O-D purposes and concluded that the sample penetration was not high enough to develop external trip tables for the area studied. Since then, the GPS sample penetration has improved due to more widespread use of GPS-enabled mobile devices, navigational apps, and in-vehicle navigation systems. In 2014, the sampling rate for GPS data in the Tyler study was estimated at 1.5 percent. Currently, TTI researchers estimate the GPS sampling rate from third-party data providers to be in the range of about 0.5 percent to 2 percent of vehicles on the roadways. However, this rate is continually increasing, making GPS technology an increasingly better option for O-D data collection into the future.

Trip data from GPS data providers can be developed and extracted for almost any time period since the data collection frequency for GPS data can be as low as one second increments. However, third-party GPS data vendors will probably not provide trip flow data in increments of seconds because they would be unable to retain proper anonymity of the data. Such data could possibly be supplied where trip flows are developed based on trip end times rounded to the nearest 10- or 15-minute increments.

## 4.3 Accuracy Margins

The GPS technology on its own is very accurate under ideal conditions (e.g., small geometric dilution of precision, no urban canyon effect). Recent studies reported the accuracy of GPS devices on popular smartphones to be in the range of 5–8 meters. Even without the GPS, the 3G iPhone with WiFi can be located with the accuracy of about 74 meters.<sup>(23)</sup> In a current TTI study using third-party GPS data, latitude and longitude coordinates were provided to four decimal places, or accurate to about 30 feet. The accuracy that GPS devices can achieve is suitable for a wide variety of O-D applications.

Trip data obtained from the third-party O-D providers will (with few exceptions) be anonymized either temporally or spatially or both. Purchasers of the data may need to take into account anonymization measures applied to the raw data set by the provider that will impact the accuracy of trip ends. The data purchaser should discuss the anonymization criteria with the data provider in order to understand its effect on the applicability of the data set for the applications of interest.

While GPS data and waypoints points are accurate, GPS trip traces are only captured when the users activate the navigation session, particularly when using mobile apps. Unlike GPS devices in freight or fleet vehicles, non-commercial users may not immediately activate a navigation session from the beginning and may also intermittently use the application throughout their travel. This on-again-off-again use of navigation will affect the quality and validity of trip traces available for O-D analysis, especially for non-commercial trips. The collection of partial trips in GPS data is mitigated by third-party providers by analyzing trip making over long periods of time, such as 1–3 months.

*While GPS data and waypoints points are accurate, GPS trip traces are only captured when the users activate the navigation session, particularly when using mobile apps.*

#### 4.4 Considerations in Scoping an O-D Study Using GPS Data

The scoping of a study using GPS O-D data depends primarily on the type of study, its objectives, and desired outcomes. Different types of studies will call for different types and aspects of trips to be analyzed and require different approaches and methods to be used in development of the O-D trip data. With few exceptions, most studies using GPS O-D data today should use third-party data due to its advances in sample penetration and product offerings within the past couple of years. The GPS sample penetration will continually increase and the current product offerings are still very new, so these should evolve and improve in years to come. Key factors, elements, and decisions to be considered in scoping an O-D study using GPS data include the following:

*The GPS sample penetration will continually increase and the current product offerings are still very new, so these should evolve and improve in years to come.*

- Study area and GPS data coverage area. The purchaser of the data will need to work with the data provider to determine whether GPS data should be purchased just for the study area or for a larger coverage area. The size of the GPS data capture area depends on many factors. Examples could include an MPO study area, a county, multiple counties, or an entire state.
- Time period and duration of GPS data. The purchaser can purchase data for the time period that best suits the study such as as-recent-as-possible to study current conditions, a period in the past to be representative of a model's base year, or a period when a major/special event occurred. The duration of the time periods available may vary by provider and be anywhere from a few days, to one month, to a few months, to a year.
- The types and form of O-D data to be procured. Currently, pre-processed GPS O-D data can be provided in the following forms:
  - Trip O-D matrices based on a study area TAZ layer. This option would provide a matrix with a percentage of total trips that occurred between each O-D pair in the study area. It may or may not include the individual trip records, depending on the provider.
  - Individual trip records that include the trip start and end points. These data provide O-Ds that relate to the trip starts and ends of vehicles within a user defined region.
  - Individual trip records that include the trip start, end points, and waypoints or TMCs between the trip start and end points. Waypoints and TMCs provide additional data that allow for the routes of trips to be studied.

Any of the above options can provide O-D data on E-I, I-E, E-E, and I-I trips that could be applicable for external or corridor studies. The addition of waypoints or TMCs increases the amount and flexibility of the analyses that can be performed on the data. Each of the above options has different cost implications:

- Vehicle types and classes. GPS trip data can be provided for all vehicles or for only certain types of vehicles such as non-commercial vehicles or commercial vehicles. For commercial vehicles, an option to categorize these into different weight classes may also be available.
- Device types and sources. GPS O-D data can include identifiers for the source of the data such as from in-vehicle navigation systems, after market GPS devices, fleet tracking systems or sources, mobile device apps, and potentially others.
- Time periods needed for analysis. Acquiring GPS O-D data by time periods such as average weekdays, weekends, or peak periods may or may not be an option. Users

would need to discuss this with the data provider prior to purchase. With the exception of trip matrix data, typically the trip records are not pre-processed into time periods and the user must process the data into the time units desired for analysis.

- Analysis of the GPS O-D Data. The data set provided will be big and contain many interrelated elements and files. Processing and analyzing the data may be complex and time consuming and may require an experienced analyst with advanced GIS and programming skills. Some users may need or desire to hire a consultant to perform the analysis.
- Cost of the data. The current general pricing structure of commercial O-D products generally depends on the:
  - Size, population, and the number of zones in the study area.
  - Type of deliverables – the price increases with the levels of post processing required and the number of additional (non-basic) options selected by the user.
  - Length of the study period.
  - Number of categorical factors to be included in the study (e.g., day part, time part, trip types, waypoints).
- Data Elements. The GPS data elements to be included in the data set and in what form they will be provided will need to be discussed between the data purchaser and provider prior to acquisition. Typical data elements will include items such as trip identification (ID), device ID, provider type, longitude and latitude coordinates of trip ends, trip start and end times, and possibly elements related to when and where a device crossed the study area boundary (if applicable).
- Classification of Trip Types. In developing trip O-Ds in the GPS data, the user and the data provider must ensure that they both have the same understanding and definition of certain trips types. For example, for trip types for an external survey are defined as follows:
  - E-I trips. Trips that begin outside of the study area and end inside the study area.
  - I-E trips. Trips that begin inside of the study area and end outside the study area.
  - E-E trips. Trips that pass through the study area.
  - I-I trips. Trips that begin and end inside of the study area.

## 4.5 Processing and Analyzing GPS Data

### 4.5.1 Data Processing and I-E, E-I, and E-E Trip Classifications

This section describes key steps for processing GPS data into a usable O-D data. Some troubleshooting and checking techniques are also discussed at the end.

A TAZ polygon layer for the study area should be used to identify trip ends and the corresponding TAZ number for each trip end. Identifying the accurate TAZ number for each trip end may need to include developing a process to account for anonymization of the data. The reported trip ends may be offset several minutes or more such that the actual location of the trip end will remain unknown. In some cases, this offset will cause trip ends to be located in the wrong internal TAZ and cause some to be incorrectly located inside or outside of the study area. Such mislocations could cause some trips to be identified as E-I or I-E when they are E-E, or E-E when they are actually E-I or I-E. Measures to account for these offsets will need to be put into place in processing the data to improve accuracy.

*The reported trip ends may be offset several minutes or more such that the actual location of the trip end will remain unknown.*

*Measures to account for these offsets will need to be put into place in processing the data to improve accuracy.*

The results for the number and percentage of E-E trips in the study area can be determined with or without imposing constraints on the amount of time it takes to travel between external pairs. Whether or not to do this may depend on a number of factors such as how E-E trips are defined in the area, the time period to be examined, or how the area's model handles E-E trips. For example, if a trip begins and ends outside of the

study area but makes a stop inside of the area, it is two separate trips—an E-I trip coming in and then an I-E trip going out. To prevent such E-I+I-E trips from being counted as E-E trips, the data can be analyzed where only trips that travel between external pairs within a specified time limit, such as the those in the model's travel time skims, are counted as E-E trips. Applying travel time constraints will help remove E-I+I-E trips from the E-E totals.

There are two alternatives in dealing E-E trips with travel times exceeding the skim time threshold. The first option is to split one E-E trip into one E-I and one I-E trip. The second option is to exclude these trips from the analysis. TTI adopted the latter in the Tyler study as there was no way to identify a specific internal TAZ if a through trip was split into two local trips.

#### 4.5.2 Development of E-E O-D Matrices

The trip records can be used to build E-E O-D matrices for all vehicle types, freight, and non-freight (cars plus mobile applications). Time constraints, if desired, can be applied in developing the matrices. The O-D matrices should then be balanced so that the trips observed for the O-D pair  $(i,j)$  are equal to the O-D pair  $(j,i)$ . The balancing process can be expressed in matrix form as:

$$M_{Balanced} = 0.5M + 0.5M^T \quad (1)$$

**Figure 8. Equation. Balanced O-D matrix.**

Where:

$M_{Balanced}$  = Balanced O-D matrix.

$M$  = Observed O-D matrix.

$M^T$  = Transpose of matrix  $M$ .

GPS trips for the O-D pairs represent just samples of the population. The sample obtained needs to be expanded to match the traffic counts collected at the external stations. One popular technique to factor up the sample O-D matrix is the Fratar method (also known as bi-proportional matrix balancing). In this method, each row and column of the O-D matrix is assigned a scaling factor. Each sample entry in the O-D matrix is then multiplied by the corresponding row (origin) and column (destination) scaling factors to produce the population estimate. Scaling factors are adjusted iteratively until the total number of trips associated with each origin and each destination match the total observed traffic volumes to and from that zone as closely as possible. Because the procedure is multiplicative, its main limitation is that any O-D pair that has zero trips in the sample will remain zero after expansion unless the analyst manually overrides a zero.<sup>(24)</sup>

The Fratar scaling process can be implemented in Microsoft® Excel using the Solver engine to adjust the row and column scaling factors, with an objective of minimizing the sum of the squared difference between the factored row and column sums and the corresponding observed traffic volumes. The factored trips from origin  $i$  to destination  $j$  can be written as:

$$T_{ij} = F_i \times F_j \times \tau_{ij} \quad (2)$$

**Figure 9. Equation. O-D matrix factoring.**

The optimization objective is to:

$$\min \left[ \sum_{i=1}^n \left( V_i - \sum_{j=1}^n F_i F_j \tau_{ij} \right)^2 + \sum_{j=1}^n \left( V_j - \sum_{i=1}^n F_i F_j \tau_{ij} \right)^2 \right] \quad (3)$$

**Figure 10. Equation. O-D matrix factoring optimization.**

subject to  $F_i$  and  $F_j > 0$

Where:

$\tau_{ij}$  = sample GPS trip counts from origin  $i$  to destination  $j$ .

$T_{ij}$  = factored trip counts from origin  $i$  to destination  $j$ .

$V_i$  = total observed volume counts at origin  $i$ .

$V_j$  = total observed volume counts at destination  $j$ .

$F_i$  = Fratar factor for origin  $i$ .

$F_j$  = Fratar factor for destination  $j$ .

#### 4.5.3 Troubleshooting and Reasonableness Checks

There are several techniques that can be used to troubleshoot the process and check the reasonableness of the results. In many cases, the O-D data obtained may be the only data set available and cannot be directly compared with others. The analyst may need to review other information obtained from the data set to check for its reasonableness. Techniques in the following paragraphs can be considered and applied for this purpose.

Distribution of data by sources and time parts. The analyst can first check the distribution of the data sources by time parts (e.g., commercial versus non-commercial) and determine if the observed patterns make sense. Using the knowledge about the study area, this analysis can compare and examine how the distributions of commercial vehicles observed in the data set vary by time-of-day (e.g., percentage of trucks during night versus day) or by day of week (e.g., percentage of trucks on weekdays versus weekends).

Comparison of percentages of local versus through trips by data sources. From the O-D tables, the analyst can group the data by trip types whether it is a through trip (E-E) or a local trip (E-I/I-E). The data can be further broken down by vehicle types (commercial versus non-commercial). Using the knowledge about the study area, the analyst can investigate the pattern of the local versus through trip splits to determine if the results are reasonable. For example, users may expect to see higher local trip percentage for non-commercial vehicles than commercial vehicles but the splits may change when examined by specific time-of-day or day of week.

