SPR-744 APRIL 2019

Optimizing Technology for Collecting Arizona Long-Distance Travel Data

Arizona Department of Transportation Research Center

ADOT

Optimizing Technology for Collecting Arizona Long-Distance Travel Data

SPR-744
April 2019

Prepared by:
Edwin N. Hard, Byron T. Chigoy, Lisa K. Green, Ph.D, Rafael M. Aldrete, Ph.D, and Stephen P. Farnsworth
Texas A&M Transportation Institute
3135 TAMU
College Station, TX 77843-3135

Published by:
Arizona Department of Transportation
206 South 17th Avenue
Phoenix, Arizona 85007
In cooperation with
U.S. Department of Transportation
Federal Highway Administration

This report is subject to the provisions of 23 USC § 409. Any intentional or inadvertent release of this material, or any data derived from its use, does not constitute a waiver of privilege pursuant to 23 USC § 409, which reads as follows:

23 USC § 409 — Discovery and admission as evidence of certain reports and surveys

Notwithstanding any other provision of law, reports, surveys, schedules, lists, or data compiled or collected for the purpose of identifying, evaluating, or planning the safety enhancement of potential accident sites, hazardous roadway conditions, or railway-highway crossings, pursuant to sections 130, 144, and 148 of this title or for the purpose of developing any highway safety construction improvement project which may be implemented utilizing Federal-aid highway funds shall not be subject to discovery or admitted into evidence in a Federal or State court proceeding or considered for other purposes in any action for damages arising from any occurrence at a location mentioned or addressed in such reports, surveys, schedules, lists, or data.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-AZ-19-744		
4. Title and Subtitle Optimizing Technology for Collecting Arizona Long-Distance Travel Data		5. Report Date April 2019
		6. Performing Organization Code
7. Author Edwin N. Hard, Byron T. Chigoy, Lisa K. Green, Ph.D, Rafael M. Aldrete, Ph.D, and Stephen P. Farnsworth		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas A&M Transportation Institute 3135 TAMU College Station, TX 77843-3135		10. Work Unit No.
		11. Contract or Grant No. SPR-PL-1(189)744
12. Sponsoring Agency Name and Address Arizona Department Of Transportation 206 S. 17th Avenue Phoenix, AZ 85007		13.Type of Report & Period Covered FINAL
		14. Sponsoring Agency Code
15. Supplementary Notes Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration		

16. Abstract

This report provides an implementation plan and recommendations for developing long-distance travel data, defined as trips 50 miles or longer, for the Arizona Department of Transportation (ADOT) statewide travel demand model. The report evaluates origin-destination (O-D) data from cellular, global positioning system (GPS), Location Based Services (LBS), and Bluetooth technologies and compares their strengths and weaknesses considering sample penetration, accuracy, and the ability to estimate long-distance travel for different trip types and categories.

The report recommends that ADOT use LBS data as the primary source for passenger travel due to the relatively high penetration rate, better location accuracy than cellular data, and ability to estimate flows at a large geographic scale. It also recommends that ADOT use GPS O-D data for truck travel, since commercial truck fleets have a significant sample penetration in GPS data. The report concludes that ADOT would benefit from acquiring trip matrices for zone-to-zone truck flows or trip records with waypoints for detailed study of truck flows on the highway network. The report provides an implementation plan for integrating “new technology” O-D data into ADOT’s statewide model. The plan outlines preparatory activities, tasks, and decisions to be addressed prior to acquiring the data.

17. Key Words		18. Distribution Statement		23. Registrant's Seal
origin-destination, O-D, cellular, GPS, Bluetooth, Wi-Fi, Location Based Services, LBS, passive, long-distance, statewide, traffic analysis zones, TAZ, travel demand, travel demand model, data collection, forecasting, inductive loop signature		Document is available to the US public through the National Technical Information Service, Springfield, Virginia 22161		
19. Security Classification Unclassified	20. Security Classification Unclassified	21. No. of Pages 73	22. Price	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ADOT	Arizona Department of Transportation
ATRI	American Transportation Research Institute
ATR	automatic traffic recorder
ATS	American Travel Survey
AZTDM	Arizona Travel Demand Model
CSV	comma separated variable
CTPP	Census Transportation Planning Package
DOT	department of transportation
E-E	external-to-external (trip)
E-I	external-to-internal (trip)
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
GPS	global positioning system
ILST	inductive loop signature technology
I-E	internal-to-external (trip)
I-I	internal-to-internal (trip)
IPF	iterative proportional fitting
ISTDM	Indiana Statewide Travel Demand Model
ITD	Idaho Transportation Department
iTRAM	Iowa Statewide Traffic Analysis Model
LBS	location based services
MAC	media access control
MAG	Maricopa Association of Governments
MPO	metropolitan planning organization
NCHRP	National Cooperative Highway Research Program
NELDT	nationwide estimate of long-distance travel
NHTS	National Household Travel Survey
O-D	origin-destination
ODME	origin-destination matrix estimation
RAZ	Regional Analysis Zone
TAC	technical advisory committee
TAZ	traffic analysis zone
TMIP	Travel Model Improvement Program
TRB	Transportation Research Board
TSTM	Tennessee Statewide Travel Model
TTI	Texas A&M Transportation Institute
VMT	vehicle miles traveled
WIM	weigh-in-motion

EXECUTIVE SUMMARY

This report provides an implementation plan and recommendations for capturing long-distance, origin-destination (O-D) travel data to support the Arizona Department of Transportation (ADOT) statewide Arizona Travel Demand Model (AZTDM). The AZTDM supports analysis of highway corridors throughout the state and serves as a tool to assist in evaluating the impacts of new or proposed transportation projects, such as new routes or improvements in capacity. It models average weekday travel by passenger and truck/freight categories.

Long-Distance Travel and the AZTDM

Long-distance travel, defined as trips greater than 50 miles, is important in the AZTDM as its chief focus is longer intrastate trips between urban areas, regions, and major attractions. A representative sample of this travel serves as critical input to the model for estimating existing flows and forecasting future traffic levels. Currently, the AZTDM uses data from the 2002 National Household Travel Survey (NHTS) and a national commodity flow model as sources for long-distance travel. However, these traditional sources do not sufficiently sample travel in Arizona or target long-distance travel. In addition, O-D data from the NHTS does not represent current travel, since the survey is conducted approximately every seven years and the data are not available until a year or more after survey completion. O-D data from new technologies can better represent current travel since they can be obtained specifically for Arizona and for the desired study period.

The AZTDM models travel that occurs within the state, comes into or out of the state, or passes through the state. The AZTDM divides trips into passenger and truck/freight.

Assessing New Technology for Long-Distance Travel Data

In recent years, non-survey travel data from new technologies, including cellular, Global Positioning System (GPS), Location Based Services (LBS), Bluetooth, and Wi-Fi, have evolved as essential sources for O-D data. Data from these sources represent observed travel behavior and are considered “passive,” since they are obtained without direct interaction with an individual. How “new technology” O-D data are obtained are summarized as:

- **Cellular O-D data:** Derived from the interaction of mobile devices within a cellular network, are a measure of estimated movements of a cell device over a period of weeks or months.
- **GPS O-D data:** Obtained via satellite trilateration, provide a device’s location and tracking its movements using time-stamped coordinates.
- **LBS O-D data:** Position information derived from mobile device software applications (apps) that collect or request the location of the device user. The data include random device location information collected via cellular networks and/or Wi-Fi and Bluetooth networks such as in public spaces and retail locations. The technology used to collected LBS data varies depending on the app used to acquire the device’s location.

- **Bluetooth and Wi-Fi data:** Collected by roadside sensors that detect Bluetooth and Wi-Fi enabled devices in vehicles when they pass a sensor. Bluetooth and Wi-Fi are wireless technologies used for exchanging data over short distances.

The researchers evaluated the technologies individually and collectively as potential sources of long-distance data for the AZTDM, considering each type and category of travel included in the model. The assessments considered each technology's attributes and properties such as accuracy, sample penetration, and ability to estimate long-distance trips.

Recommended Sources of "New Technology" Data

Considering the current capabilities of each technology, this study's researchers recommend that LBS and GPS data serve as sources for developing long-distance O-D data for the AZTDM. ADOT would benefit from using LBS data as the primary source to develop passenger O-D travel and trip tables because of their high penetration rate and ability to estimate flows and activities at a large geographic scale. LBS data will produce a better population and distribution of trip ends for the AZTDM than any other technology source.

On the other hand, it is recommended that ADOT use GPS O-D data for truck travel, since commercial fleets/trucks have a significant sample penetration in GPS data and since LBS (and cellular) data focus more on non-commercial travel. GPS data can be acquired as trip records, trip matrices containing zone-to-zone movements, and trip records with waypoints. If ADOT needs only zone-to-zone truck flows, it would benefit from acquiring GPS trip matrices. However, if ADOT wishes to examine truck flows on the highway network or an individual corridor, this can be done by acquiring trip records with waypoints.

The researchers recommend that ADOT consider Bluetooth on a limited, strategic basis for benchmarking O-D estimates obtained from LBS and GPS sources. The implementation of these recommendations will require the use of ADOT vehicle classification counts, supplemented by count data from metropolitan planning organizations (MPOs) and local jurisdictions to expand GPS data and adjust cellular data.

Implementation Plan for Obtaining Long-Distance Data for the AZTDM

Key steps for implementing data collection procedures from new technologies, listed in chronological order, include:

- Establish an active advisory committee to develop guidance on supplementing new technology sources to the existing long-distance elements of the AZTDM
- Obtain an update of O-D data sources and prices to determine whether any changes to the recommendations of this study are appropriate.
- Aggregate the AZTDM's internal traffic analysis zones (TAZs) to Census tracts or groups thereof, and reassess external zones and Regional Analysis Zones (RAZs).
- Aggregate MPO TAZs into larger zones considering the limitations of LBS data and work directly with the selected data vendor to consider gaps in data.

- Acquire LBS data with a long-distance trip filter, matrices covering all types of long-distance trips, average weekday travel, and the option to identify travel inclusive of an overnight stay.
- Acquire GPS trip matrices if ADOT needs only zone-to-zone flows.
- Purchase GPS trip records with waypoints should ADOT wishes to use the data for additional purposes, such as routing studies or select link analyses.
- Plan for data storage and analytical software for processing large data sets.

CHAPTER 1. INTRODUCTION

STUDY BACKGROUND AND NEED

Origin-destination (O-D) data and information are critical in many transportation planning and modelling 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 and license matching, have with a few exceptions, been replaced with passively collected "new technology" O-D data.

The study has two primary objectives:

- Evaluate new technology O-D sources and methods for capturing long-distance O-D travel data to support Arizona Department of Transportation's (ADOT's) statewide Arizona Travel Demand Model (AZTDM).
- Utilize the findings from the evaluations of new technologies and methods to develop an implementation plan that will assist ADOT in the use and integration of new technology O-D data in its statewide model.

The current AZTDM model (AZTDM2) supports analysis of highway corridors throughout the state, providing traffic flows by automobile and truck, as well as flow forecasts. The model can be used to evaluate the existing transportation network using various demographic and network scenarios. These scenarios might include population shifts and growth, changes to highway capacity, or highways at new locations. Scenarios can be modeled independently or in combinations. Future versions of the AZTDM (AZTDM3 and AZTDM4) will incorporate the ability to model transit and land use scenarios (ADOT 2011b).

In this context, long-distance O-D travel data consists of:

- Intrastate person and truck trips, which generally equal or exceed 50 miles.
- Interstate person and truck trips, which travel either to, from, or through Arizona.

The current AZTDM models these trips as long-distance person travel and long-distance single-unit and multi-unit trucks. The data and analytical support of the long-distance models is derived from the 2002 National Household Travel Survey (NHTS) and the Freight Analysis Framework 3 (FAF3). In addition to these sources, ADOT's traffic count and monitoring infrastructure are integral to developing the long-distance models, as traffic volumes on rural highway sections are used to calibrate these models (ADOT 2011a).

Historically, data on long-distance travel have been difficult to collect because most travel surveys are designed to collect data for a metropolitan region or urban area. The 2009 NHTS is slightly less focused on urban areas and includes some information on travel in rural regions, but has very little long-distance travel data (Schiffer 2012). In recent years, however, non-survey travel data derived from new or emerging technologies evolved as viable sources for O-D and other types of data needed for planning and operational studies. These data are generally referred to as “passive,” as they are collected without the explicit participation of the traveler. Passive data technologies include:

- **Cellular:** This technology is based on patterns of personal cell phone use. The technology uses various components of the cellular network to estimate the general geographic area where the device is used.
- **Global Positioning System (GPS):** This technology tracks movements very precisely using satellites. Although used for several decades, its use in devices owned by the general population and industries has recently exploded. These devices include onboard vehicle navigation technology, cell phone applications, and truck fleet tracking. In recent years, various companies started to aggregate data from these devices for use in evaluating travel behavior such as O-D flows.
- **Location Based Services (LBS):** LBS data are position information derived from software applications (apps) that log device location information. Such apps could include those for weather, navigation, retailers, shopping, and social connections. The data also include device location information collected via cellular networks and/or Wi-Fi or Bluetooth networks in retail locations or public spaces. The geographic precision and frequency of LBS data depend on the type of app and the technology used by the application to determine the location of the device.
- **Bluetooth and Wi-Fi:** These wireless technologies are used for exchanging data over short distances and are frequently embedded in mobile devices, computers, and in-vehicle applications such as navigation systems. Both are considered point sensor data as they detect signals omitted from Bluetooth or Wi-Fi enabled devices when the device is in range of the detector.

As part of this study, each of the above technologies is assessed individually and collectively as a potential source, or aid, in determining methods and approaches for developing long-distance O-D data for the AZTDM. The assessments consider each technology’s attributes and properties such as accuracy, sample penetration, and ability to estimate long-distance trips.

Cellular data are used for external surveys to capture travel flows into, out of, or through a region; long-distance corridor studies; travel between geographic areas; travel to and from major events and special generators (such as national and state parks); and assessment of travel *within* a region. The sample size associated with cellular data can be very large. Cellular data cannot discern the travel mode (driver, passenger, pedestrian), but provide estimates of trip purpose and whether travelers are residents or visitors to a study area (Hard et al. 2016a).

LBS data share many of the same characteristics and attributes as cellular data. Their location accuracy is better than that of cellular data, but not as good as GPS. LBS data can be used in the same types of travel studies as cellular data, such as external, corridor, and population flow studies. Like cellular data, LBS data cannot discern the travel mode, but can provide estimates of basic trip purposes and if travelers are residents or visitors to a study area. LBS served as an alternative source to cellular data in its first several years of use, but has now effectively replaced cellular data in transportation studies due to its better location accuracy. In September 2017, the only provider of cellular-based O-D data in the

United States, AirSage, changed its passive travel data analysis platform from cellular to LBS data (AirSage 2017). LBS data are heavily used in the mobile advertising industry, since the “always-there, always-on nature of devices combined with the ability to know device user locations” gives advertisers the ability to deliver relevant ads to consumers based on their real time location (Mobile Marketing Association 2015).

GPS data are generally associated with more accurate device tracking than cellular and LBS data. Depending on the provider, GPS information can be obtained with raw trace data, which give detailed movement of the device over a day, or using trip data with waypoints. The waypoint data consist of trip records along with intermittent geographic points along the travel route. GPS sampling rates (in number of devices sampled) are generally much lower than cellular sampling rates, but higher than Bluetooth sampling rates regionally. Current sources of GPS data are known to have a bias toward commercial vehicles, as much of the data are provided from the tracking of commercial fleets. On the other hand, this bias makes GPS valuable for applications dependent on truck O-D data (Hard et al. 2016a).

Bluetooth, and to a lesser extent Wi-Fi, are often used for capturing and re-identifying transportation data along highways and thoroughfares using permanent or temporary roadside sensor equipment. These technologies are ideal for corridor analyses of travel time and speed, and development of sensor-to-sensor O-D matrices. Bluetooth and Wi-Fi data collection is limited to sensor locations and does not allow for the collection of widespread, large-scale O-D data to support statewide travel demand models. However, Bluetooth and Wi-Fi can be used as tools to help ground-truth or benchmark travel data developed from cellular and GPS sources (Hard et al. 2016a).

These advances in O-D technology have not reduced the need for ground-truthing data collection, which uses more traditional traffic counting technologies. Reliable traffic count data are often used in the weighting and expansion of both GPS and Bluetooth/Wi-Fi data. Count data also can be used to assess the accuracy of cellular data using traffic assignment and region-to-region traffic flow estimates.

This research does not consider any of these technologies to be a “one size fits all” solution to effectively capture O-D long-distance data. Investments in new technology data can be used to support a range of transportation planning activities and may inform regional travel demand models and corridor studies, whether developed by ADOT or other agencies.

Researchers for this report also assessed Inductive Loop Signature Technology (ILST) for estimating O-D data. They found that ILST is not viable for estimating long-distance travel; thus it is not included as a technology assessed herein. A technical memorandum summarizing the assessment of ILST is included in the Appendix.

CHAPTER 2. LITERATURE REVIEW

OVERVIEW

This chapter details the findings of a literature review performed in support of this research. It describes the purpose of the literature review and reports its results.

Purpose of Literature Review

The purpose of the literature review is to provide background information on topics pertinent to the goal of the research. The results provide background information on statewide travel demand models, O-D data collection technologies, and the specific need—both broadly and specifically for ADOT—to better understand long-distance O-D travel. The researchers drew upon their understanding of the study objectives and experience in O-D data collection to select topics for inclusion in the literature review.

Organization of Literature Review

The body of the literature review is organized by topic, as outlined below.

- Statewide Travel Demand Models
  - Nationwide Overview
  - History of Arizona Statewide Travel Demand Model
  - Anticipated Data Needs for Future Arizona Statewide Travel Demand Models
- Long-Distance O-D Data Collection
  - Motivation for Better Long-Distance O-D Data Collection
  - Long-Distance Travel Surveys
  - External Travel Data
  - Long-Distance Freight Data
- AZTDM Long-Distance Trip Models and Data Sources
- Case Studies Using O-D Data Collection Technologies – Statewide or Long-Distance Travel
- Building Models from Passive Data
- Summary

STATEWIDE TRAVEL DEMAND MODELS

There is a wide spectrum of methodologies and approaches to statewide travel demand modeling across the country. This section begins with an overview of statewide modeling from a national perspective. A discussion of Arizona statewide travel demand modeling follows, including its history and anticipated future statewide modeling needs.

