The Effects of Ultra-Low Sulfur Gasoline on Emissions from Tier 2 Vehicles in the In-Use Fleet

Final Report

United States
Environmental Protection
Agency

The Effects of Ultra-Low Sulfur Gasoline on Emissions from Tier 2 Vehicles in the In-Use Fleet

Final Report

Assessment and Standards Division
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

NOTICE

This technical report does not necessarily represent final EPA decisions or positions. It is intended to present technical analysis of issues using data that are currently available. The purpose in the release of such reports is to facilitate the exchange of technical information and to inform the public of technical developments.

EPA United States
Environmental Protection
Agency

EPA-420-R-14-002
March 2014

Table of Contents

Acronyms 3

1. Executive Summary 4
2. Introduction 8
   - 2.1. Background 8
   - 2.2 Motivation for this Study 11
3. Study Design 11
   - 3.1. Measurement of Reversible In-Use Loading (Clean-Out Effect) 12
   - 3.2 Effect of Sulfur Level 13
4. Test Vehicle Selection, Recruitment, and Delivery 13
   - 4.1 Choice of Makes and Models 13
   - 4.2 Vehicle Recruitment Criteria 15
   - 4.3 Initial Checks and Test Vehicle Delivery 15
5. Test Fuel Specs and Procurement 16
6. Test Procedures 17
   - 6.1 Initial fuel exchange and vehicle prep 17
   - 6.2 Test procedure description 18
7. Data Analysis and Results 22
   - 7.1. Data Preparation 24
     - 7.1.1. Imputation of measurements with low concentration 24
     - 7.1.2. Detection of outliers 27
   - 7.2. Modeling Methodology 28
   - 7.3. Statistical Analysis and Results 30
     - 7.3.1. Effect of clean-out at 28 ppm 30
     - 7.3.2. Effect of clean-out at 5 ppm 35
     - 7.3.3. Effect of Sulfur level 39
     - 7.3.3.1. Tier 2 Vehicles 39
     - 7.3.3.2. Tier 3-like Vehicles 53
     - 7.3.4. Sensitivity Analysis 56
8. Summary and Conclusions 59
9. References 62

Appendix A. Details of On-Road Mileage Accumulation Route

Appendix B. Emission Concentration Plots

Appendix C. Additional Information on Emission Measurements
Appendix D. Discussion of Univariate and Multivariate Analysis of Variance
Appendix E. Plots of All Pollutants and Bags
Appendix F. Sensitivity Analysis of Influential Vehicles

Acronyms

A/F Ratio	Air-to-Fuel Ratio
BIC	Schwarz Bayesian Criterion
CH4	Methane
CO	Carbon Monoxide
CRC	Coordinating Research Council
CS	Compound Symmetry Covariance
EPA	Environmental Protection Agency
F	Fahrenheit
FID	Flame Ionization Detector
FTP	Federal Test Procedure
HC	Hydrocarbons
LOQ	Limit of Quantification
MIL	Malfunction Indicator Lamp
ML	Maximum Likelihood
MSAT	Mobile Source Air Toxics
NMHC	Non-Methane Hydrocarbons
NMOG	Non-Methane Organic Gases
NO_x	Oxides of Nitrogen
NVFEL	National Vehicle and Fuel Emission Laboratory
PGM	Platinum Group Metals
PM	Particulate Matter
PPB	Parts per Billion
PPM	Parts per Million
PSI	Pounds per Square Inch
PZEV	Partial Zero-Emissions Vehicle
REML	Restricted Maximum Likelihood
RLD	Restricted Likelihood Distance
RVP	Reid Vapor Pressure
SAS	Statistical Analysis Systems
THC	Total Hydrocarbons
UN	Unstructured Covariance

1. Executive Summary

Fuel sulfur content has long been understood to affect the performance of emission aftertreatment catalysts in light duty vehicles, where the sulfur and its oxides occupy active precious metal sites and oxygen storage materials, reducing the catalyst's efficiency in destroying pollutants. Numerous studies have shown the direct impact of fuel sulfur levels above 30 ppm on emissions, data that formed the basis of the sulfur controls in EPA's Tier 2 rulemaking for light duty vehicle emissions, published in 2000.[1]

Following the successful implementation of the Tier 2 sulfur standards, new research has focused on the emission reduction potential of lowering sulfur levels below 30 ppm, particularly on Tier 2 and newer technology vehicles, under the hypothesis that increased reliance on the catalytic convertor will result in a higher sensitivity to fuel sulfur content. A 2005 study conducted jointly by EPA and several automakers found large decreases in NOx and HC emissions from vehicles meeting Tier 2 Bin 5 emission levels when operating on 6 ppm versus 32 ppm sulfur test fuel.[2] In order to gain further understanding of how these emission reductions would translate into the in-use fleet, EPA conducted the study described here to assess the state of sulfur loading (poisoning) in typical in-use vehicles, as well as the effect of fuel sulfur level on these vehicles during subsequent mileage accumulation. It was designed to take into consideration what was known from prior studies on sulfur build-up in catalysts over time and the effect of periodic regeneration events that can occur during higher speed and load operation in day-to-day driving.

The main study sample described in this analysis consisted of 93 cars and light trucks recruited from owners in southeast Michigan, covering model years 2007-9 with approximately 20,000-40,000 odometer miles. The makes and models targeted for recruitment were chosen to be representative of high sales vehicles covering a range of types and sizes. The main study vehicle selection was also consistent with the EPAct/V2/E-89 study to allow potential linking of the two programs. While the main study vehicle selection did not specifically target cleaner emission vehicle technologies, the supplemental study acquired additional vehicles with "Tier 3-like" emission levels and technologies as discussed later. Test fuels were two non-ethanol gasolines with properties typical of certification fuel, one at a sulfur level of 5 ppm and the other at 28 ppm, the higher level chosen to be representative of retail fuel available to the public in the vehicle recruiting area. All emissions data was collected using the FTP cycle at a nominal ambient temperature of 75°F.

Using the 28 ppm test fuel, emissions data were collected from vehicles in their as-received state, and then following a high-speed/load "clean-out" procedure consisting of two back-to-back US06 cycles intended to reduce sulfur loading in the catalyst. A statistical analysis of this data showed highly significant reductions in several pollutants including NO_x and hydrocarbons (Table ES-1), suggesting that reversible sulfur loading exists in the in-use fleet and has a measurable effect on aftertreatment performance. For example, Bag 2 NO_x emissions dropped 31 percent between the pre- and post-cleanout tests on 28 ppm fuel.

Table ES-1 Percent Reduction in In-Use Emissions After the Clean-Out Using 28 ppm Test Fuel

	NO_x (p-value)	**THC (p-value)**	**CO (p-value)**	**NMHC (p-value)**	**CH_4 (p-value)**	**PM (p-value)**
Bag 1	–	–	6.0% (0.0151)	–	–	15.4% (< 0.0001)
Bag 2	31.4% (0.0003)	14.9% (0.0118)	–	18.7% (0.0131)	14.4% (0.0019)	–
Bag 3	35.4% (<0.0001)	20.4% (<0.0001)	21.5% (0.0001)	27.7% (<0.0001)	10.3% (<0.0001)	24.5% (<0.0001)
FTP Composite	11.4% (0.0002)	3.8% (0.0249)	6.8% (0.0107)	3.5% (0.0498)	6.0% (0.0011)	13.7% (<0.0001)
Bag 1 – Bag 3	–	–	7.2% (0.0656)	–	–	–

The clean-out effect is not significant at $\alpha = 0.10$ when no reduction estimate is provided.

To assess the impact of lower sulfur fuel on in-use emissions, further testing was conducted on a representative subset of Tier 2 vehicles on 28 ppm and 5 ppm fuels with accumulated mileage. A first step in this portion of the study was to assess differences in the effectiveness of the clean-out procedure under different fuel sulfur levels. Table ES-2 presents a comparison of emissions immediately following (<50 miles) the clean-out procedures at the low vs. high sulfur level. These results show significant emission reductions for the 5 ppm fuel relative to the 28 ppm fuel immediately after this clean-out; for example, Bag 2 NO_x emissions were 34 percent lower after a clean-out on the 5 ppm fuel compared to that following the clean-out on the 28 ppm fuel. This indicates that the catalyst is not fully desulfurized, even after a clean out procedure, as long as there is sulfur in the fuel. This further indicates that current sulfur levels in gasoline continue to have a long-term, adverse effect on exhaust emissions control that is not fully removed by intermittent clean-out procedures that can occur in day-to-day operation of a vehicle and demonstrates that lowering sulfur levels to 10 ppm on average will significantly reduce the effects of sulfur impairment on emissions control technology.

Table ES-2 Percent Reduction in Emissions from 28 ppm to 5 ppm for the First Three Repeat FTP Tests Immediately Following Clean-Out

	NO_x (p-value)	THC (p-value)	CO (p-value)	NMHC (p-value)	CH_4 (p-value)	PM
Bag 1	5.3% (0.0513)	6.8% (0.0053)	6.2% (0.0083)	5.7% (0.0276)	14.0% (<0.0001)	–
Bag 2	34.4% (0.0036)	33.9% (<0.0001)	–	26.4% (0.0420)	49.4% (<0.0001)	–
Bag 3	42.5% (<0.0001)	36.9% (<0.0001)	14.7% (0.0041)	51.7% (<0.0001)	28.5% (<0.0001)	–
FTP Composite	15.0% (0.0002)	13.3% (<0.0001)	8.5% (0.0050)	10.9% (0.0012)	23.6% (<0.0001)	–
Bag 1 – Bag 3	–	–	–	–	–	–

The effectiveness of clean-out cycle is not significant at $\alpha = 0.10$ when no reduction estimate is provided.

To assess the overall in-use reduction between high and low sulfur fuel, a mixed model analysis of all data as a function of fuel sulfur level and miles driven after cleanout was performed. This analysis found highly significant reductions for several pollutants, as shown in Table ES-3; the reductions for Bag 2 NOx were particularly high, estimated at 52 percent between 28 ppm and 5 ppm. For all pollutants, the model fitting did not find a significant miles-by-sulfur interaction, suggesting that the effect of sulfur level does not depend on miles driven after the fuel change, and therefore, the emission benefits of lower fuel sulfur occurred immediately and continued as miles were accumulated.

Table ES-3 Percent Reduction in Emissions from 28 ppm to 5 ppm Fuel Sulfur on In-Use Tier 2 Vehicles

	NO_x (p-value)	THC (p-value)	CO (p-value)	NMHC (p-value)	CH_4 (p-value)	NO_x+NMOG (p-value)	PM[a]
Bag 1	7.1% (0.0216)	9.2% (0.0002)	6.7% (0.0131)	8.1% (0.0017)	16.6% (< 0.0001)	N/A	–
Bag 2	51.9% (< 0.0001)	43.3% (< 0.0001)	–[a]	42.7% (0.0003)	51.8% (< 0.0001)	N/A	–
Bag 3	47.8% (< 0.0001)	40.2% (< 0.0001)	15.9% (0.0003)	54.7% (< 0.0001)	29.2% (< 0.0001)	N/A	–
FTP Composite	14.1% (0.0008)	15.3% (< 0.0001)	9.5% (<0.0001)	12.4% (< 0.0001)	29.3% (< 0.0001)	14.4% (<0.0001)	–
Bag 1 - Bag 3	–[a]	5.9% (0.0074)	–[a]	–[b]	–[b]	N/A	–

[a] Sulfur level not significant at $\alpha = 0.10$.

[b] Inconclusive because the mixed model did not converge.

Major findings from this study include:

- Reversible sulfur poisoning is occurring in the in-use fleet of Tier 2 vehicles and has a measureable effect on emissions of NO_x, hydrocarbons, and other pollutants of interest.
- The effectiveness of high speed/load procedures in restoring catalyst efficiency is limited when operating on higher sulfur fuel.
- Reducing the fuel sulfur levels from 28 to 5 ppm is expected to achieve significant reductions in emissions of NO_x, hydrocarbons, and other pollutants of interest from a broad range of common in-use Tier 2 vehicles.
- Relatively large effect of sulfur on Bag 2 NOx were found to be robust based on the sensitivity analyses examining the impact of influential vehicles and measurement uncertainty at low emission levels.
- Lower-emitting Tier 3 vehicles are expected to show similar, if not greater, sensitivities to the fuel sulfur levels compared to the conventional Tier 2 vehicles currently in-use, based on the analyses of “Tier 3-like” vehicles.

A draft version of this report underwent an independent peer review covering the design, analysis methods, and results. This process was conducted according to guidelines described in EPA’s Peer Review Handbook, and did not produce any significant adverse findings. A detailed description of the process and results is available on the EPA Science Inventory website.[3]

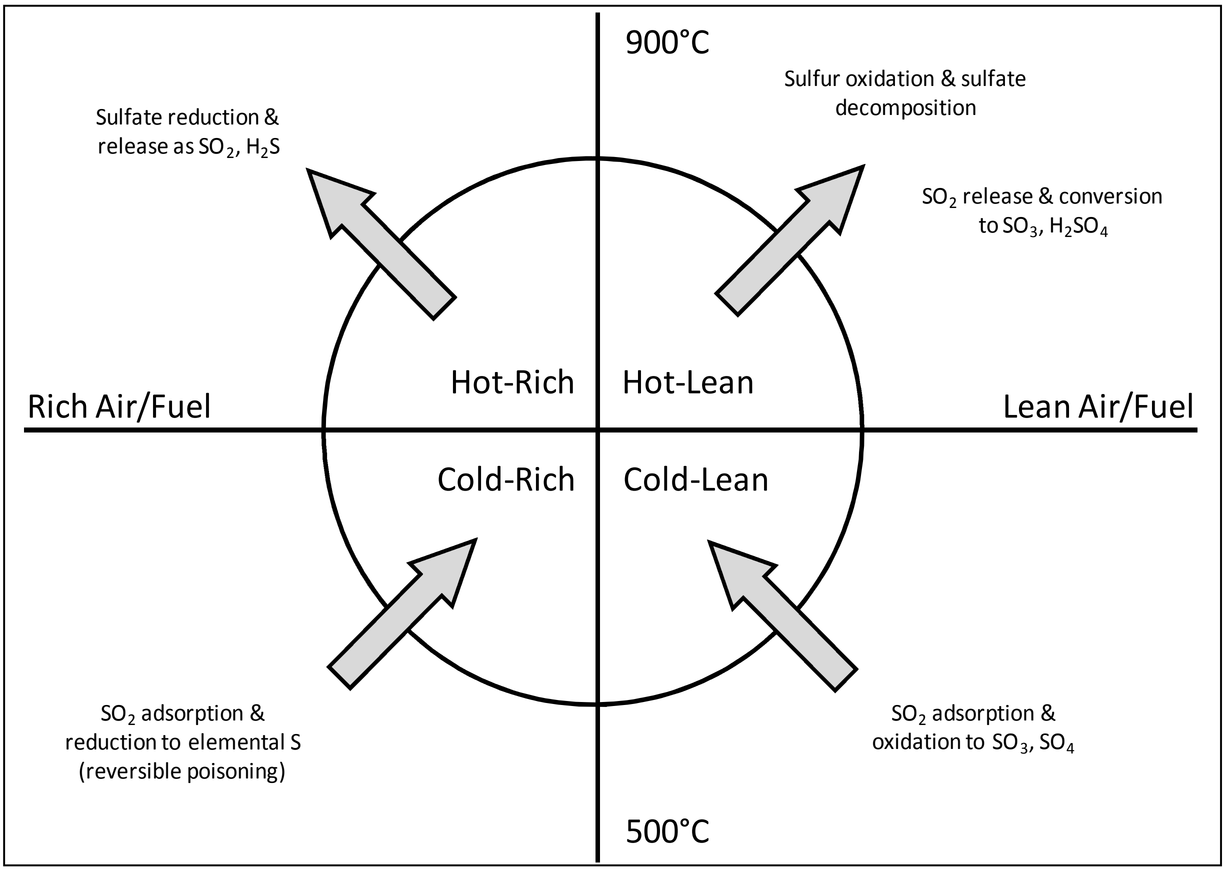
2. Introduction

2.1. Background

Sulfur in gasoline has long been found to reduce the conversion efficiency of automotive three-way catalysts, with some studies suggesting an increase in catalyst sensitivity (in terms of percent conversion efficiency) to sulfur as tailpipe emission levels decrease.[4] The Tier 2 light duty emission standards recognized the importance of sulfur and required lower gasoline sulfur levels in accordance with the relative level of the new emission standards. Since that time, little work has been published on the effect of further reductions of fuel sulfur level, especially on in-use vehicles that may have some degree of catalyst deterioration due to real world operation. This study investigates the effects of lower sulfur gasoline on in-use Tier 2 vehicles that have been operated on commercial gasoline (with relatively higher sulfur levels) under real world driving conditions.

Aftertreatment systems utilize catalysts containing precious metals and metal oxides to selectively oxidize hydrocarbons and reduce nitrogen oxides in the exhaust gases. Sulfur oxides from fuel combustion preferentially bind to active sites in the catalyst, inhibiting their ability to participate in the intended conversion reactions (this is often referred to as sulfur poisoning). The amount of sulfur retained by the catalyst is a function of the type and arrangement of active materials and coatings used within the catalyst, its operating temperature, as well as the air-to-fuel ratio and concentration of sulfur in the exhaust gas.[5,6] Modern vehicle engines operate with rich-lean oscillations that maintain the proper oxidation-reduction condition of the catalyst, but under typical driving conditions, there is a nonzero equilibrium level of sulfur retained. Regularly operating the catalyst at a high temperature under net reducing conditions can release much the sulfur oxides from the catalyst (Figure 2-1), and could minimize the effects of fuel sulfur on catalyst efficiency. However, producing these conditions at sustained and/or regular intervals would accelerate thermal degradation of the catalyst and may also raise other challenges for emissions control and fuel economy.

Figure 2-1 Catalyst-Sulfur Interaction by Temperature and A/F Ratio (Adapted from Information from Ford Motor Company)



Additionally, not being able to maintain these catalyst temperatures (e.g., cold weather, idles), and the rich air-to-fuel ratios can result in increased PM, NMOG and CO emissions. Therefore, reducing fuel sulfur levels has been the primary regulatory mechanism to minimize sulfur contamination of the catalyst and ensure optimum emissions performance over the useful life of a vehicle.