Comparison of counts from GPS-derived O-D and actual counts by vehicle types. The trip counts in the O-D tables from both balanced and unbalanced data sets can be aggregated by zones to generate trip counts observed at each zone. The unbalanced data set should be checked to determine whether the distribution of counts by each zone is consistent with the counts obtained from available stations. Users would expect to see higher traffic counts at count stations along major freeways than local roadways even from the unbalanced data set. After the balancing process, the overall distribution of traffic counts from the balanced O-D table across

external stations can be checked against the distribution of actual counts if the count data are available.

Trip length and trip time frequency distributions. Each trip record provides the time-stamped coordinates of trip origins and destinations. These coordinates can be mapped to TAZ layers where trip length and trip time for the TAZ pairs can be obtained from the skim value tables. The trip times reported by the data provider may be affected by the rounding and anonymization process. While technically it is possible to obtain the trip length directly from the GPS traces, the data provider generally neither provides actual trip traces nor trip lengths developed from the trip traces. This is because of the computational overhead required in tracking all of the points to calculate the true trip length. Instead the data provider will only extract trip ends and/or waypoints along the trip from the entire trip traces and only use these data during subsequent processing. The analyst can construct trip length and trip time frequency distributions from trip length and trip time associated with each TAZ pair. The trip count in each cell in the balanced O-D table represents the frequency of such trip length and time. The frequency distributions can be developed by vehicle types, time-of-day, and day of week. The results can then be evaluated against the knowledge about the study area. For example, one may expect to see a heavier tail on the TLFD from commercial than non-commercial ones.

*While technically it is possible to obtain the trip length directly from the GPS traces, the data provider generally neither provides actual trip traces nor trip lengths developed from the trip traces.*

Tour identification for unique device IDs. Depending on the GPS data provider, this analysis is possible if the unique device ID is not changed for at least 24 hours and the anonymization process does not significantly offset the trip ends temporally and spatially. A chain of trips from a unique device ID is considered a tour if the first trip origin and the last trip destination of the day coincides. One can expect to observe a higher percentage of non-commercial devices complete a tour within a day when compared to commercial data sources.

## 4.6 GPS-Based O-D Products

In spring 2015, INRIX introduced its Insights™ Trips product, which allows consumers to purchase their GPS-based O-D data. INRIX processes raw GPS data internally from dozens of sources such as commercial fleets, in-vehicle navigation systems, after market GPS devices, and mobile application users. Their trip algorithm first determines the type of incoming source data to include only vehicles and then evaluates if the data are adequate for converting into trips. The Trips product currently offers three types of deliverables for consumers:

- Trip records – time-stamped trip start and trip end locations.
- Trip records and waypoints – waypoints are provided in addition to the trip records.
- O-D matrices – the matrices are built based upon the consumer-provided zone structure.

During the preparation of this report, several studies around the United States were underway using INRIX's Trips product, but none were complete and final reports on these projects were not yet prepared. INRIX used a preliminary beta version of Trips to provide GPS O-D data for the spring 2014 study in Tyler sponsored by TxDOT.

Waypoints or TMCs from trip traces can be used to infer route information. The waypoints or TMCs are provided by INRIX as time-stamped coordinates in a separate file. When waypoints or TMCs are included, the entire lengths of trips (or trip tours) will be included in the data, even the portions of the trip that are outside of the study area. If waypoints or TMCs are not included

in the data, an additional data capture area around the periphery of a study may be needed to capture device trips that begin or end outside of the study area.

In January 2016, HERE introduced its O-D product called Trip Data. Similar to other GPS data providers, HERE was previously using its GPS probe data to study travel times, speeds, and congestion. The HERE Trip Data product algorithmically stitched the probes back together, to reveal the origin and destination of journeys.<sup>(25)</sup> While TomTom does not have a GPS O-D product, they can work with users to develop O-D information from their data on a case-by-case basis.

Another option for O-D data is StreetLight Data. StreetLight is a company that uses INRIX data to develop a mapping platform tool and uses a combination of both GPS and Airsage cell data in developing O-D results. The tool can be used most effectively to show travel behavior to stakeholders, based on an analysis of origin and destination pairs. Using aggregated data, the user can see before and after effects on traffic. Varying levels of metrics and number of zones can be considered in the analysis.<sup>(26)</sup> The analyses can use either GPS or cell-derived data for O-D, or both sources, depending on the user-defined parameters and scale of the study.

## 5.0 Understanding Bluetooth O-D Data

### 5.1 What They Are and Represent

#### 5.1.1 Wireless Technology for Exchanging Data over Short Distances

The Bluetooth protocol is a widely used, open standard, wireless technology for exchanging data over short distances. The technology is frequently embedded in mobile telephones, GPS, computers, and in-vehicle applications such as navigation systems. Each Bluetooth device uses a unique electronic identifier known as a MAC address. Conceptually, as a Bluetooth-equipped device travels along a roadway, it can be anonymously detected at multiple points where the MAC address, time of detection, and location are logged.

By determining the difference in detection time of a particular MAC address, the travel time and average travel speed between locations can be obtained. In addition, the resulting information can be used to estimate the O-D patterns of the Bluetooth monitored vehicles. A significant advantage of the use of Bluetooth MAC addresses for travel time monitoring and O-D data is that typically only one inconspicuous roadside installation is necessary (consisting of a field processor with appropriate software and an antenna) to capture the unique address of Bluetooth devices traveling in all directions of flow. The effective range of TTI's Bluetooth readers is about 100 meters. Figure 11 illustrates how Bluetooth data are collected.

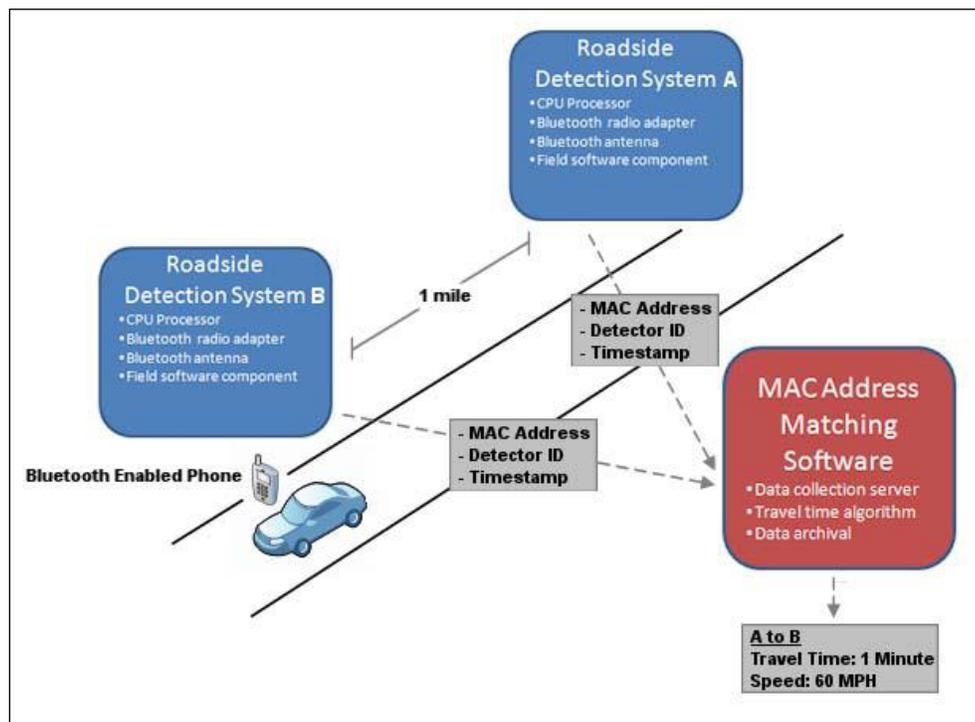


Figure 11. Illustration. Concept of how Bluetooth data are collected.

As TTI staff evaluated different types of field process controller devices as well as developed and refined field and host software systems, it became apparent that the TTI product was different than other commercially available products. The TTI software processes consistently results in an increased number of anonymous Bluetooth devices reads and more importantly valid matches when compared to other products. The resulting TTI developed intellectual

properties enabled The Texas A&M University System to submit a patent application and a subsequent commercial license for the method and process of utilizing Bluetooth technology for collecting traffic information. This system has been successfully deployed along IH-35 between San Antonio and Hillsboro, along IH-45 between Galveston and Dallas, as well as for a 650+ node arterial network in Houston. These operational deployments provide real-time travel time and speed information for the Houston region for active traffic management, public traveler information systems, and a wealth of data available for O-D pattern assessment and transportation planning studies. TxDOT has recently completed the replacement of 95 percent of the AVI readers in the Houston region with Bluetooth technology using the TTI intellectual property for real-time freeway travel-time monitoring. Including the product commercialization, more than 3,000 field units have been installed worldwide.

Texas is not alone in the pursuit of using Bluetooth technology for the purpose of O-D travel estimates. State DOTs in Oregon, Florida, and Pennsylvania have all sponsored studies that used Bluetooth O-D data collection efforts.<sup>(27,28,29)</sup> Additionally, municipalities are also using this lower-cost approach to traditional O-D data collection. Studies have been conducted in many areas throughout the country, including but not limited to Phoenix, AZ; Seattle, WA; Charleston, SC; and Jacksonville, FL. (See references 30, 31,32, and 33.) Outside of the United States, Bluetooth is also being used to help planners, engineers, and decision makers assess transportation conditions in their respective areas. This includes multiple efforts in Canada in areas such as Toronto, Vancouver, Montreal<sup>(31)</sup> and Calgary. Australia, Turkey, and Scotland have also used the technology for various O-D related purposes. (See references 34, 35, 36, 37, 38, and 39.)

### 5.1.2 Estimating O-D and Travel Time from MAC Addresses

The general premise of Bluetooth used for travel time or O-D data collection is that multiple readers are deployed along a corridor or in an urban area. The reader devices record Bluetooth signals emitted from cell phones and in-vehicle navigation systems. The recorded results are processed to develop estimates of travel time, speed, and/or matches between locations.

### 5.1.3 Acquisition and Sources

Bluetooth data can be acquired via permanent or portable reader devices. For Bluetooth studies in Texas, the primary means for collecting the data for O-D studies are via the use of portable readers that are deployed at select locations (Figure 12). The reader set-up includes a hard-case container that houses batteries, a hard drive, and an antenna that detects the Bluetooth signals. As soon as a signal has been detected, it is immediately transmitted to the host using a cellular modem. The data are also stored locally on each of the field devices as a backup in case there is an interruption in the cellular communication.



**Figure 12. Photo. Installation of a TTI mobile Bluetooth reader.**

A portion of a MAC address contains a manufacturer code that can be used to ascertain the source of the Bluetooth signal. For O-D studies, the primary source has been trending away from phone signals and toward signals from in-vehicle navigation systems.

#### 5.1.4 Privacy and Anonymization

Data collection and privacy are always a concern. Bluetooth MAC addresses detected and collected in the field cannot be linked to any person or business as they are random alphanumeric strings. MAC addresses by nature are not associated with any specific user or vehicle. The TTI system does not attempt to create a two-way connection to a user device, but rather it listens for devices that may be broadcasting. Additionally, the TTI data collection process includes the stripping of parts of the MAC address and encryption of the data before transmitting the data back to the server(s).

#### 5.1.5 Chief Purpose and Uses

In Texas, Bluetooth O-D data collection has been used to provide data for various purposes. This includes E-E trip table development, corridor studies, bridge and highway alignment studies, and ferry wait time assessments. Outside of Texas, Bluetooth is being used for many of the same purposes. Washington, D.C., and Indianapolis, IN, have used the technology to assist in the analysis of post-event traffic.<sup>(40)</sup> The Florida DOT sponsored a study of the IH-75 and SR 826 managed lanes in south Florida and the Oregon DOT funded a project geared toward the development of a wireless roadside data collection system utilizing Bluetooth technology.<sup>(27)</sup> The Maricopa Association of Governments (MAG) in Phoenix used Bluetooth data to assist in the development of an airport ground travel model<sup>(30)</sup> and Bluetooth studies in Baltimore, MD, have assisted in the evaluation of signal timing improvements and coordination plans.<sup>(40)</sup>

#### 5.1.6 Types of Trips Collected with Bluetooth-Based O-D Data

The use of Bluetooth data collection for O-D purposes has typically been used for E-E trip table development within a modeling area. Another possible use of the data collection methodology is

long distance travel between cities. While more difficult to accomplish, Bluetooth data can be used for E-I/I-E trip table development if there are enough Bluetooth reader devices to saturate the area.