Nationwide Overview

Statewide travel demand models attempt to quantify average daily personal and commercial travel throughout all regions of a state. Depending on the analysis needs, some models may quantify only motor vehicle travel, while others quantify travel by other modes such as transit and commercial rail shipments. Models of the former type are defined as three-step models, while the latter are four-step. Statewide travel demand models are not as prevalent or universal as more localized, regional travel demand models, and there is a wide range of practices. As of January 2015, 20 states had an operational statewide model, six were developing such a model, two had an inactive statewide model, 12 were revising their model, two had a partial model, and eight did not have a model (Travel Forecasting Resource 2016). With the majority of states having a statewide model in some stage of development, there is much to be learned from the experiences of others. In this literature review, emphasis is placed specifically on what Arizona can learn from the findings and best practices of studies using passive data across the country, especially for development of long-distance O-D data.

Several studies were conducted to address issues related to statewide travel demand forecasting. *NCHRP Synthesis 358-Statewide Travel Forecasting Models* (Horowitz 2006) documents the findings from survey responses obtained from 49 of the 50 states about various aspects of their statewide travel demand models and/or future statewide modeling plans. The goal of the synthesis was to provide information to states wanting either to initially develop a statewide model or improve their existing model. The report highlights the need to define uses of the model and advises that goals associated with a statewide model should be developed prior to performing further modeling work (model development or model improvements) (Horowitz 2006). Long-distance trip making is listed among the top issues hindering progress in statewide model development (Horowitz 2006). Efforts to save money on modeling have led some states to use synthetic O-D data for statewide modeling, despite their lesser sensitivity for policy analysis than data developed from actual studies (Horowitz 2008). Review of recent studies, documentation of the AZTDM, and commissioning of this research study indicate that ADOT is acutely aware of the need to reasonably model and estimate long-distance travel (ADOT 2011a).

While some degree of learning and/or transferability may be possible between states in statewide travel demand modeling, there also are aspects that make Arizona unique and should be considered in how to most effectively evaluate technologies for O-D data collection. The next section focuses specifically on Arizona and its history, plans, and needs associated with statewide planning.

History of Arizona Statewide Travel Demand Model

The first version of an AZTDM was created in 2009 and relied on borrowed trip rates from Indiana and the Quick Response Freight Manual. The second version (AZTDM2) incorporated improved zonal details and used 2009 NHTS data for calibration (Transportation Model Improvement Program [TMIP] 2014). Plans are in place for a third and a fourth version of the AZTDM, with each version anticipated to improve and provide for more detailed modeling. The third-generation model is being developed to incorporate transit mode choice (ADOT 2016a).

Anticipated Data Needs for Future Arizona Statewide Travel Demand Models

Erhardt (2012) lists external travel, commodity-based truck flows, and long-distance travel as areas that should be priorities in future Arizona statewide modeling efforts. He states that if the vehicle miles traveled (VMT) associated with these three priorities are captured, approximately two-thirds of the VMT on Arizona state highways would be outside the largest population center (metropolitan Phoenix) in

Maricopa County (Erhardt 2012). This further supports the need for Arizona to effectively capture long-distance travel for use in statewide modeling.

LONG-DISTANCE O-D DATA COLLECTION

This section contains information about what constitutes long-distance travel and why such travel is important for statewide travel demand modeling—especially in a state like Arizona. Generally long-distance travel is not captured in local or regional travel surveys, so special care needs to be taken to collect these data for effective statewide planning (Hard et al. 2016a). *NCHRP Synthesis 735 – Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models* (Schiffer 2012) distinguishes long-distance travel as two groups, long-distance (trips coming from other states or regions within the state) and interurban or rural. These groups further divide into travel by passenger and freight vehicles. Long-distance travel in Arizona includes commercial and non-commercial travel by automobile, truck, rail, and air.

Motivation for Better Long-Distance O-D Data Collection

NCHRP Synthesis 358-Statewide Travel Forecasting Models states that, “Beyond computational and data problems associated with the detail of networks in statewide models, the greatest obstacle for forecasting is good representation of long-distance trip making” (Horowitz 2006). One reason for this obstacle is that capturing long-distance travel data may be unaffordable. However, the need to understand long-distance O-D information for a state such as Arizona is critical, given the large volume of traffic that moves through Arizona between California and the rest of the country (ADOT 2011a).

Both long-distance person travel and long-distance freight travel are considered in the AZTDM2. Long-distance person travel is defined as trips longer than 50 miles made for personal reasons, excluding commute trips. Long-distance freight travel is truck trips made between metropolitan regions. Three of four residents traveling long-distance remain in Arizona, and many in-state trips are predominantly recreational (ADOT 2011a). Much of the recreational travel is highly seasonal.

The literature stresses the need to dedicate resources to obtain good data for better understanding long-distance travel. In *Development of the Arizona Statewide Travel Demand Model: Phase 2 (AZTDM2)*, ADOT advises that “A primary recommendation for the future development of AZTDM is to collect adequate data to truly understand the travel markets on the facilities and areas of interest” (ADOT 2011a). Improved data collection efforts may want to focus on filling in the gaps associated with limited travel survey data available for rural areas. *NCHRP Synthesis 358-Statewide Travel Forecasting Models* observes that “Data collection can be a large component of the development of a statewide model. For example, Ohio has devoted almost 70 percent of its travel modeling budget to data acquisition” (Horowitz 2006). For Arizona, the use of secondary data sources may take the form both of more efficiently using data that are already collected across the state (such as data collected at permanent count sites) and employing new technologies for gathering O-D data.

Current ADOT Data Collection and Technology Use

The methodologies used in O-D data collection are continually evolving, with point sensors and crowd sourcing technologies used at a growing rate. Traditional methods, such as travel surveys and traffic counts, will continue to be important in benchmarking the quality of new technology O-D data and for use in expanding these data. ADOT's current data collection efforts, in purchasing add-on surveys for the 2017 NHTS (Federal Highway Administration [FHWA] 2016b) and making improvements to its traffic data collection infrastructure, will be useful.

Several technologies are currently used to collect O-D data in a passive (unobtrusive) manner. ADOT has in-field sensors to capture travel information across the state. Improvements to in-field data collection are continually developed. One example is a vehicle identification system (Jeng 2007) that can produce loop signatures for 60 types of vehicles, in conjunction with machine learning techniques, and has a low estimated cost of implementation (Mark Catchpole, ADOT employee, personal communications, July 7, 2016). While this technology has several desirable features, it is not as good at capturing long-distance travel as some others (Praprut Songchitruksa, Texas A&M Transportation Institute (TTI), personal communications, July 7, 2016). Cellular, GPS, LBS, and Bluetooth are among the emerging technologies used in O-D data collection and capture, with several studies reporting efforts to vet technology applications in different settings.

Long-Distance Travel Surveys

The FHWA conducts the NHTS approximately every seven years; the 2017 NHTS was recently completed (FHWA 2016c). Additionally, states and regions conduct travel surveys. While there are many differences between the NHTS and regional surveys, both typically capture interurban travel data (Schiffer 2012). The largest survey to focus on long-distance data was the 1995 American Travel Survey (ATS), which focused on trips longer than 100 miles. One major concern about this data set, besides its age, is that it did not capture mid-range trips longer than 50 miles. Trips of 50 to 100 miles are significant for interurban and rural trips. The 2001 NHTS surveyed long-distance trips as part of the daily person trip survey. As a good source of long-distance data, it became the basis of the AZTDM's current long-distance estimates (ADOT 2011a). The 2001 survey has its limitations, however. Among the most significant is that it captured only 45,000 long-distance trips instead of the 550,000 trips collected in the 1995 ATS (Schiffer 2012).

External Travel Data

Travel surveys typically inform travel model components related to trip generation that is internal to a region; internal travel consists of travel where both ends of the trip occur within the region. This travel is termed Internal-to-Internal (I-I). In the AZTDM, many of these trips are modeled in the urban areas in a short-distance person travel model (ADOT 2011a). For travel where at least one trip end occurs outside the region or state, the model also relies on O-D data that include travel crossing the region's study area boundary (Hard et al. 2016a). This type of data typically reflects nonstop travel through a region, termed External-to-External (E-E); travel that enters the region externally but makes a stop in the region, termed

External-to-Internal (E-I); and travel that begins in the region and then exits it, termed Internal-to-External (I-E).

Long-Distance Freight Data

The primary source of long-distance freight data is the Commodity Flow Survey conducted by the Bureau of Transportation Statistics (Kriger et al. 2011). The Commodity Flow Survey is used in the development of two secondary sources of freight data, the FAF (FHWA-Freight Management and Operations 2016a) and IHS Global Insight's Transearch (Transportation Research Board 2008) FAF (current version FAF4). The FAF4 models freight movement by truck, rail, air, and water modes between Bureau of Economic Analysis regions in the United States and major regions of the world (FHWA 2016a). FAF4 has several limitations, including the coarseness of its zones (with only three zones in Arizona) and the lack of estimates of freight traffic that passes through a region. Transearch is a county-to-county model of freight for each state in the United States. It has the benefit of providing estimates of freight traveling through a county, but lacks detailed trade data and requires significant quality checks (Beagan et al. 2007). In addition to these sources, truck freight travel data and E-E and E-I data are collected through traditional means, including commercial vehicle surveys and external surveys.

AZTDM LONG-DISTANCE TRIP MODEL AND DATA SOURCES

The current AZTDM accounts for long-distance trip making using the Nationwide Estimate of Long-Distance Travel (NELDT) (ADOT 2011a). NELDT is a four-step model of long-distance air taxi, commercial air, and automobile trips (Moeckel and Donnelly 2011) based on the 1995 ATS. This model is adapted in the AZTDM using the 2001 NHTS. The AZTDM NELDT model (ADOT 2011a) considers the following parameters:

- Total population
- Rural/small town population
- University enrollment
- Total employment
- Leisure employment
- Seasonal dwelling units

The AZTDM long-distance model also includes 48 special generators, of which 46 are state or national parks (Table 1). The AZTDM models long-distance freight movements by disaggregating FAF3 into regions and zones associated with Arizona's three FAF zones (Phoenix, Tucson, and the remainder of Arizona) by 42 commodity groups.

The AZTDM accounts for most statewide E-E and E-I/I-E travel based on its long-distance model. However, the long-distance model was found to underestimate traffic counts at some interstate border roads with development within 50 miles of the state line. The AZTDM report states that this underestimation is likely due to short-distance E-I/I-E and E-E travel that the long-distance model does not anticipate. The model addresses this with an exogenous (directly input value), short-distance, E-I trip purpose variable (ADOT 2011a). In this case, passive data should be considered as a source for calibration.

Table 1. National and State Parks Identified as Special Generators in AZTDM2 Travel Model.

National Park Service Units	State Park Units
Canyon de Chelly	Alamo Lake
Casa Grande Ruins	Boyce Thompson Arboretum
Chiricahua	Buckskin Mountain
Coronado	Catalina
Fort Bowie	Cattail Cove
Grand Canyon	Dead Horse Ranch, Jerome, and Verde River Greenway
Hubbell Trading Post	Fool Hollow Lake
Lake Mead	Fort Verde
Montezuma Castle	Homolovi
Navajo	Kartchner Caverns
Organ Pipe Cactus	Lake Havasu
Petrified Forest	Lost Dutchman
Pipe Spring	Lyman Lake
Saguaro	McFarland
Sunset Crater	Oracle
Tonto	Patagonia Lake
Tumacacori	Picacho Peak
Tuzigoot	Red Rock
Walnut Canyon	Riordan Mansion
Wupatki	Roper Lake
Yuma Crossing	Slide Rock
	Tombstone Courthouse
	Tonto Natural Bridge
	Tubac Presidio
	Yuma Territorial Prison

Not all properties with national or state park designation are included.

CASE STUDIES USING O-D DATA COLLECTION FOR STATEWIDE OR LONG-DISTANCE TRAVEL

The case studies here provide context and a guide as ADOT considers technologies for long-distance travel. They provide real-world examples highlighting factors that will be important to consider as ADOT contemplates the most effective method of capturing long-distance travel in Arizona.

When the Idaho Transportation Department (ITD) developed its statewide travel demand model in 2013, it used cellular phone data to develop O-D trip matrices to synthesize travel. A zonal system consisting of more than 4,600 traffic analysis zones (TAZs) was aggregated into 750 super zones for the analysis. A recommendation derived from that project was to use this type of O-D matrix estimation for short-term forecasting because a better understanding of how O-D cellular flows differ from those obtained from traditional methods was needed (Coladner, Stabler, and Sikder 2015).

The Maricopa Association of Governments (MAG) compared AirSage data with the 2011 MAG travel demand model, which was developed using the 2008 NHTS data. In using a 30x30 aggregated zone configuration, researchers found that the interzonal trip counts were comparable between the two methods, but that the cellular data had a difficult time capturing shorter trips, contributing to some disparities between intrazonal trips. Given that ADOT's focus in the current research study is on long-distance travel, this finding indirectly supports the use of cellular data for long-distance travel capture. However, the AirSage cellular data did produce higher estimates than the model during peak periods (Zhang et al. 2015).

Lee, Sener, and Mullins (2016) performed a literature review highlighting the uses of passive data, specifically considering mobile phone positioning, GPS, and Bluetooth re-identification. Additionally, they conducted a survey and interviews with transportation professionals responsible for travel demand modeling. These methods of O-D data collection allow passive data collection continuously, and can be more cost-effective than traditional travel surveys. However, the authors (Lee, Sener, and Mullins 2016) note that there are some concerns over using these collection methods for travel modeling—including sampling bias, lack of associated demographic information, and inexperience with using the data collected via these emerging technologies. The authors also note that "Overall, this research revealed that passive data are mainly being used for model validation or the development of model components, but rarely were used as standalone data sets" (Lee et al. 2016). This statement highlights the fact that ADOT is advised to maintain a big-picture perspective when considering what type of technologies to use in capturing long-distance O-D data for statewide modeling, as the best approach will likely include the use of multiple data sources.

Bekhor, Cohen, and Solomon (2013) considered the use of cellular data to develop O-D tables for long-distance trips in Israel, which has a very high cellular phone market penetration. After performing a pilot study to test data collection and processing, cellular data were collected for a 16-week period from March 7 to July 2, 2007, from the provider "Orange." Long-distance trips were defined to be those that were 2.5 kilometers (1.6 miles) or longer. The authors (Bekhor, Cohen, and Solomon 2013) found that the use of cellular data helped in addressing the fact that many household travel surveys do not capture long-distance trips very well. They note that within the data, there were some instances of zig-zagging between zones, recommending further research to develop smaller zones to address this issue. This highlights the importance of zone size in capturing long-distance travel via cellular data. The fact that travel purpose and traveler characteristics are not attached to cellular data collected for long-distance travel is a drawback (Bekhor et al. 2013). Zone size will be an important consideration for ADOT as it considers the potential purchase of cellular data for use in capturing long-distance O-D data.

Lu and Zhang (2015) studied methods to impute the trip purpose associated with long-distance trips. Using machine-learning techniques to impute the trip purpose associated with passively collected simulated data, they found the resulting accuracy to be higher (95 percent compared to 77 percent) when only two trip purposes were considered (business and non-business) than when four were considered (business, personal business, social visit, and leisure) (Lu and Zhang 2015). However, that study focused on the highly urbanized MAG region that encompasses metropolitan Phoenix. Imputation of trip purpose entails different challenges in rural areas with many recreational destinations. Nonetheless, this finding may have potential implications for Arizona as ADOT contemplates the use of passively collected cellular data in capturing long-distance travel across the state by trip purpose.

Wang et al. (2012) studied the use of cellular phone network data in tracking long-haul truck travel using AirSage data obtained from Iowa, Kansas, Missouri, and Nebraska truck stop databases. Tracking and freight identification algorithms were created to track freight movement. As part of a case study, the authors selected a truck stop as a test facility to analyze. They found that only 20 percent of the freight trucks traveled 80 or more miles for their tracked trip. The authors advise that this methodology—using AirSage cellular data in tandem with truck stop database information and the data analysis algorithms—could be applied in other areas as a means of reducing the need for other data collection efforts needed to capture long-haul trucking trips (Wang et al. 2012). ADOT may be able to tailor this recommendation to more effectively meet its needs in capturing long-haul truck travel.

Cutting-edge studies using GPS O-D data are being performed in Maryland and Minnesota. For the Maryland study, researchers are studying freight movements and freight corridor using INRIX GPS data with waypoints. The expansive data set (20 million trips, 1.4 billion waypoints, and 5.5 million unique devices from 148 data source providers) was collected in 2015 from commercial fleet sources (60 percent), consumer sources (31 percent), and mobile sources (9 percent). INRIX GPS speed (travel time) data were obtained from the Vehicle Probe Project Suite's Massive Raw Data Downloader tool. Speed data were combined with volume data in estimating delay (Eisele et al. 2015). There are plans to consider both local delivery fleets and private trucking fleets. Ultimately, the researchers plan to develop an interactive map that shows freight movement between ZIP codes. They also would like to study congested freight corridors and identify underused routes that may be considered as an alternative to congested freight routes (Eisele 2015). ADOT may be interested in developing a similar map as a way of visually displaying and interacting with O-D data.

The Minnesota GPS O-D study used three months of INRIX GPS data from spring 2015 in performing a congestion relief study for I-494/Highway 52 in the Twin Cities. Trip waypoints were used in assessing routes and ArcGIS was used to make the GPS data understandable in regional travel demand modeling software. Route movement and corridor travel patterns were studied and O-D tables developed. This type of study allowed large-scale, real-word travel data to be used in calibrating a travel model (INRIX Q&A Session, January 20, 2016, obtained from Paul Czech, Minnesota DOT Planning Director, System Planning). ADOT may find this study insightful if it decides to use GPS data in calibrating its statewide model.

BUILDING TRANSPORTATION MODELS FROM PASSIVE DATA

The literature review of the use of passive data for the development of travel demand models suggests their use technology is in its nascent stage. While a number of studies compare passive data to survey and model results (Hard et al. 2016a), the actual implementation of passive data in the model development process is sparse. Passive data represent observed travel behavior but without detailed demographic (personal travel) or industry (truck travel) information. Some of the first models to use passive data for model estimation are those for trucks.

MAG has undertaken significant research in this area in its development of a tour-based truck model using GPS data. A tour-based model replicates a vehicle's movement throughout the day, while a trip model only replicates distinct, unconnected trips. The MAG model uses GPS data to estimate tour generation, number of stops, purpose and location of stops, and time of day distributions (Kuppam et al. 2013). Another approach is to use Origin-Destination Matrix Estimation (ODME) to fit passive data. ODME is a mathematical method often used to fit an O-D matrix based on known data points. These data may take the form of a seed matrix, which is the observed O-Ds of the passive data, and control totals such as traffic counts. In ODME, confidence levels are assigned to either the seed matrix or the control totals, and the matrix is iteratively adjusted until the control totals are matched. The literature suggests that ODME techniques be used with caution as results may be over-fit or naively applied (Coladner et al. 2015). However, this approach is used to expand truck GPS data for the Indiana, Iowa, and Tennessee state travel models (Bernardin, Trevino, and Short 2015).