In 2005, EPA and several automakers jointly conducted a program that examined the effects of sulfur, benzene, and volatility on gaseous emissions from a fleet of nine Tier 2 compliant vehicles (referred to as the Mobile Source Air Toxics or MSAT study[7]). The fuel matrix consisted of four non-ethanol fuels blended in a stepwise manner, starting with a purchased base fuel containing the lowest levels of the three properties of interest. The base fuel contained 6 ppm sulfur, and after additions of butane (to arrive at 9 psi RVP) and benzene (to arrive at 1.0 vol. %), sulfur was increased to 32 ppm using a small amount of doping agent. Thus, the sulfur effect could be deduced by comparing emission results between the final fuel and the one just before it in the blending sequence. The test vehicles were production samples owned by the manufacturers, which had emission catalysts and oxygen sensors bench-aged to the equivalent of 120,000 miles. The dataset consisted of tests performed at four different labs, with three manufacturers testing two of their own vehicles, and EPA testing the remaining three at its Ann Arbor, Michigan, laboratory. Tests conducted on high-sulfur fuel were preceded by an extended cruise period to simulate sulfur loading, and tests on low-sulfur fuels were preceded by a high-speed/load cycle used in previous programs performed by the Coordinating Research Council.[8]

Mixed model analyses were performed on the data using vehicle and lab as random effects and fuel properties as fixed effects. Reductions in FTP-weighted emissions for 6 compared to 32 ppm sulfur were 33 percent for NOx, 11 percent for THC, 17 percent for CO, and 32 percent for methane (all statistically significant at $\alpha = 0.10$). While the test procedures may have produced sulfur loading in some vehicles beyond what would normally be expected in-use, these data suggested that there were likely to be significant emission reductions possible with further reductions in gasoline sulfur level.

A study published in 2011 by Umicore Autocat USA describes tests performed on a model year 2009 Chevrolet Malibu calibrated to PZEV levels (20 mg/mi NOx), intended to examine the hypothesis that very-low emitting vehicles are more sensitive to fuel sulfur content. To examine the impact of sulfur on the underfloor catalyst efficiency, the study used two test fuels, at 3 and 33 ppm sulfur, over FTP, US06, and a high speed/load clean-out cycle similar to one used in the MSAT study described above,. The authors observed consistent degradation of catalyst performance when repeat FTPs were performed using the 33 ppm fuel, versus none with the 3 ppm test fuel. Emission measurements showed a NOx reduction of 41 percent between averages of three replicate FTPs on low versus high sulfur level.[9]

2.2 Motivation for this Study

The MSAT study described in the previous section motivated the design of a new study to further investigate the effects of lower sulfur gasoline on a larger number of vehicles. Specifically, we wanted to assess what the benefits of sulfur control might be under in-use driving conditions for typical modern vehicles, as previous studies have generally not provided data that can be utilized for this type of analysis in a straightforward manner. Many studies have designs that look for a change in emissions at a single point in time shortly after a fuel sulfur change and/or a clean-out cycle, while in fact, catalysts of in-use vehicles are normally operating at a point somewhere between being clean of sulfur (e.g., following high speed, high load and therefore high catalyst temperature operation) and fully loaded with sulfur (e.g., following low speed, low load, cool catalyst temperature operation). Others have measured the effects of aging catalysts on different fuel sulfur levels, which do not give any information on what happens when fuel sulfur is changed partway through a vehicle's life. Thus, the design of this study sought to assess the impact of reversible sulfur loading on in-use Tier 2 vehicles recruited from their owners, as well as the emission performance after a change to lower sulfur fuel and subsequent mileage accumulation. Additionally, the supplemental testing of "Tier 3-like" vehicles sought to assess the impact of fuel sulfur levels on vehicles and technologies at the lowest emission levels not yet prevalent in the Tier 2 fleet but similar to vehicles expected in coming years due to LEV III and Tier 3 requirements.

3. Study Design

This section outlines the hypotheses and theoretical bases for the specific procedures that follow in the testing section.

In order to characterize the effect of sulfur on the in-use fleet, vehicles were recruited from their owners and tested in EPA's National Vehicle and Fuel Emission Laboratory (NVFEL) in Ann Arbor, Michigan. Given this arrangement, it was not feasible to directly measure sulfur loading in the catalysts themselves by means that involved damaging or destroying them.

Instead, the behavior of emissions during various test procedures was used as a proxy for sulfur loading.

The program design used two fuels differing only in sulfur level, with the higher level of 30 ppm chosen to match the current average Tier 2 sulfur limit. All other fuel properties were similar to Tier 2 certification gasoline used today to demonstrate compliance with Tier 2 emission standards (see Table 5-1).[i]

3.1. Measurement of Reversible In-Use Loading (Clean-Out Effect)

The level of reversible in-use sulfur storage (or loading) and release within an exhaust catalyst system can be assessed by measuring emissions from the vehicle as received, performing a high-speed, high-load clean-out cycle, then measuring emissions again. This change in emissions represents the amount of reversible deactivation relative to a "clean" baseline state. Prior to the start of this test program, we consulted previous studies as well as vehicle calibration and emission controls experts to select an appropriate clean-out cycle for this test program. An example of such a cycle is presented in Appendix C of the CRC E-60 emissions study report.[10] The US06 certification cycle, while not as extreme as the E-60 cycle, does provide similar conditions that are favorable for desulfurization (rich-lean switching at high temperatures) and is commonly used in the auto industry to desulfurize catalysts prior to emissions testing. Though the E-60 cycle would have been preferred, there were concerns that it would cause undesirable wear and tear on recruited vehicles. As a compromise, two back-to-back US06 cycles were selected for use in this test program.

The level of sulfur loading in a vehicle catalyst is primarily a function of the fuel sulfur level and catalyst temperature, the latter of which is a function of vehicle and catalyst design and driver behavior. A vehicle with a relatively high exhaust temperature at the catalyst location, and/or significant excess loading of certain platinum group metals (PGM) and other active

---

[i] Since sulfur's effect on (non-sulfur) emissions is understood to occur solely in the catalyst, a non-ethanol fuel was chosen as a base for practical reasons. Moreover, since sulfur was varied independently of other fuel properties, we do not anticipate that the relative difference in emission results would have been adversely affected by this design choice.

materials in the catalyst may be relatively insensitive to sulfur loading regardless of driver behavior. These effects of vehicle design were accounted for in the study by recruiting different makes and models covering a range of engine and vehicle sizes. We attempted to account for the effects of the owner's behavior by testing several samples of the same make and model.

3.2 Effect of Sulfur Level

Following vehicle operation that raises catalyst temperature and removes sulfur, accumulation immediately resumes as temperature declines under more typical (milder) driving conditions. This loading continues over a period of time with vehicle operation, and can be observed as an increase in emissions (sometimes referred to as "NOx creep"). To capture this effect, repeated emission tests were performed at two different fuel sulfur levels. Given that the exhaust stream contains different concentrations of sulfur according to the fuel sulfur level, the rate of reloading is expected to be lower for the lower sulfur fuel. Additionally, if an equilibrium sulfur loading (represented by a stable level of emissions) were to be reached after many miles of driving, this procedure could investigate whether this equilibrium loading is also lower for the lower sulfur fuel.

4. Test Vehicle Selection, Recruitment, and Delivery

4.1 Choice of Makes and Models

The test fleet was chosen to be representative of latest-technology light duty vehicles being sold at the time the program was launched. The study did not attempt to analyze or model details of aftertreatment design specific to each vehicle model such as catalyst position, precious metal types and quantities used, or related engine control strategies such as timing advance at cold start or fuel cut during deceleration. While these things undoubtedly influence the behavior of emissions and may interact with the fuel sulfur effects being investigated, including them in an analysis requires correctly assessing and parameterizing them for all vehicles in the study. Instead, this program's aim was to characterize overall effects of sulfur on emission inventories by observing the aggregate behavior of a representative fleet of vehicles.

The main and largest group of vehicles was intended to conform on average to the Tier 2 Bin 5 exhaust certification level and employ a variety of emission control technologies. These goals could be achieved by including a range of vehicle sizes, engine displacements, and manufacturers. A list of 19 high-sales-volume makes and models based on 2006-8 sales data and projections had been used for test fleet selection in the EPAct/V2/E-89 study that was launched shortly before this study.[11] Given that we would be targeting recruitment of vehicles 1-3 years old, this list seemed relevant; another benefit would be that the emission behavior of these same models would also be characterized in the other study's results. Grouping sales data by engine family allowed additional transparency and flexibility in choosing test vehicles that represent a wider group with identical powertrains without targeting one specific make and model. The resulting target list of 19 vehicle models for recruitment is shown in Table 4-1.

Table 4-1 Test Vehicles Targeted for Recruitment

Make	Brand	Model	Engine Size	Engine Family	Tier 2 Cert Bin
GM	Chevrolet	Cobalt	2.2L I4	8GMXV02.4025	5
GM	Chevrolet	Impala FFV	3.5L V6	8GMXV03.9052	5
GM	Saturn	Outlook	3.6L V6	8GMXT03.6151	5
GM	Chevrolet	Silverado FFV	5.3L V8	8GMXT05.3373	5
Toyota	Toyota	Corolla	1.8L I4	8TYXV01.8BEA	5
Toyota	Toyota	Camry	2.4L I4	8TYXV02.4BEA	5
Toyota	Toyota	Sienna	3.5L V6	8TYXT03.5BEM	5
Toyota	Toyota	Tundra	4.0L V6	8TYXT04.0AES	5
Ford	Ford	Focus	2.0L I4	8FMXV02.0VD4	4
Ford	Ford	Taurus	3.5L V6	8FMXV03.5VEP	5
Ford	Ford	Explorer	4.0L V6	8FMXT04.03DB	4
Ford	Ford	F150 FFV	5.4L V8	8FMXT05.44HF	8
Chrysler	Dodge	Caliber	2.4L I4	8CRXB02.4MEO	5
Chrysler	Dodge	Caravan FFV	3.3L V6	8CRXT03.3NEP	8
Chrysler	Jeep	Liberty	3.7L V6	8CRXT03.7NE0	5
Honda	Honda	Civic	1.8L I4	8HNXV01.8LKR	5
Honda	Honda	Accord	2.4L I4	8HNXV02.4TKR	5
Honda	Honda	Odyssey	3.5L V6	8HNXT03.54KR	5
Nissan	Nissan	Altima	2.5L I4	8NSXV02.5G5A	5

Following the main group of vehicles representing the Tier 2 in-use fleet, a smaller group of vehicles meeting lower emission standards was tested using the same fuels and procedures as applicable. The purpose of this follow-on work was to evaluate the sulfur sensitivity of the newest, lowest-emission technologies that could represent a portion of the future in-use fleet. A detailed list of these vehicles is shown in Table 7-12.

4.2 Vehicle Recruitment Criteria

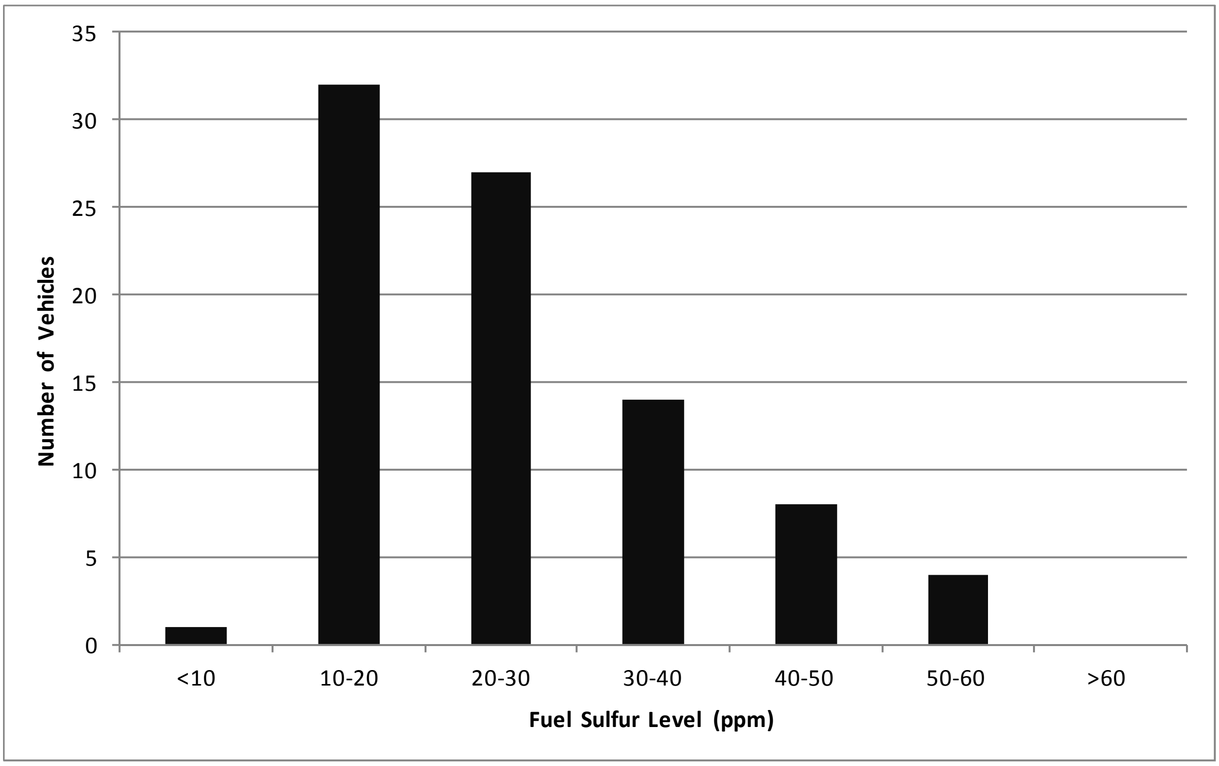
The vehicles recruited from private owners were selected to have odometer mileage between 12,000 and 40,000 miles and model year less than three years old. The recruitment process used the State of Michigan vehicle owner registration database to identify candidate vehicle owners within 50 miles of NVFEL. Solicitation letters were sent to these owners to find participants willing to allow their vehicles to be held and tested for up to six weeks, in exchange for cash and/or a loaner vehicle. Once an owner agreed to participate, the vehicle was scheduled for testing.

4.3 Initial Checks and Test Vehicle Delivery

Recruited vehicles were driven to the test facility either by their owners or NVFEL personnel sent to retrieve them. NVFEL drivers were instructed to avoid hard accelerations and high speeds in an effort to preserve the existing state of the emission controls. Once in custody, vehicles were inspected for damage or malfunction. The condition of the MIL (malfunction indicator, or “check engine”, lamp) was noted, and if illuminated, the vehicle was rejected from the study. Minor exhaust leaks were repaired as needed; however, damage to the catalyst, muffler, or sections of exhaust pipe disqualified the vehicle from the study. Gas caps and other fuel systems components were visually inspected and repaired as needed. All repairs were documented.

Once the inclusion criteria were met, the existing fuel was drained and a sample was analyzed for its sulfur content to provide information about whether any vehicles that had been operating on unusually high or low sulfur levels (no vehicles were removed from the study based on this criterion). Federal fuel sulfur standards applicable in the recruiting area specified a 30 ppm annual refinery average with an 80 ppm per-gallon cap. Figure 4-1 shows the distribution of sulfur levels found in the fuel tanks of test vehicles as received. The mean of the fuel samples for which data were available (86 of 93 vehicles) was 25.4 ppm with a standard deviation of 11.2 ppm. At the time of the study, fuel in the recruiting area typically contained 10% ethanol.

Figure 4-1 Fuel Sulfur Levels Recruited Vehicles' Fuel Tanks



5. Test Fuel Specs and Procurement

A bulk quantity of non-ethanol, low-sulfur, certification-type gasoline was purchased from a commercial fuel blender (Haltermann Solutions, Houston, TX) and divided between two underground storage tanks prior to start of the program. One tank was spiked with a small amount (on the order of a liter of volume) of dibutyl disulfide to increase the sulfur content from the as-received 5 ppm to 28 ppm. Due to the small volume of sulfur agent added, changes in other fuel properties were negligible. Other properties of the test fuels are listed in Table 5-1, and fall in the range of what is encountered in typical commercial gasoline in the vehicle recruiting area. Relative to national average 2010 summer conventional gasoline (much of which contained ethanol) as reported in the Auto Alliance U.S fuel survey database, these test fuels are somewhat higher in aromatics (by about 6 vol. %), somewhat lower in olefins (by about 6 vol. %), somewhat higher in T50 (by about 20°F), and slightly lower in T90 (by about 3°F).

Table 5-1 Test Fuel Properties

Fuel Property	ASTM Method	Low S Test Fuel	High S Test Fuel[†]
Sulfur	D2622	5 ppm	28 ppm
Benzene	D5769	0.34 Vol. %	0.34 Vol. %
Total Aromatics	D5769	31.2 Vol. %	31.2 Vol. %
Olefins	D1319	0.5 Vol. %	0.5 Vol. %
Saturates	D1319	68.3 Vol. %	68.3 Vol. %
Oxygenates	D5599	0.0 Vol. %	0.0 Vol. %
T50	D86	221°F	221°F
T90	D86	317°F	317°F
RVP	D5191	9.0 psi	9.0 psi

[†] Sulfur content was confirmed for the higher-sulfur test fuel, while other properties were assumed to be the same as the base fuel given the small amount of dopant added.

6. Test Procedures

6.1 Initial Fuel Exchange and Vehicle Prep

Fuel drains were generally done via the fuel rail by energizing the vehicle's pump. Next, the key was turned to the "on" position, without starting the vehicle, for 30 seconds before keying off. This allowed the vehicle to register the new fuel level to trigger any fast fuel trim learning procedures the vehicle may have.[ii] Next, the vehicle was filled to 50 percent capacity with test fuel. This fuel drain and fill process was then repeated a second time to ensure that the fuel system had been flushed fully prior to testing.

