

Data Processing, Analysis, and Databasing of Heavy-Duty Activity Study Results

ERG:Michael Sabisch

Doug Jackson

Sandeep Kishan

US EPA:

Carl Fulper

UCR / CE-CERT:

Kent Johnson

Thomas Durbin

Kanok Boriboonsomsin

George Scora

Daniel Sandez

Presented at the UCR / CE-CERT 7th Annual PEMS Conference
March 31, 2017



1

Study Overview

- Activity data collection of 100 local-service heavy-duty vehicles over a one to three month period each, 1 Hz, approximately 160 SAE J1939 fields
- Study focused on collection and characterization of activity and operational data from local trucks that could contribute a disproportionately large share of NOx emissions to local inventory
- Stop / go and low speed operation may challenge effectiveness of truck's selective catalytic reduction (SCR) system.



2

Study Overview, contd.

California Air Resources Board (CARB), Center for Environmental Research & Technology at the University of California, Riverside (CE-CERT), InfoWedge conducted field study

- In-field fleet characterization, instrumentation, data collection

US EPA – Provided equipment, data collection, data analysis support directly and also through contract with ERG

CE-CERT gathered, processed, analyzed data

- Dataset corrections and enhancements - time stamp corrections, GPS coordinates scrubbing (first / last trip mile) for confidentiality, GPS / ECU vehicle speed corrections, GIS merge for in-state / out-of-state road type assignments, and idle / extended idle determinations
- Performed analysis necessary to meet study objectives (some presented elsewhere)

ERG – Support under contract to EPA, focus of this presentation



3

Data Description

- CE-CERT transferred data to ERG (each vehicle single CSV containing all operation over period of instrumentation)
- Data included 160+ SAE J1939 parameters, GPS, and the new fields created by CE-CERT
- Data consisted of the following:
 - SAE J1939 Engine operational data
 - SAE J1939 Emission control system data
 - Trip activity data (speed, load, elevation, time, etc.)
 - J1939 diagnostic data (DMs, FMIs, stability PIDS)
 - Corrected and additional fields from CE-CERT



4

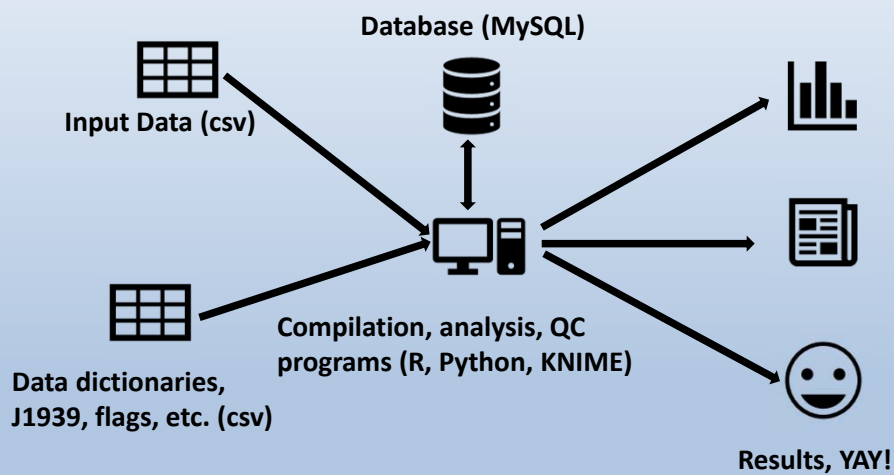
ERG Data Handling Process

- Open Source Software Packages (R, Python, MySQL, KNIME)
- Due to memory limitations, stored and accessed data (≈ 180 GB) from within MySQL database
 - Analyzed / processed entire vehicle dataset (1 – 4 GB of data for each vehicle) for some operations, pulled individual variables for others
- Data processed and analyzed using R and Python (MySQL queries from within programs)