SUMMARY

This literature review summarizes an emerging field that is quickly evolving. The body of work and proposed methodologies for collecting, processing, and modeling long-distance travel using passive data technologies are expected to improve. The field has yet to congeal around standard or consensus methodologies for using passive data in travel modeling practice.

CHAPTER 3. EVALUATION OF TECHNOLOGIES AND DEVICES

OVERVIEW

This chapter describes the measures and criteria used in evaluating technologies and devices to use, in tandem with existing ADOT data collection tools, to develop O-D data on long-distance travel for the AZTDM. As of 2018, the current version of the AZTDM is version 2 – or AZTDM2, which is referred to herein as the AZTDM, recommendations will likely be implemented with future versions of the model. This chapter also discusses the possible technology or technologies to use based on the evaluation of types of long-distance travel modeled by the AZTDM. As established, the available technologies evaluated as potential sources of long-distance O-D data are cellular, GPS, LBS, Bluetooth, and Wi-Fi.

The evaluations and assessments discussed in this chapter are organized as follows:

- Assess the categories and aspects of long-distance O-D data needed for the AZTDM model.
- Compile the technical characteristics, capabilities, and limitations of available technologies for development of long-distance O-D data and summarize their strengths and weaknesses.
- Review the use of cellular and LBS data as the primary source for passenger travel and GPS data as the primary source for truck travel.
- Develop cost estimates by technology for acquiring the data needed and discuss potential options for data purchases.

ASPECTS OF LONG-DISTANCE DATA NEEDED FOR THE AZTDM

The type of technology, or technologies, to develop long-distance O-D data for the AZTDM depends on the aspects of long-distance travel needed in the model and the capabilities and attributes afforded by each technology. The comments in this section draw from survey responses obtained from the ADOT modeling team (personal communication, e-mail correspondence from Keith Killough and Baloka Belezamo, ADOT, June 28, 2016). For this study, ADOT is interested in long-distance passenger and weekday truck travel. Long-distance O-D data for other modes such as aviation, rail, and transit, are not a need at this time. Per the definitions in Chapter 2, ADOT stratifies long-distance highway travel into:

- I-I non-work passenger vehicle trips longer than 50 miles
- E-I/I-E passenger vehicle trips between other states and Arizona
- E-E passenger vehicle trips
- I-I passenger trips to and from special generators (mainly national and state parks)
- E-I/I-E passenger trips between other states and these special generators
- E-I/I-E passenger trips of less than 50 miles that cross the state's borders
- I-I truck trips longer than 50 miles
- E-I/I-E truck trips
- E-E truck trips

For each of these long-distance categories, the AZTDM relies on different trip expansion factors and disaggregation criteria, such as population and employment data, to distribute trips within Arizona and to and from regions outside the state. To this end, the primary needs to improve the AZTDM are long-distance passenger and truck O-D data that have greater detail and better reflect actual travel patterns. Currently, the state of the practice is to explore and implement techniques to develop trip-flow matrices derived from passive data. Trip-flow matrices represent TAZ travel from zone to zone. These matrices are typically based on mathematical calculations and disaggregation assumptions over several travel model steps; they are model outputs. The final trip-flow matrices are used to assign trips to a travel model network.

For long-distance travel, trip-flow matrices (derived from passive data) can be used as a direct model input or to calibrate calculation parameters of matrices produced in the model stream. Currently, empirical data to develop long-distance passenger travel matrices come from the NHTS and residence locations collected from park visitors. Beyond the NHTS data, the AZTDM also relies on data obtained from the Census Transportation Planning Package (CTPP). The CTPP is used to review and calibrate county-to-county and sector-to-sector trip flows. Improving data and sources to inform long-distance travel estimation is imperative, as data from the 2009 NHTS survey indicate that long-distance trips (longer than 50 miles) make up 2 percent of vehicle trips in the United States, but 23 percent of total VMT. The long-distance proportion of VMT is even higher in Arizona (41 percent), because long-distance travel − including E-E trips − dominates the large non-urban portions of the state (Cambridge Systematics et al. 2012; Erhardt 2012). Thus long-distance trips, while relatively few in number, account for a large proportion of the network traffic that needs to be modeled, especially on rural state highways that are intended specifically to accommodate such trips. The AZTDM long-distance truck model relies on the FHWA's FAF3.

One of the biggest deficiencies of long-distance data sources is the lack of *recent* data. With the NHTS performed about every seven years, and a lag between the date of the survey and availability of its results, the data used in an existing model may no longer reflect *current* travel demand. The use of new technologies in collecting trip O-D information will help address the data lag issue.

CHARACTERISTICS, STRENGTHS, AND WEAKNESSES OF TECHNOLOGIES

O-D data can be derived from cellular, GPS, LBS, Bluetooth, and Wi-Fi technology, but the data from each technology has different characteristics and attributes that lend themselves to different uses. Because of their unique characteristics, the technologies have differing capabilities and limitations in estimating various aspects of long-distance (and short-distance) travel. The key characteristics to be compared and evaluated between technologies include:

- Collection method: point sensor versus ubiquitous/crowd sourcing
- Data unit collected
- What the data represent
- Positional accuracy
- Sample saturation/penetration and frequency
- Continuity of data stream
- Trip/trip end development
- Available attributes, such as ability to discern vehicle type and residency status

Characteristics of Each Technology

Cellular O-D Data Characteristics

Cellular O-D data are a measure of estimated device movements between TAZs or other pre-defined geographies of the study area. Such movements may be person trips or population flows, depending on the context. The data unit collected from cellular technology is termed a “sighting,” although it involves no visual observation. The large majority of cellular sightings are generated when a mobile device interacts with a cellular network, which can occur when a call starts or ends; a text message is sent or received; or during a data exchange. When a device is stationary for a period, its location may be reflected as a cluster of many sightings. Sighting clusters are termed “activity points,” which can represent a trip end or stop, or one of the device’s primary resting places, such as a home, workplace, or school.

Device movements are developed by analyzing activity points over a period of weeks or months. AirSage’s proprietary algorithms impute the estimated trips of a device based on its patterns between home, work, and other activity locations. Trips developed from cellular O-D data are estimated from an analysis of a device’s movements for a day or day type (e.g., average weekday or weekend day) over the study period (Hard, Chigoy, Songchitruksa et al. 2016).

AirSage indicates that, on average, the accuracy of its data is within about 300 meters (roughly 1,000 feet), but varies depending on factors such as device type, device quality, and density of the cellular network (AirSage). The smallest time interval over which cellular data can be reported is three hours because of the frequency of cellular data collection. However, the sample penetration of cellular data is its best characteristic due to the ubiquitous use and high market penetration of cellular devices. The 2014 O-D study conducted in Tyler, Texas, found 198,000 unique devices included in the data, which represented a 17.1 percent sampling rate of the population (personal communication, Vijay Sivaraman, e-mail from AirSage to TTI’s Ed Hard, Subject: Tyler Data, August 25, 2015). The sampling frequency of unprocessed cellular data is random and haphazard because cellular sightings occur randomly based on the activity of the device user.

Cellular data sightings that occur by way of a call, text, or data exchange are considered “event-driven” interactions and also are known as call data records. Event-driven sightings are the primary source for developing cellular O-D data because they are the most numerous and can best determine device activity locations. Cellular sightings also occur when a device is moving, such as in a vehicle, from one cellular coverage area to another. These types of interactions are termed “handover” sightings. In rural settings, handover sightings may play an important role in cellular O-D data collection. In urban settings, handover sightings are less useful as they occur less frequently and represent transient points where a device is moving in a vehicle. Cellular data allows identification of selected trip purposes, including work, and a device’s general home location (by city or county). However, such data cannot discern between vehicle types (passenger versus freight).

If used for the AZDTM, cellular O-D data will be a measure of the estimated device movements between TAZs in a modified statewide zone structure optimized for cellular data capture. The movements represent travel flows between the TAZs based on the analysis of mobile device sightings and activity locations for pre-selected time period aggregations.

GPS O-D Data Characteristics

GPS data for O-D studies are available as primary GPS data or through third-party GPS data. Third-party GPS data can be obtained from companies such as INRIX or HERE Technologies (formerly Nokia), and consist of time-stamped location readings obtained from GPS signals. Data are generally purchased as aggregated O-Ds, aggregated O-Ds with waypoints, or trip matrices for a requested TAZ structure. Currently, GPS O-D data are derived primarily from mobile devices/phones, in-vehicle navigation systems, and GPS data obtained from commercial truck, freight, and delivery fleets. Data from mobile devices and in-vehicle navigation systems are referred to as consumer data. In-vehicle navigation system data are from passenger vehicles and are primarily non-commercial in nature.

The data unit collected with GPS data is a GPS transmitted location, or “ping,” that results in a time-stamped coordinate. The positional accuracy is more exact than for cellular data, and is in the range of one to 10 meters (3.3 to 33 feet). GPS O-D data for passenger travel have a small sample penetration rate (estimated to be in the range of 0.5 percent to 2.0 percent), but the rate continually improves as GPS data become more widely captured. GPS O-D data for truck/freight travel have an estimated sample penetration rate of about 11 percent (INRIX 2016).

With GPS data, trips are estimated using the GPS data stream. The data stream can consist of pings related not only to the trip ends, but also to waypoint information that coincides with the trip route. Data are processed by identifying trip ends based on dwell times when a device is stationary. Unlike cellular data, GPS data can include commercial and non-commercial vehicle trip disaggregation, although they are currently biased toward commercial data because of the higher commercial penetration rate. GPS data do not allow determination of device-owner residency status (Hard, Chigoy, Songchitruksa et al. 2016).

LBS O-D Data Characteristics

LBS data are crowd-sourced data obtained from any type of app that requests a device’s location or collects it in the background. Such apps and websites could include those for weather, navigation, retailers, shopping, and social connections. LBS data includes data from “non-navigation” GPS apps where a device user’s location is randomly collected based on the device use or location. The sample size of LBS data is less than that of cellular data, but higher than for GPS sources, especially for personal travel.

In transportation planning and for the purpose of studying travel, LBS O-D data are a measure of estimated (device or trip) movements between TAZs or other pre-defined geographies of the study area. Depending on the scale and context, these movements may be person trips, vehicle trips, or population flows. The accuracy of LBS data varies depending on the app and the technology used to locate the device, which can be GPS, Bluetooth, or Wi-Fi. Based on information from Streetlight and AirSage, two major providers of LBS data, LBS data can range in accuracy from a few meters to about 50 to 100 meters (164 to 328 feet).

LBS data are processed and analyzed using the same general methods as used for cellular data. The data are aggregated and processed to identify device (trip) end points and activity points, which include recurring resting places such as homes, schools, and workplaces. Device movements are developed by analyzing data end points and activity points over a period of weeks or months, using proprietary algorithms that estimate movements of a device based on its patterns between home, work, and other activity and end point locations.

Streetlight indicates that its LBS data sample size is approximately 12 percent of the US population (Schewel 2017), while AirSage states that its sample size is similar to its (prior) cellular-based O-D technology, which was typically 15 to 20 percent of the study area population. However, the researchers believe that both sample size estimates are reflective of urban regions, and the sample size of LBS data will be less in rural and remote areas. The smallest time interval over which LBS data can be reported is three hours, similar to cellular data.

Bluetooth O-D Data Characteristics

Bluetooth technology is a point sensor data collection technique that uses readers to capture media access control (MAC) address information (Hard, Chigoy, Songchitruksa et al. 2016). It acts much like a cord between two devices by creating a wireless personal area network to enable communication (Struyk 2017). Bluetooth technology makes active inquiries from an in-field reader and can attempt to start a two-way correspondence with Bluetooth-enabled devices. The effective range of Bluetooth readers depends on the situation and strength of the antennas used. Bluetooth equipment developed by the researchers has a range of about 300 feet, but this range can be increased (up to 1 mile) with a more powerful antenna (personal communication, conference call with TTI Senior Research Engineer Tony Voigt, October 20, 2016).

Study researchers found that Bluetooth is a reliable means of estimating E-E trips for regional O-D studies. In these studies, E-E trips are developed by matching the MAC address of devices at external cordon locations along highways around the periphery of a study area or region. E-E trips are the most typical type of trip collected using Bluetooth data, and are determined using travel time thresholds to distinguish from E-I/I-E trips. In processing Bluetooth data, the researchers take several precautions to ensure the privacy of the data, including stripping part of the MAC address and encrypting the data before transmission to a server (Hard, Chigoy, Songchitruksa et al. 2016).

Wi-Fi O-D Data Characteristics

Wi-Fi technology is similar to Bluetooth but relies on passive sensing of MAC addresses using Wi-Fi signals. Wi-Fi connects devices to Internet or Ethernet networks and can have a range up to 300 feet, but Wi-Fi hotspots can exist over long stretches if they overlap (Struyk 2017). Unlike Bluetooth, which prompts readers in range of the detector to emit a signal, Wi-Fi technology functions more like an ear, waiting for devices to ask if it is there; it does not allow active inquiry, which contributes to a higher probability of Wi-Fi matches occurring in slower traffic conditions. This trait makes Wi-Fi biased toward congested conditions, which contributes to a higher probability of matches occurring in slower traffic that typically occurs in urban settings. During free-flow traffic conditions in a highway corridor, Wi-Fi typically detects fewer vehicles traveling along the highway than (usually slower) vehicles turning from the highway. This is probably due to vehicles slowing or stopping at intersections prior to turning (personal communication, conference call with TTI Senior Research Engineer Tony Voigt, October 20, 2016).

The researchers found little research on the use of Wi-Fi technology for O-D studies to support a regional or statewide travel demand model. Dunlap et al. (2015) describes the implementation of this technology for estimating O-D data on transit routes. Acyclica, a Denver-based company, recently developed Wi-Fi-based sensors (for installation at traffic signals) that can collect O-D data between signals in a traffic signal network or along a signalized arterial or corridor. However, the primary purpose of this Wi-Fi product is to collect travel times and speeds to help monitor traffic congestion and improve signal timings (Yankus 2016). The City of Seattle installed Wi-Fi sensors at traffic signals for this purpose in 2015 (Kroman 2015). Similarly, in 2015, ADOT's Traffic System Management and Operations Group compared and evaluated the use of Wi-Fi and Bluetooth re-identification sensors for travel time data collection along a stretch of Interstate I-17 (I-17) in Phoenix (Paul and Murray 2015). The City of Phoenix is in the process of deploying Acyclia's Wi-Fi-based sensors for arterial travel time and congestion data (Paul and Murray 2015). Wi-Fi sensors developed by Acyclica also are used in Boulder and Greeley, Colorado, but as in Seattle and Phoenix, their primary purposes are for travel time, congestion monitoring, and signal timing enhancement, not for collecting O-D data (Acyclica 2017).

Strengths and Weaknesses of Each Technology

This section describes the strengths and weaknesses of cellular, GPS, LBS, Bluetooth, and Wi-Fi technologies with respect to understanding long-distance travel. Much of the information in this section is drawn from the TMIP report, *Synopsis of New/Emerging Methods and Technologies to Collect Origin-Destination (O-D) Data* (Hard et al. 2016a).

Cellular O-D Data

The large sample size associated with cellular data is one of its most desirable attributes. Cellular data also have a wide range of applications and can be used for several different types of studies. Other strengths of cellular data include its ability to estimate basic trip purposes and differentiate between residents and non-residents of a study area.

The key weaknesses associated with cellular data include the inability to discern vehicle type and the lack of positional accuracy of the data. This lack of accuracy makes it difficult or impossible to identify the exact route used between a trip's origin and destination at an urban scale. Because cellular data are anonymized, they are not directly linked to subscriber demographic information (Hard, Chigoy, Songchitruksa et al. 2016).

GPS O-D Data

Unlike cellular and LBS data, GPS data allow differentiation between passenger and truck/freight travel. GPS data are well-suited to a number of different studies, including corridor comparative studies, select link analyses, O-D matrix development, and truck/freight studies. GPS data also can be acquired with waypoint data, which allow the route used between trip origins and destinations to be determined. Waypoint data are especially useful for assessing long-distance trip routes and the sensitivity of traffic to congestion.

Among GPS data's weaknesses is the inability to distinguish between resident and non-resident travel. GPS data are biased toward commercial vehicles and currently have a low sample penetration rate for passenger travel. However, this can be overcome if sample penetration rates for passenger travel improve over time as expected (Hard, Chigoy, Songchitruksa et al. 2016).

LBS Data

The strength of LBS data is its considerable sample size combined with its moderately good location accuracy. Its use of multiple technologies to collect location data based on device use and location also is a strength. Like cellular data, LBS data also has the ability to estimate basic trip purposes and differentiate between residents and non-residents.

The weaknesses of LBS data include its inability to discern vehicle type and its limited ability to collect GPS data when a device does not have cell service. While LBS data accuracy is better than that of cellular data, it still is not good enough to reliably estimate routing between trip O-Ds at an urban scale. Like cellular data, LBS data are not directly linked to subscriber demographic information.

Besides applications related to the AZTDM, ADOT may wish to perform select zone analyses (between urban areas and special generators) and regional corridor studies, if it acquires LBS O-D data for statewide modeling. Due to its wide coverage and better accuracy than cellular data, LBS data are the best new technology option for I-I and E-I/I-E O-D data collection on passenger vehicles.

Bluetooth and Wi-Fi O-D Data

Bluetooth data collection is performed using roadside sensors. Rather than capturing trip ends or trip traces as do LBS, cellular, and GPS technologies, Bluetooth data reflect point-sensor readings detected from enabled devices that pass by the Bluetooth reader during the time of data collection. One of the strengths of Bluetooth technology is its ability to be used as a benchmarking tool to check against the data from LBS, cellular, and GPS sources.

Bluetooth is best suited for small-scale studies, since each data collection point requires a roadside installation. It generally has a longer battery life than Wi-Fi, but is slower and less secure (Predictable Designs LLC 2016). Bluetooth generally has a smaller effective range than Wi-Fi, although range estimates vary with antenna strength (Predictable Designs LLC 2016; [personal communication, conference call correspondent with Tony Voigt, October 20, 2016]). The data transfer rate associated with Bluetooth is much slower than the transfer rate of Wi-Fi data (2 to 3 megabytes per second for classic Bluetooth technology versus 100-250 megabytes per second for Wi-Fi direct technology) (Predictable Designs LLC 2016). Bluetooth is generally better suited than Wi-Fi for use in O-D studies, largely because it does not have the same congestion biases that are associated with Wi-Fi and it typically produces a higher match rate (personal communication, conference call with TTI Senior Research Engineer Tony Voigt, October 20, 2016).

Wi-Fi technology is well suited for congested settings and for capturing turning vehicles at intersections, but less so for O-D studies where data are typically collected along high-speed highways and freeways. Wi-Fi is generally faster than Bluetooth and has a larger effective range of detection. However, this is largely dependent on the strength of antenna used. Wi-Fi is considered more secure than Bluetooth (Mercedes-Benz USA 2012; Predictable Designs LLC 2016) and is compatible with multiple frequency types, whereas Bluetooth only works with one frequency specification (Mercedes-Benz USA 2012).