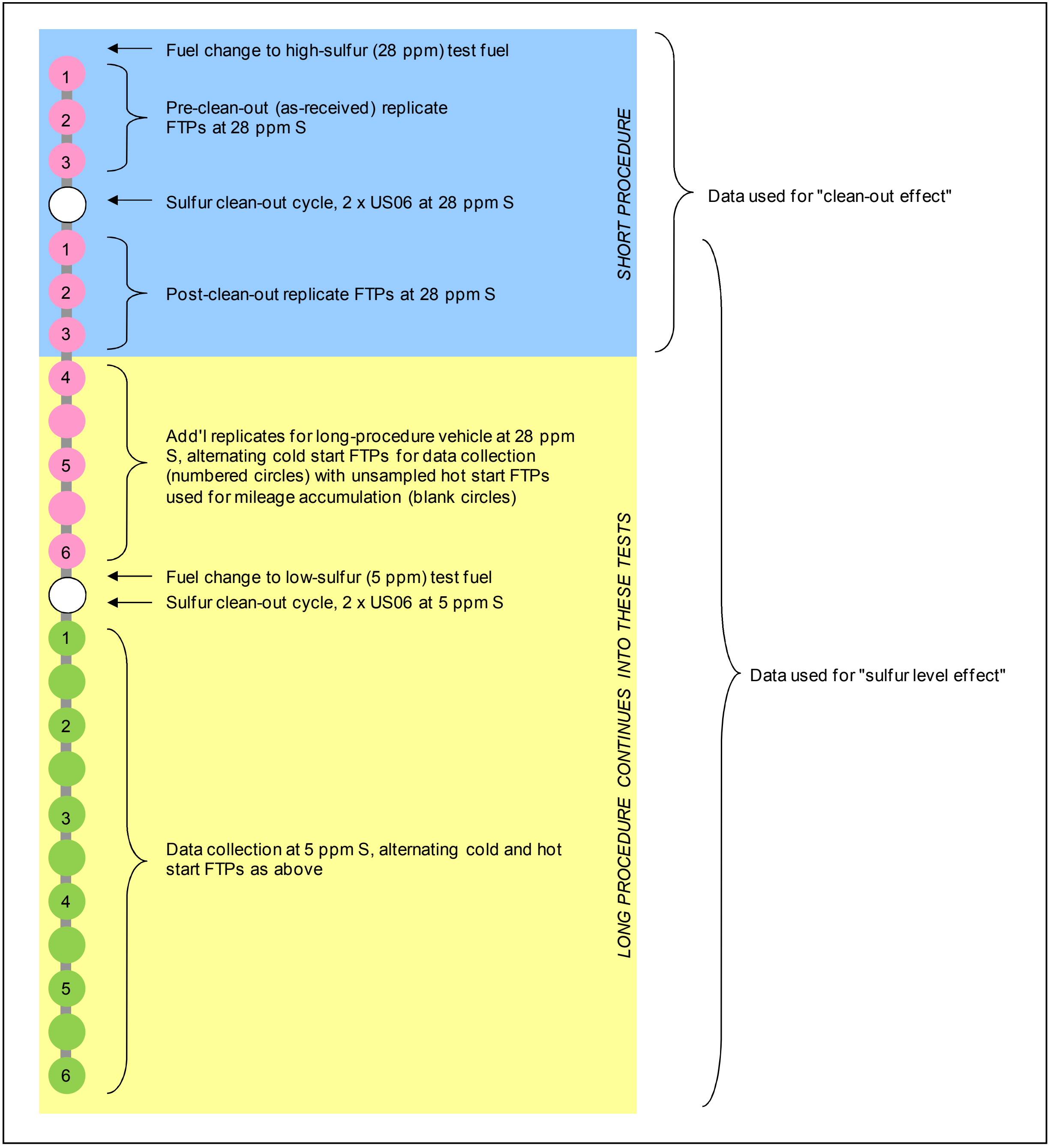
Following the fuel change the vehicle was moved to the dynamometer to conduct an LA4 preconditioning drive cycle.[iii] At the completion of this prep cycle, the vehicle was idled in park for two minutes before engine shutdown, a step intended to ensure fuel trim data was updated for idle operation. The vehicle was then moved to a storage (“soak”) area using an electric towing device for a 12-36 hour holding period at approximately 70°F prior to testing.

6.2 Test Procedure Description

Vehicles recruited for testing in this program fell into one of two test groups: “Long Test Procedure” vehicles or “Short Test Procedure” vehicles. Among the main group of Tier 2 vehicles, for a given vehicle model class (or family ID), one of the five vehicles was tested under the Long procedure while the other four were tested under the Short procedure. The choice of vehicle to undergo the Long procedure was made at random from willing participants in a given vehicle class. The original Long and Short procedures used at the beginning of the program are shown in Figure 6-1 and discussed in greater detail below. They are identical in structure for the first six emission tests. The follow-on group of “Tier 3-like” vehicles followed the modified Long procedure as described below with the exception of the manufacturer-owned vehicle. It was not expected to have a use history relevant to the in-use loading question, and therefore underwent a somewhat abbreviated procedure focused only on the effect of sulfur level.

[ii] Fuel trim refers to the ability of the engine control system to adapt to a change in fuel properties, which may affect vehicle or emissions performance. Some vehicles may accelerate this “learning” process if a fuel change is detected.
[iii] The LA4 drive cycle covers the first 7.5 miles of the FTP test cycle, consisting of typical urban driving behavior. For more details on these test cycles see 40 CFR Part 86.

Figure 6-1 Original Short (S) and Long (L) Procedures



All recruited vehicles were tested over three cold-start FTPs with high sulfur fuel to assess "as received" emissions levels prior to performing a sulfur clean-out cycle (two back-to-back US06 cycles) to remove sulfur accumulated in-use from the catalyst.[iv] Emissions were then sampled again over three cold-start FTPs on the same high sulfur test fuel to assess the effect of the sulfur clean-out cycle on emissions. This data was used to determine the "clean-out" effect and is referred to as the clean-out dataset in the analysis section of this report. Testing of a vehicle undergoing only the Short procedure was complete at this point and the vehicle was returned to its owner.

Vehicles undergoing the Long procedure continued testing for an additional length of time to determine the effects of both high and low sulfur fuel following mileage accumulation. Beginning with the fourth post-cleanout FTP test on high sulfur fuel, the test sequence began to alternate between sampled, cold-start FTPs and un-sampled hot-start FTPs (immediately following the cold-start test) in order to accumulate additional miles between emission tests. Following the sixth sampled FTP, the fuel was changed to low sulfur test fuel and a clean-out cycle was run to remove accumulated sulfur from the catalyst and establish a new "cleaned-out" emissions baseline on low sulfur test fuel. Six additional FTPs (with un-sampled FTPs in-between) were conducted on low sulfur fuel. Together with the six high sulfur fuel tests, this data is used to determine the "sulfur level effect," which is the assessment of high and low sulfur fuel use starting from a "clean" catalyst.

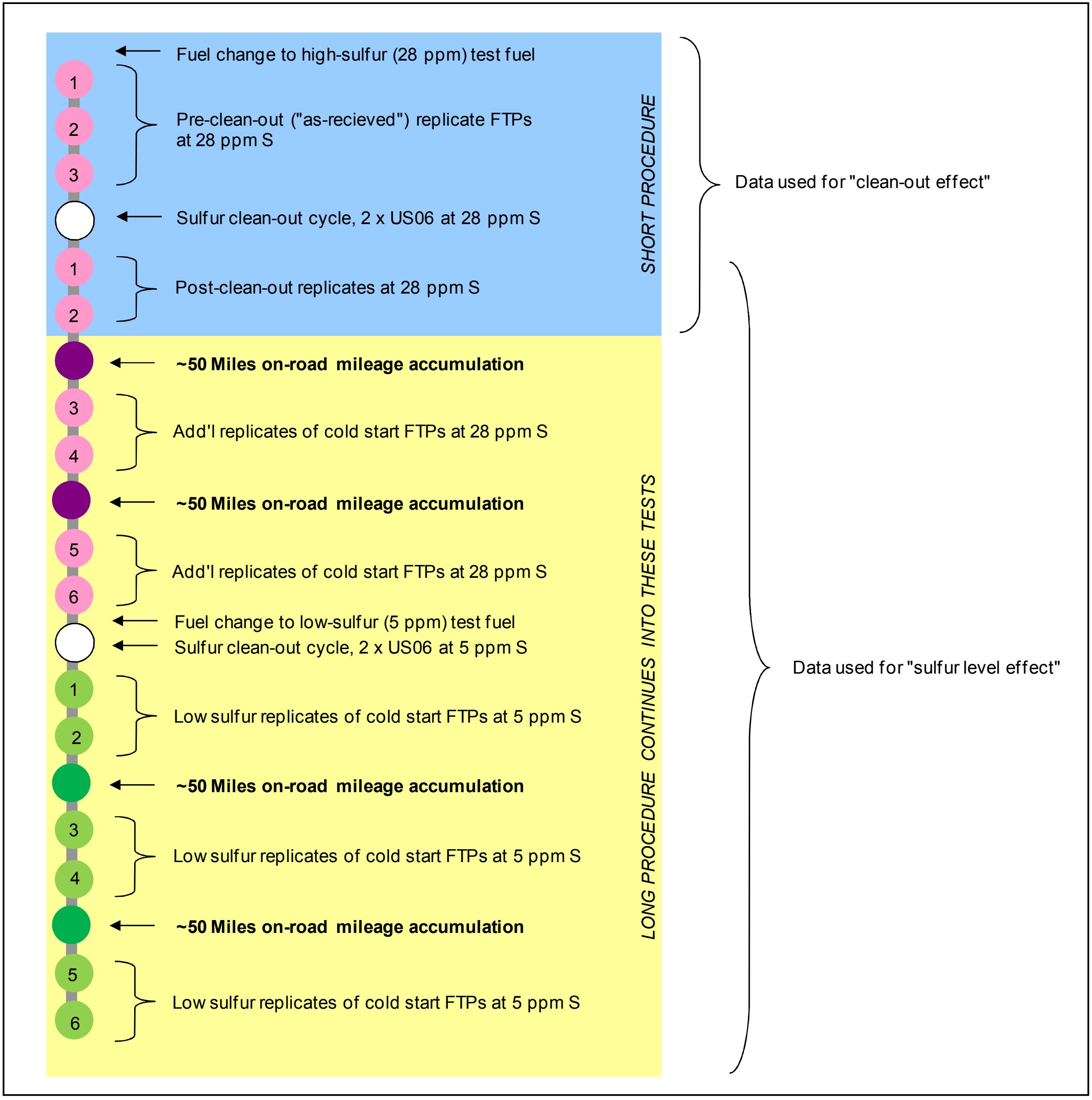
Under this original protocol, vehicles tested on the Long procedure were run for approximately 100 miles of operation on each fuel following a cleanout. However, after an interim data review mid-way through the program, we were concerned that 100 miles may not be sufficient to cover loading levels associated with a wide variety of in-use driving. Therefore, we modified the Long procedure to incorporate two 50-mile on-road mileage accumulation intervals during the sequence of emission tests on each fuel. This modified Long procedure, the L/M procedure, allowed us to accumulate an additional 50 miles on each fuel without occupying

---

[iv] A cold-start FTP refers to the FTP drive cycle being performed after the vehicle has been sitting with no engine starts for between 12-36 hours at approximately 70°F. This is useful because emissions are generally highest in the first few minutes of cold engine operation. A hot-start FTP refers to a test driven within a short time after a previous test, such that the engine and aftertreatment are still well above room temperature. Hot-start tests were only used for mileage accumulation and did not have emissions sampled.

additional test cell time. A flowchart of the modified Long procedure (L/M) is shown in Figure 6-2.

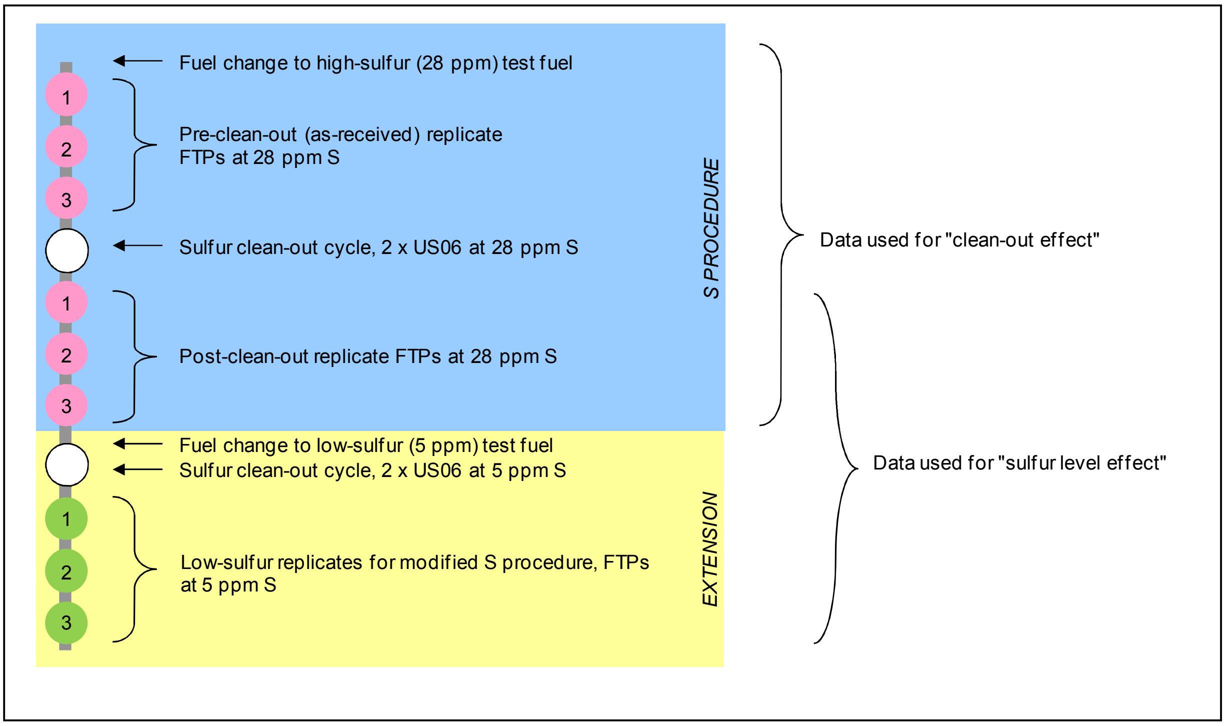
Figure 6-2 Modified Long (L/M) Procedure



On-road mileage accumulation for the L/M procedure was performed over a pre-defined driving route selected for its resemblance to the FTP in terms of speed and load distribution. Refer to Appendix A for more details on this drive route.

Following the mid-point data review the Short procedure was also modified to include a second clean-out and subsequent set of emission tests performed on low sulfur fuel. This change was made based on the observation that sulfur level had an effect on emissions immediately after the cleanout for some vehicles.[v] Figure 6-3 shows the modified Short procedure containing two fuel sulfur levels, which was implemented beginning with vehicle Family ID N513.

Figure 6-3 Modified Short Procedure



7. Data Analysis and Results

The data generated by this test program will be discussed as three distinct but overlapping datasets: “clean-out at 28 ppm” data, “clean-out at 5 ppm” data, and “sulfur level” data. The “clean-out at 28 ppm” data consists of the measurements from original and modified ‘short’ procedures; “clean-out at 5 ppm” data consists of a subset of original and modified ‘long’ procedures and modified ‘short procedures; “sulfur level” data consists of the complete data

---

[v] This information became the basis for the “clean-out at 5ppm” conclusions.

from original and modified 'long' procedures and modified 'short' procedures. All vehicles included in the "sulfur level" dataset had a subset of their tests included in the "clean-out" datasets, by design.

Pollutants included in the analysis were total hydrocarbons (THC) as reported by the FID analyzer, carbon monoxide (CO), oxides of nitrogen (NO_x), methane (CH_4), as well as particulate matter (PM) mass. The following calculated emissions were also included: non-methane hydrocarbons (NMHC, defined as THC minus methane) and oxides of nitrogen plus non-methane organic gases (NO_x + NMOG), a measurement relevant in regulatory contexts.[vi] Each bag, 'Bag 1 minus Bag 3', and the composites from the FTP test cycle were analyzed separately. The first bag captures the initial "cold start", meaning the emissions produced when the vehicle is started after cooling to room temperature overnight. The "hot stabilized" operation is captured in the second bag (a portion of the test that begins after approximately 8 minutes of driving), and the emissions from "hot-restart" are measured in the third bag (a repeat of the cold start drive cycle, but after the vehicle has been turned off for only 10 minutes). The 'Bag 1 minus Bag 3' emission value was used to isolate cold-start emissions for each FTP test.

The statistical methodologies described in the following section were applied consistently in the analysis of all pollutants and all bags. However, the analysis of nitrogen oxides (NO_x) from Bag 2 is presented in greater detail to assist the reader in understanding the analytical approaches and to illustrate the statistical methods being used. The final model results from all other pollutants are also presented in this report.

The goals of the statistical analyses of the "clean-out" and "sulfur level" data were to assess the reversible sulfur loading in the catalysts of the in-use fleet, to examine the differences in the effectiveness of the clean-out cycle by comparing emissions immediately following the clean-out

---

[vi] NMOG refers to non-methane organic gases. The flame ionization analyzer used to produce the THC result produces an inaccurate measurement of oxygenated emission species, which can be corrected using speciated results for those compounds along with their analyzer response factors. The ratio of NMOG/NMHC can be predicted with good accuracy using the fuel ethanol content (zero in this case), which was the approach used here to generate the NMOG results. This method is described in Sluder S.C, West B.H., "NMOG Emissions Characterizations and Estimation for Vehicles Using Ethanol-Blended Fuels". Oak Ridge National Laboratories, 2011. Publication number ORNL/TM-2011/461.

procedure at two sulfur levels, and to characterize the effects of fuel sulfur level on emissions as mileage was accumulated.

7.1. Data Preparation

Prior to proceeding with the statistical analyses, the issues associated with the dataset, namely, very low emission measurements and outliers were examined. The following sections describe how these issues were addressed.

7.1.1. Imputation of Measurements with Low Concentration

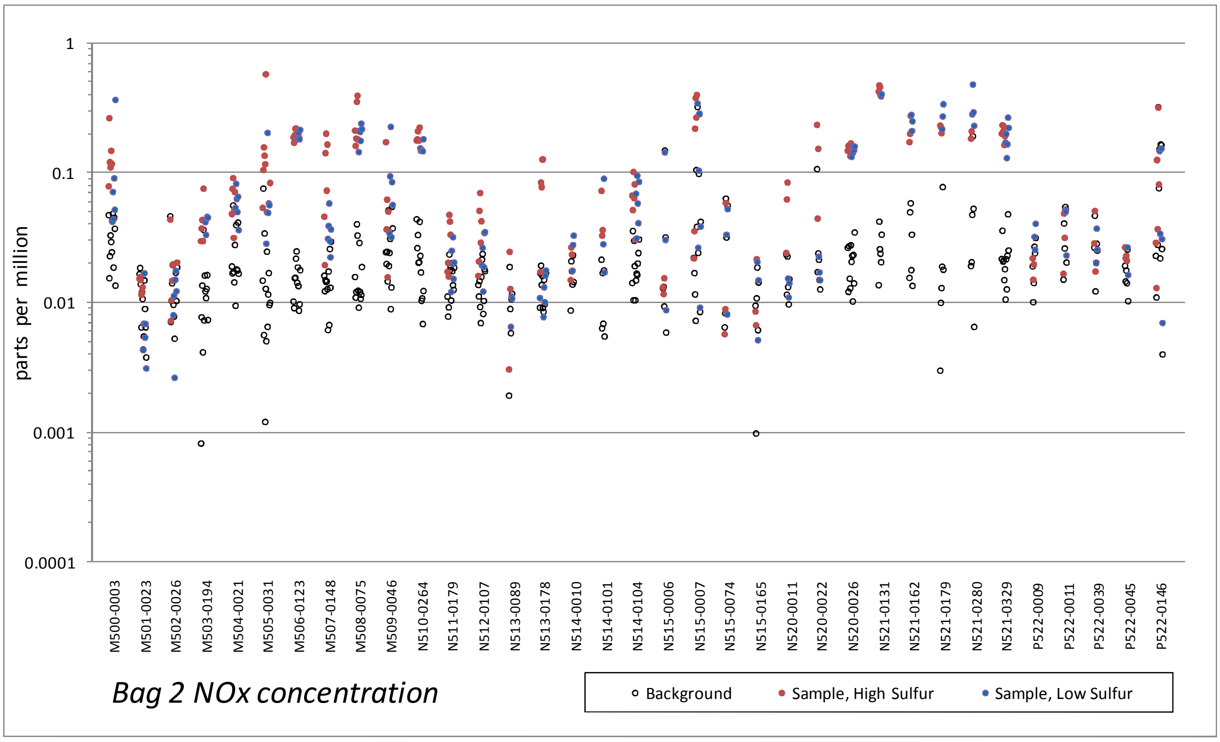
The graphical examination of both the "clean-out" and the "sulfur level" dataset revealed the presence of very low emission measurements from some pollutants and bags including NO_x Bag 2. Since uncertainty associated with these low measurements could potentially affect the outcome of the analysis, it was important to understand the measurement process and evaluate the impact of any uncertainty.