### 5.2 Sample Size, Collection Timeframe, and Saturation

The amount of data collected and the duration is determined by the agency collecting the data. In Texas, generally a three-day (72 hour) weekday period is used when collection E-E data. Longer periods have been used if more weekday data are needed or weekend data are requested. Generally, daily averages are used in an effort to minimize variance.

Table 5 provides a summary of sample sizes from several Bluetooth data collection efforts. The table shows the average daily traffic count for all the sites included in the study and the average daily number of Bluetooth observations for those same sites. In an effort to illustrate the comparability of Bluetooth observations to that of previously used roadside intercept surveys, Table 6 shows an overview of data from a few external surveys performed in Texas. It includes the traffic count total for all the surveyed sites, the total number of surveys collected, and the percent of vehicles surveyed. The traffic counts are 24-hour totals but the surveys were only collected during daylight hours (typically between 7:00 a.m. and 7:00 p.m.).

**Table 5. Bluetooth reads and match percentages.**

Area	Sites	Avg. Daily Count	Avg. Daily Reads	Percent Match
Tyler, TX (2014)	20	175,343	12,975	7.4%
Austin, TX (2013)	14*	2,219,908	118,365	5.3%
Omaha, NE (2013)	22	235,526	10,470	4.4%
Corpus Christi, TX (2013)	9	226,098	20,059	8.9%
Bryan/College Station, TX (2011)	13	92,076	7,003	7.6%

\*Permanent Bluetooth reader locations.

**Table 6. Percentages of traffic surveyed from roadside interviews.**

Area	Sites	Avg. Daily Count	Avg. Daily Surveys	Percent Surveyed
Waco, TX (2007)	15	66,896	4,557	6.8%
Dallas-Ft. Worth, TX (2007)	32	173,867	12,642	7.3%
Austin, TX (2004)	22	111,113	8,298	7.5%
Tyler, TX (2003)	18	75,775	5,124	6.8%

### 5.3 Accuracy of Bluetooth Reads

TTI's Bluetooth equipment reads are all or nothing, meaning that if the equipment captures a MAC address, it captures 100 percent of the alphanumeric string making up the MAC address. The readers do not capture partial reads, so the MAC addresses collected are complete and 100 percent accurate. By the reads being accurate, it means that the matches between readers will not be underestimated by virtue of the same MAC address being correctly read at one location, but incorrectly read at another location with the study area or corridor.

## 5.4 *How Bluetooth Data Are Processed and Analyzed to Develop E-E Matrices*

### 5.4.1 Processing and Analyses of Raw Data

Raw Bluetooth data primarily consists of a MAC address (that has been anonymized), a time stamp when the address was observed, and a signature of the Bluetooth reader device. The process for retrieving the data depends on the equipment used to collect/detect the data. Some equipment allows for real-time transmittal of data back to a server via cellular signals while others store the data on internal hard drives and are retrieved manually.

For studies in Texas, data from all Bluetooth readers in the field are compiled in one location. Typically, both permanent and portable Bluetooth readers transmit the collected data wirelessly back to a main server. At that point, algorithms are used to process the data and output those MAC addresses that were observed at two or more reader locations. Those matched MAC addresses are then compiled into an E-E matrix. The TTI process uses a first-to-first methodology of matching the first read of each MAC address at the origin with the first read occurrence at the destination.

### 5.4.2 Estimating E-E Trips Using Travel Time Thresholds

Travel time thresholds are used to remove those trips that are suspected to be E-I/I-E trips, leaving only E-E trips. The general premise is that if it takes longer than a pre-determined amount of time for a vehicle to travel between two data collection locations, then an assumption is made that the vehicle stopped between the two locations and is an E-I/I-E trip. Travel time thresholds can be developed via various means including manual travel time runs or using the travel time skim matrix values from the travel demand model.

### 5.4.3 O-D Matrix Development, Expansion, and Balancing

After E-I/I-E trips are removed from the analysis, the remaining matched trips need to be expanded in order to reflect an estimate of the total amount of E-E trips for the area being evaluated. While there are various methods to expand the raw results, the method used by TTI involves using traffic count data collected at each Bluetooth data collection location. An expansion factor is developed for each location. The expansion factor is based on the proportion of Bluetooth MAC addresses recorded as compared to the overall traffic count. If the E-E results are to be used for model development, the expanded data are balanced so that they do not create migration problems in the model.

The use of traffic count data provides control totals to help ensure that the expanded Bluetooth results are reasonable. To date, there have been no anomalies with Bluetooth data that were collected as part of the various TTI studies.

## 6.0 Studies Using Cell, GPS, and Bluetooth O-D Data

### 6.1 Cellular O-D Studies

Over the past 5–6 years, many state and local agencies have purchased cell data to provide O-D information and matrices for regional models or corridor studies. Cell developed O-D results have been used for model calibration and validation and compared to model or prior survey results for trips by purpose, trips by time period, trip distribution, and TLFDs. Another important use of cell data has been to estimate resident versus non-resident/visitor travel and flows. This section provides summaries of select studies across the United States in recent years that have used cell O-D data.

#### 6.1.1 Chattanooga Cell Phone External O-D Matrix Development-Process and Findings

In 2010, the Chattanooga–Hamilton County (TN) Regional Planning Authority conducted an external O-D study using cellular data. Data collection occurred for a five-day period in December 2010. The study area was divided into 900 grid cells and included 38 external zones. The grid cells corresponded to cell phone tower coverage areas. The initial results using the disaggregated 900 grid cell set-up were unsatisfactory as several major O-D movements were missing or low. As a result, the grid cells were aggregated into 21 mega zones to improve the cell data accuracy. The resulting total number of E-E trips obtained using the mega zone approach method was found to be much lower than that obtained using a frateded approach (18,855 versus 35,512), which may have largely been because the cell data used were only a sample of the total number of trips made and only vehicles with a cell phone connected to the network were captured. The grid cell size and the 12-hour I-E cutoff assumption may have led to some known lower-class facilities having volume estimates that were too high. Small grid cells may have been more desirable. While a full, more expensive E-E study would have been more desirable, the cellular O-D methodology was found to be a good starting point.<sup>(41)</sup>

#### 6.1.2 Mobile County O-D Study

Cell data were used by the South Alabama Regional Planning Commission (SARPC) to calibrate their travel demand model. A total of 312 traffic zones were used for the study. The study included data collected on Tuesdays through Thursdays in May 2012. A total of 192,107 devices and 1,560,501 trips were included in the sample with an average of 67 locations being recorded per device per day. A comparison of internal trips by purpose was performed using the area's 2007 model data, Airsage 2012 data, and National Cooperative Highway Research Program 2009 data ranges. Using an iterative process, the model was calibrated by matching the modeled data to the observed cellular phone data based on TLF, average trip length by trip purpose, and area-to-area trip flows. The model was also validated to ensure that the assigned volumes and counted volumes were comparable. Relative to cell data, the results found a slightly lower number of HBW trips and higher totals for HBO and NHB trip purposes. These results could be explained by the fact the cell data represent trip chains, so portions of the home to work tour are reflected in higher levels of HBO and NHB trips. Visitor trips not normally captured in a household survey would show up in the cell data as NHB trips. Additionally, researchers determined that the cell data sample was not of sufficient size or density to develop link travel times.<sup>(42)</sup> Positive takeaways from this study included the ability to use cell data in capturing visitor trips and in developing trip matrices by purpose. Most importantly, SARPC was able to calibrate and validate its base year model using cell data in

tandem with travel time information, census/American Community Survey data, and traffic count data.<sup>(43)</sup>

### 6.1.3 O-D Study and Corridor Analysis for Moore County, NC

In 2012, the North Carolina Department of Transportation (NCDOT) used cellular data to assess the O-D characteristics and composition of traffic on the Route 1 corridor in Moore County, NC, and to develop trip tables for the area's travel demand model. The purpose of the corridor study was to gain information to address concerns related to a proposed controversial bypass planned in the area. Cell data were used because they were believed to be an unbiased, non-intrusive option that was cheaper than performing a traditional O-D survey.<sup>22</sup> Cell data were acquired for a one-month period in September/October 2012, but data from only 12 weekdays in this timeframe were used in developing the O-D matrix. Grid cells were generally defined to be 1,000 meters by 1,000 meters and were used in determining trip information relative to the cellular device movements. To aid in data visualization, the TAZs were grouped into districts and associated desire lines were developed. Select link analyses were performed for key stretches of the Route 1 corridor to estimate the resident versus non-resident and pass-through composition of traffic along the corridor. The study results helped communicate and quantify to residents the composition of traffic along the region's primary north-south corridor and helped establish the corridor as a one of regional and statewide importance rather than just local. In this study, the perceived objective nature of cell phone data allayed resident concerns of biases in other sources of data. The researchers stress that it is important to begin with the end in mind when developing these types of studies and to take unique regional occurrences into account.<sup>(44)</sup>

### 6.1.4 Reconciliation of Regional Travel Model and Passive Data Tracking

In 2013, a comparison of modeling results obtained as part of the Triangle Regional Model (TRM) and cellular data obtained from Airsage was performed in the Research Triangle area of North Carolina. The Airsage data were processed to represent the morning peak hour and subsequently assigned to the highway network. The TRM data were also assigned to the network. To make the resulting models more comparable, the external trips from the TRM model were added to the Airsage model, because only internal trips were originally associated with the Airsage data. Likewise, the Airsage trips were originally person trips and had to be converted to vehicle trips using vehicle occupancy information. Multiple comparisons were performed between the Airsage data results and the TRM data results. The authors report that, "The metrics reported and compared include trip length distribution, district-to-district flows, highway assignment summarized by functional classification and volume group, estimated to observed plots, and system wide traffic flow comparisons." Overall, the authors found the two methods to produce similar results. These findings support further use of passively collected cellular data for use in modeling—especially given the relatively low cost associated with the resulting large sample size. However, the information available from cellular data does not provide the detail needed for adequate development of behaviorally based models. A recommendable approach may be to combine the use of passively collected cellular data with more traditional approaches for modeling purposes.<sup>(45)</sup>

### 6.1.5 Using Mobile Phone Location Data to Develop External Trip Models

NCDOT used cell data to develop an external travel model for the French Broad River Metropolitan Planning Organization in western North Carolina. The data were obtained from Airsage for weekdays during the month of May 2013. The goal of the study was to evaluate the usefulness of cell data in developing external trip tables as a potential lower-cost alternative to traditional and synthesized methods that had been used in the past. The study area's TAZs

were aggregated into 139 larger districts and 11 large external districts were created around the periphery of the study area. These 150 districts were used for cell data capture and analysis. Summary results from the cell data on trips by purpose, trips by time period, and resident versus non-resident status were compared to the area's household survey and found to be similar. Initial review of the data revealed that trips between adjacent external stations were included in the results. Further processing of the cell data was needed to remove the erroneous through trips. A two-step disaggregation process was used to factor data from the external districts to individual external stations. E-E trips were reduced so results more accurately reflected results from counts. After estimating E-E trips, E-I and I-E trips were estimated. Productions were assumed to occur at the external stations and all attractions were assumed to occur in the TAZs. With this in mind, linear regression was used in developing trip attraction rates for the E-I/I-E trips, and the gravity model was used in estimating E-I/I-E trip distribution. Travel patterns modeled with the cell data validated well against observed traffic counts. The study found that cell data are a useful source for the development and estimation of external trip models that represent observed local travel patterns.<sup>18</sup> The study's authors suggest that the lower cell data cost makes it more accessible for small and medium-sized communities.