While Wi-Fi is good at detecting devices, it is not as good at matching them because it searches for signals only once every 8 to 20 seconds (personal communication, conference call with TTI Senior Research Engineer Tony Voigt, October 20, 2016). Wi-Fi has a high power consumption rate (Predictable Designs LLC 2016) and unlike Bluetooth, is not compatible with certain older smart phones (Predictable Designs LLC 2016).

A TTI field test found that Bluetooth is more suitable than Wi-Fi for collecting O-D data. In a closed system where the same traffic passed two consecutive Bluetooth and Wi-Fi equipped readers, an internal trial of equipment found that approximately 90 percent of Bluetooth devices were re-identified, versus only around 20 percent of Wi-Fi devices. Wi-Fi is generally associated with good *detection* rates, but it will not get as high a proportion of *matches* as Bluetooth on a roadway with uncongested, free-flow conditions. This advantage makes Bluetooth better suited than Wi-Fi for O-D studies along highways and freeways (personal communication, conference call with TTI Senior Research Engineer Tony Voigt, October 20, 2016).

Table 2 summarizes the strengths and weaknesses associated with cellular, GPS, Bluetooth, and Wi-Fi technologies.

Table 2. Strengths and Weaknesses of Technologies Relative to Long-Distance Data Capture.

Technology	Strengths	Weaknesses
Cellular	• Large sample size • E-I/I-E data can be used for regional model purposes • Can distinguish between some trip purposes • Can distinguish between non-resident commuters and residents	• Cannot discern vehicle type • Cannot easily ground-truth to urban-scale travel routes • Provides no direct link between cellular data and subscriber demographics
GPS	• Can distinguish between passenger and truck/freight vehicles • Is well suited for truck/freight studies (which may be a critical aspect of understanding long-distance travel)	• Unable to distinguish between resident and non-resident travel • Is biased toward truck/freight vehicles (though this bias can be mitigated) • Has low penetration rate for passenger travel (although expected to improve over time)
LBS	• Uses multiple technologies to collect data (GPS, Bluetooth, and Wi-Fi) • Moderately large sample size (though not as large as cellular) • E-I/I-E data can be used for regional model purposes • Can distinguish between non-resident commuters and residents	• Cannot discern vehicle type • Cannot easily ground-truth to urban scale travel routes • Lack of cellular service/connectivity limits LBS data collection via GPS
Bluetooth	• Can use to validate or benchmark data collected using other technologies • Has a better match rate than Wi-Fi • Has long battery life relative to Wi-Fi • Prompts nearby devices for response to obtain signal	• Unable to detect trip ends • Is not well suited for large-scale studies • Is less secure than Wi-Fi • Is slower than Wi-Fi • Has smaller effective range than Wi-Fi
Wi-Fi	• Is more secure than Bluetooth • Is faster than Bluetooth • Generally has a larger effective range than Bluetooth	• Unable to detect trip ends • Is not well suited for large scale studies • Has a shorter battery life than Bluetooth • Suffers from congestion bias • Waits to hear signals, does not prompt devices to respond

APPLICATIONS OF LONG-DISTANCE O-D DATA IN STATEWIDE MODELS

The use of new technology to estimate O-D data for short-distance or long-distance travel in urban and statewide modeling is an emerging field. States that have applied, or are working on, methods to apply long-distance O-D data to statewide travel models include Idaho, Indiana, Iowa, and Tennessee. These states have varying methods of incorporating these data sources for estimating long-distance travel. The methods include using data from new technologies as a calibration source, as a non-forecast model, and as information to develop truck trip matrices.

An update to the Tennessee Statewide Travel Model (TSTM) is underway. It will incorporate FHWA's recently developed rJourney model for estimating long-distance passenger travel (Bernardin, Ferdous et al. 2015). The rJourney model, which is in the public domain, synthesizes intercity passenger travel for an entire week across several modes for the entire United States (FHWA 2015). Current development of the TSTM is adapting the rJourney framework to the use of cellular data as a model calibration source. In calibration, model results are compared to known data sources such as journey-to-work flows from the CTPP. Depending on the amount of error in the comparison, model estimation parameters are reviewed and adjusted to achieve a better estimation versus the calibration data.

ITD is using cellular data to develop passenger trip-flow matrices for existing year estimations (Stabler 2014). These trip flows are considered non-forecastable, as the estimates are not used to develop model parameters that can be applied to socioeconomic data. In the ITD model, cellular data are incorporated through ODME. These data may take the form of a seed matrix, which is the observed O-Ds of the cellular (or GPS) data, and control totals such as traffic counts. In ODME, confidence levels are assigned to either the seed matrix or the control totals and the matrix is iteratively adjusted until the control totals are matched. The literature suggests that ODME techniques be used with caution, as results may be over-fit or naively applied (Coladner et al. 2015).

Cellular data do not provide estimates of truck trips. Accordingly, "new technology" long- and short-distance truck trip estimates use primarily GPS data. The use of GPS to estimate truck trips is being applied to the TSTM and the Iowa Statewide Traffic Analysis Model (iTRAM) (Bernardin, Trevino, and Short 2015), as well as the Indiana Statewide Travel Demand Model (ISTDM) (Bernardin et al. 2011). To produce truck trip estimates, a large sample of GPS trips is expanded using ODME and truck counts collected for various purposes (e.g., corridor studies and the Highway Performance Monitoring System) in each state. However, while current efforts using the ODME approach show promise, they have not produced methodologies to forecast truck or passenger vehicle travel using the results.

Because studies to estimate O-D travel in various states have yet to settle on standardized procedures, how ADOT implements these sources will influence the overall state of the practice. Each of these studies reiterates that the development of long-distance O-D data from new technology is shaped primarily by the different characteristics of cellular, GPS, Bluetooth, and Wi-Fi data and their ability (or inability) to capture certain types and components of travel.

LBS DATA AND PASSENGER TRAVEL

LBS data are currently the best option to use as the primary source to develop passenger O-D travel and trip tables because of:

1. LBS data's good sample penetration rate – in the 15 to 20 percent range
2. GPS data's low sample penetration rate for passenger vehicles – 0.5 to 2.0 percent
3. LBS data's better positional accuracy than cellular data
4. The ability of LBS technology to estimate activities, movements, and flows at larger geographies

Results from studies using LBS data are typically reported as "all traffic" because vehicle type cannot be determined precisely with LBS data. However, it is known that such data are processed at the device level and grouped by a device's home, which is imputed based on its nighttime resting place. This resting place is further defined based on residential land uses within the study area. Accordingly, the data provider (AirSage) and the researchers believe that truck travel is a small portion of the sample and the large majority of LBS data come from passenger travel. For large study areas, however, LBS data likely contain some truck travel, as workers who drive local delivery and service trucks keep their personal cellular device with them during the workday. As of yet, no studies have tested and demonstrated a mechanism for quantifying such trip making. Additionally, the resident data set excludes truckers and visitors who may stay overnight in a hotel for one or more days as they travel through or to and from the state. Travel data from devices for which a permanent "home" nighttime resting place cannot be established are provided in AirSage's visitor database.

Based on its May 2018 product documentation, AirSage developed a new long-distance trip matrix product that ADOT could use (AirSage 2018). The product is an extension of AirSage's standard trip matrix product, where long-distance O-D data are extracted by applying a filter to the standard O-D product that filters out all trips and tours that are at least 50 miles long. It distinguishes between two subsets of long-distance travel: one for daily trips that occur over a 24-hour period and another for trips that include overnight travel (personal communication, Ed Hard phone conversation with AirSage Senior Data Scientist, Vijay Sivaraman, October 27, 2016).

While the sample penetration rate of passenger travel from GPS data is low, it does have value. As the development of O-D data from new technology is still an emerging field, the comparison of GPS-based passenger travel to that derived from LBS and cellular data would be insightful and valuable. At least one vendor indicates that its sampling from sources of passenger travel data is slowly increasing, possibly raising the sampled percent of such travel in the future.

The current AZTDM uses various data and criteria in sub-models to synthesize, generate, and disaggregate the components of long-distance data including the NHTS, Special Generator surveys, and FAF3. In the model stream, the outputs of each of these models are eventually aggregated to a trip-flow matrix of long-distance trips only.

However, since new technology sources can now be used to sample and develop estimates of these flows at an individual state level for TAZs designed for that state, these long-distance sub-models are replaced with observed data that contain one observed trip-flow matrix of I-I, E-I/I-E, and E-E trips. This matrix, depending on the passive data used, can be subdivided into passenger and truck/freight (or special generator-related) components.

Since LBS O-D data are provided as zone-to-zone movements, they are better suited for developing I-I than E-I/I-E trips. To develop E-I/I-E trips, modelers must identify the external cordon location through which the trips passed. Identifying the specific highway (e.g., external cordon location) where LBS-derived trips cross the study area can be difficult, depending on the location. Difficulties arise because LBS O-D data are provided as zone-to-zone movements, and estimating the specific route of a short trip may not be possible because of LBS data's varying and marginal positional accuracy. Obtaining accurate locations is especially difficult when an external station is located in a densely developed area (where alternative routes are close to the external cordon location).

At a statewide scale, the difficulty in distributing LBS-based flows to specific highways is diminished because the external stations are typically farther apart and often located in rural areas with no other major roadways nearby. Most of ADOT's statewide external highway locations are in rural or undeveloped areas with few alternative routes, making it easier to assign LBS O-D flows. However, lack of separation and an urban setting at a few cordon locations will require additional analyses to better estimate the distribution of LBS O-D flows.

GPS AND TRUCK/COMMERCIAL TRAVEL

GPS is the best technology option for truck O-D data, since commercial fleets/truck travel have a significant sample penetration in GPS data and because LBS and cellular data focus more on non-commercial travel. GPS data aggregators can provide truck/freight O-D data by various weight classes or by type of fleet such as local delivery or private trucking fleet. Analyses of the truck travel patterns in total and then by these categories will provide insights on the habits and impacts of each group. Like LBS and cellular data, GPS data are compatible with large-scale studies and available for purchase.

GPS data can be acquired in the following forms:

- Trip records – a file containing individual trip O-D records.
- Trip matrices – trip matrix files where the trip records have been processed by the provider into O-D matrices based on a TAZ polygon layer provided by the purchaser. Separate matrices would be provided for E-E, E-I/I-E, and I-I trips (and various subcategories if desired).
- Trip records with waypoints – a file containing individual trip O-D records, plus a related file that includes waypoint data for O-Ds in the trip file.
- Vehicle GPS traces for American Transportation Research Institute (ATRI) truck data only.

GPS trip records are provided in comma-separated variable (CSV) format and users can aggregate and analyze the data as needed. Key fields in the data file are trip ID, device ID, start date and time, end date and time, longitude and latitude of trip start and end points, and a code for the type of provider.

Trip records with waypoints are provided in three CSV files: a trip records report file, a trip records report waypoint file, and a trip records report provider details file. The first contains the mode, latitude/longitude coordinates of trip start and end locations, the type of trip (I-I, E-I/I-E, E-E), and the type of data provider (consumer, fleet, mobile phone service). The second contains the waypoint sequence, the capture date and time, the latitude/longitude of the waypoints, and the zones that the waypoints intersect. The last provides details about the provider type (consumer or fleet), the vehicle category (consumer vehicle, taxi/shuttle/town car, field service/local delivery fleet, for-hire private trucking fleet), and the vehicle weight class.

Because of the positional accuracy of GPS data, the external cordon/highway location where a truck trip entered or departed the state can be easily identified. GPS waypoint data, which essentially show the route of the trip, can be obtained. Therefore, the setting (urban or rural) and relative spacing of external stations do not affect GPS's ability to decipher which external station was used for entering or exiting the state. The positional accuracy of GPS data allows accurate O-D assignment to the zone system when the data are overlaid with the statewide TAZ polygon layer developed for the study. The development of I-I, E-I/I-E, and E-E trips for GPS data (both passenger and truck/freight) are derived after GPS O-D and waypoints have been geocoded to the roadway network, TAZ system, and study area.

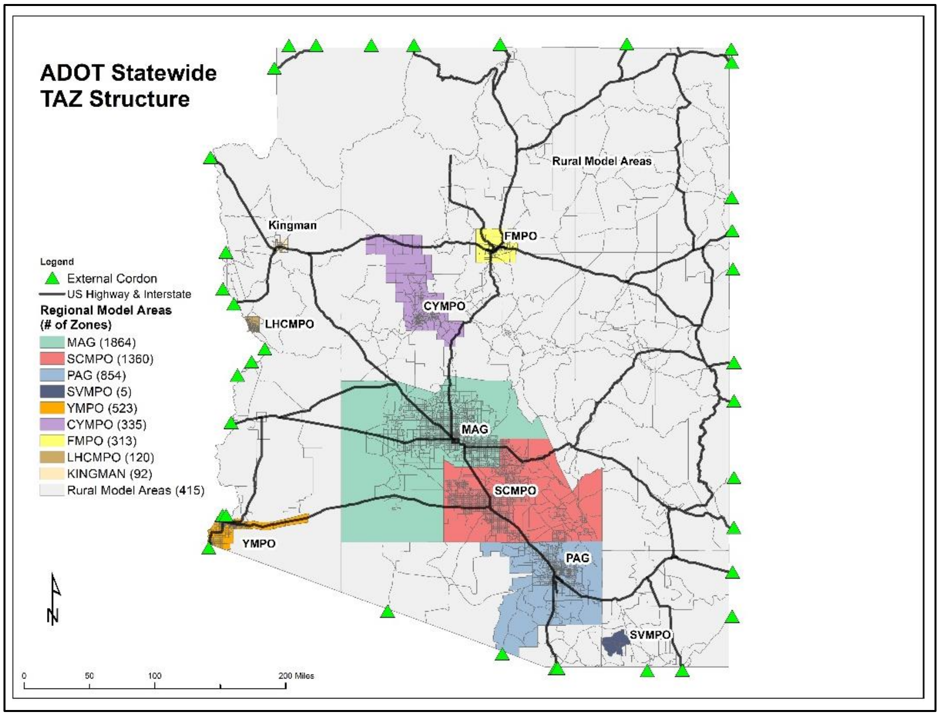
TAZ STRUCTURE DESIGN AND NEW TECHNOLOGY DATA

The AZTDM TAZ system for Arizona incorporates the urban TAZs of the state's MPOs as well as the zonal boundaries of many national and state parks. The model contains 5,881 zones. In addition to TAZs in the state, the model incorporates an external TAZ system with 108 periphery zones for states that border Arizona (including Colorado) and 44 state-based zones for the remainder of the United States, along with five zones for Mexico and one for Canada. Figure 1 shows the Arizona TAZ structure and Figure 2 shows the external TAZ structure surrounding the state. The Arizona TAZ structure incorporates regional model areas (Figure 1). Most, but not all of these areas, are part of an MPO. Additionally, the AZTDM incorporates Regional Analysis Zones (RAZ), which represent groupings of one or many of the model's TAZs.

When using LBS and GPS sources to estimate statewide O-D data, a zone system around the periphery of the state must be developed to capture external trip ends. The existing AZTDM zone structure incorporates geographies which support these peripheral zones including:

- The 16 generally county and state RAZs (in five states) that border Arizona
- Each remaining US state (44 total)
- Numerous TAZs for Mexico, including multiple TAZs in the Mexican state of Sonora
- One zone for Canada

Detail on potential refinement to these zones for the use of new technology to develop long-distance O-D data is provided in Chapter 4 under *Develop TAZ Structure and Network for LBS Data Capture*.



Source: ADOT Geographic Data Files, 2017, TTI Mapping and Graphics, 2017.

Figure 1. AZTDM Statewide TAZ Structure.

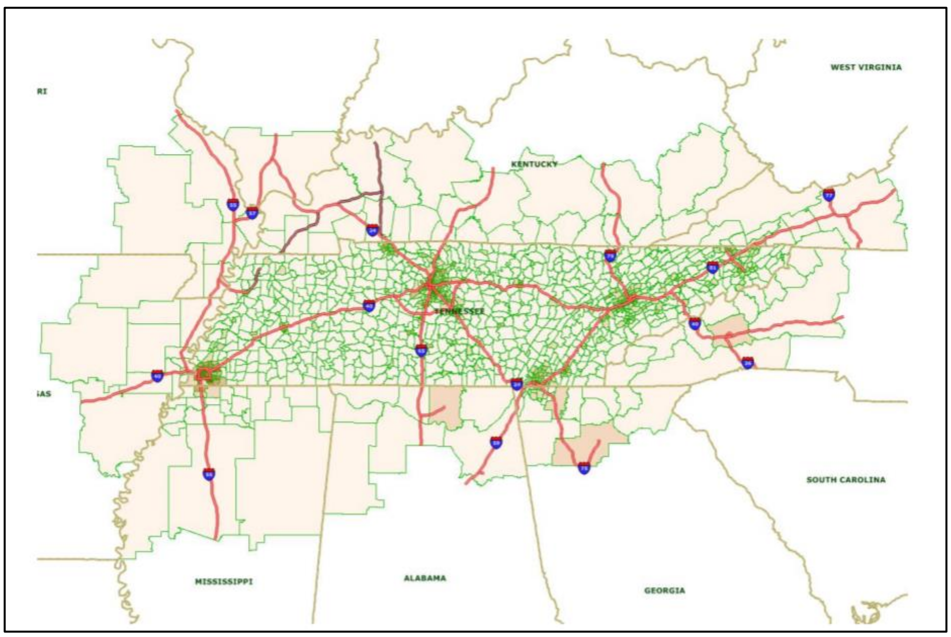
Source: ADOT, Geographic Data Files, 2017; TTI Mapping and Graphics, 2017.

Figure 2. AZTDM External TAZ Structure.

