Assessing the Effect of Five Gasoline Properties on Exhaust Emissions from Light-Duty Vehicles Certified to Tier 2 Standards:

Analysis of Data from EPAct Phase 3 (EPAct/V2/E-89)

Final Report

EPA United States Environmental Protection Agency

Assessing the Effect of Five Gasoline Properties on Exhaust Emissions from Light-Duty Vehicles Certified to Tier 2 Standards:

Analysis of Data from EPAct Phase 3 (EPAct/V2/E-89)

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-13-002
April 2013

Contents

Appendices

Executive Summary

Since the early 1990's, a large body of data has demonstrated that the properties of gasoline fuels have measurable effects on exhaust emissions from cars and trucks. Since that time, vehicle technologies have changed substantially and increasingly stringent emissions standards have been implemented, leading to marked reductions in exhaust emissions from motor vehicles. In model year 2004, cars and light trucks certified to Federal Tier 2 emissions standards entered the market. By 2017, we project that 70 percent of the car and light truck fleet will be comprised of Tier 2 vehicles, accounting for 80 percent of total vehicle miles travelled (VMT).

Existing fuel-effects models, such as the EPA Predictive Model and the Complex Model[1], were developed using data representing 1990s-technology vehicles meeting the Tier 0 and Tier 1 emission standards, levels an order of magnitude higher than current (Tier 2-compliant) vehicles[2]. With the fleet turning over to much lower-emitting vehicles, the Agency and stakeholders were interested in generating a coherent body of updated fuel-effects data, to provide the basis for generation of updated fuel effects models representing the gasoline vehicle fleet at the time of the study. In addition, in the Energy Policy Act of 2005 (EPAct), Congress required EPA to conduct the necessary research and develop updated models.

To carry out this effort, EPA entered a partnership with the Department of Energy (DOE) and the Coordinating Research Council (CRC) to undertake the largest fuels research program conducted since the Auto/Oil program in the early 1990s[3]. The program is aimed specifically at understanding the effects of fuel property changes on regulated and selected unregulated exhaust emissions from later technology vehicles certified to Tier 2 standards.

To allow estimation of selected fuel effects across their respective ranges, a statistically optimal study design was developed to represent variation in five fuel properties: ethanol volume, aromatic content, RVP, T50 and T90. These five parameters were selected based on previous studies as having potential to affect exhaust emissions[a]. Ethanol, T50 and T90 were included in the design at four, five and three levels, respectively, to allow assessment of potential nonlinear effects on emissions. The remaining two fuel properties, aromatic content and RVP, were measured at two levels each. A critical feature of the study design is that the properties of the test fuels are assigned to span the ranges of in-use fuel properties, with the intent of providing a basis for the development of statistical models capable of predicting emissions for the majority of in-use fuels.

An initial sample of 19 vehicles was chosen with the intent of representing the latest-technology light-duty vehicles sold at the time the program was launched (model year 2008). In terms of regulatory standards, the sample was to conform on average to Tier-2 Bin-5 exhaust levels and

[a] Sulfur also affects exhaust emissions, but due to its impact on vehicles' catalysts, it is necessary to assess the effects of sulfur separately from those of other fuel properties.

employ a variety of emission control technologies, realized by including a range of vehicle sizes and manufacturers. No additional criteria were used to select the individual test vehicles. Due to budget constraints, the sample was reduced from 19 to 15 vehicles for the Phase-3 program. A power analysis was performed using data from 15 vehicles retained from Phase 1, and results suggested a power in the range of 0.7-0.8 for detecting a 25 percent relative difference at a 95% confidence level. (During analysis, the confidence level was relaxed to 90%), increasing power to detect smaller effects. After considering sales levels, vehicle and engine sizes, representation of manufacturers and emissions sensitivity to ethanol, a set of 15 vehicles were used to generate the full dataset over the 27 test fuels.

Phase 3 data collection was completed in June 2010. Emissions measured include carbon dioxide (CO_2), carbon monoxide (CO), total hydrocarbons (THC), methane (CH_4), non-methane hydrocarbons (NMHC), oxides of nitrogen (NO_x), and particulate matter ($PM_{2.5}$). Emissions were measured on the LA92 test cycle at a nominal temperature of 75°F. In addition, hydrocarbons were speciated for subsets of vehicles, allowing calculation of derived parameters such as non-methane organic gases, as well as independent analyses of specific compounds including acetaldehyde, formaldehyde, acrolein, benzene and 1,3-butadiene.

