ERG
www.erg.com

Final Version
4/30/2014

Memorandum

To: CRC A-88 Project Team
Brent Bailey, CRC
Betty Taylor, CRC

From: John Koupal, ERG
Timothy DeFries, ERG
Cindy Palacios, ERG
Heather Perez, ERG

Date: April 30, 2014

Re: Evaluation of Data Sources for Improving NEI Inputs:
CRC A-88 Task 1 Memo (Final)

1. Introduction

This memo serves as the final deliverable for Task 1 of CRC-A-88, "Evaluation of Data Sources for Improving NEI Inputs", presenting ERG's evaluation of potential on-road inputs to improve for Version 2 of the NEI, and work to proceed under Task 2. This final version reflect comments received from the CRC project panel and EPA is response to the March 30th draft memo. The memo presents 1) initial screening of a broad pool of on-road inputs, narrowing to a list of candidates for detailed evaluation, 2) results of this evaluation for final candidates, focusing on viable data sources and methods to update current defaults, and 3) work to proceed under Task 2, based on the CRC panel's comments on the draft memo.

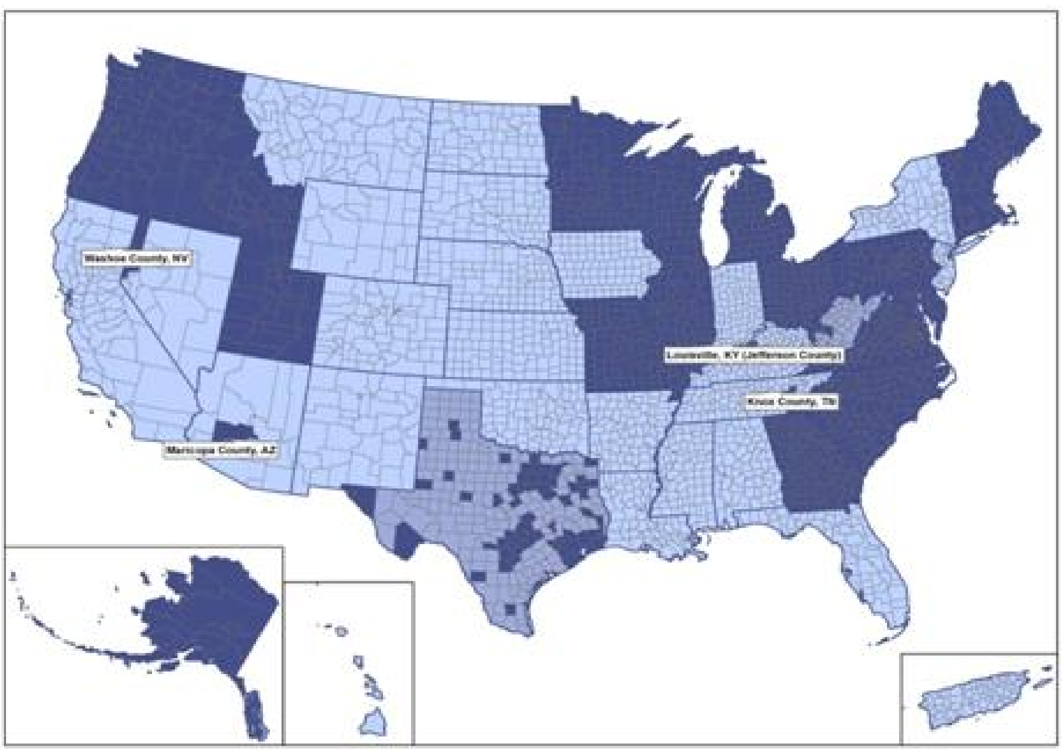
The focus of A-88 is to provide improved default on-road inputs at the local level for Version 2 of the NEI, and for EPA's broader inventory and air quality efforts. A corollary objective is to identify and evaluate data sources that could be used by modelers at the regional, state and municipal level to improve inventory and air quality modeling. Our conclusions from Task 1 are that promising national data sources exist to make meaningful improvements in the defaults for several on-road inputs.

2. Initial Screening of Candidate NEI Inputs

As the focus on A-88 is improving inputs for the NEI, the pool for potential inputs numbers in the dozens and covers MOVES itself, and the SMOKE-MOVES model used to actually generate the on-road NEI. This pool of dozens of inputs is much broader than can be addressed within the scope of A-88, so it is necessary to narrow down the list of inputs to address in Task 2. ERG's first step in this was an initial screening process that drew on what was learned in the A-84 project, stated priorities of CRC and EPA, initial investigation of data sources, and ERG's experience on other projects.

The pool of potential inputs to update was defined by a) those that states would potentially update through the MOVES county database; and b) inputs requested by CRC and/or EPA via the RFP or in communications early in the project. Some of the stated priorities of CRC and EPA broadened the focus of the evaluation beyond default MOVES inputs for non-submitting states (Figure 1), to MOVES and/or SMOKE inputs that are not requested of states as part of the NEI. The scope of these updates would therefore included all states, and carry into EPA's broader inventory and air quality work. The scope of the evaluation therefore considered improvements to national MOVES defaults that would otherwise be used in every county in the U.S., regardless of whether states submitted data for other inputs. Through communication with EPA, it is clear that in addition to Version 2 of the NEI, the improvements to be made under A-88 would also serve to improve the on-road inputs EPA's air quality modeling platform, and MOVES defaults.

Figure 1 – Version 1 NEI MOVES Submissions (states not submitting data shown in light blue)



Initial screening based on A-84 and stated CRC/EPA priorities are discussed in the following sections.

2.1. Most Influential Inputs from A-84

A-84 focused on the MOVES county database (CDB) inputs that states are requested to provide for the NEI, entered through an interface known as the county data manager (CDM). A central outcome of A-84 was an assessment of the most influential inputs on total MOVES emissions. This process began with a qualitative assessment of the influence that every MOVES CDM input would have on NEI emissions (Table 1).

Table 1: Qualitative Assessment of MOVES CDM Inputs From A-84

Table	General Description	Expected Influence on NEI
avft	Fuel technology fractions (gas/diesel/CNG)	Medium – large variations in state-submitted data not expected
avgspeeddistribution	Distribution of average speeds	High – state-submitted data likely, significant variation from default likely, MOVES results highly sensitive to changes
dayvmtfraction	VMT distribution by weekday/weekend	Medium – not significant factor in annual inventories
fuelformulation	List of possible fuels in area	Low – do not expect states to submit new formulations
fuelsupply	Market share of fuel formulations	Medium – do not expect states to have comprehensive information on fuel market share
hourvmtfraction	Distribution of VMT by hour of the day	Medium – not a significant factor in annual inventories
hpmsvtypeyear	Total Vehicle Miles Travelled (VMT) by vehicle class	High – state-submitted data likely, significant variation from default likely, MOVES results highly sensitive to changes
imcoverage	I/M program parameters	Medium - states not expected to have significant changes
monthvmtfraction	Distribution of VMT by month	Medium - while more important for annual inventories than day/hour fractions, large variations from default not expected
roadtypedistribution	Distribution of VMT across road types	High - state-submitted data likely, significant variation from default likely, MOVES results highly sensitive to changes
sourcetypeagedistribution	Fleet age distribution	High - state-submitted data likely, significant variation from default likely, MOVES results highly sensitive to changes
sourcetypeyear	Vehicle populations	High - state-submitted data likely, significant variation from default likely, MOVES results highly sensitive to changes
zonemonthhour	Meteorology	Medium – state-submitted data likely, but do not expect significant variations from defaults

In A-84, this qualitative assessment led to a focus on five inputs thought to have the highest influence on national, annual emissions: VMT, vehicle population, age distribution, average speed and road type distribution. These became the "primary" inputs studied extensively through A-84. The range of state-submitted data was analyzed and compared to MOVES defaults, and the sensitivity of MOVES emissions to variability in each assessed. From this a relative ranking of the influence of each input, by MOVES source type (vehicle class), was determined. This ranking, which varied by pollutant, is shown in Table 2.

Table 2. Most Influential Inputs by Source Type / Cluster From A-84

	HC	
Source Type/Cluster	**Input Varied**	**Increase in Total Emissions**
Passenger Car	Age Distribution	23.5%
Passenger Truck	Age Distribution	22.3%
Passenger Truck	Population Fraction	15.7%
Passenger Truck	VMT Fraction	13.9%
Passenger Car	Population Fraction	12.4%
	CO	
Source Type/Cluster	**Input Varied**	**Increase in Total Emissions**
Passenger Truck	VMT Fraction	29.8%
Passenger Truck	Age Distribution	24.9%
Passenger Car	VMT Fraction	21.8%
Passenger Car	Age Distribution	21.3%
Light Commercial Truck	Population Fraction	10.8%
	NOx	
Source Type/Cluster	**Input Varied**	**Increase in Total Emissions**
Combination Long Haul Truck	VMT Fraction	39.0%
Passenger Truck	VMT Fraction	21.5%
Combination Short Haul Truck	VMT Fraction	19.9%
Urban Unrestricted_Day	Average Speed Distribution	18.0%
Passenger Car	VMT Fraction	12.4%
	PM	
Source Type/Cluster	**Input Varied**	**Increase in Total Emissions**
Combination Long Haul Truck	VMT Fraction	78.6%
Combination Short Haul Truck	VMT Fraction	34.7%
Urban Unrestricted_Day	Average Speed Distribution	18.3%
Rural Unrestricted_Day	Average Speed Distribution	14.0%
Combination Long Haul Truck	Age Distribution	12.7%

Table 2 pinpoints the MOVES CDM inputs that will have the biggest effect on total emission inventories if they are improved, at the source type level. This is a logical starting point for prioritizing the inputs to focus on for A-88.

