CRC Report No. E-131

STUDY OF CAPABILITIES AND LIMITATIONS OF VEHICLE TELEMATICS DATA FOR EMISSION INVENTORIES

Final Report

October 2020

COORDINATING RESEARCH COUNCIL, INC.
5755 NORTH POINT PARKWAY • SUITE 265 • ALPHARETTA, GA 30022

The Coordinating Research Council, Inc. (CRC) is a non-profit corporation supported by the petroleum and automotive equipment industries. CRC operates through the committees made up of technical experts from industry and government who voluntarily participate. The four main areas of research within CRC are: air pollution (atmospheric and engineering studies); aviation fuels, lubricants, and equipment performance; heavy-duty vehicle fuels, lubricants, and equipment performance (e.g., diesel trucks); and light-duty vehicle fuels, lubricants, and equipment performance (e.g., passenger cars). CRC's function is to provide the mechanism for joint research conducted by the two industries that will help in determining the optimum combination of petroleum products and automotive equipment. CRC's work is limited to research that is mutually beneficial to the two industries involved. The final results of the research conducted by, or under the auspices of, CRC are available to the public.

STUDY OF CAPABILITIES AND LIMITATIONS OF VEHICLE TELEMATICS DATA FOR EMISSION INVENTORIES

CRC PROJECT E-131

FINAL REPORT

Prepared for:

Coordinating Research Council Emissions Committee

Prepared by:

John Koupal
Allison DenBleyker
Gopi Manne
Maia Hill Batista
Thomas Schmitt
Sandeep Kishan
Colby Richardson
Eastern Research Group, Inc.

October 30, 2020

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ACRONYMS

CARB – California Air Resources Board
CDOT – Colorado Department of Transportation
CHPHE – Colorado Department of Public Health and Environment
COVID – Corona Virus Disease (COVID-19)
DB4IoT – Database for Internet of Things
EMFAC – CARB Emission Factor Model
EPA – U.S. Environmental Protection Agency
EV – Electric Vehicle
eVMT – Electric Vehicle Miles Travelled
GPS – Global Positioning System
HD – Heavy Duty Vehicle
H/EV – Hybrid Electric Vehicle
ICE – Internal Combustion Engine
I/M – Inspection / Maintenance
LBS – Location Based Services
LD – Light Duty Vehicle
MOVES – Motor Vehicle Emissions Simulator
NAS – National Academy of Sciences
NCHRP – National Cooperative Highway Research Program
NEI – National Emissions Inventory
NFL – National Football League
OEM – Original Equipment Manufacturer
OBD – On-Board Diagnostics
PHEV – Plug-in Hybrid Electric Vehicle
SOC – State of Charge
VIN – Vehicle Identification Number
VMT – Vehicle Miles Travelled

EXECUTIVE SUMMARY

CRC initiated project E-131, "Studying the Capabilities and Limitations of Vehicle Telematics Data for Emission Inventories", to shed light on the application of vehicle telematics data towards compiling emission inventories. Under contract to CRC, Eastern Research Group, Inc (ERG) evaluated the current state of the telematics field with respect to emission inventory development; identified capabilities and limitations of the data for emission modelers; and examined how telematics data could better align with the needs of vehicle emission inventory models. ERG's prior work in characterizing telematics defined two basic groups of data: location-based (e.g. GPS pings from connected vehicles and mobile phones), and engine-based (e.g. OBD data loggers). This bifurcation was carried forward in E-131 to help evaluate the current telematics market and consider how data could be applied to improve vehicle emission inventories. ERG first conducted a market survey of telematics data use in emission models by regulatory agencies and the current data market from the perspective of inventory development. In the second phase of the project, ERG then conducted case studies of different emission inventory applications with data from three different vendors vetted and selected during the first phase.

ERG's market survey included a literature review, a survey of regulatory agencies, and direct communication with several telematics firms. The market survey confirmed a growth in the application of telematics, in particular location-based service (LBS) data culled from cell phones. Our literature review identified 40 studies that used telematics data in emissions or transportation-related studies. The majority of studies used LBS, and included an effort to estimate total vehicle miles travelled. EPA and CARB staff involved with vehicle emissions inventory modeling were surveyed on current and planned uses of telematics in the MOVES and EMFAC vehicle emissions models, as well as their telematics "wish list". EPA specifically expressed interest in comparing different telematics sources and information on representativeness of telematics samples, while CARB provided specific data fields of interest. This input from EPA and CARB, in conjunction with our review of prior studies and current data offerings, provided a roadmap for current and potential uses of telematics in emission inventories, summarized in Table ES-1.

Table ES-1. Current & Potential Telematics Use in U.S. Vehicle Emission Models

	Engine-Based	**Location-Based**
Current Uses (including upcoming model releases)	Trip starts/ends/soaks Idle time HD mileage accrual HD malfunction rate	Avg speed distributions VMT temporal distributions

	Engine-Based	Location-Based
Potential Uses	OBD code frequency OBD code duration Engine temperature Catalyst temperature Drive cycle H/EV: engine-on, SOC Isolating non-traffic idle	Total VMT Trip starts Spatial allocation of activity “Hot-spot” activity

The market survey identified several candidate sources for location-based and engine-based data on personal vehicles, the focus of this study. Out of a desire to evaluate and compare multiple sources, including at least one of each type, ERG purchased access to telematics data from three different vendors (StreetLight Data, Moonshadow Mobile, Otonomo), and assessed their capabilities and limitations with regard to generating inputs for emission inventories via several case studies conducted in the Denver metro area. The case studies were informed by the relative strengths of each dataset in terms of spatial and temporal detail, available data fields, and data resolution. Case studies with StreetLight Data’s InSight data platform (accessed at the lowest cost tier) were conducted to evaluate the feasibility of replicating county-level vehicle activity inputs for the 2017 National Emissions Inventory (NEI), and evaluating the extent of travel from Non-I/M areas into Denver’s I/M region. During the project another case study emerged for StreetLight, to estimate the VMT impact of COVID-related shutdown and gradual re-open through Spring 2020. Case studies with Moonshadow’s DB4IoT data platform, populated with INRIX data, were conducted to estimate vehicle activity near Denver’s football stadium following an NFL game for a project scale “hot spot” analysis; and to estimate total VMT based on comparison of INRIX data to traffic counter observations. Otonomo’s engine-based data were analyzed to define trip and idle events , and to evaluate the distribution of fuel fill levels for vehicles in their sample.

From the perspective of emissions inventory development, the case studies showed that the telematics datasets were most influential with respect to the ability to estimate VMT, and tracking large changes in VMT during the COVID shutdown and initial re-opening phase. Trip data from StreetLight were used to estimate total VMT within the 10-county Denver metro area in 2017, including trips that passed through the area, and found to be 34 percent higher than estimates used by EPA in the 2017 NEI. As a check on the StreetLight algorithm for scaling up telematics trip counts to total VMT, ERG developed an independent estimate of total VMT by scaling up Moonshadow/INRIX data to Colorado DOT traffic counter data, coming within two percent of StreetLight’s estimate for the same time period. Since VMT is the most important factor in estimating total emissions from vehicles, our NEI case study results suggest EPA’s NEI emission estimates for the Denver area are significantly underestimated. Because the case study only focused on Denver, this finding can’t assess whether NEI VMT is underestimated more broadly, but its significant merits follow-up study Aside from estimating total VMT, the ability to rapidly analyze the relative impacts of COVID shutdown and initial re-opening during

Spring 2020 relative to prior years (Figure ES-1) highlights the capability of telematics to readily provide data for atypical events.

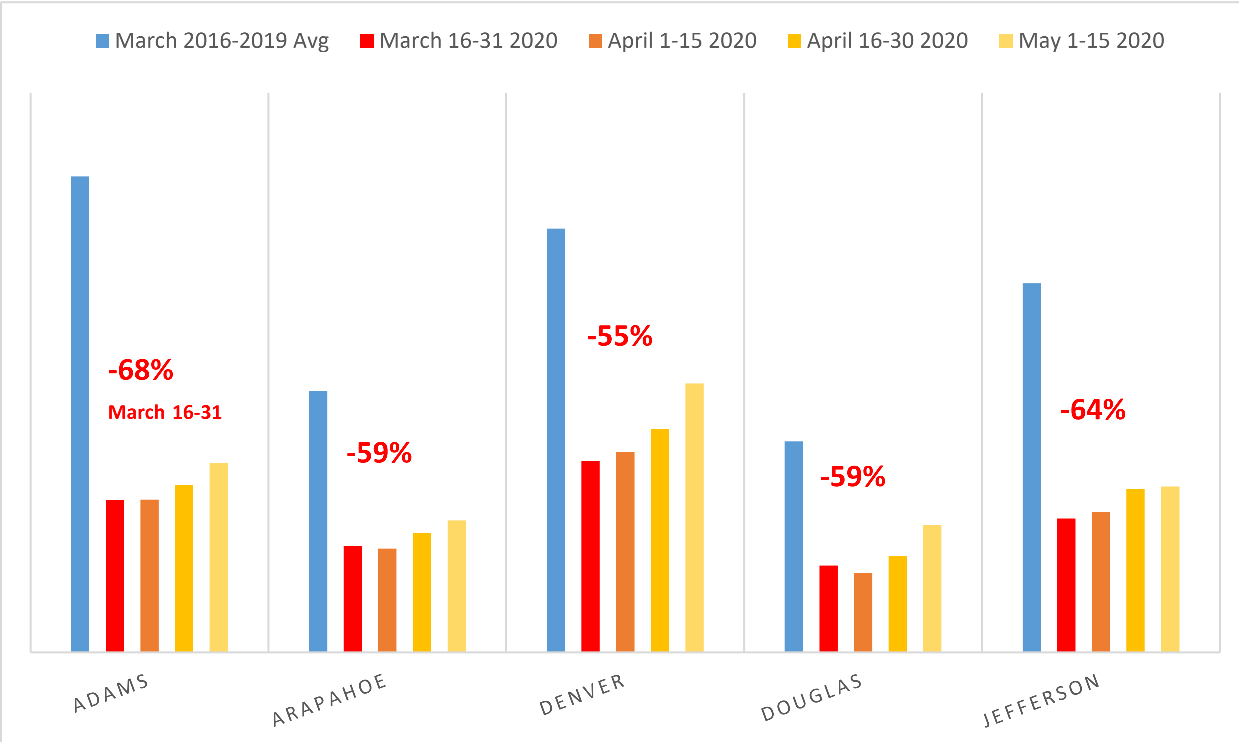

Figure ES-1. Change in Daily VMT during COVID Shutdown Denver-area Counties

The case studies confirmed that the location-based telematics can serve a growing range of emission inventory use cases, and help improve emission inventory accuracy. Key capabilities include more accurate spatial and temporal distribution of start emissions; estimation of total VMT; and the ability to pinpoint vehicle activity at specific times and locations, which will facilitate more accurate project scale “hot spot” analysis and event-specific emissions. Though not an intended case study at the outset of the project, the ability of telematics to rapidly track changes in travel patterns from COVID shutdown and gradual re-opening during Spring 2020 proved one of the most illustrative benefits of telematics data. Key limitations of location-based telematics stem from lack of specific vehicle information, resulting in difficulty assessing the representativeness of activity data. Other identified limitations are cost, and lack of focus from the telematics firms on providing data direction for emission inventory models.

On the whole, we conclude that the capability of telematics continues to advance for the purposes of emission inventory modeling. The growth in location-based data will only continue to improve capabilities for spatial and temporal activity distributions that prior studies have focused on, while the emerging capacity to estimate total activity is a significant advancement. Identified limitations keep the application of telematics in check, however. It is our hope that telematics firms will begin to cater more directly to the needs of emission inventory modelers, to better harness the enormous potential of these data for refining vehicle emission inventory estimates.

1. INTRODUCTION

This report documents the method and findings for CRC project E-131, "Studying the Capabilities and Limitations of Vehicle Telematics Data for Emission Inventories". The project objectives as defined by CRC were to provide a foundation for education on telematics databases, and provide an opportunity to improve collaboration between CRC and regulatory agencies on this topic. The project was divided into two tasks; Task 1 conducted a broad review of telematics data sources, focused on private firms that compile and sell access to telematics data. Based on this review, under Task 2 ERG purchased access to three telematics datasets and conducted case studies with each to assess capabilities and limitations of different telematics datasets for common emission inventory applications, such as the National Emissions Inventory (NEI) and project scale "hot spot" analysis.

This project built on several analyses of telematics data that have been undertaken to support default inputs for EPA's emission inventory model MOVES, California's inventory model EMFAC, and default data used for the NEI, including:

- CRC Project A-106, which evaluated the potential for telematics data to improve inventories and spatial allocation of vehicle start emissions (CRC 2017). This project included an evaluation of multiple types of telematics data that are directly relevant to this E-131 project.
- CRC Project A-100, the first large-scale use of vehicle telematics data to produce county-level default inputs for MOVES, for average speed and temporal VMT distributions (DenBleyker 2019).
- NCHRP Project 8-101, leading an effort to pilot the use of telematics data to develop MOVES start-related inputs for trucks (NAS 2019).
- ARB's use of telematics data to improve mileage accrual and, for hybrid vehicles, miles travelled on electric power (eVMT) rates in EMFAC (CARB 2020).
- EPA's use of commercial telematics data in order to develop MOVES inputs related to vehicle speeds and trip patterns, and evaluative OBD MIL-On behavior (EPA 2019).

Taken together, these prior projects have established that telematics data are not uniform , but rather encompass a myriad of data types (e.g. location data from connected vehicles, location data from mobile phones, or engine data from specific vehicles) compiled and bundled via multiple sources, primarily private firms. The landscape of telematics data continues to change, as data sources, firms, and privacy constraints have evolved rapidly in the last five years. The fact that no firm is focused on providing data for MOVES or EMFAC continues to pose a challenge for emission inventory developers who want to tap into very large sample sizes of telematics data. As presented in the study, while the capabilities of the data are indeed promising, there are still many limitations that need to be considered in using telematics data for emissions models.

2. MARKET SURVEY OF VEHICLE TELEMATICS DATA

For the first phase of the project ERG conducted market research of telematics firms and identified a short list of candidate data to evaluate for this project. ERG conducted surveys of regulatory agencies to understand current applications of telematics, concerns, and “wish list” priorities with respect to their emissions inventory models. To understand broader use of telematics in transportation field and expand the list of candidate firms beyond those used previously, a literature and online search of telematics vendors was conducted. Interviews of select firms were then conducted and quote requested based on criteria needed for case studies presented in Section 3. From this a short list of telematics data was identified to purchase.

2.a Literature Search

To identify the use of telematics in emissions- and transportation-related studies beyond the studies previously listed, we conducted a literature search of academic journals (via Science Direct), proceedings of recent Transportation Research Board (TRB) annual meetings, and Society of Automotive Engineers (SAE International) technical papers. Using keywords “telematics”, “big data”, “location-based services”, and the names of individual firms, the search identified 40 studies that used telematics data in emissions or transportation-related studies. The majority of studies used location-based service (LBS) data (i.e. GPS “pings” from connected vehicles or phones converted to vehicle trips) for transportation origin-destination analysis, though these studies aren’t directly relevant to emission inventory development. The most common firms providing data for these studies were INRIX, followed by Airsage.

The most relevant study from this group estimated total VMT, the most fundamental activity input for emission inventories, using LBS data (Fan et al. 2019). The study used INRIX GPS waypoint data used to develop vehicle counts by link over entire state of Maryland. The data correlates well with traditional AADT from traffic counts on select roads, though at a fraction of total volume. The study found that the INRIX data accounts for an estimated 2-10 percent of vehicles on the road, meaning that even large-scale telematics datasets require significant upscaling to derive total VMT.

2.b Agency Surveys

To address CRC’s stated project objective of improving collaboration between CRC and regulatory agencies with respect to telematics, U.S. EPA and CARB modeling groups were asked to describe their plans for telematics in upcoming releases of MOVES / EMFAC; and to describe a telematics “wish list” to help inform the data evaluation and purchase for E-131. A summary of responses is shown in Table 1. One “wish list” item was to compare different telematic sources by data type (engine and location), and vendor; this informed our design of the case study around three datasets gathered at the same time/location.