During emission testing, the vehicle exhaust stream is collected and diluted with background air to avoid water condensation and other stability problems. A small sample of this mixture flows into a collection bag for analysis after the test. The concentration of emission species in the bag is determined by flowing the contents through a properly-calibrated gas analyzer and reading the result. This method provides a time-weighted result via physical integration of the emission stream produced over the course of a transient driving cycle.

The measurement process has uncertainty associated with it as a result of the physics of mixing and sampling from a gas stream as well as noise in analyzer components such as optoelectronic detectors and signal amplifiers. This means that repeated measurements taken under identical process conditions will produce a range of results, their average being the true (intended) response of the instrument and the range around it representing the measurement variability.

For the analyzers used in this program, the size of the measurement error is expected to increase relative to the measured value as the concentration decreases. Moreover, the dilute bag method used here requires measurement of concentrations in both the sample and background bags, followed by a subtraction between the two, such that the net result contains variability from two measurements. To assess whether these issues may affect this dataset, we examined plots of the measured concentrations for each test by vehicle by pollutant and bag. Figure 7-1 shows the Bag 2 NO_x dataset for the vehicles providing the "sulfur level" data, which contains a number of very low values, as well as tests where sample and background are of similar magnitude (the vehicle codes refer to the Family IDs listed in Table 7-7; additional plots are available in Appendix B). Given these findings, we performed sensitivity analyses to evaluate the impact of these low emission measurements on the study results (presented in Section 7.3.4).

Figure 7-1 Bag 2 NO_x Concentration Measurements by Vehicle



When a dilute emission measurement is lower than the measured background level, the net result is reported as zero (this calculation is performed on a test-by-test basis). However, since it is unlikely that tailpipe emissions are truly zero during a test, it was assumed that a zero is a result of the actual emission level being smaller than the sum of the measurement errors occurring on the sample and background measurements. The emission level was thus considered to be below the limit of quantification (LOQ), a level below which we are not confident in the accuracy of quantitative value.

In this situation, the data point can be left as zeroes, deleted, or replaced with an imputed value. However, because it was necessary to apply log-transformation, as described in more detail in Section 7.2, the zero values could not be left as is in the data. Table 7-1 summarizes the number of measurements with zero values and the percentages in parenthesis for each dataset. Given that observations below the LOQ appear to be randomly distributed across sulfur levels and vehicles, and since excluding such observations would result in reduced sample size, less statistical power, and larger standard errors,[12] they were imputed in the analysis.

Since an imputation method involving each vehicle's own longitudinal data would be superior to methods using no information about the vehicle,[13] a commonly-used single-imputation method, using half the minimum of a valid measurement from a given mileage bin for the vehicle with zero values, was performed. This imputation method recognized the fact that emission measurements below the limit of quantification must be smaller than any quantified value.

Although using vehicle-specific imputation minimizes the likelihood of artificially reducing the natural variance of the data, there is a potential to inflate the reliability estimates as the number of imputed values increase. However, since the number of measurements with imputed values are less than 20 percent (Table 7-1), we can expect good estimates of the reliability of measurements.[14] Nonetheless, it is important to determine the effect of these imputed values on the resulting test statistics and corresponding conclusions. Thus, the results from the statistical analysis with and without the imputed values were compared once the model was finalized to assess the potential for introducing bias. This is discussed further in Section 7.3.4.

Table 7-1 Number of Measurements With Zero Values

	Clean-out at 28 ppm data (N = 541)					
	NO$_x$	THC	CO	NMHC	CH4	PM
Bag 1	0	1 (0.2%)	0	1 (0.2%)	1 (0.2%)	1 (0.2%)
Bag 2	32 (5.9%)	11 (2.0 %)	33 (6.1%)	48 (8.9%)	4 (0.7%)	2 (0.4%)
Bag 3	0	1 (0.2%)	21 (3.9%)	38 (7.0%)	1 (0.2%)	1 (0.2%)
FTP Composite	0	1 (0.2%)	0	1 (0.2%)	1 (0.2%)	2 (0.4%)
Bag 1 – Bag 3	0	1 (0.2%)	0	1 (0.2%)	1 (0.2%)	1 (0.2%)
	Clean-out at 5 ppm data (N = 183)					
	NO$_x$	THC	CO	NMHC	CH4	PM
Bag 1	0	0	0	0	0	0
Bag 2	15 (8.2%)	7 (3.8%)	4 (2.2%)	5 (3.8%)	4 (2.2%)	0
Bag 3	2 (1.1%)	0	2 (1.1%)	17 (9.3%)	0	0
FTP Composite	0	0	0	0	0	0
Bag 1 – Bag 3	0	0	0	15 (8.2%)	0	0
	Sulfur level data (N = 322)†					
	NO$_x$	THC	CO	NMHC	CH4	PM
Bag 1	0	0	0	0	0	0
Bag 2	21 (6.5%)	14 (4.3%)	10 (3.1%)	33 (10.2%)	5 (1.5%)	2 (0.9%)
Bag 3	3 (0.9%)	0	4 (1.2%)	28 (8.6%)	0	0
FTP Composite	0	0	0	0	0	0
Bag 1 – Bag 3	7 (2.2%)	0	1 (0.3%)	0	0	15 (6.5%)

† The sulfur level data for PM had 232 measurements.

7.1.2. Detection of Outliers

Next, before proceeding to the full analysis, preliminary models were fit to detect outliers. The residual plots were visually inspected for outlying observations and the outliers were identified using the screening criterion value of ±3.5 for the externally studentized residuals. Generally, one can expect about 95% of the externally studentized residuals to be within ±3.5 standard deviations, and applying the ±3.5 criterion to detect potential outliers has been widely used in statistics. When the outlying observation represented an actual measurement, it was examined to assess its validity. Since none of the outliers representing actual measurements showed a clear indication of measurement error, it was assumed that the outlying observations were real and thus they were included in the dataset for the analysis. However, there were instances where a very low imputed value was identified as an outlier. In such instances, the imputed values were removed from the dataset. Table 7-2 summarizes the number of outliers as well as the number of imputed measurements that were removed in parenthesis.

Table 7-2 Number of Outliers

	Clean-out at 28 ppm data (N = 541)					
	NO_x	**THC**	**CO**	**NMHC**	**CH4**	**PM**
Bag 1	2 (0)	4 (0)	2 (0)	5 (0)	2 (1)	2 (0)
Bag 2	3 (1)	7 (4)	5 (2)	3 (0)	6 (1)	5 (0)
Bag 3	3 (0)	2 (0)	5 (2)	5 (4)	1 (0)	2 (0)
FTP Composite	3 (0)	3 (0)	2 (0)	3 (0)	0 (0)	1 (0)
Bag 1 – Bag 3	13 (4)	5 (0)	1 (0)	6 (0)	5 (0)	7 (3)
	Clean-out at 5 ppm data (N = 183)					
	NO_x	**THC**	**CO**	**NMHC**	**CH4**	**PM**
Bag 1	0 (0)	1 (0)	2 (0)	1 (0)	1 (0)	1 (0)
Bag 2	1 (0)	0 (0)	1 (1)	1 (0)	2 (0)	1 (0)
Bag 3	3 (0)	2 (0)	1 (0)	2 (1)	0 (0)	1 (0)
FTP Composite	2 (0)	1 (0)	0 (0)	1 (0)	0 (0)	2 (0)
Bag 1 – Bag 3	3 (2)	1 (0)	1 (1)	2 (0)	3 (0)	3 (0)
	Sulfur level data (N = 322)†					
	NO_x	**THC**	**CO**	**NMHC**	**CH4**	**PM**
Bag 1	0 (0)	4 (0)	4 (0)	3 (0)	3 (0)	2 (0)
Bag 2	0 (0)	1 (1)	4 (1)	3 (0)	3 (0)	1 (0)
Bag 3	1 (0)	4 (0)	3 (0)	2 (0)	0 (0)	1 (0)
FTP Composite	4 (0)	3 (0)	3 (0)	3 (0)	1 (0)	2 (0)
Bag 1 – Bag 3	2 (0)	2 (0)	6 (1)	2 (0)	1 (0)	4 (0)

† The sulfur level data for PM had 232 measurements.

7.2. Modeling Methodology

The following section describes the statistical approaches and the model fitting methodologies applied in the analysis of all three datasets – "clean-out at 28 ppm", "clean-out at 5 ppm", and "sulfur level" data.

First, the emission measurements were log-transformed. In the current study, the distributions of emissions exhibited positive skewness (log-normal), and thus, transforming emission measurements by the natural logarithm was necessary to stabilize the variance, to obtain a linear relationship between the mean of the dependent variable and the fixed and random effects, and to normalize the distribution of the residual. The log-transformation of emission measurements has been well-established in previous studies analyzing vehicle emissions data.[15,16,17]

Both the "clean-out" and "sulfur level" data are a classic example of a "repeated measures data" where multiple measurements were taken from a single vehicle at different accumulated mileages. The conventional methods for analyzing "repeated measures data" are the univariate and multivariate analysis of variance (discussed further in Appendix D). However, the linear mixed model was selected for the analyses of both "clean-out" and "sulfur level" data for the following reasons: The mixed model approach uses generalized least squares to estimate the fixed effects, which is considered superior to the ordinary least squares used by the univariate and multivariate procedure.[17] It is a more robust and flexible procedure in modeling the covariance structures for repeated measurements data and better accounts for within-vehicle mileage-dependent correlations.[15,16] In addition, the mixed model is capable of including vehicles with missing data and handling irregularly spaced measurements.

The MIXED procedure in the SAS 9.2 software package was used to fit the model. The mixed model is represented in Equation 7-1 as:

$$\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_iu_i} + \boldsymbol{\varepsilon_i} \qquad \textbf{Equation 7-1}$$

where β and u_i are fixed and random effects parameters, respectively, and ε_i is the random residuals. The mixed model accounts for correlation in the data through the inclusion of random effects and modeling of the covariance structure. β represents parameters that are the same for all vehicles and u_i represents parameters that are allowed to vary over vehicles reflecting the natural heterogeneity in the vehicle fleet. In other words, the model considers the differences in the effect of sulfur level on emission for each vehicle. The distributional assumptions for the mixed model are: u_i is normal with mean 0 and variance G_i; ε_i is normal with mean 0 and variance R_i; the random components u_i and ε_i are independent.

In developing the mixed model, a top-down model fitting strategy, similar to previously established methods,[18,19] was used. The first step was to start with a saturated model, including all candidate fixed effects to ensure unbiasedness of the fixed effect estimates. Next, we selected a model with the most optimal covariance structure, which specifies the variation between vehicles as well as the covariation between emission measurements at different accumulated mileages on the same vehicle. Finally, the fixed effects portion of the model was reduced to fit the final model.

7.3. Statistical Analysis and Results

The statistical analyses were performed on all three datasets ("clean-out at 28 ppm", "clean-out at 5 ppm", and "sulfur level" data) for the conventional Tier 2 vehicles. However, for the "Tier 3-like" vehicles designed to supplement the Tier 2 vehicles, only the "sulfur level" data was analyzed, considering the small sample size of the "Tier 3-like" vehicles.

7.3.1. Effect of Clean-Out at 28 ppm

The main objective of analyzing the "clean-out at 28 ppm" data was to assess the in-use reversible sulfur loading in the catalysts at the fuel sulfur level of 28 ppm. In addition, the results from the "clean-out" data were used to supplement the "sulfur level" data in estimating the differences of in-use sulfur loading between high and low fuel sulfur levels, as described in Section 7.3.3.

The "clean-out at 28 ppm" data consists of the as-received emission measurements (pre-cleanout) and the measurements after the back-to-back US06 cycles (post-cleanout) at the fuel sulfur level of 28 ppm from original and modified 'short' procedures. The change in emissions between pre- and post-cleanout represents the amount of reversible in-use sulfur loading at 28 ppm. The "clean-out at 28 ppm" data at the time of the analysis comprised 19 vehicle families, for a total of 93 unique vehicles. Vehicles from the same engine family had the same engine size, vehicle configuration, and weight. The average starting odometer reading of the 93 unique vehicles was 32,049 ± 6,136 miles. Additional details of this test fleet are shown in Table 7-3.

A total of 541 measurements were taken – 275 pre-cleanout and 266 post-cleanout measurements.

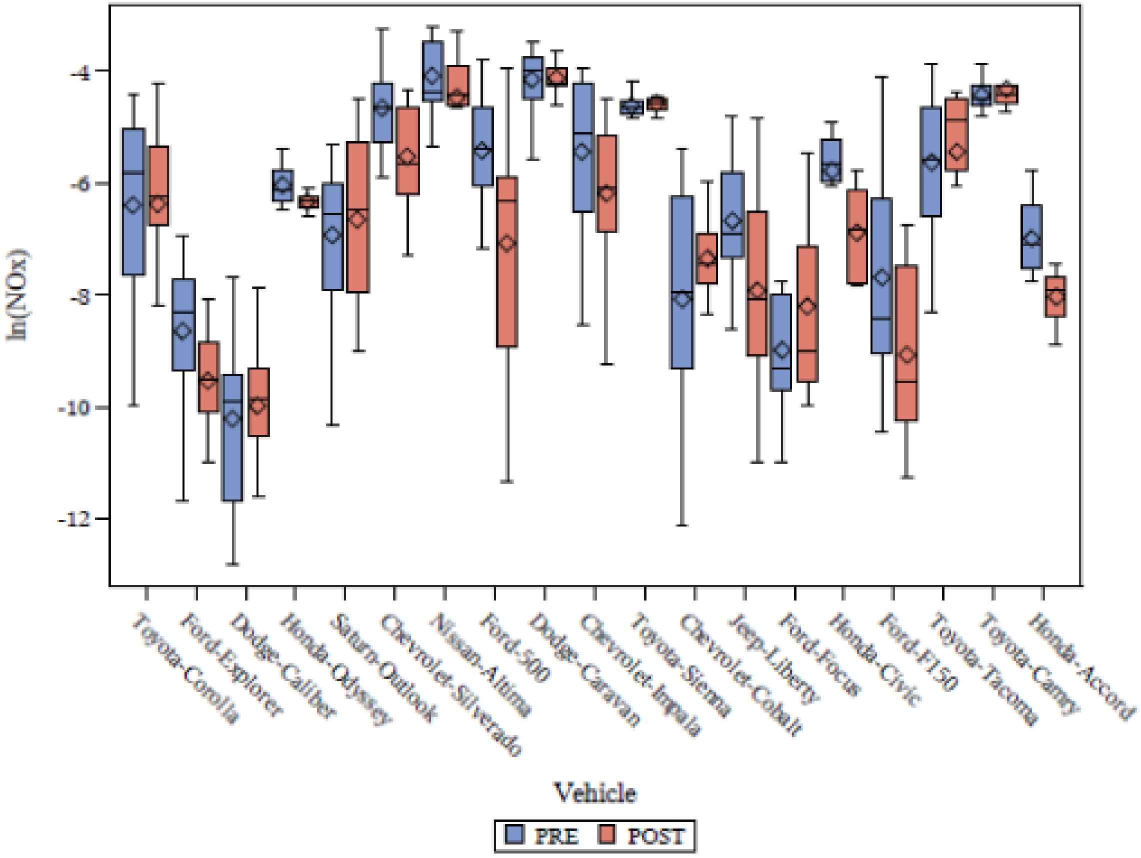
Table 7-3 Description of Vehicles in the "Clean-Out at 28 ppm Data"

Vehicle Family ID	Make	Model	Model Year	Tier 2 Cert Bin	Number of Vehicles	Average Starting Odometer
M500	Toyota	Corolla	2007	5	5	32,578
M501	Ford	Explorer	2007	4	5	33,605
M502	Dodge	Caliber	2007	5	5	31,184
M503	Honda	Odyssey	2007	5	5	35,954
M504	Saturn	Outlook	2007	5	5	35,762
M505	Chevrolet	Silverado	2007	5	5	37,401
M506	Nissan	Altima	2007	5	5	32,283
M507	Ford	Taurus	2007	5	5	29,442
M508	Dodge	Caravan	2007	8	5	34,371
M509	Chevrolet	Impala	2007	5	5	26,183
N510	Toyota	Sienna	2007	5	5	30,996
N511	Chevrolet	Cobalt	2008	5	6	29,023
N512	Jeep	Liberty	2008	5	5	27,530
N513	Ford	Focus	2008	4	5	26,843
N514	Honda	Civic	2008	5	4	37,988
N515	Ford	F150	2008	8	5	31,719
N520	Toyota	Tacoma	2009	5	3	28,964
N521	Toyota	Camry	2008	3	5	28,506
P522	Honda	Accord	2008	3	5	29,601

The box-plot of the "clean-out at 28 ppm" data (Figure 7-2) for log-transformed NO_x bag 2 shows the spread of the data for each vehicle family by clean-out status. The diamond and the line inside the box represent the mean and the median, respectively. The box represents the interquartile range between 25th and 75th percentile and the error bars show the full data range.

The box-plot suggests that the variance in the pre-cleanout group tends to be larger than the variance observed in the post-cleanout group. Generally, there is a tendency for the pre-cleanout measurements, representing the as-received sulfur level, to have higher NO_x than the post-cleanout measurements. In addition, the plot illustrates that the emission profiles are substantially different between vehicle families. Some of the vehicle families also show the presence of within-family variability, as demonstrated by each error bar, reflecting the influence of individual driving patterns on in-use sulfur loading. These findings from the graphical examination of the data assisted in selecting the appropriate statistical model for the analysis.