2.2. CRC & EPA Priorities

In addition to A-84, the evaluation of potential Task 2 improvements also reflect stated priorities from CRC and EPA. From CRC, the RFP provided an initial list of project priorities, which were further defined in an email sent by Brent Bailey at the outside of the project, which defined top priorities as "VMT mix …(trucks versus cars for example)", and "VMT by day of week". It also provided input on EPA priorities, excerpts as follows:

- *Items that help us estimate the quantity and spatial and temporal distribution of extended idling.*
- *Speed distribution item is of lesser importance. We suggest replacing the speed item with an analysis of populations using the FHWA data. There is a lot of variability state to state in the ratio of VMT to vehicle population. This ratio affects starts and evaporative emissions.*

To clarify EPA's priorities and plans for NEI updates, ERG coordinated a conference call with EPA/OTAQ and EPA/OAQPS modeling staff on February 6th. During this call, EPA clarified that an analysis of population using the FHWA data was not necessary if direct population data (e.g. from state registration data) could be obtained. EPA also confirmed that improving start activity and the allocation of heavy-duty extended idle were top priorities, as was improving temporal VMT distributions. Although the latter was judged in A-84 not to have much influence on annual emissions estimated in the NEI, EPA's interest in improved temporal VMT distribution is for episodic modeling via the EPA air quality modeling platform.

2.3. Initial Screening

ERG did an initial screening of the most influential inputs from A-84 and priorities expressed by CRC and EPA to determine the feasibility for improving these inputs within the scope of A-88. This process began in proposal phase, and was continued in the effort to narrow down inputs that showed the most promise for carrying forward to Task 2. The initial screening was focused on the question of whether national datasets existed that could potentially improve local defaults, and whether they could be obtained within the timing and cost of the contract.

Of all of the inputs considered in this initial screening, only speed distribution did not meet the criteria, primarily because of cost. "Telematics" datasets are being compiled by a number of commercial entities including cellular providers and commercial GPS vendors, and are available for purchase, if not directly than aggregated in such a way that no personal information can be retrieved. For example, on behalf of EPA, ERG worked with TomTom to provide speed distribution data averaged across the entire U.S. to update the MOVES national defaults. Link-specific speed data is also compiled by INRIX, Inc. who provides real-time traffic data to mapping programs from telematics and other data sources. From this, we believe that telematics data will be an excellent source of data to improve local activity inputs for MOVES, and cost effective if pursued for individual areas. For example, ERG has had initial discussions with Verizon Telematics, a company that compiles these data on behalf of insurance companies; these discussions have confirmed that data on speed distribution and a number of other trip-related activities are available by location across the U.S. For heavy trucks, telematics systems are increasingly used for fleet management, and can provide speed distribution data for long-haul trucks.[1] However, based on previous ERG work with telematics data, the projected cost of location-specific data for tens or potentially hundreds of specific locations around the U.S. would be well beyond the scope of this contract. This, combined with EPA input that speed distribution was of low priority, led us to remove speed distribution from further consideration in A-88

2.4. Inputs Chosen for Detailed Evaluation

Based the initial assessments described in Section 2.1-2.3, ERG settled on the following inputs to move ahead for detailed evaluation:

- Passenger car & truck age distribution
- Passenger car & truck population
- Combination long-haul VMT
- Temporal VMT distribution (e.g. allocation of VMT by day & hour)
- Heavy-duty extended idling
- Start activity

A detailed evaluation was then performed for each to identify viable datasets and an approach to update current defaults. These evaluations, along with a recommended approach for improving each, are presented in Section 3.

3. Detailed Evaluation of Candidate Inputs

The purpose of the detailed evaluation was to develop initial recommendations as to which inputs to proceed with updating under Task 2. The evaluation of each input involved assessing the data source of current defaults, identifying candidate data sources that could update the defaults, and assessing potential approaches to improving current

[1] Boriboonsomsin, et. al "Generating Heavy-Duty Truck Activity Data Inputs for MOVES based on Large-Scale Truck Telematics Data", TRB Paper 12-3528, January 2013

defaults. Per direction from the CRC project team, the focus in identifying data sources was on datasets maintained at a national level, e.g. by federal agencies or companies focused on national data compilation, rather than states. The data sources should cover the entire U.S. while providing location-specific data for MOVES that would be an improvement to the current defaults. ERG identified potential datasets for all inputs except temporal VMT distribution; as discussed below, the best available "national" database for this only covers a subset of states.

A discussion of candidate datasets and thoughts on how they could be applied to MOVES (or SMOKE-MOVES) is detailed for each input in the subsections below. For each input, we review the relevant inputs in MOVES and/or SMOKE, the source of current default inputs, the data source identified to improve default inputs, and recommended approach to updating default inputs. CRC project panel comments on the initial recommendations in the draft memo provided the final direction on which of these inputs to pursue under Task 2. The comments are addressed with a discussion of whether, and how, each input will be pursued in Task 2.

3.1. Passenger Car & Truck Age Distribution

Relevant NEI input: Age distribution is a county-specific input in the MOVES CDB, in the sourceTypeAgeDistribution table. Age distribution is provided for each of the MOVES source types. Age distribution is one of the factors used to determine representing counties in the SMOKE-MOVES framework used for the NEI and EPA's air quality modeling platform.

Source of current default data: Version 1 of the NEI used a combination of state-submitted age distribution data and national defaults developed from a national compilation of state registration databases, purchased from R.L. Polk (now IHS Automotive). As detailed in A-84, age distribution was submitted for ~1,400 counties for Version 1 NEI. Representing counties were established for Version 1 based in part on the state-submitted age distribution data.

Data source identified for improvement: ERG evaluated purchasing data from the database of vehicle registrations compiled by IHS Automotive. The source of this database is state vehicle registration databases, which is consistent with the approach in EPA MOVES guidance and the majority of states that compiled and submitted data in Version 1. The primary IHS database is focused on light vehicles, and ERG focused on age distribution for the passenger car, passenger truck and light commercial truck source types (source types 21, 31 and 32). ERG spoke with an IHS representative who confirmed that their database does provide coverage for all U.S. states, and is compiled each year as of July 1. The IHS database contains vehicle population to the county, model year, and vehicle make level. However, due to IHS data restrictions, ERG will need to purchase data from IHS and calculate age distribution fractions, which can then be distributed publicly as "derived" data. These derived age fractions can be used directly for age distribution inputs for passenger cars, passenger trucks and light commercial trucks in MOVES. The derived data would be fractions of total population in each model year from 1982 through 2012, as of July 1 2011, for each county purchased.

The draft memo discussed the options of either purchasing data for the 47 representing counties in the states that use MOVES but did not submit MOVES input data or completed emissions, or all 3,222 U.S. counties. Though the latter option was not included in ERG's cost proposal bid, this level of would provide a rich dataset covering all counties (representing and non-representing), and enable an improved assessment of SMOKE-MOVES representing counties. Comments received from the CRC panel on the draft memo overwhelmingly favored the latter option. ERG is therefore planning to proceed with purchase of data for all counties, and based on input from the CRC panel, conduct the following analyses with it:

- Derive age distribution for each county (derived data that can be shared publicly);
- Perform cluster analysis of age distributions within each state to determine if separate distributions are warranted (e.g. urban/rural). This analysis will be useful for re-analysis of representing counties in the SMOKE-MOVES framework.

ERG's proposed effort for the project will be adjusted accordingly by dropping work on other inputs, as discussed in later sections.

Approach to updating NEI inputs: Purchasing and deriving age distribution data from IHS for passenger cars and trucks is a straightforward way to improve the default age distribution data. County-specific distributions will be a significant improvement over the national default age distributions currently planned for the NEI, and based on the results of A-84 would have a meaningful effect on emissions. The source of the IHS data, state registration databases, is comparable to that used by states for NEI submissions and local SIP/Conformity modeling.

The resulting population numbers by county can be fed directly into MOVES CDB for the passenger cars (source type 21). We would propose to use the light truck data directly for passenger truck (31) and light commercial truck (32) source types. This reflects some simplification, because (according to VIUS) passenger trucks includes some heavier weight classes not included in the IHS database. However, well over 90 percent of passenger trucks and light commercial trucks would be covered by the IHS light truck database, and the improvement of using actual local registration data in the place of national defaults would justify this simplified approach.

Future updates to these data would be straightforward as well, assuming IHS continues to offer the registration database and analysis as a product. As this would require an ongoing financial commitment on the part of EPA and/or other agencies supporting air quality work, an alternative to purchasing the data would be to work with individual states to obtain vehicle registration data, and compile data from multiple states. This can be problematic, however, due to turnaround time and dealing with format or vehicle classification inconsistencies state-to-state. The cost of purchasing the IHS data is relatively modest, and would be a good ongoing investment for the emission inventory and air quality community to make.