#### 6.1.6 Using Cellphone O-D Data for Regional Travel Model Validation

Airsage cellular data from October 2013 were used by RSG, Inc. in performing a study to validate a regional four-step travel model in the Syracuse, NY, area. The modeling area consisted of 1,185 TAZs, which were aggregated into 788 larger zones for cellular data capture. The study found that the cell data were better than model results at differentiating trip purposes associated with HBO trips than with HBW trips or NHB trips. Additionally, while work trip destination was correlated with zone level employment locations in the model, the same could not be said for the cell data. A case study of the University of Syracuse was performed, with findings indicating that the total number of trips obtained from the cell data was too low for an area of this importance, resulting in only about a third as many total trips being estimated as were estimated from the model. The time of year the data are collected may impact trip purpose splits. The cell data and model trip length distributions matched fairly well, as did the combined HBW and NHB TLFDDs. Some of the takeaways include deciding as an agency what the greatest level of acceptable data aggregation is and to design the zones accordingly. Additionally, it is important to strike a good balance when determining zone size. If zones are too small, it can lead to missing many E-E, E-I, and I-E trips; whereas, if zones are too large, it can lead to an excess of E-E trips being captured. Additionally, when performing a select link analysis, it is recommended that a long link length be selected and that particular care be taken when performing these types of analyses.<sup>(46)</sup>

#### 6.1.7 Development of the Idaho Statewide Travel Demand Model Trip Matrices Using Cell Phone O-D Data

A Statewide Travel Demand Model (STDM) was begun for Idaho in 2013. The project had two phases, and the first phase (which was the focus of the cited presentation) involved using cell phone O-D data to synthesize travel demand. Auto and truck trips were included in the O-D matrix estimation and analysts used volumes by user class, level of service measures, and the TREDIS economic model to evaluate performance. The statewide zone system of over 4,600 TAZs was aggregated into 750 super zones for cell data capture and analysis. One month of cell data was acquired for the study. Estimating the O-D matrix involved an iterative process until the observed and estimated traffic volumes (by user class) converged to an acceptable level. In terms of the O-D matrix estimation results, the analysts reported a "reasonable goodness-of-fit between synthesized travel demand and limited observed data across multiple

dimensions-user class, facility type, geography.” However, analyst indicated that ODME naively adjusts the travel demand to match the traffic counts and this can result in over-fitting. Analysts warn that this type of O-D matrix estimation should only be used for short-term forecasting. Analysts say that cell phone O-D data are a reasonable starting point for generating statewide trip matrices, but a better understanding is needed on how cell data flows are different than traditional travel modeling data sets.<sup>(47)</sup>

### 6.1.8 Understanding Cellular-based Travel Data

In this study, MAG compared 2013 Airsage cell data to the 2011 MAG travel demand model, which was developed using 2008 National Household Travel Survey data. An O-D trip matrix was developed using Airsage data with a 30 × 30 zone configuration for weekday travel in October 2013. Resident and visitor data were considered separately and the data were grouped by morning, mid-day, evening, and daily data. Based on the comparison, it was determined that the daily total trips matched pretty well between the Airsage data and the MAG model; however, for some areas the number of peak period trips varied between the estimates. The number of trips by zone matched fairly well between the Airsage results and the MAG model. On the other hand, the Airsage data did not appear to capture as many short-distance trips as the model, which led to differences in intrazonal trips being estimated between the two methods. Airsage data were also found to be consistently higher during peak periods than the model estimates.<sup>(48)</sup>

### 6.1.9 Preliminary Evaluation of Cellular O-D Data as a Basis for Forecasting Non-Resident Travel

The Metropolitan Washington Council of Governments (MWCOC) used cell data to study external and visitor/tourist O-D patterns for input to MWCOC model. The study included 3,722 internal cell zones and 12 external cell data capture areas around the modeling area. It included cellular data for Tuesdays, Wednesdays, and Thursdays from April 2014. The study noted that with a mobile device trip a lot of information is not known, including household characteristics of the traveler, who the individual trip-makers are, the vehicle type, the mode type, and the path of the O-D. It also referred to trips resulting from cell data as aggregate O-D flows. The study compared the 2015 MWCOC model output to 2014 cellular trip data. The total number of trips obtained using the two methods was comparable, though cell trips had a higher share of home-based trips and a lower share of non-home trips. Comparisons were also performed on internal trip lengths by purpose, jurisdictional daily trip flows, home-based productions versus households, HBW attractions versus employment, external/through trips, and E-I trips by jurisdiction in terms of cell results, and various other model or survey results. On multiple occasions, data aggregation was found to help reduce some noise in the data. In comparing cellular O-D trips to modeled person trips, it was found that the global trips and trip lengths were comparable. However, when disaggregated by model purpose there are more noticeable differences. In terms of O-D external and through trips, the cellular data exceeded the count data by 30 percent. The authors note that there may be limited ability to ground-truth cellular O-D movements with ground counts. The cellular results obtained for non-resident visitor trips generators were to be expected. In using cellular data in the future, it is important to remember that they are different than typical data used in modeling. Therefore, it is important to understand how they are different and to address known uncertainties associated with cellular data.<sup>(49)</sup>

*The study noted that with a mobile device trip a lot of information is not known, including household characteristics of the traveler, who the individual trip-makers are, the vehicle type, the mode type, and the path of the O-D.*

## 6.2 GPS O-D Studies

Several cutting-edge GPS O-D application studies are in their infancy, as attested to by the fact that IH-494/TH62 Congestion Relief Study and the Maryland Freight Fluidity Study summarized within this section had not yet been published at the time of this report. This speaks to the continually evolving nature of O-D technologies.

### 6.2.1 Early TTI Studies Using GPS Data from Data Providers

For many years prior to development of GPS based O-D products, TTI partnered with several data providers and worked under non-disclosure agreements using their data to examine the quality and feasibility of using third-party GPS data to develop E-I/I-E and E-E trip tables. The purpose of this research was to determine if such data could be used to augment or replace external tables developed from traditional travel survey means. Key items included in the research were to examine:

- The saturation of the data, its geographic coverage, and its temporal distribution.
- The quality of the trip traces and the extent of gaps in the data, which was highly dependent on the ping frequency rate of the GPS point data.

In a 2011 study, TTI partnered with a private data provider and developed a methodology and technical criteria to process its GPS database to develop external trips for a large urban area. The approach involved identifying and isolating GPS data streams that crossed-specified external stations for a pre-defined period of time, then separating these data streams by direction. Nine months of data were analyzed based on a pre-defined definition a trip. For legal and privacy reasons, the provider was unable to allow TTI access to the raw data. The lesson learned from this effort was that GPS data could be used to develop and analyze external trips, but the data sample/penetration was much too low (at that time) for it to be used to develop trip tables.

In a 2013 study, TTI obtained a nationwide data set of raw GPS data from another data provider to assess external trip making. The data set included 1.2 billion GPS probe records from numerous mobile GPS data sources. TTI developed algorithms and procedures in Python/R with geospatial analysis to process the data set (~20–30 GB/day) and reduce it to a specific time period and geographic region of interest. The analysis converted raw GPS points to trips based on stop dwell times and spatial criteria and re-identified and re-linked parts of trips from unique IDs. The re-identification and re-linking was necessary because the IDs were temporary and only persisted for 20 minutes. Trip ends identified within the study area were geocoded to the area's TAZ system. If no trip ends were identified in the study area, the external station where the trip exited the study area was identified. The resulting trips were subsequently processed to develop E-I/I-E and E-E trip tables for the urban area. Similar to the 2011 study, the lesson learned from this effort was that GPS data could be used to develop and analyze external trips, but the data sample/penetration was still low.

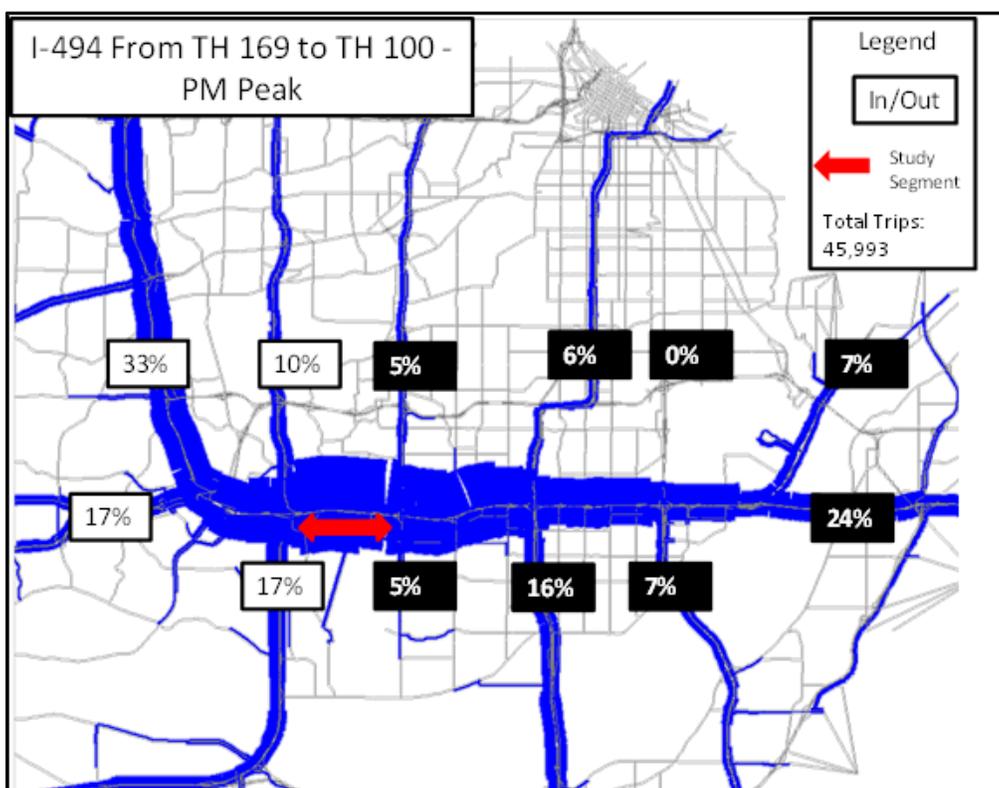
### 6.2.2 IH-494/TH62 Congestion Relief Study

This study, sponsored by the Minnesota DOT, was still in process at the time this report was prepared (March 2016). It used three months of INRIX GPS data from spring 2015 to provide trip O-D and routing patterns to support a congestion relief study along IH-494/Highway 62 in the Twin Cities Metropolitan Area of Minnesota.

The GPS data set included about 800,000 trips within the study area for filtering, about 300,000 trips were used in the final analysis. Each GPS trip waypoints to provide information about route, with the average trip

waypoints. The section of highway studied was segmented to better being physically inconsistent, for variations in volumes and patterns in corridor, and to allow for different pricing alternatives to be implemented. A multistep route-matching process using ArcGIS was used to convert the that was understandable to the regional travel demand modeling software. to study regional movements by route, travel patterns by corridor segment, tables. According to the consultant, SRF Consulting Group, the GPS O-D world trip data into the project on a scale not previously feasible. The data qualitatively assess travel markets and routing choices by travelers and patterns to calibrate the travel model.<sup>(50)</sup>

Figure 13 shows the sources of Os and Ds stemming from a segment of the IH-494/TH 62 study corridor.



Source: (51).

**Figure 13. Map. Travel routes of vehicles from segment of IH-494/TH 169 corridor.**

### 6.2.3 Maryland Freight Fluidity Study, Fall 2014–Spring 2017

This on-going study, sponsored by the Maryland DOT-State Highway Administration (MDOT-SHA), is using INRIX O-D data containing GPS trip records with waypoints to study freight movements and freight corridors in Maryland. With their unique expertise in the freight mobility area, TTI is assisting MDOT-SHA through the University of Maryland. The O-D data are for the months of February, June, July, and October 2015, including about 20 million trips, 1.4 billion waypoints, 5.5 million unique devices from 148 different data source providers. A total of

60 percent of the trips were from commercial fleet sources, 31 percent were from consumer sources and the remaining 9 percent were from mobile sources.

The study will focus on commercial fleet trip quantities and movements for the highway network to study performance measures such as travel time, delay, how much freight is moved, and the costs of delay. Researchers anticipate analyzing commercial trips by vehicle weight class and provider profiles such as local delivery fleets and private trucking fleets. One deliverable planned for the project includes development of an interactive map that shows flows to and from zip codes that provides an assessment of freight zone-to-zone accessibility. This will answer questions such as where did most freight trips from each respective zip code begin and end, how many trips stayed internal to the zone, how many were traveling through, and which zones carry the bulk of freight traffic. Other potential work includes analyzing the O-D data for the states most congested corridors to identify what segments are being under and over used and potential alternative routes.<sup>(52)</sup>

### 6.3 Bluetooth O-D Studies

#### 6.3.1 Generating Route-Specific O-D Tables Using Bluetooth Technology

Bluetooth data were collected using 14 sensors placed along a 15-mile corridor in Jacksonville, FL. Data collection occurred from February 18, 2011, through February 25, 2011. The minimum distance between sensors was slightly less than one mile, and the maximum distance between sensors was 5 miles. Both O-D matrices and travel times were developed using the data; however, the O-D matrices development was the focus of this paper. A hybrid approach to Bluetooth sensor placement was used to enable both types of studies to be performed. The authors were especially interested in developing a route-specific O-D matrix. Several steps of cleaning, and sequential trip link development, were required to enable this analysis. Through the data cleaning process, the number of device scans was reduced from 253,367 to 124,624; additionally, the number of unique MAC IDs was reduced from 33,789 to 23,614. An expansion factor was developed based on the capture rate of 6.135 percent. The authors concluded that Bluetooth can be used for an O-D corridor analysis with multiple routes. They also note that they were able to use the data collected in after-model validation for toll revenue forecasts.