Table 1. Input from Regulatory Agencies on Telematics in Vehicle Emission Inventories

	Current Use of Telematics (including upcoming model releases)	"Wish List" for Telematics Datasets
U.S. EPA	• Soak distribution (Verizon) • Trip starts/day & distribution (Verizon) • Idle time (Verizon) • National-level default average speed distribution (TomTom)	• Data is representative (or stratified to allow scaling) of national activity in terms of geographic variation, vocational variation, vehicle age, technology, fuel type etc. • Represents inactive vehicles • Can distinguish different vocation of vehicles (Uber vs. personal, parcel delivery truck vs. utility truck) • Can distinguish EVs from conventional ICEs • Provides the location (county, road type) of driving (not just the zip code of residence) • Identifies grade, class (e.g. LD, HD), speed and acceleration • Telematics data for off-road sources

CARB	• HD mileage accrual (Geotab)
	• HD malfunction rate (Geotab)
	• PHEV eVMT and activity (OEM-provided).
	• Rideshare trip activity (Uber/Lyft)
		• Fill current data gaps such as mileage accrual rates for LD vehicles younger than eight years and spatial distribution of truck travel.
		• Desires trip level data with...
		  ○ VIN
		  ○ Trip ID
		  ○ VMT (for PEVs also provide eVMT)
		  ○ Date/time for trips start and end
		  ○ Lat/long for trips start and end
		  ○ Fuel consumption (gallons)
		  ○ Electricity consumption (kWhr)
		  ○ MIL status along with all active SPNs (preferably emission related)
		  ○ Percent VMT < 25 mph / 25 – 50 mph / > 50 mph
		  ○ Percent time idling (speed <1 mph)
		  ○ For trucks, average bhp-hr/mi for that trip
		  ○ Vehicle odometer

A separate survey was disseminated via the informal MOVES Multi-jurisdictional Organization (MJO) workgroup, comprised mostly of state and local environmental agencies across the U.S. state and local model agencies were asked whether they use telematics for NEI submissions; and if yes, identify data sources and specific MOVES inputs populated. Of the 16 respondents, only one local agency (Pima, AZ) reported obtaining and processing new telematics data (as opposed to existing datasets such as CRC A-100), for average speed distribution.

Based on the agency survey and literature review, Table 2 provides a summary of telematics data used in inventory models, and potential additional applications.

Table 2. Current and Potential Uses of Telematics Data in Emission Inventories

	Engine-Based	**Location-Based**
Current Uses (including upcoming model releases)	Trip starts/ends/soaks Idle time HD mileage accrual HD malfunction rate	Avg speed distributions VMT temporal distributions
Potential Uses	OBD code frequency OBD code duration Engine temperature Catalyst temperature Drive cycle H/EV: engine-on, SOC Isolating non-traffic idle	Total VMT Trip starts Spatial allocation of activity “Hot-spot” activity

2.c Survey of Telematics Firms

Drawing from prior studies, literature search, and Agency input, a list of candidate sources for telematics data purchase was identified, as follows:

- Verizon Connect (verizonconnect.com)
- Moonshadow Mobile (moonshadowmobile.com)
- Geotab (geotab.com)
- StreetLight Data (streetlightdata.com)
- Otonomo (otonomo.io)
- INRIX (inrix.com)
- IMS (ims.tech)

These firms were engaged via phone or email to understand the data they compiled, data coverage, and available fields (an eighth firm, Wejo, was contacted through their website but did not respond so were not included in direct interviews). Based on responses interviews and/or email questions, a top-level summary of each firm's data source, capability and coverage is shown in Table 3.

Table 3. Summary of Telematic Firm Surveys

Firm	Data Overview	Personal Veh. Sample
IMS	OBD data from insurance safe driver program	~700k-1M U.S.-wide
Verizon Connect	OBD data from insurance safe driver program; would not be able to replicate prior EPA data buy per new privacy policies.	~800k U.S.-wide
Moonshadow Mobile	Database platform for one-stop access & comparison of different datasets (DB4IoT). Can handle LBS or OBD. Offers pre-loaded data from multiple vendors (e.g. INRIX, Wejo, UnaCast, X-Mode Social).	Millions U.S.-wide
Geotab	OBD via Geotab GO, fleet-focused.	U.S.-wide fleet vehicles (e.g. rental, municipal, commercial).
StreetLight Data	LBS has included OEM, but mobile devices now majority of sample. Via StreetLight Insight web access platform, 3 tiers of access and price points, varying by level of data aggregation (Essentials, Advanced Analytics, Multi-Mode).	~1.5 billion trips per month
Otonomo	OBD from several OEM + rental. Up to 250 data fields, varies by OEM, make, even trim. Has data platform, can accommodate custom data draws.	18 million worldwide, per website. U.S.-specific not disclosed.
INRIX	LBS from several OEM/ fleets /mobile devices. Has developed analytical packages for specific use cases, but prefers to vend via 3rd party handlers.	Millions U.S.-wide; tenfold increase in sample size in 2019.

In conducting surveys with telematics firms, the following general observations where made:

- There is no one-size fits all telematics dataset to serve emissions inventory modeling.
- There is little commonality across datasets; each brings something unique.

- Many vendors have a web-based interface, which is easier and most cost effective to access vs. working with vendor on custom draw.
- Engine-based data is limited – nearly all sources quoted are for LBS from mobile phones and/or connected vehicles.
- Unlike earlier studies, the emerging business model is purchasing limited time access to a firm's data, from which aggregate statistics can be derived, vs. purchasing data itself.

In discussion with firms listed above, three removed themselves from consideration as a data source for Task 2: Verizon (internal privacy restrictions); Geotab (fleet-oriented data not a good fit for the project); and INRIX (stated preference was to supply through other vendors). This left StreetLight Data, Otonomo, Moonshadow, and IMS as potential candidates for data purchase.

2.d Data Quotes

The original scope and budget of E-131 focused on the purchase of one dataset for Task 2 analysis. However, give the variety of datasets identified, Task 2 was adapted to accommodate the Table 1 "wish list" item to compare multiple datasets, including LBS and OBD. To narrow down the candidate sources from Table 3 and fit multiple datasets within the existing budget, a case study was devised to allow comparison of datasets in a common location and time. The Denver Metropolitan Statistical Area (MSA) was chosen (counties of Arapahoe, Jefferson Adams, Douglas, Broomfield, Elbert, Park, Clear Creek, and Gilpin), for calendar year 2017 or more recent year if data were cheaper and/more abundant. A price quote was requested from the two firms providing LBS only (StreetLight, Moonshadow), and two firms providing some degree of engine data (Otonomo, IMS). In addition to cost, firms were asked to provide details on temporal coverage, sample size, available vehicle information, data "ping" rate, data fields, geographic coverage, how pass-through trips were accounted for, and access protocol. Reponses were received from Moonshadow, StreetLight, and Otonomo; IMS ultimately did not provide a quote. Underscoring the different focus of firms, Moonshadow's quotes were based on the source of telematics data, since the Db4IoT platform can work with data from multiple sources. Multiple follow-up calls with held with each firm to discuss details on data coverage, availability and price options. Quote results are summarized qualitatively in Table 4.

Table 4. Quotes Received for Telematics Data

Company / Platform	Data Source	Duration	Relative cost
Moonshadow DB4IoT	INRIX LBS (OEM/mobile)	1 month	\$\$\$\$\$
	Wejo LBS (OEM)	1 month	\$\$\$
	UnaCast LBS (mobile)	2 months	\$\$
	X-Mode Social LBS (mobile)	2 months	\$\$
	ERG-provided 3rd party data		\$
StreetLight InSight - “Essentials”	LBS (mobile)	4 years	\$
Otonomo	OBD	12 months	\$
IMS	OBD	Did not respond to request for quote	

In order to meet project objectives within budget of data purchase, data were selected for purchase from each vendor but at lower cost points, causing some restriction of data. StreetLight InSight - Essentials (lowest cost access point); Moonshadow Db4IoT populated with Wejo data; and Otonomo were selected. Wejo was unable to secure approval from OEM data sources in time however, so Moonshadow substituted INRIX data at comparable cost. These three datasets formed the basis of Task 2 assessment and case studies discussed in Section 3.

3. DATA PURCHASE, EVALUATION, AND ANALYSIS PLAN

Under Task 2, ERG purchased access to three telematics datasets (StreetLight InSight - Essentials, INRIX via Moonshadow DB4IoT, and Otonomo), evaluated the data, and conducted several proof-of-concept analyses for generating activity inputs for emissions modeling. Original goals of the case studies included:

- Compare different telematics datasets in same time & place;
- Develop MOVES inputs (e.g. speed, VMT);
- Quantify pass-through and ride-share activity;
- Compare to independent data – e.g. traffic counts;
- Evaluate interfaces for telematics data platforms.

These objectives were discussed with vendors in the process of developing quotes. Because vendors do not have familiarity with the needs of emission inventory modelers, or with MOVES in particular, recalibration of case study objectives was required after gaining access to vendor platforms and initial direct work with the datasets to assess their capabilities and limitations. Each vendors' business model dictated that the purchase was for access the vendor's data platform, rather than a direct purchase and ownership of trip data.

3.a Evaluation of Datasets

Upon gaining access to data, ERG was able to better evaluate each vendor's data with respect to suitability for emission inventory modeling. This evaluation turned up some data limitations for inventory modeling that weren't apparent in initial discussion with vendors. For example, INRIX data in Moonshadow DB4IoT did not include scaled estimates of total volume. StreetLight InSight - Essentials reports aggregate trip metrics only, which was not useful for some MOVES inputs, notably average speed distribution.[1] Otonomo data was limited with respect to available engine-based data; while providing fuel fill level and odometer, it did not include data that would render it more useful for inventory modeling, such as key-on / key-off times or OBD MIL codes. More broadly, Otonomo confirmed that the fleet was all rental vehicles, making the data unrepresentative of personal vehicle travel. The data are useful for evaluating rental fleets.

These issues affected the scope of the case studies planned for the data. For instance, though we had hoped Otonomo data could provide robust engine-based OBD data to compare directly to LBS, the lack of key-on/off times, speeds, or OBD codes did not allow for this comparison. Taking into account the limitations listed above, analysis plans were modified accordingly, with final scope shown in Table 5.

[1] it is important to note that some limitations identified with "Essentials" are addressed at higher price points in StreetLight's cost structure. For example, an "Advanced Analytics" level adds capability to analyze individual road segments, specific events, and commercial vehicles; a "Multimode" level adds bicycle and pedestrian metrics.

Table 5. Final Case Study Overview

Company	Time Period	Useable Data Fields (select)	MOVES Case Study	Other analyses
Moonshadow w/ INRIX (DB4IoT)	October 2019	Trip Distance Time Speed Origin Destination Road	**Project scale analysis** Denver Broncos game traffic	Comparison
StreetLight ("Essentials" web-based platform)	March 2016 – March 2020 +	(Aggregate Data) Distance Trip Speed Origin Destination Total volume	**NEI 2017 inputs** **Speed** (avgSpeedDistribution) **VMT** (sourceTypeDayVMT, sourceTypeYearVMT) **Temporal distribution** (monthVMTFraction, dayVMTFraction, hourVMTFraction) **Starts** (starts, startsPerDay, startsHourFraction)	Comparison COVID 19 shutdown impact non I/M vehicles in I/M area
Otonomo (web-based platform)	April 2019 – March 2020	Time/Date Latitude Longitude Car Make Fuel Level Odometer	**Ozone day Sensitivity analysis:** Fuel level distribution SampleVehicleTrip	Comparison

3.b Comparison of Datasets

To compare the three datasets, a distribution of trips for an October 2019 day (surrogate for VMT) was generated for all three datasets. Trips on Tuesday, October 15th 2019 across the Denver MSA was the basis for comparison; for StreetLight Insight - Essentials, specific day data were not available – the finest level of resolution available was the average across five Tuesdays in October 2019. Table 6 shows a direct comparison of metrics from the three platforms; the raw INRIX sample size is largest, with over twice the raw trips as StreetLight Insight – Essentials, and over ten times the trips as Otonomo.

Table 6. Raw Trip Comparison for Denver MSA

Dataset	Metric	October 15 Estimated Raw Trips
StreetLight Insight - Essentials	Unique device trip count – average October 2019 Tuesday (specific date not available in Essentials)	342,225
INRIX via Moonshadow DB4Iot	Sum of unique trip IDs	786,636
Otonomo	Sum of Individual trips in Otonomo trip summary report	67,441

[a] as discussed in Section 5, our analysis suggests Otonomo's trip summary report undercounts the number of vehicle trips by ~60 percent.

Because reported INRIX and Otonomo data are not scaled to estimate total trips, the only direct means of comparison is the hourly distribution of trips across a specific day. The distribution was the only common element as Moonshadow and Otonomo did not provide a basis for scaling to total trips or VMT, as StreetLight Insight - Essentials did (this was addressed in a case study discussed in Section 4). This comparison is shown in Figure 1, which shows the volume of trips as a percentage of daily total. StreetLight and Moonshadow align well, considering the variability inherent in StreetLight being an average of five days while Moonshadow represents one specific day. The difference in Otonomo trip patterns stems for the all-rental fleet. While personal vehicles exhibit a typical bi-modal peak in morning and afternoon rush hours, the Otonomo fleet peaks later in the day.

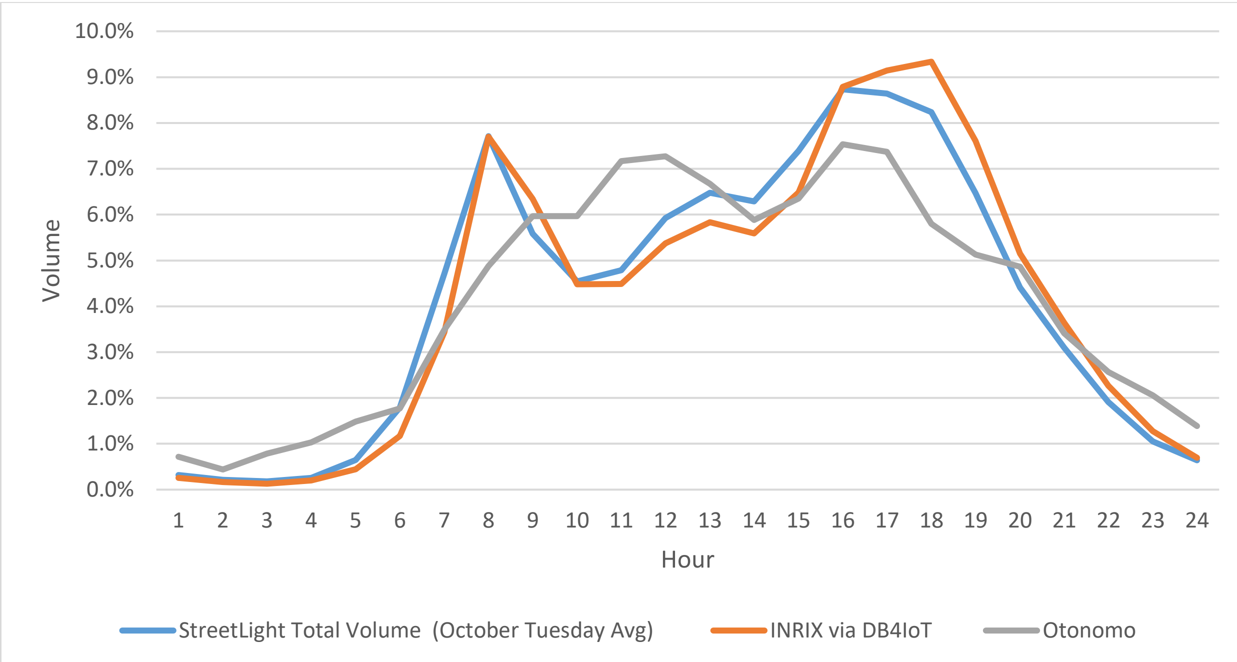

Figure 1. Hourly Distribution of Trips for Three Datasets

4. CASE STUDIES WITH STREETLIGHT INSIGHT - ESSENTIALS

The breadth of data from StreetLight (2016 through present) allowed analysis of broader temporal scope than the other two sources. StreetLight was the only dataset which could be used to replicate annual vehicle activity inputs used for the 2017 NEI, the latest NEI iteration published by EPA as of the analysis. (U.S. EPA 2020)

StreetLight InSight - Essentials was also used to quantify the degree of travel in I/M counties from non I/M areas. The Origin-Destination function of the Essentials platform allows isolation of travel from one county to another. Outside the Denver MSA, travel can be isolated entering from specific roads (e.g. interstates).

Finally, during the course of the project the COVID shutdowns of Spring 2020 went into effect. StreetLight Data responded by releasing data in half-month increments (vs. the usual one month increment) to allow focused analysis of changes in travel trends immediately following shutdown, and during the initial "opening up" stages. Though the transportation impacts of a major health crises were not foreseen at the outset of the project, the ability to include this analysis as part of E-131 highlights additional significant capability of telematics: quick turnaround for data (a few weeks for general access, vs. annual cycle of traditional traffic data).

Each analysis is detailed in the following sections.

4.a NEI Inputs

The U.S. EPA compiles the National Emissions Inventory (NEI) to provide a comprehensive nation-wide estimate of annual air emissions of criteria and hazardous pollutants from all sectors. The NEI is developed on a three-year cycle, reporting annual emissions every third year; the most recently published cycle at time of the E-131 project was for calendar year 2017. During the development cycle EPA works closely with state, local and tribal environmental agencies to compile emissions inventories for each county in the U.S., down to very detailed subsector levels. The resulting compilation provides the official U.S. emissions inventory and serves as the basis for numerous efforts including trends analysis, air quality planning, regulation development and health exposure analyses.