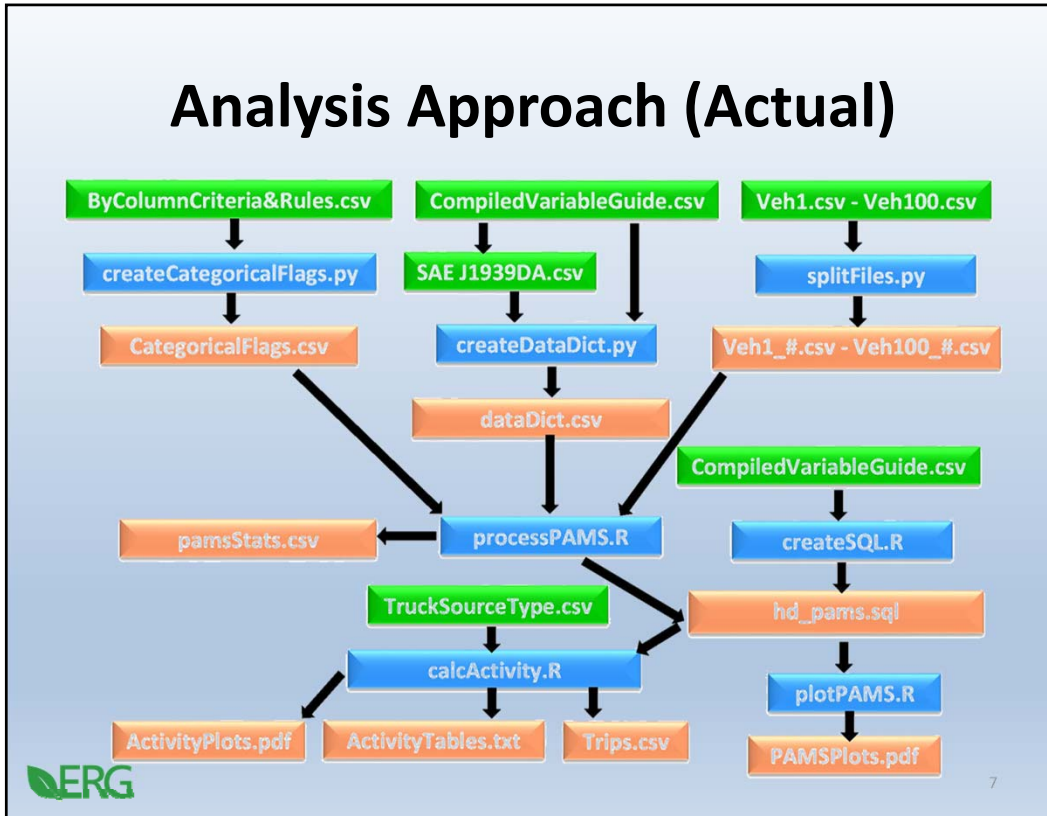
5

Analysis Approach (Theoretical)



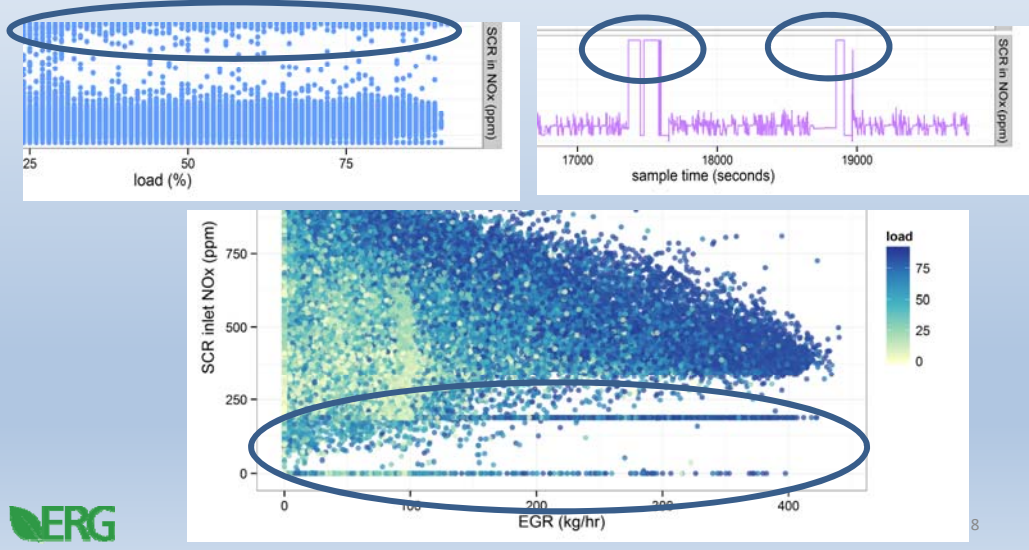
6

Analysis Approach (Actual)



Initial Data QC Process

Initial process involved engineering review of data plots and statistics to identify outliers



Revised Data QC Process

- Amount of data and number of identified data issues necessitated more “automated” approach involving development of “flags” within database
- Data Analyzed in R and assigned “flag” fields associated with each variable
 - “Categorical” Flags – Used to identify “suspect” data based on operating status or other SAE J1939 status / FMI PIDS
 - “Variable out of Range” (VOOR) Flags – Applied based on behavior / disposition of each variable during operation
- All original data retained – database “flag” variables identify specific variables (and observations) that are suspect
- Filtered dataset created by MySQL query on flags (in R or otherwise)
- Corrected variables created, as needed