Development and Examples of External Periphery Zones

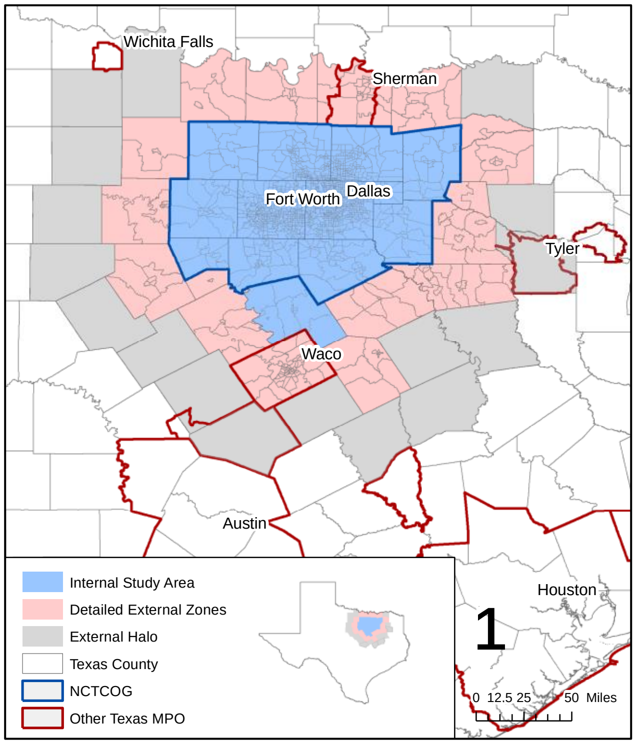
Processing of cellular, LBS, and GPS data is based on point and event locations of the source. Locations from LBS data can be filtered by AirSage based on the TAZs in which they occur. For E-E trips, the only knowledge needed is to determine whether the trip has at least one sighting in Arizona and two trip ends outside of the state. For E-I and I-E trips, the determination is based on whether the trip has at least one trip end outside of Arizona and a trip end inside of Arizona. GPS data use similar criteria in establishing internal/external classifications, but can have greater accuracy for trips traveling through Arizona at all points along the route (Hard, Chigoy, Songchitruksa et al. 2016).

External zones for capturing cellular, LBS, and GPS data are typically designed to include significant regions that interact with the state, such as nearby major cities in bordering states and significant highway corridors. In general, AirSage recommends that external zones designed to capture cellular data buffer a region at least 40 to 45 minutes from a study area border. For GPS data, this minimum buffer is unnecessary, as GPS points can be located in close proximity to a border. Figure 3 and Figure 4 show examples of external zones for Tennessee and the Dallas/Fort Worth metropolitan area based on the use of cellular data. External zones using LBS data must be designed to capture the last trip end (stop) before entering or after exiting the study area; they are designed to include nearby cities, towns, and activity centers outside the study area.



Source: (Bernardin, Ferdous et al. 2015)
http://tfresource.org/images/2/26/ITM16_Integration_of_the_National_Long_Distance_Passenger_Model_with_the_Tennessee_Statewide_Model_and_Calibration_to_AirSage_Data.pdf.

Figure 3. Tennessee AirSage Travel Analysis Districts.



Source: Dallas-Ft. Worth New Technology External O-D Study, TTI, 2017.

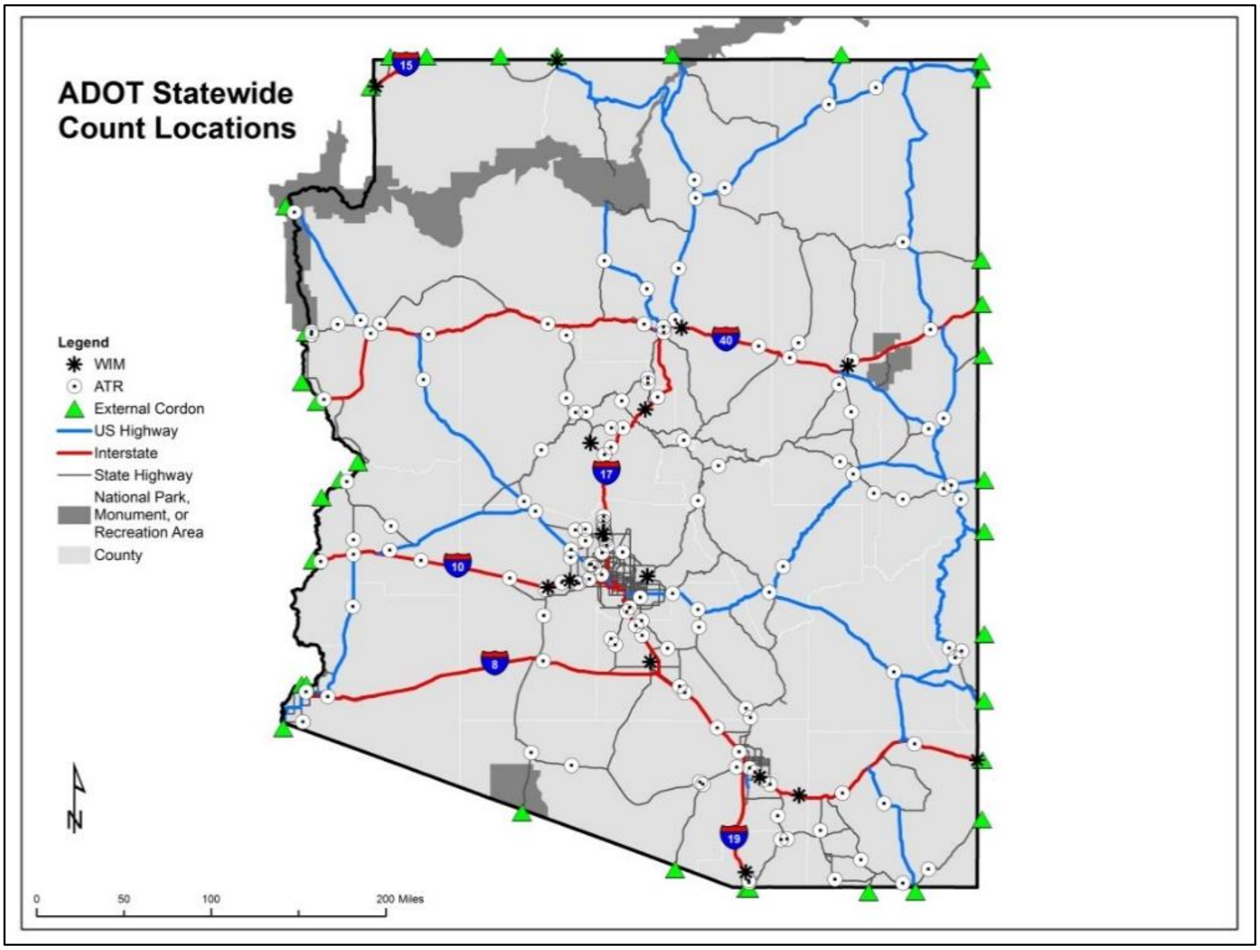
Figure 4. Dallas-Fort Worth External Study Zones.

External data from passive data sources can be integrated into the model in several ways. For small urban models, external zones can be associated with points where a major roadway crosses from the internal study area to the external study area. In this case, the external zone is sometimes referred to as a "travel shed." More complex networks and proximity of cities often warrant the development of a more detailed travel shed. Details on the development of these travel sheds are described in FHWA's TMIP webinar *New Technology Sources for Origin-Destination (O-D) Data: Overview and Lessons Learned* (Hard, Chigoy, and Songchitruksa 2016). Large urban regions and states need additional consideration as they offer many paths from an external zone to the internal study area. In these cases, it is advisable to associate the external zones with an external highway network that is integrated with the internal network. External travel networks are typically less detailed than internal ones and include only major highways. Using the external AZTDM network, trips can be routed directly to the internal network or pre-processed and associated with the most probable highway point that crosses the boundary.

Integration of Field Sensor Equipment and Bluetooth

Existing expansion and data fitting methods for passive O-D data require extensive data collection and management of primary data sources, such as traffic counts and socioeconomic information. These are often augmented by geospatial data on roadways and land use. US Census data are publicly available and socioeconomic TAZ data are available from regions that maintain travel models. The availability of traffic data is typically a function of the robustness of a region's or state's traffic counting program.

Currently, ADOT maintains 175 continuous traffic counting stations. These stations, illustrated in Figure 5 as ATR (automatic traffic recorder) and WIM (weigh-in-motion), collect traffic data by various vehicle classifications schemes, including axle configuration, length, and weight (ADOT 2016b). These stations are augmented by approximately 600 Intelligent Transportation System Freeway Management System detectors in metropolitan Phoenix and Tucson, along with others located primarily in major urban areas. Additionally, ADOT periodically collects traffic data at another 1,400 locations on the state highway system. Local jurisdictions also collect traffic data elsewhere as needed. As research continues, the demand for additional traffic counts and count estimates will likely increase. Arizona has large sections of rural highways where additional traffic counts would benefit expansion methods.



Source: ADOT, Geographic Data Files, 2017; TTI Mapping and Graphics, 2017.

Figure 5. ADOT Traffic Count Stations.

Recent work expanding truck GPS data in Iowa used a truck count data set of more than 63,000 counts, while in Tennessee more than 6,000 truck counts were used (Bernardin, Trevino, and Short 2015). In both cases, however, researchers suspect that many of these counts may have been "count estimates" conducted for the Highway Performance Monitoring System. Regardless, as research into this area continues, the demand for additional traffic counts and count estimates will increase. Arizona has large sections of rural interstates and highways where additional traffic counts would benefit expansion methodologies. Furthermore, numerous traffic counts are conducted throughout the state each year by counties, municipalities, and state agencies for routine traffic impact, signal, and highway corridor studies.

Benchmarking cellular, LBS, and GPS O-D data can be accomplished by installing Bluetooth sensors at selected ATR locations along key highway segments and then comparing the cellular, LBS, and/or GPS O-D results to those obtained using Bluetooth. The latter is better suited than Wi-Fi for benchmarking because it has a higher vehicle match rate on highways with free flow traffic conditions.

Bluetooth also is commonly used for travel time and speed data on urban arterials and freeways, as well as on long stretches of rural freeways nationwide. A properly designed and deployed system of Bluetooth readers has the potential to provide useful and comparative trip information for important corridors throughout Arizona, such as the rural interstate highways.

COST ESTIMATES OF ACQUIRING LONG-DISTANCE O-D DATA

This section provides cost estimates to acquire LBS and GPS data needed for the development of long-distance O-D data for the AZTDM. It also estimates the cost of limited collection of Bluetooth O-D data that could be used for benchmarking LBS- and GPS-derived O-D results. As of 2018, cellular data are only available as a legacy data source for years prior to 2017. In light of this, cost estimates are provided for LBS data in lieu of cellular.

Acquisition of LBS Data

The cost of LBS data is largely a function of the number of TAZs in the study area and to a lesser extent its population. It also is impacted by the following options and attributes that the data can provide:

- Use of internal and external zones
- Number of consecutive and non-consecutive months of data acquired
- Whether long-distance trips are filtered
- Full day and partial day aggregation options selected
- Level of trip purpose disaggregation selected, if any
- Level of residence class disaggregation selected, if any
- Level of demographic attributes selected, if any
- Types of additional reports requested

To include all 1,526 Census tracts, parks, and external zones, a statewide zone structure of about 1,800 TAZs may be needed. The cost of purchasing LBS data for 1,800 zones in Arizona, with its 2017 population of about 7 million, is estimated to be $150,000 to $200,000. This estimate is based on recent purchases of LBS data along with cost estimates for various regions in Texas, one month of data, and the same basic attributes needed in the AZTDM data set.

Acquisition of GPS Data

The cost of GPS O-D data is based primarily on the population of the study area. For GPS data, unlike cellular data, the number of TAZs in the study area does not affect the cost. However, the cost of GPS data, like that of LBS data, is based on attributes included in the data set. As discussed under *GPS and Commercial/Truck Travel* in this chapter, GPS O-D data can be purchased in the form of trip matrices, trip records, or trip records with waypoints. Acquisition of only the trip records will suffice for the basic O-D data needed for the AZTDM, but acquiring the GPS trip records with waypoints will provide many more uses for the data and a more accurate means of filtering the data for long-distance trips.

The cost of GPS trip O-D matrix for the AZTDM is estimated at $150,000 to $200,000, or about the same as the cost of LBS data. This estimate was provided by a GPS data provider and is based on a matrix with 1,800 TAZs and a population of seven million. The unit costs of both LBS and GPS data may vary.

Collection of Bluetooth Data

The cost for collection of Bluetooth data depends on the number of locations selected, and if ADOT purchases Bluetooth readers for continuous data collection or hires a third party for one-time collection. For developing cost estimates, an example consists of collecting O-D data along stretches of 10 highways around the state. This example would require the placement or installation of Bluetooth sensors at 20 of ADOT's existing ATR locations.

The cost for purchasing Bluetooth readers for 20 ATR locations is estimated at $2,800 per unit, for a total of $56,000. ADOT would have the option of purchasing either permanent readers or portable units that could be moved periodically between ATR locations. The estimate does not include the cost of installing the equipment, but it does include training and software set-up. The source of the estimate is a national provider of ITS equipment that has significant Bluetooth installations in several Arizona cities.

In lieu of purchasing Bluetooth equipment, ADOT could hire a third party to collect Bluetooth data on a one-time basis to benchmark LBS and GPS O-D results and obtain an alternative source for O-D, travel time, and speed data. Only a few companies in the United States collect Bluetooth O-D data, which is done using their portable Bluetooth readers. The primary costs for such a third party to collect the data include costs to ADOT such as rental of the Bluetooth equipment, travel, and data processing and analysis. The researchers are not aware of any companies in Arizona that collect O-D data using Bluetooth. Therefore, such a service may need to be performed by an entity from out of state, raising the travel cost. The researchers estimate the cost for one-time Bluetooth O-D data collection at $150,000 to $200,000.

CHAPTER 4. RECOMMENDATIONS AND IMPLEMENTATION PLAN

OVERVIEW

This chapter provides recommendations and an implementation plan on how ADOT can acquire and use new technology O-D data for modeling long-distance travel in the AZTDM, including:

- Recommendations for sources of new technology O-D data, including how and where ADOT would use new O-D data sources.
- Recommendations on implementation, including an overall plan as well as tasks and subtasks.

COMBINATION OF NEW TECHNOLOGIES NEEDED

Considering the current capabilities of new technology sources and methods, ADOT is advised to consider a combination of LBS, GPS, and point sensor data, with the latter including up-to-date classification counts. LBS and GPS technologies are needed to adequately sample and estimate long-distance travel by passenger vehicles and trucks/freight. ADOT may want to consider the use of Bluetooth on a limited basis for benchmarking O-D estimates obtained from LBS and GPS sources. Classification counts from ADOT's statewide system of ATRs, and perhaps counts from MPOs and local jurisdictions, will be needed to expand GPS data and adjust LBS data.

Figure 6 illustrates the recommended framework to develop long-distance O-D data for the AZTDM, using "new technology" data. The approach uses LBS data as the primary source to capture passenger trips and GPS data as the primary source to capture truck/freight trips. It segregates O-D travel into three trip categories (E-E, E-I/I-E, and I-I trips) for both passengers and trucks/freight. This segregation is necessary because different technologies are needed to obtain data on different types of travel.

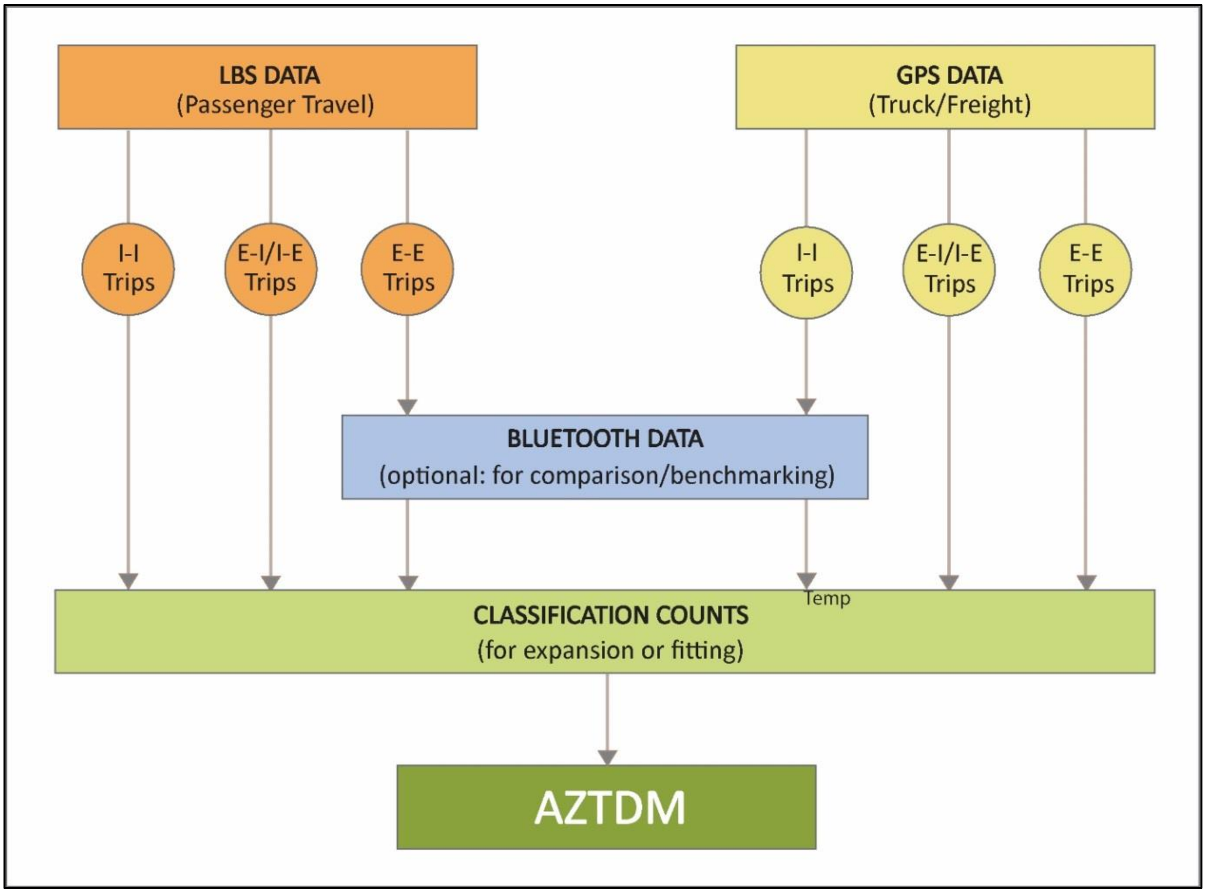


Figure 6. Implementation Framework for Long-Distance O-D Data for the AZTDM.

LBS Data: Primary Source for Passenger Travel

Due to the high LBS penetration rate and the ability to estimate flows and activities at a large geographic scale, it is recommended that ADOT use LBS data as the primary source to develop passenger O-D travel and trip tables. At a small urban scale, LBS data's lack of positional accuracy and generally low collection frequency may be a detriment, but at a statewide scale focusing on longer-distance trips for a 24-hour period, these characteristics are acceptable.

Further recommendations on the use of LBS data include consideration of:

- Use of AirSage's long-distance O-D product that includes only trips of 50 miles or more.
- Use of AirSage's resident data set for estimating passenger travel by Arizonans, and its visitor database (if desired) for analyzing travel by non-residents (e.g., truckers and visitors) that involve an overnight stay.
- Comparing long-distance passenger travel from LBS data to that developed from GPS data. This comparison may find data from GPS useful, depending on whether the GPS sample penetration rate for passenger vehicles has increased since this research was conducted.