For mobile sources, the inventory compilation is extensive because it requires vehicle activity data representing the entire year – allocated by month, day type and hour in some cases. Since the transition to MOVES for the NEI starting with the 2011 inventory year, state and local agencies have the option of providing custom MOVES inputs at the county level for vehicle activity (e.g. total VMT, temporal VMT and start allocations, starts per day, average speed distributions) and fleet data (e.g., vehicle population, age distribution). Since this is optional, state and local agencies also have the option of using default county-level data developed by EPA. For 2017, all counties in the Denver MSA used EPA default data. For this case study, we evaluated how the StreetLight Insight - Essentials platform could be used to develop NEI activity inputs. StreetLight data were analyzed for starts, average speed, and VMT. NEI-quality inputs were derived from starts and VMT, and compared to the EPA defaults used for the 10 Denver MSA counties in the 2017 NEI; these analyses are presented in the following sections. As discussed in 4.a.3, speed data from the Essentials platform was too aggregate to be used for MOVES NEI inputs.

4.a.1 Working with Streetlight Insight - Essentials Platform

With the StreetLight InSight platform, users have the option to upload custom GIS shapefiles that contains polygon or line segment zones. Users can also create zones using the Streetlight graphical interface. Polygon zones can be zip codes, counties, states, or custom geographic areas. Line segments usually represents streets and highways (complete or partial segments). For this analysis, ERG uploaded a GIS shapefile consisting of the 10-County Denver MSA (Figure 2) to Streetlight.

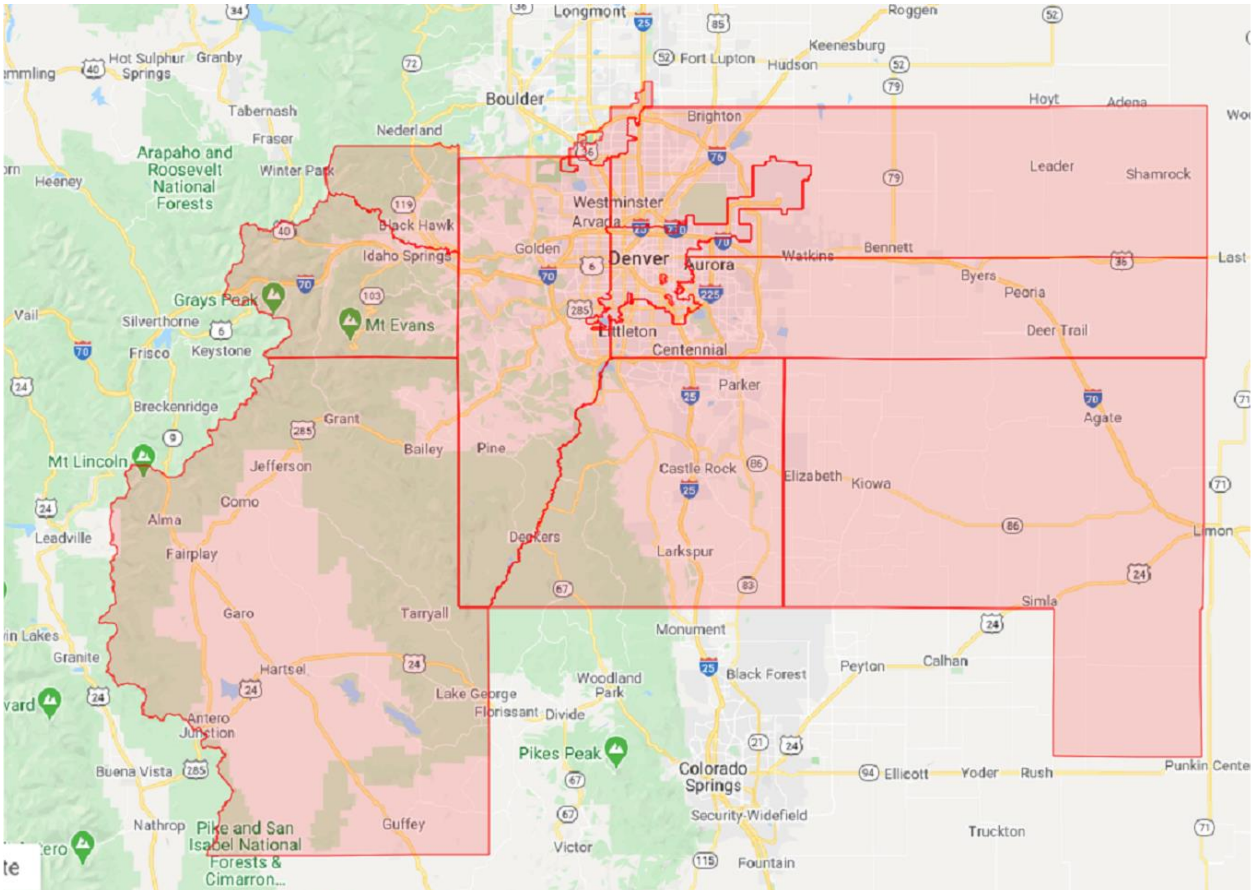

Figure 2. 10-County Denver MSA Zone (Study Area)

The Streetlight platform offers different types of analyses including modular analysis (zone activity and origin-destination analysis), average annual daily traffic analysis (AADT), exploratory analysis (e.g., top routes between origin and destination), traffic diagnostics, and segment analysis. For this project, ERG subscription only allows modular analysis and AADT.

The Streetlight platform can generate trip counts by year (individual or combined years), month (individual or combined months), day of the week (individual or group of days), and by hour of the day (or grouping of hours). Additional data metrics include average trip speed, trip length, and trip duration for all the trips included in the grouping (by year, month, day, and hour).

Analysis options in the Essentials platform are summarized below.

- **Zone activity analysis –** This analysis provides information on number of trips that start and/or end in the study area, average trip speeds, average trip durations, and average trip lengths. VMT can be estimated by multiplying the number of trips with the average trip length (in miles). However, the average trip length data available from the zone activity analysis includes trip length from start of trip to end of trip. For trips starting and/or ending outside the study area, the trip length includes entire length of the trip and not just the portion that occurs within the study area. Therefore, not all the VMT estimates generated using the zone activity analysis are accurate. This type of analysis produces reliable VMT estimates for trips that start and end in the study area (i.e., no travel outside study area). For trips that involve some portion of travel outside the study area, the VMT estimates are inaccurate.

- **AADT analysis –** Estimates the annual average daily traffic counts for a given zone (i.e., annual traffic counts divided by 365 days). Trip speed, length, and duration data are not included as part of this analysis. AADT analysis can only be performed on zones (polygon or line segment) not exceeding 0.04 square kilometers in size (i.e., zones smaller than 0.015 square miles). This type of analysis cannot be performed on the study area.

- **Origin-Destination (O-D) analysis** – Estimates the traffic flow between origin and destination zones. This analysis provides information on number of trips that start and end in the origin/destination zones. The O-D analysis also provides information on average trip speed, length, and duration. VMT estimates can be generated from the O-D analysis by multiplying the number of trips with the average trip length. Trips starting and/or ending outside the study area, but passing through, are not included in this type of analysis.

4.a.2 Estimating Trip Starts

Background

Emissions from vehicle engine starts are an important contributor to overall on-road emissions, and MOVES tracks start exhaust emissions separately from running exhaust. For a 2017 national average fleet, the emissions from engine starts make up 42 percent of total on-road volatile organic compounds (VOC), 36 percent of carbon monoxide (CO), and 14 percent of oxides of nitrogen (NO_X).

The important activity parameters to estimate engine start emissions are the number of starts by time of day and the duration of time between trips. As of MOVES2014b, the data source for engine start activity is instrumented vehicle studies. In those studies, portable activity monitors (PAMS) were installed on a sample of vehicles. The PAMS devices collected time-stamped engine on/off events (among other parameters), which produces the number of starts for the vehicle by time of day, as well as trip length and time between trips when the engine is off (known as soak time). Instrumented vehicle studies have the advantage of known individual vehicle characteristics (such as vehicle class and model year) and a complete picture of trips at the individual vehicle level. The downside of PAMS studies are that the sample size is relatively small.

Telematics data provides the total number of trips for many more vehicles and provides built-in geofencing capability, but with a trade-off of high degree of vehicle anonymity. In StreetLight Data InSight, and other platforms, the vehicle trips are aggregated without any information about earlier or later trips on the same vehicle. The anonymity means the data are not suited for evaluating engine soak periods.

Sample Size

The instrumented vehicle studies underlying the starts activity in MOVES2014 came from six instrumented vehicle studies conducted between 1992 and 2005 in eight U.S. cities. The StreetLight Data sample size over the four-year period in the Denver MSA was approximately 3,627,000 unique devices (Table 7). When we limit the data year for purposes of comparison to 2017 NEI, the sample size is approximately 795,000 devices in 2017. As a point of comparison, this 2017 device count of 795,000 is 35% of the 2.3 million registered "personal" vehicles (motorcycles, passenger cars, and passenger trucks) in the ten county area, according to the 2017 National Emissions Inventory (U.S. EPA, 2020).

Table 7. Summary of MOVES2014 Data Sources vs. StreetLight InSight - Essentials

Parameter	Instrumented Vehicle Study (MOVES2014)	StreetLight Data Platform
Sample Size	1,305 vehicles [A]	3.6 million devices [B]
Number of starts by time of day	Yes	Yes
Duration of time between trips	Yes	No

[A] The 1,305 vehicles are the vehicle count for six instrumented studies in Table 12-6 of U.S. EPA (2016).
[B] StreetLight Data platform limited to the Denver MSA over the four years 2016-2019.

Number of Starts

The MOVES model estimates the number of starts by multiplying vehicle population (either supplied by state agencies, or in the case of Denver area counties for 2017, using MOVES defaults from state registration data) and the number of starts (trips) per vehicle per day from the instrumented vehicle studies. MOVES2014b estimates the average passenger vehicle takes nearly 6 trips per day, though EPA has indicated, based on analysis of Verizon data, that this estimate will be lower in the next version of the model (EPA, 2019).

The StreetLight Platform offers two output options for trip origin analyses. The StreetLight "Index" output type reflects the number of observations from unique devices. The StreetLight "Volume" output type reflects StreetLight's estimate of the totals, using scaling factors derived from analysis of traffic monitors. Table 8 and Table 9 (Weekday and Weekend) show the ratio of StreetLight Index to Volumes covers, and that the Index covers about half the total trips. Table 8 and Table 9 also compare the percent difference between StreetLight's estimate of total trip starts from the NEI 2017 approach. Denver and Douglas show the best agreement on weekdays, while Denver, Clear Creek, and Park track the best for weekends.

Table 8. StreetLight Starts Estimated Total vs. NEI 2017 Estimated Total (Weekdays)

County	StreetLight Data, 2017			NEI 2017 Starts (Estimated Total)	Percent difference (Volumes - NEI)/NEI
	Index (Sample)	Volumes (Estimated Total)	Ratio of Index/Volumes		
Adams	1,028,258	2,673,499	0.385	2,336,635	14.4%
Arapahoe	1,449,342	2,129,984	0.680	2,986,801	-28.7%
Broomfield	157,841	227,211	0.695	308,818	-26.4%
Clear Creek	14,816	62,421	0.237	95,139	-34.4%
Denver	1,662,832	3,136,417	0.530	2,985,808	5.0%
Douglas	681,593	1,606,810	0.424	1,553,827	3.4%
Elbert	24,908	86,967	0.286	165,673	-47.5%
Gilpin	8,080	34,777	0.232	44,067	-21.1%
Jefferson	1,241,795	2,526,550	0.491	2,751,583	-8.2%
Park	11,241	98,780	0.114	118,332	-16.5%
Total	**6,280,706**	**12,583,416**	**0.499**	**13,346,683**	**-5.7%**

Table 9. StreetLight Starts Estimated Total vs. NEI 2017 Estimated Total (Weekends)

County	StreetLight Data, 2017			NEI 2017 Starts (Estimated Total)	Percent difference (Volumes - NEI)/NEI
	Index (Sample)	Volumes (Estimated Total)	Ratio of Index/Volumes		
Adams	906,350	2,342,305	0.387	2,027,030	15.6%
Arapahoe	1,251,046	1,838,548	0.680	2,610,443	-29.6%
Broomfield	134,621	193,416	0.696	269,075	-28.1%
Clear Creek	19,384	78,189	0.248	78,801	-0.8%
Denver	1,382,404	2,584,343	0.535	2,592,590	-0.3%
Douglas	624,902	1,465,175	0.427	1,344,573	9.0%
Elbert	23,670	81,450	0.291	141,775	-42.5%
Gilpin	12,192	49,023	0.249	37,989	29.0%
Jefferson	1,130,422	2,286,671	0.494	2,399,840	-4.7%
Park	11,807	99,708	0.118	100,769	-1.1%
Total	**5,496,798**	**11,018,828**	**0.499**	**11,602,884**	**-5.0%**

Overall results for StreetLight Data vs. NEI 2017 over the MSA are relatively close (5-6 percent different), though differences vary significantly by county. For weekdays, and a lesser extent weekends, smaller counties shows lower starts than assumed in the NEI, while the larger counties (Denver, Douglas, Arapahoe) show more starts. This may reflect a redistribution of start activity depending on whether a county is a trip generator (e.g. suburb / exurb) or trip attractor (e.g. urban core). This is not accounted for in EPA's NEI defaults since the same starts/vehicle estimate is applied to all counties, meaning starts only vary by vehicle population. This highlights a strength of telematics: improved spatial distribution of vehicle start and trip activity. Though start emissions over the Denver MSA would not be significantly different if StreetLight trip starts were used, the redistribution of these emissions spatially could be important for regional ozone and PM formation.

Distribution of starts by time of day

We examined the larger set of all four data years 2016-2019 to analyze the distribution of starts by hour, day of week, and month of year. Using multiple years of traffic monitor data is common practice in the development of temporal distributions of VMT by hour, day, and month.

The MOVES model distributes day total starts to the 24 hours based on instrumented vehicle studies and others as discussed in US EPA (2016). Figure 3 shows the MOVES2014b personal vehicle relative starts by hour of day for weekday and weekend day types. The weekday profiles (left) resemble an urban VMT profile with a morning and evening peak and some midday activity.

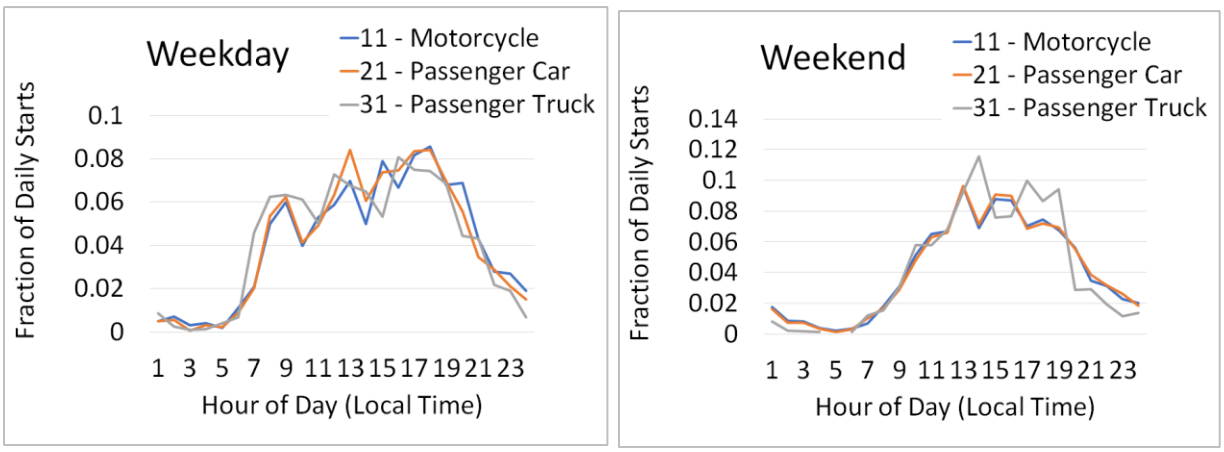

Figure 3. MOVES Default Hourly Trip Distributions

In contrast, the hourly trip distributions derived from StreetLight (Figure 4) shows that the diurnal trip start patterns can vary significantly by county even within a single MSA. StreetLight metrics were available for all seven days of the week, and patterns are apparent within the categories of weekday and weekend. Fridays show the morning peak at the same hour (Hour 8, 7:00 to 7:59 AM) as other weekdays, but Friday has higher midday starts and often an earlier afternoon peak, compared to Monday through Thursday. For weekends, Saturday trip starts are most often higher than Sunday (similar to VMT). These data show that the national average starts distribution by hour may not track well for more rural counties (Clear Creek, Gilpin, and Park).

