CRC Report No. A-106

EVALUATING THE SENSITIVITY OF MOVES2014A TO LOCAL START ACTIVITY DATA

Final Report

December 2017

COORDINATING RESEARCH COUNCIL, INC.

5755 NORTH POINT PARKWAY • SUITE 265 • ALPHARETTA, GA 30022

Evaluating the Sensitivity of MOVES2014a to Local Start Activity Data: CRC A-106

FINAL REPORT

Prepared for:

Coordinating Research Council

Prepared by:

Eastern Research Group, Inc.
StreetLight Data, Inc.

December 26, 2017

ERG Project No.: 4020.00.002.001

Evaluating the Sensitivity of MOVES2014a to Local Start Activity Data: CRC A-106

FINAL REPORT

Prepared for:

Coordinating Research Council

Prepared by:

John Koupal
Doug Jackson
Scott Fincher
Allison DenBleyker
Eastern Research Group, Inc.
3508 Far West Blvd., Suite 210 Austin, TX 78731

Laura Schewel
Neil Vaz
Neal Bowman
StreetLight Data, Inc.
135 Stillman St. San Francisco, CA 94107

December 26, 2017

Table of Contents

List of Tables

List of Figures

Acronyms and terms

ARB	California Air Resources Board
CHTS	California Household Travel Survey
CO	Carbon Monoxide
CRC	Coordinating Research Council
EMFAC	California ARB’s vehicle emissions model
EPA	U.S. Environmental Protection Agency
GPS	Global Positioning System
LBS	Location-Based Services
LDV	Light-duty vehicle (i.e. a car)
LDT	Light-duty truck (SUV, crossover, minivan, pick-up)

MOVES	U.S. EPA's vehicle emissions model
NEI	National Emissions Inventory
NHTS	National Household Travel Survey
NOx	Oxides of Nitrogen
OBD	On-Board Diagnostics
PM	Particulate Matter
TAZ	Traffic Analysis Zone
TDM	Travel Demand Model
VMT	Vehicle Miles Traveled
VOC	Volatile Organic Compounds
VSP	Vehicle Specific Power

Executive Summary

The A-106 project was developed by CRC to continue evaluation of emerging data sources that can be used to improve emission inventories with MOVES2014a, focusing on the National Emissions Inventory, but with applicability to regional, state and local inventories as well. The A-106 project builds on A-100, specifically to assess telematics data for improving vehicle start-related activity at the local level. Trip start activity, characterized in MOVES by the number of vehicle starts as well as the time between trip ends and starts (soak time, which affects emissions as well) has long been difficult to obtain on a local scale. The emergence of telematics, explored in detail as part of A-100, provides a new opportunity for gathering vehicle start activity at the individual county, or even sub-county, level. The objective of A-106 is to assess the use of telematics for improving local start activity inputs for MOVES at a pilot level, including emissions sensitivity and comparison to other data sources used to estimate start activity and to allocate start emissions spatially and temporally for air quality modeling. For this project, ERG teamed with StreetLight Data, Inc. to provide data and direct support for achieving project objectives. StreetLight Data is a mobility analytics provider who compiles various types of data derived from mobile devices to support transportation planning and policy analysis (www.streetlightdata.com). StreetLight mixes and matches location data derived from navigational GPS devices (currently from partner INRIX) and Location-Based Services, culling data from hundreds of applications operating on smart phones (currently from partner Cuebiq).

Task 1 of the project reviewed those sources of telematics data that exhibited the potential to improve MOVES inputs related to vehicle starts, such as starts per day, starts per vehicle, temporal distribution of starts, and soak distribution. Our review focused on two types of sources: GPS data that tracks movement only, and engine data loggers that collect activity and engine parameters for individually logged vehicles. Both sources have pros and cons with respect to MOVES start activity data. The main benefit of GPS-only data is sample size, and the ability to “geo-fence”, i.e. focus on trip activity within a specific geographic region, regardless of whether a vehicle is from within the area, or travelling from another area. The benefit of engine logger data is that it represents precise measures of activity and time between key-on and key-off. These can only be inferred from GPS data based on movement. However, it is difficult to use engine logger data to broadly characterize activity within a specific area, since this information is available only for individual vehicles, rather than for locations.

Under Task 2, StreetLight Data processed trip metrics at the census tract level for three counties. ERG then processed the data into MOVES inputs, and compared the results to MOVES default trip activity. As of December 2016, StreetLight’s data repositories process analytics for nearly 35 million users, or about 10 percent of the adult U.S. population, and about 12 percent of commercial truck trips. StreetLight Data currently uses one major navigation-GPS data supplier, INRIX, and one Location-Based Services (LBS) data supplier, Cuebiq. Mobile phone apps providing LDS data include those for couponing, dating, weather, tourism,

productivity, weather, and many more apps which utilize their users' location in the physical world as part of their value. The apps collect anonymous user locations when they are operating in the foreground, and also collect anonymous user locations when operating in the background.

StreetLight Data accessed LBS to provide aggregated passenger vehicle trip metrics (trip index, starts per vehicle, and dwell times) by census tract, month and hour in three urban counties: Cook (IL), Fulton (GA), and Clark (NV). ERG then processed the metrics into county level MOVES inputs and compared the sensitivity of the MOVES model emissions outputs to results obtained from using inputs derived from other independent data including MOVES defaults, Engine-based telematics data from Verizon, National Household Travel Survey, and the California Household Travel survey. This comparison found that the StreetLight Data starts per vehicle metric were consistent with NHTS and CHTS results, but about one start per day higher than the Verizon start/vehicle estimates.

Regarding comparison to MOVES defaults, Figure ES-1 provides the average starts per vehicle and dwell (soak) times based on StreetLight Data metrics for the three counties, compared to MOVES defaults (expressed as a range of where actual soak times might fall within the soak time "bins" used by MOVES). The counties exhibit similar trends, with the StreetLight Data metrics showing about one start/day less than the MOVES defaults, but longer average soak times.

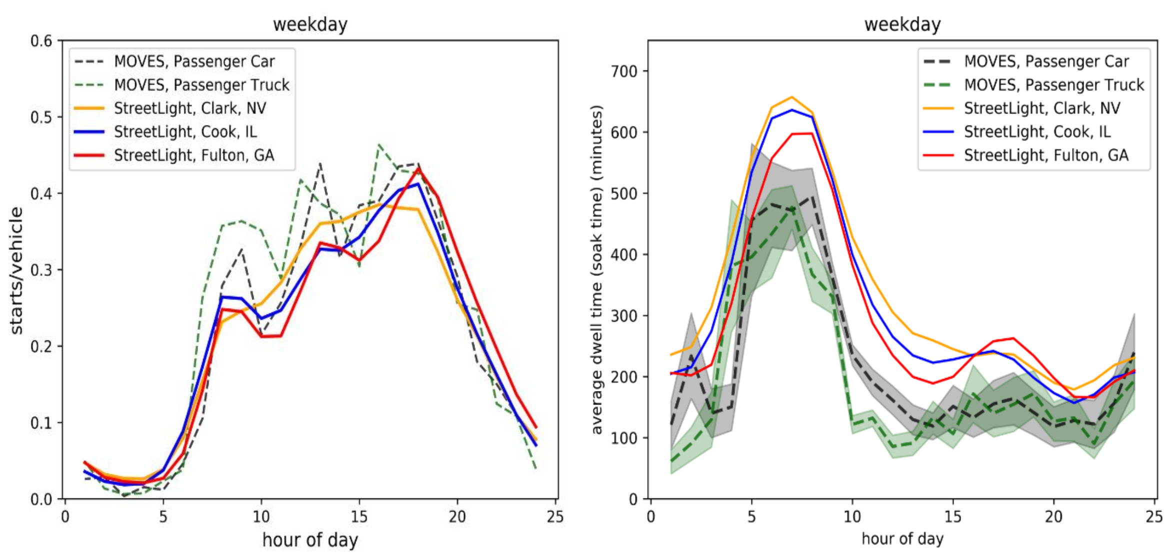


Figure ES-1. Average starts/vehicle and dwell times for 3 counties vs. MOVES defaults

Replacing default start activity with StreetLight Data metrics in MOVES resulted in an overall 10-15 percent increase in start-related emissions for cars and light trucks, as shown in ES-2, using Cook County, IL (Chicago) as an example. From this it was concluded that the emissions impact of the increased soak times represented in the StreetLight Data outweighed the impact of the lower number of starts/vehicle in this sample. While daily emissions were consistently higher in all three counties with the StreetLight Data metrics, hourly differences varied, though large percent increases were generally the rule for off-peak hours.

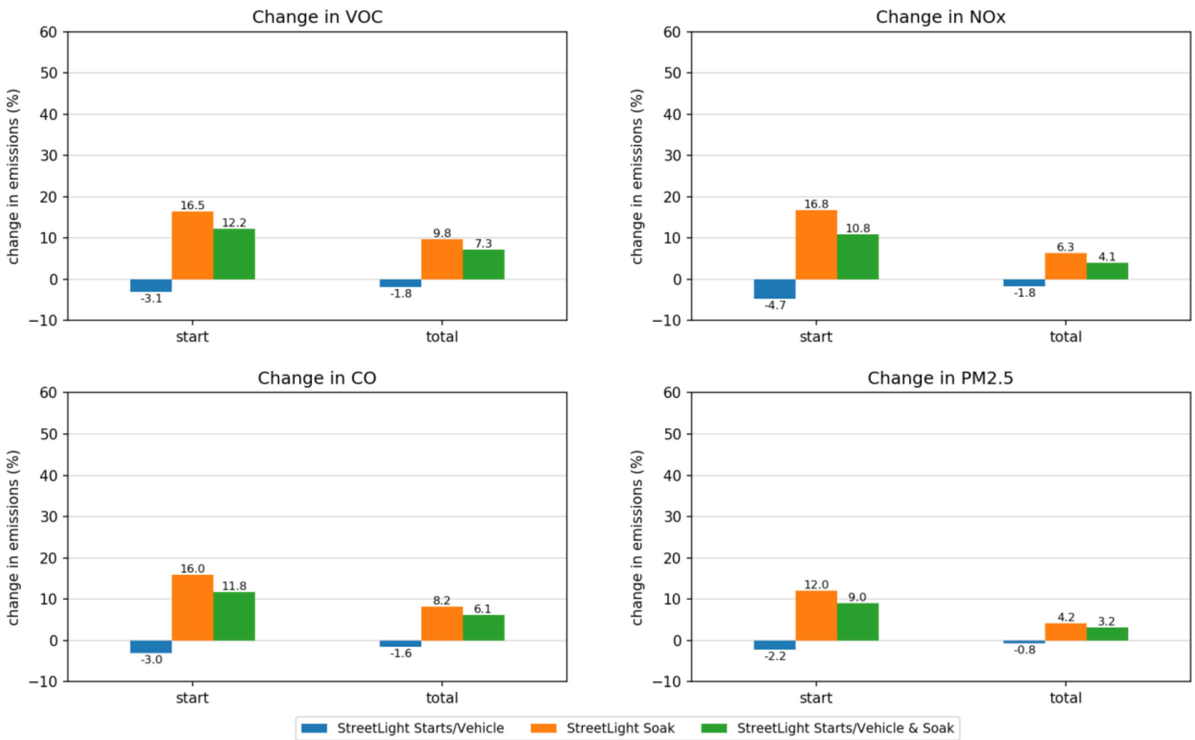


Figure ES-2. Change in Cook County weekday emissions using StreetLight Data metrics

ERG analyzed the variability in StreetLight Data metrics vs. U.S. Census data on human population, vehicle population, and housing density, as well as Travel Demand Model (TDM) trips origin estimates, all at the census-tract level. This assessment showed that the StreetLght Data correlated best with TDM trips at the census tract level, suggesting that telematics can be useful for validating TDMs, and can provide trip metrics for areas not using TDMs.

ERG also used the tract-level data to assess spatial distribution of start activity. The spatial distribution of StreetLight Data metrics was compared to human population, vehicle population and TDM trips. Again, the spatial distribution of the telematics data agreed well with TDM trips, but less so with human or vehicle population. This has important implications for EPA's air quality modeling, which allocates county-level vehicle start emissions to sub-county grid cells using human population. Figure ES-3 shows a map of the quintile difference between StreetLight Data's trip index and human population in central Clark County, which can also be read as the change in emissions intensity if the telematics data were used instead of human population to allocate emissions. As shown, relative to human population, the use of telematics data would shift start emissions to central Las Vegas which includes the airport, university and the Las Vegas Strip.

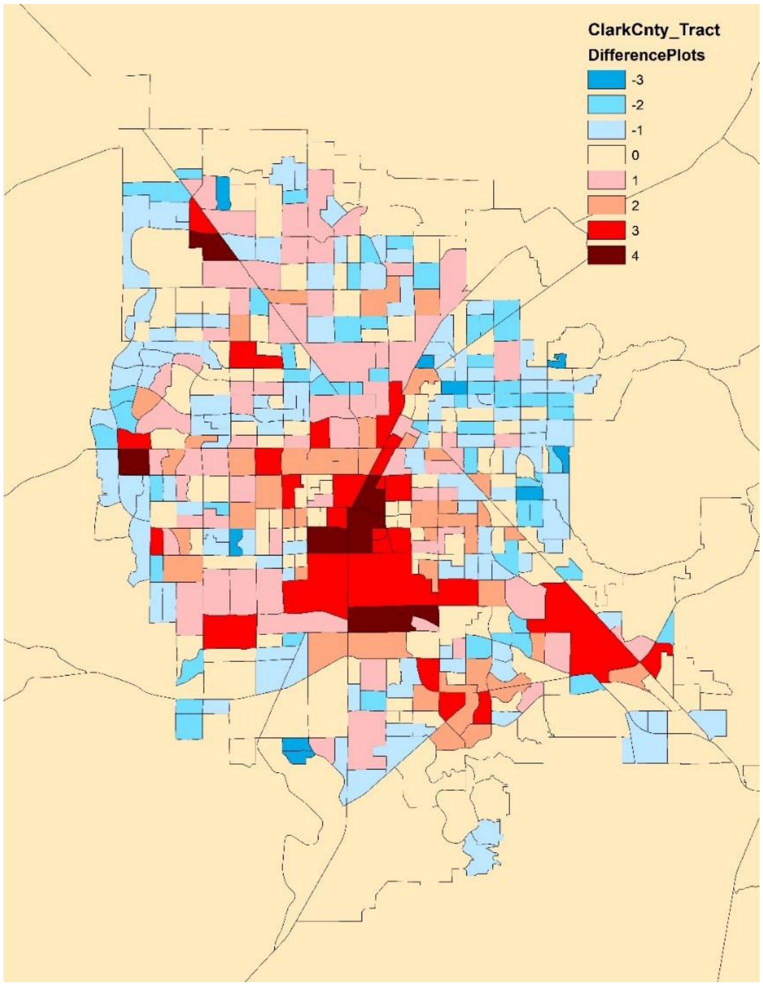


Figure ES-3. Difference in Spatial Distribution of Vehicle Starts for Telematics vs. Population

The findings of this pilot 3-county study could be extended to broader areas, given the scope of LBS data across the U.S. and abroad. Potential applications for national emissions and air quality modeling could include: developing local-level inputs for the NEI as was done for A-100 with vehicle speeds; updating air quality modeling surrogates using telematics in place of human population; or updating temporal profiles of start activity (by month, day or hour) for emissions and air quality modeling. These improvements could be applied by local, state and regional modelers as well.

1.0 Introduction

The A-106 project was developed by CRC to continue evaluation of emerging data sources that can be used to improve emission inventories with MOVES2014a, focusing on the National Emissions Inventory (NEI), but with applicability to regional, state and local inventories as well. Previous related projects have focused on assessing state-supplied inputs for the NEI (A-84);[1] improving the selection of default inputs for the NEI (A-88);[2] and using telematics data to further improve vehicle speed and temporal VMT inputs that are difficult to obtain from traditional sources at the local level for the NEI, (A-100).[3] The A-106 project builds on A-100, specifically to assess telematics data for improving vehicle start-related activity at the local level. Start emissions are estimated from MOVES2014a runs to contribute about 80 percent of car and light truck VOC exhaust emissions, and about 40 percent of CO and NOx emissions in 2017. As controls continue to reduce emissions from gasoline vehicles generated during warmed-up operation, the share of total inventory contributed by start emissions is projected to increase. Trip start activity, characterized in MOVES by the number of vehicle starts as well as the time between trip end and subsequent trip start (soak time, which affects emissions as well) has long been difficult to obtain on a local scale. The emergence of telematics, explored in detail as part of A-100, provides a new opportunity for gathering vehicle start activity at the individual county, or even sub-county, level. The objective of A-106 is to assess the use of telematics for improving local start activity inputs for MOVES at a pilot level. This assessment includes an evaluation of emissions sensitivity and comparison to other data sources used to estimate start activity and to allocate start emissions spatially and temporally for air quality modeling.

For this project, ERG teamed with StreetLight Data, Inc. (www.streetlightdata.com) to provide data and direct support for achieving project objectives. StreetLight Data compiles data from mobile devices to support transportation planning and policy analysis. StreetLight mixes and matches location data derived from navigational GPS devices (currently from partner INRIX) and Location-Based Services, culling data from hundreds of applications operating on smart phones (currently from partner Cuebiq). ERG is working with StreetLight Data to provide trip-start activity data for medium and heavy-duty trucks as part of an ongoing project sponsored by the National Cooperative Highway Research Program (NCHRP), known as NCHRP 8-101.

ERG approached the project in four separate tasks, presented as individual sections in this report. Task 1 reviewed telematics data sources with respect to their applicability for producing start-related inputs for MOVES. Under Task 2, StreetLight provided pre-aggregated telematics data for three counties used in A-100 pilot analysis (Cook County IL; Fulton County GA; Clark County NV). The data were then analyzed and further aggregated by ERG to populate MOVES start-related inputs Under Task 3, these StreetLight-derived MOVES inputs were compared to MOVES defaults and several independent sources of census, trip and spatial data. Under Task 4, ERG performed an emissions sensitivity analysis to quantify the impact the telematics data has on MOVES emissions relative to default values. This work included a sensitivity analysis to assess the emission impact of variability in start activity across one county.

2.0 Task 1: Review of Telematics for MOVES Start-Related Inputs

This section documents ERG's review of telematics data sources which have the potential to improve MOVES input data related to vehicle starts, such as starts per day, starts per vehicle, temporal distribution of starts, and soak distribution. The purpose of this review is to discuss pros and cons of different data sources for the MOVES user community. ERG has direct experience with several telematics vendors through projects for EPA, CRC and NCHRP, including providers of GPS-based data (tracking vehicle movements) and OBD-based trip data (logging engine operation), which have fundamentally different characteristics with respect to quantifying vehicle starts.

While the remaining tasks for A-106 focused on three pilot counties, this review considers broader applications such as NEI or SIP modeling. Our evaluation: (a) considers whether data sources are robust enough to provide unique data at the county level across the U.S., (b) compares the spatial and temporal resolution of different datasets, and (c) reviews what vehicle classes are covered. The review also considers whether data can be obtained directly from commercial vendors, or in aggregated form. For each data source evaluated, we discuss the supplemental work required to translate a vendor's data into MOVES start inputs.

2.1 Overview of Telematics Data Sources

We define two broad categories of telematics data for the purpose of considering MOVES start-related activity. The first, which we term "GPS only", tracks movements of GPS sensors either affixed to vehicles (e.g., connected vehicles), or "along for the ride" via GPS navigation units (e.g. Garmin, TomTom) or mobile phones. In total, this pool of data would include data culled from connected vehicles, mobile phones, and location-based services (LBS). A benefit of GPS only data is that third party providers focused on marketing information for transportation applications do the work of mapping the raw GPS signals, and parsing it into trip-based metrics.

The second category, which we term "Engine/GPS", compiles data from vehicles instrumented with portable loggers which gather engine parameters from the vehicle's Controller Area Network (CAN) Bus / On Board Diagnostic (OBD) port. These units typically include GPS sensors, so GPS data can be provided along with engine parameters such as key on/key off times, vehicle speed, engine speed and engine load. Because these monitors are linked with specific vehicles (unlike mobile phones), they also provide details on vehicle make, model year, engine parameters, etc. However, most applications within this general category are focused on monitoring vehicle activity, e.g. driver safety for insurance purposes, and do not currently map or process available GPS signals to the same extent as those which rely on GPS-only data.

The merits of both types of data sets are discussed based on the experience of ERG with specific companies: TomTom, StreetLight Data, Verizon and VNomics. This review does not cover all firms and thus the breadth of telematics capability, but provides a general representation of telematics products available, and the tradeoffs between GPS-only and Engine/GPS sources Through previous work on CRC A-100 and with EPA, ERG has reached out to INRIX, Waze and Progressive Insurance, who either declined participation or did not reply. Our overall lesson from this is that, though a mountain of telematics data now exists, not all telematics companies are equipped or interested in making these data available for custom applications such as emissions modeling.

2.1.1 "GPS-Only" Data

GPS data are collected from GPS devices in "connected" cars (cars with navigational systems), commercial trucks with GPS fleet management systems, and smart phones that have navigational apps that utilize the GPS chip in the phone. These data track movement based on GPS "pings". LBS datasets are a separate data stream that tracks the location of smart phones via apps that provide locations, such as user "check-ins" – as discussed on Section 3, LBS significantly increases the amount of data that can be considered for characterizing trip activity. StreetLight estimates their LBS data stream covered about 35 million users as of the end of 2016, with coverage continuing to increase since then.

GPS-only data are available from third-party vendors who compile raw data and process it for specific applications, such as navigation or transportation planning. In ERG's experience, these vendors are not able to sell raw GPS data, due to privacy and other restrictions, but are amenable to providing (for a fee) custom aggregated statistics as needed by MOVES. Part of the cost of the aggregated data is the custom work required by the vendor to process raw data into the desired MOVES-based aggregate statistics. Once custom programming is in place, however, in theory the price of data purchase will come down. A benefit for the vendor of undertaking the initial custom work is the prospect that other MOVES users will want to purchase data aggregated in a similar way, for their specific modeling domain.

Example vendors of GPS-only data are TomTom and StreetLight Data. A discussion of ERG's experience with each with respect to MOVES inputs is below.

2.1.1.1 TomTom

GPS data are collected by TomTom via their stand-alone navigation units, the TomTom smartphone application, and external GPS providers. Some users of TomTom units give permission for TomTom to collect and store users' personal (anonymous) data on central servers. The data are collected while the GPS unit is on, either in map or navigation mode. As long as the device is turned on, it is gathering data. A unit's GPS tracks are delivered to TomTom servers when data are collected either over the cell network as a "live" feed or as a "non-live" stream of data when the user connects to receive software or map updates.

TomTom's data collection began in January 2008. The data have been collected continuously, and the database currently has over 1 trillion data points. Since all U.S. drivers do not use a TomTom GPS unit or app and users who "opt in" are self-selected, biases could exist in the data that are collected. Anecdotally, drivers that own GPS units are less likely to use them when they drive in familiar areas in comparison with unfamiliar areas. TomTom data are obtained from units on all road types but at this time the data do not distinguish source types. Since TomTom devices are portable, they are not able to capture vehicle information. TomTom suspects that "virtually all" of their data is collected from the use of their devices in vehicles which are light-duty cars, trucks, and vans. TomTom data are obtained from all areas of North America where vehicle owners/operators with TomTom GPS units drive. There are some areas where their data counts are low, but for urban areas TomTom states that they have a large quantity of data.

TomTom City is an online application that processes GPS signals into real-time traffic data and congestion metrics, available online for multiple cities around the world (for example, New York City - see Figure 1).

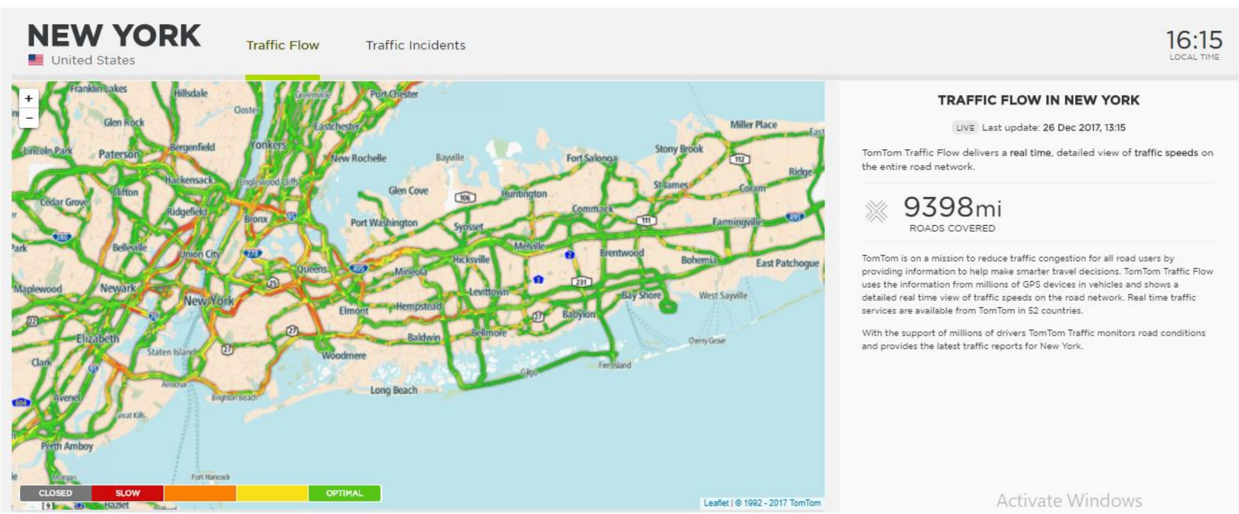


Figure 1. Real-Time Traffic Map of New York City from TomTom City
(https://www.tomtom.com/en_gb/traffic-news/new-york-city-traffic/traffic-flow)

In prior work, ERG has used TomTom data to help update average speed distributions for MOVES. TomTom appears focused on using data to assess traffic congestion, rather than trip metrics, and is not considered a promising source for MOVES start-related inputs. This is a general issue with telematics data focused on routing and traffic congestion patterns – the data may not retain trip elements necessary to feed MOVES start inputs.

2.1.1.2 StreetLight Data

StreetLight Data compiles various types of data derived from mobile devices to create analytics that support transportation planning and policy analysis. StreetLight mixes and matches location data derived from navigational GPS devices (currently from partner INRIX) and LBS (currently from partner Cubiq). StreetLight compiles data from billions of trips created from archival, anonymous, trace data from millions of GPS devices. The devices include smart phone applications, in-dashboard car navigation systems and commercial fleet management systems. The data have a spatial precision of approximately 5 meters, a high frequency sampling rate, and the ability to separate commercial and personal vehicles. StreetLight is also able to provide data for passenger vehicles, medium truck and heavy truck streams.