3.2. Passenger Car & Truck Age Populations

Relevant NEI input: Although vehicle population is an input to the MOVES CDB (in the sourceTypeYear table), for the NEI process it is a direct input to SMOKE, as "VPOP". VPOP is derived from the MOVES CDB inputs, for consistency, based on default mapping between MOVES source types and SCC vehicle classes. If updated, ERG would need guidance from EPA about how population data inputs should be provided – as MOVES CDB, SMOKE VPOP or both.

Source of current default data: Although the default MOVES database has national source type populations that are allocated to the county level, these are projected from 1999 totals and weren't used for Version 1 NEI VPOPs. Instead, updated default VPOP estimates were developed by multiplying MOVES national default estimates of miles per vehicle (by source type) to county-level Version 1 VMT estimates derived from HPMS, and mapping to SCC class.

Candidate data source identified: The IHS database discussed for age distribution can also provide county-level passenger car & light truck vehicle populations. It is the same data that would be used for age distribution, simply totaled by county separately for passenger cars and light trucks, by fuel type. Per CRC panel comments, ERG will derive the county-level population data for all U.S. counties from the data purchased from IHS.

Approach to updating NEI inputs: The IHS car population numbers by county can be fed directly into MOVES CDB and/or VPOP estimates for passenger cars (source type 21). Their truck data would require some additional work to be used for passenger truck (31) and light commercial truck (32) source types. The IHS truck data would need to be split between the two source types; a default estimate for this split would likely be required. According to the Vehicle In-Use Survey (VIUS), the data source behind many MOVES defaults for trucks, "passenger trucks" include some heavier weight classes not included in the IHS database. This is reflected in the MOVES passenger truck makeup.

Because of this, some adjustment would be required to the IHA data to account for populations of heavier passenger vehicles. In essence the IHA data would update only the "light truck" portion of the passenger truck category, which is the vast majority of these trucks. The mapping between IHA data, MOVES sources types and the SCC-based VPOP estimates could also be informed by the fuel type splits (gas/diesel) at the county level.

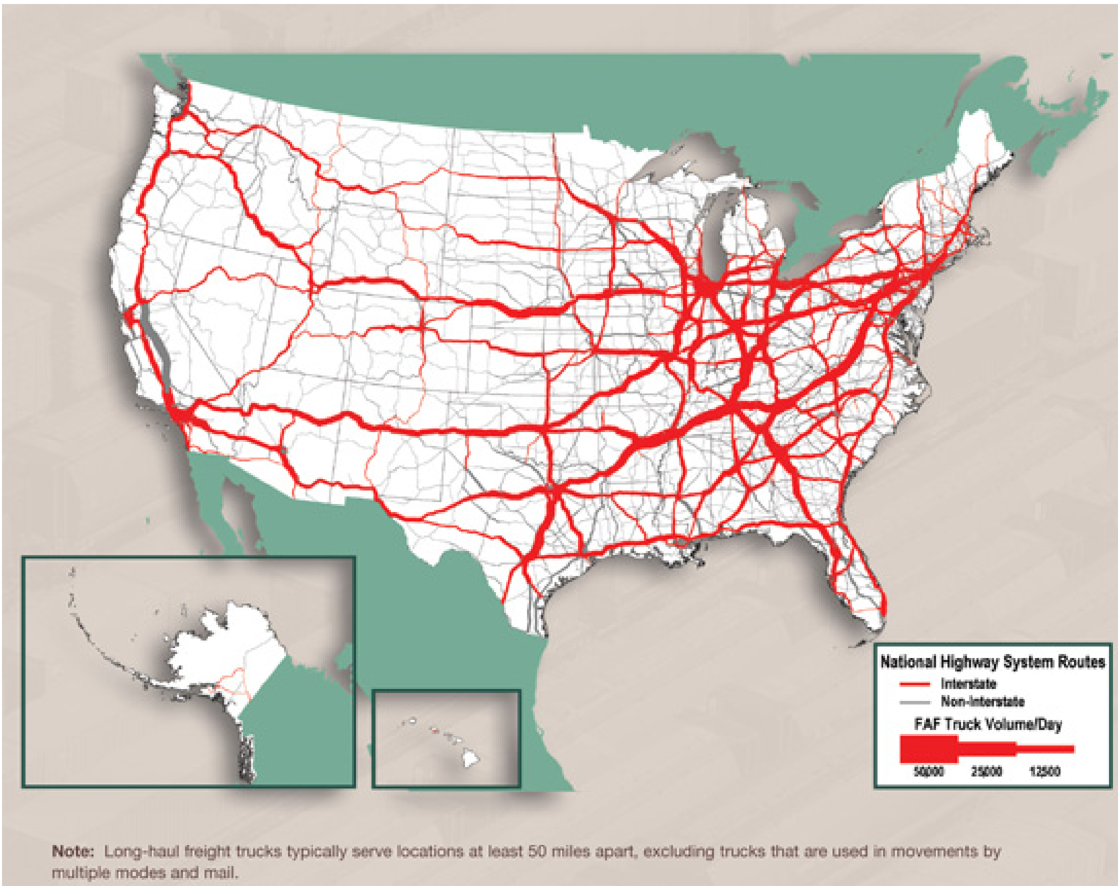
3.3. Long-Haul Truck VMT

Relevant NEI inputs: Annual VMT is a direct input to the MOVES CDB though the HPMSVtypeYear table. This is at the source type level, so combination short-haul and long-haul truck (source types 61 and 62) VMT are direct inputs to MOVES. VMT is also a direct input into SMOKE-MOVES, so for the NEI would be a focus of this update. The same data can be used to populate the inputs for both models, though some mapping between Highway Performance Monitoring System (HPMS), MOVES Source Type and SCC classification schemes is required; for example, while MOVES has long-haul VMT split out, the closest corresponding SCC class is Class 7 & 8 heavy-duty diesel trucks. These align well but not perfectly, and the mapping between HPMS data, MOVES source types and SCCs is an intricate process undertaken by EPA based on national default estimates. This process needs to be considered in how updates are made to long-haul truck VMT.

Source of Current Data: Although the default MOVES database has national long-haul truck VMT that are allocated to the county level, these are projected from 1999 national totals and weren't used for Version 1 NEI. Instead, updated estimates were developed directly from the 2011 HPMS universal dataset. Because the HPMS universal dataset is the main repository of local VMT estimates across the nation, this is already the best national source of local-specific VMT data available; however, HPMS only reports estimate of totals in the single unit truck and combination truck classifications, not broken down by short-haul vs. long-haul, or fuel type. To estimate these breakdowns, national MOVES defaults are applied uniformly across the entire U.S. A key opportunity for improving NEI default VMT estimates is therefore in the allocations of HPMS VMT, particularly for long-haul trucks.

Candidate data source identified: ERG has focused on the Freight Analysis Framework (FAF), a compendium of data and analyses maintained by FHWA (http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/), as a candidate source for updating long-haul truck VMT. The FAF integrates information from a number of transportation and commodity flow data sources to allow a more detailed assessment of travel and population related to goods movement in each county of the U.S. Included in the FAF are data on average annual daily traffic (AADT) for long haul trucks reported for specific segments of interstates, highways and major roads across the U.S. (Figure 2).

Figure 2: FAF Average Daily Long-Haul Freight Traffic on the National Highway System (2007)
Source: FHWA



T

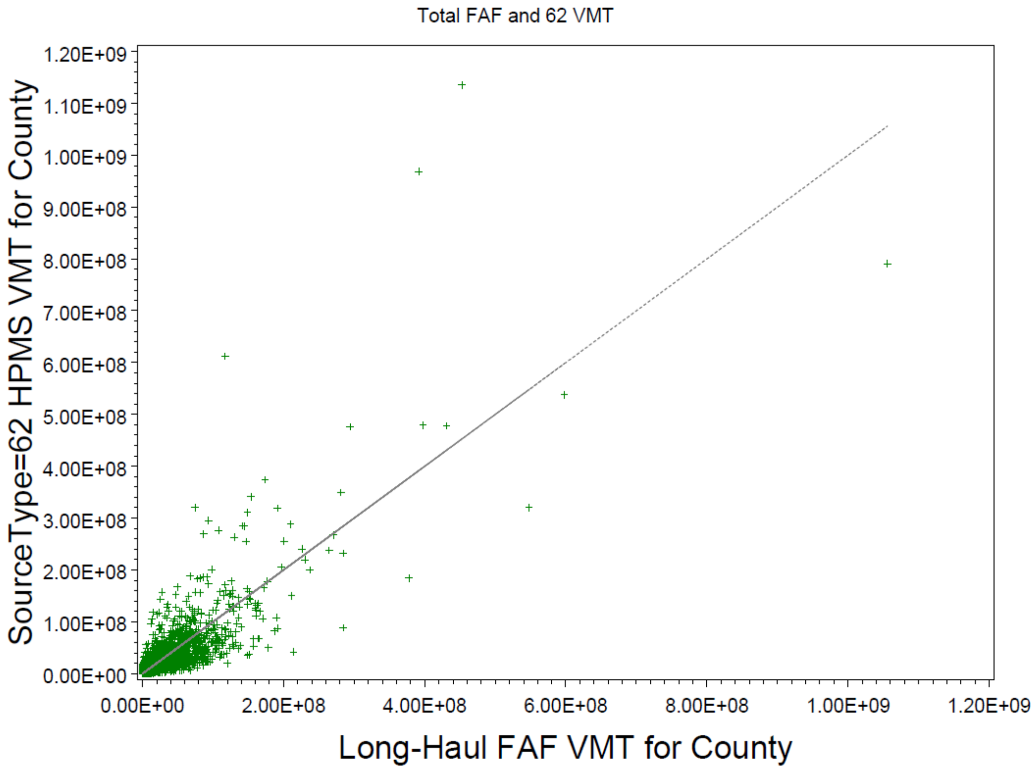
Under Task 1, ERG evaluated the FAF and confirmed that it is a viable data source for updating local default long-haul truck VMT estimates (and by association, short-haul truck VMT). It appears to be the only database focused on quantifying long-haul truck travel by specific geographic location at a national level. Telematics datasets are coming online that could provide such information, but as of yet do not appear to provide the needed coverage for a geographic national scale.