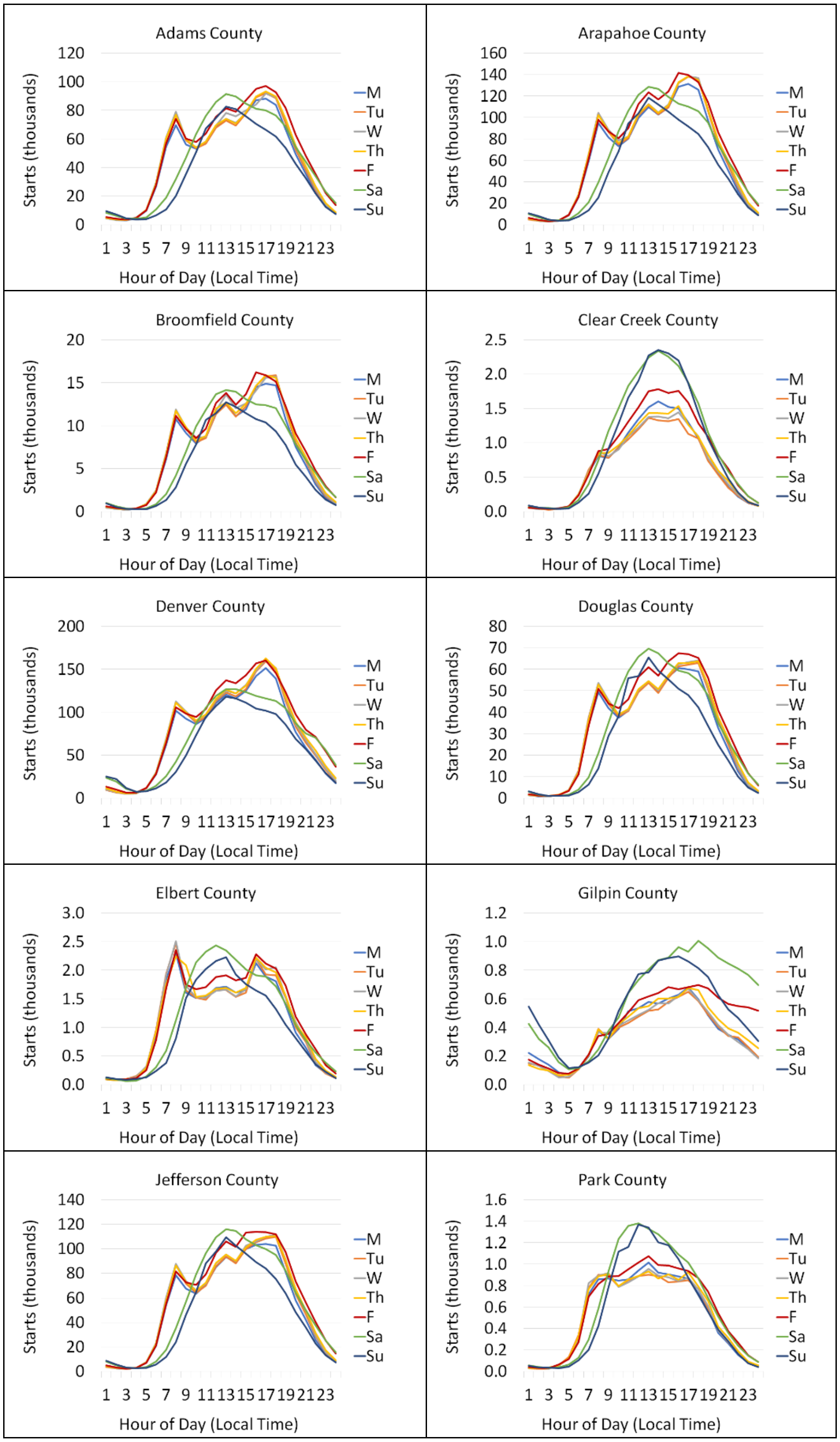

Figure 4. StreetLight 2017 Hourly Trip Distributions

Distribution of starts by day type

The MOVES model uses just two day type categories, although on-road emissions modeling for regulatory purposes and air quality planning uses an emissions processor called SMOKE, which operates at a higher resolution of seven day types of the week. Because vehicle start activity has historically been difficult to gather at the local level (often surveys or trip generation/travel demand models), SMOKE relies on the MOVES model start activity applied to multiple day types (i.e., weekday applies to five days and weekend applies to two days). These SMOKE weekly temporal profiles will be flat over Monday through Friday and step decrease to Saturday and Sunday. Figure 5 shows the diversity trip start patterns over day of week within the Denver MSA. In general, the number of starts steadily rises over Monday to Friday with a decrease on Saturday followed by another on Sunday. The exceptions to this are Clear Creek, Gilpin, and Park counties.

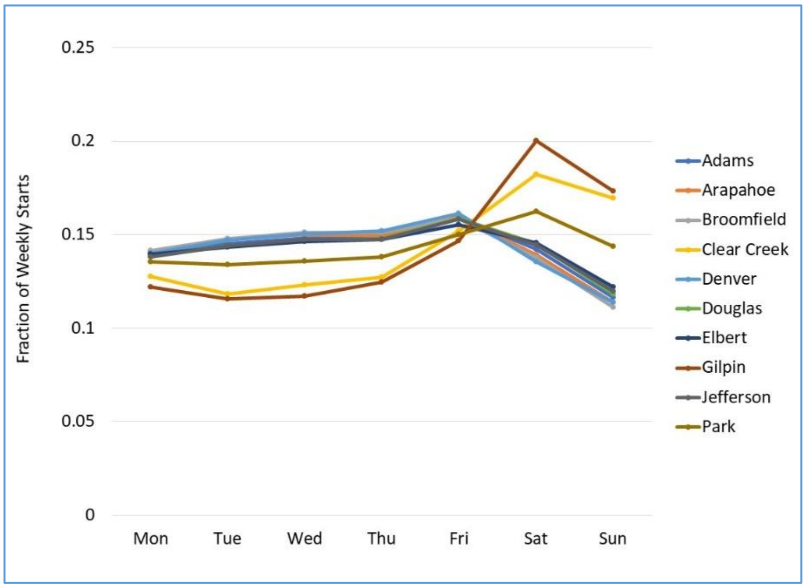

Figure 5. StreetLight Platform Starts Distribution by Day Type of Week

Distribution of starts by month of year.

As of MOVES2014b, the model does not vary the number of starts by month of year. Traditional activity data collection programs have typically only instrumented vehicles for 1-2 weeks, and therefore cannot provide activity data across multiple months. Telematics provides the temporal breadth for this however, with results shown in Figure 6. There are some differences that show generally higher starts in the summer months; this is most pronounced for Clear Creek and Park counties. The trend of higher starts in summer than winter is similar to typical VMT patterns.

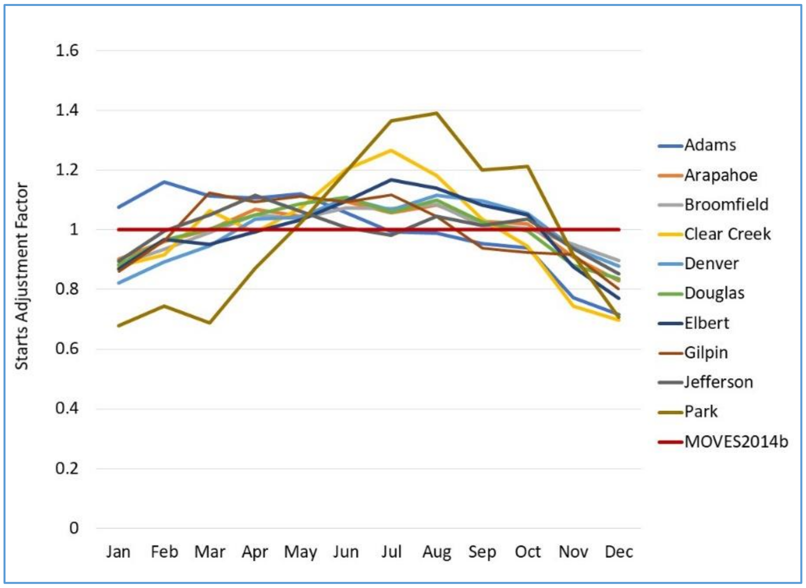

Figure 6. StreetLight Platform Derived Starts Adjustment Factor by Month

4.a.3 Average Speed Distribution

Average speed distributions for MOVES have been previously developed from telematics data for national defaults by EPA, and at a more detailed level (individual county or grouped counties) for the NEI as part of the CRC A-100 (DenBleyker et al 2019). For the latter, StreetLight data was used to generate MOVES inputs, but it required custom programming to distribute driving activity in MOVES speed bins. The StreetLight Insight - Essentials platform provides immediate access to trip data, but at a more aggregate level than the custom programming done in the A-100 project. Our evaluation thus focused on whether the Essentials platform could produce speed distributions at the resolution needed by MOVES, replicating A-100.

Speed-Distribution Data in Streetlight InSight - Essentials

ERG also investigated using the StreetLight Insight - Essentials platform to obtain speed distribution data required by MOVES. The data available from Streetlight consists of a distribution of average trip speeds for all the trips included in the analysis. The data does not include the time that the vehicle spends travelling in different speed bins. The MOVES average speed distribution is meant to reflect a mid-point of these two approaches, reflecting link-level average speeds estimated in traditional travel demand models. This is what was used in the CRC A-100 project which produced county-level average speed distributions from StreetLight Data, and required custom programming from StreetLight to produce. Discussion with StreetLight during the market survey phase indicated that their InSight platform could replicate this level of detail at higher cost points, e.g. the Advanced Analytics or Multimode levels. Our analysis confirmed the speed distributions available at the Essentials levels is too aggregate, reported as a distribution of overall average trip speeds. This distribution would be applicable in EMFAC however, which characterizes average speed distribution based on entire trips rather than road links.

Streetlight analyses provide trip average speed (mph) as one of the data metrics along with average trip duration (in seconds) and average trip length (in miles). ERG ran a zone activity analysis to review available speed data in Streetlight. Streetlight users can establish custom speed bins as part of the analysis. ERG set up speed bins that are closely aligned with speed bins used in MOVES. The speed bins were – 0-3 mph, 3-8 mph, 8-13 mph, 13-18 mph, 18-23 mph, 23-28 mph, 28-33 mph, 33-38 mph, 38-43 mph, 43-48 mph, 48-53 mph, 53-58 mph, 58-63 mph, 63-68 mph, 68-73 mph, and above 73 mph.

The zone activity analysis was setup with the following parameters:

- Zone = 10-County Denver MSA
- Year = 2017
- Months = All
- Day Types = All days (Mon-Sun)
- Hour = All Hours (12 am – 12 am)

ERG filtered out the results for Denver County for a closer review of the speed data. Table 10 presents the average daily speed distribution results for Denver County. As shown, about 60 percent of trips were estimated to have average speeds below 13 mph, which is also the overall sample average. By comparison, the annual 2017 Denver County average speed from CRC A-100 is 24.9 mph. The StreetLight speed distribution for this analysis appears biased low, which may be an artifact of the error introduced in converting LBS to specific vehicle trips. In particular, StreetLight does not assign a trip end until a device moves less than 5 meters in 5 minutes; any chained trips with stops less than 5 minutes would be grouped, and the stop time added to overall trip duration. This would serve to lower the average trip speed, and skew the overall trip speed distribution as shown in Table 10.

Table 10. Speed Distribution of Trips for Denver County for 2017.

Trip Speed Bin	Percent of total trips with trip average speed within the speed bin
Trip Speed 0-3 mph	9.5%
Trip Speed 3-8 mph	27.9%
Trip Speed 8-13 mph	23.5%
Trip Speed 13-18 mph	16.2%
Trip Speed 18-23 mph	9.6%
Trip Speed 23-28 mph	5.5%
Trip Speed 28-33 mph	3.3%

Trip Speed Bin	Percent of total trips with trip average speed within the speed bin
Trip Speed 33-38 mph	2.0%
Trip Speed 38-43 mph	1.2%
Trip Speed 43-48 mph	0.7%
Trip Speed 48-53 mph	0.4%
Trip Speed 53-58 mph	0.2%
Trip Speed 58-63 mph	0.1%
Trip Speed 63-68 mph	0.1%
Trip Speed 68-73 mph	0%
Trip Speed Above 73	0.1%
Trip Start Count = 2,977,598; Trip End Count = 3,023,499	
Average Trip Speed for all Trips = 13 mph	

4.a.4 Vehicle Miles Travelled (VMT)

VMT Estimation Methodology

VMT can be estimated from StreetLight Insight – Essentials trip metrics by multiplying the average daily trip volume with the average trip length (in miles). A rough estimate of the VMT can be generated using either the zone activity analysis or the O-D analysis. This section presents the results of VMT estimated using zone activity and O-D analyses and describes the recommended method to obtain a more accurate VMT estimate using O-D analysis with entry/exit gates. VMT estimates were generated using results from the zone activity analysis using the following parameters:

- Year = 2017
- Months = All
- Day Types = All days (Mon-Sun)
- Hour = All Hours (12 am – 12 am)

The zone activity analysis only considers trips that either started or ended in the study area, regardless or either origin or destination. With this mode of analysis, trips that start and end outside the study area and pass through the study area are not included. Trips starting and ending within the study area are generally the same trips, so it is not appropriate to sum the VMT from both sets as this will double-count VMT.

VMT estimates are presented in Figure 7 by county based on trip starts, and trip ends; both methods results in comparable VMT. VMT is calculated as the product of StreetLight trip volume (scaled) and average trip length. The average trip length data account for the total trip length (start to end) and not just the portion of trip that occurred within the 10-County area. Therefore, for trips that start outside and end inside the study area (or vice-versa), the VMT estimates include trip length outside the study area - – so the mileage from long-distance trips that happen to start or end in the study area are included, inflating VMT estimates. The corresponding VMT estimates are therefore significantly higher than those used in the 2017 NEI estimates.

Alternatively, the O-D analysis provides a better VMT estimate than the zone activity analysis. The O-D analysis takes into account origin and destination of the trips; this can ensure only VMT within the 10 county area is included, excluding out-of-area VMT. The O-D analysis was created using the following parameters:

- Origin zone = 10-County Denver MSA
- Destination zone = 10-County Denver MSA
- Year = 2017
- Months = All
- Day Types = All days (Mon-Sun)
- Hour = All Hours (12 am – 12 am)

Figure 7 also includes 2017 VMT estimates generated using data from the O-D analysis; "II" standards for internal-internal, to reflect trips occurring completely within the MSA. The VMT estimates are much lower than those based on trip starts and trip ends; though still much higher than the NEI for most counties.

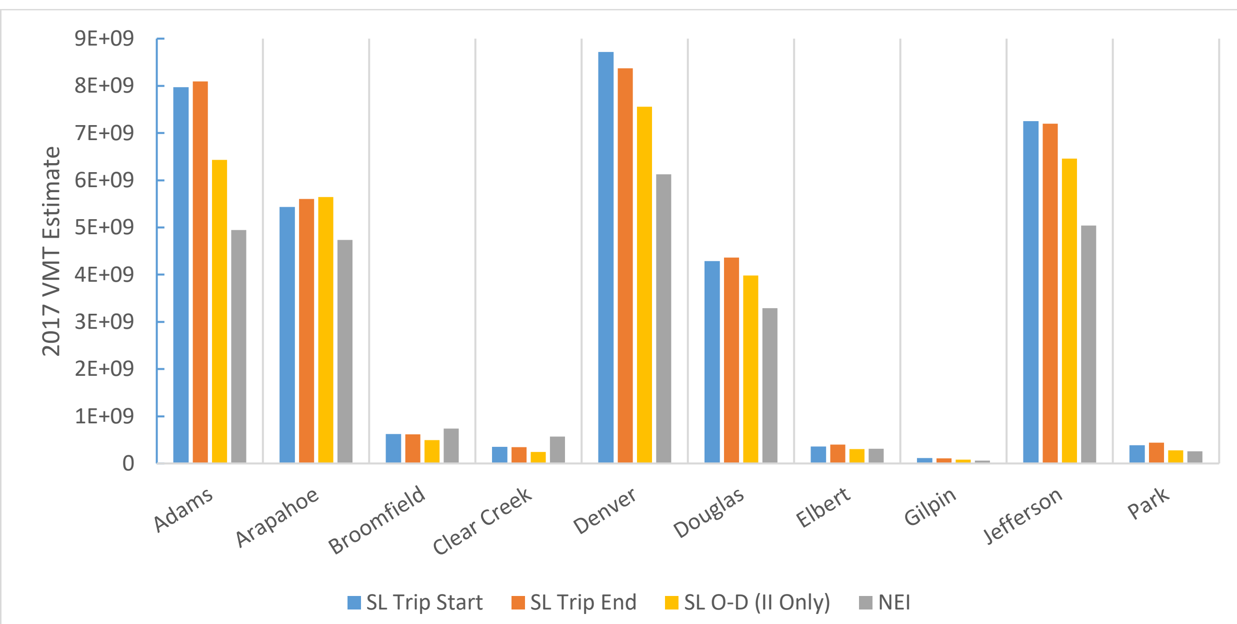

Figure 7. 2017 VMT Estimates by County: Several StreetLight Metrics & NEI

For the O-D analysis shown in Figure 7, trips that originate outside the study area and end in the study area are not included in the “II” trips. Trips that start or end outside the study, including “pass-through” trips (no stops in area) are not included. As detailed on StreetLight’s online support forum (StreetLight 2020), more accurate VMT estimate can be achieved by breaking down the travel in the study area by analyzing all internal and external traffic patterns. This method requires a more complex setup but produces a better VMT estimate by analyzing only the portions of trips that are within the study area. This method relies on O-D analysis with boundary gates (entry and exit gates) established for the study area. Ideally boundary gates should be established for all roads and highways that lead travel in and out of the study area. For the purpose of this study, it was decided to establish boundary gates for the 3 major interstate highways that lead travel in and out of the 10-County Denver MSA. Boundary gates (entry and exit gates) were setup for I-70 (E/W), I-25 (N/S), and I-76 (E/W). Entry gates (trips entering the study area) and exit gates (trips leaving the study area) were established at the intersection of the interstate highways and the study area. Four boundary gates were established for both I-70 and I-25 and 2 boundary gates were established for I-76 (I-76 enters the study area and merges with I-70). Locations of boundary gates used for this analysis are shown in Figure 8, with a breakout showing an example of an entry gate (northbound I-25) defined as a zone in StreetLight.