While some of the data sources are similar between StreetLight and TomTom, StreetLight uses the GPS data in part to process vehicles trip metrics for use in transportation planning, as well as roadway operation. For example, StreetLight's data are commonly employed to develop origin/destination matrices used in traffic planning. This emphasis on trips is a good fit for the objectives of A-106.

StreetLight provides trip results as a normalized "trip index". The trip index values do not represent absolute number of trips or vehicles. Rather, the values represent the relative amount of trips, and are weighted based on the density of reporting devices relative to human population. Trip index values are comparable within vehicle type, and across spatial and temporal categories. For U.S. Projects, the trip index value is normalized by adjusting the number of trips in StreetLight's data sample to the actual number of trips in a region around Sacramento CA, as derived from the measurements published by the California Department of Transportation; Sacramento is used because it has several well-maintained loop counters there, and does not have extreme seasonal variation in traffic.

An example of the normalization process is as follows: if in May 2017 there were 100 StreetLight trips at a normalization location, while a physical traffic sensor detected 1,000 trips, the May 2017 StreetLight trips would be normalized (scaled up) by a factor of 10. If in June 2017 there were 200 StreetLight trips at a normalization location, while the physical sensor detected 1000 trips occurred, June trips would be normalized by a factor of 5. Thus, the normalization factor smooths out variation caused by the StreetLight sample size increasing or decreasing to allow for comparisons over time.

The normalization to trip index allows StreetLight to capture monthly and seasonal variation more accurately, even as their sample grows. The implication for MOVES-based analysis is that StreetLight is not a source of total trip starts, but rather relative measures such as starts/vehicles, and distributional inputs such as soak time, temporal distributions, and spatial distribution.

Because A-106 focused only on 3 counties, StreetLight Data's analysis provided fine levels of detail both spatially (census tract level) and temporally (by hour, day type and month). However, based on their work under A-100, StreetLight has the capability to cover the entire continental U.S. The data are also readily geofenced, making spatial analysis an intuitive and direct use of the data.

2.1.1.3 GPS-Only Data Summary (Using StreetLight as an example)

Temporal resolution: Datasets are not based on a fixed fleet of vehicles; mobile phones are not associated with individual vehicles, and even connected vehicles with hard-wired GPS enter and leave the data pool. However, the sample size is large enough to provide data for each month, day and hour. For StreetLight, trip data are normalized to account for fluctuating sample size, to allow comparison across time,

Data available: Raw trip data or specific locations are not available, due to privacy concerns. Aggregated data have been provided for average speed (A-100), VMT distribution by month/day/hour (A-100), starts/vehicle by daytype/hour (NCHRP 8-101 and CRC A-106), and soak time distribution by day type/hour (NCHRP 8-101 and A-106).

Vehicle classes covered: passenger vehicles (including cars and light trucks), medium-duty trucks (14K-26K), heavy-duty trucks (>26K).

Sample rate: There are low resolution (0.01-.033 Hz, or 1 ping every 10-30 seconds) and high resolution (1 Hz) samples for Passenger vehicles. MD and HD trucks sample at 0.0167 – 0.0056 Hz (1 ping every 1-3 minutes).

Sample size: > 1,000,000

Vehicle information: Vehicle category only

Geofencing: Because GPS is the "currency" of this data source, the ability to geofence (isolate specific boundaries from which to cull data) is a natural byproduct. This is evaluated thoroughly in Task 2, since StreetLight was able to produce trip metrics at the census tract level.

2.1.2 Engine/GPS Data

2.1.2.1 Verizon

Verizon Telematics contracts with one or more vehicle insurance companies to provide telematics data for evaluating driver behavior as a way to determine risk. Drivers who subscribe to the datalogging program receive small, palm sized dataloggers that plug into their vehicles' OBD diagnostic port and log data from the vehicles' OBD data stream for a period of several months. These data are provided to the insurance company, but also are logged by Verizon Telematics, scrubbed of personal information, and marketed for use in driver and vehicle research.

Supported by ERG, EPA has worked with Verizon to purchase data on about 47,000 vehicles registered in five states. The data are in raw form, and include only engine data, no GPS. ERG and EPA have analyzed trip metrics from the data including starts per day, soak distributions and miles per day. (Figure 2 provides an example of the starts per day per vehicle information calculated using this data source.) While the data can provide precise estimates of these metrics, the only spatial feature available is where the vehicle is registered. Thus, there is no precise way to determine where vehicle activity takes place; and, even if isolated to a specific area, the data cannot account for the activity of "migrating" vehicles in or through a specific locale. EPA's analysis of Verizon data thus far has focused only at the state level, as these data cannot be used to pinpoint individual trips at even the county and certainly sub-county level.

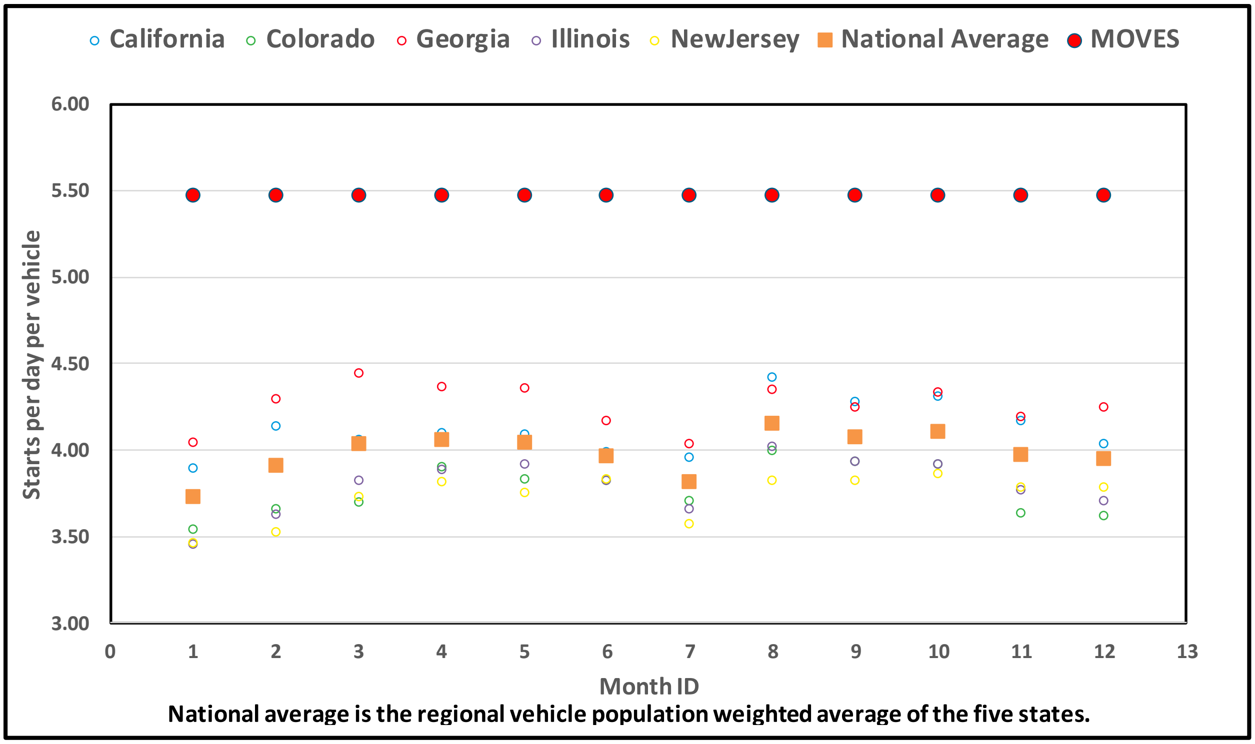


Figure 2. Starts/Vehicle/Day derived from Verizon Telematics Data for 5 States (LDV Weekdays)

Source: Brzezinski et. al, *Using Vehicle Telematics for MOVES Activity Input*, Presentation at International Emission Inventory Conference, August 2017

Through previous work, ERG also identified that the "Snapshot" datalogger program used by Progressive insurance is to that of Verizon. Progressive sells certain data summary and analysis reports from this program, but they do not sell raw data for their clients to analyze. ERG's contact at Progressive indicated that they were unwilling to sell the raw data related to their Snapshot program to ERG in support of this work.

2.1.2.2 VNomics (Heavy Duty Trucks)

VNomics (www.vnomicscorp.com) is a firm focused on helping heavy-duty truck fleets reduce fuel consumption. Using engine data loggers with GPS that connect through the electronic control module (CAN bus), information is gathered on latitude/longitude and engine parameters 1Hz level for truck speed, throttle position, fuel rate, RPM and torque. Basic vehicle information is also available. The data are routed to central servers that allow fleets to analyze fuel consumption and contributing factors to fuel wastage, including specific driver actions. The system includes real-time driver coaching to optimize fuel economy via truck speed, shifting patterns and reduced idling.

Raw data on individual trucks are available for purchase from Vnomics. Limitations of the data for assessing geographic locations are similar to Verizon – the data are truck-centric rather than location-centric. Like the Verizon data, the data can provide precise estimate of key-on and key-off times, and GPS data can be worked with in GIS or mapping applications to determine location. Because these are mostly long-haul trucks, however, a disadvantage for spatial analysis with these data is isolating trip activity to specific geographic areas. The data are ideal for determining start activity for a specific long-haul truck fleet, but since MOVES in concerned with start activity in a specific area, there is no guarantee that trucks with available data are traveling in that area.

2.1.2.3 Engine/GPS Data Summary (Using Verizon as an example)

Temporal resolution: 1-2Hz over period of logging (for example, EPA Verizon data purchased over 12 months)

Data available:

- A unique vehicle ID scrubbed of vehicle ownership information
- Make, model, year, engine size, # of engine cylinders
- Model year of engine, if available
- VIN stem (8 digits)
- Vehicle's Gross Vehicle Weight (may be from VIN)
- Vehicle's stored zip code and County
- Fuel calculation method
- Trip start, end time and date
- Miles driven per trip
- Average MPG per trip
- Average and Maximum speed binned in 5 MPH bins.
- Battery voltage, RPM engine load and coolant temperature every 40 minutes
- Check engine light status, if applicable
- DTC code, if applicable
- The data are available in a csv format, one file for each State

Vehicle classes covered: "Personal Use" data covers LDV and LDT through 10,000 lbs (Class 2b). "Commercial Use" data covers MD and HD truck from Class 4-8. VNomic covers HD long-haul trucks (Class 8).

Sample rate: speed data generally 2 Hz, other parameters less frequent – depends on the variable, and vehicle manufacturer.

Sample size: Verizon has about 200,000 active vehicles nationwide.

Vehicle information: Make, model, trimline, model year, engine displacement, number of cylinders

Geofencing: There is no location data available for individual trips, only for where a given vehicle is registered. For this reason, it is not possible to assign individual trips (e.g. starts) to specific locations, especially at sub-county levels. This is a fundamental difference from the GPS-only data sources.

2.2 Applicability to MOVES Activity Inputs

A comparison of the two approaches for relevant MOVES start inputs is shown in Table 1 below. On balance, GPS-only has the advantage for anything related to spatial analysis of activity data. Engine/GPS has the advantage for activity linked to vehicle attributes such as model or age, and for more precise estimate of key on/off. However, lack of spatial attributes limits application of these data to larger geographic areas (nation, multi-state region)

Table 1. Applicability of Telematics Data Types to MOVES Start Inputs

MOVES Input	**GPS-Only**	**Engine/GPS**
Starts per vehicle	• Trips are implied • Can geofence by region, county, subcounty (e.g. grid)	• Key on/off provides more exact accounting of trips • Cannot geofence
Temporal distribution of starts (month/day/hour)	• Yes, though indexed over fluctuating sample	• Yes, sample size more stable temporally, as it is based on physical instrumentation of individual vehicles
Soak distribution	• Yes. Cannot distinguish < 15 minute soaks. • Multiple day soaks limited, or not possible • Geofencing	• Yes. Soaks < 15 minutes and multiple day soaks • No geofencing
Spatial distribution	• Yes	• No

Table 1. Applicability of Telematics Data Types to MOVES Start Inputs

MOVES Input	GPS-Only	Engine/GPS
Start activity by vehicle attributes such as vehicle model, age	• No	• Yes

2.3 Summary of Telematics Assessment

Telematics data sources present the possibility of a step change in the quantity of vehicle activity data that can be used to feed emission models such as MOVES. With telematics, the prospect of basing estimates on data from millions of vehicles is now a reality. Our review of telematics focused on two types of data: GPS data that track movement only, and engine data loggers that collect activity and engine parameters for individually logged vehicles. Both sources have pros and cons with respect to their use in developing MOVES start activity inputs. The main benefit of GPS-only data is sample size, and the ability to "geo-fence", i.e. focus on trip activity within a specific geographic region, regardless of whether a vehicle is from within the area, or travelling from another area. The benefit of an engine logger data is a precise measure of key-on and key-off, which can only be inferred from GPS data based on movement. However, it may be difficult to use engine logger data to characterize activity within a specific area, since data are linked to individual vehicles, not locations. Section 4.1 includes a direct comparison between trip metrics derived from StreetLight and Verizon.

3.0 Task 2: Generation of Pilot Dataset

Under Task 2, StreetLight Data processed trip metrics at the census tract level for the three pilot counties. ERG then processed the data into MOVES inputs, and compared to MOVES default trip activity. This section provides detail on StreetLight's data source and process, and the steps required to process StreetLight's delivery into county-level MOVES inputs.

3.1 Telematics Data Sources & Compilation (StreetLight InSight® Technology)

3.1.1 Overview

This section presents a comprehensive description of StreetLight Data's location data sources and algorithmic processing methodology, as of Fall 2017. StreetLight Data uses multiple data sources to develop metrics, believing that a combination of navigation-GPS data and Location-Based Services data is best suited to meet the needs of transportation planners, modelers, and engineers. In the past, StreetLight has incorporated and evaluated several different types of mobile data supply, including cellular tower data and ad-network derived data. The data supply landscape changes quickly as mobile technologies emerge and evolve, so new data providers are evaluated on an on-going basis to ensure the best available data sources are available for analytics.

As of December 2016, StreetLight's data repositories process analytics for nearly 35 million users. The data supply grows each month as StreetLight Data providers deliver new data sets. StreetLight Data currently uses one major navigation-GPS data supplier, INRIX, and one Location-Based Services (LBS) data supplier, Cuebiq. StreetLight is using these partners' data for Metrics in the US and Canada. Table 2 summarizes key details for the major location data technologies on the market.

Table 2. Overview of the telematics supply options for transportation analytics.

Type	Description	Pros	Cons	Notes
Cellular Tower	Derived from cellular tower "triangulation" (100-2000m spatial precision)	Large sample size, ability to infer home/work	Very poor spatial precision (average of several hundred meters), infrequent pings for some suppliers, high cost, customers opt-out, can't differentiate personal from	We haven't seen the US cellular industry making investments to improve these weaknesses.

Table 2. Overview of the telematics supply options for transportation analytics.

Type	Description	Pros	Cons	Notes
			commercial or bike/ped. Expanding to new countries requires new suppliers.	
Navigation-GPS	From connected cars, trucks, smart phones with navigation apps (3-5m spatial precision)	Excellent spatial precision, frequent pings enabling speed and mode inference, separate personal from commercial, opt-in. New countries may use same suppliers.	Usually lower sample size, difficulties inferring home/work depending on supplier practices.	Active mode inference only possible from smart phones.
Location-Based Services Derived Data	Mix of navigation-GPS, aGPS, and sensor proximity data from apps that “background” with locational data collection (5-25m spatial precision)	Very good spatial precision ability to infer trip purpose and trip chains superior to all other options, high sample size and growing (one provider has 10% of US population), opt-in. New countries may use same suppliers.	Suppliers still ramping up	Several players are emerging in this new market with very large sample sizes, opening up the possibility of a healthy, competitive supply base
Ad-network Derived Data	When user sees an ad on their phone, their location is recorded by ad-network	Large sample size of individuals, but very few pings per month. New countries may use same suppliers.	Few pings per month mean inference of travel patterns is not feasible	This source should not be used until significant changes are made.

Because of the pros and cons noted above, only LBS data was used for the ERG A-106 Project. Cuebiq provides pieces of software (called an SDK) to makers of mobile apps to facilitate Location Based Services (LBS). Thus, we refer to this type of technology as “LBS Data.” These apps include: couponing, dating, weather, tourism, productivity, weather, finding a restaurant/bank/gas station near you, and many more apps which utilize their users’ location in the physical world as part of their value. The apps collect anonymous user locations when they are operating in the foreground. They also collect anonymous user locations when operating in the background, whenever the device begins moving. LBS software collects data with WiFi proximity, a-GPS and a few other technologies, which means the majority of data has better than

50-meter spatial precision. Figure 3 provides a visual example of unprocessed LBS data captured over one month in a set location, along with GPS data. As shown, in this example, the LBS data points provide a more robust range of spatial data. This filtered visualization shows a subset of data points to improve visibility.

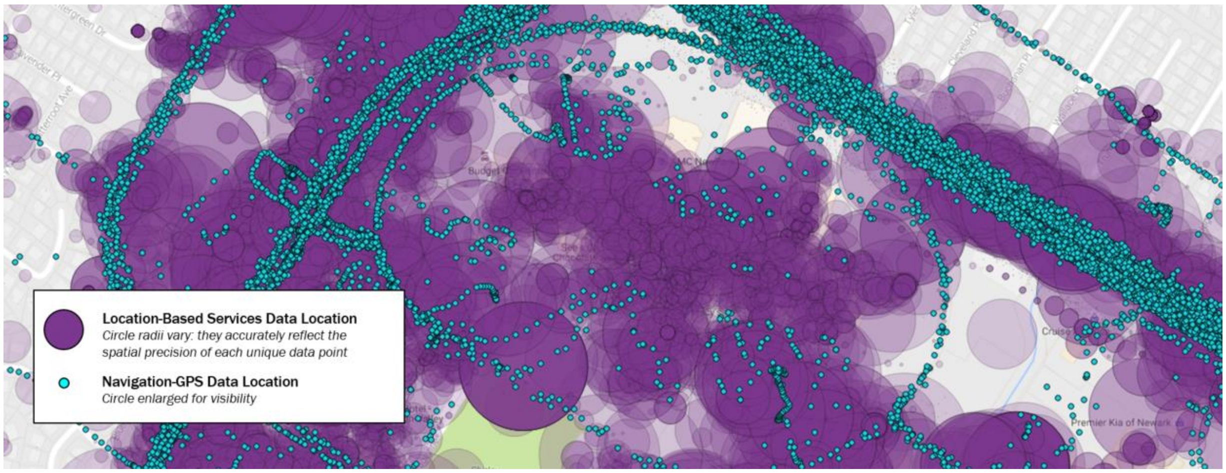


Figure 3. Visualization of unprocessed LBS and GPS data captured in one month near a mall in Fremont CA.

3.1.2 Data Processing Methodology

This section provides an overview of StreetLight's RouteScience® engine and the fundamental methodology that StreetLight uses to generate *StreetLight InSight* Metrics. Steps 1-7 are in common for all StreetLight Metrics. Steps 8 and 9 have been enhanced to give more specific methodology for the A-106 project.

3.1.2.1 Step 1 – ETL (Extract Transform and Load)

StreetLight pulls data in bulk batches from our suppliers' secure cloud environments (daily, weekly, or monthly, depending on the supplier). This data is already de-identified by suppliers in advance, before it is pulled by StreetLight. The ETL process pulls the data from one environment securely to another, eliminates corrupted or spurious points, reorganizes data and indexes it for faster retrieval and more efficient storage.

3.1.2.2 Step 2 – Data Cleaning and Quality Assurance

After each batch data ETL, StreetLight runs several automated quality assurance tests to establish key parameters of the data. For example, StreetLight checks that the volume has not changed unexpectedly, that the data are properly geolocated, and that it shares similar patterns to the previous month, etc. In addition, StreetLight staff visually and manually reviews key statistics about each data set. If anomalies or flaws are detected, they are reviewed by StreetLight in more detailed manner, and then escalated to conversations with suppliers' technical staff.

3.1.2.3 Step 3 – Create Trips/Activities

For any type of data supply, the next step is to group the data into key patterns. For example, for navigation-GPS data, a series of data points whose first timestamp is early in the morning, travel at reasonable speeds for 23 minutes, then stand still for several minutes could be grouped into a probable "trip." For LBS data, a series of pings within the same region for a few hours, followed by a gap and then a series of pings at another location, could be grouped into two probable "activities."

3.1.2.4 Step 4 – Contextualize

Next, StreetLight integrates other "contextual" data to add richness and improve accuracy of the mobile data. There are several types of contextual data including road networks, speed limits and directionality, land use data, parcel data, and census data, in addition to other data sets.

For example, a "trip" from a navigational GPS device is a series of connected dots. If the vehicle turns a corner but is only pinging every 10 seconds, that intersection might be "missed" when all the devices' pings are connected to form a complete trip. StreetLight utilizes road network information including speed limits, directionality, etc. to "lock" the trip to the road network, so that the true complete route of the vehicle is represented, even though discrepancies in ping frequency etc. may occur. Figure 4 illustrates this point.

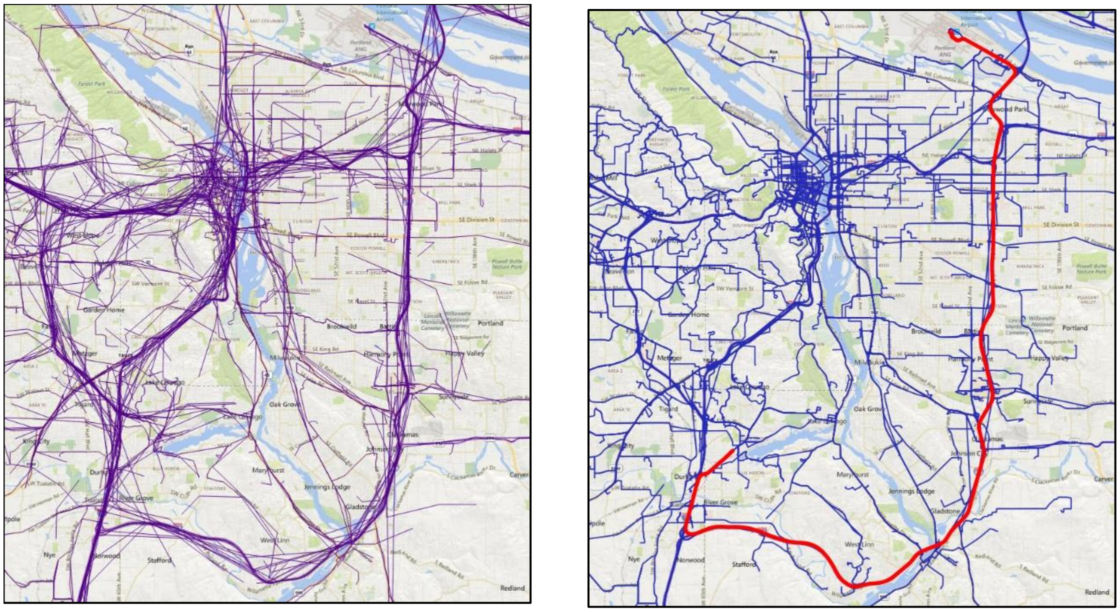


Figure 4. "Unlocked" Trips becoming locked trips

As another example, if a device has several activities on a census block with lots of residential land use, and that activity often occurs overnight, there's a high probability that that devices' owner lives on that census block or census block group. Then, the census-measured distribution of income for residents of that census block can be appended to that device and it can "carry" that distribution everywhere else it goes.

3.1.2.5 Step 5 – More Quality Assurance

After patterns and context are established, StreetLight's data engine *RouteScience* performs automatic additional quality assurance to flag patterns that appear suspicious or unusual. For example, if a trip appears to start at 50 miles per hour in the middle of a four-lane highway, that start is flagged as "bad." Flagged trips and activities are not deleted from raw databases, but are filtered out of most *StreetLight InSight* Metrics.

3.1.2.6 Step 6 – Normalize

Next, the data is normalized along several different parameters. As all data suppliers change their sample size regularly (usually increasing it), monthly normalization occurs.

For LBS/activity type data, StreetLight performs a population-level normalization for each month of data. For each census block group,[a] StreetLight measures the number of devices in that sample that appear to live there, and makes a ratio to the population reported to live there in total. To determine home location, StreetLight analyzes each device and where it spends each night in a month. If that location is a residential zoned place, it is tagged as a plausible home location. Over the course of the month, an analysis is performed on census blocks that get multiple tags. The five with the highest number of plausible tags are considered home locations. As noted in a previous example, a device from a census block that has 1,000 residents but two StreetLight devices will be scaled differently everywhere it goes compared to a device from a census block that has 1,000 residents and 200 StreetLight devices. For trips, StreetLight uses a set of permanent public loop counters at certain highway locations to measure the change in trip activity each month. Then it compares this ratio to the ratio of trips at the location, and normalizes appropriately.

StreetLight does not account for multiple phones in a car at this time, nor do they account for the case when a phone owner rides in another person's car, which adds some uncertainty to the StreetLight estimates. However, multiple phones happen rarely in StreetLight's sample because a) not every phone is in their sample, and b) having the LBS algorithm place two phones in the same vehicle would require the phones have the exact same apps (of the >300 apps that location data is culled from), in order to provide identical location data. In event it does happen, if multiple phones are from people from the same household, they will be adjusted down in the normalization step. StreetLight does exclude trips that appear to be in non-motorized travel (walk, bike) based on speed, acceleration, and other factors.

3.1.2.7 Step 7 – Store Clean Data in Secure Data Repository

After being made into patterns, checked for quality assurance, normalized, and contextualized, the data is stored in a proprietary format. This enables responses from queries via the *StreetLight InSight* web application in an efficient manner.