9

Categorical and VOOR Flags

“Categorical” Flags – Applied to selected variables

- Data acquired when the engine was not running (key on, engine off)
- Data acquired during the first minute of engine operation (spikes and erroneous data common during start-up)
- Fault Mode Indicators - SPNs with associated FMIs of 2 - 9, 11 - 14, & 19 - 21 were flagged, per SAE J1939-73, Section A.1.2
- Stability Indicators – SPNs with associated stability SPNs \neq 1 were flagged, per SAE J1939-71

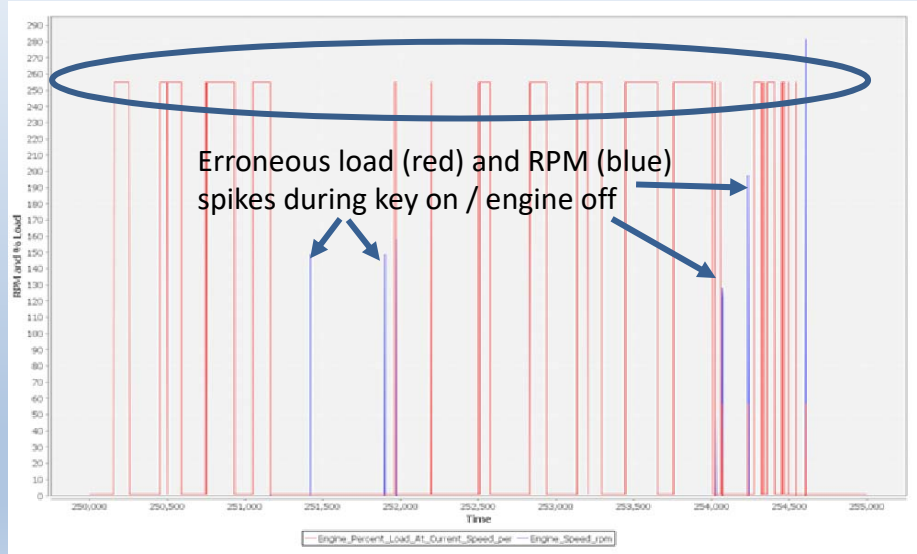
“Variable out of Range” Flags – Applied to all variables

- 0 = within SAE J1939 operational range
- 1 = data outside of SAE J1939 operational range
- 2 = “erroneously” static but within range (for entire vehicle dataset)
- 3 = “erroneously” static and out of range (for entire vehicle dataset)



10

Data Collected with Engine Off



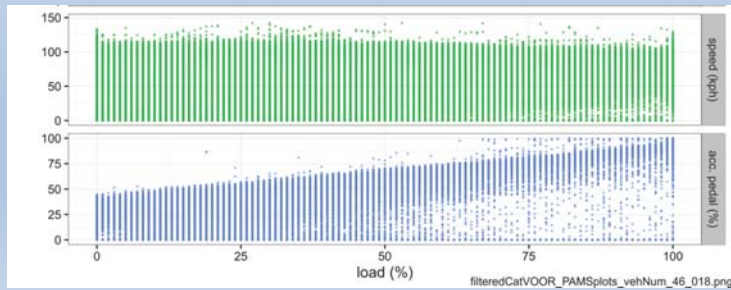
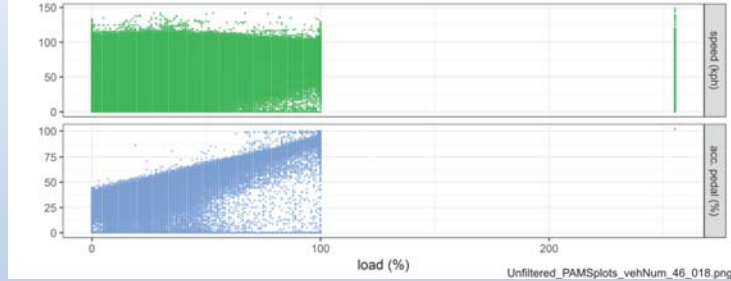
Engine Start / Stop Spikes