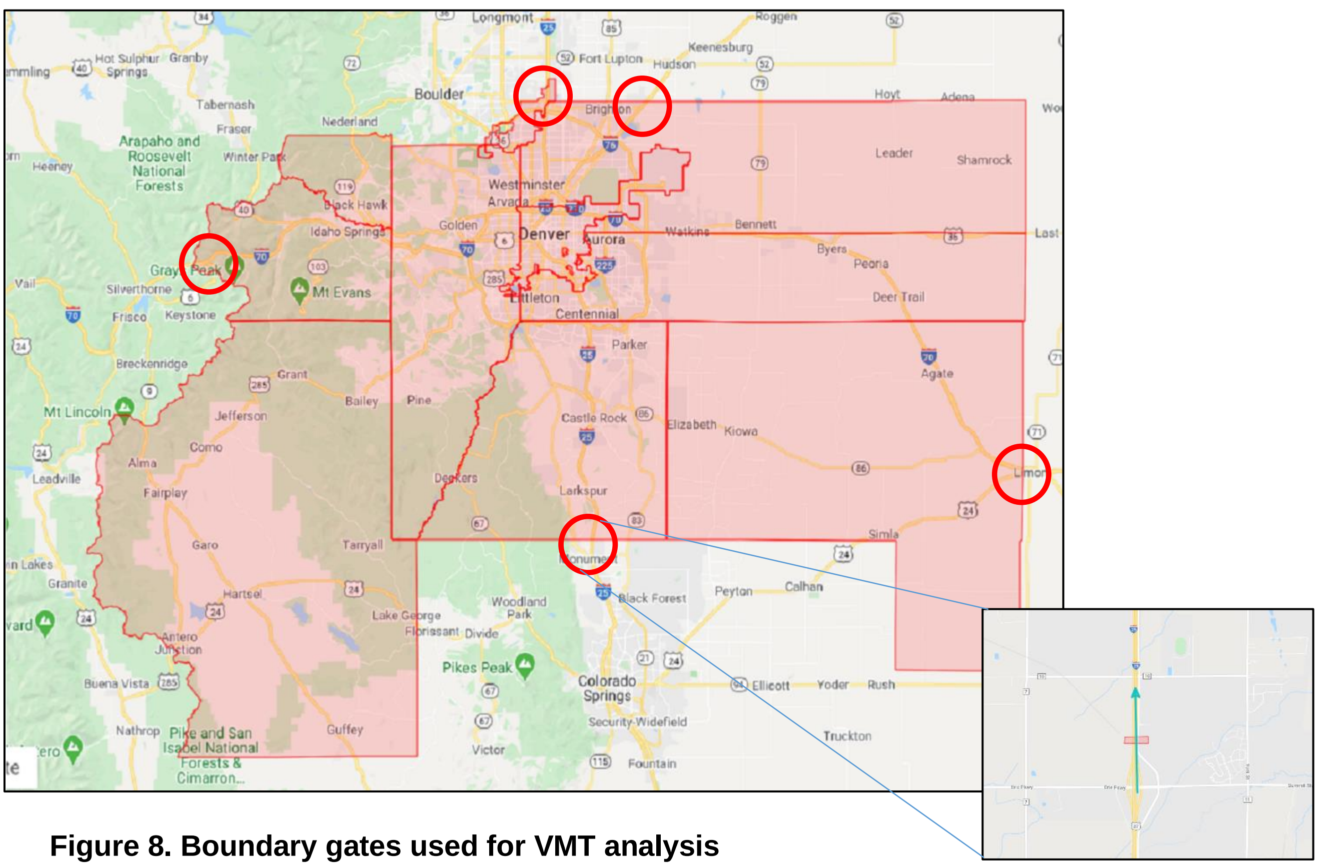

Figure 8. Boundary gates used for VMT analysis

Using the boundary gates, travel in the study area can be broken down as follows:

- Internal-Internal (II) – Trips that originate and end in the study area (i.e., origin and destination = study area).
- Internal-External (IE) – Trips that originate in the study area and travel outside the study area (i.e., Origin = study area; Destination = exit boundary gates).
- External-Internal (EI) – Trips that originate outside the study area and end within the study area (i.e., Origin = entry boundary gates; Destination = study area).
- External-External (EE) – Trips that originate and end outside the study area but pass through the study area (i.e., Origin = entry boundary gates; Destination = exit boundary gates).

The trip length data from this analysis will only include length from origin to destination. Since the origin and destination for this analysis are the study area, entry gates, and exit gates, the trip length will only cover travel within the study area. The O-D analysis with boundary gates is setup with the following parameters:

- Origin zone = 10-County Denver MSA and boundary entry gates
- Destination zone = 10-County Denver MSA and boundary exit gates
- Year = 2017
- Months = All
- Day Types = All days (Mon-Sun)
- Hour = All Hours (12 am – 12 am)

The results of the O-D analysis with boundary gates are presented in Table 11 External trips (IE, EI, and EE) account for about 10 percent of VMT in the Denver MSA. Though this analysis can be further improved by established boundary gates on more roads that travel in and out of the study area, these case study estimates account for the majority of travel that occurs within the study area.

Table 11. Denver MSA 2017 VMT Estimate w/ Boundary Gates

Trip Type	Trip Description	Estimated Annual VMT	Percent of Total VMT
II	Trips that originate and end in the study area	31,778,838,150	91%
IE	Trips that originate in the study area and travel outside the study area	1,359,135,718	4%
EI	Trips that originate outside the study area and end within the study area	1,516,482,400	4%
EE	Trips that originate and end outside the study area but pass through the study area	252,074,694	1%
Total Estimated 2017 Annual VMT for the 10-County Denver MSA		**34,906,530,961**	
2017 NEI VMT for Denver MSA		**26,080,638,162**	

For comparison, the 10-county VMT totals used for the 2017 NEI are shown in Table 11 ; the StreetLight estimate of VMT is 34 percent higher. As detailed in Section 5, StreetLight's VMT estimates are supported by independent comparison of INRIX data and Colorado DOT traffic counts. With VMT being the most influential variable on overall vehicle emissions [cite], use of the telematics-based VMT estimates would increase vehicle emissions estimates significantly relative to the NEI.

4.b Non-I/M vehicles in I/M area

The Denver I/M program includes a subset of the 10-County MSA (including all of Broomfield, Denver, Douglas, and Jefferson Counties, plus portions of Adams and Arapahoe Counties), plus counties outside the study area (e.g. Boulder County). A factor in the effectiveness of I/M programs is the extent of higher-emitting vehicles from outside the I/M area travelling (and emitting) within I/M program boundaries. Vehicles registered outside the program area that travel frequently in the I/M area are supposed to undergo inspection, but CHPHE confirms this is difficult to track and enforce. Telematics data may be helpful to at least quantify the amount of non I/M traffic in I/M areas, though not for specific vehicles. Following on the VMT case study, the Origin-Destination feature of StreetLight Insight - Essentials was also applied to quantify the extent of VMT in I/M counties that originate in Non I/M counties. Note this analysis cannot determine the specific registration area of vehicles, but is only inferring based on trip origins.

To estimate the amount of traffic from non-I/M areas in the Denver area, an Origin-Destination analysis was conducted, finding that traffic from non-I/M counties and entry gates to the I/M counties of Broomfield, Denver, Douglas, and Jefferson Counties accounted for 3 percent of VMT in the 10-county area. CDPHE also suggested looking specifically at traffic from El Paso county, a non-I/M county that is home to Colorado Springs and a source of significant traffic into the Denver area. El Paso county as a whole was not within the study area purchased from Streetlight, so couldn't be studied directly. As a proxy, the I-25 North entry gate was used; vehicles entering the Denver area through this gate accounted for nearly 0.5 billion miles travelled in 2017, about 1.5 percent of VMT in the area.

4.c Impacts of COVID-19 shutdown (March – May 2020)

During the course of the project the COVID shutdowns of Spring 2020 went into effect. StreetLight Data responded by releasing data in half-month increments (vs. their standard one month increment) to allow focused analysis of changes in travel trends immediately following shutdown, and during the initial "opening up" stages. Though the transportation impacts of a major health crises were not foreseen at the outset of the project, the ability to include this analysis as part of E-131 highlights additional significant capability of telematics: quick turnaround for data (a few weeks for general access, vs. annual cycle of traditional traffic data). To demonstrate the utility of telematics data for analyzing one-off events and trends, we assess the change in average daily VMT right after COVID shutdown in Denver MSA (official lockdown ordered on March 16, 2020) vs. the average March VMT from prior four years (2016-19) and subsequent 15 day periods of 2020 through early May. Results are shown in Figure 9. Change in Daily VMT from COVID Shutdown in Denver-area Counties for the five populous counties in the Denver MSA. VMT was estimated based on trip starts in each county, so may include VMT outside the metro area. The reductions show for the period immediately after shutdown vary from 55 percent (Denver county) to 68 percent (Adams county). The rate of increase during gradual "opening up" vary by county as well.

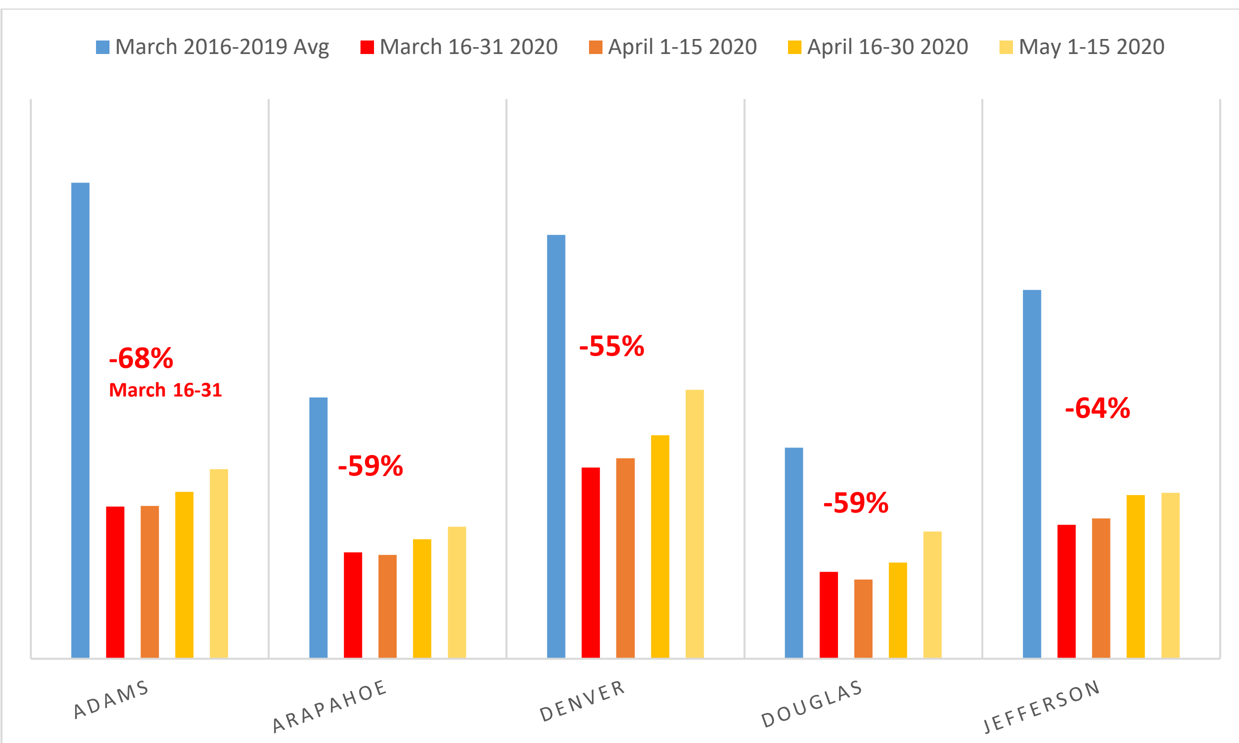

Figure 9. Change in Daily VMT from COVID Shutdown in Denver-area Counties

5. CASE STUDIES MOONSHADOW DB4IOT / INRIX

5.a Estimating Total VMT with Moonshadow DB4IoT Platform

Moonshadow loaded INRIX vehicle trip data from October 1-31 2019 into Db4IoT with. Each vehicle trip has a unique identification code, a known origin/destination, set of waypoints along the trip, and the total trip distance traveled. Since the INRIX trip data was "as is", without further scaling to reflect total trips, ERG conducted an analysis to scale up trips, and compare to StreetLight total volume results for the same period. The approach for estimating total VMT from the INRIX sample of vehicle trips involved generating summaries of the INRIX VMT at the county level and generating scaling factors that represent the vehicle coverage of the INRIX data, based on comparison to traffic counts from Colorado Dept. of Transportation (CDOT). INRIX VMT was scaled up accordingly to an estimated total VMT.

Figure 10 shows a sample of the INRIX trip waypoints for one day: Wednesday, October 2, 2019. The colors of the waypoints signify vehicle speed, and the most readily visible colors below are green and yellow, which correspond to approximately 30-60 kph (20-40 mph). The red color waypoints represent vehicle speed between roughly 90-120 kph (55-75 mph), and these are mostly visible along freeways, especially outside of the Denver urban area.

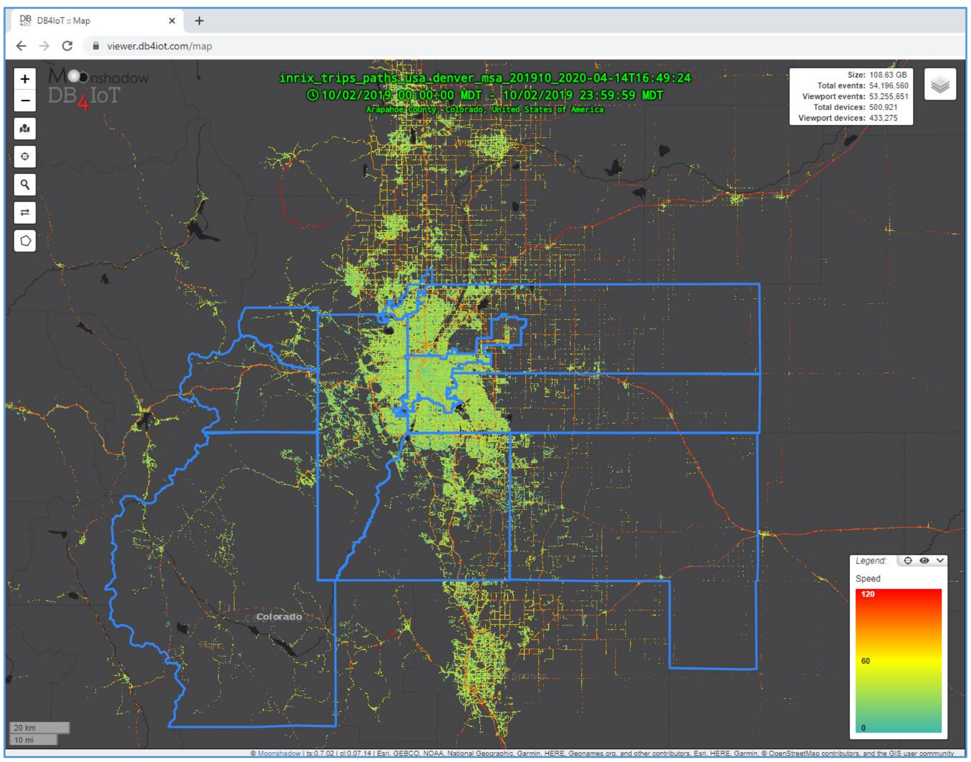

Figure 10. INRIX trip waypoints in the DB4IoT platform on October 2, 2019

In addition to the mapping feature, DB4IoT has a query builder tool that allows users to download summarized trip information. The query builder tool includes all columns available from the INRIX source data, user choices for aggregation of the data (sum, mean, min, max, unique count, etc.), grouping parameters such as date, time of day, etc., and various filter options including county boundaries. ERG used the DB4IoT query builder tool with the county filter to extract daily VMT for each of the 10 counties for October 2019. Due to the large volume of data and server response limitations, we extracted day total VMT, one county at a time for the full month of October 2019.