3.1.2.8 Step 8 – Aggregate in Response to Queries – ERG A-106 Queries

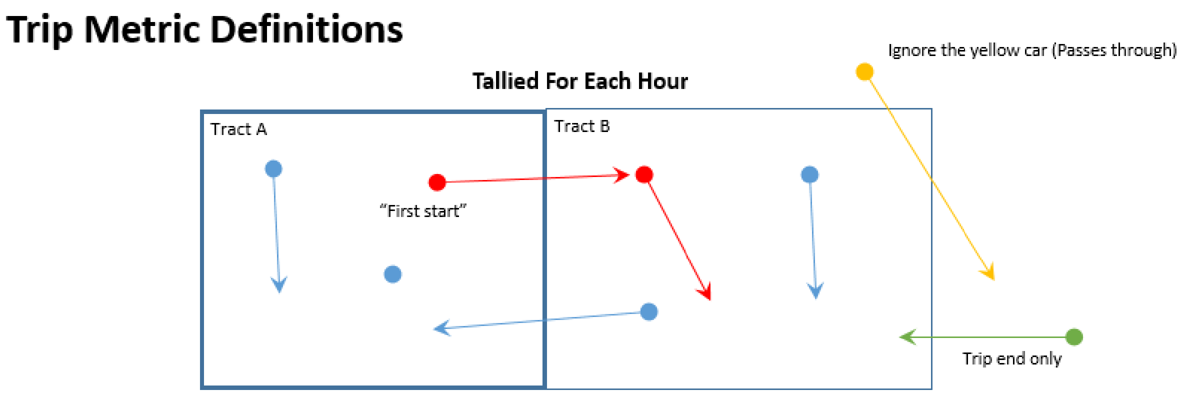
The analysis for the ERG A-106 project only used LBS data, and involved the following steps:

[a] According the U.S. Census, a block group consists of about 39 blocks per group. The U.S. has 66,438 census tracts; 211, 827 census block groups; and 8,269, 131 census blocks (www.census.gov)

1. For each month between April 2016 and March 2017, look at all "activity places" that have a previous as well as next activity place at a distance of >= 100 meters (i.e. there is a trip ending at the activity place, and a trip starting from the activity place). This threshold is in place because the minimum distance for definable trips is 100 meters; there is no limit on maximum distance, or on the minimum or maximum time of captured trips.
2. Note that each activity place is actually a polygon, depending on the number of activity points that are between the previous trip end and next trip start. Filter the places to only those that intersect the counties of Clark NV, Cook IL and Fulton GA.
3. For each activity place, determine the following attributes:
    a. Enter Time: The time at which the previous trip ended at the place.
    b. Exit Time: The time at which the next trip started at the place.
    c. Dwell Time: The duration between the Enter time and Exit time.
4. Assign each place to a Census Tract ID, based on its polygon overlap with the census tract. It is possible that a place polygon overlaps multiple census tracts, in which case it is assigned to each, based on the percent of area overlap with the corresponding census tract.
5. Then using the above, for each segment of Data Month, Census Tract, Day Type (Weekday/Weekend) and Hour, the Metrics are calculated as the following (see Figure 5 for schematic):
    a. **Number of Trip Ends:** Number of activity places, scaled by the population-based device factor, that have an Enter Time in that segment.
    b. **Number of Trip Starts:** Number of activity places, scaled by the population-based device factor, that have an Exit Time in that segment. The normalized result is known as trip index (TripIndex) in subsequent analyses.
    c. **Starts Per Vehicle**: This is equal to (Number of Trip Starts) / ("Resident Vehicles" i.e. Number of First Trip Starts + Number of Parked Devices).
        i. The Number of Trip Starts is as stated above.
        ii. The Number of First Trip Starts is the number of activity places, scaled by the population-based device factor, that have an Exit Time in that segment, and do not have an earlier Enter Time in that same segment.
        iii. The Number of Parked Devices is the number of activity places, scaled by the population-based device factor, that have a dwell time that span the entire hour for the segment. E.g. if a place has an enter time = 8:15 am and an exit time = 12:30 pm, it is counted as parked for the hours of 9, 10, 11.
    d. **Average Dwell Time**: Average dwell time of each activity place, scaled by the population-based device factor, that has an Enter Time in that segment. The minimum dwell time in the sample is 5 minutes. There is no maximum, so dwell times can span multiple days, however, a device needs to ping at least once every 30 minutes for it to keep continuity within that soak (else, it will "end" the soak

time).

e. **Distribution of Dwell Time (0-15 min, 15-30 min, etc.)**: Same as above for dwell time, but only included in the specific dwell time bin.



Dwell Time: The duration of time spent between the end of one trip and the beginning of the next trip,
Resident Vehicle: A vehicle is considered Resident in a domain (tract) for a given hour if either of the following two conditions are met:
1) the vehicle's first start within the given hour is within that tract, or
2) the Dwell Time of the vehicle includes the entirety of the given hour.
Vehicles that pass through without stopping, or have a "1st start" in another tract are NOT counted as resident vehicles
Starts Per Vehicle: This custom metric is defined for a given county, vehicle weight class, day type, and hour bin. For each bin, the metric is defined as:
(number of vehicle starts) / (number of Resident Vehicles)

Example:

Tract	A&B	A	B
(i) # of vehicles starting	4	2	2
(ii) # of veh still	1	1	0
(iii) # of trips starting	5	2	3
Starts/vehicle = (iii)/(I + ii)	1	.67	1.5

Only count the red car as starting in it's VERY FIRST tract of the hour. For this hour, the red car is a resident vehicle only in Tract A, though there is a trip start in both Tract A & B.

Figure 5. Defining Trip Metrics: Starts, Starts/Vehicle and Dwell Time

An important distinction with the location-based StreetLight Data vs. logger-based datasets (Verizon) is that StreetLight tracks vehicles entering and exiting a defined geographic area. StreetLight defined "resident vehicle" as vehicles within a geographic area for a given hour; it is not based on where vehicles are registered, but where vehicles are located in a given hour. The derivation of starts/vehicle is shown in Figure 5 for two example domains (e.g. census tracts). Resident vehicles are based on location of a vehicle at the start of a time period, while starts are defined based on trip activity. Note that some vehicles may remain stationary during a period, while trips within a chain main originate in a different domain from where the vehicle is "counted". The result may be start/vehicle less than 1, reflecting stationary vehicles; or greater than 1, which may indicate more trip starts than resident vehicles. Pass through vehicles are ignored. This tracking of "resident" vehicles can account for the influx of traffic from commuting. Figure 6 shows average weekday trip start index, trip end index, and resident vehicles for Fulton County, which shows the clearest affect from migration. Where trip ends are greater than trip starts, resident vehicles increases; these are vehicles commuting into the county. In the afternoon, the opposite happens – trip starts are greater than trip ends, and resident vehicles decreases.

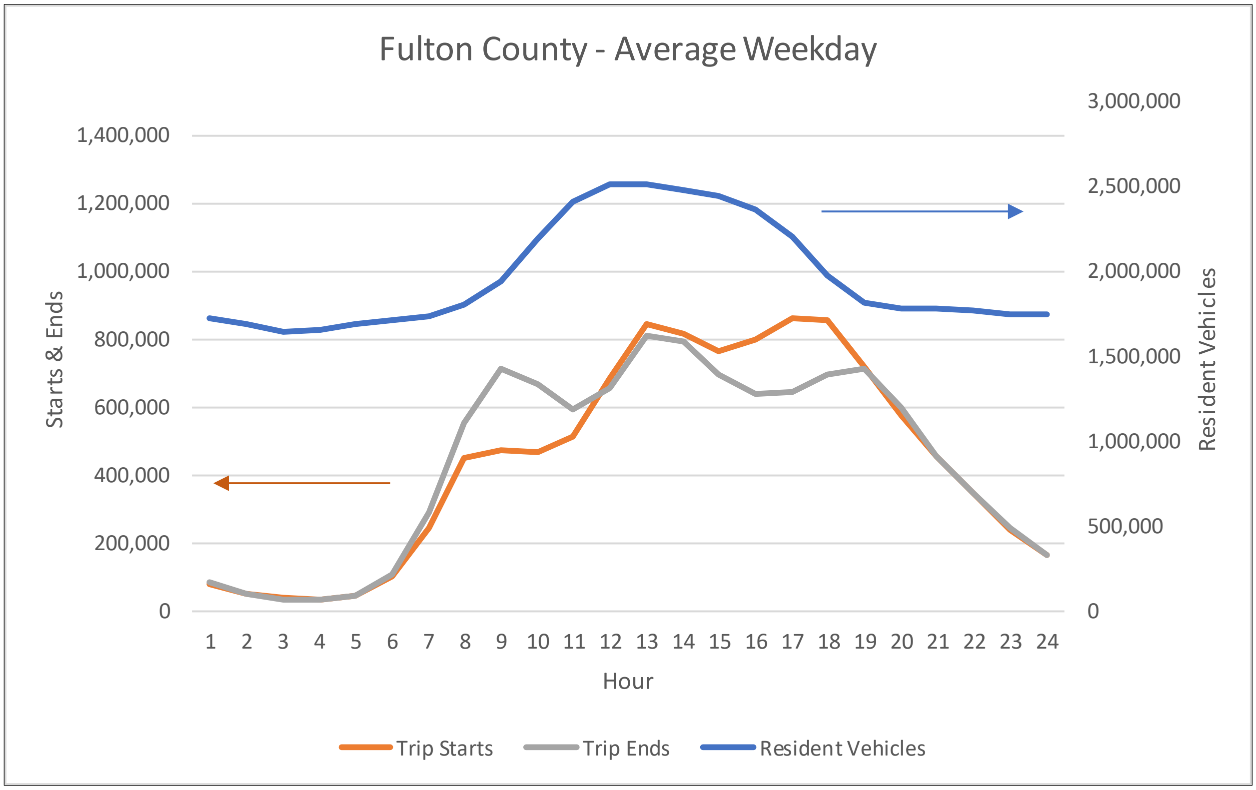


Figure 6. Fulton County Trip Start & End Index, Resident Vehicles for Average Weekday

The resident vehicles implicit in the StreetLight data are not consistent with MOVES and SMOKE/MOVES static population. What this inconsistency points out is that by assuming a static population, MOVES may mis-apportion starts geographically, especially in areas with inter-county commuting. For example, using a static population based on county registration, coupled with starts/vehicle from vehicles registered in that county, would not account for starts in the county from vehicle registered elsewhere (e.g. commuting vehicles). How to reconcile this will take some thought, but at a minimum LBS could be useful for re-apportioning start activity calculated based on registration-based populations.

3.1.2.9 Step 9 – Final Metric Quality Assurance – ERG A-106 Queries

Before delivering results, we performed the following quality assurance checks.

1. The total trip starts and trip ends, aggregated by day type and hour (i.e. distribution over a day type) is a "bell-shaped curve" that is lower at night-time hours and peaks during the day time.   Check the same for a particular census tract and data month as well.
2. The total trip starts and total trip ends aggregated for the census tract and data month are about the same.
3. The total distribution (scaled) for the dwell time adds up to the total number of trip starts (within rounding error).
4. Review the minimum and maximum values, and a few of the highest values (i.e. below the maximum) for the average dwell time.
5. Review the minimum and maximum values, and a few of the highest values (i.e. below the maximum) for the starts per vehicle.

3.2 Processing method to generate MOVES tables

Two MOVES input tables were generated from the StreetLight data: *startsPerVehicle* and *ImportStartsOpModeDistribution*. Both tables have key fields of *sourceTypeID* and *hourDayID*, which combines hour of the day and day type (weekday or weekend). The processing of the StreetLight data into these tables was minimal, since the data aggregation performed by StreetLight was configured to match these MOVES inputs as directly as possible. The starts/vehicle data could be imported directly into *startsPerVehicle* after being aggregated by county and day type. This table is not listed under the "Starts" tab of the County Data Manager, but can be accessed through the "Generic" tab and read into a custom database specified by the user. This database is then entered in the Manage Input Datasets GUI screen when setting up a runspec.

Dwell (soak) times were provided by StreetLight, expressed as an indexed trip count in each dwell time bin. These bins were set up to match the MOVES start operating mode distributions as closely as possible. The main difference was that the data for the shortest soak period could only be provided for periods of 0-15 minutes instead of the 0-6 minutes defined by MOVES. StreetLight bins of 0-15 and 15-30 minutes were combined to approximate the 6-30 minute bin in MOVES. The index trips by bin were normalized to total indexed trips to create the operating mode fraction, by hour and day type. These were then imported into the table *ImportStartsOpModeDistribution* through the County Data Manager "Starts" tab.

The StreetLight data were applied to both the Passenger Car (*sourceTypeID*=21) and Passenger Truck (*sourceTypeID*=31) vehicle types.

4.0 Task 3: Data Analysis & Comparison to Independent Sources

Task 3 focused on comparison of StreetLight's trip metric data to independent sources of trip data. Comparisons were performed at the county level to MOVES defaults, summarized Verizon telematics data, and the National Household Travel Survey (NHTS), to provide a general sense of how aggregate trip metrics of starts/vehicle and average soak time compared between StreetLight Data and these sources. Analyses were also performed at the census tract level to assess correlations between StreetLight data, census statistics, and travel demand model trip origins. Analysis of spatial distribution was also conducted at the census tract level, to help assess how telematics data might be used to improve the allocation of county-level start emissions to sub-county level, for use in air quality modeling. County and tract-level analyses are summarized in the following sections.

4.1 County-Level Analysis

4.1.1 Summary & Comparison to MOVES Defaults

The hourly starts/vehicle distributions derived from the StreetLight data are shown in Figure 7 (starts/vehicle/hour summed over 24 hours is the more typical starts/vehicle/day metric presented in MOVES documentation). Default MOVES starts/vehicle distributions are shown for comparison; these were generated by running MOVES2014a with the default database and then calculating starts/vehicle from the population and activity data contained in the *movesActivityOutput* table. The default MOVES starts/vehicle distributions are identical for all three counties. As shown, the three pilot counties are similar in starts/vehicle, and show roughly the same trend over both weekdays and weekend days. For weekdays, all three counties average 5.1 starts/vehicle/day, vs. 5.5 for MOVES (average of passenger car and truck). On weekends, StreetLight Data averaged 4.6, 4.7 and 5.1 starts for Clark, Cook and Fulton respectively, vs. 4.85 for MOVES (average of passenger car and truck).

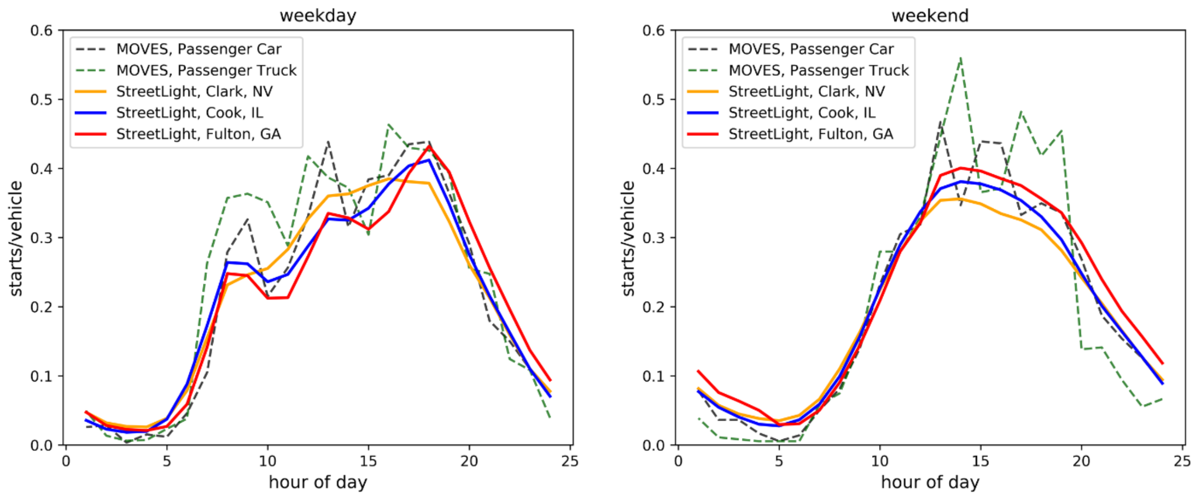


Figure 7. StreetLight and default MOVES starts/vehicle/hour

The StreetLight average dwell times (soak time) for each county are shown in Figure 8, along with the average soak times calculated from the same MOVES runs used to generate the default MOVES starts/vehicle distributions in Figure 7. Although MOVES uses a distribution of soak times, only the average is shown here for presentation purposes. The MOVES default soak time distributions were calculated from the *startOpModeDistribution* table that is saved to the *movesExecution* database. This database contains an *opModeFraction* for each *sourceTypeID, hourDayID,* and *opModeID*. In order to translate this table into average soak times, the range of soak times represented by each *opModeID* must be converted into a single point value. For example, *opModeID* 102 encompasses a range of 6-30 minutes, so the midpoint value of 18 minutes could be used as the single point value to represent this *opModeID*. For comparison purposes, we calculated the MOVES default soak time distributions using the 10th percentile point of each bin; the midpoint of each bin; and the 90th percentile point. The midpoint assumption is shown by the dashed lines in Figure 8, and the ranges encompassed by the 10th percentile to 90th percentile assumptions are shown by the shaded regions.

As shown, for the StreetLight data, the average dwell time has a maximum at the beginning of morning rush hour (7-8am), as vehicles have their first start after sitting overnight. The average decreases through the day, as vehicles in use during the day have a shorter time between trips throughout the day, with a weekday bump at the afternoon rush hour, reflecting vehicles that have been parked for the entire work day. The average soak periods for StreetLight are longer than MOVES defaults throughout the day; as detailed in Section 4, this difference is consequential for emissions.

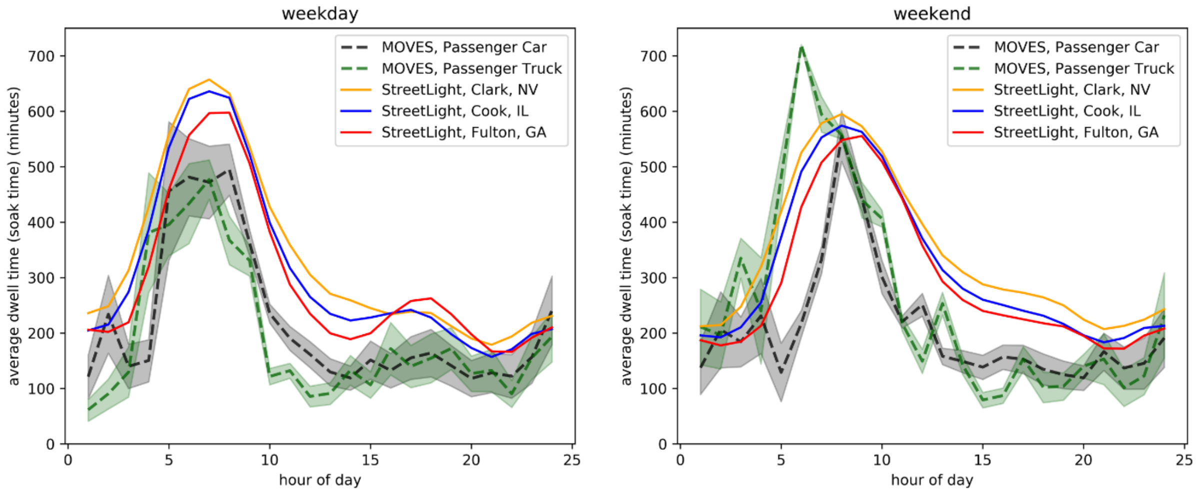


Figure 8. StreetLight and default MOVES average dwell (soak) times.
Shaded areas around MOVES defaults lines show ranges of values obtained using the 10th percentile to 90th percentile of each operating mode.

A key feature of the StreetLight dataset was the ability to isolate start activity at the sub-county level, a stated objective of CRC in for this study. ERG and StreetLight Data, in consultation with CRC, decided that data would be provided at the census tract level. This was the smallest unit at which StreetLight felt they could generate reliable results for comparison with travel demand models (TDMs). Census tracts represent aggregations of travel analysis zones (TAZs) used in TDMs. Therefore, analysis at the Census tract level enables direct comparison betweenStreetLight data and TDMs. To give a sense of variation across tracts, Figure 9 shows weekday starts/vehicle/hour for each tract in Fulton county. The average of this plot is from Figure 7. Figure 9 gives a sense of the variability at the tract level, with starts/vehicle varying from the mean by roughly ±50 percent.

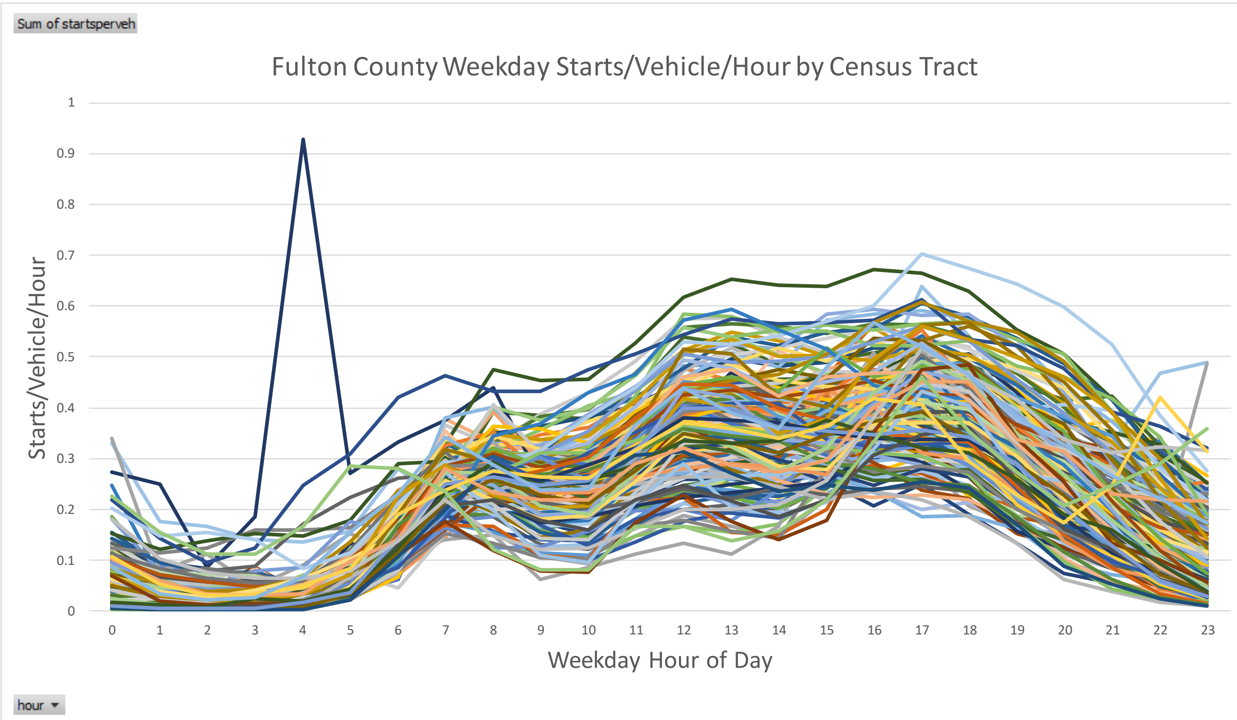


Figure 9. Fulton County Starts/Vehicle/Hour by Census Tract

To assess the sensitivity of the MOVES emissions outputs to this variability, ERG identified two tracts in Fulton, GA, that were at the 10th and 90th percentiles in terms of starts/day/vehicle (4 starts/day and 7 starts/day, respectively): tract 13121010114 and tract 13121011421. StreetLight Data's trip metrics for these tracts were used to develop alternate *startsPerVehicle* and *ImportStartsOpModeDistribution* tables. The start/vehicle and average dwell times for these tracts are shown in Figures 10 and 11. The tracts used as the 10th percentile and 90th percentile cases were then run through MOVES in the same way that the county averages were.

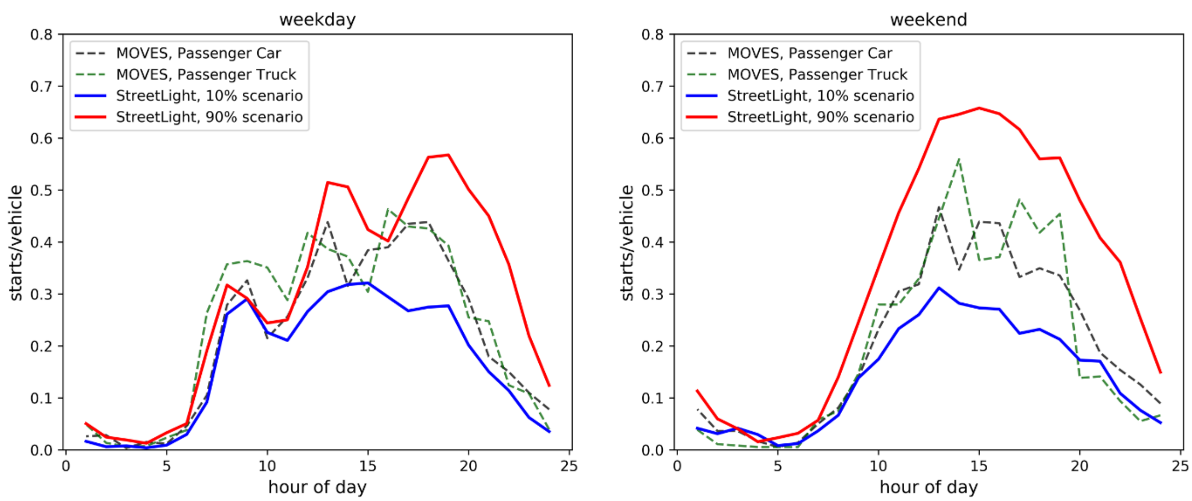


Figure 10. StreetLight starts/vehicle for 10th and 90th census tract vs. MOVES default

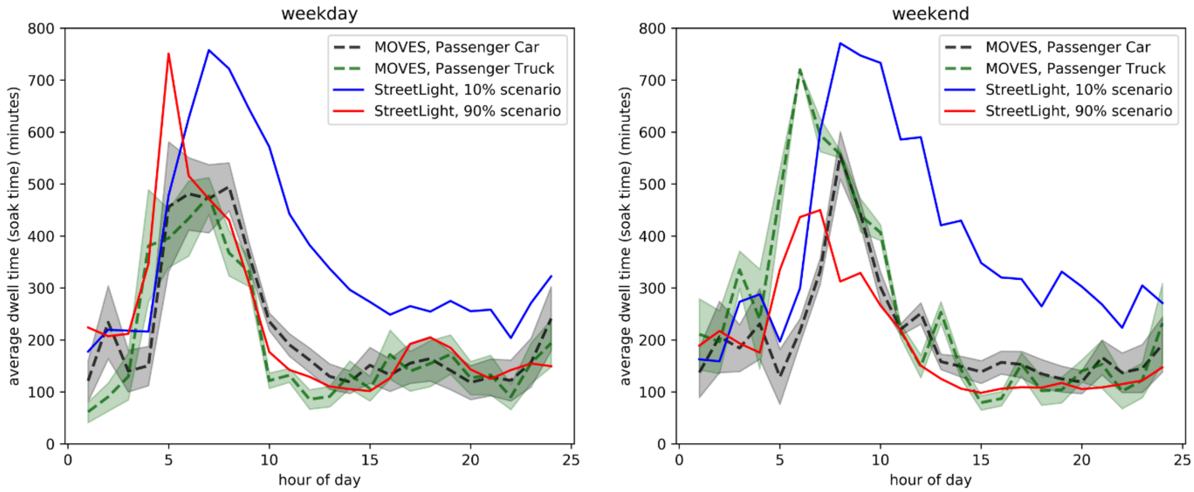


Figure 11. StreetLight average dwell time for 10th and 90th census tract vs. MOVES default

Shaded areas around MOVES defaults lines show ranges of values obtained using the 10th percentile to 90th percentile of each operating mode.