Figure 7-2 Box-Plot of Vehicle Families by Pre- and Post-Cleanout at 28ppm (NO_x Bag 2)



The dependent variable (Y_i) of the mixed model (Equation 7-1) was the natural logarithm of emissions. The fixed effect (X_i) included in the model was the cleanout status (pre- and post-cleanout), vehicle type (car vs. trucks), and the interaction term. The random effects (Z_i) were each vehicle family in the study assuming that vehicles from the same vehicle family have similar emission profiles. For the degrees of freedom, the Satterthwaite approximation was used for tests of fixed effects.[20]

First, a test for the significance of the between-family variation was performed for inclusion of the random intercept for each family using the likelihood ratio test. For all pollutants and bags, the result was significant and the random intercept for each family was included for all subsequent models. The significance of between family-variation was observed graphically in Figure 7-2 and is also supported by recent automotive emissions studies.[21,22]

The covariance structure was estimated using restricted, or residual, maximum likelihood (REML)[23] and selected by comparing the Schwarz Bayesian Criterion (BIC) values for various potential covariance structures. BIC, considered a consistent criterion, accounts for the total number of observations in the dataset and has generally been known to perform relatively better for small size datasets.[24,25,26]

In modeling the covariance structures, the variance associated with the pre-cleanout was permitted to differ from the variance associated with the post-cleanout. As an initial step, a saturated model with fixed effects was attempted to be fit with "unstructured" (UN) covariance, which makes no assumptions regarding equal variances or correlations. One advantage of starting with the unstructured covariance is the fact that it allows selection of the most optimal covariance structure based on the data by visually examining the matrix for patterns. However, the model failed to converge and it was not possible to fit an unstructured covariance matrix. It may be explained by the fact that UN structure requires estimation of a large number of variance and covariance parameters, resulting in computational problems.

Since the "clean-out" data involved three repeat tests of FTP for both pre- and post-cleanout without mileage accumulation, the covariance matrix of compound symmetry (CS), in Equation 7-2, was selected, which assumes that the measurements from the same vehicle have the homogeneous variance and the correlation among measurements is constant regardless of how far apart the measurements are.

$$\boldsymbol{R_i = Var(\epsilon_i)} = \begin{bmatrix} \boldsymbol{\sigma^2 + \sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} \\ \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma^2 + \sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} \\ \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma^2 + \sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} \\ \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma_1{}^2} & \boldsymbol{\sigma^2 + \sigma_1{}^2} \end{bmatrix} \qquad \textbf{Equation 7-2}$$

The covariance structure was modeled using a combination of compound symmetry structure within families and a random effect between families. This combination structure specifies an inter-vehicle random effect of differences between families, and a correlation structure within families that are the same between emission measurements. It implies that the only aspect of the covariance between repeated measures is due to the vehicle contribution.

Following the selection of structures for the random effects and the covariance structure for the residuals, the fixed effects in the model were tested using the approximate F-test with the Satterthwaite degrees of freedom. A non-significant effect was removed from the model and the reduced model was re-fit and re-tested until all fixed effects in the model were significant. The significance level of 10 percent ($\alpha = 0.1$) was used to test the null hypothesis. Statistical hierarchy was kept in removing insignificant fixed parameters – first order terms were included when the second order term was significant.

We then performed the likelihood ratio test between the reference and the nested models using Maximum Likelihood (ML)[23,27] as the estimation method to examine whether the model can be reduced further without influencing the model fit. To perform the likelihood ratio test, the -2 Res Log Likelihood Scores from two separate models were considered and the chi square test statistic was computed by subtracting the -2 Res Log Likelihood from the model with more parameters from the model with fewer. The Chi-Square test statistic had the degrees of freedom equal to the difference of the number of parameters. When this test statistic was significant, the model with the greater number of parameters was chosen. However, when the test statistic was not significant, there is no significant difference in the fit of the two models and the model with the fewer number of parameters was selected. Once the fixed effects were finalized based on the chi square test, no further simplification of the fixed-effects component of the model was performed. For all pollutants and bags, only the cleanout effect was significant – the emissions levels did not differ significantly between cars and trucks.

The change in emissions between pre- (as-received level) and post-cleanout at 28 ppm was estimated using the differences of least squares means from the final model. The Tukey-Kramer adjustment was used in calculating the p-values for the least squares means. The difference of least squares means between pre- and post-cleanout were reverse-transformed to grams per mile space to estimate the percent reduction in emissions. Table 7-4 shows the reductions in emissions for all pollutants between pre- and post-cleanout at the fuel sulfur level of 28 ppm, reflecting the level of reversible sulfur loading in catalysts. The statistical analysis of the “clean-out at 28 ppm” showed highly significant reductions in emissions for several pollutants including NO$_x$ and hydrocarbons. The findings here suggest there is a degradation of emission performance for Tier 2 vehicles in-use due to sulfur loading in the catalyst, and it has a measureable effect on aftertreatment performance.

Table 7-4 Percent Reduction in In-Use Emissions After the Clean-Out Using 28 ppm Test Fuel

	NO$_x$ (p-value)	THC (p-value)	CO (p-value)	NMHC (p-value)	CH4 (p-value)	PM (p-value)
Bag 1	–	–	6.0% (0.0151)	–	–	15.4% (< 0.0001)
Bag 2	31.4% (0.0003)	14.9% (0.0118)	–	18.7% (0.0131)	14.4% (0.0019)	–
Bag 3	35.4% (<0.0001)	20.4% (<0.0001)	21.5% (0.0001)	27.7% (<0.0001)	10.3% (<0.0001)	24.5% (<0.0001)
FTP Composite	11.4% (0.0002)	3.8% (0.0249)	6.8% (0.0107)	3.5% (0.0498)	6.0% (0.0011)	13.7% (<0.0001)
Bag 1 – Bag 3	–	–	7.2% (0.0656)	–	–	–

The clean-out effect is not significant at $\alpha = 0.10$ when no reduction estimate is provided.

7.3.2. Effect of Clean-Out at 5 ppm

To study the differences in the effectiveness of the clean-out procedure between 28 ppm and 5 ppm fuel sulfur levels, a dataset was constructed using the subset of “sulfur level” data which include the measurements from a subset of original and modified ‘long’ procedures and modified ‘short’ procedure with mileage accumulation less than 50 miles for both sulfur levels. This dataset essentially represents the first three repeat tests of FTP from each sulfur level following a high-speed/load “clean-out” procedure consisting of two back-to-back US06 cycles.

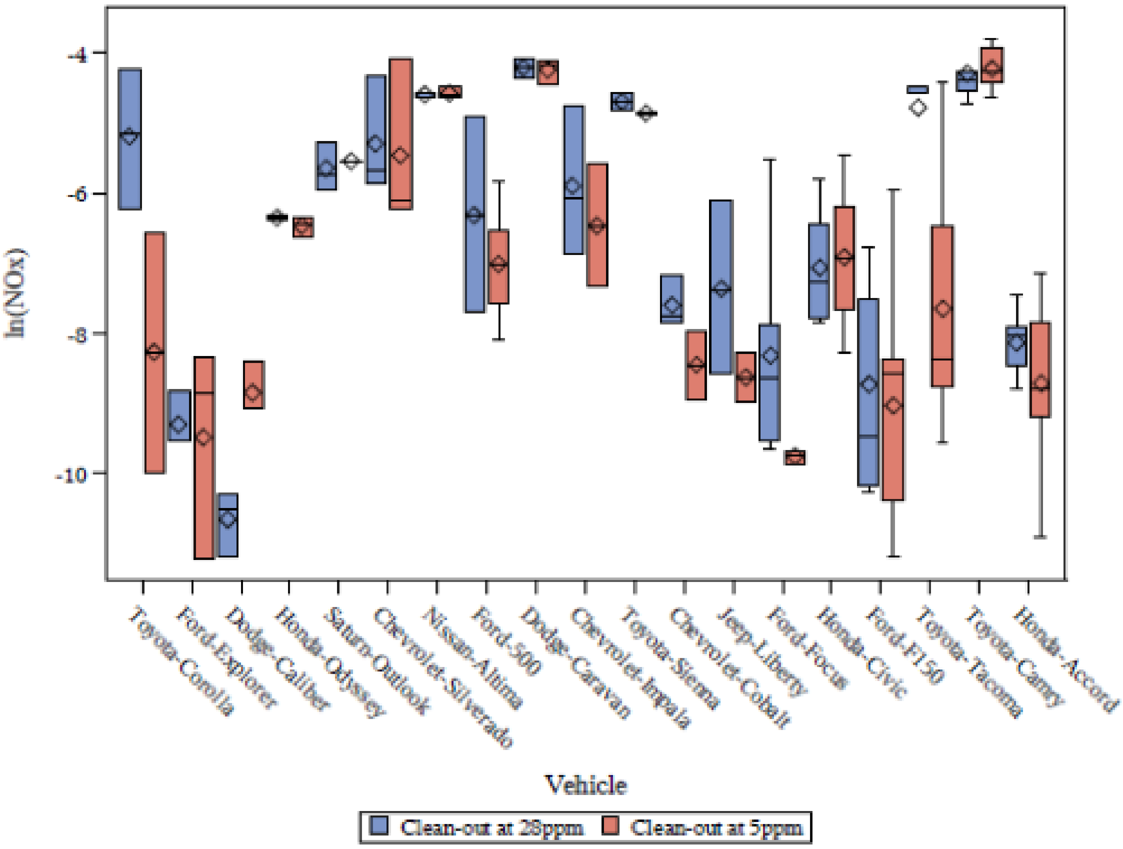
The dataset included 17 vehicle families, 35 individual vehicles with 183 observations – 88 measurements from clean-out at 28 ppm and 95 measurements from clean-out at 5 ppm. The average starting odometer reading of 35 unique vehicles was 31,178 ± 6,351 miles. Additional details of the test fleet are shown in Table 7-5.

Table 7-5 Description of Vehicles in the "Clean-Out at 5 ppm Data"

Vehicle Family ID	Vehicle ID	Make	Model	Model Year	Tier 2 Cert Bin	Number of Vehicles	Average Starting Odometer
M500	0003	Toyota	Corolla	2007	5	1	33,122
M501	0023	Ford	Explorer	2007	4	1	27,562
M502	0026	Dodge	Caliber	2007	5	1	29,097
M503	0194	Honda	Odyssey	2007	5	1	35,816
M504	0021	Saturn	Outlook	2007	5	1	43,733
M505	0031	Chevrolet	Silverado	2007	5	1	27,891
M506	0123	Nissan	Altima	2007	5	1	39,936
M507	0148	Ford	Taurus	2007	5	1	28,802
M508	0075	Dodge	Caravan	2007	8	1	41,117
M509	0046	Chevrolet	Impala	2007	5	1	37,734
N510	0264	Toyota	Sienna	2007	5	1	38,464
N511	0179	Chevrolet	Cobalt	2008	5	1	38,722
N512	0107	Jeep	Liberty	2008	5	1	24,614
N513	0089, 0178	Ford	Focus	2008	4	2	24,726
N514	0010, 0101, 0104	Honda	Civic	2008	5	3	32,931
N515	0006, 0007, 0074, 0165	Ford	F150	2008	8	4	29,738
N520	0011, 0022, 0026,	Toyota	Tacoma	2009	5	3	28,964
N521	0131, 0162, 0179, 0280, 0329	Toyota	Camry	2008	3	5	28,506
P522	0009, 0039, 0146, 0045, 0011	Honda	Accord	2008	3	5	29,601

Figure 7-3 shows the box-plot of log-transformed emissions from Bag 2 NO_x by vehicle family and by clean-out sulfur level at 28 ppm and 5 ppm. The diamond and the line inside the box represent the mean and the median, respectively. The box represents the interquartile range between 25th and 75th percentile and the error bars show the full data range. The data generally shows that the NO_x emissions from clean-out at 28 ppm fuel sulfur are higher compared to the emissions from clean-out at 5 ppm, suggesting that the fuel sulfur level affects the effectiveness of the clean-out cycle in reducing sulfur loading in the catalyst. Furthermore, the data illustrates that there are significant between- and within-vehicle family variability.

Figure 7-3 Box-Plot of Vehicle Families by Clean-Out Sulfur Level at 28 ppm and 5 ppm (NO_x Bag 2)



The statistical approach described in Section 7.3.1 was applied in modeling the fixed effects and the covariance structure of this dataset, including modeling the vehicle family as a random effect. Table 7-6 summarizes the percent reduction in emissions from 28 ppm to 5 ppm fuel sulfur level only for the first three test replicates following the clean-out cycle. The results from bag 2 NO_x, for example, shows that immediately after the clean-out, a reduction of 34 percent was achieved by lowering the fuel sulfur level. This suggests that the effectiveness of high-temperature and high-speed cleanout cycle to reduce catalyst sulfur loading is limited by the level of sulfur in the fuel. Also of note here is the absence of any statistically significant PM reductions for these procedures; there is more discussion of PM effects in 7.3.3.

Furthermore, the reduction in emissions from cleanout shown in Table 7-4 would likely be larger if the low sulfur test fuel at 5 ppm had been used for the cleanout procedure and the tests that immediately followed the as-received baseline emissions. This is demonstrated in Figure 7-4, which compares the pre-cleanout (as-received) emissions level to the emissions level after the clean-out using 28 ppm and 5 ppm fuel sulfur. The data generally shows that the lower fuel sulfur level increases the effectiveness of the cleanout cycle. In order to eliminate the vehicle variability, only the pre-cleanout results from 35 unique vehicles that received the cleanout at both 28 ppm and 5 ppm are included (rather than comparing these results on 35 vehicles to the pre-cleanout results from 93 vehicles in Section 7.3.1).

Figure 7-4 Box-Plot of Vehicle Families by Pre-Cleanout and Post-Cleanout at 28 ppm and 5 ppm (NO$_x$ Bag 2)

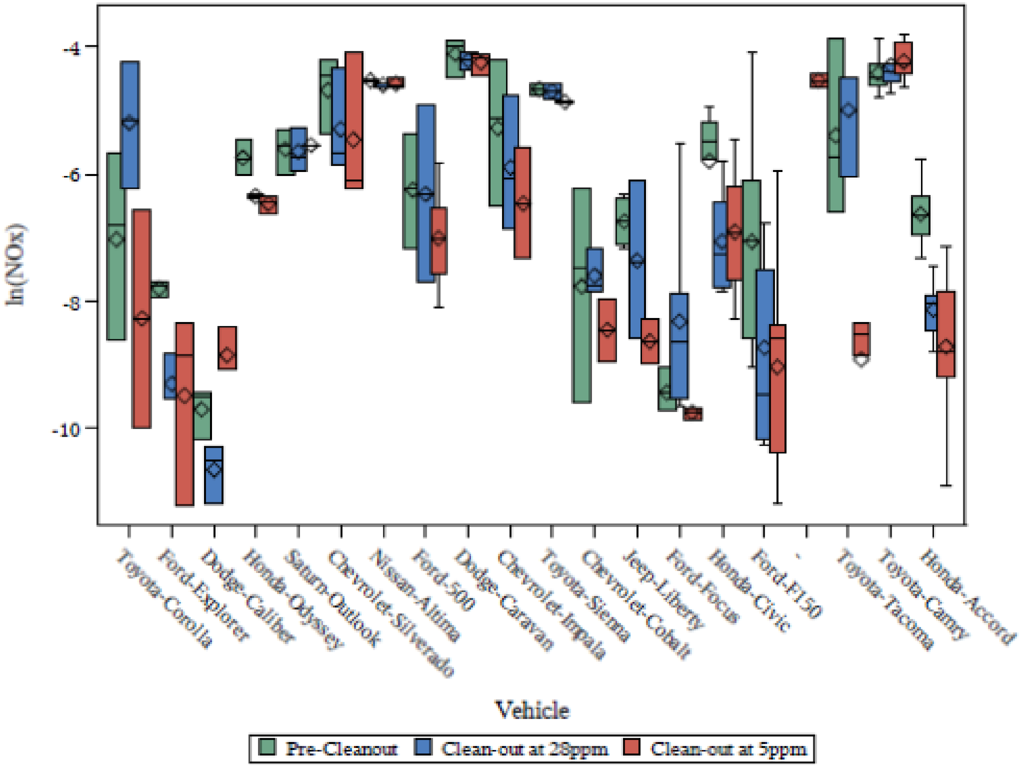


Table 7-6 Percent Reduction in Emissions From 28 ppm to 5 ppm for the First Three Repeat FTP Tests Immediately Following Clean-Out

	NO_x (p-value)	THC (p-value)	CO (p-value)	NMHC (p-value)	CH4 (p-value)	PM
Bag 1	5.3% (0.0513)	6.8% (0.0053)	6.2% (0.0083)	5.7% (0.0276)	14.0% (<0.0001)	–
Bag 2	34.4% (0.0036)	33.9% (<0.0001)	–	26.4% (0.0420)	49.4% (<0.0001)	–
Bag 3	42.5% (<0.0001)	36.9% (<0.0001)	14.7% (0.0041)	51.7% (<0.0001)	28.5% (<0.0001)	–
FTP Composite	15.0% (0.0002)	13.3% (<0.0001)	8.5% (0.0050)	10.9% (0.0012)	23.6% (<0.0001)	–
Bag 1 – Bag 3	–	–	–	–	–	–

The effectiveness of clean-out cycle is not significant at $\alpha = 0.10$ when no reduction estimate is provided.

7.3.3. Effect of Sulfur level

The "sulfur level" data represents the emission measurements from the repeated FTP test cycles following the clean-out. This dataset provided key information for assessing the in-use effect of sulfur level on emissions over time as vehicles operate on fuels with different sulfur levels. The dataset is comprised of emission measurements from two different fuel sulfur levels at 28 ppm and 5 ppm, following vehicle operation that removed sulfur from the catalyst (clean-out procedure).

7.3.3.1. Tier 2 Vehicles

The "sulfur level" data for Tier 2 vehicles included all measurements from vehicles tested on both sulfur levels, for a total of 35 vehicles from 19 vehicle families (Table 7-7). The average starting odometer of the 35 vehicles was 31,178 ± 6,351 miles. A total of 322 measurements were taken from Long (S/L and L/M) and modified Short procedures – 161 measurements each for both high and low fuel sulfur levels.