Although the FAF datasets holds promise for improving long-haul truck VMT estimates, there are several caveats with the dataset that need to be considered in determining how to apply the data for location-specific emissions modeling. The primary caveat is that FAF long-haul truck volumes are not based on direct measurement, but are modeled from a series of estimates of network capacity, commodity flow and truck trip lengths. Some of these inputs, especially truck trip lengths (from the 2002 VIUS survey) are dated. The FAF estimates of total truck volumes are validated against measured vehicle counts in aggregate, though the FAF documentation itself cautions the user on the accuracy of the estimates at a fine level of detail, e.g. on any given network link. Additional caveats are that the FAF only covers a subset of interstates and major roadways, and the FAF is only updated in accordance with the five-year cycle of the U.S. Economic Census. The most recent published year is 2007, so the last year of complete FAF data was 2007. Subsequent years are projected forward from the economic census data. For these reasons, ERG plans to use the FAF estimates to inform *relative* contribution of long-haul trucks to VMT, rather than using direct estimates of FAF long-haul VMT to update the 2011 estimates.

ERG evaluated FAF long-haul VMT estimates against different permutations of the national default VMT used in Version 1. This is not a perfect comparison, because FAF data reflects 2007, while Version 1 is for 2011 (nationally, combination truck VMT was about 12 percent lower in 2011 vs. 2007); and, FAF is only for freeways and major arterials. These initial assessments were intended just to gauge how the FAF numbers compare to HPMS estimates in a "big picture" fashion. As a first check, Figure 3 shows a comparison of 2007 FAF long-haul truck VMT (x-axis) vs. NEI default 2011 VMT, split into the MOVES combination long-haul source type (y axis), which travel mostly on the major roads covered in the FAF. Each point represents a single county, with a 1:1 line superimposed to show the relative scatter of the data. This chart shows a good scatter around the 1:1 line; the national totals of FAF long-haul and MOVES combination long-haul are actually within one percent of one another. This plot highlights the high degree of difference between FAF and default on a county-specific basis; many counties show a relative difference of a factor of five or more vs. the default. Based on the A-84 results, this level of variability in long-haul VMT would have dramatic effects on total PM and NOx emissions.

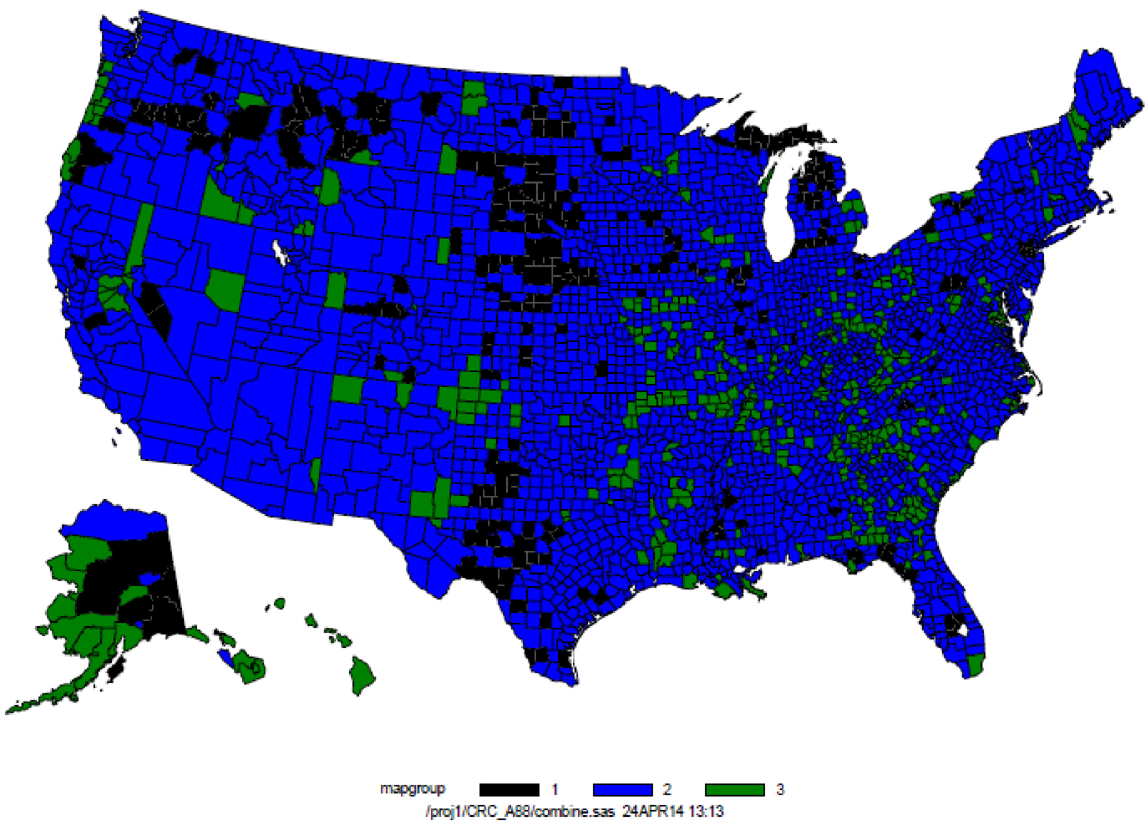
1785 West Stadium Boulevard, Suite 202, Ann Arbor, MI 48103 • Phone: 508 868-0786 • Fax: 734 994-5709
Ann Arbor, MI • Arlington, VA • Atlanta, GA • Austin, TX • Boston, MA • Chantilly, VA • Lexington, MA
Nashua, NH • Prairie Village, KS • Research Triangle Park, NC • Sacramento, CA

Figure 3: FAF Long-Haul VMT vs. Default 62 VMT by County



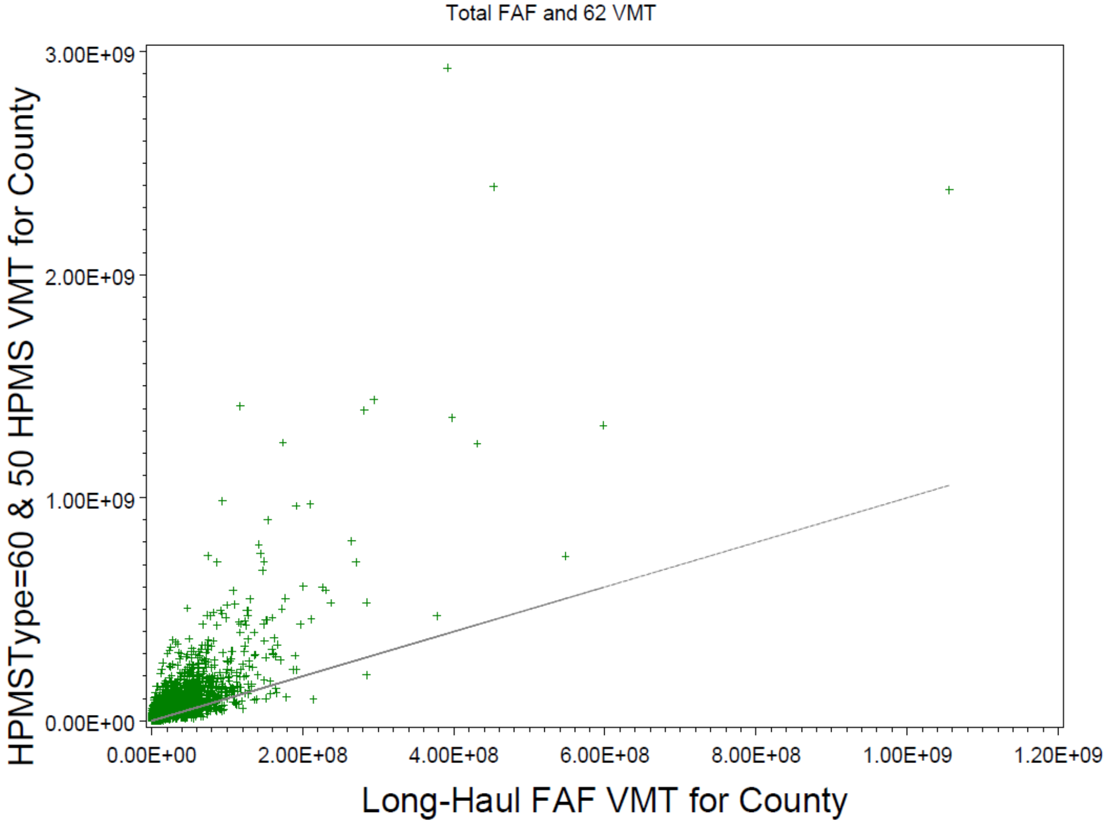
The CRC panel comments requested a mapping of FAF vs. default data to understand geographic trends. To address this request, the same data are also represented in Figure 4 below. This map shows very broad groupings of "FAF > 3x default" (black), "0.33x default < FAF < 3x default" (blue), and "FAF < 0.33x default" (green). This shows where the extreme values from Figure 3 are located. The map shows that the more extreme differences between FAF and default long-haul VMT (black and green counties) are spread across the country, and seem to be in rural areas as opposed to major cities.