4.1.2 Comparison to Verizon & NHTS

As discussed in Section 2, Verizon Telematics compiles data from loggers installed on thousands of vehicles nationwide, mainly from owners participating in insurance programs which reward safe driving. EPA's Office of Transportation and Air Quality has been purchasing and analyzing a subset of these data for several years, in part to update MOVES default activity assumptions.[4] EPA has recently presented results of an analysis of Verizon data for 5 states (Illinois, Georgia, California, Colorado, New Jersey) with the intent of scaling trip activity from these states up to a representative national default for MOVES.[5] Because the Verizon activity data is culled directly from vehicle OBD ports, which provide precise engine operation, comparison of StreetLight's location-based data to Verizon's engine-based is very useful to consider reconciliation of telematics data sources. It is fortuitous that the Verizon samples in Illinois and Georgia overlap with StreetLight's data in Cook and Fulton counties. EPA's analysis has focused on state level analysis; ERG has supported EPA's work with Verizon, and was able to produce summary statistics of the Verizon data specifically for Cook and Fulton counties, for a more direct comparison.

Another source of data on individual vehicle trips in the National Household Travel Survey (http://nhts.ornl.gov/), or NHTS. This survey is conducted and published every 8 years, with 2009 being the most recent year available. Though transitioning to using GPS instrumentation at least in part, the currently available NHTS data were gathered through travel diaries. Data were collected in major metropolitan areas, providing another source of direct comparison to StreetLight and Verizon. NHTS 2009 tracked travel via diary on each vehicle for one day per household. An advantage of this dataset is that it tracked trips individually on multiple vehicles per household, which provides a sense of primary vs. secondary vehicle trip activity. The sample sizes used for the comparison are shown in Table 3.

Table 3. Verizon & NHTS Sample Sizes for Comparison to StreetLight

	Fulton County	**Cook County**	**Clark County**
Verizon	2,376	5,432	-
NHTS	8,911	1,079	311

Figure 12 shows average starts/vehicle/day estimated from the three sources. StreetLight and Verizon are annual averages, while NHTS reflects the one recorded day of travel per household. NHTS is divided into 1st and 2nd household vehicles. When looked at this way, StreetLight falls in the NHTS range, while Verizon is closer to the 2nd NHTS household vehicle.

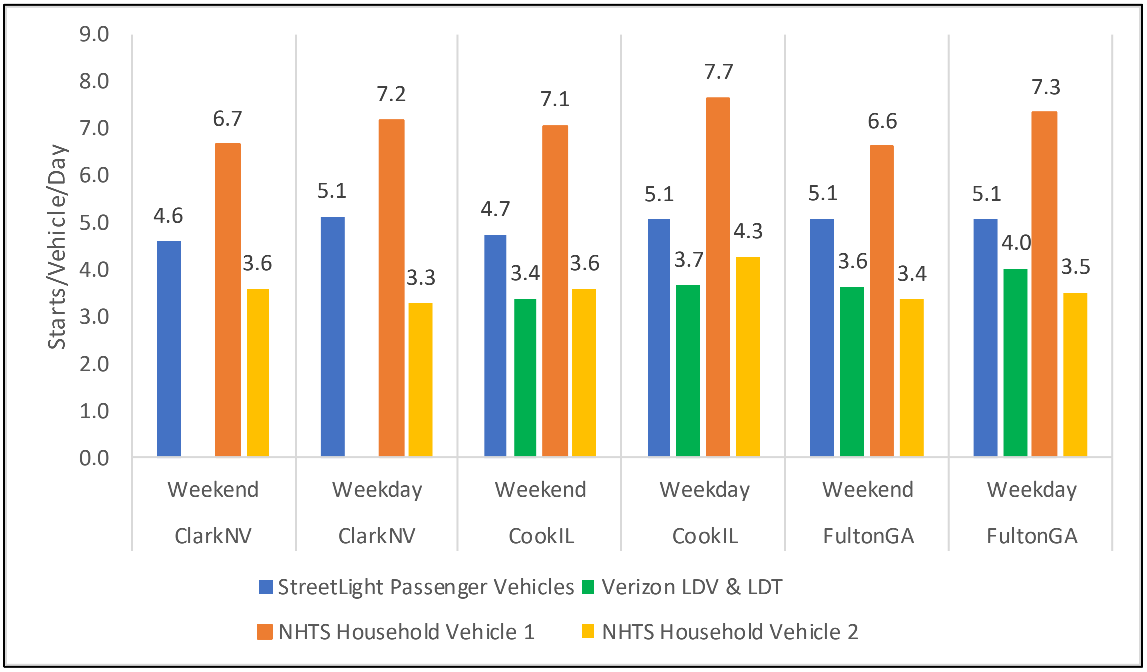


Figure 12. Average Starts/Vehicle/Day for StreetLight, Verizon & NHTS

Though not shown on the chart, another data point to consider is the California Household Travel Survey (CHTS), which instrumented vehicles in 1,440 household for 7 days each, covering a period from January 2012 – January 2013. The overall average starts/veh/day for the sample was 4.75, which is being used by ARB to update the EMFAC vehicle emission model activity assumptions.[6] This is about 0.7 starts/veh/day higher than Verizon data culled statewide in California (Figure 2), and in the range of the StreetLight data presented for the three counties.

A hypothesis for why the Verizon starts/vehicle estimate is lower than the averages using either StreetLight, NHTS or CHTS is because a) it may be weighted towards secondary vehicles and/or b) the monitoring of vehicles affects the amount of driving, because the insurance programs using the data provide incentives to drive less. For example, State Farm provides a 20 percent discount in insurance premiums for policy owners who do not drive their vehicles in excess of 12,000 miles per year.[7] Such discount programs also may not require all vehicles in a household to be monitored, so drivers may have the incentive to selectively instrument vehicles that would qualify for low mileage discounts. While beyond the scope of this study, we recommend further analysis of Verizon dataset to determine how mileage accrural rates compare independent sources, such as vehicle registration records.

Conversely, a reviewer raised the question as to whether StreetLight may overcount vehicle trips by ignoring multiple devices per vehicle; StreetLight Data observes in their data that this happens very rarely, so is ignored in estimates. Continued comparison and evaluation of the representativeness of different telematics datasets is recommended.

4.2 Tract-Level Analysis

The variability of trip metrics within a county, as demonstrated in Figure 9 above, provides the opportunity to assess if it can be explained with tract-level data available from the U.S. Census, or from trip origins generated by travel demand models (TDMs). Establishing such a correlation could be useful for predicting trip metrics in the absence of telematics data. Conversely, the analysis shows the potential for telematics data to help validate TDMs. This analysis is detailed in Section 4.2.1.

Another function of tract-level data is to evaluate the spatial distribution of start activity within a county. This is useful because air quality modeling requires emissions at the sub-county level (generally 4km or 12km grid cells), and in the SMOKE / CMAQ framework used by EPA, start emissions are allocated from the county level to grid cell based on human population. StreetLight trip metrics are therefore compared to human population at the census tract level, along with vehicle population and TDM trips, to assess how these different sources might affect spatial allocation of emissions. This analysis is detailed in Section 4.2.2.

4.2.1 Regression of Trip Metrics

In an effort to determine which available predictor variables are most closely associated with the StreetLight TripIndex, as well as the daily starts per vehicle, ERG created and analyzed several linear regression models. The models were created on a by-county basis for Fulton County GA, Cook County IL, and Las Vegas County NV. The data consisted of approximately 2,000 observations of the variables of interest, aggregated Table 4 below.

Table 4. Response & Predictor Variables in Tract-Level Regression

Response Variable	Predictors
StreetLight TripIndex	Trips
	Population
	Vehicles
Starts Per Vehicle	Population Density
	Housing Unit density
	Vehicle Density
	Vehicles per Capita
	Vehicle per Housing Unit

In general, each analysis followed a series of steps. First, ERG read in the data of interest and examined diagnostic plots for the untransformed values. For a linear model to be valid – that is, for confidence intervals, hypotheses testing, t-tests, or other standard statistical methods to have any meaning – the model must meet certain criteria. Specifically, the errors should be independently and identically distributed, normally, with mean 0 and constant variance. In most cases modeled here, the normal Q-Q plot indicated deviation from normality, and the scale-location plot showed variability that was non-constant with increasing fitted values. Further, the Cook's D plot often identified one or more outliers of concern.

To mitigate such problems, ERG next used the standard Box-Cox transformation algorithm to determine the power transformation most likely to produce a valid statistical model. Although the exact parameter chosen varied between models, a transformation was required in all cases. After applying the necessary power transformation to the data, ERG was able to produce diagnostic plots indicating a valid model, in that normality and equal variance requirements were satisfied, and outlier intensity was mitigated.

A table summarizing results for all models is shown in Table 5.

Table 5. Regression Analysis Results

Response Variable	County	Box-Cox Transformation	Significant Predictors	Adjusted R-square	Max VIF
TripIndex	ALL	0.0697 (log)	Trips, Population, Vehicles	0.72	5.2
	Fulton County	0.444	Trips	0.83	7.3
	Cook County	0.132 (log)	Trips, Population	0.73	6
	Las Vegas County	0.052 (log)	Trips	0.81	3.8

Response Variable	County	Box-Cox Transformation	Significant Predictors	Adjusted R-square	Max VIF
Starts Per Vehicle	ALL	-0.421	Population Density, Unit Density, Vehicles per Housing Unit	0.097	23.2
	Fulton County	-0.539	Vehicles per Housing Unit	0.199	28.7
	Cook County	-1.09	Population Density, Unit Density	0.068	25.1
	Las Vegas County	-0.0387	Vehicles per Housing Unit	0.197	25.6

From the results of the analysis, we can conclude that for the StreetLight TripIndex, the number of trips is likely a useful predictor at the p=0.05 level, with population and vehicles sometimes of benefit. The adjusted R-square values are greater than 0.7, which indicates the transformed models do a reasonably good job of capturing variability in the TripIndex. With some variance inflation factors (VIFs) above 5.0, multicollinearity is a minor concern – that is, we have an indication that there is a bit of correlation among predictors. This may mean that the estimated coefficients have some additional uncertainty, but probably not too much.

For starts per vehicle, the conclusions are less clear. While vehicles per housing unit shows up as significant most frequently at the p=0.05 level, the overall model performance is poor, with adjusted R-square values less than 0.2. Further, the results have very high VIFs, indicating that multicollinearity among the predictors is a problem. That is to say, even if these models had better predictive power, we would not be able to calculate very precise estimates of the model coefficients. Note that validation was not performed on these models, and thus estimates of specificity and sensitivity are not presented. Rather, given the transformed nature of the variables and the difficulty of interpreting them in transformed space, the results are merely indicators of trends in the data. For example, as number of trips increases in each census tract, TripIndex increases.

The diagnostic plots, marginal model plots (MMP), and added variable plots (AVP) from analyses involving all counties combined are displayed in Appendix A, along with related output from the R statistical software.

4.2.2 Spatial Analysis

The telematics trip data used for this study have a location accuracy of about 5 meters, meaning that the raw data could readily roll up to the resolution needed by air quality models (AQMs), usually 1-km, 4-km, or 12-km grid cells. AQMs use spatial surrogates for most sources (except point sources) because the precise emissions locations aren't known at a sub-county level. For on-road vehicle start emissions, the traditional spatial surrogate has been human population. The human population data source is freely available, thoroughly vetted/quality assured, and regularly updated, and covers the entire US. Population has been used to allocate start emissions most recently in EPA's 2011 modeling platform that supported ozone transport modeling for the 2008 National Ambient Air Quality Standards (NAAQS), the 2015 NAAQS for ozone, and other special studies.[8] Focused AQ modeling over nonattainment areas that are urban and have their own travel demand models (TDMs) sometimes use trip origins and destinations by traffic analysis zone (TAZ) to allocate off-network emissions from vehicles, but these spatial data are not widely available for larger modeling domains.

This spatial distribution assessment of the telematics data is limited by the scope of the study to focus on census tracts within 3 counties as opposed to grid cells in a modeling domain. As such, the approach to understanding the impacts of telematics data is qualitative in nature. Nonetheless it is important to consider where emissions are placed, particularly in cities where mobile sources dominate the emissions and even more so for the pollutants VOC and CO because a large portion of these occur during vehicle starts. In this study's baseline scenario (NEI 2014 inputs), Fulton, Cook, and Clark County start emissions contribute 45%, 56%, and 35% of on-road VOC and 36%, 48%, and 30% of on-road CO emissions, respectively. The contribution of starts to other pollutant emissions is smaller at 14-20% for NO_X and 7-15% for $PM_{2.5}$, though the telematics vehicle trips data can improve the spatial resolution of all pollutants input to an AQM.

4.2.2.1 Qualitative Assessment

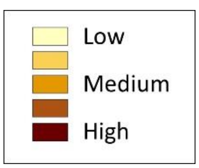
The qualitative approach to understand differences among potential spatial surrogate methods was to prepare color-coded maps of the census tracts in the 3 counties, allowing side-by-side comparisons of where the start emissions would be placed using the data sources of (1) human population, (2) vehicle population, (3) TDM trips by TAZ, and (4) telematics trips. Each county has a block of 4 maps, following the layout summarized in Figure 13. The top row contains the population density by tracts with human on the left and vehicle on the right. The bottom row shows the relative trip density by tract from TDMs (Left) and telematics (Right). Each county map follows the same layout.

Human Population	Vehicle Population
TDM Trip Origins	Telematics Trip Starts

Figure 13. Layout of Census Tract Maps for each County

Human population and household vehicles by census tract came from the 2011-2015 American Community Survey (ACS) 5-Year estimates available from the U.S. Census Bureau. Travel demand model (TDM) trips by travel analysis zone (TAZ) were graciously provided by local transportation planning agencies in each county – the Atlanta Regional Commission (Fulton County, GA), the Chicago Metropolitan Agency for Planning (Cook County, IL), and the Regional Transportation Commission of Southern Nevada (Clark County, NV). The population and vehicle counts were available at the census tract level, but the TDM trips required geographic information system (GIS) analysis to transform the TAZ level activity into census tracts.

The population in the ACS datasets do not contain any temporal variability because the data are simply a snapshot in time of where humans were living and where vehicles were domiciled. The TDM trips provided by transportation planning agencies for this work are presented as daily totals for a weekday, though it is worth noting that multi-hour time periods are often available in the TDM trip distributions (i.e., AM and PM peaks, midday, and overnight periods). The telematics data trip start spatial distributions presented here are also weekday totals (for July), for consistency with the TDM-based trip presentation, although the telematics trips are available for all 12 months, weekday and weekend day types, and 24 hours of day.



The set of four maps per county uses a color scheme of five bins shown to the left. Qualitatively, the lighter colors correspond to areas in the county that would receive less of the total start emissions, and darker means the area would receive a high amount of start emissions. There was a quantitative basis for the binning: percentiles. From top down, the bins correspond to the 20th percentile (pale yellow), the 20-40th (dark yellow), 40-60th (tan), 60-80th (brown), and 80-100th (maroon) percentiles in the tract level density values.

4.2.2.2 Results

Figure 14 shows the results for Fulton County, Georgia. The main differences are apparent in North Fulton, where the human population and vehicle population density hot spots seem focused around cities located in this part of the county (Alpharetta, Roswell, and Johns Creek). In contrast, both the TDM trip and telematics trip densities appear centrally focused around the highway US 19. The differences between population and trips in general is subtle for Fulton, but the TDM trips appears to best model the location of actual trips compared to the other potential surrogate sources of human or vehicle population. Figure 15 shows a difference plot between human population and StreetLight trip start index – the differences are expressed as delta quintiles – when considering the allocation of county level emissions, a positive difference illustrates an increase in emissions intensity, shifted from the areas with negative difference. These plots give some sense of where emissions would be re-allocated if StreetLight metrics were used as a surrogate, instead of human population.

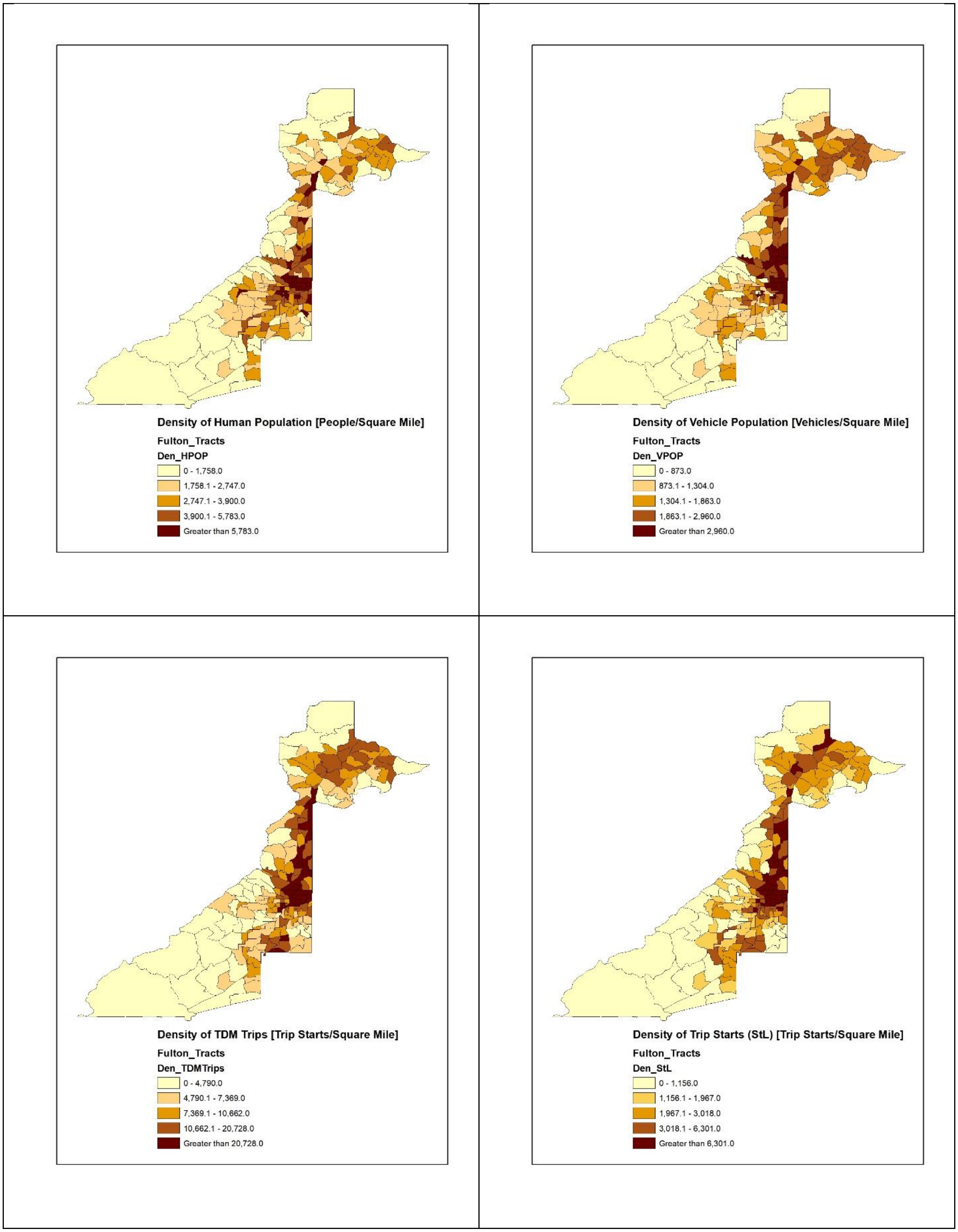


Figure 14. Potential Methods to Spatially Distribute Start Emissions – Fulton County, Georgia

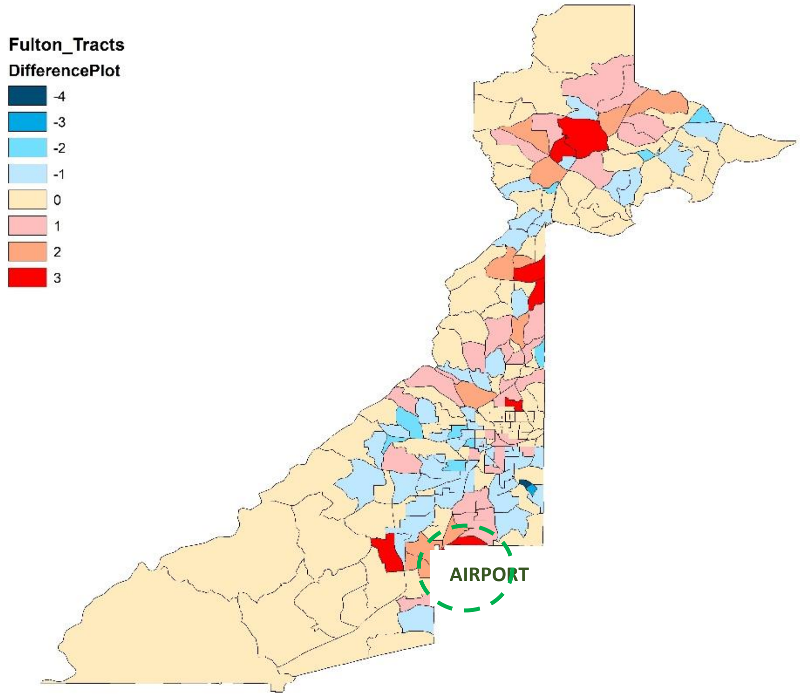


Figure 15. Difference Plot for Fulton County (StreetLight minus Human Population Quintile)

Figure 16 shows the results for Cook County, Illinois. Like the Fulton County set (Figure 14), Cook's population data for humans and vehicles closely aligns. A noticeable feature common to all 4 data sources is the low density of population and trips along the Chicago Sanitary and Ship Canal that runs from Lake Michigan southwest through the county. The telematics trip dataset shows a higher density of activity at Chicago O'Hare International Airport than the TDM trips or the population datasets. Other than the airport, there aren't any obvious differences at major landmarks. Figure 17 illustrates this more clearly as the shift of emissions intensity between human population and StreetLight data surrogates.

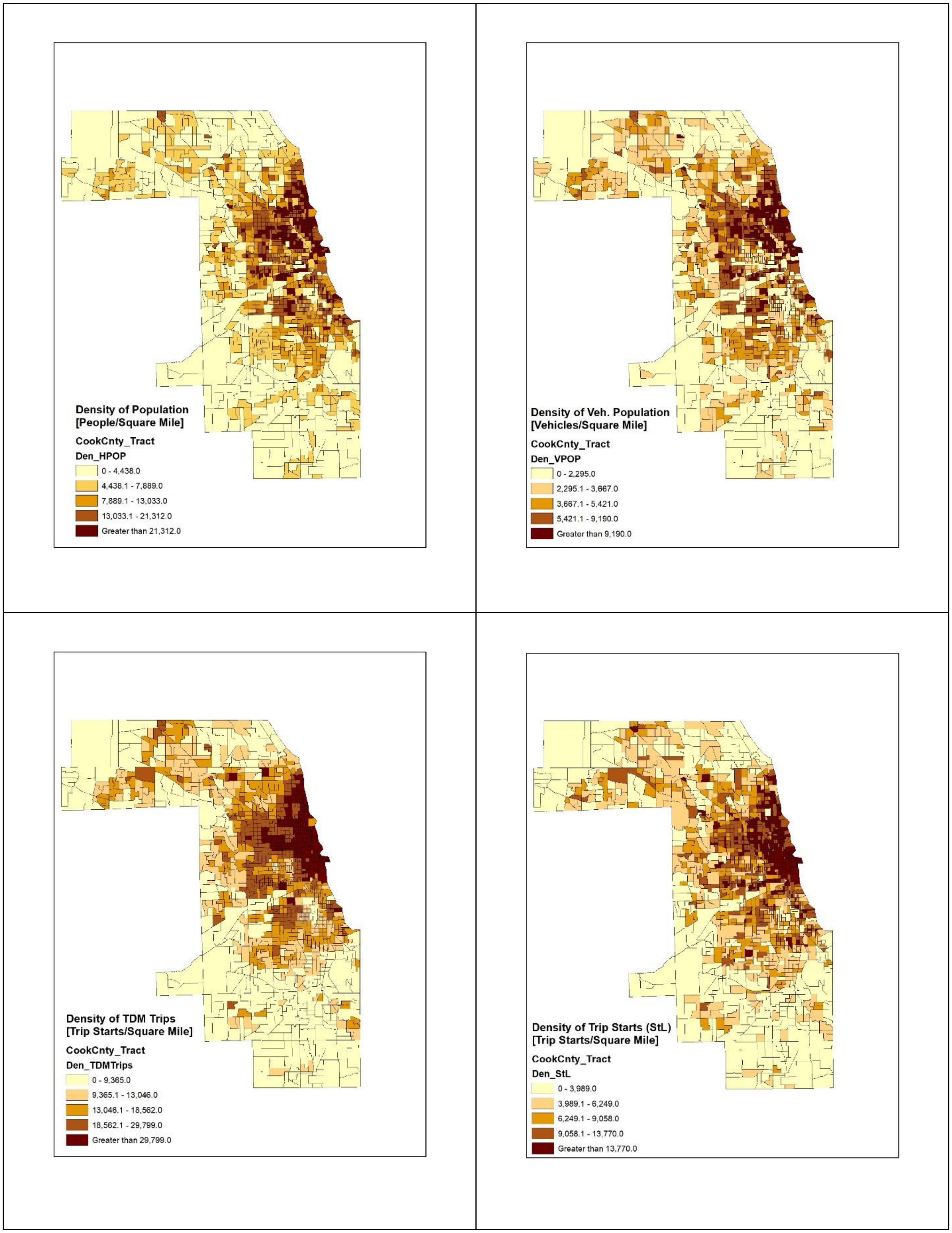


Figure 16. Potential Methods to Spatially Distribute Start Emissions – Cook County, Illinois

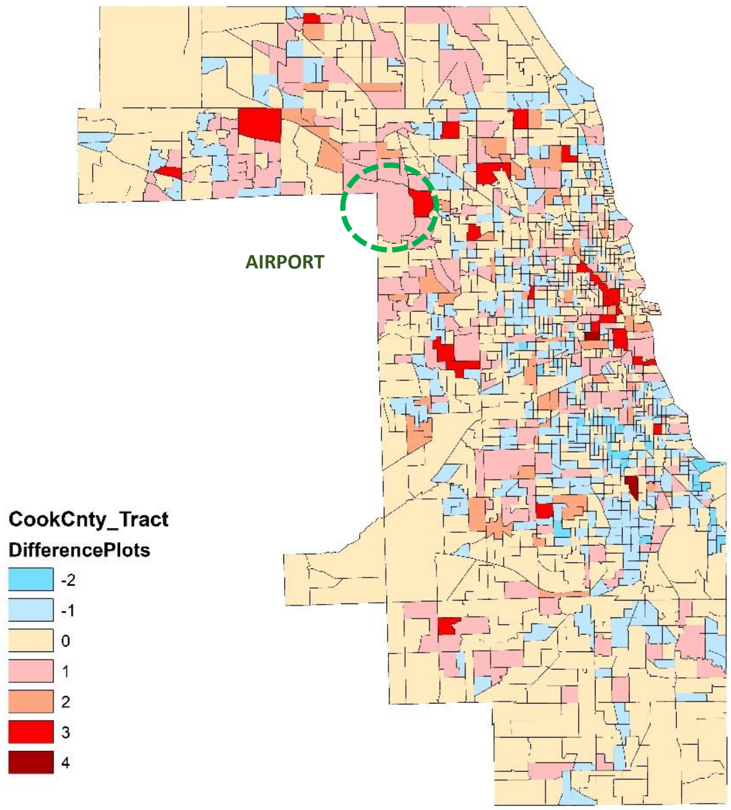


Figure 17. Difference Plot for Cook County (StreetLight – Human Population Quintile)

Figure 18 shows the Las Vegas Valley rather than the whole of Clark County, Nevada. Most of Clark County's land area has a low relative density population and trips, with most of the density and variation in density focused in Las Vegas and Henderson. Of the three cities, Las Vegas shows the biggest differences in spatial distribution of population and trip density. There a noticeable gap in the human and vehicle population maps in the commercial area that includes McCarran International Airport, the University of Nevada Las Vegas and hotels and shopping malls just east of The Strip (which runs along Las Vegas Blvd). In contrast, the TDM trip origins and telematics trip starts show this area as having the most starts. The TDM trips also do a better job than population of reflecting the starts in the town of Henderson, located to the southeast of Las Vegas. Figure 19 shows the shift in emissions intensity as difference plot.