Time	RPM	EGR Flow Rate (kg/hr)	Intake MAF (kg/hr)	Exh Temp (C)	Intake_MAP_kPa	DEF Tank Level (%)	Engine_Fuel_Temp (C)	Engine_Oil_Temp (C)
2015-04-01T09:30:41.0	0	62.6	0	167.72	94	63.6	51	95
2015-04-01T09:30:42.0	0	62.6	0	167.72	94	63.6	51	95
2015-04-01T09:30:43.0	0	3276.75	3276.75	1774.97	510	102	215	1774.97
2015-04-01T09:30:44.0	0	0	0	119.91	94	59.6	14	79
2015-04-01T09:30:45.0	0	0	0	121.25	94	59.6	57	75
2015-04-01T09:30:46.0	0	0	0	121.59	94	59.6	58	75
2015-04-01T11:39:54.0								
2015-04-01T11:39:55.0	0	3276.75	3276.75	1774.97	510	102	215	1774.97
2015-04-01T11:39:56.0	0	0	0	128.09	94	54.8	14	83
2015-04-01T11:39:57.0	0	0	0	128.91	94	54.8	60	80
2015-04-01T11:39:58.0	0	0	0	128.34	94	54.8	61	80
2015-04-01T11:39:59.0	0	0	0	128.81	94	54.8	61	80
2015-04-01T11:40:00.0	0	0	0	128.47	94	54.8	61	80



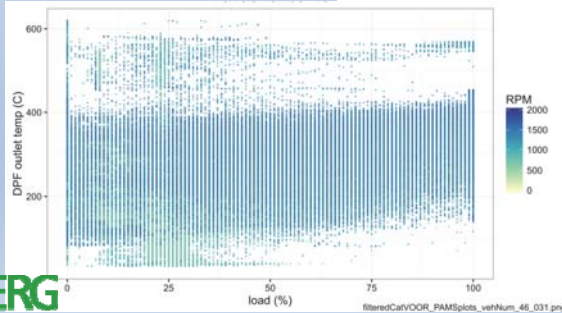
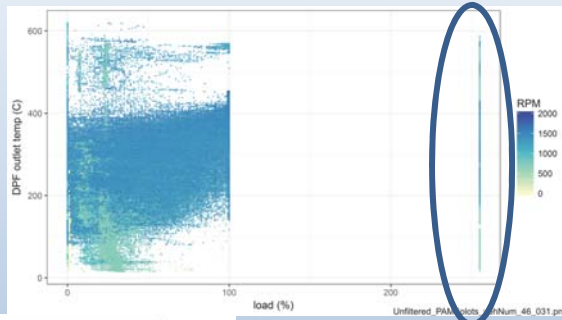
Automated Data Filtering