#### 6.3.2 Bryan-College Station Bluetooth External Survey

Bryan/College Station (BCS) was selected as the site for a test to determine if Bluetooth technology could be used to determine E-E trips for urban areas in Texas. BCS was selected because it would allow for comparison with an external study that was performed in the area in 2002. A total of 13 external station locations were identified for data collection. Those trips with a travel time higher than a specified threshold were considered to be external-local trips. Adjustments were made in the analysis to ensure that Bluetooth data collected from multiple devices in the same vehicle were not both included. Vehicle classification count (VCC) data were used for expansion purposes. Data were collected the week of August 15, 2011, with data from Tuesday through Thursday being analyzed. Approximately 8 percent of vehicles were estimated to contain Bluetooth devices. After expanding the results, TTI found that the 2011 Bluetooth results contained 60 percent more E-E trips than what was obtained in the 2002 traditional external survey. Part of this was a reflection of increased traffic volumes between 2002 and 2011. The fact that data were not collected at night in the 2002 survey may have also contributed to the discrepancy in E-E trips between the two surveys.

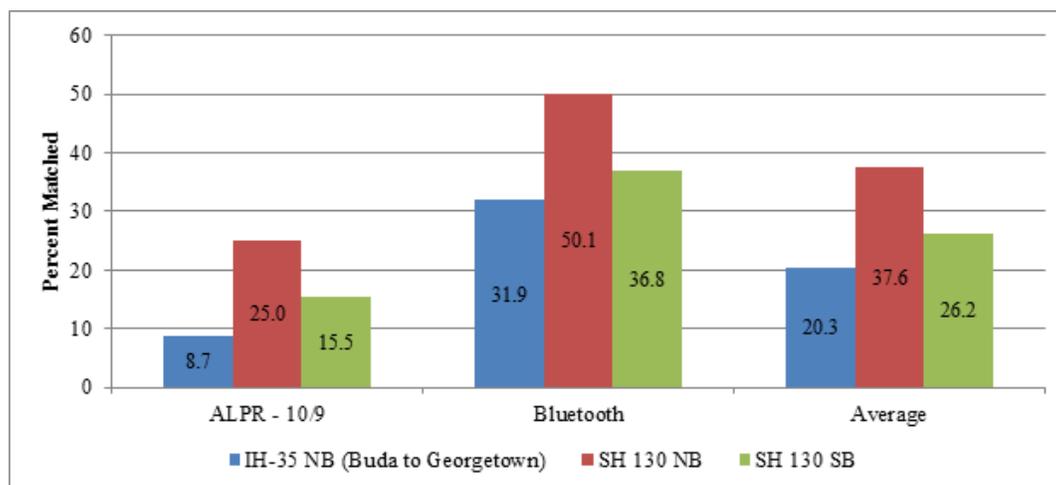
### 6.3.3 TTI Bluetooth O-D Studies for Route Information

In addition to using Bluetooth for external O-D studies, TTI has used Bluetooth data to conduct other types of studies to aid in making decisions about tolling along a corridor, the alignment of a major bridge, and the need for a bypass route around a community. Summary information on each case study is provided in the following paragraphs.

The SH 130/IH-35 Traffic Diversion Study in the Austin, TX, Region. In spring 2013, a one-year truck toll reduction period was implemented on SH 130 east of Austin to improve safety and reduce congestion on IH-35 through central Austin by diverting more (commercial) trucks from IH-35 over to SH 130. To measure the effectiveness of this strategy, this study was conducted to collect O-D and traffic count data using Bluetooth readers, ALPRs, and VCCs to provide estimates of traffic being diverted from IH-35 to SH 130. The study focused on commercial truck diversion, but also obtained results for non-commercial vehicles.

The Bluetooth data were collected over a three-day 72-hour weekday period in fall 2013 at 16 locations along IH-35 and 13 locations along SH 130. ALPR cameras were used at the IH-35/SH 130 interchange on the north and south sides of the Austin area to collect data on directional and through movements by commercial and non-commercial vehicle classes. A total of 16 ALPR cameras were needed to capture all movements. Due to their expense, the cameras collected data for only 24 hours of the 72-hour period when the Bluetooth units were deployed.

The three key results from the study included the percentage of vehicles that traveled all the way through the Austin area, the percentage of southbound vehicles on IH-35 that divert to SH 130, and the percent of northbound vehicles on IH-35 that divert to SH 130. Figure 14 shows the percentages of vehicles traveling through the Austin area on IH-35 based on Bluetooth and ALPR. ALPR results were significantly lower than the Bluetooth results. The lower ALPR results could be a result of inaccurate reads and ALPR cameras only capturing about 70 percent of the license plates of the passing traffic (when compared to traffic counts).



**Figure 14. Graph. Comparison of through trips on IH-35 in Austin by Bluetooth and ALPR.**

The study found that only a small percentage of commercial vehicle traffic was diverting from IH-35 through the heart of the Austin area to use SH 130. For the southbound direction, the results found that about 4,900 vehicles per day (15.4 percent of the traffic) on IH-35 diverted to SH 130 and that about 380 of these were commercial trucks. Of these southbound vehicles that



Highway Bypass Study in Alice, TX. Alice is located along SH 44 in South Texas about 50 miles west of Corpus Christi. SH 44 is part of an east-west route used by commercial traffic traveling between the Port of Corpus Christi and international Texas/Mexico bridge crossings in Laredo, TX. This study was conducted to determine the need for a relief bypass route around Alice. The study used Bluetooth readers in combination with ALPR cameras and VCCs to assess O-D patterns around and through the Alice area. Key objectives of the study were to: 1) estimate the amount of non-commercial and commercial traffic on SH 44 that travels through the Alice area; and 2) estimate the amount of commercial east-west traffic traveling through the Alice area that bypasses the city using routes to the north or south.

TTI collected Bluetooth data at 20 different locations during the Tuesday through Thursday timeframe of March 19–21, 2013. Overall, the Bluetooth detection rate was 8 percent of vehicles that passed a device location. ALPR data were collected at seven locations for a 24-hour period, with two cameras (one for each direction) used at each location. Using the ALPR data, it was possible to determine the distribution of state and county license plates. The various data collection efforts allowed for the following comparisons:

- ALPR and VCC Methods: ALPR cameras were associated with 17,800 fewer observations than the VCC method. In comparing vehicle type results, ALPR methods had a higher percentage of non-commercial vehicles than the VCC method (83 percent versus 80 percent).
- ALPR versus Bluetooth O-D Trip Tables: In comparing the Bluetooth data collected during the same 24-hour time period as the ALPR data, the Bluetooth method was associated with 5,200 more O-D matches than the ALPR method. Overall, the two methods were found to produce very similar O-D matrix estimates. Similar results were found for both technologies in comparing the through trips on SH 44.

## 6.4 Studies Comparing Technologies

### 6.4.1 Tyler External Survey Comparing Cell, GPS, and Bluetooth O-D Data

This study was conducted by TTI in coordination with TxDOT in spring 2014. The scope was to collect, analyze, and compare external O-D data for the Tyler MPO study area (Smith County, TX) using Bluetooth technology provided by TTI, cellular data provided by Airsage, and GPS data provided by INRIX. The chief purpose of the study was to determine if cell and GPS data were viable for use in TxDOT's external surveys so that TxDOT could resume collecting external data for Texas MPOs.

At the time, Airsage recommended zones for cell data capture be a minimum of 500 meters x 500 meters in size. Based on this criterion, the Tyler study area's 420 TAZs were aggregated into 307 larger zones. Figure 16 illustrates the internal study area with 307 TAZs, the 17 travel sheds around the study area for cell data capture, the 20 external stations for the study, and an approximate 10-mile buffer around the study for GPS data capture.

One month of cellular data was acquired for the study. It included data for average weekdays and weekends, AM and PM peak data, and 24-hour totals. The cell data acquired included 198,344 unique devices, which represented a 17.1 percent sampling rate of the study area population. Airsage reported an average of 180 device sightings per day for the study area.

Three months of pre-processed GPS data from INRIX was acquired for the study. INRIX provided separate data sets for GPS data sources from cars, freight/commercial fleet vehicles, and mobile applications. The data set included 492,023 usable trip records. Each record contained a unique device ID and the trip ID with time-stamped trip end locations. The GPS trip

ends were pre-processed by INRIX using a 10-minute dwell time threshold and the time stamps for trip ends were rounded to the nearest hour. The distribution of GPS data by sources for freight, cars, and mobile applications were 57 percent, 16 percent, and 27 percent, respectively. Using the frequency of the GPS passes and the traffic counts collected at the external stations, TTI calculated the sampling rate of the GPS data to be about 1.6 percent for the study area.

TTI collected Bluetooth data at 20 external stations for a two-week period and used data from Tuesdays, Wednesdays, and Thursdays to develop average weekdays results. Over 170,000 Bluetooth observations were captured during the survey period. Across all external stations Bluetooth detection ranged from about 4–11 percent. There were nearly 4,100 weekday matches and over 1,800 weekend matches. Using the expansion factors developed for each Bluetooth location, the raw data were expanded and balanced to produce O-D matrices.

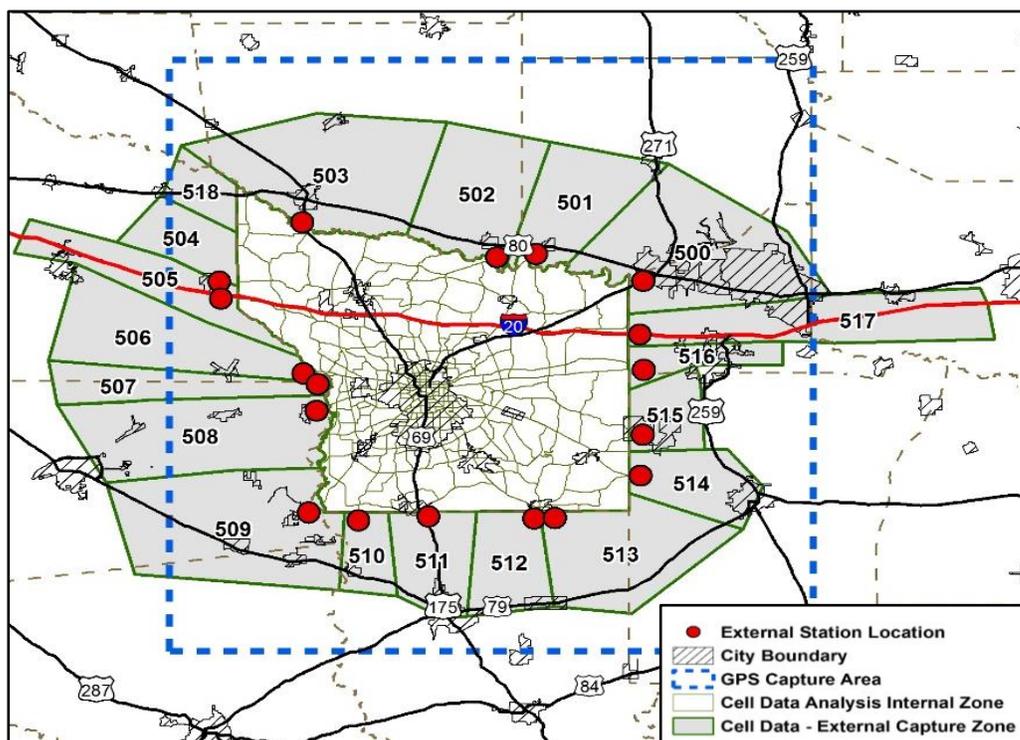


Figure 16. Map. Study area design for collection of cell, GPS, and Bluetooth data in Tyler, TX.