Figure 4: FAF Long-Haul VMT vs. Default 62 VMT by County – Extreme Differences (Black = extreme high, Green = extreme low, Blue = not extreme)



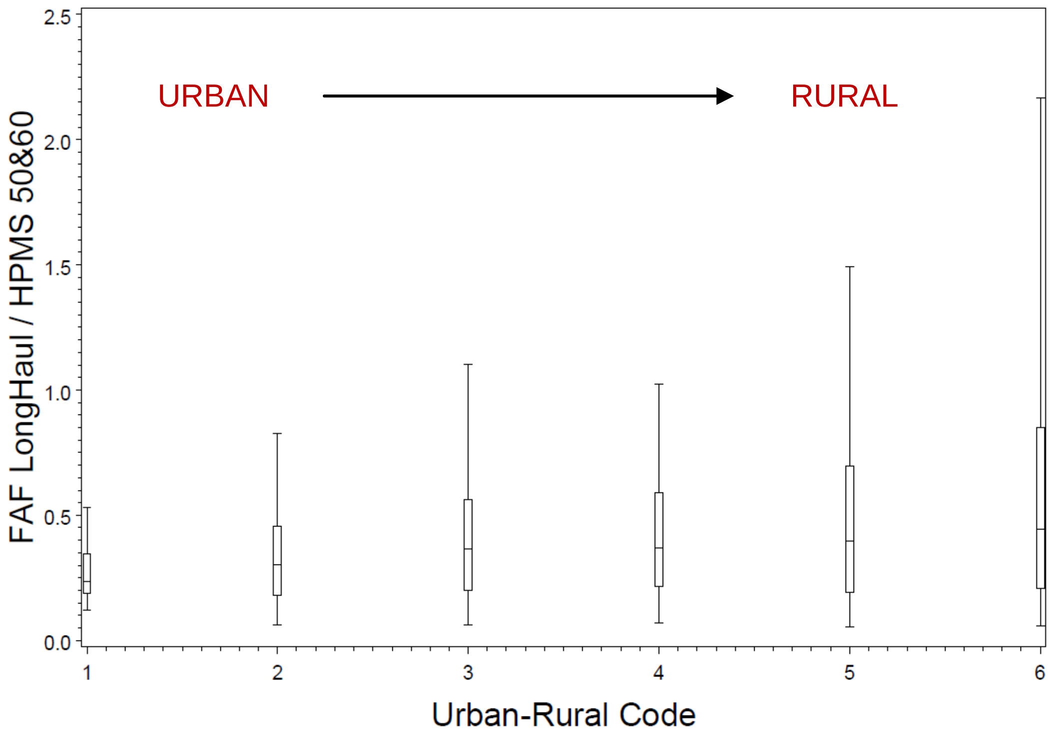
Given the wide spread of FAF estimates compared to defaults, a further reasonableness check was performed against more direct HPMS estimates. While HPMS does not estimate long-haul trucks directly, it does provide direct estimates for single unit and combination trucks. Although the exact percentage is uncertain, all trucks considered "long haul" trucks should fit within these HPMS single unit or combination categories (According to both FAF and MOVES default estimates, about 10 percent of long haul trucks are single unit). We would therefore expect FAF long-haul VMT to be less than HPMS single unit + combination truck VMT, especially since FAF is only for a subset of roadways. This comparison by county is shown in Figure 5, with a 1:1 line superimposed. In the majority of counties, FAF long-haul VMT is lower than HPMS single unit + combination, but does appear higher for some counties.

Figure 5: FAF Long-Haul VMT vs. Total HPMS Single & Combination Truck VMT by County



ERG investigated reasons behind the variability in FAF long-haul VMT vs. total single unit+combination VMT, particularly for counties which showed higher FAF long-haul estimates than the HPMS superset. ERG binned each county according to a 6-level urban/rural classification code used by the CDC National Center for Health Statistics (1=large central metro. i.e. most urban; 6 = non-core, i.e. most rural) and looked at the spread in the ratio of FAF long-haul : HPMS single + combination VMT in each. This is shown in Figure 6, with box plots by each urban/rural code (whiskers are $5^{th}/95^{th}$ percentile). This shows that the ratio of long-haul truck VMT to HPMS single+combination truck VMT is higher in more rural counties, with the values greater than 1 (i.e. below the 1:1 line in Figure 5) generally falling in the most rural counties. This approach may provide insight into ways to aggregate county-level FAF estimates to smooth out the noise of individual county extremes.

Figure 6: Spread of FAF Long-Haul : HPMS Single + Comb. VMT Ratio by Urban/Rural Level



Approach to updating NEI inputs: Given the overall dearth of information on long-haul vs. short-haul truck VMT mix, the FAF provides a viable national dataset to improve this information on a location-specific basis. The CRC project panel approved of pursuing updates to long-haul VMT using FAF, so it will be included as part of Task 2. Because of the caveats listed above, however, it is not recommended that FAF long-haul VMT estimates be used directly to replace national defaults on a county-by-county basis. Rather, an approach which uses FAF to better allocate long-haul VMT within the broader HPMS truck VMT categories is recommended. This approach acknowledges that the county-level HPMS estimates are the “gold standard”, but that the FAF can be used to better allocate VMT to long-haul vs. short-haul trucks in specific areas. This will take advantage of the variability in long-haul VMT fraction without making the wholesale switch to absolute FAF estimates.

ERG’s planned approach under Task 2 will be to use the relative contribution of long haul VMT from the 2007 FAF to re-allocate the 2011 HPMS-based VMT. This will preserve overall 2011 VMT estimates, while updating long-haul VMT based on FAF.

Because there is some concern about the accuracy of FAF data at the individual county level, this approach will investigate using FAF long-haul data aggregated to regional or state level, in order to avoid errors introduced by individual county extremes. As the FAF developers caution against using FAF estimates at too fine a level of detail, it may turn out the they are best applied higher than county level.

3.4. Heavy-Duty Truck Idling

For Version 1 of the NEI, extended heavy-duty truck idling was estimated based on default activity estimates and allocations to counties and grid cells; local data were not requested or submitted. For this project, CRC and EPA have indicated a desire for improved information on the quantity and spatial/temporal allocation these emissions, which would serve not just the NEI but inventory and air quality modeling by EPA, regional and local modelers.

Extended idling is limited to combination long-haul trucks in MOVES, and total idle activity (ExtendedIdlingHours) is a function of national default ratios of hours of extended idle per hours in operation, informed by DOT regulations concerning driver rest. National datasets that can establish location-specific inputs for total idle activity do not yet exist. Heavy-duty telematics may provide these data in time, but coverage is currently not broad enough. ERG's evaluation of this for A-88 was therefore on improving the spatial distribution of idling activity and emissions, through better allocations of total activity to the county and grid cell level. CRC project panel comments did not want to pursue this work under Task 2, however, citing overlap with similar efforts being undertaken by EPA and states. The discussion below includes the initial evaluation and proposals for Task 1, although this work will not be continued. ERG will deliver a cleaned-up version of the truck/rest stop database purchased as part of the Task 1 evaluation, but will not pursue the analysis further at this time.

Relevant inputs / source of current data: In MOVES, total heavy-duty truck idle activity is calculated at a national level as a function of long-haul truck VMT, then assigned to each county using allocation factors derived from county-level VMT estimates. For the NEI, SMOKE-MOVES further allocates activity from county to grid cell level using human population. The detailed spatial allocation of these emissions is therefore not based on specific locations of truck stops or rest stops where idling will occur.