Table 7-7 Description of Tier 2 Vehicles in the “Sulfur Level” Data

Vehicle Family ID	Vehicle ID	Make	Model	Model Year	Tier 2 Cert Bin	Number of Vehicles	Average Starting Odometer
M500	0003	Toyota	Corolla	2007	5	1	33,122
M501	0023	Ford	Explorer	2007	4	1	27,562
M502	0026	Dodge	Caliber	2007	5	1	29,097
M503	0194	Honda	Odyssey	2007	5	1	35,816
M504	0021	Saturn	Outlook	2007	5	1	43,733
M505	0031	Chevrolet	Silverado	2007	5	1	27,891
M506	0123	Nissan	Altima	2007	5	1	39,936
M507	0148	Ford	Taurus	2007	5	1	28,802
M508	0075	Dodge	Caravan	2007	8	1	41,117
M509	0046	Chevrolet	Impala	2007	5	1	37,734
N510	0264	Toyota	Sienna	2007	5	1	38,464
N511	0179	Chevrolet	Cobalt	2008	5	1	38,722
N512	0107	Jeep	Liberty	2008	5	1	24,614
N513	0089, 0178	Ford	Focus	2008	4	2	24,726
N514	0010, 0101, 0104	Honda	Civic	2008	5	3	32,931
N515	0006, 0007, 0074, 0165	Ford	F150	2008	8	4	29,738
N520	0011, 0022, 0026,	Toyota	Tacoma	2009	5	3	28,964
N521	0131, 0162, 0179, 0280, 0329	Toyota	Camry	2008	3	5	28,506
P522	0009, 0039, 0146, 0045, 0011	Honda	Accord	2008	3	5	29,601

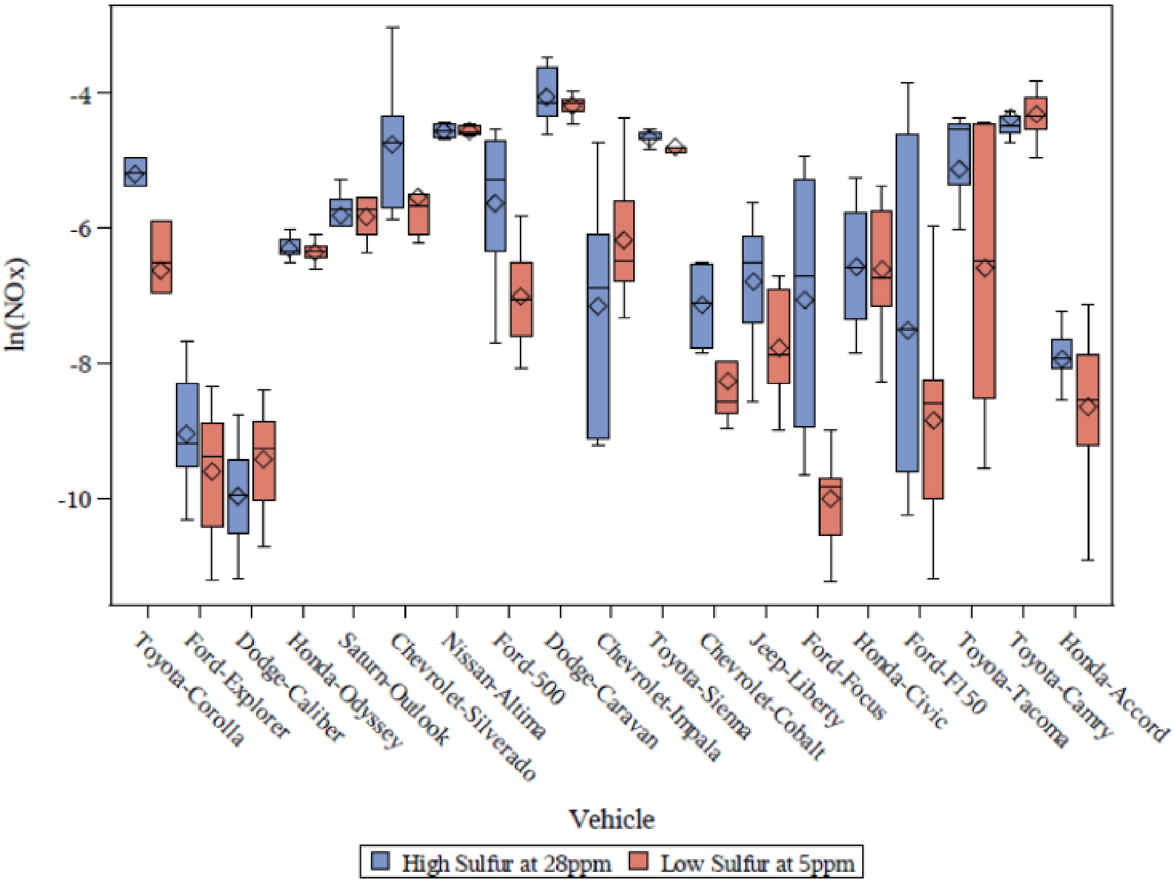
The box-plot of the log-transformed emissions from Bag 2 NO_x “sulfur level” data (Figure 7-5) shows the spread of the data for each vehicle family and sulfur level across all mileages. The diamond and the line inside the box represent the mean and the median, respectively. The box represents the interquartile range between 25th and 75th percentile and the error bars show the full data range. Generally, there is a tendency for the vehicles running on high sulfur fuel to emit more NO_x than the vehicles running on low sulfur fuel. However, the effect of operation on higher sulfur fuel certainly varies by vehicle family, suggesting the presence of substantial between-vehicle family variability. For example, Toyota Corolla, Ford

Focus, and Chevrolet Cobalt clearly show a large effect of fuel sulfur level on emissions while the effect is more marginal for the Nissan Altima and Honda Civic.

As a result of the changes in testing procedures described in Section 6.2, the number of tested vehicles is not the same across vehicle families in the “sulfur level” data. Considering the differences in number of unique vehicles in each vehicle family and the presence of variability between vehicle families illustrated by Figure 7-5, each vehicle family was considered as a random effect in constructing the statistical model, similar to the analyses done for the “clean-out at 28 ppm” and “clean-out at 5 ppm”.

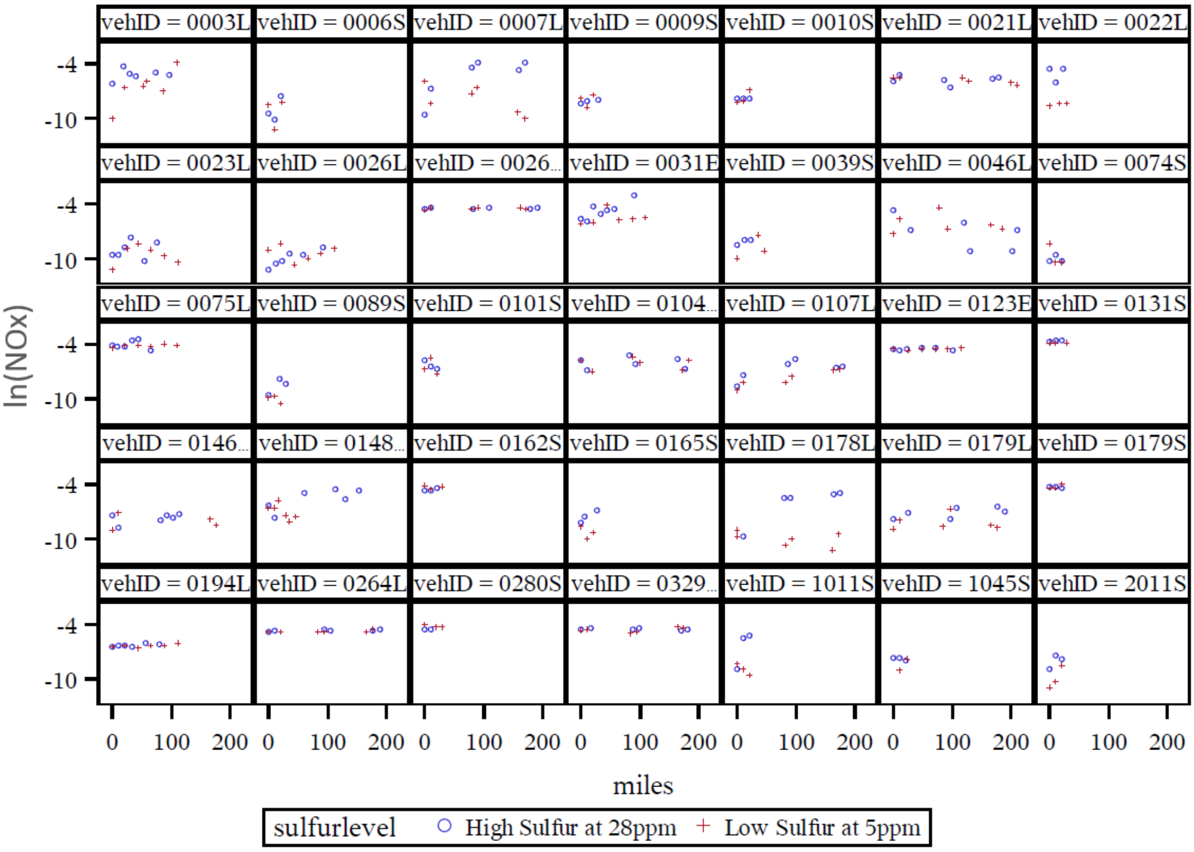
Figure 7-6 presents the log-transformed emissions from individual vehicles by sulfur level. The plot shows that the increase in emissions as vehicles accumulate mileage for the high sulfur level is more significant compared to the low sulfur level, contributing to the increased variance for some vehicles and suggests that the rate of sulfur loading might be different for the two fuel sulfur levels. Thus, the interaction between sulfur level and the accumulated mileage was included in the statistical modeling of the data. These findings from the graphical examination of the data assisted in selecting the statistical modeling approach discussed in this section.

Figure 7-5 Box-Plot of Individual Vehicle Families by Sulfur Level (NO$_x$ Bag 2)



We refrained from looking at the simple descriptive statistics, such as the mean and the standard deviation, to assess the relationship between the sulfur level and emissions even as a preliminary step, because reaching any conclusion from them can potentially be very misleading due to the presence of repeated measurements, and both between-vehicle and within-vehicle variability. In addition, the mileage accumulations varied from vehicle to vehicle, and simple descriptive statistics would not be able to accurately capture the significant amount of variability inherent in the dataset.

Figure 7-6 Log-Transformed Emissions from Individual Vehicles by Sulfur Level (NO$_x$ Bag 2)



In analyzing the "sulfur level" data, a similar top-down model fitting statistical approach to that applied to the "clean-out" data, as described in Section 7.3.1, was applied to characterize the effects of fuel sulfur level on emissions as a function of accumulated mileages since cleanout.

The dependent variable (Y_i) was the natural logarithm of emissions. The fixed effects (X_i) included in the model were sulfur level, accumulated mileage, vehicle type, and the interaction terms. The random effects (Z_i) were each vehicle family in the study. The likelihood ratio test for the significance of between-vehicle variation was statistically significant for all pollutants and bags, and thus, the random intercept for each vehicle family was included in the model. The significance of the between-vehicle variation was observed graphically in Figure 7-5.

All of the measurements from the same vehicle family will have the same between-vehicle family errors; their within-vehicle family errors will differ, and can be correlated within a vehicle family. The measurements from the same vehicle family are assumed to be correlated because they share common vehicle characteristics and have similar emission profiles. Also, measurements on the same vehicle close in time are often more highly correlated than measurements far apart in time as observed in Figure 7-6 – the covariation within vehicles. Both within- and between-vehicle errors are assumed independent from vehicle to vehicle. Since measurements on different vehicles are assumed independent, the structure refers to the covariance pattern of measurements on the same subject. For most of these structures, the covariance between two measurements on the same vehicle depends only on the differences in mileage accumulation between measurements, and the variance is constant over mileage.

The covariance structure was modeled by first fitting the "unstructured" (UN) covariance matrix with a saturated model including all fixed effects. For similar reasons provided in the analysis of the "clean-out" data, the unstructured model failed to converge. Next, since emissions were measured irregularly, where the mileage intervals between measurements are more or less unique to each vehicle, the spatial covariance structure, which allows for a continuous representation of mileage, was fit. However, the model failed to converge for the spatial covariance matrix as well. Thus, we proceeded to fit the compound symmetry (CS) structure which specifies that measurements at all mileage have the same variance, and that all measurements on the same vehicle have the same correlation. The BIC value for the compound symmetry was 803.36.

Lastly, the first-order autoregressive structure was modeled. It assumes that the variances are homogeneous and the correlations decline exponentially with time, i.e., the variability in measured emissions is constant regardless of mileage for each vehicle and the measurements next to each other are more correlated than the measurements further apart. The BIC value for the first-order autoregressive structure was 764.90. Since the BIC value of the first-order autoregressive structure was lower than the BIC value for the compound symmetry of 803.36, the first-order autoregressive structure (Equation 7-3) was selected as the covariance matrix.

$$\boldsymbol{R_i} = \boldsymbol{Var(\epsilon_i)} = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \cdots & \sigma^2\rho^{n-1} \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho^{n-1} & \sigma^2\rho^{n-2} & \sigma^2\rho^{n-3} & \dots & \sigma^2 \end{bmatrix} \quad \textbf{Equation 7-3}$$

A combination of first-order autoregressive structure within vehicles and a random effect between vehicles was used to model the covariance structure which specified an inter-vehicle random effect of differences between vehicles, and a correlation structure within vehicles that decreases with increasing mileage lag between emission measurements. Furthermore, the variance associated with the low sulfur level was permitted to differ from the variance associated with the high sulfur level. Since the first-order autoregressive structure was selected due to limited available options, we acknowledge that there might be some limitations inherent in the assumption of constant distance between two measurements. However, the estimates of fixed effects, such as the differences between sulfur level means, may be the same for different covariance structures, differing only in the standard errors of these estimates.

Once the structures for the random effects and the covariance structure for the residuals were selected, the fixed effects in the model were tested using the approximate F-test with the Satterthwaite degrees of freedom. The step-wise backward elimination approach was used to remove any non-significant fixed effects (shown in red in Table 7-8), starting from the saturated model. The significance level of 10% ($\alpha = 0.1$) was used to test the null hypothesis while keeping statistical hierarchy.

Table 7-8 Type 3 Tests of Fixed Effects (NO_x Bag2)

	Effect[†]	Num DF	Den DF	F Value	Pr > F[‡]
Model 1	slevel	1	254	7.66	0.0061
	miles	1	271	0.10	0.7499
	vehclass	1	18.2	0.18	0.6761
	slevel * miles	1	170	0.79	0.3743
	miles * vehclass	1	280	1.20	0.2748
Model 2	slevel	1	259	7.63	0.0062
	miles	1	264	17.07	< 0.0001
	vehclass	1	17	0.40	0.5363
	slevel * miles	1	175	0.72	0.3982
Model 3	slevel	1	259	7.66	0.0061
	miles	1	264	17.08	< 0.0001
	slevel * miles	1	174	0.70	0.4028
Model 4	slevel	1	219	18.28	< 0.0001
	miles	1	270	17.54	< 0.0001

[†] slevel = sulfur level (high and low); miles = accumulated mileage since clean-out; vehclass = vehicle types (car and truck); [‡] Pr > F represents the p-value associated with the F statistic;

Then, the likelihood ratio test using maximum likelihood was performed to examine if the model can be reduced further without compromising the model fit. For example, comparing model 4 and 5 (Table 7-9), since the result of the likelihood ratio test was not statistically significant, we concluded that accumulated mileage does not have an effect on Bag 2 NO_x, and thus, model 5 was selected as the final model.

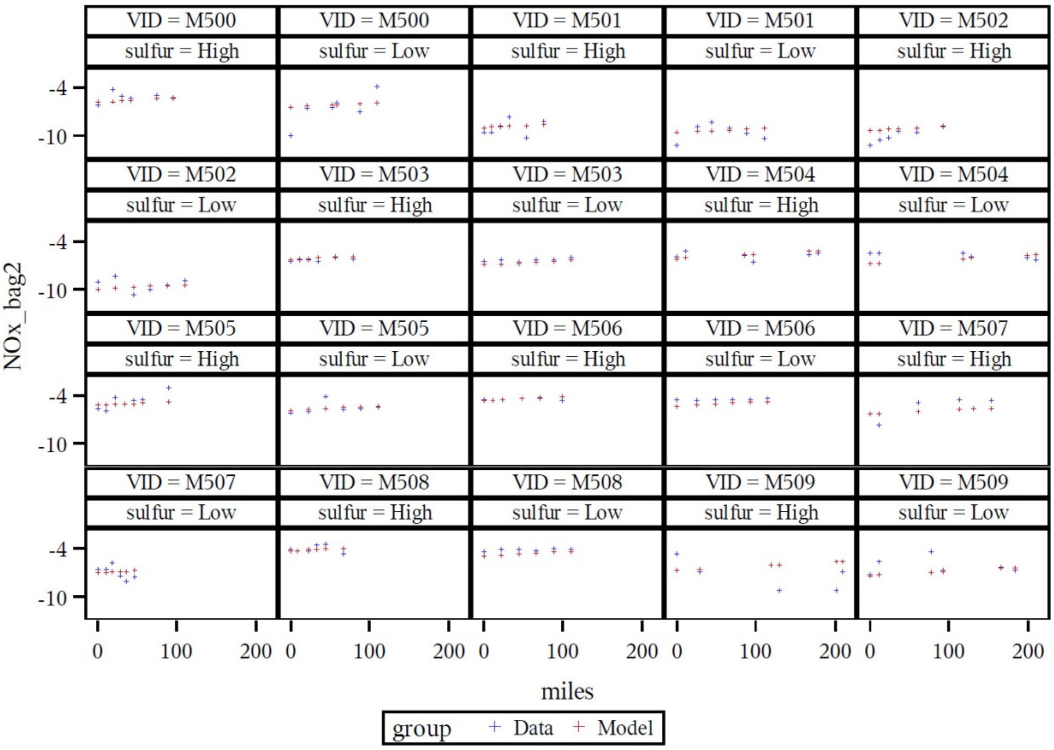
Table 7-9 Likelihood Ratio Test for Bag 2 NO_x Model

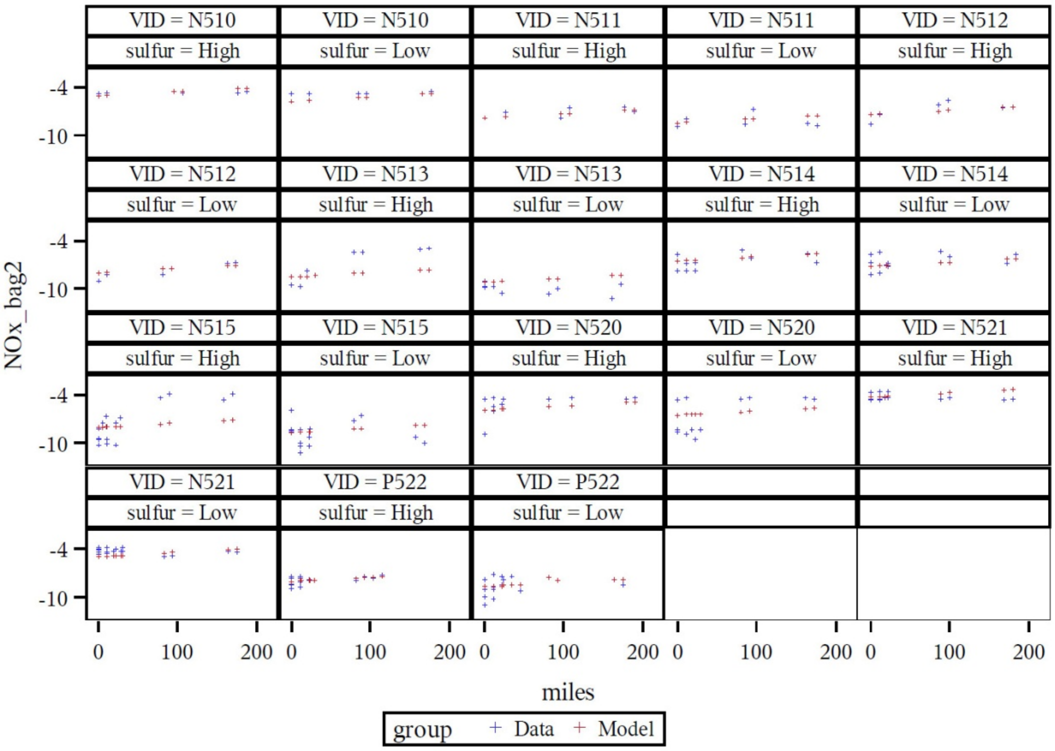
	Fixed effects in model	-2 Res Log Likelihood	p-value (χ^2)
Model 4	slevel, miles	991.6	0.1213
Model 5	slevel	994	

The final NO_x Bag 2 model (model 5) had sulfur level as the fixed effects. Thus, the model finds that there is a statistically significant difference in emissions from high and low fuel sulfur levels that do not differ between vehicle types (car vs. truck) since the sulfur level and vehicle type interaction term was not significant. Also, since the mileage term is not significant, it can be concluded that the mileage accumulation after the clean-out does not increase emissions independent of the fuel sulfur level. In addition, since the sulfur level and the accumulated mileage interaction term was not significant, the model suggests that the rate of sulfur loading does not vary by accumulated mileages after the clean-out between high and low fuel sulfur levels. In other words, the effect of high fuel sulfur on Bag 2 NO_x exists immediately after clean-out and remains essentially constant on a percentage basis, during subsequent driving of a vehicle.