Speed / Accelerator Position vs. Load



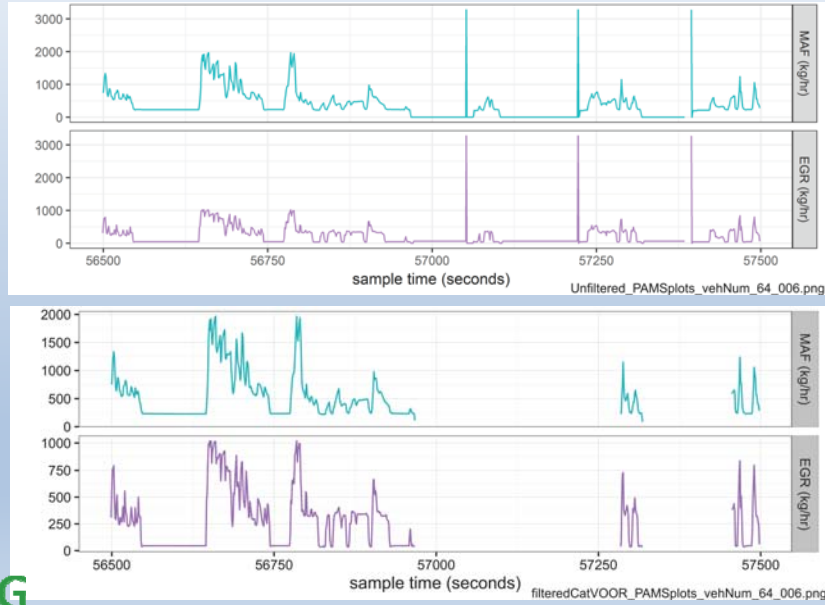
Automated Data Filtering

DPF Outlet Temperature vs. Load



Automated Data Filtering

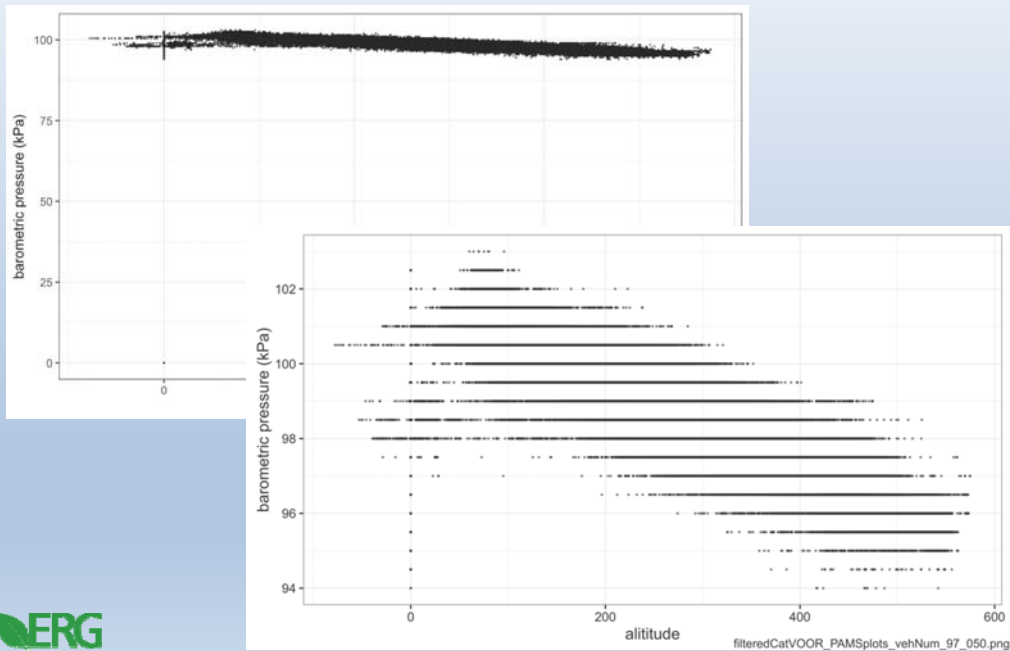
MAF and EGR Startup Spike Removal



15

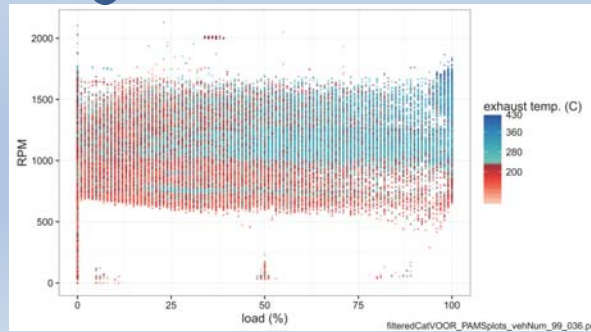
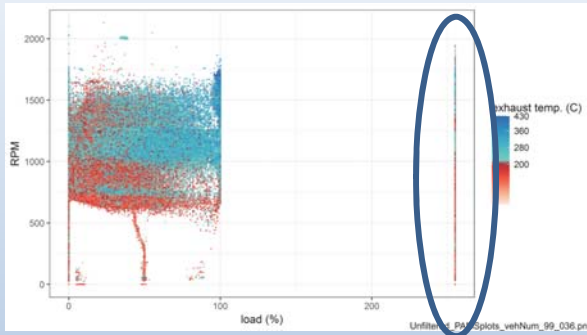
Automated Data Filtering

Barometric Pressure vs. Altitude



Automated Data Filtering

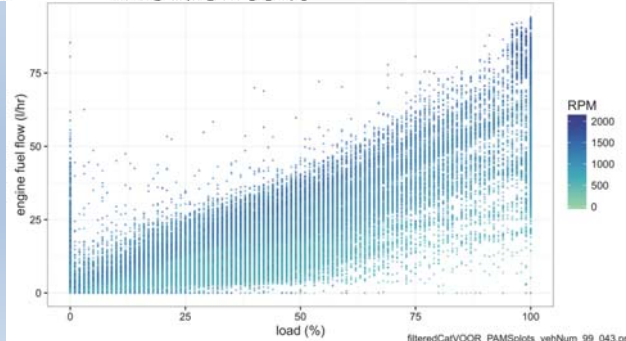
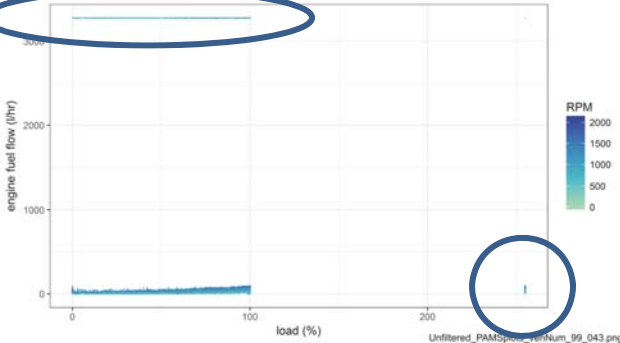
RPM vs. Load



17

Automated Data Filtering

Engine Fuel Rate vs. Load



18

Activity Characterization

- Once data was filtered, “validated” data and “flag” variables stored in MySQL database for data extraction and analysis
- focus shifted to characterizing operational activity for use in emissions modeling
- Second-by-second data was aggregated into “trip-based” file for use in activity characterization analysis of starts, soaks, and VMT distributions
- Development of “trip-based” file reduced dataset from over 100,000,000 records to a dataset with roughly 40,000 records, so database management not required