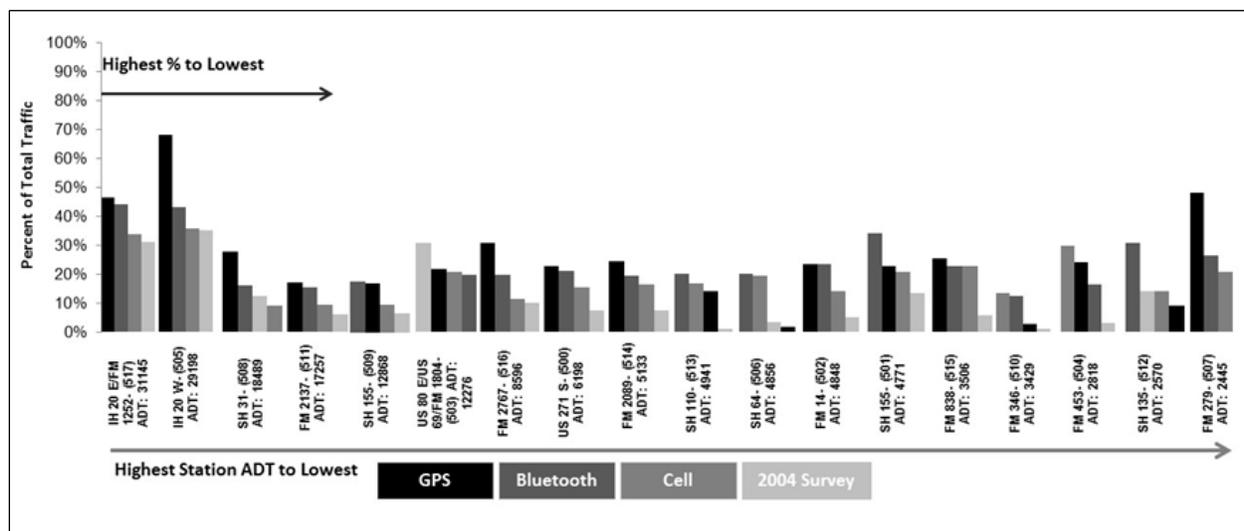
Based on TTI’s extensive use and experience with Bluetooth technology and E-E data collection, agency researchers believe it is a reasonably accurate means for estimating E-E trips for, so they used it as a benchmark comparison to the cellular and GPS data. Based on this premise, if the E-E estimates derived from the cellular and GPS data compared well to the Bluetooth E-E estimates, the E-I/I-E estimates from those same sources could be considered reasonable.

Table 7 shows a summary comparison of the total and percentage E-E trips between the 2004 survey data and the 2014 Bluetooth, GPS, and cell data. The table also shows E-E results for Bluetooth and cell data broken down by non-commercial and freight/commercial categories. Results for cell data in these categories are not provided since cell data cannot distinguish between vehicle classes.

**Table 7. Bluetooth, GPS, and Cell E-E results in Tyler, TX.**

Vehicle Category	2004 Survey		Bluetooth		GPS		Cell	
	EE Trips	% EE	EE Trips	% EE	EE Trips	% EE	EE Trips	% EE
Non-Commercial	18,714	15.5%	29,925	25.9%	28,864	25.0%	na	na
Commercial/Frieght	8,907	25.0%	18,626	31.2%	30,258	50.7%	na	na
Total	27,620	17.6%	48,551	27.7%	59,122	33.7%	34,386	18.3%

Figure 17 shows the percent of E-E trips to total traffic by external station for the 2004 survey and the 2014 Bluetooth, GPS, and cell data. The chart shows predominantly low levels of E-Es for the 2004 survey and mostly higher levels of E-E for GPS, which are consistent with the E-Es totals shown in Table 7.



**Figure 17. Graph. Comparison of Bluetooth, GPS, and cell E-E results by station in Tyler, TX.**

Figure 17 shows that the majority of cell E-E results by station are lower than those of Bluetooth and GPS. These results are especially low since they were developed based on a 24-hour period, unlike the Bluetooth and GPS results where travel time constraints were applied. The time constraint used was the time it takes to travel between each external station, plus about a 20 percent cushion.

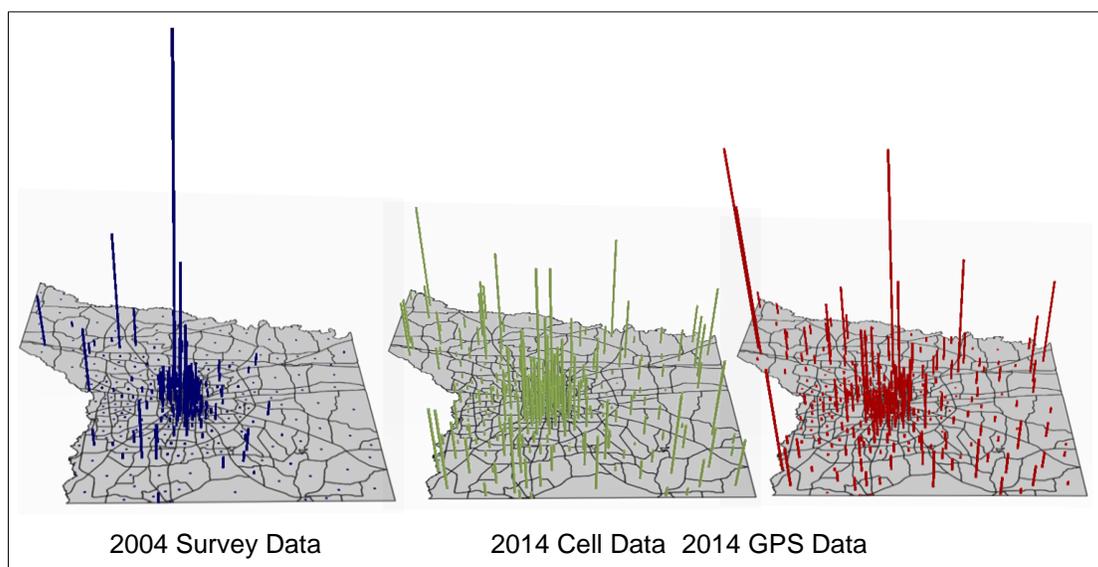
The Bluetooth and GPS results show a considerable increase in E-E trips through the Tyler study area from 2004 to 2014. The higher number and percent of E-Es for GPS is due in large part to a bias toward commercial vehicles. With few exceptions, commercial vehicles make a higher percentage of E-E trips than non-commercial vehicles. The GPS data for the Tyler study was comprised of 70 percent freight/commercial vehicles, 20 percent mobile applications, and 10 percent passenger vehicles.

The percent of total cell E-Es of 18.3 compares better with the percentages of E-E of Bluetooth and GPS for non-commercial vehicles, which were 25.9 percent and 25.0 percent, respectively. Researchers believe this may be a more accurate comparison, since the commercial bias

*The higher number and percent of E-Es for GPS is due in large part to a bias toward commercial vehicles. With few exceptions, commercial vehicles make a higher percentage of E-E trips than non-commercial vehicles.*

in the GPS data is mitigated and since cell data may have a bias toward non-commercial vehicles.

Figure 18 shows the distribution of E-I/I-E trips ends across the internal TAZ structure based on the expanded data for the 2004 survey, and the 2014 cell and GPS data. Bluetooth data are not included since it cannot be used to develop E-I/I-E trips on an area wide basis. It shows that the external trip ends for the 2014 cell and GPS data are clearly better distributed than those in the 2004 survey. This better distribution is likely due to increased traffic volumes, substantially more suburban development, and better sampling rates. The cell graphic shows unusually high trip ends in some of the rural periphery zones near the study area boundary. These results could be due to poor cell coverage and/or a lack of positional accuracy of cell data. The numerous high spikes in trip ends near the top of the GPS graphic are probably a result of the commercial bias in this data. These spikes are located along a major interstate (IH-20) that runs through the area and likely represent trip ends at truck stops.



**Figure 18. Illustration. Comparison of cell and GPS E-I/I-E trips in Tyler, TX.**

The E-I/I-IE trips and the Tyler model skims were used to develop and compare TLFDs between the GPS and cell data. A Kolmogorov-Smirnov test was used to test the statistical similarity of the cell and GPS TLFDs. TLFDs were not statistically similar, despite their appearing to be when charted.

## 6.4.2 NAPA County Travel Behavior Study; NCTPA Board Meeting Presentation; December 17, 2014

*Five different data collection methods (VCCs, winery regression analysis, license plate matching, surveys, and mobile device data) were used in performing a travel behavior study for Napa County, CA.*

Five different data collection methods (VCCs, winery regression analysis, license plate matching, surveys, and mobile device data) were used in performing a travel behavior study for Napa County, CA. There is a high level of visitor travel in the area due to its large number of wineries. One of the elements collected in the study was the travel patterns of visitors within the area, an element that few studies had captured previously. However, it was also important to understand the travel behavior of residents as well, so the study considered the travel

behavior of both residents and visitors. One of the goals of the study was to use the results to expand transit and paratransit services, as well as to improve travel demand modeling. In analyzing mobile device data, INRIX and Streetlight data were collected for a 61-day period from September 1, 2013, to October 31, 2013. StreetLight data used algorithms to infer the trip purpose, as well as origin and destination locations, which were tagged geographic layers of interest. Based on the results of the study, 55 percent of the sample was found to be internal trips, with the rest either being external or pass-through trips. The advantage of using mobile device data, compared to the other data collection methods, is that it is associated with a very large sample size. However, the trip purpose and O-D had to be inferred. In developing O-D trip tables, the analysts began with the mobile device data (given its large sample size), and then used the data obtained from the other four methods in refining it for a single day of data. This technique was necessary given that the cellular data obtained were representative of relative trips rather than absolute trips. It is possible to aggregate the O-D data to show flows between cities. The mobile device data were also found to provide large amounts of data for use in model calibration and validation.

*In comparing the different data collection methods, each had its strengths and weaknesses.*

In comparing the different data collection methods, each had its strengths and weaknesses. VCCs and winery regression analysis helped to provide control totals for comparison purposes for the other survey data collection methods, though being associated with little to no information on trip and demographic information. The license plate matching data were a good source for external data, but not for internal or inter-regional travel. The survey data provided insights into trip and demographic information but was expensive and associated with only a small sample. The mobile data obtained from INRIX and StreetLight Data provided valuable information on different types of trips—including internal trips—at a relatively low cost and with a large sample. However, the need to infer information and lack of demographic information associated with this mobile data were limiting factors. Given these strengths and weaknesses, the results obtained using the different data collection methodologies were all used in making comparisons to, and integrating with, existing model results.<sup>(54, 55)</sup>

## Appendix A: Key Takeaways from Select Studies

**Table 8. Takeaways from select cell O-D studies.**

	Cell O-D Study and Date of Report	Key Takeaways and Lessons Learned
1	Chattanooga Cell Phone External O-D Matrix Development  November 2011	<ul style="list-style-type: none"> <li>- Use of 900 grid cells was too disaggregate to capture cell data, but subsequent aggregation into 21 megazones, in tandem with 12-hour I-E cutoff, may have led to some lower-class facilities having too high of volume estimates.</li> <li>- E-E trips from megazones were much lower than fratedared approach.</li> <li>- Cell data collection period of only five days was probably too short; a longer data collection period was needed.</li> <li>- Compared to a full-scale traditional external survey, the cellular O-D approach was thought to be a good, lower-cost option.</li> </ul>
2	Mobile (AL) O-D Study to Calibrate and Validate Travel Demand Model  January 2013	<ul style="list-style-type: none"> <li>- Cell data can be used to capture visitor trips and in developing trip matrices by purpose.</li> <li>- Cell data, in tandem with travel time data, census/ACS, and traffic count data can be used to calibrate and validate model.</li> <li>- Cell data were found to produce fewer HBW trips, but higher HBO and NHB trips.</li> <li>- It is thought that cell data represent trip chains, and thus portions of the home to work tour, are reflected in higher levels of HBO and NHB trips.</li> <li>- Trip distribution can be skewed by service area.</li> <li>- The sample of cell data for the study was not sufficient to develop link travel times.</li> </ul>
3	O-D Analysis for Moore County, NC for Corridor and Select Link Analyses  July 2013	<ul style="list-style-type: none"> <li>- Cell data can be a faster and lower cost alternative to traditional O-D survey methods.</li> <li>- Cell data can be used in corridor studies to study pass-through and resident vs. non-resident make-up of traffic along a corridor.</li> <li>- Begin with the end in mind when scoping these types of studies.</li> <li>- Cell data are a useful source for the development and estimation of external trip models.</li> </ul>
4	Reconciliation of Regional Travel Model and Passive Data Tracking  Triangle Region, NC  January 2014	<ul style="list-style-type: none"> <li>- This study compared trip length distribution, district-to-district flows, assignment by functional class, and system-wide traffic flows between cell and model results.</li> <li>- Overall, the model and cell data produced similar results.</li> <li>- Cell data do not provide the detail needed for adequate development of behaviorally based models.</li> <li>- The findings support further use of cell data for use in modeling, especially considering their low cost and high sample size.</li> </ul>
5	Using Mobile Phone Location Data to Develop External Trip Models  Western North Carolina  January 2015	<ul style="list-style-type: none"> <li>- Check data for erroneous trips between external districts that do not cross the study area boundary.</li> <li>- If needed, use a disaggregation process to factor data from the external districts to individual external stations.</li> <li>- Travel patterns modeled with the cell data validated well against observed traffic counts.</li> <li>- Cell data are a useful source for the development and estimation of external trip models.</li> <li>- The use of cell data can be a lower cost alternative to traditional methods of obtaining external data.</li> </ul>
6	Using Cellphone O-D Data for Regional Travel Model Validation  Syracuse, NY  May 2015	<ul style="list-style-type: none"> <li>- Strike a good balance in determining cell data zone size; if too small, it can lead to missing many E-E, E-I, and I-E trips, but if too large, zones can lead to an excess of E-E trips being captured.</li> <li>- Cell data are better than model results at differentiating HBO trips than with HBW or NHB trips, thought the time of year data are collected may impact trip purpose splits.</li> <li>- Cell data under-estimated trips for the University of Syracuse.</li> <li>- Use long links for select link analyses and take care when doing this type of analysis.</li> </ul>