Analysis

Following the completion of data collection, construction of the dataset involved intensive evaluation and quality assurance. Successive rounds of statistical modeling were applied to the data, to achieve several goals, including identification of potential candidate models, identification and review of outlying observations, identification and review of subsets of data from influential vehicles, and identification of models including subsets of terms that best explain the results obtained.

The analysis process involved ongoing consultation among EPA, DOE and CRC staff and contractors. However, it should be noted that this report describes analyses performed and conclusions reached by EPA independently of its partners, except where noted.

The models reported in this section are as parsimonious as the data and subject-matter knowledge allow. That is to say, they do not include all possible terms, but rather subsets of terms considered to give the best fit to the dataset. This approach was followed for several reasons: (1) the candidate fuel effects identified for inclusion were selected because we anticipated that they could be important for one or more emissions. However, we did not anticipate that all fuel properties would be meaningful for all the compounds selected for measurement. (2) Insofar as possible, it is highly desirable that the models selected be intelligible and interpretable in terms of physical and chemical processes, and that the models include effects describing important processes affecting emissions. (3) It is important to avoid “overfitting” of models by including terms that may prove to represent study artifacts or random variation.

Results for Regulated Emissions, Total Hydrocarbons and Methane

The Phase 3 study was conducted to assess the effects of fuel properties on the emissions of vehicles certified to Tier-2 standards, primarily the Bin-5 standards. Reviewing the results of statistical modeling the measured emissions, it is clear that such effects exist and are measurable.

It is important to note that the effects of different fuel properties are not cleanly separable. It is difficult to modify one property in an actual fuel without affecting one or more of the others. The study design and analysis of the data are structured so as to allow assessment of fuel effects *as though they were independent* of each other. However, in interpreting or applying the models, it is critical to *consider the effects of all five fuel properties in conjunction* with each other. Consideration of single coefficients in isolation can easily result in misleading conclusions.

Tables ES-1 and ES-2 summarize model coefficients for the regulated pollutants, total hydrocarbons and methane. The values in Table ES-1 represent results for "Cold-start" emissions, based on results for Bag 1 of the LA92 cycle. Similarly, the values in Table ES-2 represent results for "hot-running" emissions, based on results from Bag 2 of the LA92. Results for Bag 3 emissions are not presented, as review of results suggests that the models for Bag 3 may be less reliable than those in Bags 1 and 2, especially for PM and NO_x. In addition Figures ES-1 and ES-2 give qualitative summaries of the direction and size of the coefficients presented in the tables.

It is important to note that the coefficients represent abstract quantities that cannot be directly interpreted in terms of fuel properties themselves (% ethanol, psi, °F, etc.)[b]. However, the coefficients for different fuel properties can be directly compared, allowing assessment of the relative importance of the effects of the fuel properties on the emissions constituent modeled. A positive coefficient indicates an increase in emissions with an increase in the fuel property, or a decline in emissions with a decrease in the fuel property. Similarly, a negative coefficient indicates a decrease in emissions with an increase in the fuel property, or an increase in emissions with a decrease in the fuel property.

[b] The values presented are "standardized coefficients," representing the change in the natural logarithm of emissions due to a change in the fuel property of one standard deviation, calculated with reference to the fuel matrix used in the project.

Table ES-1. Models representing "Cold-start" Emissions for the Regulated Pollutants[1].

Model term	Notation	THC	CH_4	NMOG	NMHC	CO	NO_x[1]	PM
Intercept	Intercept	-0.8664	-3.0074	-0.95209	-1.0315	1.3466	-2.8594	0.6559
etOH	Z_e	0.0548	0.06994	0.08019	0.03094	-0.1049	0.06750	0.1582
Arom	Z_a	0.0676	-0.1053	0.08782	0.09461	-0.01242	0.1339	0.3833
RVP	Z_r	-0.0445	-0.03275	-0.04224	-0.04568	-0.00762		
T50	Z_5	0.1288	0.07554	0.1345	0.13689	-0.03273	0.04783	0.0550
T90	Z_9	0.0183			0.02160	-0.1571		0.2923
etOH × etOH	ZZ_{ee}	0.0436	0.02844	0.04432	0.04612	0.07304		
T50 × T50	ZZ_{55}	0.0736	0.05170	0.07579	0.07534	0.05358		0.0935
etOH × Arom	ZZ_{ea}	0.0179	0.02088	0.01693	0.02045	0.02086	-0.02369	
etOH × RVP	ZZ_{er}		0.01082			0.01596		
etOH × T50	ZZ_{e5}	0.0445	0.03048	0.04653	0.04729	0.1064		
etOH × T90	ZZ_{e9}	0.0214			0.02441			
Vehicle variance	σ^2_{veh}	0.1325	0.2855	0.1224	0.1266	0.3920	0.5925	0.4251
Residual error	σ^2_{ε}	0.06872	0.03014	0.07538	0.07624	0.07214	0.1458	1.0359

[1] Models fit on basis of 11-term design model, representing results for Bag 1 on LA92 cycle.
[2] Fit excluding the Ford Focus. See 6.1.1.