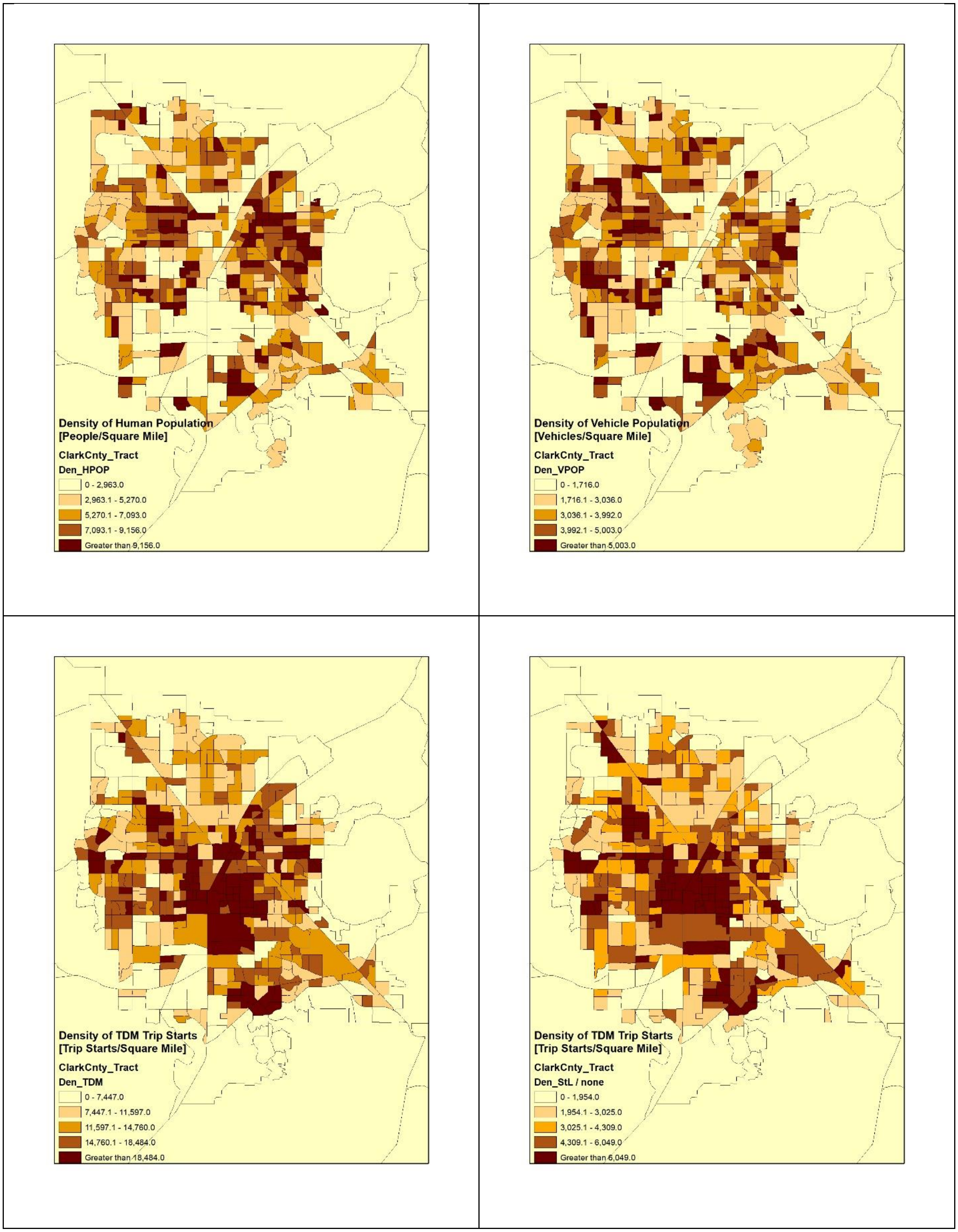


Figure 18. Potential Methods to Spatially Distribute Start Emissions – Clark County, Nevada

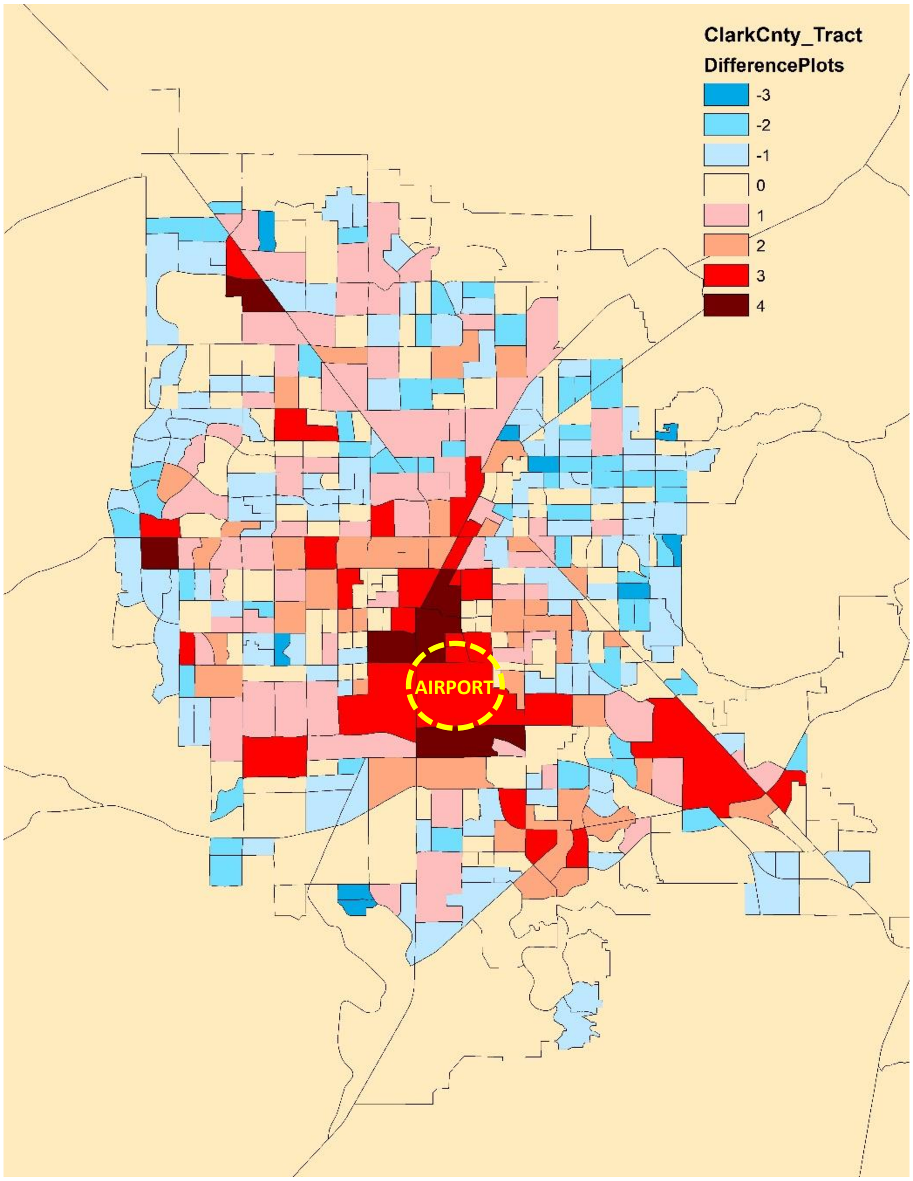


Figure 19. Difference Plot for Central Clark County (StreetLight - Human Population Quintile)

4.2.2.3 Spatial Analysis Conclusions

Out of the three alternative surrogates of humans, vehicles, and TDM, the TDM trip density by census tract most closely follows the spatial patterns of the sample of actual trip starts from the telematics data. While population datasets have the benefit of being no-cost and widely available, they tend to miss hot spots of start activity in dense commercial areas and at airports. The location of airports is highlighted in Figure 15-17, showing an increased concentration of start activity with StreetLight data in all three counties.

5.0 Task 4: Emissions Analysis

To evaluate emissions inventory impacts, MOVES was run at the County Domain/Scale in inventory calculation mode at the hourly level for the year 2014. The pollutants included volatile organic compounds (VOC), carbon monoxide (CO), oxides of nitrogen (NO_X), and fine particulate matter ($PM_{2.5}$). The inputs for the model runs included the ambient meteorology, vehicle fleet descriptions, fuel parameters, inspection & maintenance programs, and activity (VMT, population, and speeds) that were included in the county databases (CDBs) that EPA publicly released for Version 1 of the 2014 National Emissions Inventory (NEI).[9]

The three CDBs were downloaded from EPA's website, and were run directly as-is to produce the 2014 annual baseline emissions estimates, with one exception: the Fulton, GA, CDB included a populated *startsPerDay* table, which precluded the use of the new StreetLight-based *startsPerVehicle* table. Therefore, the submitted *startsPerDay* table was cleared, causing MOVES to use to the MOVES2014a default starts/vehicle for the baseline case and the new *startsPerVehicle* table for the other scenarios.

The primary scenario run for each county included both of the new tables derived from the county-specific StreetLight data (*startsPerVehicle* and *ImportStartsOpModeDistribution*). In addition, we ran one scenario with just the *startsPerVehicle* table replaced and another with just the *ImportStartsOpModeDistribution* table replaced to assess the relative impacts of these inputs, although in reality these inputs are linked since more frequent starts implies shorter soak times. We also performed two sensitivity analysis runs using StreetLight data from individual tracts representing the 10th percentile and 90th percentile, in terms of starts/vehicle/day, for Fulton, GA, as described in Section 3.

With the exception of the *startsPerVehicle* and *ImportStartsOpModeDistribution* table, all other inputs (e.g., age distribution, fuels, etc.) were the same between the baseline and scenario runs.

5.1 Results

For an overall comparison of the scenarios, annual average day emissions were calculated for each pollutant of interest. First, the monthly average day emissions for each MOVES run were calculated as a weighted average of the weekday (*dayID*=5) and weekend (*dayID*=2) emissions, using the number of days per week (5 and 2, respectively) as the weighting factor. These were then scaled to the annual level using the *noOfDays* field of the MOVES *monthOfAnyYear* table.

For Fulton, GA, the baseline Passenger Car and Passenger Truck emissions totals for VOC, CO, NO_x, and $PM_{2.5}$, are shown in Table 6. The corresponding emissions for the MOVES run using the StreetLight starts/vehicle and soak time distributions are also shown, along with the change in emissions from the baseline scenario on a percentage basis.

Table 6. Fulton County Emission Results – Primary Case

County	Source Type	Pollutant	Baseline Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions: Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions: Change from Baseline
Fulton, GA	Passenger Car (*sourceTypeID*=21)	VOC	8.9	9.5	7%
		CO	86.5	91.1	5%
		NO_x	9.0	9.5	6%
		$PM_{2.5}$	0.3	0.3	2%
	Passenger Truck (*sourceTypeID*=31)	VOC	6.1	6.6	8%
		CO	75.8	80.9	7%
		NO_x	8.5	8.8	4%
		$PM_{2.5}$	0.2	0.2	2%

The Passenger Car and Passenger Truck emissions totals for VOC, CO, NO_x, and $PM_{2.5}$ for the sensitivity cases, along with the change in emissions from the baseline scenario on a percentage basis, are shown in Table 7.

Table 7. Fulton County Emission Results – Sensitivity Case

Source Type	Pollutant	Baseline Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions, 10th percentile case: Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions, 10th percentile case: Emissions (Change from Baseline)	StreetLight Starts/Vehicle and Soak Time Distributions, 90th percentile case: Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions, 90th percentile case: Emissions (Change from Baseline)
Passenger Car (*sourceTypeID*=21)	VOC	8.9	8.9	0%	10.6	19%
	CO	86.5	86.8	0%	98.8	14%
	NO_x	9.0	8.8	-3%	10.9	20%
	$PM_{2.5}$	0.3	0.3	-1%	0.3	7%
Passenger Truck (*sourceTypeID*=31)	VOC	6.1	6.1	1%	7.4	22%
	CO	75.8	76.6	1%	88.4	17%
	NO_x	8.5	8.3	-2%	9.8	16%
	$PM_{2.5}$	0.2	0.2	0%	0.2	7%

Figure 20 and 21 show the total change in annual average day emissions and the change in just the start emissions. The full StreetLight scenario (starts/vehicle and soak times) is shown (green bar), along with the results from the runs with only the *startsPerVehicle* table (blue) or the *ImportStartsOpModeDistribution* table (orange) replaced. The 10th percentile (red) and 90th percentile (purple) sensitivity cases are also included. As these figures show, the overall effect of the full StreetLight inputs is approximately equal to the additive effects of the individual inputs.

The sensitivity cases highlight a different trend than that of the county averages. County averages suggest little emissions sensitivity to a slight drop in starts/vehicle vs. MOVES defaults, but a large increase in emissions due to longer soak durations. The sensitivity cases suggest that for wider variations in starts/vehicle, soak distribution does not necessarily vary accordingly, so changes in emissions are more sensitive to the changes in starts/vehicle. Using the activity inputs from Figures 10 and 11 presented earlier, the 10^{th} percentile case (based on a census tract with 4 starts/day) shows a drop in NOx and $PM_{2.5}$ start emissions but a very slight increase in VOC and CO start emissions; this suggests that the drop in starts/vehicle is offset by the increase in soak times, but not to the degree seen in the average case, where longer soak times led to a larger overall increase in emissions. Emission results from the 90^{th} percentile case (based on a tract with 7 starts/day) reflect the significant increase in starts; and for this particular tract, as shown in Figure 11, the soak distribution is similar to MOVES default but with longer average soak times for the morning peak; this adds further emissions, resulting in large increase in start emissions (35 to 55 percent) and total emissions for this case. Results in other locations will depend on the start/soak tradeoff. These results reinforce the importance of using consistent start and soak inputs from the same data source.

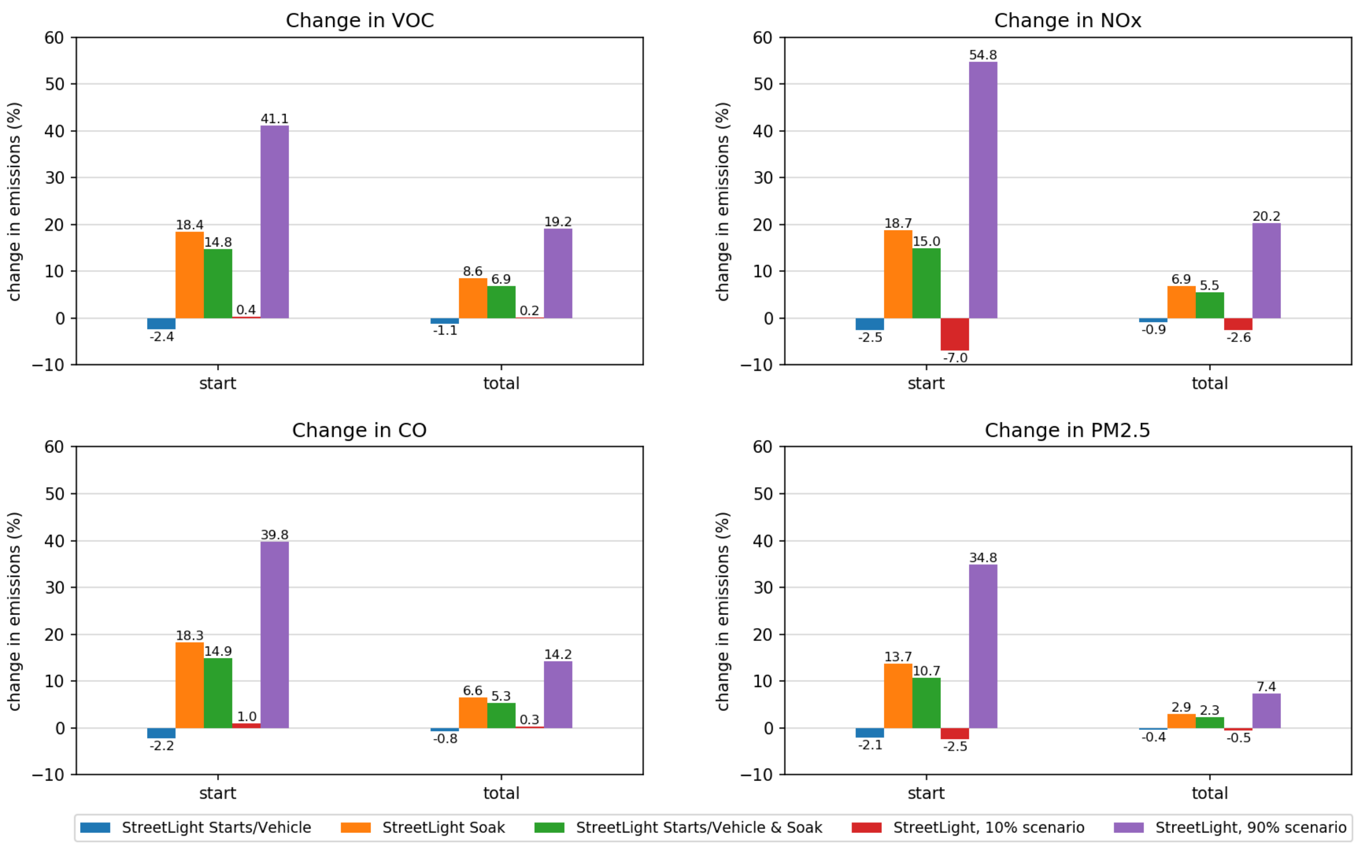


Figure 20. Fulton County Passenger Car change daily emissions by process and scenario

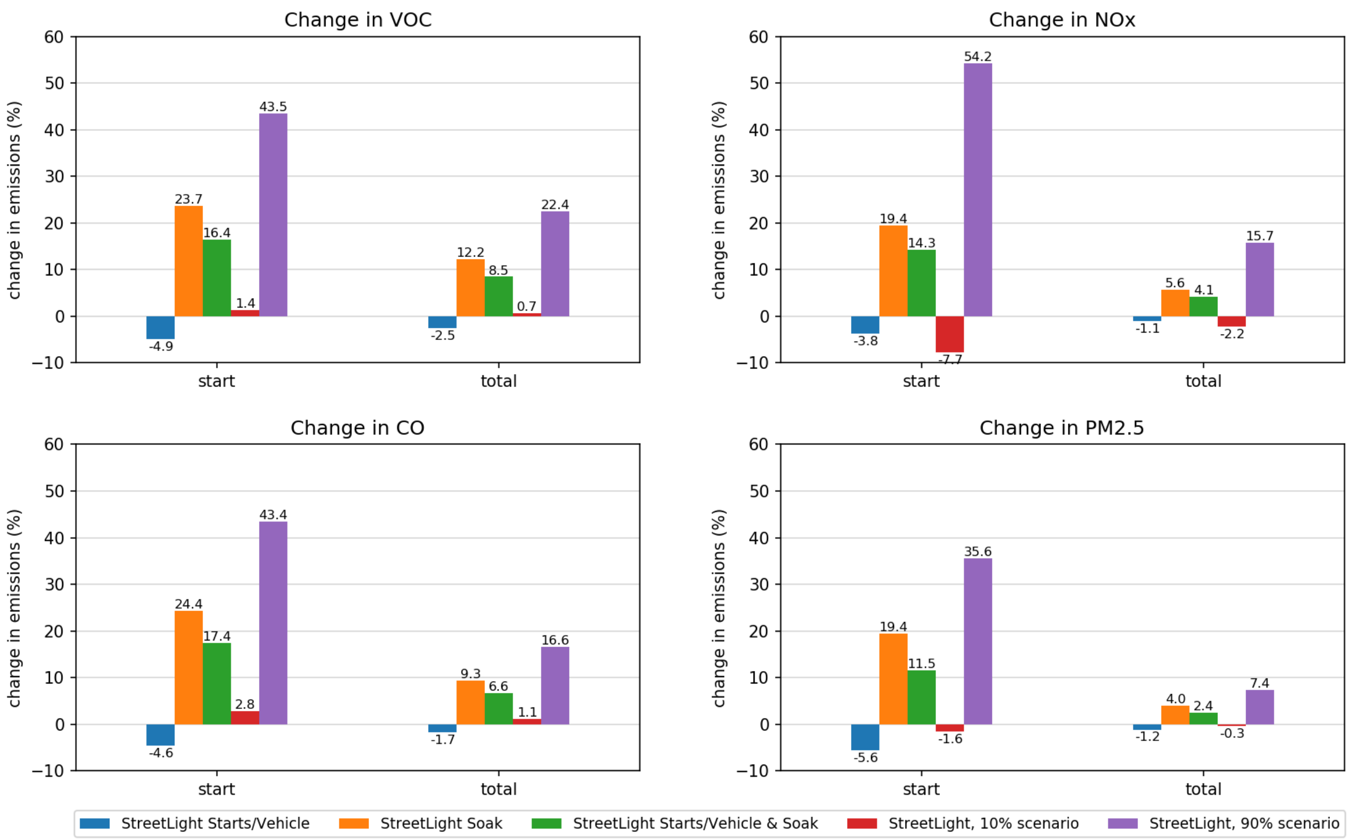


Figure 21. Fulton County Passenger Truck change daily emissions by process and scenario

For air quality modeling, it is also important to consider the hourly differences in emissions between scenarios. Representative plots for Passenger Cars on a July weekday and weekend, showing hourly differences on a percentage basis, are provided in Figure 22 and Figure 23, respectively. Analogous plots for Passenger Trucks are shown in Figures 24 and 25.