Figure 7-7 shows the data vs. predicted plots based on the final model for NO_x Bag 2. There are two paired plots next to each other with the same vehicle ID showing emissions from both high and low sulfur. There are some instances (e.g., VID M502) where the model overestimates the effect of sulfur by over-predicting the emission levels of high sulfur and under-predicting the emission levels of low sulfur. In contrast, there are other instances (e.g., VID M513) where the model underestimates the effect of sulfur by under-predicting the emission levels of high sulfur and over-predicting the emission levels of low sulfur. However, this is to be expected given the variability in the emission testing. In general, the model predictions are in agreement with the data.

Figure 7-7 Data vs. Predicted (Log-Transformed Bag 2 NO$_x$)





Furthermore, the one-to-one plot of data vs. model predictions in Figure 7-8 shows that the points generally lie close to the 1:1 line and has the adjusted R-square of 0.71, demonstrating reasonable accuracy in model predictions for Bag 2 NO_x.

Figure 7-8 Data vs. Predicted (Log-Transformed NO_x Bag 2)

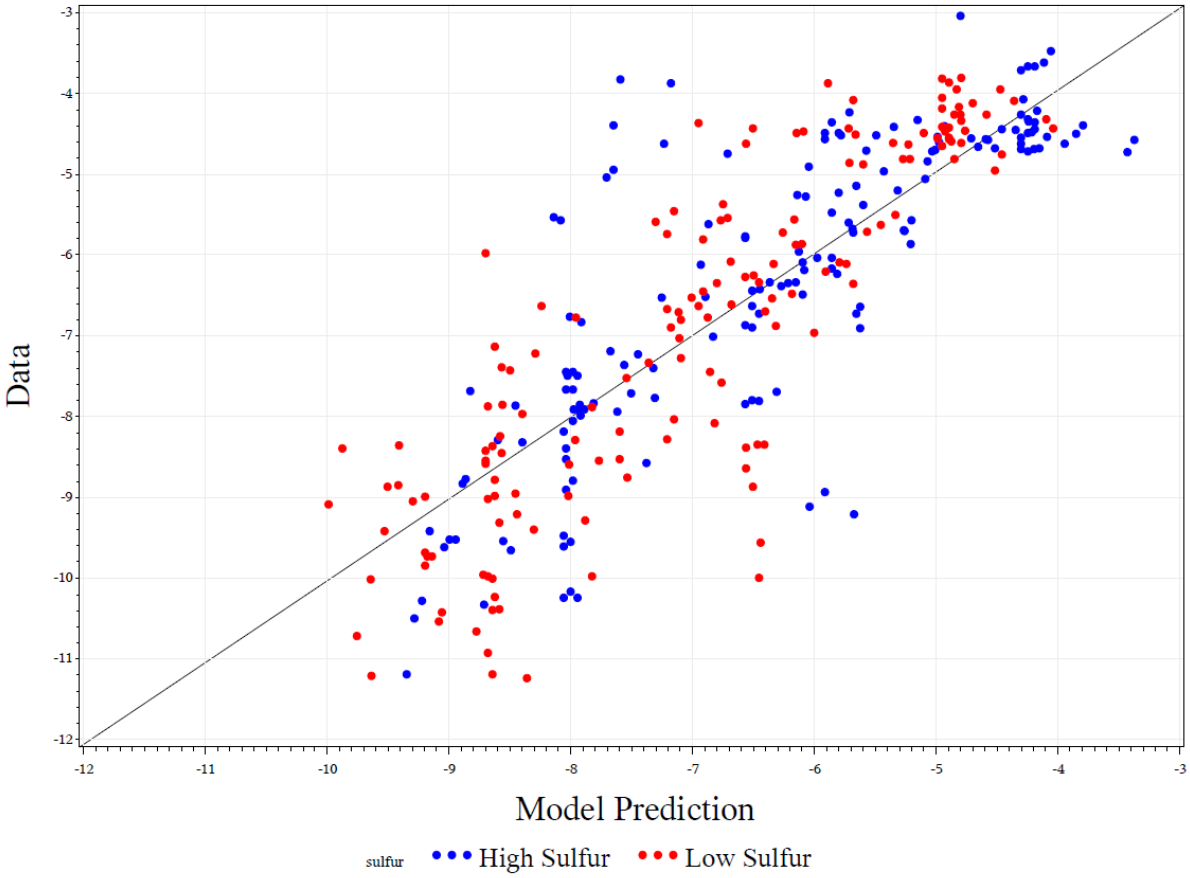


Figure 7-9 presents the model predictions for individual vehicle by sulfur level for Bag 2 NO_x in grams/mile. The model-predicted values depicted in this figure were generated from Model 5 in Table 7-8 for NO_x bag 2. The model for Bag 2 NO_x shows that sulfur loading resumes immediately after the clean-out (at or near zero accumulated mileage) with high fuel sulfur level emitting significantly higher emissions compared to low fuel sulfur level. In addition, the sulfur loading continues with vehicle operation, causing an increase in emissions. The rate of sulfur loading was not statistically different between high and low sulfur levels on a percentage basis.

Figure 7-9 Model Predictions for Individual Vehicle by Sulfur Level (NO_x Bag 2)

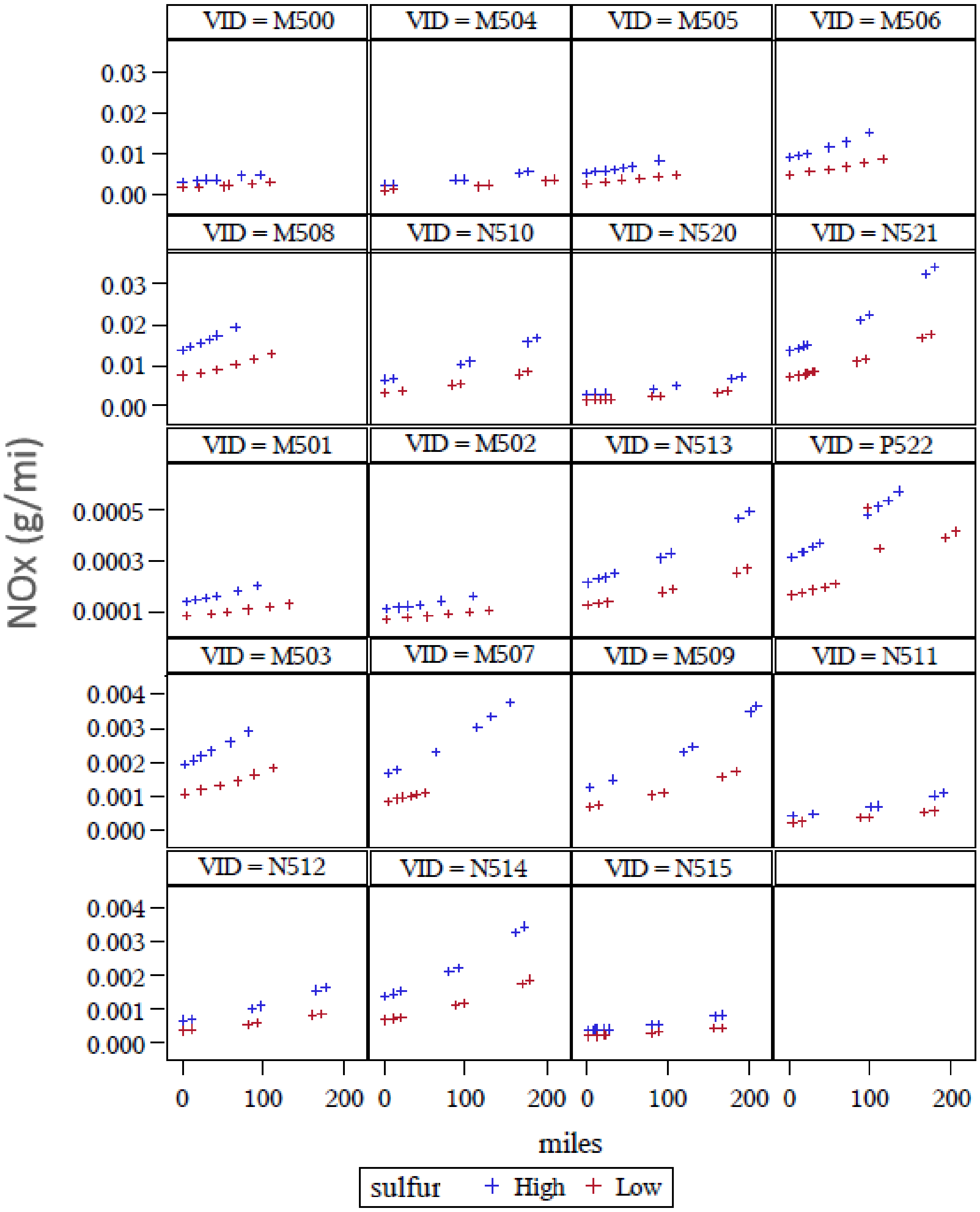


Table 7-10 summarizes the final models that were selected for all pollutants and bags, applying the same statistical methodology described for Bag 2 NO_x. For all models, the sulfur level and mileage interaction term was not significant, and the change in emissions from reducing the fuel sulfur from 28 ppm to 5 ppm was estimated using the differences of least squares means from the final model, adjusting for other effects in the model. The Tukey-Kramer adjustment was used in calculating the p-values for the least squares means. The differences of least squares means between high and low fuel sulfur level were reverse-transformed to estimate the percent reduction in emissions (Table 7-11). When the sulfur level and mileage interaction term is not significant, the percent differences in emissions between high and low fuel sulfur levels stay constant across accumulated mileage after clean-out (the sulfur loading curves for high and low sulfur are parallel) and thus, using the least squares means to quantify the reduction in emissions without considering the as-received in-use sulfur loading was sufficient.

Table 7-10 Final Models of All Pollutants

Pollutant	Bag	Fixed Effects[†]
NO_x	Bag 1	slevel, miles
	Bag 2	slevel
	Bag 3	slevel, miles
	FTP Composite	slevel, miles
	Bag 1 – Bag 3	–
THC	Bag 1	slevel, miles
	Bag 2	slevel, miles
	Bag 3	slevel, miles
	FTP Composite	slevel, miles
	Bag 1 – Bag 3	slevel
CO	Bag 1	slevel, miles
	Bag 2	–
	Bag 3	slevel, miles
	FTP Composite	slevel, miles
	Bag 1 – Bag 3	–
NMHC	Bag 1	slevel
	Bag 2	slevel, miles
	Bag 3	slevel, miles
	FTP Composite	slevel, miles
	Bag 1 – Bag 3	–
CH_4	Bag 1	slevel, miles
	Bag 2	slevel, miles
	Bag 3	slevel, miles
	FTP Composite	slevel, miles
	Bag 1 – Bag 3	–
PM	Bag 1	–
	Bag 2	–
	Bag 3	–
	FTP Composite	–
	Bag 1 – Bag 3	–

[†] slevel = sulfur level (high and low); miles = accumulated mileage since clean-out;

Table 7-11 summarizes the percent reduction in emissions from 28 ppm to 5 ppm fuel sulfur for all pollutants and all bags. The results suggest that significant reductions in emissions can be achieved by reducing the fuel sulfur levels from 28 to 5 ppm in the in-use fleet of Tier 2 vehicles. Unlike the gaseous pollutants, there was no effect of sulfur level found for PM. A potential explanation is that the majority of PM mass as measured in this program (that is, from normal-emitting Tier 2 vehicles operated at low and moderate loads) was likely soot produced shortly after cold start (bag 1).[28] Once formed in the combustion chamber, oxidation of soot requires residence time at high temperature with lean air-fuel conditions. Since modern gasoline engines are calibrated to operate at or very near a stoichiometric fuel/air mixture over all operating modes, minimal destruction of soot occurs in the catalyst regardless of its relative efficiency. As a result, sulfur would not be expected to have a significant effect on directly-emitted PM (other than the very small amounts of sulfate), which is consistent with these results. A clean-out effect on PM was observed in the initial portion of the procedures but not later, suggesting it may not have been an actual effect of sulfur on catalyst efficiency but something else related to release and accumulation of material typically measured as PM. Since there were no analyses of PM composition in this program, we are not able to draw more definitive conclusions.

Table 7-11 **Percent Reduction in Emissions from 28 ppm to 5 ppm Fuel Sulfur on In-Use Tier 2 Vehicles**

	NO_x (p-value)	**THC (p-value)**	**CO (p-value)**	**NMHC (p-value)**	**CH4 (p-value)**	**NO_x+NMOG (p-value)**	**PM[a]**
Bag 1	7.1% (0.0216)	9.2% (0.0002)	6.7% (0.0131)	8.1% (0.0017)	16.6% (< 0.0001)	N/A	–
Bag 2	51.9% (< 0.0001)	43.3% (< 0.0001)	–[a]	42.7% (0.0003)	51.8% (< 0.0001)	N/A	–
Bag 3	47.8% (< 0.0001)	40.2% (< 0.0001)	15.9% (0.0003)	54.7% (< 0.0001)	29.2% (< 0.0001)	N/A	–
FTP Composite	14.1% (0.0008)	15.3% (< 0.0001)	9.5% (< 0.0001)	12.4% (< 0.0001)	29.3% (< 0.0001)	14.4% (<0.0001)	–
Bag 1 - Bag 3	–[a]	5.9% (0.0074)	–[a]	–[b]	–[b]	N/A	–

[a] Sulfur level not significant at $\alpha = 0.10$.

[b] Inconclusive because the mixed model did not converge.

7.3.3.2. Tier 3-Like Vehicles

Following the main test program of Tier 2 vehicles, a set of vehicles meeting lower "Tier 3-like" emissions standards were tested to evaluate the effect of sulfur on these newer and cleaner vehicles. These vehicles were tested using the same fuel and test procedures described in Chapters 5 and 6. The "sulfur level" data for "Tier 3-like" vehicles consisted of all measurements from the five vehicles tested on both 28 and 5 ppm sulfur fuels. A total of 64 measurements were taken from Long (L/M) and modified Short procedures – 33 measurements from high fuel sulfur levels and 31 measurements from low fuel sulfur levels. The description of the vehicles tested in the supplemental program is shown in Table 7-12.

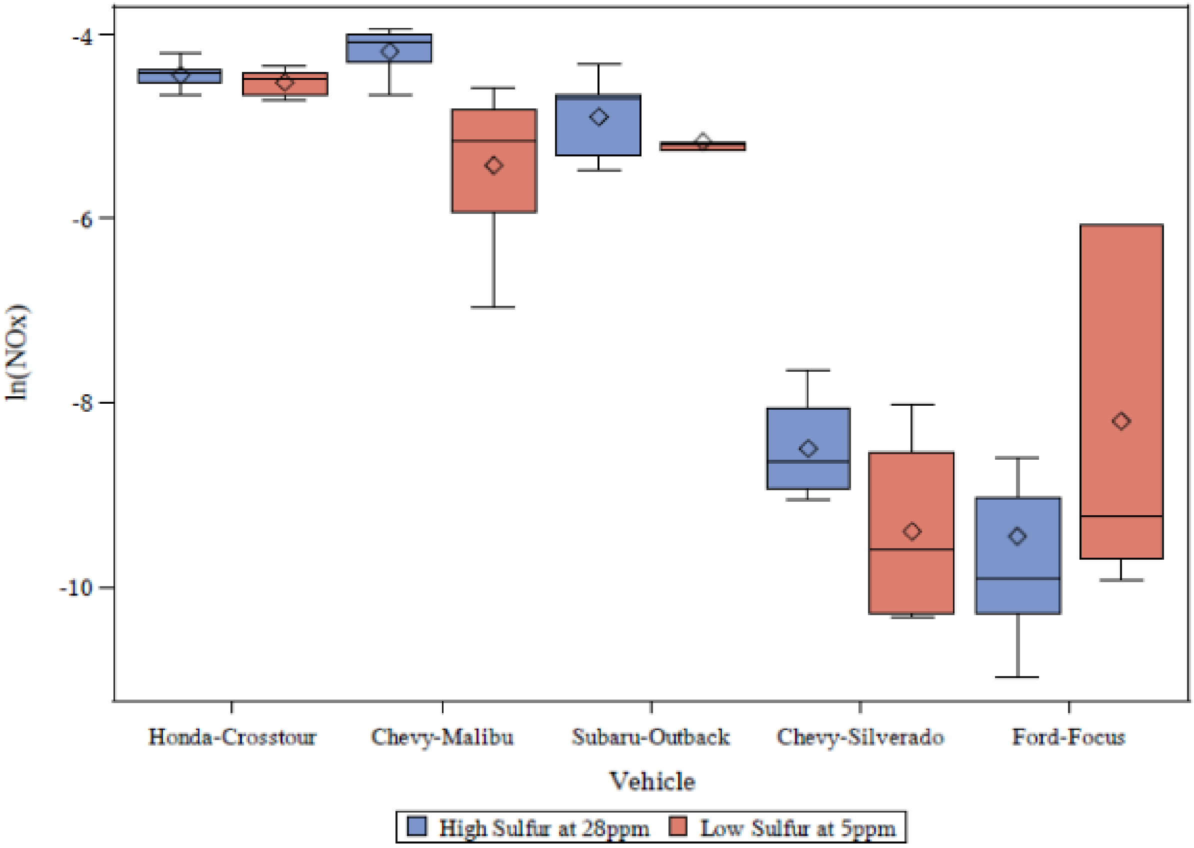
Table 7-12 Description of "Tier 3-like" Vehicles in the "Sulfur Level" Data

Vehicle Family ID	Vehicle ID	Make	Model	Model Year	Emission Standards	Starting Odometer	Vehicle Origin
P528	0001L	Honda	Crosstour	2011	ULEV	12,827	Recruited
P530	0001	Chevy	Malibu	2010	SULEV	10,285	Manufacturer [a]
P531	0001L	Subaru	Outback	2008	SULEV	36,635	Recruited
R532	0001L	Ford	Focus	2010	SULEV	28,673	EPA-owned
P532	0001L	Chevy	Silverado	2011	T2 B4	714	EPA-owned

[a] This vehicle was loaned by Umicore Autocat USA, and is the same vehicle used in their 2011 study.