Table ES-2. Reduced Models representing "Hot-running" Emissions for the Regulated Pollutants[1].

Model term	Notation	THC[2]	CH_4	NMOG[2]	NMHC[2]	CO	NO_x[3]	PM
Intercept	Intercept	-4.6533	-5.7075	-5.2360	-5.3253	-1.3893	-4.5692	-1.3107
etOH	Z_e	0.0327	0.05860	0.02673			0.06299	0.1126
Arom	Z_a	-0.0195	-0.09836	0.03634	0.03987	0.0913	0.04407	0.1662
RVP	Z_r	-0.0355	-0.02049	-0.04786	-0.05881	0.0299		
T50	Z_5	0.0501	0.04394	0.04915	0.04548	0.0261		
T90	Z_9	0.0514	0.02575	0.07252	0.08202	0.0440		0.1072
etOH × etOH	ZZ_{ee}							
T50 × T50	ZZ_{55}	0.0337	0.01227	0.05349	0.04774			
etOH × Arom	ZZ_{ea}		0.008769	0.02171				
etOH × RVP	ZZ_{er}							
etOH × T50	ZZ_{e5}			0.02586				
etOH × T90	ZZ_{e9}							
Vehicle variance	σ^2_{veh}	0.8384	1.1108	0.8502	0.9691	1.9187	0.4720	0.7827
Residual error	σ^2_{ε}	0.06717	0.02518	0.1310	0.1708	0.1256	0.1836	1.1337

[1] Models fit on basis of 11-term design model, representing results for Bag 2 on LA92 cycle.
[2] Fit excluding the Honda Odyssey and Toyota Sienna. See 6.1.3.
[3] Fit excluding the Chevrolet Cobalt.

Figure ES-1. Qualitative Summary of the Sign and Magnitude of Linear-Effects Coefficients for "Cold-Start" (Bag 1) Reduced Models, based on the 11-term Design Model (NOTE: This figure does not attempt to represent interaction terms).

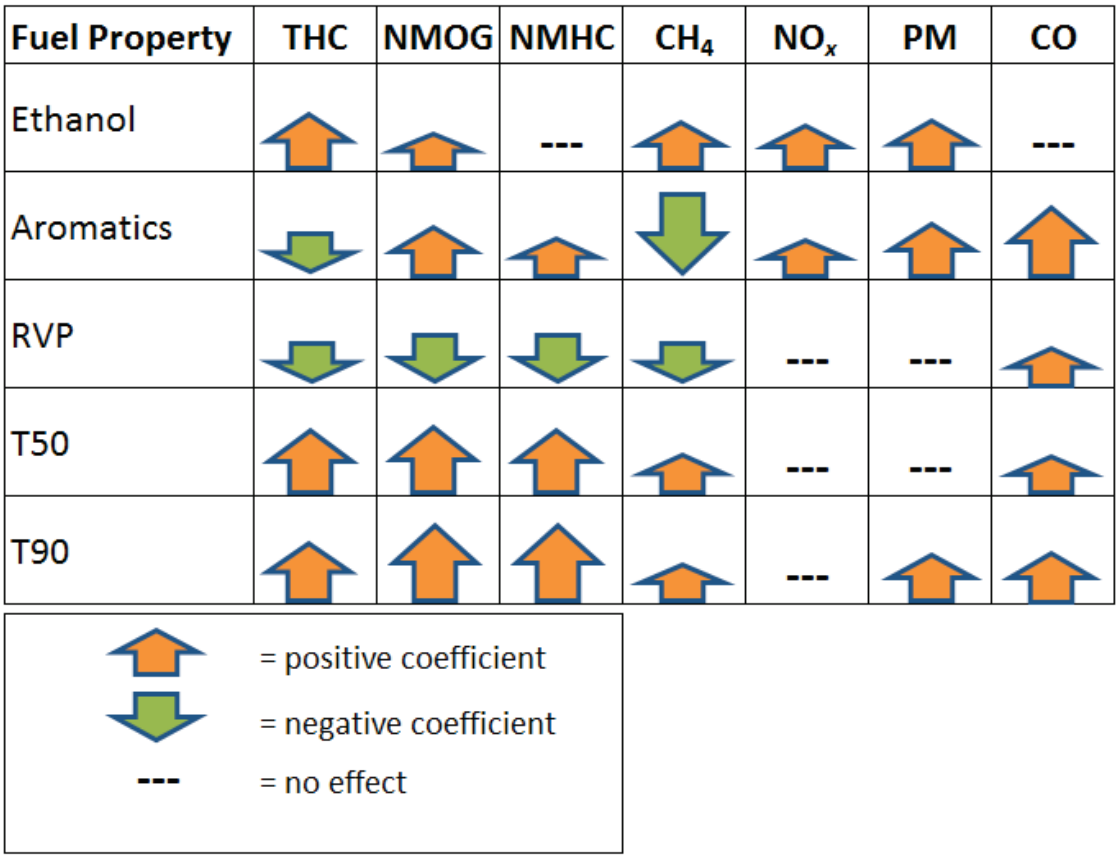
Fuel Property	THC	NMOG	NMHC	CH_4	NO_x	PM	CO
Ethanol	↑	↑	↑	↑	↑	↑	↓
Aromatics	↑	↑	↑	↓	↑	↑	↓
RVP	↓	↓	↓	↓	---	---	↓
T50	↑	↑	↑	↑	↑	↑	↓
T90	↑	---	↑	---	---	↑	↓