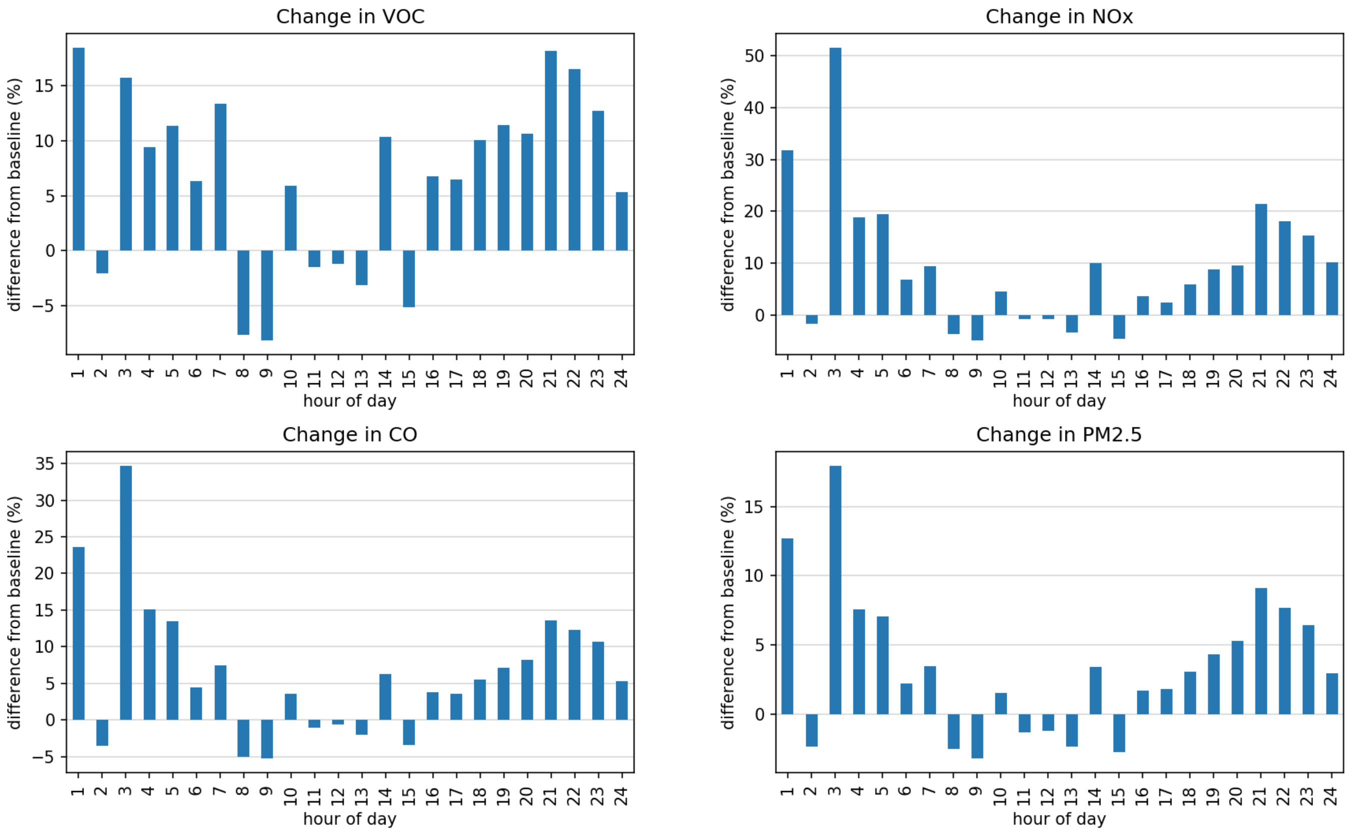


Figure 22. Fulton County Passenger Car weekday hourly impacts of StreetLight Metrics

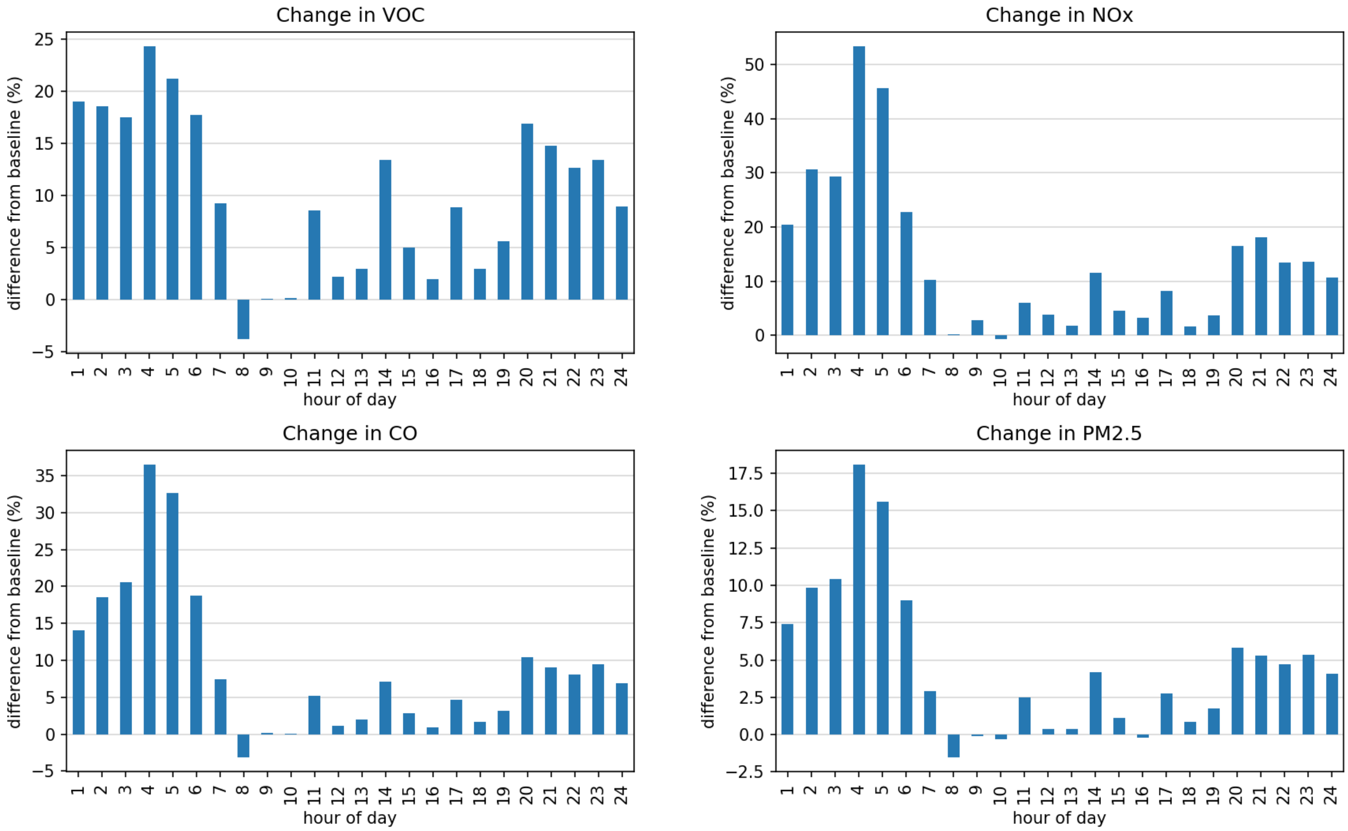


Figure 23. Fulton County Passenger Car weekend hourly impacts of StreetLight Metrics

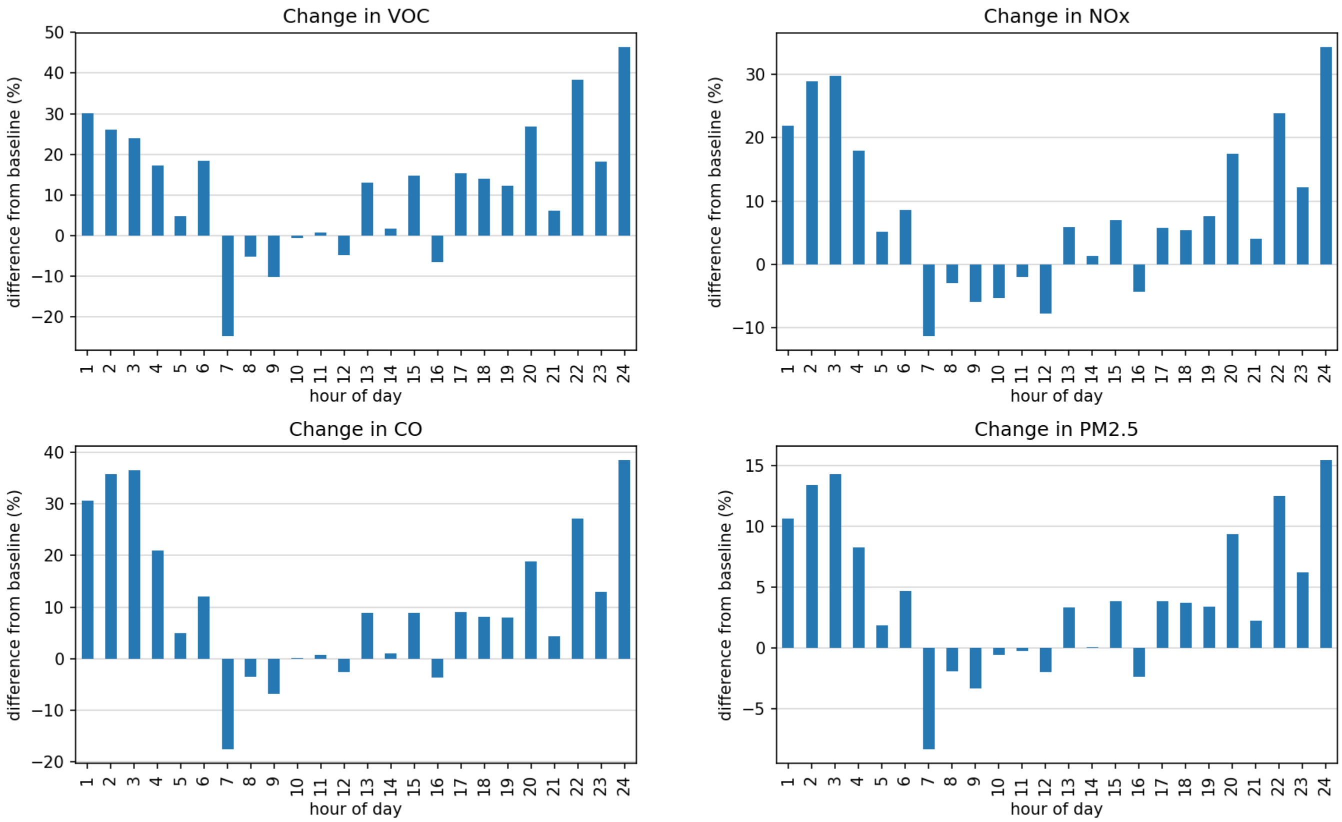


Figure 24. Fulton County Passenger Truck weekday hourly impacts of StreetLight Metrics

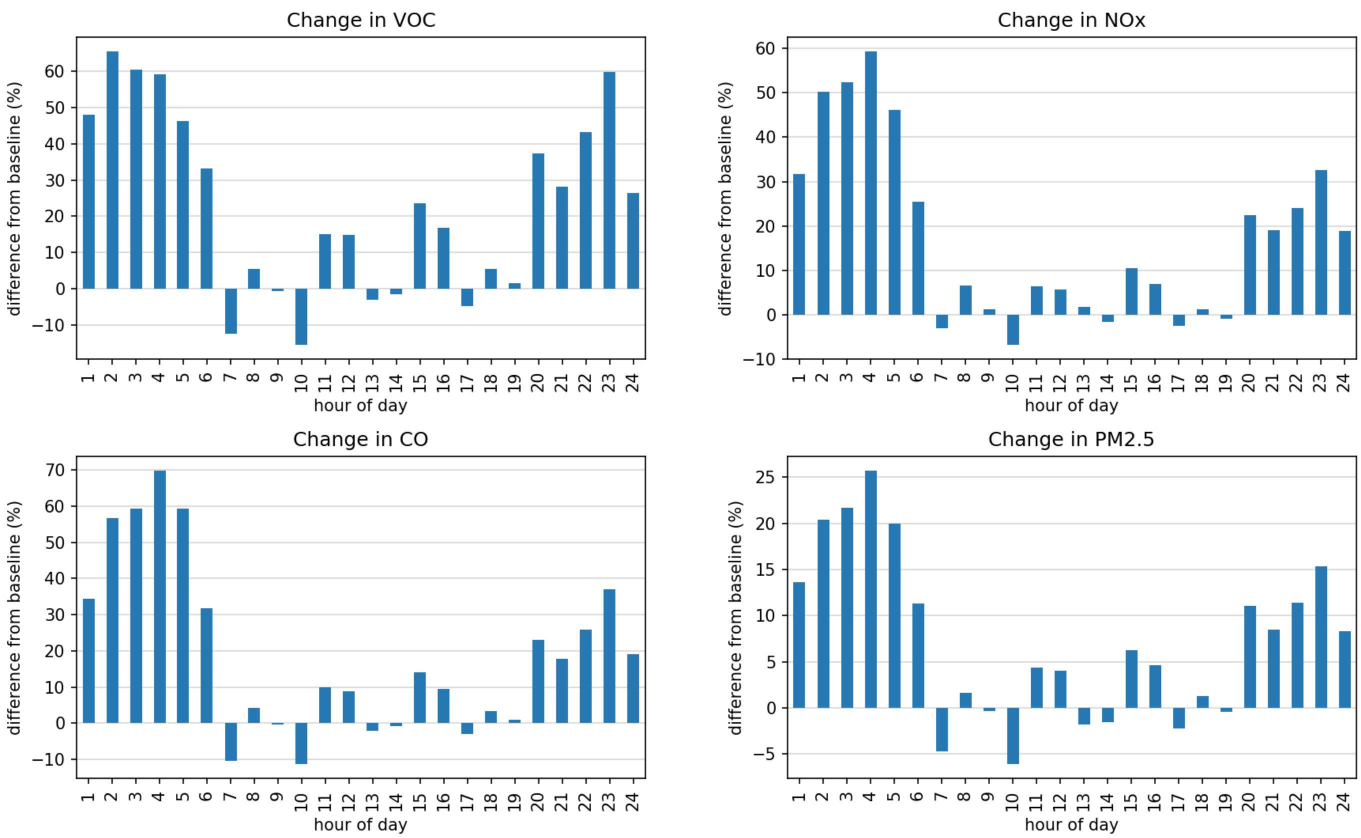


Figure 25. Fulton County Passenger Truck weekend hourly impacts of StreetLight Metrics

Results from the Cook, IL, MOVES runs are shown below: annual average day emissions totals for the baseline and full StreetLight scenarios (Table 8); percent change in start and total emissions for all scenarios (Figures 26 and 27); and hourly differences between the baseline and full StreetLight scenarios for representative July days (Figures 28-31).

Table 8. Cook County Emission Results

County	Source Type	Pollutant	Baseline Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions	
				Emissions (TPD)	Change from Baseline
Cook, IL	Passenger Car (*sourceTypeID*=21)	VOC	30.6	32.8	7%
		CO	261.7	277.6	6%
		NO_x	29.0	30.2	4%
		$PM_{2.5}$	1.0	1.0	3%
	Passenger Truck (*sourceTypeID*=31)	VOC	23.4	25.4	9%
		CO	257.7	274.3	6%
		NO_x	32.5	33.5	3%
		$PM_{2.5}$	0.7	0.8	3%

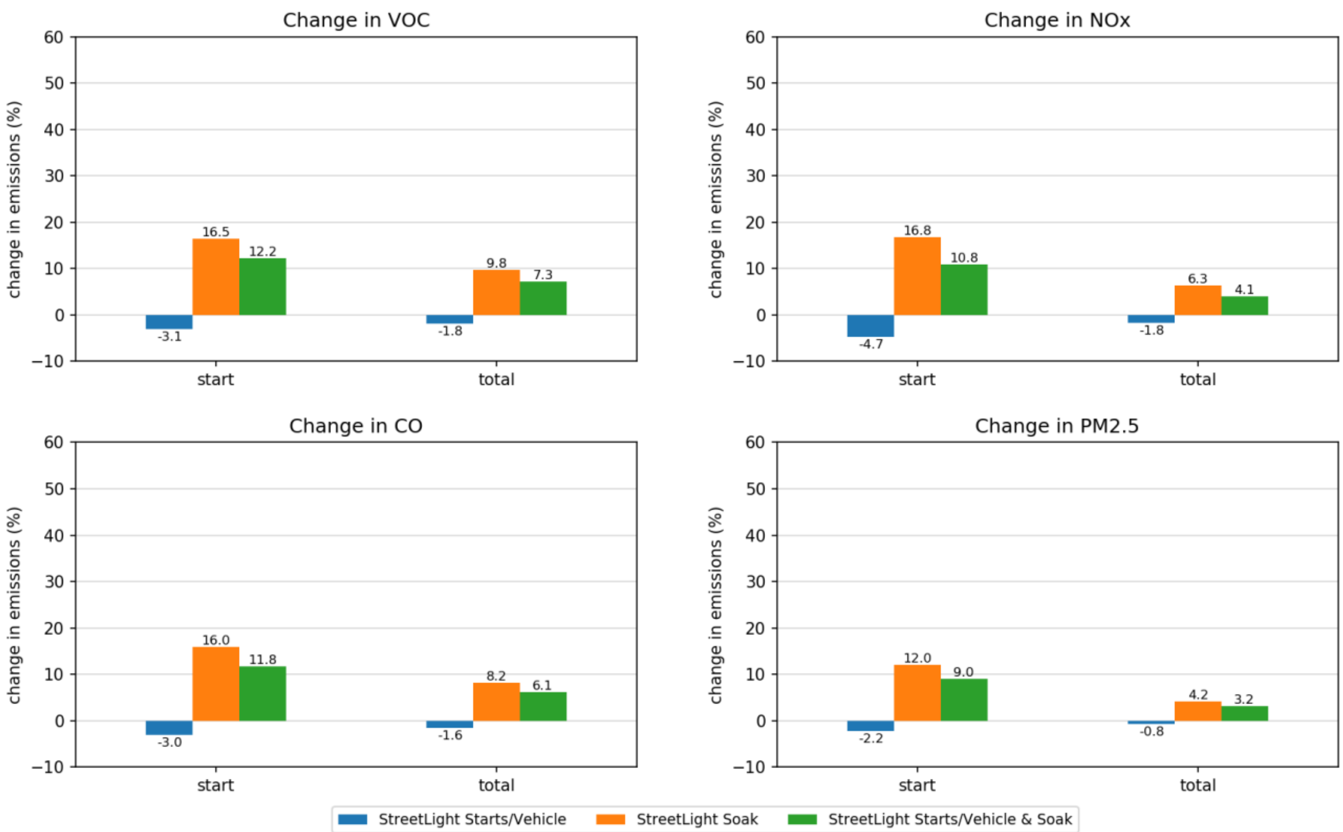


Figure 26. Cook County Passenger Car change daily emissions by process and scenario

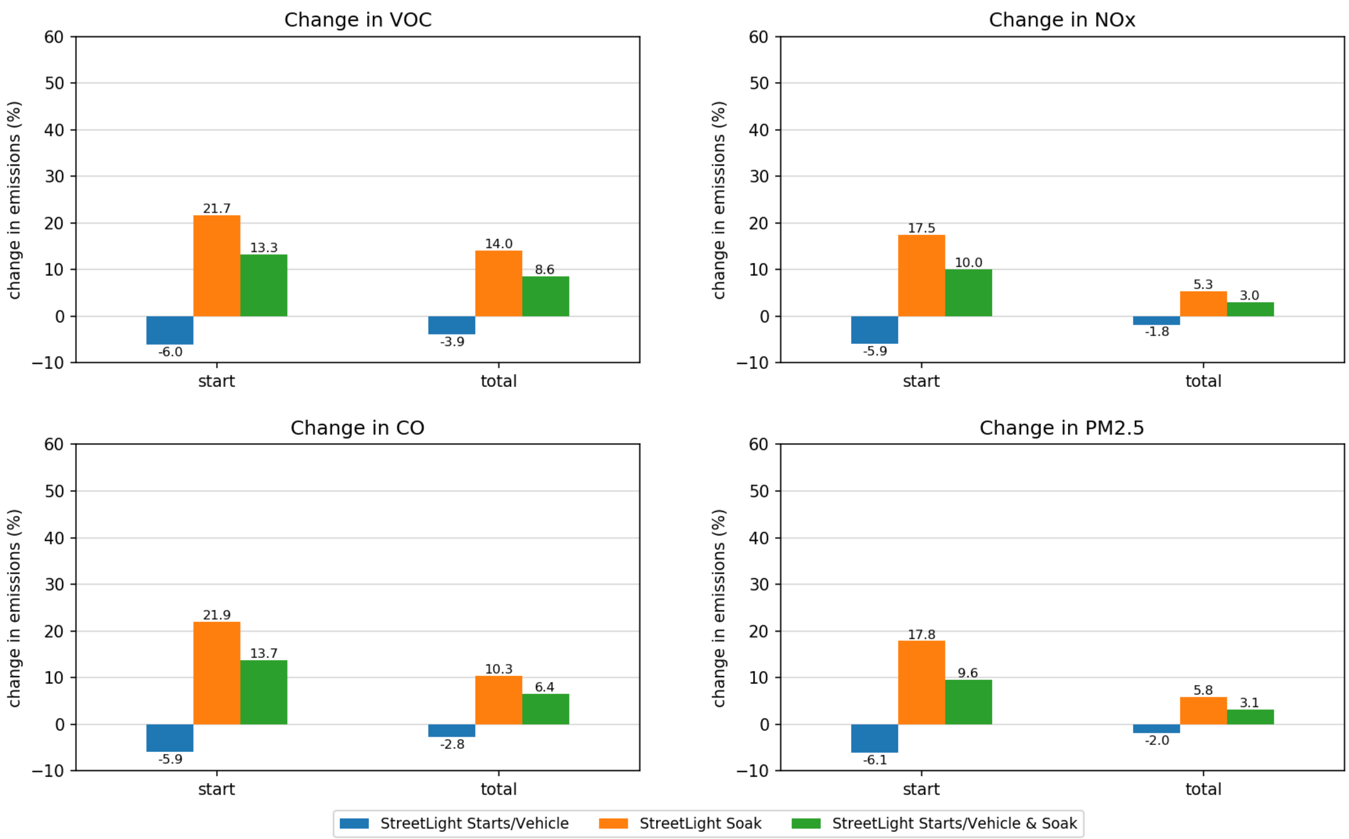


Figure 27. Cook County Passenger Truck change daily emissions by process and scenario

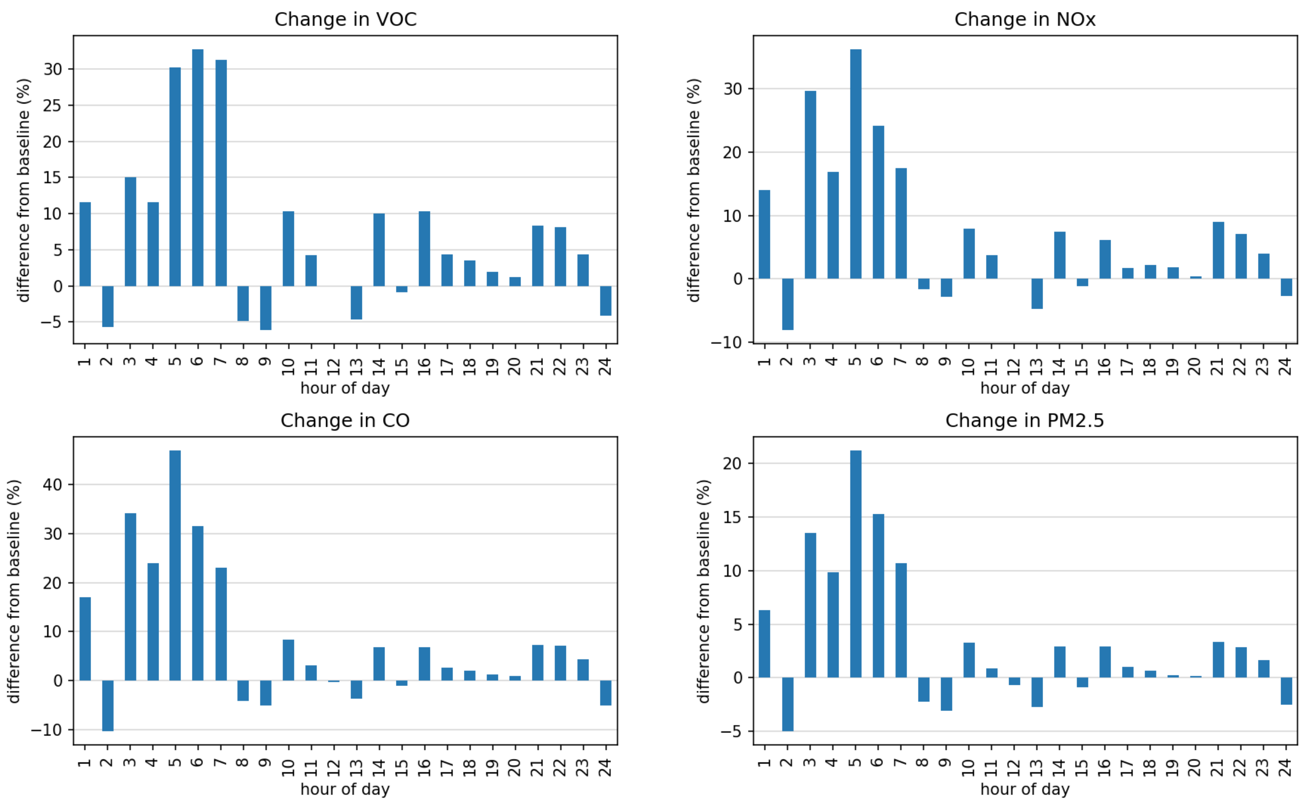


Figure 28. Cook County Passenger Car weekday hourly impacts of StreetLight Metrics

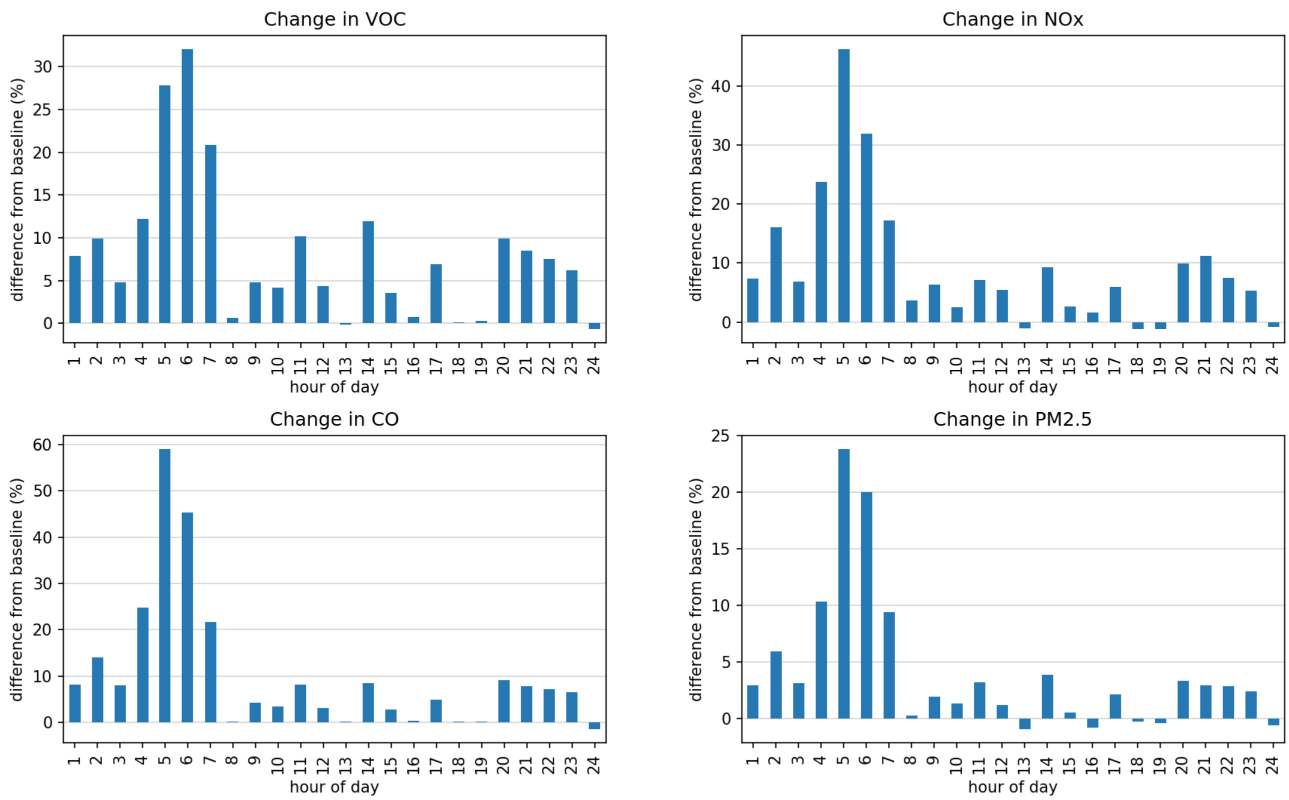


Figure 29. Cook County Passenger Car weekend hourly impacts of StreetLight Metrics