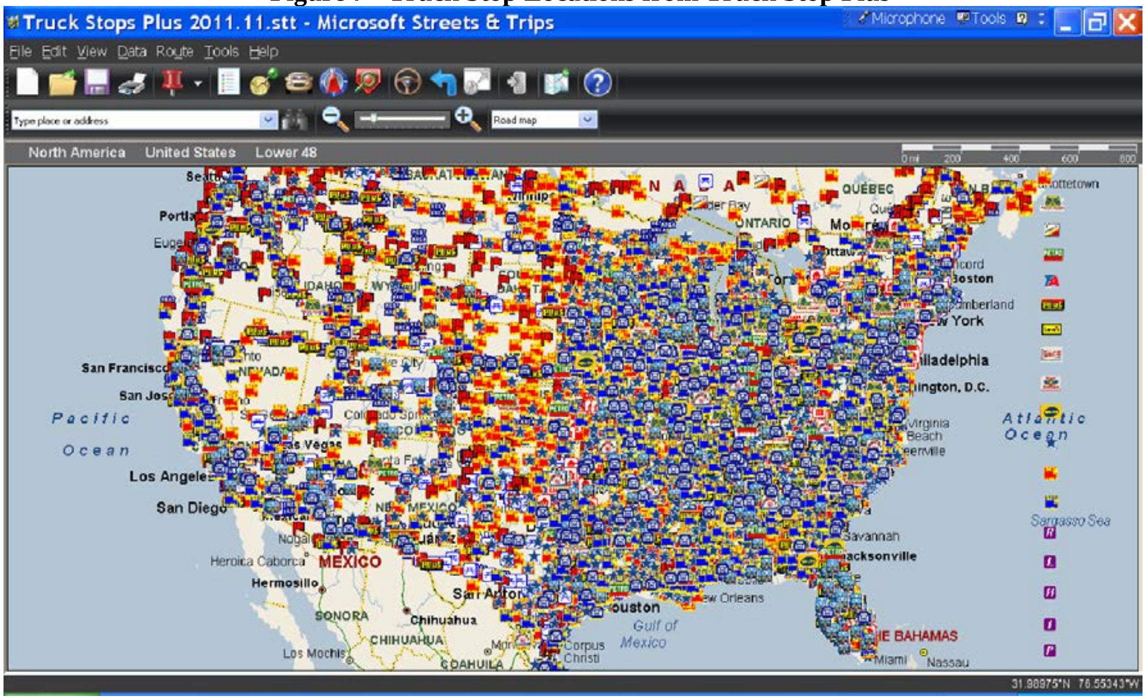
Approach to updating NEI inputs: ERG proposed to generate a consistent set of idle allocation factors for MOVES at the county level, and for SMOKE at the grid-cell level, using a GIS-based analysis of truck idling demand and supply. The simplest approach to the supply side is to develop allocations factors based on geocoded locations of truck and rest stops. This by itself would be a significant improvement over the current approach, which doesn't account for truck/rest stop location. As this is similar to work undertaken by EPA and states, it was deemed low priority for A-88 by the CRC project panel. For future consideration, a further step would be to assess not just the location of truck and rest stops, but their available supply of idling vis a vis number of parking spaces, restrictions or charges on overnight parking, etc. Idling demand can then be overlayed on supply to estimate how many spaces will be filled with idling trucks in any given location. Demand is proposed to come from the updated long-haul VMT discussed in the previous section. Further refinement could account for meteorology (e.g. when demand for heat or air conditioning is higher), and trucker preferences for given idle spots all other things being equal.

Candidate data source identified: A key element of this approach is finding a national database of truck and rest stops that can be put into a GIS framework to estimate idling supply. Such a database would ideally also have detailed data on attributes of each (number of spaces, etc.) to further refine the supply estimate. Fortunately there are many candidate databases with this level of detail, developed to provide truckers with location-specific and real-time information on rest areas, fuel prices, facilities etc. These databases have a strong online and mobile app component to better serve the trucker population, making incorporation of national databases into GIS software very feasible.

ERG evaluated four such databases for use in the analysis: 1) Truck Stops Plus (truckstopsplus.com), 2) Truck Stop Report (truckstopreport.com), 3) Truck Stop Pro (truckstoppro.com), and 4) POI Megafile (msstreets.com/2011/04/07/poi-megafile-truck-stop-guide-for-st/) . For this evaluation, ERG has concluded that Truck Stops Plus provides the most comprehensive and detailed database for the analysis, at a nominal cost (under fifty dollars). The Truck Stops Plus add-on template for Streets and Trips includes 7,347 trucks stops and includes all major chain truck stops, as well as independent truck stops (Figure 7). This dataset includes the size category of each truck stop (less than 20 parking spaces, 20-70 parking spaces, and more than 70 parking spaces), and details such as whether or parking is allowed, whether there is a charge for overnight parking, and what services and facilities (e.g. showers) are offered.

ERG purchased the Truck Stop Plus database as part of Task 1 evaluation, and performed some cleanup of the database to enable further analysis. Although this work will not be continued in Task 2, CRC has requested that ERG deliver this cleaned-up database to aid in QA of related work that states and EPA are undertaking.

Figure 7 - Truck Stop Locations from Truck Stop Plus



3.5. Start Activity

For Version 1 of the NEI, start exhaust emissions were also estimated based on default activity estimates; local data were not requested or submitted. For this project, CRC and EPA initially indicated a desire for improved information start activity at the local level; however, subsequent comments on the draft memo did not support pursuing this work for Task 2.

Relevant NEI inputs: Two primary activity inputs dictate vehicle start emissions: daily starts per vehicle, and the distribution of soak times (time between the end of one trip and beginning of the next). These are explicit inputs in MOVES. Though not part of the MOVES CDBs, county level data on starts/vehicle and soak distribution can be entered into the model with custom input database consisting of a startsPerVehicle table, and opModeDistribution tables. Starts/vehicle is input at the hourly level, which provides the default temporal allocation of starts as well. For SMOKE-MOVES, start activity is embedded in start emissions of "rate per vehicle" (RPV), and updates with local data would be reflected in these rates as well. Starts per vehicle is used to estimate total number of starts in MOVES; MOVES2014 will accept starts directly through the CDM.

Source of current defaults: Start activity data is currently calculated from individual vehicle trips in the sampleVehicleTrips table. This table is constructed from instrumented vehicle studies conducted in several cities around the U.S., and provides key-on and key-off times for each vehicle trip, on an individual vehicle basis. This provides a consistent activity basis for start and park activity needed for start as well as evaporative emissions processes. National default estimates of starts/vehicle and soak distribution are aggregated from this collection of driver surveys.

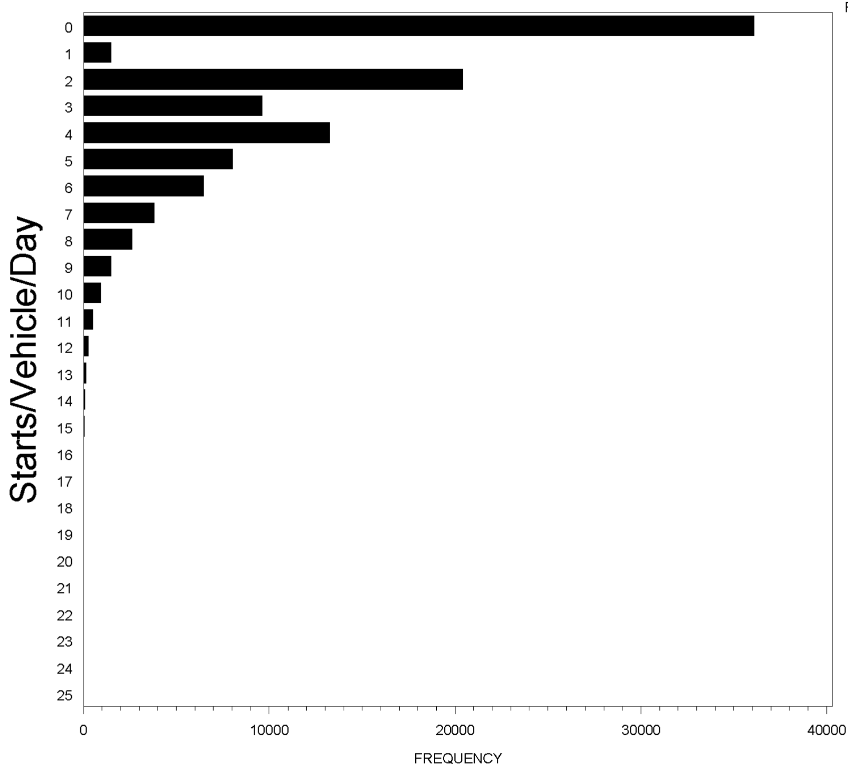
Candidate data source identified: Telematics data also holds promise for improving local estimates of start activity in MOVES. As with speed distribution, telematics may be an excellent source of data for local modelers, but ERG's judgment is that coverage and cost preclude it from being used to develop local defaults across the nation under A-88. ERG therefore focused on the most viable national source for trip activity data, the National Household Travel Survey (NHTS). The NHTS is conducted every several years by FHWA (update cycles in the past few decades have ranged from 5-8 years), and compiles extensive data on vehicle trip activity for individual households (e.g. trips per household, trip length, trip time of day) by location across the U.S. based on surveys. While this focuses on the travel of passenger vehicles, these vehicles have been shown to produce the majority of overall start emissions in inventory. An investigation of using the NHTS for updating MOVES soak distribution has already appeared in the literature.[2] As part of Task 1, ERG evaluated the most recent NHTS trip database (2009) with a focus on whether there are detectable variations in start activity in specific locations that could translate into location-specific defaults. In general the results were a) there are variations, though not large; and b) the NHTS results are quite different from MOVES national defaults for starts/vehicle. However, subsequent analysis and comments from the CRC panel called into question how well the NHTS trip survey data represent actual key-on events (what MOVES counts as "starts"), leading to the decision not to pursue using NHTS to update MOVES start activity as part of Task 2.