19

Activity Characterization

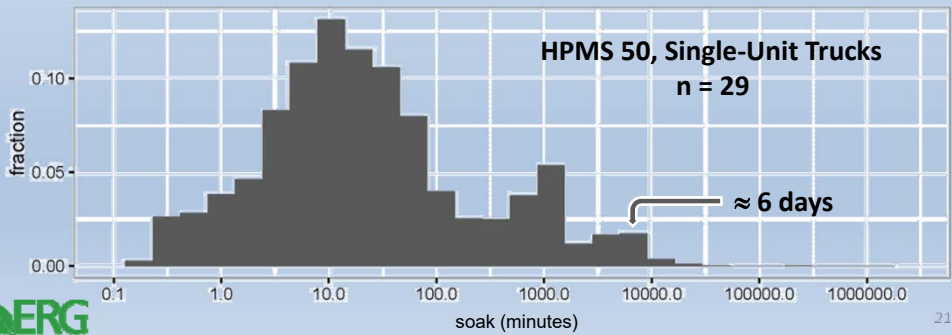
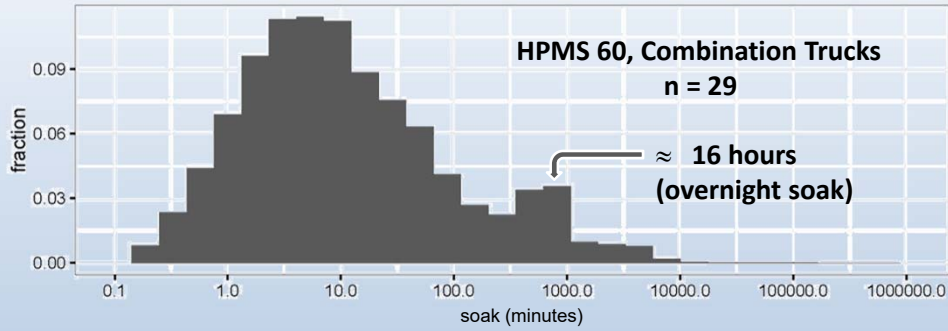
- Results developed by truck source types, HPMS groupings, and by overall / weekday / weekend
 - Source types assigned based on review of truck characteristics, reported usage, lifetime mileage, and mileage accumulated during PAMS instrumentation
 - HPMS groupings eliminate source type uncertainty

Vehicle Type	Source Type ID	HPMS Group
Light Commercial Truck	32	30
Intercity Bus	41	40
Transit Bus	42	40
School Bus	43	40
Refuse Truck	51	50
Single-Unit Short Haul Truck	52	50
Single-Unit Long Haul Truck	53	50
Combo Short Haul truck	61	60
Combo Long Haul Truck	62	60



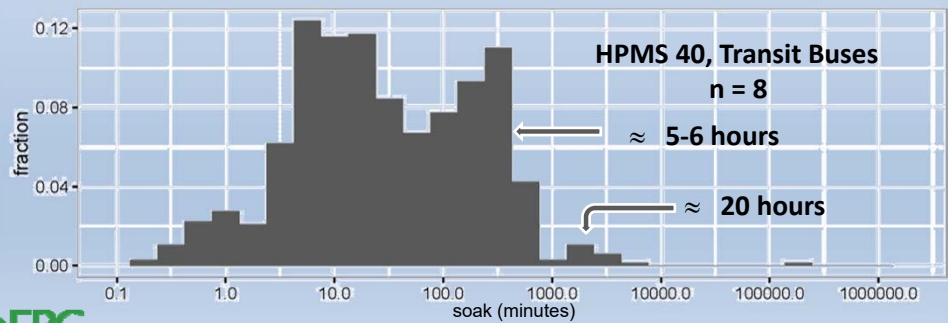
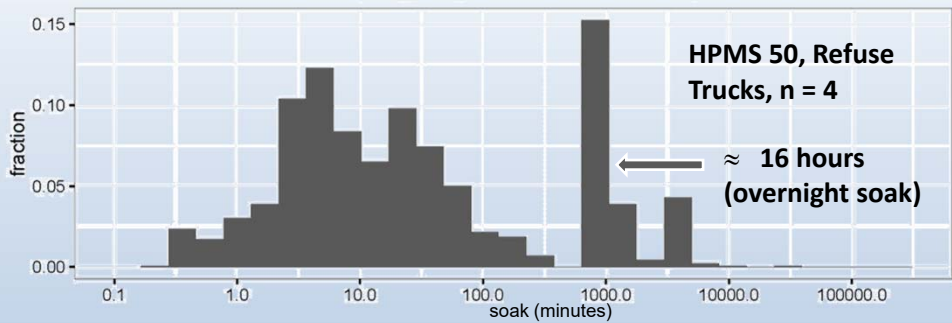
20