ERG also obtained hourly traffic counts for October 2019 from CDOT traffic volume monitors, which Moonshadow staff loaded into their platform for this work. In addition, Moonshadow added a new parameter to our merged dataset called a "volume factor," which they calculated as the hourly INRIX volume divided by hourly CDOT volume multiplied by 100. The volume factor can be thought of as the coverage level. Figure 11 shows the distribution of volume factors (expressed as percent), showing the most common volume factor is 6%. The volume factor values vary by traffic monitor, hour of the day, and day type of week, while the vast majority fall between 3 and 12%.

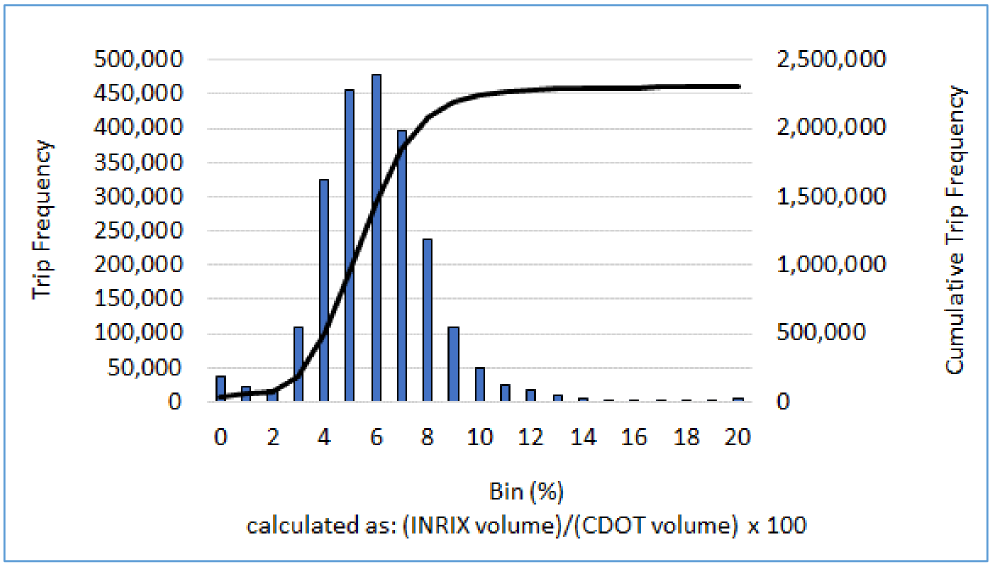

Figure 11. Distribution of the Number of Trips by Volume Factor (Percent Coverage)

Figure 12 shows the locations of the CDOT monitors. Note that there were no monitors located in Elbert County (lower rightmost county boundary).

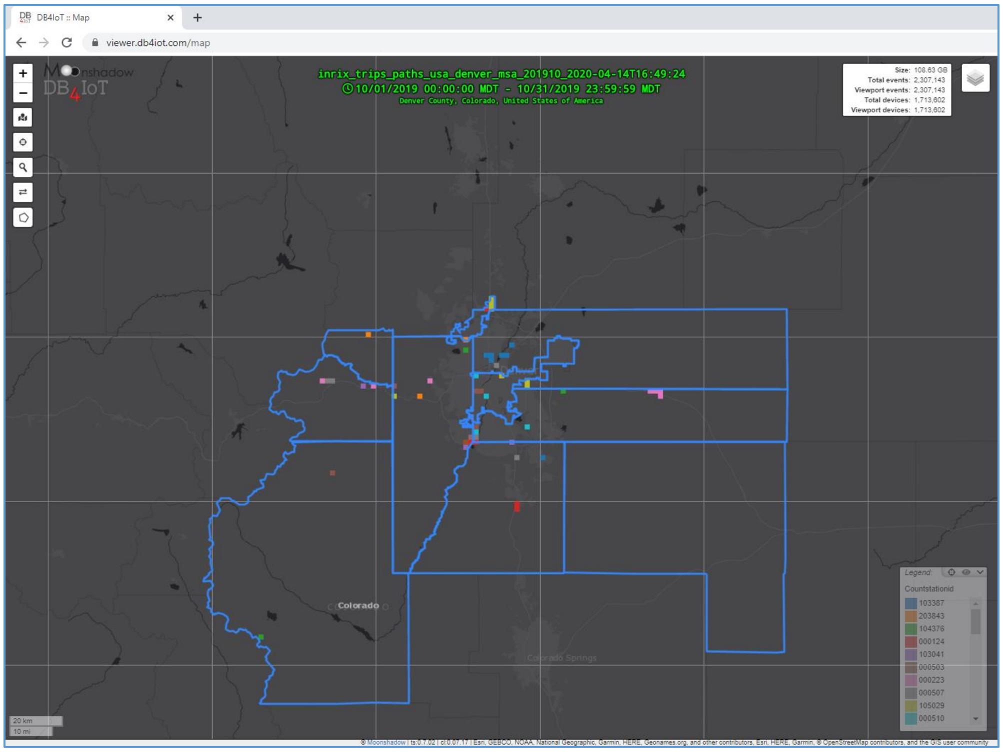

Figure 12. Location of Colorado DOT monitors, viewed in the DB4IoT platform

Estimating INRIX VMT

ERG extracted the distance traveled by trips in each of the counties of the 10-county MSA. One limitation of using the DB4IoT platform to estimate total VMT is that there aren't tools built in to parse the individual trips temporally or spatially. Because DB4IOT doesn't parse trips by date, the trip IDs and their VMT is duplicated when the trips occur over midnight. DB4IoT instead reports the full length of the trip on both dates. Similarly, the total trip distance is reported without ability to divide the distance by road type (freeway vs. non-freeway) or by county (e.g., a trip that starts in Denver but ends in Arapahoe).

An example of trip distance duplicate reporting can be seen in Table 12, where a trip began in Adams County before midnight on Monday, October 14 and traverses into neighboring county Arapahoe, and shortly after the date changed to Tuesday the 15th and the trip ended. The total distance was just over three miles, but it is reported three times in the extracted dataset.

Table 12. Example of Duplicated Trips Across County and Date

County	Date	Destination	Origin	Unique Trip ID	Reported VMT
Adams	Oct 14	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	3.037767
Adams	Oct 15	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	3.037767

County	Date	Destination	Origin	Unique Trip ID	Reported VMT
Arapahoe	Oct 15	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	3.037767

Ideally, we could have divided the trip into dates according to the distance driven on each day and within each county, but that precise information is not currently available. Instead we took a simplified approach to first allocate the trip to dates and locations. First, trips that spanned consecutive days (i.e., trips near midnight) were divided by two (e.g., 3.037767 miles / 2 days = 1.518883 miles). Next, for local trips (i.e., begin and end within the state) the trip VMT was divided by the number of MSA counties traversed. Table 13 summarizes how this approach allocates the local trip VMT into dates and counties.

Table 13. Correction of Duplicated Trips Across County and Date

County	Date	Destination	Origin	Unique Trip ID	Adjusted VMT
Adams	Oct 14	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	1.518883
Adams	Oct 15	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	0.759442
Arapahoe	Oct 15	Arapahoe	Adams	ffe757c8b8b8d8204905471b8fc0f365	0.759442
Total					**3.037767**

Trips spanning multiple MSA counties was a common occurrence (42% of trips were duplicated across at least 2 counties). Trips spanning overnight periods was uncommon (< 1% of trips).

Lastly, another challenge in tabulating county total VMT are long distance trips, or any trips with an origin or destination outside of Colorado. There weren't many of these; they made up about 1% percent of the total trips (18 million) in October 2019. Nonetheless, these long trips spanned from 70 to 1,404 miles. Without the information to precisely cut the trip distance into MSA and non-MSA travel, we again adopted a simplified approach to reduce VMT overcounting. The longest travel distance in the MSA from a pass-through trip is approximately 220 miles (the distance between towns of Buena Vista and Woodrow). Therefore, we capped the out-of-state trips at a maximum of 220 miles to exclude some of the non-MSA travel.

Table 14 shows the INRIX daily VMT summarized by day of the week and county after allocating duplicate trips to dates and counties and setting a maximum VMT for the long-distance trips.

Table 14. INRIX Average Day VMT by County and Day Type, October 2019

County	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Adams	1,025,486	890,746	1,065,442	1,077,222	1,326,131	1,192,744	893,018
Arapahoe	824,366	728,664	863,605	704,336	776,941	733,553	706,922
Broomfield	337,103	292,961	349,281	346,652	444,082	441,530	306,815
Clear Creek	130,667	103,620	136,915	139,906	229,792	222,177	210,564
Denver	896,829	769,958	929,792	967,084	1,194,284	1,078,064	783,982
Douglas	686,382	640,381	706,574	736,446	952,890	968,144	626,877
Elbert	141,712	123,000	151,929	164,841	186,865	160,481	150,231
Gilpin	8,168	6,826	7,364	7,012	11,978	15,885	10,938
Jefferson	523,387	593,452	692,306	682,755	671,659	650,550	509,458
Park	48,895	41,414	46,306	45,558	82,614	70,025	71,757
Total	**4,622,993**	**4,191,022**	**4,949,513**	**4,871,810**	**5,877,237**	**5,533,153**	**4,270,562**

Table 15 shows the volume factors by county and day type. Total VMT in Table 16 was estimated as *Total VMT = (INRIX Daily VMT)/(Volume Factor/100)*.

Table 15. Volume Factors by County and Day Type, October 2019

County	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Adams	5.3	5.2	5.2	5.3	5.4	5.9	5.7
Arapahoe	8.2	8.0	8.7	8.9	8.4	8.6	9.2
Broomfield	7.1	7.2	7.1	7.4	7.6	8.4	8.1
Clear Creek	7.3	7.2	7.6	7.9	8.0	7.6	8.3
Denver	5.2	5.1	5.3	5.2	5.3	5.2	5.2
Douglas	7.2	7.2	7.4	7.4	7.7	9.0	8.2
Elbert	N/A *						
Gilpin	4.2	4.1	4.0	4.1	3.8	4.7	4.2
Jefferson	4.7	4.7	5.4	6.7	4.5	5.6	5.4
Park	N/A *		5.7	5.4	N/A *		
Average	**6.2**	**6.2**	**6.4**	**6.5**	**6.6**	**7.2**	**6.9**

* Monitor data not available. In these areas, we scaled to total VMT using the average volume factors in the final row.

Table 16. Calculated Total Daily VMT by County and Day Type, October 2019

County	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Adams	19,526,824	16,994,749	20,384,770	20,396,456	24,351,218	20,261,640	15,688,299
Arapahoe	10,072,656	9,152,770	9,897,250	7,946,625	9,206,396	8,521,288	7,660,271
Broomfield	4,770,375	4,053,344	4,948,237	4,671,549	5,833,669	5,242,176	3,795,142
Clear Creek	1,796,497	1,443,819	1,791,424	1,770,390	2,868,915	2,914,578	2,526,435
Denver	17,371,662	14,952,302	17,467,140	18,625,527	22,732,310	20,690,045	15,109,973
Douglas	9,468,235	8,914,055	9,604,241	9,973,512	12,315,706	10,801,201	7,649,527
Elbert	2,291,443	1,991,077	2,383,905	2,541,400	2,849,043	2,238,160	2,187,301
Gilpin	194,447	167,557	182,299	172,056	313,599	336,037	257,556
Jefferson	11,160,534	12,610,785	12,735,675	10,169,268	14,902,351	11,575,990	9,445,466
Park	790,612	670,397	810,410	836,846	1,259,573	976,604	1,044,756
Total	**77,443,284**	**70,950,856**	**80,205,350**	**77,103,628**	**96,632,779**	**83,557,722**	**65,364,726**

Using the number of each day type (4 days or 5 days) in October 2019 the total VMT for the month is estimated to be 2.4 billion miles (2,433,293,214). For comparison, an O-D analysis was conducted for StreetLight in October 2019, with a resulting total of 2.2 billion miles (2,177,051,096) for "II" trips only (i.e. completely within 10-county boundary). If external trips (IE, EI, EE) are included at the same rate as estimated for 2017 (10 percent), the StreetLight estimate would increase to 2.4 billion miles (2,391,317,804). Considering our StreetLight estimate does not account for external travel on roads other than interstates, the scaled INRIX estimate and StreetLight estimates appear quite comparable. This lends support to StreetLight-based VMT for 2017 presented earlier, showing significantly higher VMT than estimated for the 2017 NEI.

5.b Project Scale "Hot Spot" Analysis

A second case study leveraged the capability of Moonshadow's DB4IoT platform to provide detailed travel data for specific locations (individual road links) at specific times (15 minute increments). This is useful for project scale "hot spot" analysis to consider local concentrations of hourly $PM_{2.5}$ and NO_2, regulated by NAAQS. ERG utilized DB4IoT's polygon tool using waypoint geometry to define the study area for the analyses. The polygon tool allows users to draw a unique shape around specific map features or geographic areas to define the study area. The data are then filtered so that only information from data points within the limits of the polygon are used in the analyses. As the focus of this analysis was comparing traffic congestion and emissions on a NFL Denver Broncos game day at Mile High Stadium (Sunday, October 13th) versus a non-game day (Sunday, October 6th), ERG drew a rectangular polygon to define a study area that includes Mile High Stadium, surrounding parking lots and segments of the North Valley Highway and W Colfax Ave (Figure 13).

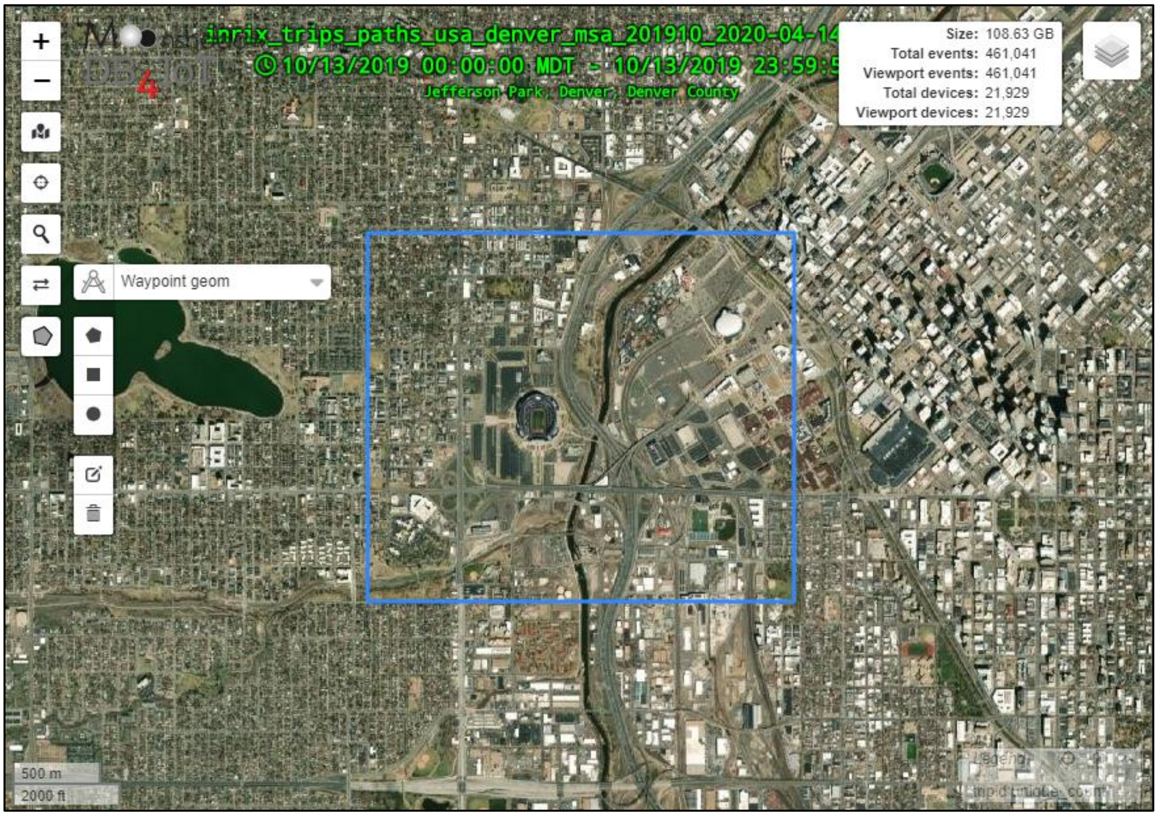

Figure 13. Mile High Stadium Polygon (Study Area)

ERG used Moonshadow to run analyses on trip counts, average speed, and trip origins during and immediately after the Broncos game (4:00 PM - 8:00 PM), and compared to the same time for the non-game day. The resulting data was then used to run a MOVES2014 + AERMOD analysis to estimate emissions concentrations between the hours of, which captures the post-game traffic "bump". Figure 14 depicts an example analysis for trip count and average speed data from within the Mile High Stadium polygon study area on October 13th, 2019. Variations in trip count are represented by line segments of varying thickness (i.e. more heavily trafficked roads are represented by thicker lines), and average speed is represented by a gradient of colors defined in the legend. The thick line in the middle of the figure is I-25.