LBS data's high sample penetration and wide geographic coverage will produce a better population and distribution of trip ends throughout the AZTDM TAZ system than any other technology source. LBS data are a better source than GPS for developing long-distance passenger I-I and E-I/I-E trips. This is especially true at the state-wide level as the TAZs will be much larger than those used at an MPO scale.

GPS Data: Primary Source for Truck Travel

Third party GPS data from proprietary sources such as INRIX, HERE, or ATRI should be used for developing truck O-D travel and trip tables for the AZTDM. ADOT is advised to use GPS O-D data for truck travel, since commercial fleets/trucks have a significant sample penetration in GPS data and since LBS data focus more on non-commercial travel.

GPS data can be acquired as trip records, trip matrices containing zone-to-zone movements, and trip records with waypoints. If ADOT needs only the trip matrices containing zone-to-zone movements, it may acquire GPS trip matrices. However, if ADOT needs or desires to examine truck flows on the highway network or in an individual corridor, it is advised to acquire trip records with waypoints. This option would give ADOT the ability to analyze truck routes on any highway segment in the state. It works better than zone-based data for external trips, since it can determine the specific highway used by vehicles that cross the state's borders.

Sources for I-I, E-I/I-E, and E-E Trip Types

Table 3 provides recommendations for O-D data sources by trip type. It is based on the current state of O-D practice and products, and might change if advances in methods and technology occur. The recommendation also considers updates to each long-distance element of the AZTDM simultaneously. In some cases, ADOT may benefit from updating the long-distance components (trucks and passenger vehicles) separately, depending on the urgency of the respective needs.

Table 3. Long-Distance Trip Type and Recommended Data Sources.

<table>
<tr><th rowspan="2">Trip Stratification</th><th colspan="2">Recommended Data Source</th><th rowspan="2">Comments</th></tr>
<tr><th>Primary</th><th>Secondary</th></tr>
<tr><td>I-I non-work passenger, >50 miles</td><td rowspan="3">LBS</td><td rowspan="3">INRIX GPS</td><td>50-mile threshold may be needed.</td></tr>
<tr><td>I-E/E-I passenger, including from bordering states</td><td>Will need detailed review of cellular network coverage.</td></tr>
<tr><td>E-E passenger</td><td>Will need AirSage external filter and home location report. May lose detail of trip origin/destination state.</td></tr>
<tr><td>I-I/E-I/I-E special generator</td><td>Combined LBS and GPS</td><td>N/A</td><td>Will need detailed review of cellular network coverage. GPS sample may be small.</td></tr>
<tr><td>I-I truck/freight, >50 miles</td><td rowspan="3">INRIX GPS</td><td rowspan="3">ATRI</td><td rowspan="3">May lose detail of trip origin/destination state.</td></tr>
<tr><td>I-E/E-I truck/freight</td></tr>
<tr><td>E-E truck/freight</td></tr>
</table>

IMPLEMENTATION PLAN: TASKS AND SUBTASKS

Figure 7 illustrates the recommended implementation plan for acquiring and developing O-D data from new technologies. It comprises five generally chronological tasks that may overlap. As the plan shows, some important planning and preparatory activities must be conducted prior to data acquisition. The plan's main tasks and related subtasks are described in the following sections. While not explicitly shown in Figure 7, linkages between tasks and subtasks mean that earlier activities in the plan will inform those that occur later.

ADOT can use the knowledge gained from this research study to gain practical experience with O-D data from emerging technologies.

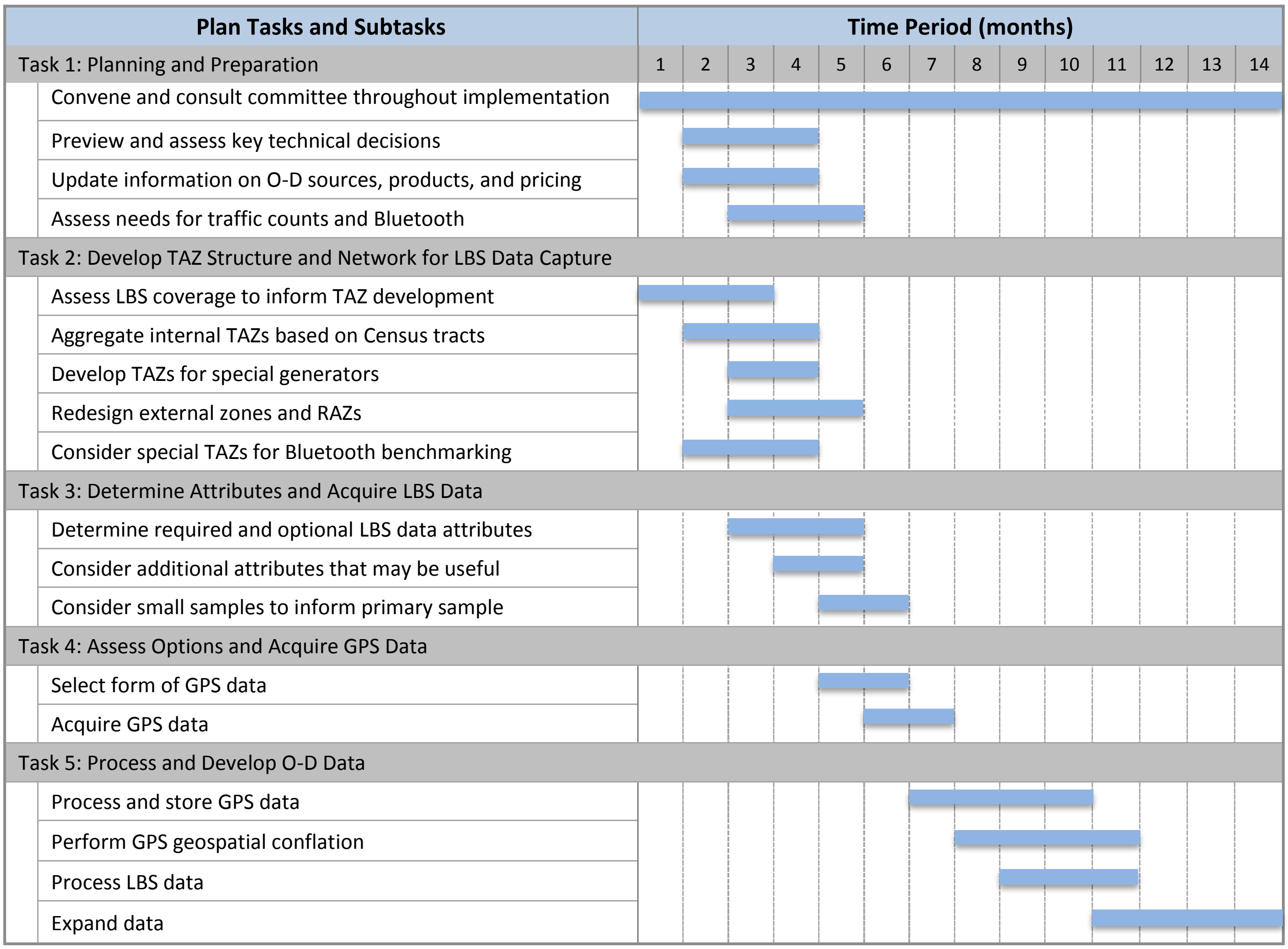


Figure 7. Implementation Plan Tasks and Subtasks.

TASK 1: PLANNING AND PREPARATION

This section proposes a step-by-step approach that to enable ADOT to integrate the newly obtained passive data to maximize value.

Convene and Consult an Advisory Committee throughout Implementation

It is recommended that ADOT form an advisory committee to develop guidance on supplementing new technology sources to the existing long-distance elements of the AZTDM. To support modeling, the new sources can provide metrics similar to those derived from traditional data collection activities, such as surveys and external cordon studies. Although the metrics may be similar, the data to derive them will be different in type, form, and sample population. New technologies offer rich sources of data that cannot be obtained by traditional means.

ADOT would select the members of the committee, which may consist of key personnel from ADOT (such as the Multimodal Planning Division's Transportation Analysis group and Transportation Systems Management and Operations), selected MPOs, and other stakeholders. The committee would be informed by, and provide guidance to, data providers as well as technical personnel tasked with processing new technology data into the AZTDM model stream. The committee also would review the results of any benchmarking and summary studies, such as comparisons between existing long-distance, O-D data sources and data from the:

- NHTS
- CTPP
- FAF
- Regional Travel Demand Models Census region to region trip flows
- Current AZTDM trip tables using Census region to region or county to county trip flows

Preview and Assess Key Technical Decisions

Implementation of this study's recommendations may occur in phases. As part of planning and preparation, ADOT, in coordination with the advisory committee, will review and assess technical decisions that will need to be made such as development of a TAZ structure suitable for new technology data and determining if O-D data should be acquired in the form of trip records, trip matrices, or trip records with waypoints. This review may help ADOT and the committee gain a more thorough understanding of the project and the implications that key technical decisions will have for the project's scope and budget. It also will help the advisory committee provide guidance to ADOT.

Since techniques and modeling practices for integrating LBS and GPS O-D data into travel demand models are not yet standardized, the ADOT statewide modeling team and advisory committee would need to review and assess changes to the model to efficiently use the new data. These will include changes to TAZ boundaries, the network outside Arizona, and potentially the modeling stream. ADOT also may review how different trip types and markets, such as consumers/passengers, trucks/freight, residents, commuters, and visitors will be handled in data acquisition and modelling.

Other areas for review of key technical matters and decisions will include:

- Developing an LBS-based TAZ structure, including aggregating TAZs, developing TAZs for special generators, and optimizing external TAZs. Prior to developing a new TAZ structure, an

assessment of cellular coverage in the state should be conducted to identify uncovered (dead) areas, especially along rural highways and around external cordon locations.
- Determining how best to integrate Arizona's in-state network with the highway network outside Arizona, including how to collect long-distance travel data from Mexico.
- Making choices related to acquisition of LBS data, including core (required) and optional attributes that can expand the value and uses of the data.
- Considering available sources and forms of GPS data, including the use of INRIX and possibly ATRI sources and options to acquire the data as pre-processed O-D matrices or as unprocessed trip records, with or without routing information on each trip.
- Processing and storing LBS and GPS O-D data, including data server and hosting needs, trip start and end criteria, matching trips to external cordons, and map matching − the process of matching coordinate data to roadway features.
- Using traffic counts for expansion, including compilation of traffic count data, the potential need for additional count data in rural areas, and alternative sources of imputed count data.
- Potentially using Bluetooth for benchmarking: identifying the best corridors or highway segments for comparing Bluetooth O-D to LBS and GPS O-D results, selecting existing Weigh-in-Motion (WIM) and Bluetooth locations, and assessing the costs of temporary and permanent Bluetooth installations (if needed).
- Selecting and using data matrix expansion techniques, software, and statistical error metrics.

While the decisions on data will not need to be made until later in implementation, an early review will help ADOT and the committee understand the forms and attributes of LBS and GPS data. Decisions on data forms and attributes will have a significant impact on implementation costs. An early review and discussion of key technical decisions may bring about changes in how future tasks are implemented.

Update Information on O-D Sources, Products, and Pricing

Since the "new technology" O-D data and practices are continually changing, and because data acquisition could occur a year or more after completion of this study, ADOT may consider conducting a status review of O-D data and sources to learn whether any advances occurred that would change the recommendations of this report. The status review might be one of the initial steps of implementing the recommended plan shown in Figure 7.

INRIX's *Insights™Trips* is recommended as the primary source for truck/freight travel due to its oversampling of these trip types and good routing/waypoint data. These recommendations may change in a year or more, since many other data providers like HERE, TomTom, Streetlight, ATRI, and Cuebiq are working to improve data quality and product offerings. During fall 2017, ADOT's Multimodal Planning Division entered into an on-call contract with INRIX and HERE for statewide travel time and performance analytics (Ed Hard phone conversation with Reza Karimvand of ADOT Systems Management and Operations, December 19, 2017). These sources of passive GPS data could be used for statewide O-D data collection if this project is implemented.

In addition to data quality and coverage, cost and pricing are important. Current providers of passive O-D data price them *a la carte*. This pricing is typically based on study area size, population, number of metrics (details of the data), and other factors. Discounts and other terms (such as obtaining a sample of the data) can be negotiated with most vendors. Each vendor of LBS and GPS data has a different level of flexibility, because of its unique business model and characteristics of the source data. The source data typically requires significant technological resources to store, archive, process, and access.

Assess Needs for Traffic Counts and Bluetooth

Traffic counts are an important element of the implementation plan as they serve as the basis for expanding GPS data. At external cordons, the counts will serve as control totals. Counts on Arizona's internal highway network will be used to fit GPS and LBS data O-D aggregations to better match counts and other observed data. Field collection of Bluetooth data will be an element of the plan if ADOT decides to use Bluetooth for benchmarking LBS or GPS O-D results.

As discussed in Chapter 3, ADOT maintains 175 continuous counting stations across the state and periodically collects traffic data at about 1,400 other locations. Additional traffic count data may be obtained from corridor studies and regional data collection activities, such as those conducted by urban counties, large cities, councils of governments, and MPOs. These existing count sources can provide foundational data for expanding and fitting passive O-D data, but supplemental count data will be needed in areas with no recent counts.

The researchers for this study recommend collection of vehicle classification counts on each highway segment included in the AZTDM that crosses the state border. New counts would be needed on sections of rural interstates and highways where no count data are available. They also may be needed at MPO boundaries for expansion purposes. Counts should generally be conducted during the period for which the O-D data are acquired. ADOT maintains a platform to distribute and process these data in its Statewide Traffic Count Database System. It is recommended that an assessment and compilation of all currently available sources of counts be taken at the beginning of implementation to determine where additional field data collection will be needed.

O-D data collected by Bluetooth sensors can be used for comparison to O-D results obtained using LBS and GPS data. ADOT, in coordination with local agencies, may be able to use existing Bluetooth installations to perform these comparisons. As part of developing a field plan to use Bluetooth for this purpose, the locations of all existing Bluetooth installations on ADOT highways would be identified, compiled, and illustrated on a map for consideration as locations for Bluetooth data collection. If any of these locations are at or near traffic count/data sites and on a section of highway where ample LBS or GPS samples will be collected, they may be candidates for Bluetooth sites. WIM locations with Bluetooth installations would be good candidate locations as they would allow comparisons between different types of trucks.

The combination of existing Bluetooth sites, cellular signal coverage, traffic data collection sites, and the LBS-based TAZ structure must be considered in determining the activity areas or highway segments for Bluetooth benchmarking comparisons. The cost of new Bluetooth sensor locations is provided in Chapter 3.

TASK 2: DEVELOP TAZ STRUCTURE AND NETWORK FOR LBS DATA CAPTURE

The existing internal and external/peripheral TAZ structure for the AZTDM was designed to accommodate traditional O-D data sources such as NHTS and FAF3. A new statewide TAZ structure must be developed before purchasing LBS-sourced O-D data, because LBS-based trip ends must be assigned to larger geographic areas designed to take into account the positional imprecision of the underlying data.

Assess LBS Coverage to Inform TAZ Development

As part of re-assessing AZTDM internal and external zones, ADOT is advised to examine statewide cell phone coverage maps as a guide to LBS TAZ and RAZ design, and to identify areas with no cellular service. These dead areas will affect LBS-based TAZ size and configuration, especially in rural areas. To ensure the most effective design of TAZs or RAZs, cellular towers and dead areas should be mapped:

- Along rural stretches of major highways inside the state and in external peripheral areas.
- Along and around designated external station locations.
- In and around national and state parks and other special generators.

If ADOT cannot obtain proprietary cell tower location and coverage information for direct review and assessment, AirSage can internally review tower locations, coverage, and signal strength to help guide design of rural TAZs.

Aggregate Internal TAZs Based On Census Tracts

It is recommended that the AZTDM's existing internal TAZs be aggregated into larger zones, coinciding with Arizona's 1,526 Census tracts, to better capture LBS data. In some densely populated areas, aggregates of Census tracts might be used to form LBS TAZs. Benefits of aggregation are that it improves the accuracy of LBS O-D data, conforms to the demographic and other data sources (e.g., CTPP) used to expand and enrich such data, and reduces the cost of LBS data acquisition. The equivalency of TAZs to Census tracts would allow comparison of data from new technologies with data from well-known sources. Without the zone aggregation and resulting reduction in the number of TAZs, the purchase and use of LBS data may be unaffordable.

Develop TAZs for Special Generators

As in the existing AZDTM, TAZs for national and state parks in Arizona should be included in the new internal TAZ system for LBS data. The coverage and configuration of the existing park TAZs in remote areas should be reviewed and reconfigured based on cellular coverage and tower locations. Other Arizona special generators that ADOT deems important should be represented by their own TAZs.

As an example, TAZs for large trucking distribution centers that substantial truck traffic (and little passenger traffic) could be represented as TAZs to isolate the truck traffic related to these sites. TAZs for major trucking sites and intermodal facilities would provide ADOT valuable information on the relative amounts and distribution of truck traffic at these sites. In addition, they would allow the LBS-based truck trips from these sites to be removed from the LBS data to ensure that trucks are not double-counted in O-D truck matrices, since the primary source of truck trip data will be GPS.

Currently, the AZTDM models truck trips from Mexico and Canada using FAF3. O-D data from new technologies are not currently available in sufficient sample sizes to provide comparable estimates to the existing model. Therefore, the two countries may be treated as special generators, with current procedures remaining in place.

Redesign External Zones and RAZs

The existing AZTDM external zones that cover neighboring states may be reconfigured as needed to better represent nearby cities and highway corridors. It is recommended that the new external zones do not cover an extensive area beyond the adjacent states. Redesigned external zones may be associated with an external highway network integrated with the Arizona network.

External zones serve several purposes: to distinguish between resident and non-resident travel; to capture vehicles/devices before they enter and after they exit the study area; and depending on forecasting needs, each external zone can be forecasted independently, allowing modeling of different assumptions of growth in external traffic.

Consider Special TAZs for Bluetooth Benchmarking

If ADOT decides to make limited use of Bluetooth to benchmark LBS and GPS results, then TAZs developed specifically for comparing Bluetooth data to LBS and GPS data may need to be included in the new TAZ structure. These TAZs should be considered case by case between activity centers and along major highway corridors identified as suitable locations for the technology comparisons.

TASK 3: DETERMINE ATTRIBUTES AND ACQUIRE LBS DATA

When acquiring LBS O-D data, ADOT may choose from numerous options and attributes that will be required to meet its objectives at a reasonable cost. Some attributes will be required for the AZTDM, while others will be optional. Figure 8 shows an example of the table of data attributes and options that AirSage uses in its pricing calculator sheet.

The selection of options and attributes would allow ADOT to customize the LBS O-D data set and, if desired, add information. Discussions of the required versus optional (desirable) attributes would need to occur between the LBS data vendor, the advisory committee, and ADOT technical staff before an LBS data purchase order is submitted.