Soak Distributions per Trip



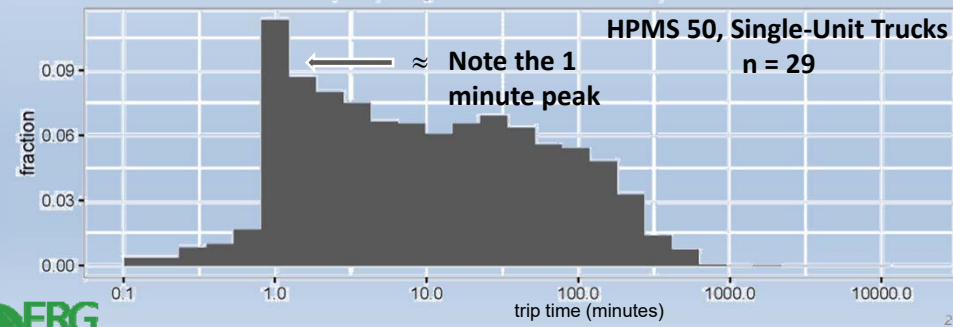
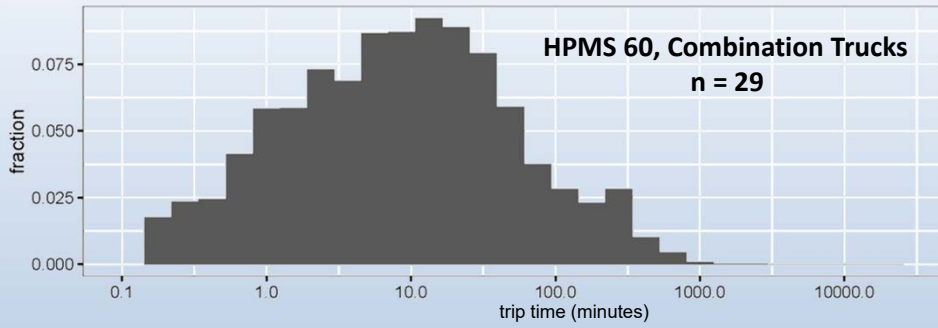
21

Soak Distributions per Trip



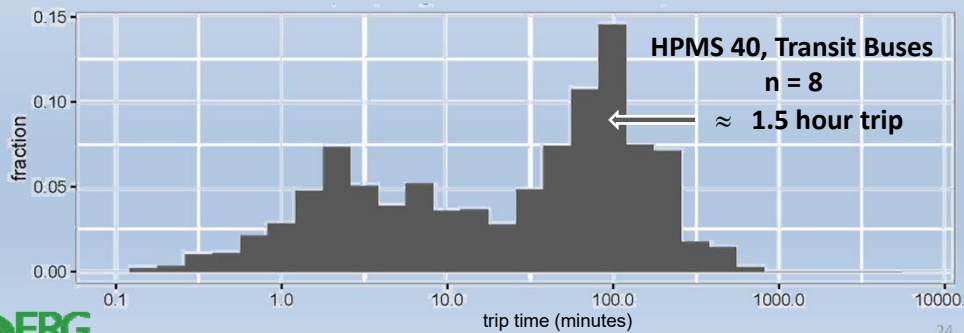
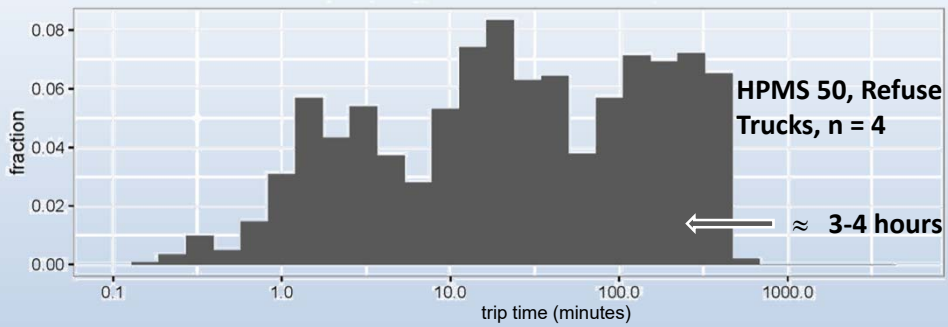
22