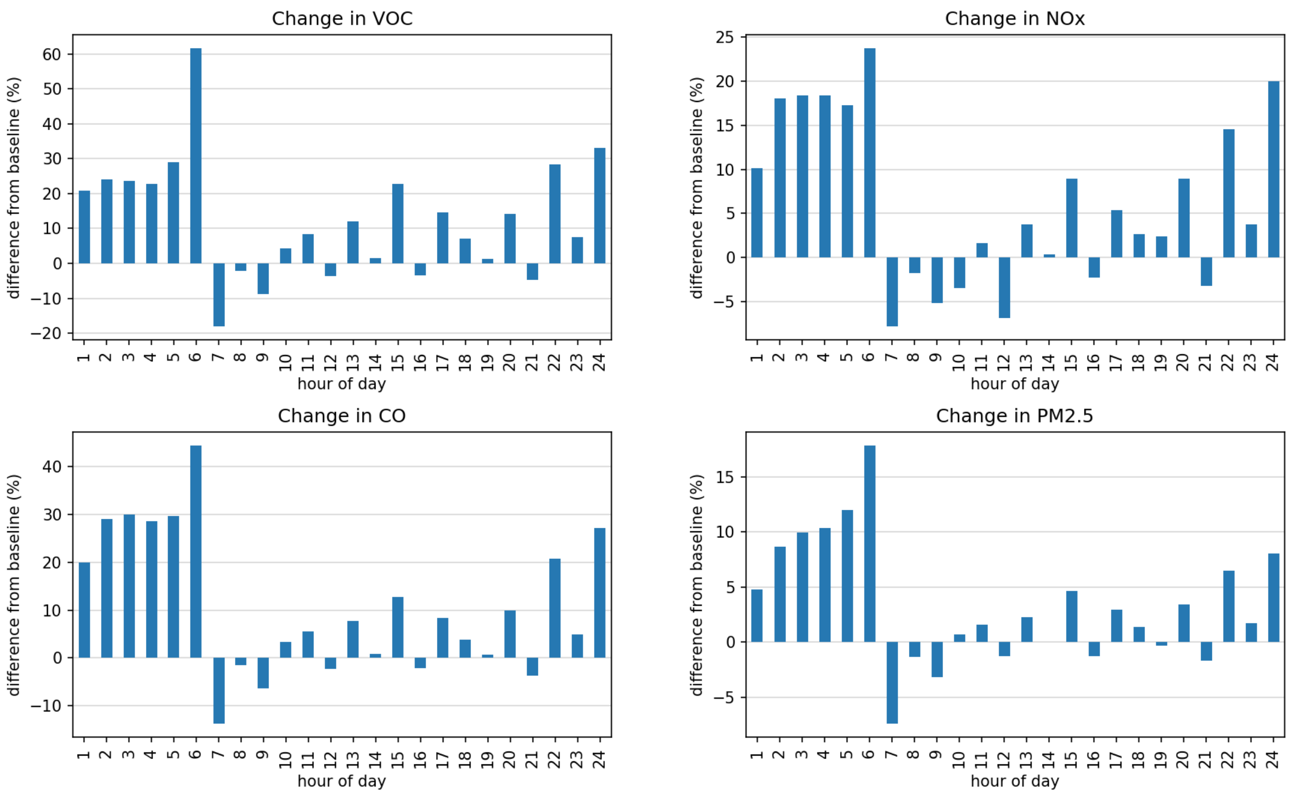


Figure 30. Cook County Passenger Truck weekday hourly impacts of StreetLight Metrics

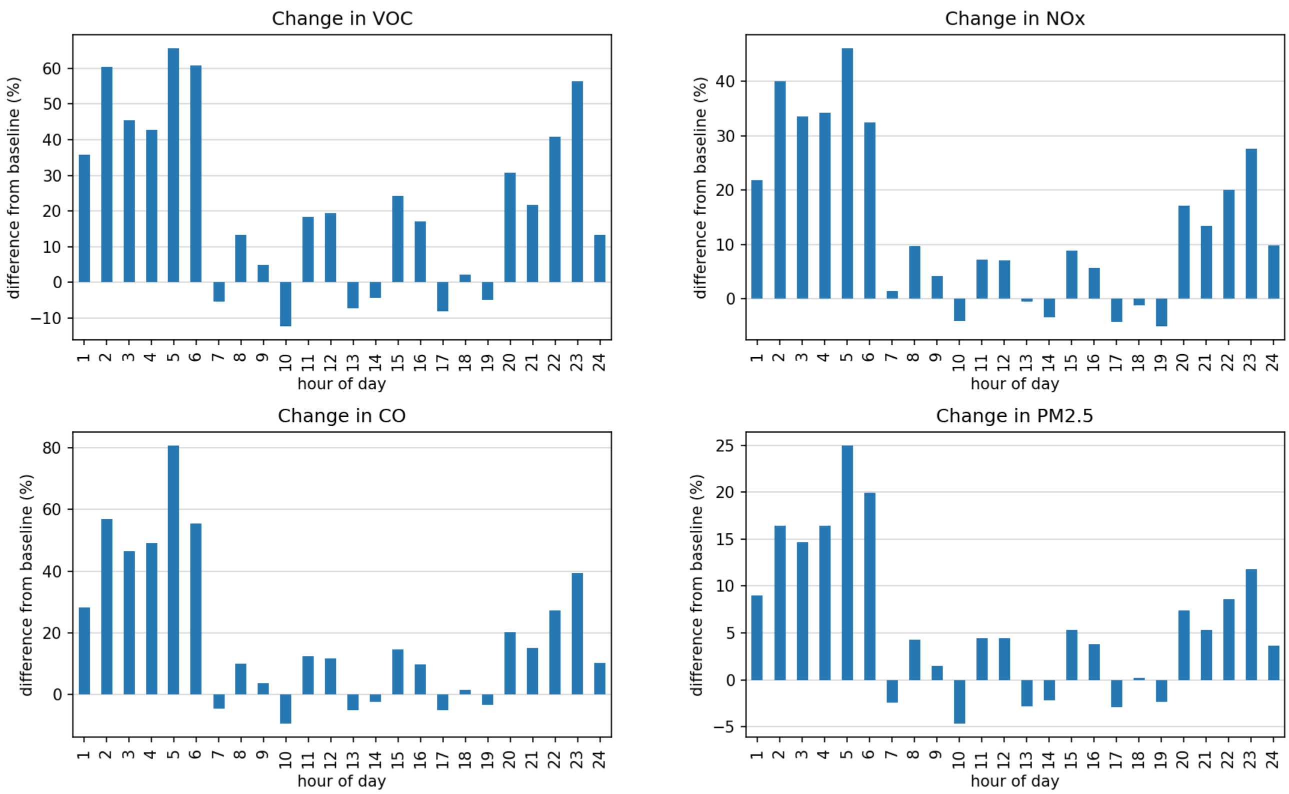


Figure 31. Cook County Passenger Truck weekend hourly impacts of StreetLight Metrics

Results from the Clark, NV, MOVES runs are shown below: annual average day emissions totals for the baseline and full StreetLight scenarios (Table 9); percent change in start and total emissions for all scenarios (Figures 32 and 33); and hourly differences between the baseline and full StreetLight scenarios for representative July days (Figures 34-37).

Table 9. Clark County Emission Results

County	Source Type	Pollutant	Baseline Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions: Emissions (TPD)	StreetLight Starts/Vehicle and Soak Time Distributions: Change from Baseline
Clark, NV	Passenger Car (*sourceTypeID*=21)	VOC	10.4	10.9	4%
		CO	108.7	112.0	3%
		NO_x	14.2	14.6	3%
		$PM_{2.5}$	0.3	0.3	1%
	Passenger Truck (*sourceTypeID*=31)	VOC	12.8	13.4	5%
		CO	159.7	166.2	4%
		NO_x	23.6	24.1	2%
		$PM_{2.5}$	0.3	0.3	1%

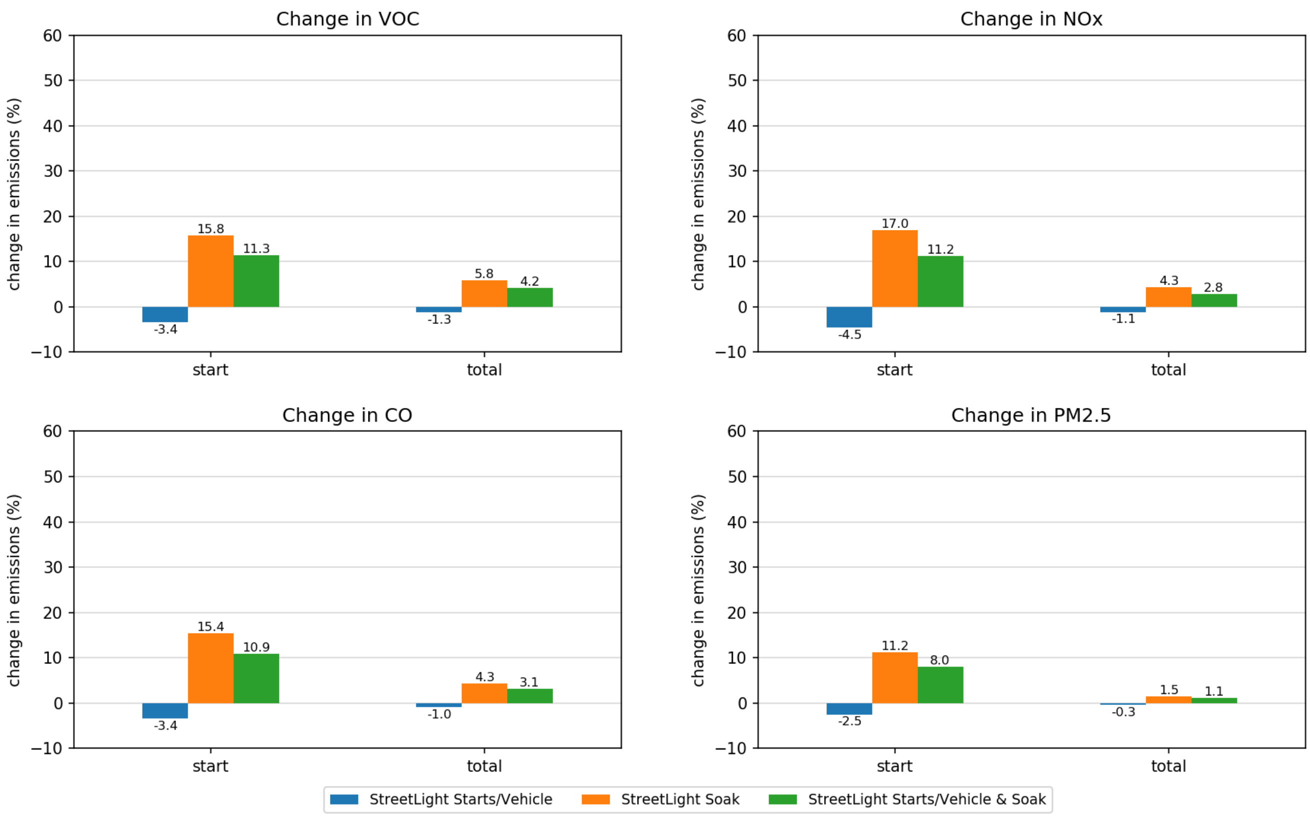


Figure 32. Clark County Passenger Car change daily emissions by process and scenario

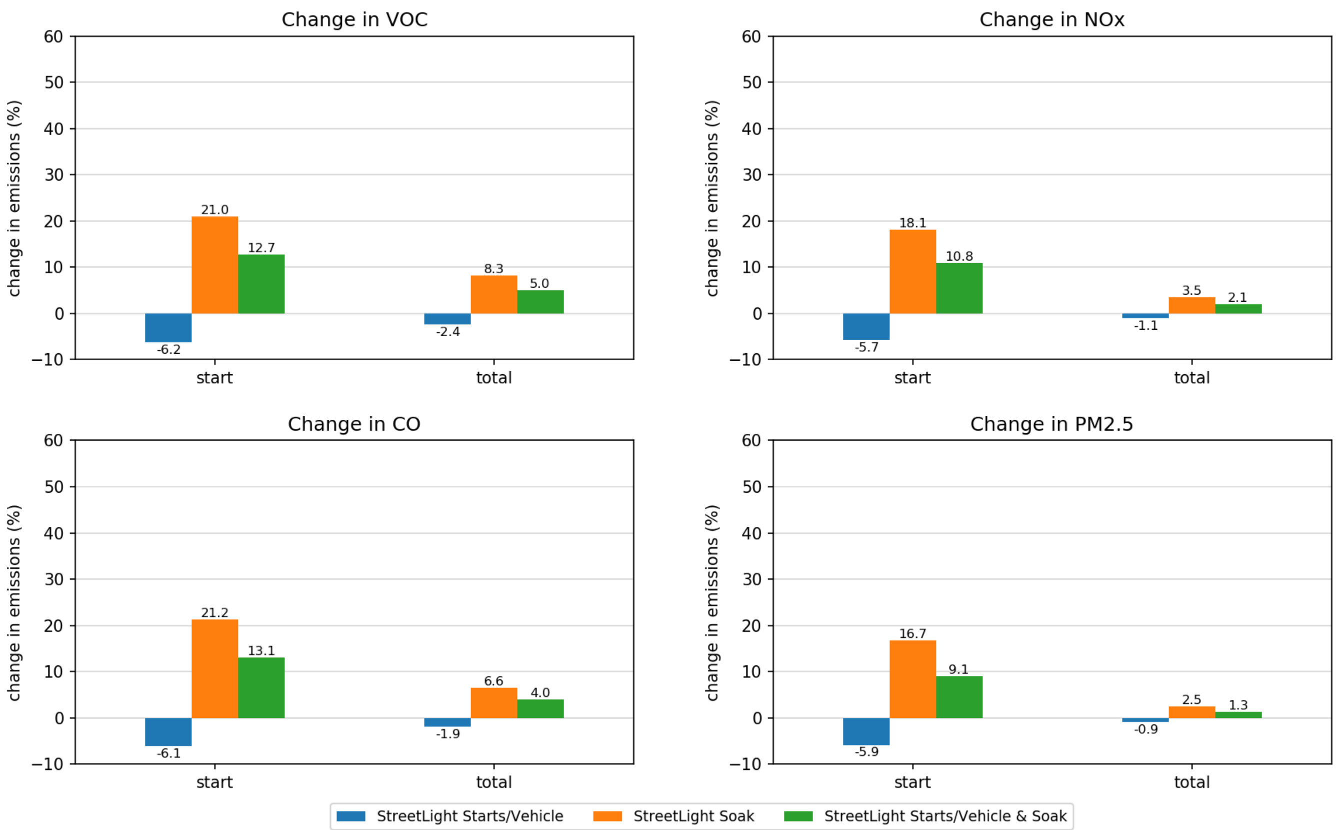


Figure 33. Clark County Passenger Truck change daily emissions by process and scenario

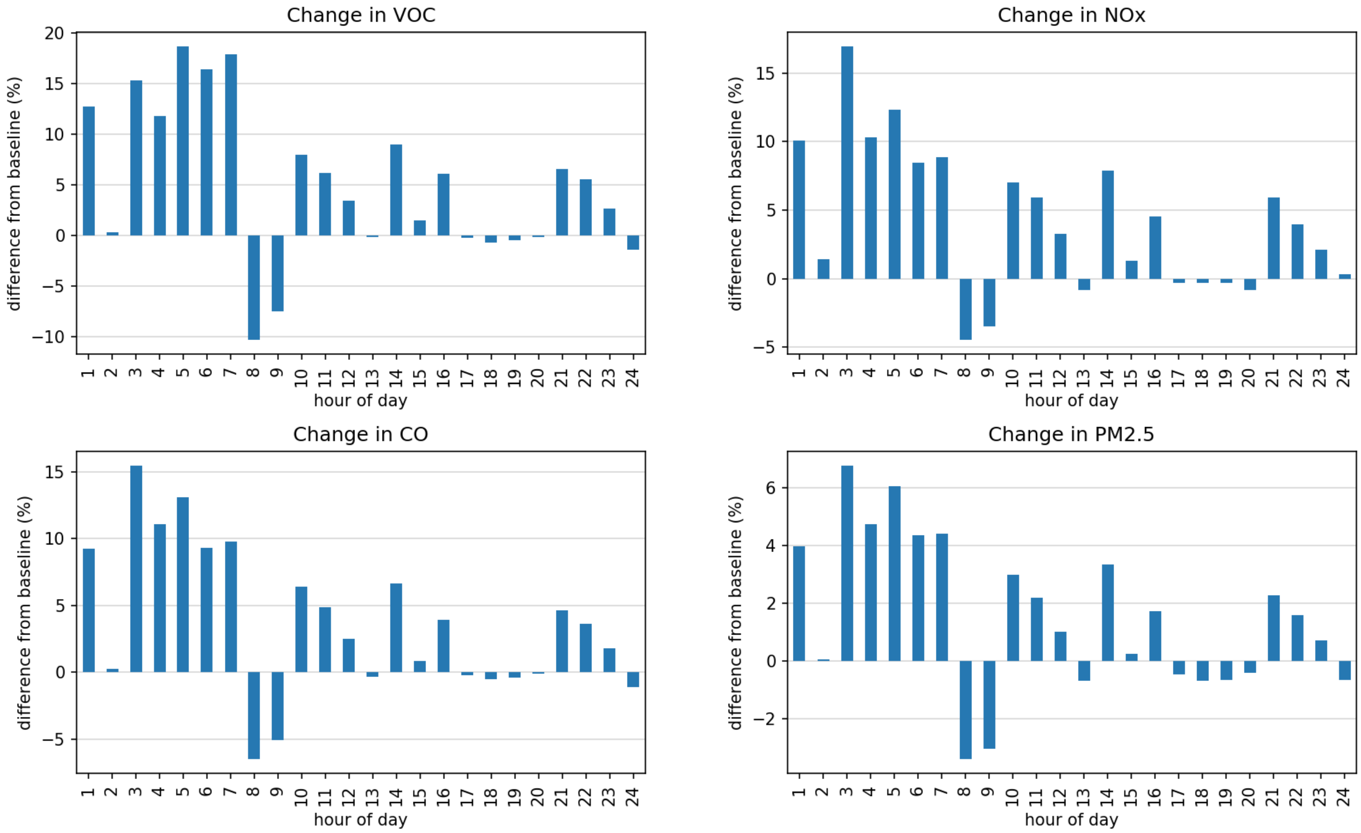


Figure 34. Clark County Passenger Car weekday hourly impacts of StreetLight Metrics

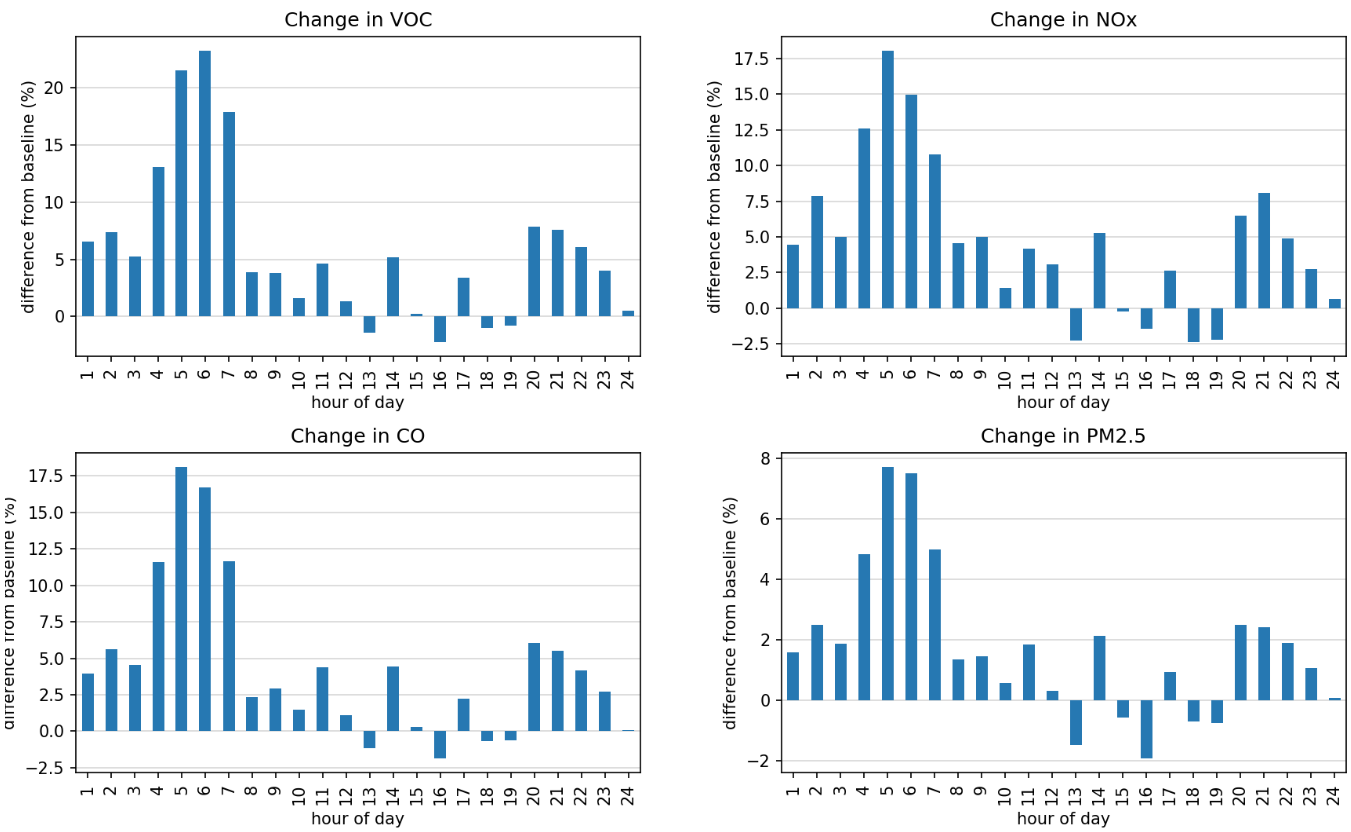


Figure 35. Clark County Passenger Car weekend hourly impacts of StreetLight Metrics

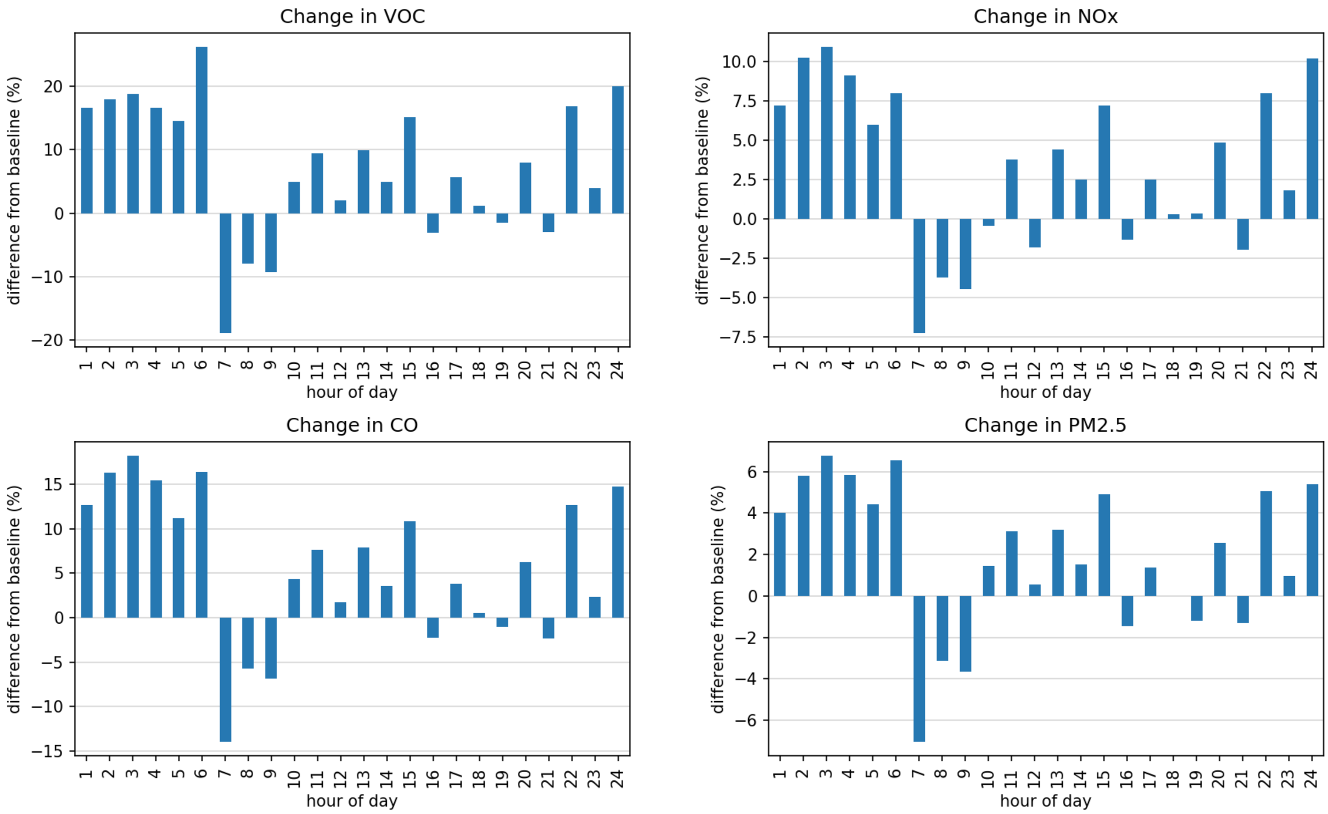


Figure 36. Clark County Passenger Truck weekday hourly impacts of StreetLight Metrics

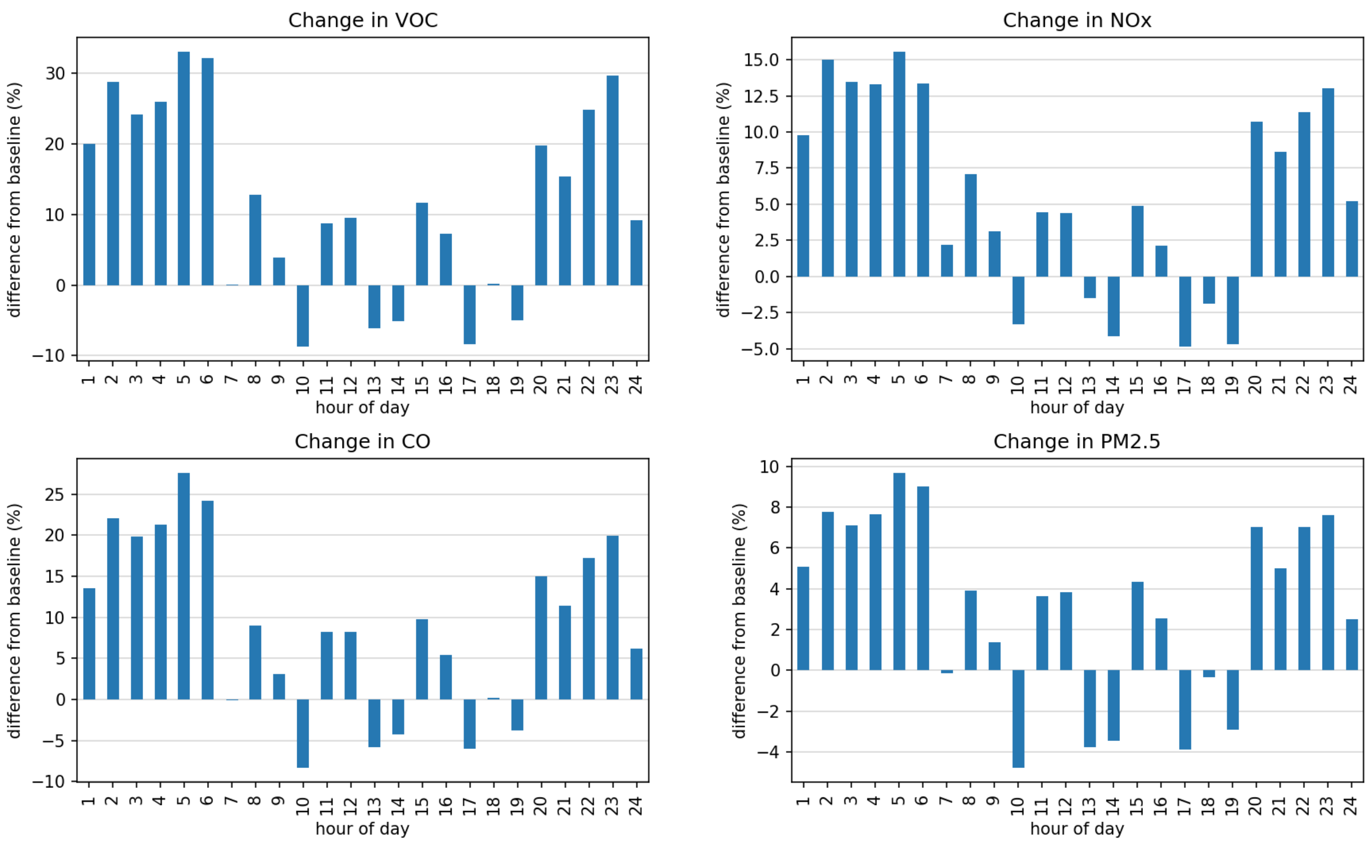


Figure 37. Clark County Passenger Truck weekend hourly impacts of StreetLight Metrics

5.2 Discussion of Emission Analysis Results

The three counties show similar trends in emissions when the StreetLight Data was used in place of the MOVES2014a start-related inputs employed for the 2014 NEI. At the county/daily level, fewer starts per vehicle in the StreetLight metrics led to a small drop in overall emissions, but when the updated soak distributions were also used, emissions went up. When updating MOVES start-related inputs, both starts per vehicle and soak distribution should be updated together, since they vary with each other. In terms of MOVES2014a emission effects, for the county average cases the longer soak times were more influential than a slight drop in starts. For this case, start emissions increased 10-15 percent for all analyzed pollutants with the updated inputs, resulting in an increase in total emissions (all processes) of up to 7 percent (for Fulton County VOC). The 90th percentile sensitivity case with higher starts showed much higher emission increases (35 to 55 percent for start emissions), though applying trip metrics from a single census tract to the entire county is not a realistic case.