Required and optional attributes to be considered with the purchase of LBS data include:

- Zone options such as internal zones only, internal and external zones, and external to external zones that require a long-distance trip filter.
- Day aggregations such as average weekday, average weekend, and average specific day of week.
- A.m. peak period, p.m. peak period, 24-hour total, and others.

- Resident class attributes that can distinguish between residents, short-term visitors, and long-term visitors.
- Trip purpose options and demographic information based on Census tract data of a device's nighttime resting place.

While the number of attributes selected will affect the cost of LBS data, the most important factor determining the price is the number of zones in the study area.

ADOT may consider the seasonality (time of year) for which LBS data are purchased. Data from non-summer months, when public schools are in session, are typically used when acquiring data for average weekday travel. However, if ADOT wishes to acquire data for different types of seasonal travel, it can purchase data for specific months, such as June and July during peak vacation season.

Product Trip Matrix	
No. of zones	
No. of months covered per report	
No. of reports	
Population covered	
Long-distance trip filter	
Internal and External Zones Options:	
Internal zones only	
Internal and external (all I-I, E-E, I-E and E-I trips)	
Internal and external (only I-I and E-I/I-E trips)	
Day Aggregation Options:	
Average weekday (Tu-Th)	
Average weekend day (Sat-Sun)	
Average full week (Sun-Sat)	
Total weekdays	
Total weekends	
Total full week	
Average specific day of week	
All individual days (enter no.)	
Total specific day(s) (enter no.)	
Total	
Day Part Aggregations (3-hour minimum)	
Early AM county to county	
AM peak (6:00 AM-9:00 AM)	
Midday	
PM peak (4:00 PM-7:00 PM)	
Late PM	
24-hour total	
Total	

Trip Purpose Attributes	
3-class (HBW, HBO, O)	
9-class (HBW, HBO, WBO, WBH, WBW, HBH, OBO, OBH, OBW)	
Total	
Residence Class Attributes	
2 categories (residents, visitors)	
6 categories (resident worker, home worker, in-commuter, out-commuter, short-term visitor, long-term visitor)	
Total	
Demographic Attributes	
Bundle includes the 3 below:	
Annual household income (Census bins)	
Age (Census bins)	
Autos (Census bins)	
Total	
Optional Add-on Reports	
Home-Work Matrix Report	
County to county	
Use study zones	
Home Location Report	
Home locations only in study area	
Home locations nationwide	

Key
B = based, E = external, H = home, I = internal, O = other, W = work
Source: Bill King, AirSage, personal communications, August 8, 2018.

Figure 8. Example Required and Optional Attributes for Pricing LBS Data.

Determine Required and Optional LBS Data Attributes

The decision to purchase LBS data may be based in part on unique travel characteristics in which ADOT may have an interest. The data purchase may have added value with the inclusion of optional attributes. It is recommended that these options be presented to the advisory committee for input. ADOT may wish to consider the following options:

- The daily long-distance trip filter
- Internal and external zone options developed to include all I-I, E-I/I-E, and E-E trips
- Average weekday travel based on a 24-hour period
- Resident class attributes with the six residence categories in Figure 8

The daily long-distance trip matrix is an extension of AirSage's standard trip matrix product, but uses a filter to provide only trips and tours at least 50 miles long. As part of this filter, regular daily trips are chained, where appropriate, to form long-distance tours. With this option, ADOT would receive the standard daily matrix of all trips, along with the long-distance trip matrix (since the standard matrix must first be created before the long-distance matrix can be developed).

LBS data can be purchased with either two or six resident class categories (defined in Table 8). The six-class option has a higher price, but can distinguish between short- and long-term visitors. Both resident class options collect information on visitors, but the six-class option includes visitors who stay overnight in the state while the two-class option does not.

The final recommendation on LBS data acquisition should be made only after ADOT and the committee have reviewed and assessed the LBS products and options that are available when the data are needed.

Consider Additional Attributes that May Be Useful

In addition to data for an average weekday, ADOT also may want to consider acquiring data for an average weekend. This option would allow comparison of long-distance travel between weekdays and weekends. Such data are especially important for state routes where travel patterns and vehicle types differ markedly between weekdays and weekends during certain seasons. Prominent summer examples are I-17 and State Route 69 between Phoenix, Prescott, and Flagstaff; State Route 87 between Phoenix and Payson; and the highways to and from Sedona.

ADOT also may have an interest in a home location report. Under this option, the home location (typically Census tract) of the device owner making trips is provided. This option would allow ADOT to study where travelers to any destination (TAZ) in the state reside, in Arizona or elsewhere in the United States.

Consider Small Samples to Inform Primary Sample

If ADOT is uncertain about the best month(s) for which to purchase LBS data, it could first purchase a group of smaller, less costly, samples for different months and compare contents. The comparison could help ADOT make its choice.

The cost of LBS data is based primarily on the number of aggregated LBS TAZs in the study area. For obtaining sample data sets, the TAZs could be aggregated to a small number, obtainable for a modest cost. For example, sample data sets for a TAZ structure aggregated to just the 15 counties in Arizona and a few counties, or combined counties, outside the state would show summary data of O-D interactions between counties. These county-to-county sample data sets could be purchased for several months or for all 12 months. They represent average daily total O-D interactions for each month and would allow determination of the best geographic size and period for purchasing the final, more detailed data set for long-distance O-D estimations. The sample summary data sets could be used to compare the number and distribution of external trip types, such as resident and visitor travel in different months. They also could be compared to regional travel flow estimates derived from other sources. The Florida DOT used this approach of purchasing small samples to inform decisions on the primary sample in a recent statewide O-D study. (Personal communication, Ed Hard phone conversation with AirSage Senior Business Development Executive, Bill King, August 28, 2017.)

TASK 4: ASSESS OPTIONS AND ACQUIRE GPS DATA

Secondary source GPS data are the best source for development of long-distance O-D truck travel and trip tables for the AZTDM. The data also may be used as a secondary or supplemental source for passenger vehicle O-D data. There are numerous providers of secondary source GPS data and some have a product that allows pre-processed GPS trip records with origins and destinations to be sold to an end user. O-D data from these products can be provided with or without routing information. Vendors have different ways to process O-D data into E-I/I-E and E-E trip types. One product provides an E-E trip matrix as part of its standard output, while another requires the end user to further process and analyze the data to develop E-E trips.

Unlike all other GPS providers, ATRI offers only truck data from fleets. These data focus on nationwide long-haul trucking, are available only in the form of raw trip traces, and must be processed and analyzed by the end user to identify origins and destinations for trip records. Regardless of source, raw GPS trip traces need significant post-processing by the user. This processing in turn requires large computing resources and data management to extract and impute important details.

Select Form of GPS Data

Secondary source data can be acquired as raw GPS trip traces, trip matrices, trip records, and trip records with waypoints. ATRI, the only source that provides GPS data in the form of raw trip traces, may not be a good candidate for ADOT's primary source of GPS data, since it may not provide as good a sample of local and regional truck carriers as other firms. However, ATRI might be a good supplemental source of GPS data due to its long-haul trucking emphasis. The forms of GPS data that ADOT is advised to consider for the AZTDM include trip matrices, trip records, and trip records with waypoints. The following briefly describes these three options.

Trip Matrices

Acquiring trip matrices will provide ADOT TAZ-to-TAZ, O-D flows for the AZTDM, but under this option, ADOT does not actually receive any GPS data. Instead, it receives TAZ-to-TAZ flows based on the provider's processing of the GPS data to ADOT's TAZ system. This option would probably be the least costly, but at the information received would have less value than that from any other option. Acquiring just the trip matrices does not allow other uses, since the matrices exclude GPS trip end and waypoint data. Additionally, the trip matrices are provided unexpanded and will require expansion by ADOT.

Trip Records

Acquisition of trip records is another option to develop O-D data matrices needed for the AZTDM. Under this option, ADOT receives the trip records and then must process the data and assign the trip ends to the TAZs to develop O-D matrices. Acquiring trip records would allow ADOT to perform additional zone-based studies and queries to examine O-D amounts and distributions between zones. For example, a select zone analysis could be performed to determine the origins of all trips to Grand Canyon National Park or the destinations of trips that enter the state on a specific highway. This option, like the trip matrices, does not include data on the routes used between the trip origins and destinations.

Trip Records with Waypoints

Acquiring the GPS trip records with waypoints is an option ADOT may want to consider seriously, although the off-the-shelf purchase price may be relatively high. The addition of waypoints would allow the data to be used not only for the statewide model, but also for planning and operational studies throughout the state. For example, ADOT could use the data to conduct corridor studies, routing studies, and select link analyses on any highway or freeway of interest. The addition of waypoints will improve the accuracy of state border crossing points and enhance the development of E-I/I-E matrices. The researchers believe that the waypoint option offers ADOT the best value for the money, since it can be repurposed and used for many other transportation studies or projects. It is recommended, however, that ADOT obtain cost estimates for GPS data with and without waypoints to assist in making the decision between options.

Acquire GPS Data

Based on the current abilities and products of GPS data providers, it is recommended that ADOT consider acquiring secondary GPS data from INRIX's *Insight Trips* O-D product as the primary source of truck/freight O-D data for the AZTDM. Trip data from this product also can serve as a supplemental source of passenger O-D data. In order to ensure a sufficient sample of long-distance trips (including school trips) for an average weekday, ADOT is advised to acquire GPS data for four to five non-summer months. This could be one consecutive fall/winter or winter/spring period, or split between spring and fall.

INRIX believes its GPS data sourced from mobile devices and navigation apps to be primarily from (non-commercial) passenger vehicles. Although the sample penetration of the data is currently low, ADOT can use it to supplement the LBS-based passenger O-D data if the sample is large enough.

As with the LBS data, a final recommendation on GPS data acquisition should not be made until ADOT and the advisory committee assess the GPS products and options available when the data are needed. The costs of all GPS products and options (in comparison with their expected value to ADOT) can then be reviewed and considered. The use of GPS data for travel modeling purposes is evolving and improvements are likely. For this reason, all secondary-source GPS data providers and products should be re-examined before acquiring the data.

TASK 5: PROCESS AND DEVELOP O-D DATA

Regardless of the data source – LBS or GPS − the goal is to normalize all data sources so that they share similar O-D conflations and are attributed to similar roadways.

Process and Store GPS Data

The behavioral complexities of GPS data include whether the movement of a vehicle from one location represents a trip. For example, the vehicle could travel from an origin to a destination, but stop at a traffic signal during the trip. GPS data must be processed so as not to interpret this stop as a trip end. Similar problems occur when vehicles stop due to congestion, or when loss of signal occurs under bridges or at other dead spots. Translating GPS data to represent distinct trips involves processing the data through an algorithm that examines vehicle speed, stop times, and sometimes land use. GPS providers of trip data apply a proprietary algorithm to segment these trips. Providers also will reduce the GPS signals from raw time intervals to intervals of one or more minutes, anonymize all trips with a unique ID (to protect privacy), and remove the first and last portions of the trip for additional anonymization. This significantly reduces the size of the data sets, but depending on the coverage, scale, time period, and place of origin of the data purchased, the GPS data sets can be very large. Similarly, commercial vehicle data provided by ATRI can be large, but represent the movement of a vehicle over an entire day. ATRI does not predetermine trips in its data products, and it is up to the purchasing agency to develop an algorithm suitable for its needs.

Regardless of the product or mix of products, ADOT would need to plan for data storage and analytical software for processing large data sets. Storage options include cloud-based services via Amazon Web Services, Dropbox, or other providers. Analytical software can include the R open source statistical platform, SAS, and Matlab as well as relational SQL databases or NoSQL databases. Depending on the data size, analyses can use Hadoop Distributed File System and overlay a data access and processing library such as Apache Spark. The latter relies on multiple computers to process a task, and works well for aggregation and sorting of very large data sets.

Perform GPS Geospatial Conflation

Regardless of the data source, GPS data should be conflated to other geographic sources. At a minimum, these include origin and destination TAZs, land use, and matching to regional roadways. The latter process is termed “map matching,” the process of matching coordinate data to roadway features. Most mobile device users see this process when using directions on mapping apps. Map matching is a computationally intensive process as each GPS point can be associated with several roads, depending on the detail of the geographic line work of the roadway used. For example, GPS points along dense highway corridors can be associated with the main lanes, feeder roads, or ramps. Figure 9 shows an example of matching GPS data to road segments.

Source: TTI.

Figure 9. Matching GPS Data to Road Segments.

As part of the data purchase, GPS providers also can conflate geographic elements, such as TAZs. This can save time and resources, but ADOT may wish to consider its own processes to conflate geographic data, as this will provide the greatest flexibility in use of the data and avoid potentially costly reprocessing.

Process LBS Data

The LBS data set from AirSage is relatively small and represents a matrix of trips based on the electronic TAZ shapefile provided by the purchasing agency. LBS data can be processed in a number of ways to add detail similar to GPS data. The most common method is to directly associate the data with roadways using a transportation network and traffic assignment algorithms. However, one of the most important aspects of processing LBS data is development of the geographic TAZ structure.

Expand Data

Once GPS and LBS data sources are sufficiently normalized to common roadway and TAZ features (whether areas or boundaries, natural or artificial), they can be summarized in an O-D table or trip matrix. The trip matrix is used for mathematical data expansion procedures and for combining the results. Each data set may need segmented expansion procedures, such as between passenger and commercial vehicles, and to control for differing error between them.

Data expansion is the process of growing (expanding) sample data to match a known population. Part of data expansion is the calculation of different weights representing each category of the population relative to the sample, and vice versa. This process controls for biases in the sample data that may cause overrepresentation of some population groups. For external analysis, the basis for expansion is traffic counts at the boundary of the study area, which represent the known population to which data samples will be expanded. Mathematical procedures to calculate the factors include Iterative Proportional Fitting (IPF) and Fratar.

An alternative method of data expansion is ODME. This process is more sophisticated than data weighting, which seeks to fit input matrix data of O-Ds to match numerous control totals in the form of traffic counts.

External Long-Distance Trips

IPF is the recommended procedure for fitting external GPS and LBS data to traffic counts. This procedure iteratively fits trips in an O-D matrix to the traffic counts at the external cordon stations in Arizona. It is recommended that traffic counts be segmented, at a minimum, into passenger and commercial vehicles. The process is iterative because initial weights are estimated based on the ratio of the total trips from all origins to the traffic count at the destination. The result of this first estimation will produce an error estimate of weighted trips versus the total traffic counts. Using this error, the weights are recalculated iteratively until the error is minimized. The resulting weights can then be applied to the aggregate O-D trips in the GPS and LBS data.

Internal Long-Distance Trips

IPF can be implemented with relative ease using statistical programs or Microsoft® Excel. ODME expansion is a more complex mathematical method often used to fit an O-D matrix to known data points. These data may take the form of a seed matrix, which is the observed O-Ds of the passive data, and control totals such as traffic counts. ODME techniques have traditionally been used to adjust travel demand model outputs to more reasonably fit traffic counts. ODME is especially important for detailed operational models where accurate traffic estimates are integral to design decisions (Patman et al. 2016). For some models, this process may be the most reasonable method for fitting several layers of ground-truth data, such as traffic data and observed O-D data from GPS and LBS sources. In contrast to IPF, ODME must match hundreds or thousands of control totals represented by traffic counts using a seed matrix. Confidence levels are assigned to either the seed matrix or the control totals, and the matrix is iteratively adjusted until the control totals are matched across the entire network. ODME using traffic count data considers the counts to represent the known total of trips, so the trip-flow matrices are iteratively adjusted to match the counts. Data preparation for ODME, more than for IPF, is an intensive exercise that must be approached with care while accounting for the results of the external long-distance estimates.

Although the literature suggests that ODME techniques be used with caution (Coladner et al. 2015; Cambridge Systematics et al. 2012), such methods using GPS data are currently applied for the TSTM, the iTRAM, and the ISTDM (Bernardin, Ferdous et al. 2015; Bernardin, Trevino, and Short 2015; Bernardin et al. 2011). The techniques also are being used to adjust trip-flow matrices from LBS data for the TSTM and the Idaho statewide model (Stabler 2014).

Use of passive O-D data for modeling of long-distance travel is an emerging field and methods to expand and fit such data are in a state of flux. Passive O-D data, while having large samples of travel data, lack the richness of traditional surveys and other data on socioeconomic and trip characteristics. There are current proposals for research to address this, but so far they are only ideas. Passive O-D data are integral, not only to travel models that reflect existing travel patterns, but also to those that forecast patterns. Additionally, GPS data represent a large geospatial data set. Working with these data requires computational and data management expertise that many organizations may lack, so an organization may choose to invest in specialized training or in hiring staff with advanced database and programming skills.

SUMMARY OF RECOMMENDATIONS

The recommendations for the use of "new technology" sources to estimate long-distance travel for the AZDTM relate to:

- The best technologies and associated O-D data to use for different types of passenger and truck/freight travel.
- The implementation plan setting forth tasks and subtasks for integrating these O-D data into the AZTDM.

Considering their current capabilities and limitations, LBS and GPS technologies are appropriate sources for developing long-distance O-D data for the AZTDM. The primary sources are LBS data for passenger travel and GPS data for truck/freight travel. The researchers recommend the AirSage long-distance O-D product for LBS data and a combination of sources, such as INRIX, HERE, and ATRI, for GPS data. ADOT also is advised to consider Bluetooth to a limited extent for benchmarking O-D estimates obtained from LBS and GPS sources. Implementation will require the use of ADOT vehicle classification counts, supplemented by count data from MPOs and local jurisdictions to expand GPS data and adjust LBS data.

The recommended implementation plan consists of five chronologically overlapping tasks, namely:

- Task 1: Planning and preparation
- Task 2: Develop TAZ structure and network for LBS data capture
- Task 3: Determine attributes and acquire LBS data
- Task 4: Assess options and acquire GPS data
- Task 5: Process and develop O-D data

In Task 1, ADOT would recruit an advisory committee consisting of key personnel from ADOT, MPOs, and other stakeholders to provide guidance on technical matters and help make technical decisions. These decisions may involve data acquisition, potential changes to the model and TAZ structure (as needed to integrate LBS and GPS data), and the possible use of Bluetooth as a source for comparing results. In addition, ADOT would obtain an update on O-D data sources and prices as part of this initial task. New technology O-D data and practices continue to evolve, so an update may uncover advances that would change the recommendations of this research study. (In fact, technological advances have occurred between the initial research for this study and its publication.)

In Task 2, ADOT would aggregate internal TAZs to single or multiple Census tracts, redesign existing TAZ boundaries for parks, and reassess external zones and RAZs. ADOT also would aggregate MPO TAZs into larger zones that account for the limitations of LBS data and cellular network coverage. When creating new TAZs, ADOT could work directly with the LBS data vendor to address data gaps (areas without cellular service). The researchers recommend that external zones contain only neighboring states, in order to limit the cost of acquiring LBS data.