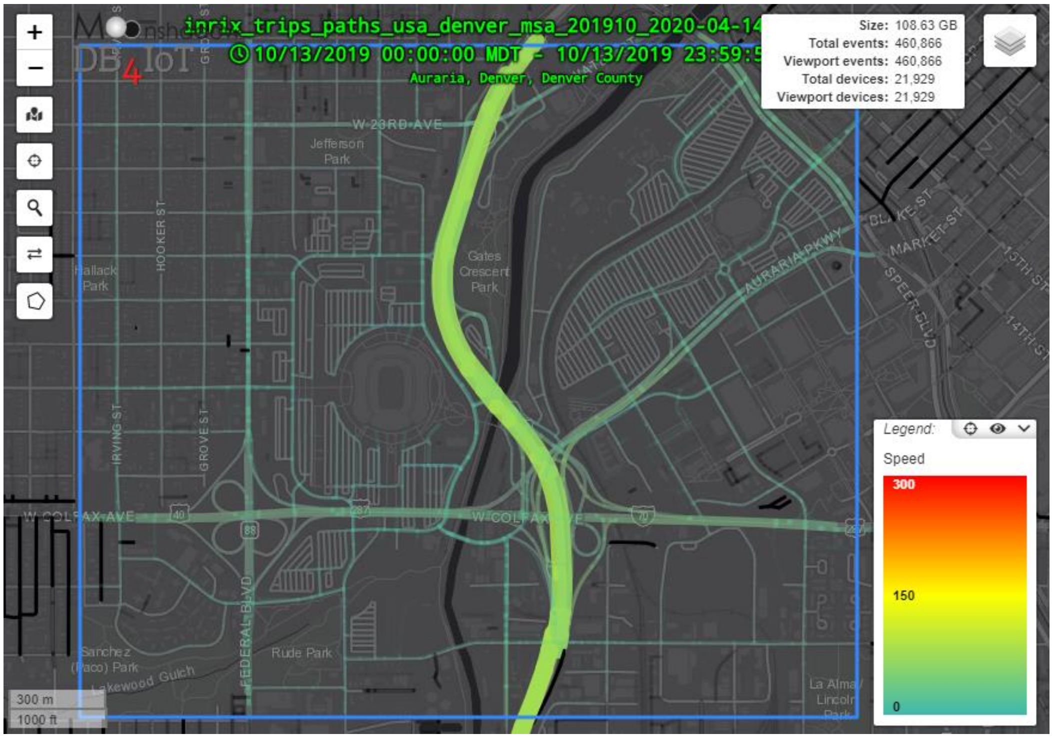

Figure 14. Trip Count and Average Speed within Study Area, Post-game (10/13 4-8pm).

ERG used DB4IoT to estimate changes in trip origins (starts), traffic volume, and vehicle speed, due to the NFL game, within the study area between the hours of 4:00 PM and 8:00 PM. These results were then used to estimate the increase in NO_2 and $PM_{2.5}$ along I-25 due to game traffic.

Trip Starts in Study Area

As described above, the trip origin geometry analysis generates the trip count of unique trip identifiers that originate within the study area during the specified time frame. Figure 15 shows the results of the trip origin geometry analysis, which used the following parameters:

- Year = 2019
- Dates = October 6th and October 13th
- Time Frame = 4:00 PM to 8:00 PM
- Study Area = Mile High Stadium polygon

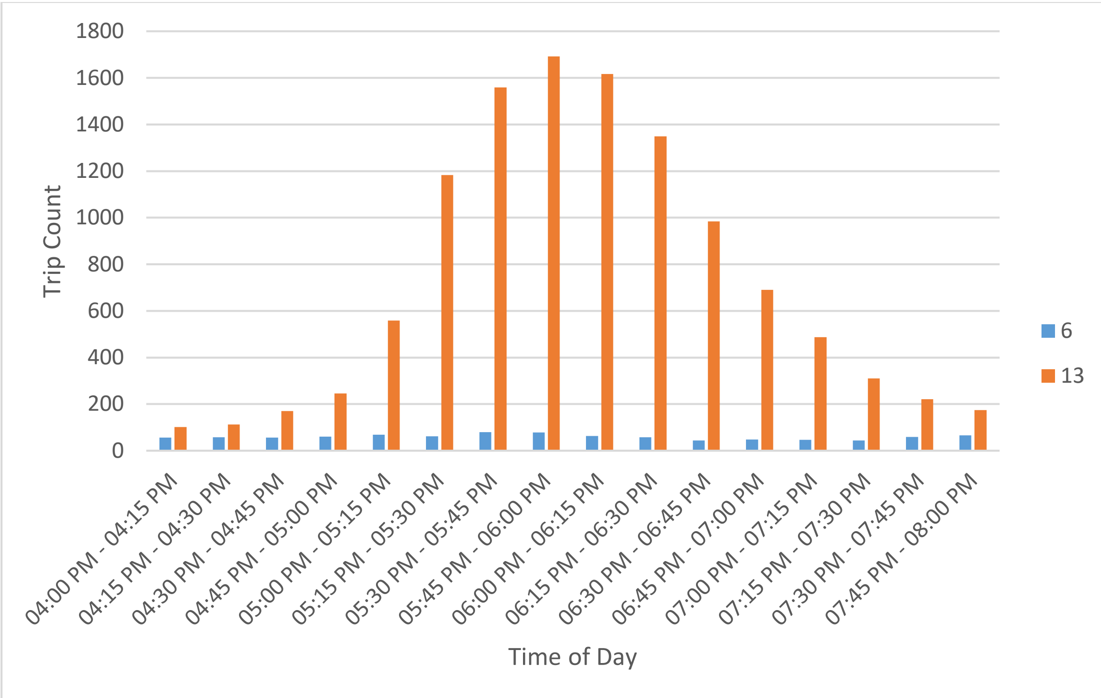

Figure 15. Number of Trips Originating within Study Area on October 6^{th} vs. October 13^{th}: Trip Origin Geometry Analysis

The origin geometry analysis generates the trip count of trips that originate within the study area between the hours of 4:00 PM and 8:00 PM. The results indicate that the number of trips originating within the study area increased by about 1,212% overall from October 6^{th} to October 13^{th}. These results again align with the increased traffic on game days (October 13^{th}), when thousands of people are traveling to and from Mile High Stadium, versus non-game days (October 6^{th}).

Emissions "Hot Spot" Analysis: I-25

An emissions "hot spot" analysis using MOVES and AERMOD was conducted on I-25 near the stadium (the major North/South highway bisecting the study area in Figure 14, to estimate the increase in hourly $PM_{2.5}$ and NO_2 concentration due to higher post-game traffic volume and corresponding traffic congestion. Raw INRIX trip volumes from 5-6pm on I-25 within the study area were estimated for October 13^{th} and October 6^{th} by summing unique trip counts on I-25 (North Valley Freeway) within the study area, which would include north and southbound traffic. The INRIX trips were then scaled to total volumes estimates using a Volume Factor of 5.2 percent, which was the aggregate average scaling factor for CDOT traffic monitors in Denver County on Sundays. This approach estimated that the game added about 14,000 vehicles to I-25 near the stadium relative to a non-game day. DB4IoT also provided a distribution of speeds on I-25 (Figure 16), reflecting increased congestion on game day. Using these inputs, the change in near-road concentrations of $PM_{2.5}$ and NO_2 were estimated with MOVES and AERMOD. The game traffic increased near-road NO_2 concentrations up to 23 ppb (on NAAQS 1-hr standard of 100 ppb), and $PM_{2.5}$ concentrations up to 5 µg/m^3 (on NAAQS 24-hr standard of 35 µg/m^3). Though not included in this project scale analysis, the impact of post-game trip starts presented in Figure 15 will compound the increases in emission concentrations.

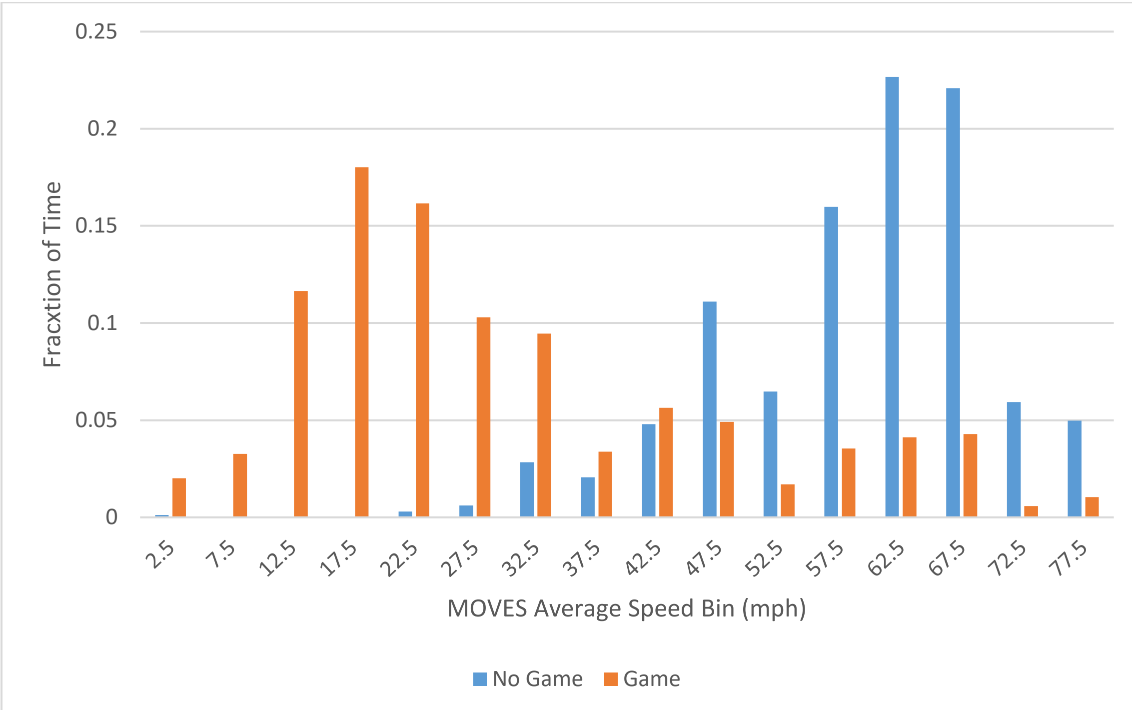

Figure 16. I-25 Speed Distribution 5-6pm: Game vs. No Game

This project scale analysis with Moonshadow DB4IoT demonstrates a major strength of telematics: the ability to pinpoint vehicle activity at a high spatial and temporal resolution, and to isolate traffic, emission, and air quality impacts of specific events. As the need to quantify hot-spot emissions continues to evolve, this case study shows that telematics can fill the need for improved localized data to support these analyses.

6. CASE STUDY ANALYSES WITH OTONOMO

The usefulness of Otonomo's data towards the E-131 project objectives proved limited, for the following reasons:

- The data provided were for all rental vehicles, making it unrepresentative for the purpose of broad vehicle emissions inventory;
- Otonomo's trip summaries did not include much of the data fields that appeared in the raw "points" file, and did not appear to include events identified as trips by ERG, calling into question their accuracy.
- It was difficult to associate data from the raw "points" file with what the vehicle was actually doing – e.g. turned off, turned on, stationary, moving.

Our work with Otonomo therefore focused primarily on trying to identify trips from raw "points" data, to generate a more accurate trip summary that included data fields of interest, such as odometer and fuel fill level.

6.a Creating a Trip Summary

Otonomo output is either in terms of raw pings of data (e.g., every two minutes), or summarized trips. Our evaluation of Otonomo's trip summary found many cases where seeming trips were not included; the Otonomo summaries also excluded many data fields that would be useful to have on a trip-summary basis for inventory modeling. ERG therefore developed an alternative method to identify trips from Otonomo's raw data, and to summarize data fields of interest. ERG's trip summary file increased the trip count by ~ 60 percent vs. Otonomo's summaries. ERG also attempted to identify vehicle idle events from raw points data.

Some caveats for the case study: almost all vehicles represented in the data are from car rental fleets and as such the data is not representative of the general highway-driving vehicle population. Also, it was unclear from the data if a vehicle ID in the database consistently represented a single vehicle, as Otonomo's data processing algorithms may assign multiple vehicle IDs to the same vehicle throughout a given time period. There was nothing in the data to indicate clearly if this occurs or not, but there are instances of similar vehicle make/model configurations with different vehicle IDs at nearly the same timestamp and location (lat/long), raising suspicion that vehicle ID was not a stable parameter in the database.

Trip Identification

Our case study focused on defining trips from raw Otonomo data on August 6th, 2019. Based on analysis of raw points data, ERG identified 30,087 candidate trip events based on changes in GPS lat/long; of these 18,676 events had corresponding changes in vehicle odometer, and were identified as likely trips. Notably, Otonomo's trip summary for the same time period produced zero trips, underscoring significant issues with Otonomo's trip summary process (as detailed below other days did produce trips, but at a much lower rate than trips identified by ERG).

ERG's algorithm in identifying trips is the following:

- Trips are defined as a chronological sequence of pings of a unique vehicle ID that indicate a change in location (change in latitude/longitude).
- Stops between trips are assumed if no change in location occurs in 5 minutes (assuming absolute stops at traffic lights or in jams take less than 5 minutes).
- Trips captured either start or end in the geographic 10-county area around Denver.
- Many captured trips are likely stationary events: They produced a slight variation in location from the trip start to end (lat/long), but no change in odometer (or a nonsensical negative change in odometer in one case). Eliminating these reveals moving vehicle events that can be reasonably interpreted as trips, or *likely* trips.

The lack of trips in the Otonomo trip summary file for August 6th, 2019 may be an extreme anomaly. For other days, the Otonomo trip summary file was populated, though with significantly lower trips than estimated using several variations of the ERG algorithm detailed above. This is exemplified with a comparison of a four-hour period (6-10am) on October 15th, shown in Table 17 The Otonomo trip estimate was generated for a geofenced rectangle approximating Denver MSA (i.e. the trip either starts, ends, or is entirely within in the defined area). ERG developed an alternate trip summary based on raw points in the same area, applying different filters to address uncertainty in whether the raw data indicated trips or not. From 4,236 candidate events, events with no significant change in distance within 5 minutes were filtered out (leaving 2,804 events); then a second filter was applied to select remaining events of at least one minute. The resulting ERG trip estimate, 2,759, is 60 percent higher than the 1,894 trips identified by Otonomo.

Table 17. ERG Processing of Otonomo Trips

Model (Oct 15, 6AM-10A)	In Denver Metro Area	Filter Out Likely Stationary Events	Filter Out Events < 1 minute
ERG Algorithm	4,236	2,804	2,759
Otonomo	1,894	1,894	1,894

Idle Event Identification

For the same set of raw data, ERG also worked to identify non-traffic idle events. This would be useful for emission inventories since idle-specific data, since MOVES plans to add “short duration idle” as an emissions process in future versions of MOVES, distinguished from in-traffic idle. From the August 6th points file, ERG identified 1,040 “stationary events”, with 77 of these being identified as likely idle events.

Identification of idle events was complicated by the fact there was not a good indicator of engine status in the raw data; though it was included as a field, our evaluation found that ping data wasn’t received when the engine was off, making it difficult to discern during stationary events whether the engine was truly on, or just reflecting the most recent engine-on ping.

Stationary events were therefore estimated by ERG as those events with distance < 0.1 mile (to allow for GPS error)[2], and duration greater than 5 minutes (to allow for traffic light stops and stopped congestion). To further isolate engine-on idle events, drops in fuel level of at least 1 percent over a 5 minute idle were assume to denote engine operation. Some filtering was required to exclude nonsensical data, such as anything above 30% fuel tank level consumed per hour seemed too extreme to be accurate.

[2] In this dataset of 77 events, only 2 events displayed a GPS change, and of those the median difference was 6.5 ft.

Summary statistics and observations for identifying idle events are summarized below:

- Summary statistics:
    - *Likely* idling events: 77
    - Duration (minutes): Median: 23.0 Average: 80.4
    - Fuel tank level decrease (percent of full): Median: -1.6% Average: -3.4%
    - Rate of consumption (percent of tank/hr): Median: 3.8% Average: 4.9%
- Isolation of stationary events difficult to discern because odometer readings might stay constant while calculated distance changed, making the event dependent on which metric was assumed to be accurate.
- Idling somewhat difficult to determine because fuel level changes do not always correspond to a significant change in odometer reading or calculated distance.
- Due to the above difficulties, reasonable idling events captured in a single day from this data set would not provide a strong dataset for analyzing idling. However, capturing likely idling events from a week or month might provide a solid sample size. The dataset would need further refinement to reduce outliers. For example:
    - Even a 10-20 percent-tank-level/hr consumption rate might be considered too fast a rate of idling consumption to be accurate; and/or,
    - A duration of > 30 minutes seems unlikely and may be an outlier from inaccurate distance/location data.