	Cell O-D Study and Date of Report	Key Takeaways and Lessons Learned
7	Development of the Idaho STDM Trip Matrices Using Cell Phone OD Data State of Idaho May 2015	<ul style="list-style-type: none"> <li>- Cell O-D data are a reasonable starting point for generating statewide trip matrices.</li> <li>- Use of cell data in ODME produced reasonable results, but ODME should only be used for short-term forecasting.</li> <li>- A better understanding is needed on how cell data flows are different that traditional travel modeling data sets.</li> </ul>
8	Understanding Cellular-based Travel Data Maricopa County, AZ May 2015	<ul style="list-style-type: none"> <li>- Total daily trips matched fairly well between cell and model data, but there were differences in peak period trips for some areas.</li> <li>- Cell data did not appear to capture as many short distance trips as the model.</li> <li>- Cell data were found to be higher during peak periods than model data.</li> </ul>
9	Preliminary Evaluation of Cellular O-D Data as a Basis for Forecasting Non-Resident Travel Washington, D.C. May 2015	<ul style="list-style-type: none"> <li>- In comparing model results to cell results, the total number of trips was comparable but the cell results had higher share of HB trips and a lower share of NHB trips.</li> <li>- Aggregation of cell data was found to reduce noise (i.e., there were more differences in cell verses modeled data when data are disaggregated).</li> <li>- External travel from cell data exceeded counts by 30 percent.</li> </ul>

**Table 9. Takeaways from select Bluetooth and/or ALPR O-D studies.**

1	Generating Route-Specific O-D Tables Using Bluetooth Technology Jacksonville, FL 2012	<ul style="list-style-type: none"> <li>- The collective capture rate of the 14 Bluetooth sensors used in the study was 6.14 percent.</li> <li>- Bluetooth can be used for an O-D corridor analysis with multiple routes.</li> <li>- It was possible to use the data collected in after-model validation or toll revenue forecasts.</li> </ul>
2	Bryan-College Station Bluetooth External Survey August 2011	<ul style="list-style-type: none"> <li>- The 2011 Bluetooth results contained 60 percent more E-E trips than what was obtained in the 2002 traditional external survey, but part of the increase was due to an increase in traffic volumes between 2002 and 2011.</li> <li>- The fact that data were not collected at night in the 2002 survey may have also contributed to the discrepancy in external-through trips between the two surveys.</li> <li>- The detection rate of the 13 Bluetooth sensors in the study was about 8 percent.</li> <li>- Bluetooth is a viable means for estimating E-E trips as part of an external survey.</li> </ul>
3	SH 130/IH-35 Traffic Diversion Study in the Austin, TX region 2013	<ul style="list-style-type: none"> <li>- ALPR results of pass-through traffic were significantly lower than the Bluetooth results. ALPR captured about 70 percent of license plates.</li> <li>- In this study, ALPR was more susceptible to human error than Bluetooth.</li> <li>- The study included 35 Bluetooth locations, which had detection rates that ranged from 3–10 percent.</li> </ul>
4	The Harbor Bridge Alignment Study in Corpus Christi, TX Spring 2013	<ul style="list-style-type: none"> <li>- Bluetooth in combination with traffic count data was a cost effective and reliable means of assessing which proposed bridge alternative most directly linked with prevailing traffic patterns.</li> </ul>
5	Highway Bypass Study in Alice, TX March 2013	<ul style="list-style-type: none"> <li>- Bluetooth readers in combination with ALPR cameras and VCCs were a viable means to assess O-D patterns for the study.</li> <li>- The collective detection rate of the 20 Bluetooth sensors in the study was about 8.3 percent.</li> <li>- Bluetooth data collected during the same 24-hour time period as ALPR data were associated with 5,200 more O-D matches than the ALPR method.</li> <li>- Similar results were found for both technologies in comparing the through trips on primary highway bisecting the area.</li> </ul>

## Appendix B: Comparison of Characteristics by Technology

This section describes capabilities and limitations of each technology in relation to the following key characteristics:

- Positional accuracy of data unit.
- Sample saturation/penetration.
- Sampling frequency.
- Continuity of data streams.
- How trips and trip ends are estimated and defined.
- Measures and processes used to anonymize data.
- Types of suitable studies and geographies.

### Positional Accuracy of Data Unit

The positional accuracy GPS trip data are the most refined of the three technologies, with accuracy being achieved to within 1–10 meters (under ideal conditions). GPS accuracy is considered to be very good for O-D applications. The accuracy of trip ends of GPS data obtained from third-party sources will be reduced due to anonymization of the data. The accuracy of trip ends developed from cell data generally vary from about 150–500 meters. According to Airsage, an individual sighting in an urban area is accurate to about 300 meters; however, information is used from multiple sightings to refine a location.<sup>(11)</sup> It is difficult to ground-truth cellular data and accuracy can vary significantly by device type.<sup>(20)</sup> Airsage has the ability to offer cell O-D data with enhanced accuracy, but it may result in a decreased sample size.<sup>(56)</sup> Since Bluetooth data do not provide the location of trip ends, a comparison of its accuracy to that of GPS or cell data are not relevant. TTI Bluetooth readers have a range of detection of about 100 meters.

### Sample Saturation/Penetration

Cell data have very good sample penetration rates and are estimated to be in the range of 15–25 percent of the population, with generally higher samples in urban areas and lower samples in rural areas.<sup>(11)</sup> The sampling rate is dependent on the market penetration rates of the carrier(s) with whom the cell data provider has agreements. GPS data are currently estimated to be in the range of about 0.5–2 percent of vehicles on the roadway, which is much lower than the penetration rate reported for cellular data. However, this rate is continually increasing (albeit slowly), making GPS technology an increasingly better option for O-D data collection into the future. Bluetooth sampling rates vary by location, but are generally in the range of about 5–15 percent of the traffic that passes by the Bluetooth device.

### Sampling Frequency

Most GPS devices collect raw data in 1-second increments, but this level of temporal resolution is not provided from third-party data providers due to privacy concerns. GPS data obtained from these providers will be pre-processed and provided in larger time increments, ranging from seconds to minutes, depending on study requirements and what is purchased. Some data providers will not provide the actual GPS data for O-D, but instead provide an O-D matrix based on the TAZ geography provided by the purchaser. The sampling frequency of cell data is less than that of GPS data and highly variable depending on a number of factors. According to Airsage, sighting frequency can vary widely based on the quality of the device, with some being sighted 50–60 times per day and others 1,000 or more times per day.<sup>(13)</sup> Airsage also states that

cell device identification is defined dynamically (for them) every 10 minutes and devices average about 180 sightings per day.<sup>(56)</sup> Bluetooth readers collect data in 1-second increments when a device has been detected and remains within range of the Bluetooth reader.

### Continuity of Data Streams

It is difficult to assess and compare the continuity of data streams between cell and GPS trip data. The assessment could vary widely depending on a study's scope and geographic scale, length of time and travel distance, and use of raw versus processed data.

If data from individual cell and GPS devices were compared when data were being collected, the GPS data would be more continuous almost certainly for short time periods and distances, but possibly not for long time durations and distance. GPS data's greater sampling frequency and better location accuracy are more likely to produce continuous data streams than cell data, since its sampling frequency and location accuracy is poorer and varies widely. However, a significant portion of the continuous data streams in GPS could be for only portions of trips based on when a GPS device is in use for a navigation session.

The continuity of both cell and GPS data is improved through processing where long periods of data are analyzed to develop average trips for smaller time durations. For example, data from four weeks of weekday data are analyzed to develop average weekday trips or average trips for each weekday for each device. Such processing allows gaps in raw data streams to be filled or smoothed by evaluating a device's movement over time.

Data continuity is more of a factor for localized or operational type studies. GPS data's finer granularity both in time and space make it better suited for studies of this type where greater specificity is needed, such as to evaluate network routing or turning movements. Data continuity is less of a factor (in relative terms) for larger regional or statewide studies where less specificity in time and location is needed.

Since Bluetooth is point sensor data, it cannot be compared to cell or GPS data relative to data continuity because it cannot collect streams of data to estimate device movements or endpoints. However, within the 100 meter range of a Bluetooth monitor, the data are continuous.

### How Trips and Trip Ends Are Estimated and Defined

O-D data are developed by cell data aggregators using time and location data. Cell sightings during the day and evening are used to determine the work and home location of the phone, respectively, which helps in determining trip purpose. Sampled cellular data are then expanded using census tract data.<sup>(21)</sup> If a cellular device does not move over a period of five minutes, a trip destination is assumed to have occurred.<sup>(57)</sup> Similarly, the resident-nonresident status of data collected from a cell device is established by mapping where the device resides between 9 p.m. and 7 p.m.<sup>(58)</sup> For GPS data, trips are developed from GPS points by identifying the trip ends. Trip ends are established considering the dwell times of where a GPS device has remained at the same location in the data stream. The dwell time thresholds can vary and can potentially be changed at the request of the data purchaser. At a minimum, dwell time thresholds should be at least two minutes, but five minutes may be more typical. In processing GPS data, algorithms using spatial and temporal criteria are used to identify trip ends.<sup>(59)</sup> Unlike cell and GPS data, trip ends cannot be estimated with Bluetooth data.

### Measures and Processes used to Anonymize Data

Cellular O-D data provided cell data providers, such as Airsage, are anonymous.<sup>(60)</sup> Airsage uses WISE technology within the anonymization process. To assure additional privacy, Airsage changes the encrypted ID of devices every 28 days. Similarly, for GPS data, no individual data are provided, for privacy reasons.<sup>(59)</sup> GPS data are anonymized through time and/or space. The

first and last 5-or-10 minutes of movement is removed from the data in order to not reveal specific beginning and end points. Additionally, to ensure that long-term tracing of vehicles cannot occur, the unique ID of each device is scrambled periodically. An extensive protocol is followed with Bluetooth data to ensure that they are properly anonymized to protect the privacy of individuals. Included among these practices are removing digits from the MAC address field prior to sending this information for central processing, having no personal information attached to the MAC address, having some MAC addresses duplicated, encrypting data when they are sent to a central location for processing, and aggregating the data sample prior to use in O-D studies.<sup>(61)</sup> No MAC address is linked to a person or business; rather, a random alphanumeric string is assigned.

At the present time, which technology to use depends primarily on the level of spatial and temporal resolution needed to meet study objectives and desired level of accuracy. Each technology has strengths and weaknesses and no one technology can collect all data elements or characteristics of travel needed for a thorough O-D study. Which technology (or technologies) to use for an O-D study will depend on the type of study, its scope, objectives, and budget. Any of the technologies could be used in most O-D studies, but the question is not which technology will work, but which is most suitable and will provide the best results relative to the project's objectives.