ERG's evaluation of the NHTS trip data focused on calculating starts activities in the form of MOVES start activity, with a focus on determining if starts behavior varied substantially by some geographical measure. This focused on passenger cars (SourceID=21) on weekdays (DayID=5). Default MOVES estimates show a total number of starts per vehicle over the 5 weekdays of 29.4 (5.88 starts per day). To compare to this, we used the NHTS data to calculate the distribution of trips for cars that completed the NHTS survey on weekdays. ERG calculated the trip distribution 1) so that it can be compared with the MOVES internal table and 2) so that we can investigate whether geographical household location can substantially affect the trip distribution. How the fraction of vehicles that are not started for one or more days is handled by

[2] Zhang et. al, "Estimating and Modeling Soak Time Distributions with the 2009 National Household Travel Survey Data", TRB Paper 12-4562, January 2012

MOVES and is prepared for analysis using the NHTS data affects the comparison between the start distributions for the MOVES default and the NHTS data. Figure 7 shows the distribution of trips per vehicle per day for the 105,322 cars that completed the NHTS survey on a weekday. The average number of trips/vehicle/day for this dataset is 2.73, compared to 5.88 for MOVES. However, there are questions about whether trip surveys underreport the actual number of key-on starts represented by MOVES, and whether NHTS overestimates the number of zero-trip vehicles because it focuses on a single day of household activity. On the latter point, Figure 8 shows that 34 percent of the vehicles in the NHTS dataset had zero trips, which has a major effect on the trip/day averages presented here. How representative this prevalence of zero-trip vehicles are in the entire population is a significant question that would require resolution before applying NHTS data in MOVES.

Figure 8 - Distribution of Trips/Vehicle/Day Across All Households in NHTS



Starts/Vehicle/Day	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
0	36105	36105	34.28	34.28
1	1472	37577	1.40	35.68
2	20412	57989	19.38	55.06
3	9626	67615	9.14	64.20
4	13249	80864	12.58	76.78
5	8030	88894	7.62	84.40
6	6457	95351	6.13	90.53
7	3805	99156	3.61	94.15
8	2611	101767	2.48	96.62
9	1489	103256	1.41	98.04
10	946	104202	0.90	98.94
11	504	104706	0.48	99.42
12	273	104979	0.26	99.67
13	139	105118	0.13	99.81
14	82	105200	0.08	99.88
15	44	105244	0.04	99.93
16	26	105270	0.02	99.95
17	19	105289	0.02	99.97
18	17	105306	0.02	99.98
19	4	105310	0.00	99.99
20	4	105314	0.00	99.99
21	2	105316	0.00	99.99
22	3	105319	0.00	100.00
23	2	105321	0.00	100.00
24	1	105322	0.00	100.00
25	0	105322	0.00	100.00

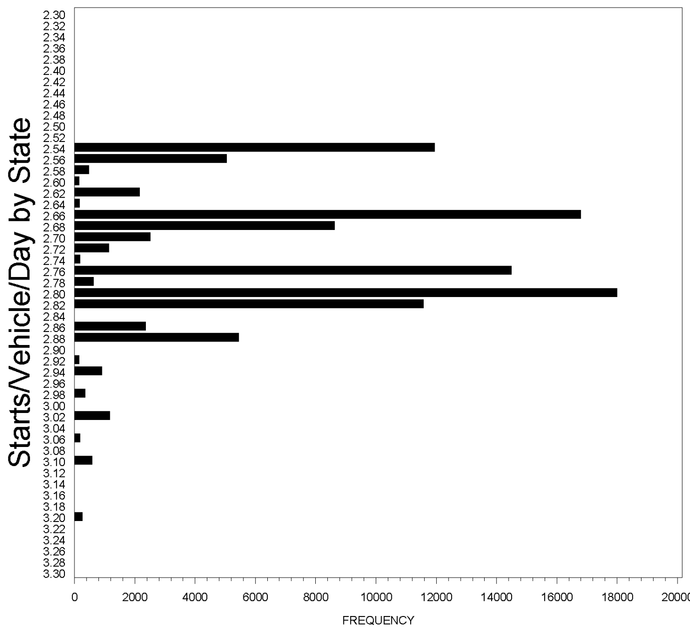
/proj1/CRC_A88/NHTS/makebins.sas 28MAR14 11:09

Although there are questions about how well NHTS could be used to estimate absolute start/day estimates in MOVES, it does provided a fertile dataset for exploring geographic differences in trip activity. As part of Task 1, ERG analyzed how trips/vehicle vary by geographic region, analyzing trips/vehicle by broad census divisions in the country (Table 3), where trips/vehicle ranged from 2.63-2.87; by state (Figures 9), with a range of 2.54-3.21, and by metropolitan area ("core-based statistical area", or CBSA), Figure 10, with a range of 2.44-3.26. We would expect from a purely statistical perspective that the distribution of average trips/vehicle/day would become wider as the number of levels of a geographical factor increases. The progression of distribution widths from Table 3 through Figures 9 and 10 demonstrates this expected trend. In the limit, where each individual vehicle is assigned its own geographical location, the trips/vehicle/day distribution would just be the distribution show in Figure 8.

Table 3: Trips/Vehicle/Day by Census Division

Census Division	Number of Vehicles	Average Trips/Vehicle/Day
EastSouthCentral	2343	2.63
SouthAtlantic	39804	2.65
Pacific	17248	2.66
NewEngland	2284	2.78
WestSouthCentral	14192	2.80
EastNorthCentral	5264	2.80
MiddleAtlantic	12385	2.82
WestNorthCentral	5823	2.85
Mountain	5979	2.87

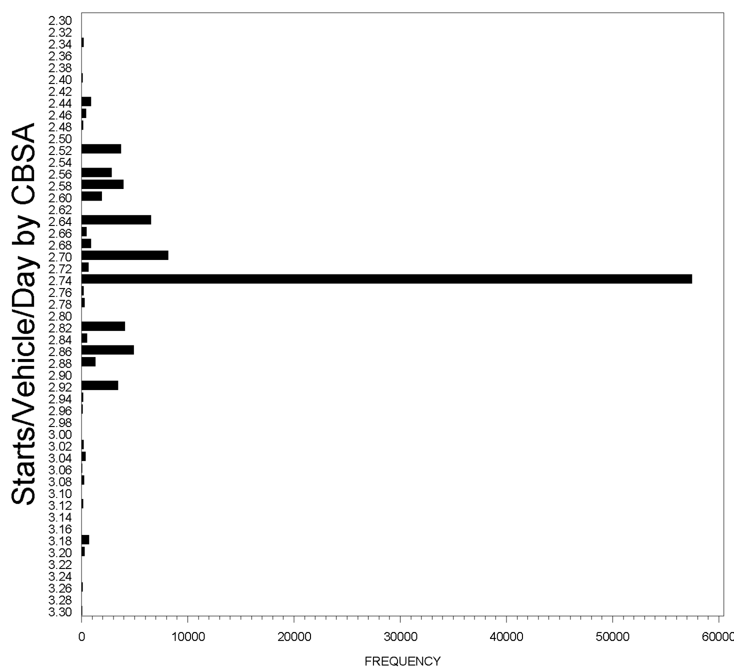
Figure 9 - Trips/Vehicle/Day by State



Starts/Vehicle/Day	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
2.30	0	0	0.00	0.00
2.32	0	0	0.00	0.00
2.34	0	0	0.00	0.00
2.36	0	0	0.00	0.00
2.38	0	0	0.00	0.00
2.40	0	0	0.00	0.00
2.42	0	0	0.00	0.00
2.44	0	0	0.00	0.00
2.46	0	0	0.00	0.00
2.48	0	0	0.00	0.00
2.50	0	0	0.00	0.00
2.52	0	0	0.00	0.00
2.54	11946	11946	11.34	11.34
2.56	5049	16995	4.79	16.14
2.58	479	17474	0.45	16.59
2.60	145	17619	0.14	16.73
2.62	2165	19784	2.06	18.78
2.64	165	19949	0.16	18.94
2.66	16783	36732	15.93	34.88
2.68	8625	45357	8.19	43.07
2.70	2518	47875	2.39	45.46
2.72	1149	49024	1.09	46.55
2.74	182	49206	0.17	46.72
2.76	14489	63695	13.76	60.48
2.78	630	64325	0.60	61.07
2.80	18003	82328	17.09	78.17
2.82	11577	93905	10.99	89.16
2.84	0	93905	0.00	89.16
2.86	2358	96263	2.24	91.40
2.88	5444	101707	5.17	96.57
2.90	0	101707	0.00	96.57
2.92	162	101869	0.15	96.72
2.94	906	102775	0.86	97.58
2.96	0	102775	0.00	97.58
2.98	357	103132	0.34	97.92
3.00	0	103132	0.00	97.92
3.02	1162	104294	1.10	99.02
3.04	0	104294	0.00	99.02
3.06	186	104480	0.18	99.20
3.08	0	104480	0.00	99.20
3.10	580	105060	0.55	99.75
3.12	0	105060	0.00	99.75
3.14	0	105060	0.00	99.75
3.16	0	105060	0.00	99.75
3.18	0	105060	0.00	99.75
3.20	262	105322	0.25	100.00
3.22	0	105322	0.00	100.00
3.24	0	105322	0.00	100.00
3.26	0	105322	0.00	100.00
3.28	0	105322	0.00	100.00
3.30	0	105322	0.00	100.00