The differences in hourly emissions were larger, though generally in the off-peak hours. Across county, pollutant and car/truck, the most consistent trend was a sharp increase in emissions between 8pm and 7am, with differences in peak hours mixed. These hourly differences would be important to account for in air quality modeling, where the temporal distribution of emissions affects the formation of ozone and secondary PM throughout the day.

6.0 Conclusions

Our assessment of location-based telematics data confirms that this emerging source holds great promise for improving start activity metrics for emissions and air quality modeling. Location-based service data (LBS) culled by StreetLight Data provides access to sample sizes that are orders of magnitude higher than those developed for activity studies that the emissions community has typically relied on. This project confirms that LBS can be used to generate the metrics of starts per vehicle and soak distribution that influence the magnitude of emissions inventories developed at the census tract level by month, day and hour.

For the pilot study, analysis of the StreetLight trip metrics found similarities at the aggregate county level between Fulton, Cook and Clark counties. The StreetLight data showed about one start/day less than the MOVES defaults for these counties and vehicle types, while average soak times were longer than MOVES defaults. This led to an overall increase in start emissions of 10-15 percent for VOC, CO, NOx and $PM_{2.5}$ compared to 2014 NEI inputs (of which only Fulton had supplied custom data). StreetLight county-level starts per vehicle compares well to the National House Travel Survey (NHTS) and California Household Travel Survey (CHTS), but are higher than averages from Verizon Telematics data from vehicle monitored for insurance purposes. These differences underscore the need for further study of telematics data sources to understand how they represent driving behavior of the entire vehicle fleet, including secondary household vehicles.

A major advantage of using LBS telematics data for vehicle activity is the spatial resolution possible. The data show that variability in trip metrics at the census tract level is significant, and that the spatial allocation of activity varies from the traditional surrogate of human population. The trip metric data also correlate well with TDM trip origins, and provides a possible means of validating these models, or filling gaps in spatial or temporal coverage.

The finding of this pilot 3-county study could be extended to broader areas, given the scope of LBS data across the U.S. and abroad. Potential applications for national emissions and air quality modeling could include: developing local-level inputs for the NEI as was done for A-100 with vehicle speeds; updating air quality modeling surrogates using telematics in place of human population; or updating temporal profiles of start activity (by month, day or hour) for emissions and air quality modeling. These improvements could be applied by local, state and regional modelers as well.

7.0 Acknowledgements

The authors gratefully acknowledge the sponsorship of the Coordinating Research Council's Atmospheric Impacts Committee, and project chairs Susan Collet and Steve McConnell; TDM data provided by Guy Rousseau (Atlanta Regional Commission), Craig Heither (Chicago Metropolitan Agency for Planning), and Beth Xie (Regional Transportation Commission of Southern Nevada); and GIS support from Gopi Manne of ERG.

8.0 References

1. Eastern Research Group, Inc., *Study of MOVES Information for the National Emission Inventory: CRC Project A-84*, CRC Final Report, October 2013.

2. Eastern Research Group, Inc., *MOVES Input Improvements for the 2011 NEI: CRC Project A-88*, CRC Final Report, October 2014.

3. Eastern Research Group, Inc., *Improvement of Default Inputs for MOVES and SMOKE-MOVES: CRC Project A-100*, CRC Final Report, February 2017.

4. Brzezinski. D and Verma, A. "Telematics and MOVES", EPA/ORISE Presentation to MOVES Model Review Workgroup, June 2017

5. Brzezinski, D. et al. "Using Telematics Data for MOVES Activity Inputs" Presentation at the 2017 International Emissions Inventory Conference, Baltimore August 2017

6. California ARB, "Vehicle Activity Profiles: Light Duty Vehicles" Presentation at EMFAC Workshop June 2017

7. www.carinsurance.com, accessed November 7 2017

8. U.S. EPA, *National Emissions Inventory, version 1 Technical Support Document*, Office of Air Quality Planning and Standards, December 2014.

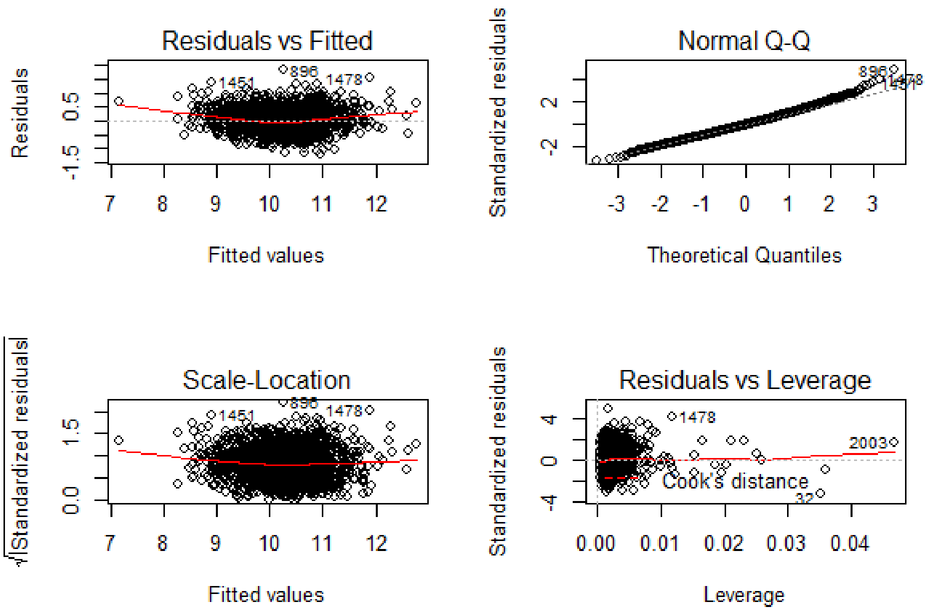
9. U.S. EPA, *Technical Support Document, Preparation of Emissions Inventories for the Version 6.2, 2011 Emissions Modeling Platform*, Office of Air Quality Planning and Standards, August 2015

Appendix A – Statistical Analysis

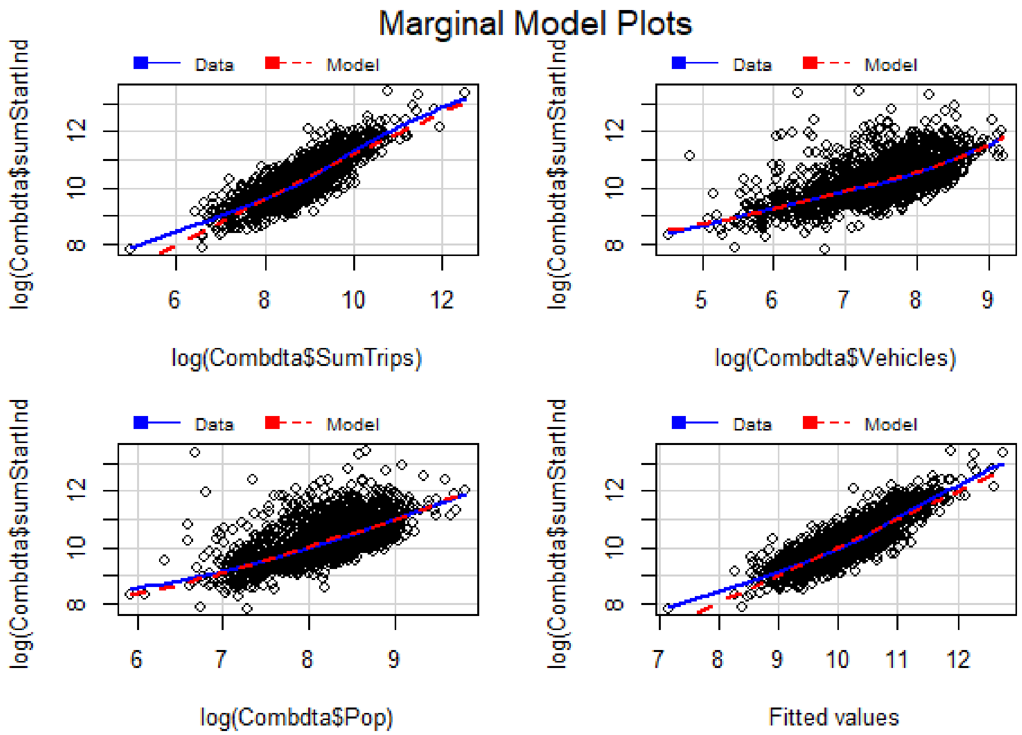
The diagnostic plots indicate how well the model meets standard linear regression conditions, while the MMP show the degree to which the fitted model values match with nonparametric estimates. The AVP displays the effect of each predictor on the overall model, having adjusted for effects of all other predictors.

TripIndex R Results and Plots, All Counties

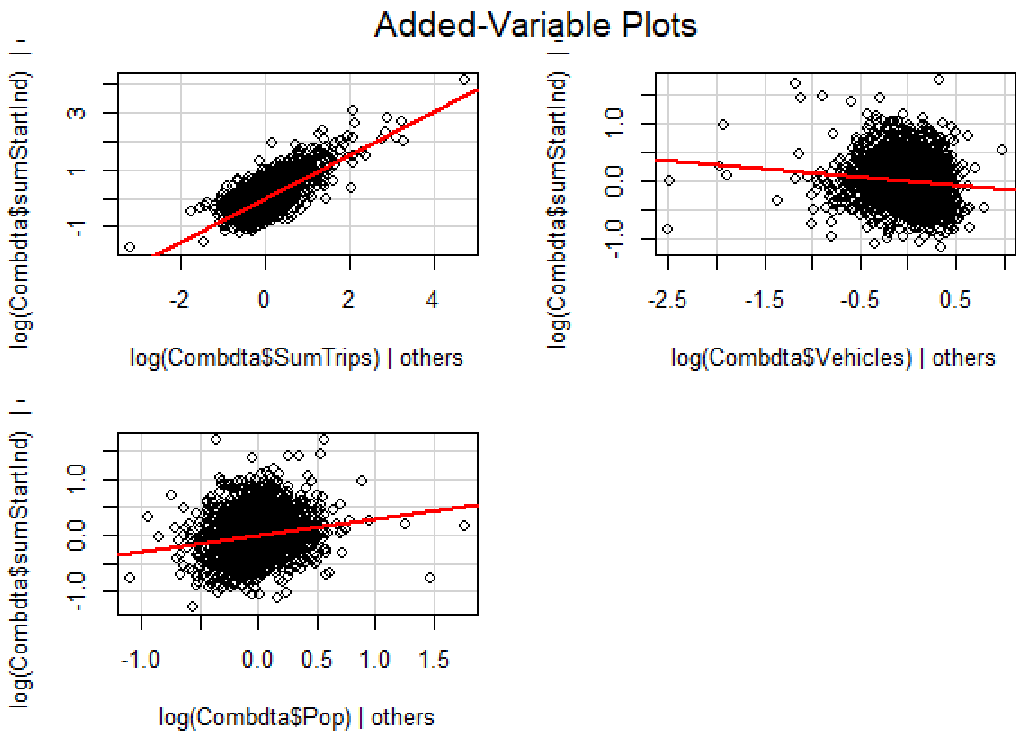
```
Call:
lm(formula = log(Combdta$sumStartInd) ~ log(Combdta$SumTrips) +
    log(Combdta$Vehicles) + log(Combdta$Pop))

Residuals:
     Min       1Q   Median       3Q      Max
-1.19027 -0.25234 -0.01676  0.22864  1.81499

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.17555    0.15830  13.743  < 2e-16 ***
log(Combdta$SumTrips)  0.75784    0.01601  47.345  < 2e-16 ***
log(Combdta$Vehicles) -0.13699    0.02807  -4.880 1.15e-06 ***
log(Combdta$Pop)       0.29354    0.03579   8.202 4.23e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3787 on 1968 degrees of freedom
  (38 observations deleted due to missingness)
Multiple R-squared:  0.7223,  Adjusted R-squared:  0.7219
F-statistic:  1706 on 3 and 1968 DF,  p-value: < 2.2e-16

> vif(m2)
log(Combdta$SumTrips) log(Combdta$Vehicles)      log(Combdta$Pop)
             2.044154              5.179601              4.431236
```



Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage



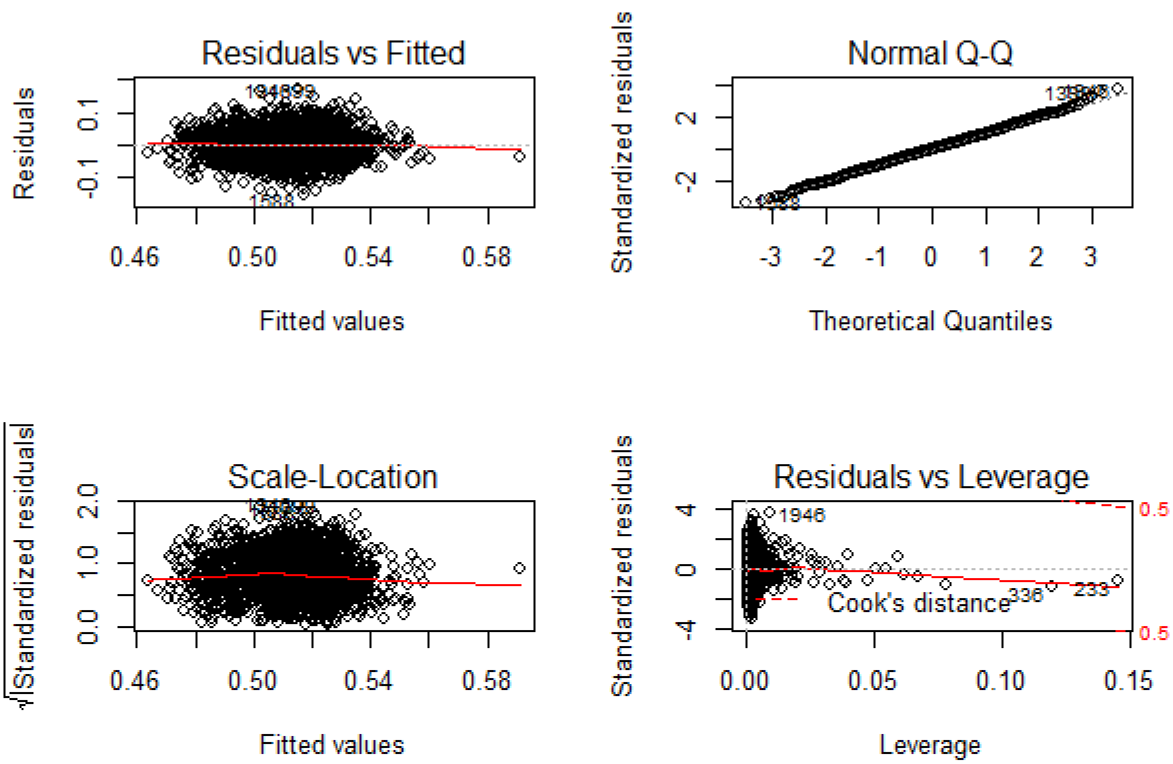
Marginal Model Plots
Data
Model
log(Combdta$sumStartInd
log(Combdta$SumTrips)
log(Combdta$Vehicles)
log(Combdta$Pop)
Fitted values

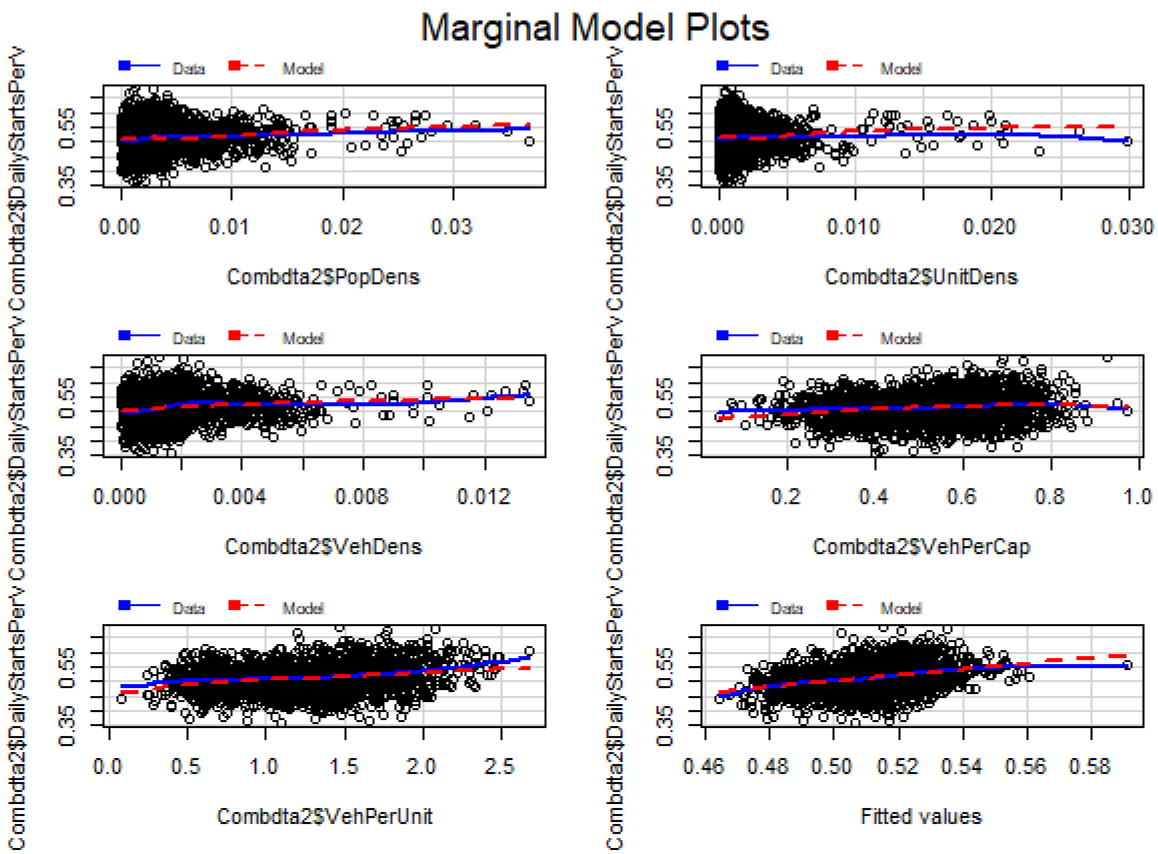


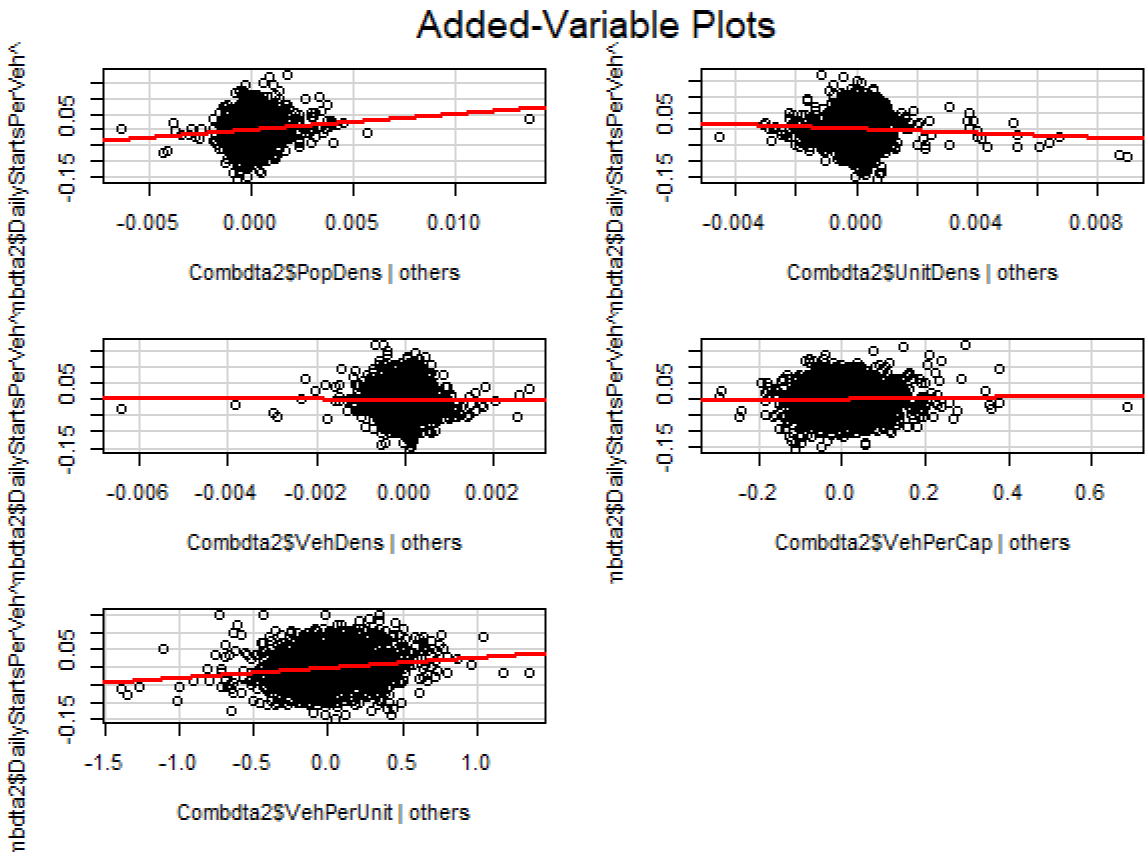
StartsPerVehicle R Results and Plots, All Counties

```
Combdta2$DailyStartsPerVeh^-0.42 ~ Combdta2$PopDens +
    Combdta2$UnitDens + Combdta2$VehDens + Combdta2$VehPerCap +
    Combdta2$VehPerUnit)

Residuals:
      Min        1Q    Median        3Q       Max
-0.151365 -0.029695  0.000957  0.030304  0.168663

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.452109   0.005698  79.342  < 2e-16 ***
Combdta2$PopDens     5.267231   1.207131   4.363 1.35e-05 ***
Combdta2$UnitDens   -3.308847   1.305888  -2.534   0.0114 *
Combdta2$VehDens    -0.909502   2.254875  -0.403   0.6867
```

```
Combdta2$VehPerCap   0.020209   0.014254   1.418   0.1564    
Combdta2$VehPerUnit  0.028264   0.004155   6.803 1.36e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04578 on 1965 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.09917,	Adjusted R-squared:  0.09688 
F-statistic: 43.26 on 5 and 1965 DF,  p-value: < 2.2e-16

> vif(StartsM2)
   Combdta2$PopDens   Combdta2$UnitDens    Combdta2$VehDens  Combdta2$VehPerC
ap 
          23.157721           10.145722           12.523044            4.5405
70 
Combdta2$VehPerUnit 
           3.434811 
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage


Marginal Model Plots
Data
Model
Combdta2$DailyStartsPerV
0.35
0.55
0.00
0.01
0.02
0.03
Combdta2$PopDens
0.000
0.010
0.020
0.030
Combdta2$UnitDens
0.000
0.004
0.008
0.012
Combdta2$VehDens
0.2
0.4
0.6
0.8
1.0
Combdta2$VehPerCap
0.0
0.5
1.0
1.5
2.0
2.5
Combdta2$VehPerUnit
0.46
0.48
0.50
0.52
0.54
0.56
0.58
Fitted values


Added-Variable Plots
Combdta2$PopDens | others
Combdta2$UnitDens | others
Combdta2$VehDens | others
Combdta2$VehPerCap | others
Combdta2$VehPerUnit | others