## References

- <sup>1</sup> Airsage Documentation, How Accurate is the Data?, <http://www.airsage.com/Technology/Accuracy/>, accessed, December 2015.
- <sup>2</sup> Cellint website, [http://www.cellint.com/news/CELLULAR\\_BASED\\_TRAFFIC\\_MONITORING\\_SOLUTION.html](http://www.cellint.com/news/CELLULAR_BASED_TRAFFIC_MONITORING_SOLUTION.html), accessed 11/17/2015.
- <sup>3</sup> Mirzaei, Arash. "Technology Tests in Trip data Collection." Presentation at TRB's 93<sup>rd</sup> Annual Meeting, Washington, D.C., January 2014.
- <sup>4</sup> Calabrese, F. "Urban sensing using mobile phone network data." Ubicomp 2011 Tutorial, IBM Research, Dublin, Ireland, 2011.
- <sup>5</sup> Airsage Documentation. Understanding Population Movements, updated version provided by Bill King, November 18, 2015.
- <sup>6</sup> INRIX Cellular Analytics: Overview, Webinar with TTI Researchers, November 2013.
- <sup>7</sup> T. Wang, C. Chen, and Ma, Jingtao. "Mobile Phone Data as an Alternative Data source for Travel Behavior Studies." Paper presented at the TRB 2014 Annual Meeting, Washington, D.C.
- <sup>8</sup> Sivaraman, Vijay. E-mail from Airsage to TTI, Subject: Tyler Study Comments/Questions, October 9, 2014.
- <sup>9</sup> King, William H. "Data Solutions for Your Transportation Studies." Airsage Presentation at FHWA Cell Phone Data and Travel Behavior Research Symposium, Washington, D.C., February 2014.
- <sup>10</sup> Schuman, Rick. "INRIX Perspectives – Coverage, Resolution, Characteristics (and Issues)." Presentation at FHWA Cell Phone Data and Travel Behavior Research Symposium, Washington, D.C., February 2014.
- <sup>11</sup> Airsage Nationwide Commute Report, Data for April 2014. Frequently Asked Questions, January 8, 2015.
- <sup>12</sup> Smith, Cy. Airsage Meeting with NCTCOG, TxDOT, and TTI, NCTCOG Offices, Arlington, TX.
- <sup>13</sup> Airsage Presentation by M. Martino, TRB Transportation Planning Applications Conference.
- <sup>14</sup> Wang, T., C. Chen, and Ma, Jingtao. "Mobile Phone Data as an Alternative Data source for Travel Behavior Studies." Paper presented at the TRB 2014 Annual Meeting, Washington, D.C.
- <sup>15</sup> Airsage website, accessed September 17, 2015.
- <sup>16</sup> TTI Conference Call with Airsage. Discussion of Cell Data Purchase on Design of Tyler O-D Study, March 28, 2014.
- <sup>17</sup> Airsage Documentation. External Analysis, updated version provided by Bill King, November 18, 2015.
- <sup>18</sup> Huntsinger, L.F., and K. Ward. "Using Mobile Phone Location Data to Develop External Trip Models." *Journal of the Transportation Research Board*, No. 2499, Transportation Research Board, Washington, D.C., January 2015.
- <sup>19</sup> Sample Airsage Output Data file, FHWA Cell Research Symposium, 2/12/14.
- <sup>20</sup> Milone, R. "Preliminary Evaluation of Cellular Origin-Destination Data as a Basis for Forecasting Non-Resident Travel." PowerPoint presentation. 15th TRB National Transportation Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015 Applications Conference May 17-21, 2015.

- <sup>21</sup> Coladner D, B. Stabler, and S. Sikder. "Development of the Idaho STDM Trip Matrices Using Cell Phone OD Data." 15th TRB Transportation Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.
- <sup>22</sup> Fussell, Rhett, Craig Gresham, and Cy Smith. "Origin Destination Analysis for Moore County, NC." NCDOT and the Moore County Transportation Committee (MCTC), July 31, 2013.
- <sup>23</sup> U.S. National Library of Medicine. "How Accurate is the GPS on my Smart Phone? (Part 2)." Community Health Maps, <http://communityhealthmaps.nlm.nih.gov/2014/07/07/how-accurate-is-the-gps-on-my-smart-phone-part-2/>.
- <sup>24</sup> Chitturi, M.V., J W. Shaw, J.R.C. IV, and D.A. Noyce. "Validation of Origin-Destination Data from Bluetooth Reidentification and Aerial Observation." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2430, 2014, pp. 116-123.
- <sup>25</sup> Delaney, Ian. HERE 360 blog, <http://360.here.com/2016/01/13/here-unveils-trip-data-for-247-intelligence-on-road-usage/>, accessed January 13, 2016.
- <sup>26</sup> Schewel, L., and R. Schuman. "Data-Driven Transportation Demand Management (TDM)." Webinar presented on September 22, 2015.
- <sup>27</sup> Kim, D., D. Porter, S. Park, A. Saeedi, A. Mohsemi, N. Bathaee. And M. Nelson. "Bluetooth Data Collection System for Planning and Arterial Management." Federal Highway Administration report FHWA-OR-15-01, August 2014.
- <sup>28</sup> "I-75 & Palmetto Expressway Origin-Destination Study: Final Report." RSG Inc. March 2012.
- <sup>29</sup> *I-95 Pennsylvania Origin-Destination Study*. Traffax, College Park, MD. <http://www.traffaxinc.com/sites/default/files/I-95%20OD%20Study%20July%202011.pdf>, accessed November 19, 2015.
- <sup>30</sup> Zhang, W., J. Samuelson, and B. Kidd. "Understanding Travel Time and Origin-Destination Characteristics at Airports Using Bluetooth Technology." Paper prepared for the 2013 ITE Western District Conference.
- <sup>31</sup> Malinovskiy, Y., N. Saunier, and Y. Wang. "Analysis of Pedestrian Travel with Static Bluetooth Sensors." *Transportation Research Record: Journal of the Transportation Research Board*, No. 2299, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 137-149.
- <sup>32</sup> Reiff, R. M. "Determination of Origin-Destination Using Bluetooth Technology." Presented at the ITE Annual Meeting, Atlanta, GA., 2012.
- <sup>33</sup> Carpenter, C., M. Fowler, and T. J. Adler. "Generating Route-Specific Origin-Destination Tables Using Bluetooth Technology." In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2308, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 96-102.
- <sup>34</sup> Doyle, S., C. Covelli, A. Lanigan, and B. Hooton. "Gardiner Expressway East Planning Study Innovative Planning Techniques." Paper presented at the 2015 Conference of the Transportation Association of Canada. Charlottetown, PEI.
- <sup>35</sup> Guevarra, D., and J. Voss. "Using Bluetooth Technology to Conduct Origin-Destination Surveys." <http://www.cite7.org/Regina2015/documents/Session4B-02.pdf>, accessed November 19, 2015.
- <sup>36</sup> Steel, P., and P. Kilburn. "Using Bluetooth Technology to Monitor Traffic Patterns around Urban Centers in Alberta." Presented at 18th ITS World Congress, Orlando, Fla., 2011.
- <sup>37</sup> Michau, G., A. Nantes, et al. "Retrieving Dynamic Origin-Destination Matrices from Bluetooth Data." In *Transportation Research Board 93<sup>rd</sup> Annual Meeting*, January 12-16, 2014, Washington, D.C.

- <sup>38</sup> Yucel, S., H. Tuydes-Yaman, O. Altintasi, and M. Ozen. "Determination of Vehicular Travel Patterns in an Urban Location Using Bluetooth Technology." Presented at ITS America Annual Meeting and Expo, Nashville, TN, 2013.
- <sup>39</sup> Kay, A., and P. Jackson. "An Appraisal of Emerging Bluetooth Traffic Survey Technology." <http://www.starconference.org.uk/star/2012/KayJackson.pdf>, accessed November 18, 2015.
- <sup>40</sup> Carrol, M., R. Dowling, and S. Ashiabor. "Current Practices in the Use of Emerging Traffic Data Sources for Transportation Planning: Technologies, Challenges and Lessons." Paper submitted to the Transportation Research Board 2012 Meeting.
- <sup>41</sup> Cambridge Systematics. "Chattanooga Cell Phone External O-D Matrix Development-Process and Findings." Presented to Tennessee Model Users Group (TNMUG), November 10, 2011.
- <sup>42</sup> Harvey, J. SARPC Trip Distribution Model Calibration, Technical Memorandum from Alliance Transportation Group to SARPC, December 28, 2012.
- <sup>43</sup> Harrison, K., "Origin-Destination Study for Mobile County – SARPC, December 31, 2012." Available at [www.sarpc.org](http://www.sarpc.org), accessed January 5, 2016.
- <sup>44</sup> Fussell R, C. Gresham, and C. Smith. "Origin Destination Analysis for Moore County, NC." Parsons Brinckerhoff, Clearbox, Airsage, July 31, 2013.
- <sup>45</sup> Huntsinger, L., and R. Donnelly. "Reconciliation of Regional Travel Model and Passive Device Tracking." TRB Annual Meeting, Washington, D.C., January 2014.
- <sup>46</sup> Bindra S. "Using Cellphone O-D Data for Regional Travel Model Validation." 15th TRB Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.
- <sup>47</sup> Coladner D, B. Stabler, and S. Sikder. "Development of the Idaho STDM Trip Matrices Using Cell Phone OD Data." 15th TRB Transportation Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.
- <sup>48</sup> Zhang, W., A. Kuppam, V. Livshits, and B. King. "Understanding Cellular-based Travel Data Experience from Phoenix Metropolitan Region." Presentation at the 15<sup>th</sup> National Transportation Planning Application Conference, Atlantic City, NJ, May 2015.
- <sup>49</sup> Milone R. "Preliminary Evaluation of Cellular Origin-Destination Data as a Basis for Forecasting Non-Resident Travel." PowerPoint presentation at the 15th TRB National Transportation Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.
- <sup>50</sup> Morris, Paul. Information on the I494/TH 62 project provided via e-mail to MnDOT's Paul Czech, February 20, 2016.
- <sup>51</sup> Morris, Paul. INRIX Q&A Session, Presentation to MnDOT Metropolitan Council By SRF, January 20 2016.
- <sup>52</sup> Eisele, W.L. *Freight Fluidity Framework, Calculation Procedures and Analysis Findings*. Technical Memorandum for the Maryland State Highway Administration from Texas A&M Transportation Institute through University of Maryland, June 1, 2015.
- <sup>53</sup> Chigoy B., S. Farnsworth, and E. Hard. "Case Studies Using Bluetooth O-D Surveys to Address Planning and Policy Issues." TRB Tools of the Trade Conference, Burlington, VT July 2014.
- <sup>54</sup> "Napa County Travel Behavior Study." NCTPA Board Meeting Presentation, December 17, 2014.
- <sup>55</sup> Fehr & Peers. "Napa County Travel Behavior Study-Draft Survey Results and Data Analysis Report." December 8, 2014.
- <sup>56</sup> Smith, Cy. Meeting with North Central Texas Council of Governments and TTI, Arlington, TX, July 15, 2015.

<sup>57</sup> Bindra S. “Using Cellphone O-D Data for Regional Travel Model Validation.” 15th TRB Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.

<sup>58</sup> Fussell, R., and C. Gresham. “No Horsing Around, A Hole in One with Mobile Phone Data.” National Transportation Planning Applications Conference, May 19, 2015, Atlantic City, NJ, May 19, 2015.

<sup>59</sup> Cohn, N. “TomTom Data for Origin-Destination.” FHWA Cell Phone Data and Travel Behavior Research Symposium, Washington, D.C., February 12, 2014.

<sup>60</sup> Smith, C., and B. King. “Location Intelligence from Cellular Signaling Data Powering Transportation Planning.”

<sup>61</sup> Borchardt, D.W. Texas A&M Transportation Institute. Traffic Data Collection – PHR District – June 22-26, 2015.” Memorandum to Texas Department of Transportation – PHR District, June 18, 2015.

## NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United State Government assumes no liability for its contents or use thereof.

The United States Government does not endorse manufacturers or products. Trade names appear in the document only because they are essential to the content of the report.

The opinions expressed in this report belong to the authors and do not constitute an endorsement or recommendation by FHWA.

This report is being distributed through the Travel Model Improvement Program (TMIP).

U.S. Department of Transportation  
Federal Highway Administration  
Office of Planning, Environment, and Realty  
1200 New Jersey Avenue, SE  
Washington, DC 20590

May 2016

FHWA-HEP-16-056



U.S. Department of Transportation  
**Federal Highway Administration**