6.b Fuel fill distribution

The ability of engine-based data to provide fuel fill information will be very useful for improving evaporative VOC inventory estimates. The amount of vapor generated in a tank is highly dependent on the fill level of the tank; EPA certification procedures require 40 percent fill, an assumption carried over to MOVES to estimate the quantity of vapor vented from the fuel tank. Fuel fill level can be modified in MOVES input, but this is not required in MOVES technical guidance, and is not common due to the lack of real-world data to put into the model. The ability to gather fuel fill data from engine-based telematics would represent an important opportunity to improve emission inventories. To demonstrate how real-world fuel fill data would affect emission inventory estimates, a sensitivity analysis of MOVES2014b VOC emissions was conducted assuming 10 percent fuel fill level (near empty), and 90 percent fuel fill level (near full) to a typical July weekday in Denver County. Results are shown in Table 18 , for vapor venting only (+/- 50%) and total VOC (+/- 12-14%).

Table 18. Sensitivity of Denver County VOC to Fuel Fill Level (8/6/19)

Fuel fill level	VOC Tons/Day (Tank Vapor Venting Only)	VOC Tons/Day (Total)
40% (default)	1.78	6.81
10%	2.70 (+ 52%)	7.74 (+ 14%)
90%	0.96 (- 46%)	6.00 (- 12%)

This sensitivity provides context for distribution of fuel fill for August 6th shown in Figure 17 by bins of 10 percent. The majority of vehicles in the Otonomo sample are at least half full; 44 percent have more than 80 percent full. This would reduce VOC emissions relative to MOVES default assumptions; however, we expect this result is biased high because it is from a fleet of rental cars, which would tend to have higher fill levels than the average personal vehicle.

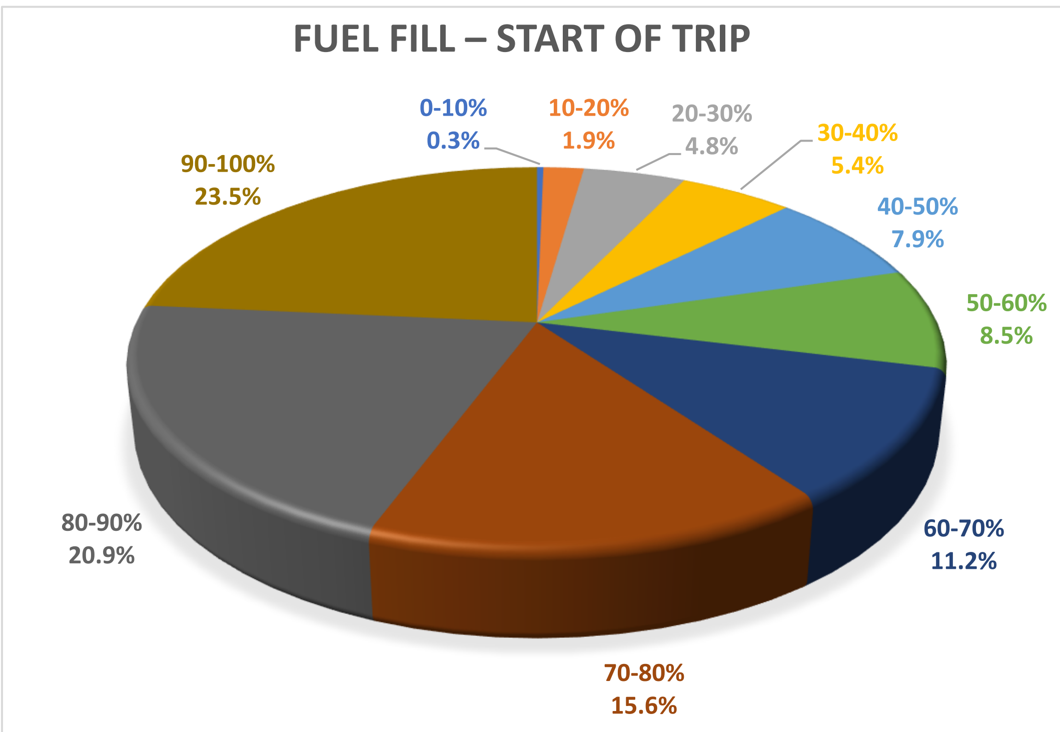

Figure 17. Fuel Fill Distribution for Otonomo Fleet, August 6th 2019

Overall our analysis of Otonomo data pointed towards the current limitations of engine-based data to help improve vehicle emissions inventories. With the right data fields, engine-based data could provide a wealth of information on both the vehicle (e.g. model year and technology, to help develop representative distributions), activity (e.g. key on/key off times), and engine parameters such as MIL codes, EV charge (where applicable), and RPM. Unfortunately, our evaluation of telematics data was not able to find a ready source of such data.

7. ASSESSING CAPABILITIES AND LIMITATIONS OF TELEMATICS

Through a broad market survey, in-depth communication with several telematics firms, and detailed analysis of telematics data from three different vendors, the work of this project sheds light on significant capabilities and limitations of telematics data for compiling vehicle emission inventories. While telematics has been applied in targeted ways for specific emission inventory inputs by federal and state modeling agencies, the broader application of telematics is hampered by a lack of focus in the telematics market on providing data targeted for compiling robust, representative emission inventories. Our general conclusions regarding capabilities and limitations from Task 1 market research and Task 2 case studies are summarized below.

Capabilities of Telematics for Emission Inventories

- Relative to traditional instrumented vehicle studies which have populated EPA and CARB emission inventories since the 1990s, location-based telematics data is a sea change with respect to sample size, geographic coverage, and temporal coverage of data. As an example, vehicle activity for EPA models through the early 2010s relied on data on a few hundred vehicles collected in 3 U.S. cities over about 1 week of time in the early 1990s; location-based telematics data can now provide an ongoing stream of relatively low-cost activity data from millions of vehicles year-round for most of 3,200+ counties in the U.S.

- Telematics can greatly improve the spatial distribution of vehicle trip activity, which will improve local emission inventories and regional air quality modeling, which is highly sensitive to this distribution. As an example, while total trip starts estimated with telematics in the Denver MSA for 2017 were very close to the 2017 NEI estimate region-wide, individual counties within the MSA varied from the NEI estimate by as much as 50 percent.

- Telematics can provide data in specific locations and times, as demonstrated in the case study for the Moonshadow DB4IoT platform. This allows ongoing capacity for high-resolution analysis of traffic and emissions impacts from specific events, and for better quantifying “hot spot” emissions as required by EPA and FHWA.

- Telematics can provide data on the impact of significant events on broader regional traffic as well – for example, severe weather or construction. By happenstance, this capability of telematics was underscored for this project by the ability to track the VMT change from COVID shutdown, and subsequent re-opening.

- Engine-based data can provide real-world data on important vehicle parameters, though the most useful parameters for emission inventories (e.g. MIL on status) could not be obtained for this project.

- Telematics firms are developing web-based front ends applications and establishing subscription-based access levels to provide a lower cost access for more aggregate (but still very useful) data.

Limitations of Telematics for Emission Inventories

- Though engine-based telematics data holds immense potential as a source of important vehicle parameter data, robust and representative sources of these data for light duty vehicles appear to be drying up. Sources of heavy-duty truck data, though outside the scope of this study, continue to proliferate for truck fleets due to electronic data logging requirements for trucks.
- For personal vehicles, LBS data sources are growing and appear to be the dominant source of telematics data into the future, especially from mobile phone data. Though the coverage of location data is immense, there are key limitations to these data for use in estimating emission inventories. These include:
    - Lack of specific vehicle information (model year, car/light truck, technology, fuel etc.)
    - Lack of engine operation data (e.g. key on/off time, engine status, instantaneous speed and acceleration, battery charge, accessory use, etc.)
    - Lack of complete traffic coverage, requiring estimates to “scale up” to total traffic volumes.
- Data samples, though large, are “passive”; unlike instrumented vehicle studies, there is no up-front study design to enable directly application of results to a larger, representative sample. In this way telematics data samples are more of a “catch all” rather than a conscious study design.
- The cost of telematics data may be prohibitive for state and local emission modelers depending on specific data needs. Based on discussions with vendors during this project, access to 1 years’ worth of telematics data at the roadway link level for an entire metro may cost well into six figures.
- No telematics firms are currently focused on providing data specifically for emission inventory development. As such, effort is needed to “shoehorn” existing data platforms into model inputs.

How Telematics Fulfills Agency “Wish Lists”

In this section capabilities and limitations are also assessed relative to the EPA and CARB “wish list” items for identified in Task 1, to provide an additional metric for evaluation. These are assessed against our findings in this study, shown in Table 19.

Table 19. Revisiting Telematics "Wish Lists" with Findings of E-131

EPA "Wish List" Item	Can telematics data be assessed for this study products address?
Data is representative (or stratified to allow scaling) of national activity in terms of geographic variation, vocational variation, vehicle age, technology, fuel type etc.	**Somewhat.** Geographic variation can be addressed. For other variables listed, key vehicle information is not available – i.e. vehicle age, technology, fuel type
Represents inactive vehicles	**No.**
Can distinguish different vocation of vehicles (gig fleet vs. personal, parcel delivery truck vs. utility truck)	**Somewhat.** StreetLight Data is working to isolate gig fleet travel, buses. Truck vocations not addressed in any source.
Can distinguish EVs from conventional ICEs	**No.** OBD should be able to distinguish this, but not available in data evaluated for this study.
Provides the location (county, road type) of driving (not just the zip code of residence)	**Yes.**
Identifies grade	**No.** Though for specific links, other datasets (e.g. TIGER) could be overlaid to provide grade in a GIS platform.
Identifies vehicle class (e.g. LD, HD)	**Somewhat.** Connected vehicle data can distinguish cars and commercial trucks. OBD should be able to distinguish this, but not available in data evaluated for this study.
Identifies speed	**Yes.** Though in aggregate only.
Identifies acceleration	**No.**

CARB “Wish List” Item	Can telematics data assessed for this study products address?
Mileage accrual rates for LD vehicles younger than eight years	**No**. Vehicle age or model year not available
Spatial distribution of truck travel	**Yes.** .
Trip level data with following fields:	
• VIN	**No.**
• Trip ID	**Yes.**
• VMT	**Yes.**
• eVMT for PHEVs	**No.**
• Date/time for trips start and end	**Somewhat**. Cannot pinpoint times, but within time block.
• Lat/long for trips start and end	**Yes.**
• Fuel consumption (gallons)	**Somewhat**. Otonomo provided fuel fill level, which would need to be paired with external tank size to estimate consumption.
• Electricity consumption (kWhr)	**No**
• MIL status along with all active SPNs (preferably emission related)	**No**

CARB "Wish List" Item	Can telematics data assessed for this study products address?
• Percent VMT < 25 mph / 25 – 50 mph / > 50 mph	**Yes**
• Percent time idling (speed <1 mph)	Somewhat. This was estimated by ERG from raw Otonomo data.
• For trucks, average bhp-hr/mi for that trip	n/a for this study

8. CONCLUSIONS

Several projects undertaken by EPA, CARB and CRC have used vehicle telematics data to update vehicle activity inputs for emission inventories. These prior projects have established that telematics data is not a uniform dataset, but rather a myriad of data types (e.g. location data from connected vehicles, location data from mobile phones, or engine data from specific vehicles) compiled and bundled via multiple sources, primarily private firms. The landscape of telematics data continues to change, as data sources, firms, and privacy constraints have evolved rapidly in the last five years. Under contract with CRC, ERG has evaluated the current state of the telematics field with respect to emission inventory development, identifying capabilities and limitations of the data for emission modelers, and identifying how telematics data could align better with the needs of vehicle emission inventory models.

ERG's evaluation of telematics for this project included a literature review, market survey including direct communication with several telematics firms, and detailed analysis of telematics data from three different vendors. Our findings confirm a growth in the application of telematics, in particular location-based data culled from cell phones. Our literature review identified 40 studies that used telematics data in emissions or transportation-related studies. The majority of studies used this location-based data (i.e. GPS "pings" from connected vehicles or phones converted to vehicle trips), and suggest that the sample sizes of such data are large enough to begin to estimate total vehicle miles travelled, a core input to vehicle emission inventories. Prior to this project, the majority of telematics applications for emission inventory models have focused on distributions of activity (e.g. speed distribution, temporal trip distributions, spatial distributions) rather than total activity. As signaled in our literature search, a significant development in telematics is sufficient coverage to estimate total activity (VMT) as well.

EPA and CARB have incorporated new telematics data into upcoming releases of MOVES and EMFAC respectively, and were surveyed for this project on future plans and telematics "wish list. EPA specifically expressed interest in information on representativeness of telematics samples, while CARB provided specific data fields of interest. However, our evaluation of telematics data sources for this project suggest that many of the Agency "wish list" items cannot yet be fulfilled with telematics data.

Case studies conducted in the Denver metro area on three telematics datasets purchased for this project confirmed that the location-based telematics can serve a growing range of emission inventory use cases, and use of these data may help improve emission inventory accuracy. Key capabilities include more accurate spatial and temporal distribution of start emissions; estimation of total VMT; and the ability to pinpoint vehicle activity at specific times and locations, which will facilitate more accurate project scale "hot spot" analysis and event-specific emissions. The most consequential finding from our case study was that in the Denver MSA, VMT from telematics data was 34 percent higher than that used in the 2017 NEI, with significant ramifications for emission inventory estimates. This raises the question as to whether telematics as a whole is able to quantify vehicle travel not well captured through traditional traffic counter or travel modeling means. Though not an intended case study at the outset of the project, the ability of telematics to rapidly track changes in travel patterns from

COVID shutdown and gradual re-opening during Spring 2020 proved one of the most illustrative benefits of telematics data.

Key limitations of location-based telematics stem from lack of specific vehicle information, resulting in difficulty assessing the representativeness of activity data. Though engine-based data showed great promise for improving emission inventories in initial telematics studies, sources of these data appear to be drying up, eclipsed by LBS. Other identified limitations are cost, and lack of focus from the telematics market on providing data directly for emission inventory models. This continues to pose a challenge for emission inventory modelers who want to tap into very large sample sizes of telematics data. As presented in the study, while the capabilities if the data are indeed promising, there are many limitations that need to be considered in using telematics data for emissions models. On the whole, we conclude that the capability of telematics continue to advance for the purposes of emission inventory modeling. The growth in LBS data will only continue to improve capabilities for spatial and temporal activity distributions that prior studies have focused on, while the emerging capacity to estimate total activity is a significant advancement. Identified limitations keep the application of telematics in check, however. It is our hope that telematics firms will begin to cater more directly to the needs of emission inventory modelers, to better harness the enormous potential of these datasets for refining vehicle emission inventory estimates.

9. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial and technical support of the Coordinating Research Council Emissions Committee, project leads Paul Loeper and Michael Viola, and CRC ; David Choi of EPA and Sam Pournazeri of CARB for providing agency input; Jim Sidebottom and Rob Dawson of Colorado Dept. of Public Health & Environment for input on Denver case studies; Colby Richardson of ERG for literature review support; and telematic firm staff for technical support including Eimar Boesjes, Wander Boesjes and Anita Van Asperdt (Moonshadow Mobile Inc.); and Ariel Bell and Kobi Elmuday (Otonomo Inc); and Jim Hubbell (StreetLight Data Inc).

10. REFERENCES

CARB 2020 *EMFAC2020x Updates*, CARB Workshop presentation, July 2020 Available online: https://ww2.arb.ca.gov/sites/default/files/2020-07/EMFAC202x_2nd_Workshop_07302020_ADA.pdf (accessed July 30, 2020)

CRC 2018. *Evaluating the Sensitivity of MOVES2014a to Local Start Activity Data*, Report prepared by ERG, December 2017. Available online: http://crcao.org/wp-content/uploads/2019/08/CRC-A106-Final-Report_Dec2017.pdf

DenBleyker et al 2019. *Using Telematics Data to Improve the National Emissions Inventory,* TRB Annual Meeting Paper 2018

Fan et al. 2019 *Using big GPS trajectory data analytics for vehicle miles traveled estimation* TRR Part C 2019

National Academies of Sciences, Engineering, and Medicine (NAS) 2019. *Guide to Truck Activity Data for Emissions Modeling*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25484.

StreetLight 2020. "How to Get Regional VMT Using O-D Analysis", online support article ,https://support.streetlightdata.com/hc/en-us/articles/360038900092-How-to-Get-Regional-VMT-using-an-O-D-Analysis (accessed Sept 29, 2020)

U.S. EPA, 2016. *Population and Activity of On-road Vehicles in MOVES2014.* Assessment and Standards Division, Office of Transportation and Air Quality. EPA-420-R-12-003a. March. Available online: https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100O7PS.pdf (accessed June 30, 2020)

U.S. EPA, 2019. *Updates to EPA's Motor Vehicle Emission Simulator (MOVES). Highlights from presentation to Coordinating Research Council Real World Emissions Workshop.* March 11. Available online: https://www.epa.gov/sites/production/files/2019-06/documents/02-moves-update-crc-2019-04-10.pdf (accessed July 1, 2020)

U.S. EPA, 2020*. Supporting Data for the 2017 National Emissions Inventory, MOVES County Databases for All Counties.* April. Available online: ftp://newftp.epa.gov/air/nei/2017/doc/supporting_data/onroad/CDBs_for_all_counties/ (accessed June 30, 2020)