The box-plot of the log-transformed emissions from Bag 2 NO_x "sulfur level" data (Figure 7-10) shows the spread of the data for each vehicle and sulfur level across all mileages. The diamond and the line inside the box represent the mean and the median, respectively. The box represents the interquartile range between 25th and 75th percentile and the error bars show the full data range. Generally, there is a tendency for the vehicles running on high sulfur fuel to emit more NO_x than the vehicles running on low sulfur fuel. However, the effect of operation on higher sulfur fuel certainly varies by each vehicle.

Figure 7-10 Box-Plot of "Tier 3-Like" Vehicles by Sulfur Level (NO_x Bag 2)



In analyzing the "sulfur level" data for "Tier 3-like" vehicles, a similar top-down model fitting statistical approach to that described in Section 7.3.1, was applied to characterize the effects of fuel sulfur level on emissions as a function of accumulated mileages since cleanout. The dependent variable (Y_i) was the natural logarithm of emissions. The fixed effects (X_i) included in the model were sulfur level, accumulated mileage, vehicle type, and the interaction terms. The random effects (Z_i) were each vehicle in the study. A combination of first-order autoregressive structure within vehicles and a random effect between vehicles was used to model the covariance structure which specified an inter-vehicle random effect of differences between vehicles, and a correlation structure within vehicles that decreases with increasing mileage lag between emission measurements. The same statistical methodologies utilized for evaluating the sulfur level effects for Tier 2 vehicles were applied to "Tier 3-like" vehicles.

Table 7-14 summarizes the percent reduction in emissions from 28 ppm to 5 ppm fuel sulfur for all pollutants and all bags. The results suggest that significant reductions in emissions can be achieved by reducing the fuel sulfur levels from 28 to 5 ppm in the in-use fleet of "Tier 3-like" vehicles.

As indicated in the analysis, the cleaner "Tier 3-like" vehicles are impacted more significantly in Bag 1 NO$_x$ and THC (and composite as a result) than what was observed in the analysis of the Tier 2 vehicles. This is not unexpected since the cleaner vehicles tend to rely more on efficient catalyst activity sooner in the operation of the vehicle following the cold start. The sulfur hinders the catalyst from performing at expected efficiency levels early in the operation, resulting in a larger penalty to these cleaner vehicles that rely more heavily on the catalyst to meet the lower emission standards. Overall, we expect lower-emitting Tier 3 vehicles to show similar or greater sensitivity to the fuel sulfur levels compared to the conventional Tier 2 vehicles.

Table 7-13 Percent Reduction in In-Use Emissions from 28 ppm to 5 ppm Fuel Sulfur on "Tier 3-Like" Vehicles

	NO$_x$ (p-value)	THC (p-value)	CO (p-value)	NMHC (p-value)	CH4 (p-value)	NO$_x$+NMOG (p-value)	PM
Bag 1	15.4% (0.0731)	13.5% (0.0388)	17.7% (<0.0001)	–	19.7% (0.0003)	N/A	–
Bag 2	40.9% (0.0074)	–	–	–	27.5% (0.0021)	N/A	–
Bag 3	–	30.1% (0.0294)	28.0% (0.0587)	–	29.2% (0.0024)	N/A	–
FTP Composite	23.9% (0.0203)	14.6% (0.0312)	21.0% (< 0.0001)	–	24.8% (0.0002)	13.9% (< 0.0001)	–
Bag 1 - Bag 3	–	–	–	–	–	N/A	–

Sulfur level not significant at $\alpha = 0.10$ when no reduction estimate is provided.

7.3.4. Sensitivity Analysis

A series of sensitivity analyses of the "sulfur level" data was performed to address some of the issues that might affect the mixed model results. They include the impacts of: low concentration measurements, censoring of measurements with zero values, and influential vehicles. The sensitivity analyses were conducted only for Bag 2 NO_x, since above mentioned issues pertain the most to Bag 2 NO_x. For example, Bag 2 NO_x showed a higher percentage of measurements with zero values than most other pollutant and bag combinations, as illustrated in Table 7-1.

Effect of Low Concentration Measurements

The issue of measurements with very low concentration from Bag 2 NO_x has been discussed in Section 7.1.1. To address uncertainty of measurements from these very low-emitting vehicles, we performed sensitivity analyses using two measurement concentration screening levels: 100 ppb (based on the lower end of the instrument manufacturer's stated calibration range for the emission analyzer), and 50 ppb (chosen at half the former limit). In each analysis, vehicles with all sample measurements falling below the screening level were removed, and models were refit. Results of these sensitivity analyses are provided in Table 7-14.

Table 7-14 Results of sensitivity modeling analysis (NO_x Bag 2)

Model Description	Number of Vehicles	Observations	Model Estimate of Bag 2 NO_x Reduction
Final NO_x bag 2 model	35	322	51.9%
50 ppb vehicle screen	28	263	48.4%
100 ppb vehicle screen	19	191	48.2%

In each of these sensitivity analyses, the sulfur level effect remained highly significant with p-value < 0.004, suggesting a meaningful sulfur effect exists regardless of removal of the lowest-emitting vehicles. Thus, we conclude that the sulfur effect is considerably larger than the uncertainty or error associated with the measurements.

Effect of Use of Imputed Values

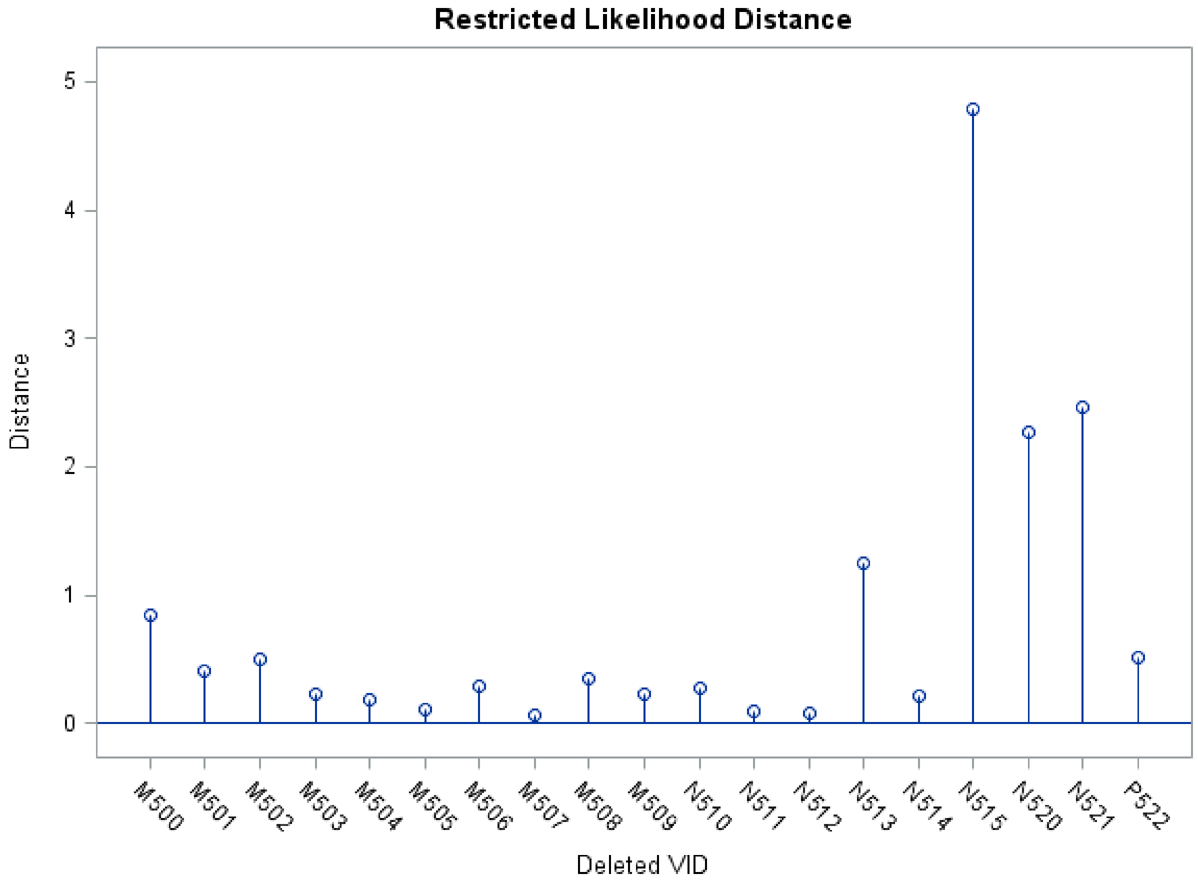
In order to assess the impact of replacing censored values, models with and without imputed values for Bag 2 NO_x were compared. For the model without imputed values, the mixed model was re-fit using a new dataset with all imputed values removed, consisting only of actual measurements. Based on the examination of the estimates of fixed effects and the standard errors from both models, we concluded that the imputed values did not significantly bias the results. The percent reduction in emissions from 28 ppm to 5 ppm fuel sulfur level was changed from 51.9% in model with imputed values to 50.0% in model without them. The sulfur level effect remained highly significant with p-value <.0001 for the model with and without the imputed values.

Effect of influential vehicles

As an additional test of robustness, we also looked at the impact of removing the influential vehicles from the dataset. Influence can be broadly defined as the ability of a single or multiple vehicles to affect the resulting outcome through the presence or absence in the data. The influential vehicles can be identified by examining the restricted likelihood distance (RLD), which is calculated after an iterative process of refitting the model with and without each vehicle.

Figure 7-11 shows the restricted likelihood distance from the influential diagnostics where vehicle family IDs N515, N520, and N521 can be considered influential vehicles affecting both the fixed effects and covariance parameter estimates based on Cook's D and COVRATIO estimates. Although we do not have specific grounds for excluding these vehicles from the mixed model analysis since the measurements from these vehicles did not fall into the category of neither low concentration measurements nor the outlying observations, these influential vehicles were removed and the model for Bag 2 NO_x was refit to examine the impacts from these vehicles.

Figure 7-11 Influence Diagnostics for NO_x Bag 2



The resulting model showed that the percent reduction in emissions from 28 ppm to 5 ppm was 52.1 percent, compared to the reduction of 51.9 percent from the final model. This analysis demonstrated that even when the influential vehicles are removed from the analysis, the reduction in emissions from reducing the fuel sulfur level from 28 ppm to 5 ppm is still highly significant with p-value <0.0001. The sensitivity analyses examining the influential vehicles for all pollutants and bags are presented in Appendix F.

8. Summary and Conclusions

This study assessed the emission reductions expected from in-use Tier 2 light duty vehicles with a reduction in gasoline sulfur content. The test fleet consisted of light-duty cars and trucks chosen to be representative of high sales models covering a range of types and sizes. Test fuels were two non-ethanol gasolines with properties typical of certification fuel, one at a sulfur content of 5 ppm and the other at 28 ppm.

Using the high-sulfur test fuel, emissions data were collected from vehicles in their as-received state, and then following a high-speed/load "clean-out" procedure consisting of two back-to-back US06 cycles to examine the existence of reversible sulfur loading in the in-use fleet. In addition, the differences in the effectiveness of the clean-out procedure at reducing emissions between the two fuel sulfur levels were assessed. Lastly, a representative subset of vehicles performed additional test replicates alternated with mileage accumulation on both high and low sulfur test fuels. This dataset was used to assess the differences in emission performance after clean-out as a function of fuel sulfur level. Major findings from this study include:

- Reversible sulfur loading is occurring in the in-use fleet of Tier 2 vehicles and has a measureable effect on emissions of NO_x and other pollutants of interest. For example, by performing a clean-out cycle on 28 ppm fuel, FTP composite NO_x was reduced by 11%, NMHC by 4%. A PM reduction of 14% was found, but it is not clear that it was a sulfur effect given other results in the program.
- The effectiveness of high speed/load procedures in restoring catalyst efficiency is limited when using higher sulfur fuel. Comparing emissions immediately following the clean-out procedure on 5 vs. 28 ppm fuel, FTP composite NO_x emissions were 15% lower, NMHC 11% lower, and CO 9% lower.
- Reducing fuel sulfur levels from 28 to 5 ppm is expected to achieve significant reductions in emissions of NO_x, hydrocarbons, and other pollutants of interest in the in-use fleet. For example, FTP composite NO_x was 14% lower, NO_x+NMOG 14% lower, and CO 10% lower. Several sensitivity analyses were performed for Bag 2 NOx and suggested the magnitude and statistical significance of the results are robust.
- Lower-emitting "Tier 3-like" vehicles are expected to show similar or greater sensitivity to the fuel sulfur levels compared to the conventional Tier 2 vehicles in-use.

The overall results of this study are in agreement with other studies conducted using low sulfur gasoline in Tier 2 vehicles. The magnitude of NO_x and HC reductions found in this study when switching from 28 ppm to 5 ppm fuel are consistent with those found in other studies done by the US EPA and automobile and catalyst manufacturers.[29,30,31]

A draft version of this report underwent an independent peer review covering the design, analysis methods, and results. This process was conducted according to guidelines described in EPA’s Peer Review Handbook, and did not produce any significant adverse findings. A detailed description of the process and results is available on the EPA Science Inventory website.[3]

9. References

---

[1] The Regulatory Impact Analysis for the Tier2/Sulfur Final Rule, Chapter III and Appendix B, EPA 420-R-99-023.

[2] The Regulatory Impact Analysis for the Control of Hazardous Air Pollutants from Mobile Sources Final Rule, Chapter 6, EPA 420-R-07-002.

[3] Four Peer Reviews in Support of the Tier 3 Rulemaking: Fuel Sulfur Effects Analysis Draft Report. U.S. Environmental Protection Agency, Washington, D.C. Documents available at http://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=240145.

[4] Hochhauser A.M. (2008). *Review of Prior Studies of Fuel Effects on Vehicle Emissions*, Report number E-84. Coordinating Research Council, Alpharetta, GA. Available at www.crcao.org.

[5] Heck R., Farrauto R.J., Gulati, S.T. (2009). *Catalytic Air Pollution Control: Commercial Technology* (3rd Ed.). John Wiley & Sons: Hoboken, NJ.

[6] Eastwood P. (2000) *Critical Topics in Exhaust Gas Aftertreatment (Engineering Design).* Research Studies Press Ltd: Hertfordshire, England.

[7] The Regulatory Impact Analysis for the Tier2/Sulfur Final Rule, Chapter III and Appendix B, EPA 420-R-99-023.

[8] Durbin T.D., *et al.* (2003). *The Effect of Fuel Sulfur on NH3 and Other Emissions from 2000-2001 Model Year Vehicles,* Appendix C. Report number E-60. Coordinating Research Council, Alpharetta, GA. Available at www.crcao.org.

[9] Ball D., Clark D., Moser D. (2011). *Effects of Fuel Sulfur on FTP NOx Emissions from a PZEV 4 Cylinder Application.* SAE 2011 World Congress Paper 2011-01-0300. SAE International: Warrendale, PA.

[10] Durbin T.D., *et al.* (2003). *The Effect of Fuel Sulfur on NH3 and Other Emissions from 2000-2001 Model Year Vehicles,* Report number E-60. Coordinating Research Council, Alpharetta, GA. Available at www.crcao.org.

[11] EPAct/V2/E-89: Assessing the Effect of Five Gasoline Properties on Exhaust Emissions from Light-Duty Vehicles Certified to Tier 2 Standards, Final Report on Program Design and Data Collection, EPA-420-R-13-004, April 2013.

[12] Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

[13] Engels, J.M.; Diehr, P. (2003). Imputation of missing longitudinal data: a comparison of methods. *Journal of Clinical Epidemiology,* 56 (2003), 968–976.

---

[14] Downey R.G., & King C.V. (1998). Missing data in Likert ratings: a comparison of replacement methods. *J Gen Psychol.* 125:175–191.

[15] Stephens, R.D. (1994). Remote sensing data and a potential model of vehicle exhaust emissions. *Journal of the Air Waste Management Association,* 44:1284-1292.

[16] Holmen, B.A., & Niemeier, D.A. (1998). Characterizing the effects of driver variability on real-world vehicle emissions. *Transportation Research Part D: Transport and Environment*, 3(2), 117-128.

[17] Beydoun, M., & Guldmann, J. (2006). Vehicle characteristics and emissions: Logit and regression analyses of I/M data from Massachusetts, Maryland, and Illinois. *Transportation Research Part D: Transport and Environment,* 11(1), 59-76.

[18] Diggle P.J. (1988). An approach to the analysis of repeated measures. *Biometrics,* 44:959-971.

[19] Wolfinger, R.D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation,* 22(4):1079-1106.

[20] Keselman, H.J., Algina, J., Kowalchuk, R.K., & Wolfinger, R.D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, 52:63-78.

[21] Deaton, M.I., & Winebrake, J.J. (2000). The Use of Mixed Effects ANCOVA to Characterize Vehicle Emissions Profiles. *J. Transp. Stat.* 3:49-64.

[22] Winebrake, J.J., Deaton, M.L., Coburn, T.C., & Kelly, K.J. (2000). *Statistical Analysis of Emissions and Deterioration Rates from In-Use, High Mileage CNG and Gasoline Ford Crown Victoria Taxicabs*; SAE Technical Paper Series No. 2000-01-1959; Society of Automotive Engineers: Warrendale, PA.

[23] Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag.

[24] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika,* 52:345-370.

[25] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika,* 52:345-370.

[26] Little, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. (1996). *SAS system for mixed models,* Cary, NC: SAS Institute.

[27] Wolfinger, R.D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics, Simulation and Computation,* 22(4):1079-1106.

---

[28] “Analysis of Particulate Matter Emissions from Light-Duty Gasoline Vehicles in Kansas City” (2008), Chapter 8. EPA420-R-08-010.

[29] Regulatory Impact Analysis for the Control of Hazardous Air Pollutants from Mobile Sources Final Rule, Chapter 6. EPA 420-R-07-002.

[30] Ball D., Clark D., Moser D. (2011). *Effects of Fuel Sulfur on FTP NOx Emissions from a PZEV 4 Cylinder Application.* SAE 2011 World Congress Paper 2011-01-0300. SAE International: Warrendale, PA.

[31] Shapiro, E. (2009). *National Clean Gasoline, An Investigation of Costs and Benefits.* Published by the Alliance of Automobile Manufacturers, Washington, DC.