Time Distributions per Trip



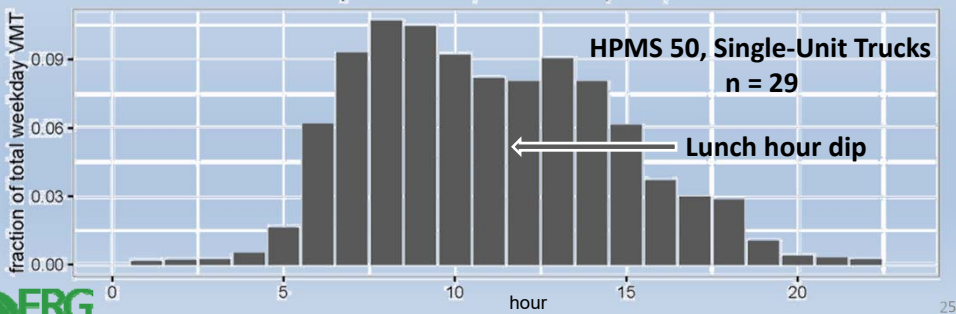
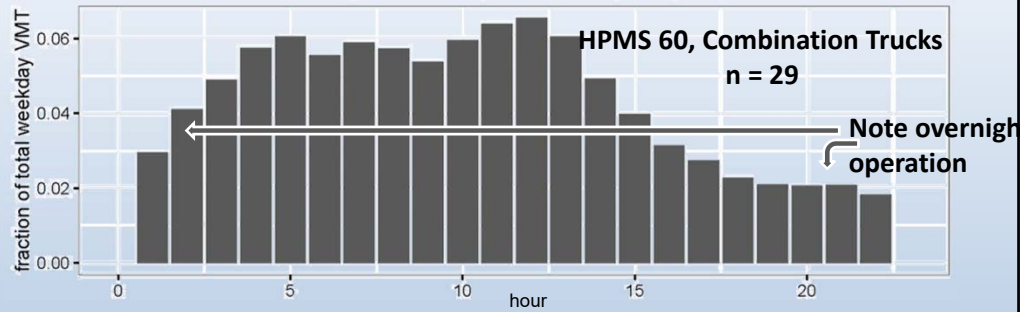
23

Time Distributions per Trip



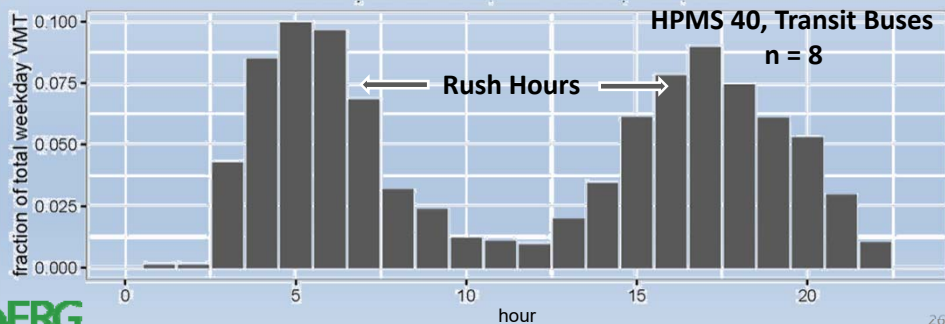
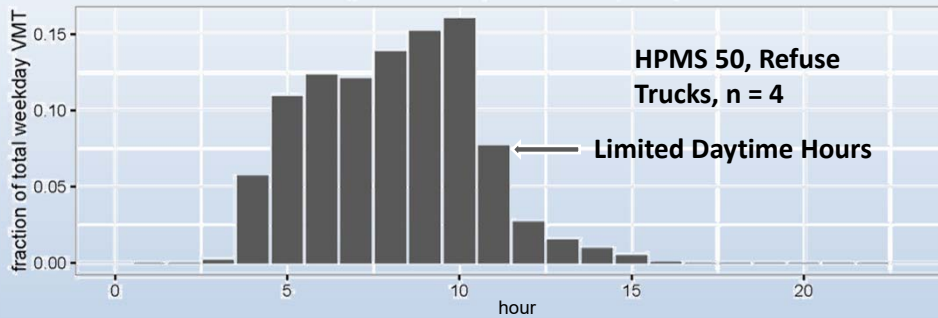
24

VMT Distributions per Hour



25

VMT Distributions per Hour



26

Acknowledgements

ERG gratefully acknowledges the following study partners and presentation co-authors :

Kent Johnson, Thomas Durbin, Kanok Boriboonsomsin George Scora and Daniel Sandez (**UCR / CE-CERT**) for conducting the activity study from which this data was acquired, for project collaboration and for providing technical support

Carl Fulper (**US EPA**) for providing ERG support funding, equipment support, and technical study support



27

Thank you for your time!

michael.sabisch@erg.com



28