Task 3 consists of specifying and then acquiring data with the attributes that ADOT wants, needs, and can afford. ADOT technical staff, the advisory committee, and the LBS data provider would discuss the attributes and options prior to acquiring the data. Based on current capabilities, the researchers recommend that LBS data attributes include a long-distance trip filter; I-I, E-I/I-E, and E-E trip matrices; average weekday travel; and the option to identify travel that involves an overnight stay. ADOT also may wish to consider acquiring LBS data for weekend travel and seasonal periods, such as the summer recreational season in cooler areas of the state.

Task 4 consists of assessing the options available for GPS O-D data, making a decision on what form of the data to purchase, and then acquiring it. ADOT would first review available GPS O-D providers and their products prior to determine the impact of recent advances. ADOT may choose to purchase only trip matrices if it merely needs zone-to-zone flows for the AZTDM. However, trip records with waypoints will allow the data to be repurposed and used for many other transportation studies and projects statewide. For average weekday travel, ADOT would benefit from acquiring data for four to five non-summer months to ensure that an adequate sample of long-distance trips, including school trips, are a part of the data set. If ADOT is uncertain on which option to acquire, it can obtain cost estimates for trip matrices and trip records with waypoints.

Task 5, the final portion of the implementation plan, consists of processing and storing the O-D data to prepare it for use in the AZTDM. The researchers recommend that ADOT plan for sufficient data storage and analytical software capacity to process large data sets. For GPS, the amount of processing and storage needed will depend on the form of the data acquired. The files for GPS trip matrix data are much smaller and require less processing than those for trip records with waypoints. LBS trip matrix files are similar in size to those of GPS; both are relatively small and represent a matrix of trips based on the electronic shape file of the TAZ structure. The researchers recommend working with ADOT technical services to identify data processing and storage options for waypoint data (if acquired), including in-house resources and cloud-based services such as Amazon Web Services or Dropbox.

Current practice by the researchers for this study is to process LBS and GPS data separately with the goal of ultimately conflating (or matching) the results to the AZTDM TAZ structure and the model's statewide roadway network. This process will result in a set of O-D matrices for expansion and weighting. The researchers recommend processing GPS trip coordinates to the highway network using a map matching process where each GPS point is associated with a road segment. ADOT would apply weights for passenger and truck travel in expansion to control for biases in the sample data. For external long-distance trips, IPF would be used to expand LBS and GPS data to traffic counts for passenger vehicles and trucks. For internal long-distance trips, LBS and GPS data would be expanded to seed matrices using ODME.

The use of new technologies to develop long-distance trip O-D information will improve the AZTDM's ability to represent current patterns. The practice of using these new sources of data to develop statewide long-distance O-D data is still evolving. How ADOT adopts and implements the recommendations of this research will be of interest and value across the country and influence the state of the practice.

REFERENCES

Acyclica. 2017. “Spotlight on Greeley Colorado.” Last Modified August 14, 2017. https://acyclica.com/relations/spotlight-cities/greeley-colorado.

Airsage. 2017. Press Release: AirSage Products and Insights Platform Enhanced with Multi-Source GPS and Location-Based Panel, September 13, 2017.

AirSage. 2018. Transportation Product Documentation, AS DOC 0013 - Version 1.0, May 22, 2018. https://airsage.com/trippatterns/.

Arizona Department of Transportation (ADOT). 2016b. “Data and Analysis-Average Annual Daily Traffic.” accessed July 26, 2016. https://www.azdot.gov/planning/DataandAnalysis.

Arizona Department of Transportation (ADOT). 2016a. “Planning-Data and Analysis-Travel Demand Modeling.” accessed May 23, 2016. https://www.azdot.gov/planning/DataandAnalysis/travel-demand-modeling.

Arizona Department of Transportation (ADOT). 2011a. Development of the Arizona Statewide Travel Demand Model: Phase 2 (AZTDM2), edited by Arizona Department of Transportation: Parsons Brinckerhoff.

Arizona Department of Transportation (ADOT). 2011b. Summary Report, Arizona Department of Transportation (ADOT), Travel Modeling Peer Review. In *ADOT Peer Review Panel Report*.

Beagan, Daniel, Michael Fischer, and Arun Kuppam. 2007. Quick Response Freight Manual II. Cambridge Systematics, Inc.

Bekhor, Shlomo, Yehoshua Cohen, and Charles Solomon. 2013. “Evaluating long-distance travel patterns in Israel by tracking cellular phone positions.” *Journal of Advanced Transportation* 47:435-446. doi: 10.1002/atr.170.

Bernardin, Jr. Vincent L., Jonathan Avner, Jeffery Short, Laurence Brown, Roy Nunnally, and Steve Smith. 2011. Using Large Sample GPS Data to Develop an Improved Truck Trip Table for the Indiana Statewide Model. In *Transportation Research Board (TRB) Innovations Paper*.

Bernardin, Jr. Vincent L., Nazneen Ferdous, Hadi Sadrasadat, and Chin-Cheng Chen. 2015b. Integration of the National Long Distance Passenger Model with the Tennessee Statewide Model and Calibration to AirSage Data. In *Travel Forecasting Resource*.

Bernardin, Jr. Vincent L., Steven Trevino, and Jeffery Short. 2015a. “Expanding Truck GPS-Based Passive Origin-Destination Data in Iowa and Tennessee.” *Sumbitted to the Transportation Research Board Conference August 1, 2014*.

Cambridge Systematics, Inc., Vanasse Hangen Brustin, Gallop Corporation, Chandra R. Bhat, LLC Shapirio Transportation Consulting, and PLLC Martin/Alexio/Bryson. 2012. Travel Demand Forecasting: Parameters and Techniques. In *NCHRP Report 716*: National Cooperative Highway Research Program.

Coladner, David, Ben Stabler, and Sujan Sikder. 2015. “Development of the Idaho STDM Trip Matrices Using Cell Phone OD Data.” 15th TRB Transportation Planning Applications Conference, Atlantic City, New Jersey.

Dunlap, Matthew, Zhibin Li, Kristian Henrickson, and Yinhai Wang. 2015. “Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit.” TRB 95th Annual Meeting Compendium of Papers, Washington, D.C.

Eisele, William L. 2015. Freight Fluidity Framework, Calculation Procedures and Analysis Findings. In *Technical Momorandum for the Maryland State Highway Administration from Texas A&M Transportation Institute through University of Maryland*: Texas A&M Transportation Institute.

Eisele, William L., Reuben M. Juster, Kaveh Farokhi Sadabadi, Tom Jacobs, and Subrat Mahapatra. 2015. Sumbitted for Review of Publication in *Transportaton Research Record* and Presented at Transportation Reserach Board's 95th Annual Meeting.

Erhardt, Greg. 2012. *Evaluating Travel Markets in a Statewide Model: An Arizona Case Study*. TRB Annual Meeting.

Federal Highway Administration (FHWA). 2015. Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework. In *Implmentation Report-Exploratory Advanced Research Program*.

Federal Highway Administration (FHWA). 2016a. “Freight Analysis Framework-FAF4 Data and Documentation.” accessed July 25, 2016. http://www.ops.fhwa.dot.gov/Freight/freight_analysis/faf/index.htm#faf4.

Federal Highway Administration (FHWA). 2016b. “The NHTS Add-on Program.” National Household Travel Survey, accessed September 28, 2016. http://nhts.ornl.gov/addOn.shtml.

Federal Highway Administration (FHWA). 2016c. “Welcome to the National Household Travel Survey.”

Hard, Ed, Byron Chigoy, and Praprut Songchitruksa. 2016a. *New Technology Sources for Origin-Destination (O-D) Data: Overview and Lessons Learned*. May 12, 2016.

Hard, Ed, Byron Chigoy, Praprut Songchitruksa, Steve Farnsworth, Darrell Borchardt, and Lisa Green. 2016b. Synopsis of New/Emerging Methods and Technologies to Collect Origin-Destination (O-D) Data. Edited by Federal Highway Administration TMIP: Texas A&M Transportation Institute.

Horowitz, Alan. 2006. Statewide Travel Forecasting Models-*A Synthesis of Highway Practice*. In *NCHRP Synthesis 358*: National Cooperative Highway Research Program.

Horowitz, Alan J. 2008. White Paper: Statewide Travel Demand Forecasting. In *Prepared as Part of NCHRP Project 08-36, Task 76c*, edited by American Association of State Highway and Transportation Officials (AASHTO).

INRIX. August 16, 2016. *Next-Gen Origin Destination Analysis Enabling Faster, More Informed Decision Making & Planning*.

Jeng, Shin-Ting. 2007. “Real-time Vehicle Reidentification System for Freeway Performance Measurements.” Doctor of Philosophy, Civil Engineering, University of California, Invine.

Kriger, David, Matthew McCumber, Allison Clavelle, Becky Gan, and Tavia Chow. 2011. Freight Transportation Surveys-A Synthesis of Highway Practice. In *NCHRP Synthesis 410*: National Cooperative Highway Research Program.

Kroman, David. September 8, 2015. "Seattle Installs New System to Track Individual Drivers." Last Modified August 11, 2017. http://crosscut.com/2015/09/seattles-new-technology-tracks-how-we-drive/.

Kuppam, Arun, Jason Lemp, Dan Beagan, Vladimir Livshits, Lavanya Vallabhaneni, and Sreevasta Nippani. 2013. "Development of a Tour-Based Truck Travel Demand Model using Truck GPS Data." TRB 2014 Annual Meeting, Washington, D.C.

Lee, Richard J., Ipek N. Sener, and James A. Mullins III. 2016. "An evaluation of emerging data collection technologies for travel demand modeling: from research to practice." *The International Journal of Transportation Research* XX (X):1-13. doi: 10.1080/19427867.2015.1106787.

Lu, Yijing and Lei Zhang. 2015. "Imputing trip purposes for long-distance travel." *Transportation* 42:581-595. doi: 10.1007/s11116-015-9595-0.

Mercedes-Benz USA. 2012. "Difference between Bluetooth and WiFi." Last Modified October 17, 2016. http://www.engineersgarage.com/contribution/difference-between-bluetooth-and-wifi.

Mobile Marketing Association. 2015. "Demystifying Location Data Accuracy." Last modified November 23, 2015. https://www.mmaglobal.com/webinars/demystifying-location-data-accuracy .

Moeckel, Rolf and Rick Donnelly. 2011. "Nationwide Estimate of Long-Distance Travel (NELDT)-Generating External Trips for Local Travel Demand Models." Paper submitted for presentation at the Annual meeting of the Transportation Research Forum 2011.

Patman, Krishna C., David B. Roden, Weihao Yin, Li-Yang Feng, Sayeed Mallick, and Scott B. Smith. 2016. A Practical Guide on DTA Model Applications for Regional Planning. In *TMIP Report*: AECOM.

Paul, Sanjay and Andy Murray. 2015. " Statistical Validations for the Evaluation of Bluetooth/Wi-Wi-Fi Technology on Urban Freeway Applications." 22nd ITS Arizona Conference.

Predictable Designs LLC. 2016. "Bluetooth or WiFi - Which is Best for Your New Product Design?", Last Modified October 17, 2016. http://predictabledesigns.com/what-type-of-wireless-is-right-for-your-product-bluetooth-wifi/.

Schewel, Laura. 2017. "Location-Based Services Data Beats Cellular of Spatial Precision." Last modified May 25, 2017. https://www.streetlightdata.com/cellular-data-vs.-location-based-services-data-spatial-precision.

Schiffer, Robert G. 2012. Long-Distance and Rural Travel Transferable Parameters for Statewide Travel Forecasting Models. In *NCHRP Report 735*: National Cooperative Highway Research Program.

Stabler, Ben. 2014. *Development of the Idaho Statewide Travel Demand Model Trip Matrices Using Cell Phone OD Data and Origin Destination Matrix Estimation*.

Struyk, Tara. April 28, 2017. "What is the difference between Bluetooth and Wi-Fi?", Last Modified October 20, 2017. https://www.techopedia.com/2/27881/networks/wireless/what-is-the-difference-between-bluetooth-and-wi-fi.

Transportation Model Improvement Program [TMIP]. 2014. Arizona Department of Transportation (ADOT) Travel Model Peer Review Report.

Transportation Research Board. 2008. Forecasting Statewide Freight Toolkit. In *NCHRP Report 606*: National Cooperative Highway Research Program.

Travel Forecasting Resource. 2016. "Statewide models." accessed July 28, 2016. http://tfresource.org/Category:Statewide_models.

Wang, Ming-Heng, Steven D. Schrock, Nate Vander Broek, and Thomas Mulinazzi. 2012. The Use of Cell Phone Network Data in Traffic Data Collection and Long-Haul Truckshed (Geographic Extent) Tracking. Mid-America Transportation Center.

Yankus, David. December 12, 2016. "Wireless Communication System Helps Keep Maricopa Traffic Flowing." Last Modified August 3, 2017. http://www.pinalcentral.com/maricopa_monitor/wireless-communication-system-helps-keep-maricopa-traffic-flowing/article_01ce75a5-bb1c-51fe-8650-b5c856ca18a1.html.

Zhang, Wang, Arun Kuppam, Vladimir Livshits, and Bill King. 2015. "Understanding Cellular-based Travel Data Experience from Phoenix Metropolitan Region." 15th TRB National Transportation Planning Applications Conference, Atlantic City, New Jersey.

APPENDIX

Texas A&M Transportation Institute
3135 TAMU
College Station, TX 77843-3135

979-845-3326
Fax: 979-845-7548
http://tti.tamu.edu

MEMORANDUM

TO: Ethan Rauch, Senior Research Project Manager, ADOT

FROM: Ed Hard, TTI Research Scientist
Byron Chigoy, TTI Associate Research Scientist

RE: Explanation: Lack of Relevancy of Task 4 in SPR-744, Optimizing Technology for Collecting Long-Distance Travel Data

DATE: February 28, 2018

In the July 6, 2017 TAC meeting the committee concluded that the work as described in Task 4, *Evaluate and Recommend Software*, was not relevant or needed due to advances in technologies and data collection methods that had taken place since the task was proposed. In light of this, the committee advised TTI to prepare a Technical Memorandum for Task 4 (in lieu a chapter write-up) that explained and clarified why the task was not applicable to the project. This explanation is provided in the paragraphs below.

The scope of Task 4 (in ADOT's Task Assignment MPD 025-16) was to 'Evaluate and Recommend Software' suitable for generating O-D data from the 'devices' that would be recommended in Task 3 (upon its completion). The task called for assessing the compatibility of software packages with ADOT's 'existing processing systems'. If no suitable software packages were found, the scope called for laying out the functions that a new software package would need to construct a statewide O-D database from the recommended count and device data.

The work set forth in Task 4 is not applicable or needed in the project for the following reasons:

- It assumes real-time 'point sensor' data collection from 'devices' would be used, not ubiquitous and pre-processed O-D data from third party data aggregators such as INRIX and Airsage;
- It assumes one technology would be used, not the combination of technologies that will be needed to develop O-D matrices; and,
- It is believed that ADOT's 'existing processing systems' may be referring to proprietary software that accompanied a prior purchase of Wi-Fi data and Wi-Fi is not a technology that is recommended for long-distance O-D data collection.

As written, the task describes software needed to assimilate and process data collected from a network of 'point sensor' devices, such as Wi-Fi or Bluetooth, located along highways throughout Arizona. The software and related approach to processing the data is not suitable for aggregated 'crowd sourced' data that are now available. Such software will not be needed for third-party cell and GPS data, which are recommended as the primary sources of O-D data for ADOT's statewide model.

Researchers believe that the task was developed assuming that a statewide network of data would be *collected* using point sensor Wi-Fi or Bluetooth devices and that software would need to be developed to process and analyze these data into O-D matrices. When it was written, it was not envisioned that large-scale, pre-processed cell and GPS data could/would be *purchased* on a statewide basis. Since TTI's recommended approach is not based on point sensor data or a single technology or data source, evaluating and recommending software as described in Task 4 is not needed.

Texas A&M Transportation Institute
3135 TAMU
College Station, TX 77843-3135

979-845-3326
Fax: 979-845-7548
http://tti.tamu.edu

MEMORANDUM

TO: Keith Killough, AICP, ADOT Director of Transportation Systems Analysis, ADOT

FROM: Ed Hard, TTI Research Scientist

RE: Review of Inductive Loop Signature Detection (ILSD) for Long Distance O-D Data Collection

DATE: March 28, 2017

One of ADOT's comments in review of TTI's Task 3 Memo was for TTI to look into the use of Inductive Loop Signature Detection (ILSD) as a potential option for collecting long distance O-D data. We have conducted this research and have concluded that, at the current time, the accuracy of vehicle re-identification using ILSD is not consistent or high enough for it to be a plausible source for long distance O-D data collection. This conclusion was reached based on review of numerous TRB papers and scholarly articles and related presentations at TRB and NATMEC, along with interviews with senior TTI researchers with expertise in advanced inductive loop technology. In light of our finding, it is TTI's recommendation that ILSD not be included as a technology in the Task 3 Memo that is discussed as one of the options/methods for long distance O-D data collection.

Conventional inductive loop detectors (ILDs) detect the presence of a vehicle and are typically used for vehicle counts, speeds, and travel time/congestion monitoring. ILDs can distinguish between passenger vehicles and heavy trucks, but cannot further classify vehicle or truck type. ILDs with inductive signature technology (or ILSD) have higher resolution sampling that produces a waveform signature that can be used to distinguish between many different heavy truck types and classes.

Different vehicle types have different waveform signatures and the same vehicle types have similar signatures. The ability to accurately classify and re-identify vehicles is dependent on the uniqueness of their signature. Heavy trucks are more easily classified than passenger vehicles due to their greater variations in size, shape, length, and number of axles that create distinct and dissimilar signatures. However, even with the greater differences in their characteristics, re-identification of trucks between different sites using ILSD is currently not accurate enough to make it a plausible option for long distance collection of O-D data. A 2015 study on a .66 mile stretch of freeway in California using ILSD integrated with a re-identification algorithm found the re-identification rate for trucks to be 65.6 percent. However, this result is based on a data collection sites located only .66 miles apart. The results would be lower for long distance O-D data collection where sites would be greater than 50 miles apart. The re-identification rate for passenger vehicles would be much lower than trucks since their signatures are too similar and non-distinct.

Compared to other more tested and vetted technologies, ILSD is not a feasible option of O-D data collection at this time. However, it is suitable for collecting enhanced truck classification data based on body configuration. While the technology cannot be used at for exact vehicle re-identification, it can be used for tracking and monitoring truck patterns by industry specific body types such as domestic versus port containers, logging trucks, fuel trucks, enclosed van trailers, and others. Research also suggest that the technology can also be used for re-calibrating WIM locations.