/proj1/CRC_A88/NHTS/makebins.sas 28MAR14 11:09

Figure 10 - Trips/Vehicle/Day by CBSA



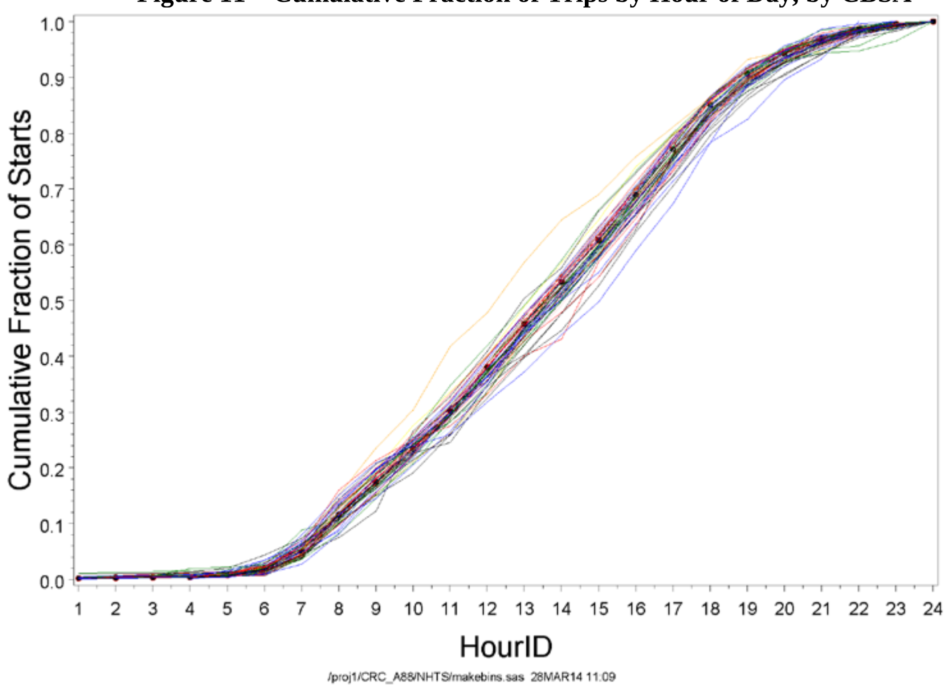
Starts/Vehicle/Day	FREQ.	CUM. FREQ.	PCT.	CUM. PCT.
2.30	0	0	0.00	0.00
2.32	0	0	0.00	0.00
2.34	153	153	0.15	0.15
2.36	0	153	0.00	0.15
2.38	0	153	0.00	0.15
2.40	110	263	0.10	0.25
2.42	0	263	0.00	0.25
2.44	886	1149	0.84	1.09
2.46	420	1569	0.40	1.49
2.48	129	1698	0.12	1.61
2.50	0	1698	0.00	1.61
2.52	3711	5409	3.52	5.14
2.54	0	5409	0.00	5.14
2.56	2836	8245	2.69	7.83
2.58	3937	12182	3.74	11.57
2.60	1892	14074	1.80	13.36
2.62	0	14074	0.00	13.36
2.64	6528	20602	6.20	19.56
2.66	452	21054	0.43	19.99
2.68	865	21919	0.82	20.81
2.70	8141	30060	7.73	28.54
2.72	674	30734	0.64	29.18
2.74	57463	88197	54.56	83.74
2.76	195	88392	0.19	83.93
2.78	281	88673	0.27	84.19
2.80	0	88673	0.00	84.19
2.82	4097	92770	3.89	88.08
2.84	503	93273	0.48	88.56
2.86	4923	98196	4.67	93.23
2.88	1303	99499	1.24	94.47
2.90	0	99499	0.00	94.47
2.92	3446	102945	3.27	97.74
2.94	151	103096	0.14	97.89
2.96	107	103203	0.10	97.99
2.98	0	103203	0.00	97.99
3.00	0	103203	0.00	97.99
3.02	198	103401	0.19	98.18
3.04	348	103749	0.33	98.51
3.06	48	103797	0.05	98.55
3.08	223	104020	0.21	98.76
3.10	0	104020	0.00	98.76
3.12	148	104168	0.14	98.90
3.14	0	104168	0.00	98.90
3.16	0	104168	0.00	98.90
3.18	704	104872	0.67	99.57
3.20	296	105168	0.28	99.85
3.22	0	105168	0.00	99.85
3.24	0	105168	0.00	99.85
3.26	113	105281	0.11	99.96
3.28	0	105281	0.00	99.96
3.30	41	105322	0.04	100.00

/proj1/CRC_A88/NHTS/makebins.sas 28MAR14 11:09

The NHTS dataset also included a Urban/Rural variable, which was assigned to every household. Rural, as well as urban, households were assigned to households whether they were within or outside of CBSAs. The rural vehicles averaged 2.48 trips/vehicle/day and the urban vehicles averaged 2.82 trips/vehicle/day. The wide spread of these two values is almost as wide as the spread of the fifty values for state or CBSA. This perhaps unexpectedly wide spread in average trips/vehicle/day is understandable in terms of differences in human/vehicle behavior in urban vs. rural geographical areas. It suggests that a factor like urban vs. rural or a factor that influences vehicle behavior might be helpful in stratifying the vehicle start distribution for more accurate emissions estimates by MOVES.

Starts/vehicle/day is an important quantity for estimating fleet emissions; this can provide location-specific data on soak distribution, as well as hourly variation in starts. The distribution of starts/vehicle/day among the 24 hours in the day also has an effect on emissions. To examine this, we looked at the effect of geography on the cumulative relative fraction of trips by hour in the NHTS database. The cumulative distribution plots by CBSA is shown in Figure 11, and demonstrates an increased dispersity that occurs as the number of levels increases with each geographical factor.

Figure 11 – Cumulative Fraction of Trips by Hour of Day, by CBSA



Approach to updating NEI inputs: In general, NHTS may be a viable data source to improve location-specific MOVES starts/vehicle, temporal allocation, and soak distribution. However, there are some significant questions about how well NHTS trip survey data represents key-on start events needed by MOVES. For this reason, the CRC panel decided not to pursue updates to start activity under Task 2. For future consideration, the next iteration of NHTS will include GPS-equipped vehicles, which should address the primary questions of reported trips vs. actual key-on events. This may make NHTS a more viable source for start activity in the future. ERG did perform some subsequent analysis on the NHTS database to look at geographic variation in trip/day activity when zero-trip vehicles are excluded, and this work will be included in the overall project report.

3.6. Temporal Distribution of VMT

A discussion of Temporal VMT distribution was included in the draft memo, with a recommendation not to pursue this work under Task 2. The CRC project panel agreed with this recommendation. The evaluation discussion is included below for context and future reference.

Relevant NEI Inputs: Though there are temporal allocations in MOVES CDBs (DayVMTFraction, HourVMTFraction), SMOKE has separate temporal allocation factors that are applied to aggregate VMT. When updated, we would recommend updating both in concert.

Source of current data: National defaults from FHWA studies conducted in the 1990s are the basis of the daily and hourly VMT allocations in MOVES. These defaults were developed from analysis of state-level traffic count data.

Candidate data source identified: The Vehicle Travel Information System (VTRIS) is a database of location and time-resolved state-level traffic count data compiled by FHWA. These data can provide the exact input needed for daily our hourly VMT allocation in MOVES and/or SMOKE. Unfortunately, data is only collected and maintained in VTRIS for 27 states; updating for just these states would still be a significant improvement over national defaults. Under a separate NCHRP project (NCHRP 25-38), ERG is working with Cambridge Systematics to develop resources and data tools for MOVES practitioners. As part of this work, Cambridge is developing a spreadsheet tool that compiled VTRIS data and provides VMT allocation data at the level needed by MOVES. The project will be complete, and tool available to users, by Summer 2014.

Approach to updating NEI inputs: Because the VTRIS tool compiled for NCHRP 25-38 is imminent, and because temporal allocation of VMT is desired not for NEI improvement as much as broader (and perhaps longer term) episodic air quality modeling, ERG recommended (and the CRC panel agree) to not include temporal VMT as part of Task 2 updates.

4. Work to Proceed Under Task 2

Based on the initial screening and detailed evaluations discussed in Section 2 and 3, and the CRC panel comments on the draft memo, ERG will proceed with the following under Task 2:

- Purchase IHS data and derive age distribution and vehicle population for cars and light trucks for each county in the U.S.

- Perform cluster analysis of age distributions within each state to determine if separate distributions are warranted (e.g. urban/rural). This analysis will be useful for re-analysis of representing counties in the SMOKE-MOVES framework.

- Update long-haul VMT using the Freight Analysis Framework (FAF). ERG will use the relative contribution of long haul VMT from the 2007 FAF to re-allocate the 2011 HPMS-based VMT.

- Package and deliver cleaned-up Truck Stop Plus database to allow states and EPA to QA their ongoing work with heavy-duty idle allocations.

- Document additional analysis of NHTS trip data in overall project report.