↑ = positive coefficient

↓ = negative coefficient

--- = no effect

Figure ES-2. Qualitative Summary of the Sign and Magnitude of Linear-Effects Coefficients for "Hot-running" (Bag 2) Reduced Models, based on the 11-term Design Model (Note: This figure does not attempt to represent interaction terms).



In reviewing the tables and figures, we can make some generalizations with respect to the individual fuel properties:

Ethanol: In most models, the linear-effect coefficients for ethanol are positive for both running and start emissions, implying that increases in ethanol content would be associated with increases in emissions (if the remaining fuel properties could be kept constant while increasing the ethanol level). A conspicuous exception to the pattern is CO, which has a negative coefficient for start emissions and no ethanol term for running emissions. Another exception is NMHC, which no ethanol term for running emissions. For start emissions, the etOH×etOH quadratic term is positive for all HC species and CO, imparting some curvature to ethanol trends for these species.

Aromatics: The patterns for aromatics are less consistent. Coefficients are positive for most models, with several exceptions for both start and running emissions. One exception is CO, which has a small negative coefficient for start emissions and a larger positive coefficient for running emissions. A second exception is THC, for which the start coefficient is positive and the

running coefficient negative. Thirdly, coefficients for CH_4 are large, negative and similar in size for both start and running emissions, which is unique in implying that changes in ethanol have similar relative effects on both start and running emissions. For start emissions, the interaction between aromatics and ethanol appears in all models except PM. The start interaction terms are consistent in size and positive in sign for all emissions except NO_x.

RVP: A linear term is included for all pollutants except NO_x and PM. The sign of the term is consistently negative with a single exception for running CO, which has a positive term. For the hydrocarbons, the size of the term is relatively consistent, although the coefficients for running models tend to be somewhat smaller than those for start emissions. The interaction with ethanol appears in two models for start emissions, but in no models for running emissions. In both start models, the terms are small and positive.

T50: Coefficients for this property consistently positive, with the single exception of start CO, and appears in all models except running NO_x and PM. For start emissions, the effects are largest for THC, NMOG and NMHC, and smaller for CH_4, NO_x and PM; for CO, the term is negative and relatively small. For the hydrocarbons except CH_4, T50 is the largest single term. For running emissions, coefficients are positive but smaller than for start emissions. For the hydrocarbon species, T50 shows a consistent reinforcement interaction with ethanol for start emissions, for running emissions, the interaction applies only to NMOG. For start CO, the interaction is present but acts as an interference, in that both linear terms are negative and the interaction is positive.

T90: This term is unique in that it appears more frequently in models for running than for start emissions, and in that it is sometimes larger in running models than start models. In the start models, the term is large and positive for PM, small and positive for THC and NMHC, large and negative for CO, and absent for the remaining models. In the running models, the term is large and positive for the hydrocarbons except methane, small and positive for PM and CO, and absent for NO_x. The interaction between T90 and ethanol is retained in only two models for start emissions, THC and NMHC, in which it is positive and similar in size. In both models, the linear and interaction terms are all positive, qualifying this effect as a reinforcement interaction. The T90 coefficient is largest for start PM, where it has a reinforcement interaction with the even stronger aromatics effect.

In addition, it is possible to make some general points about the responses of exhaust emissions to changing fuel properties that apply across the measured compounds and species, and for both start and running emissions.

- Other factors being equal, increasing ethanol is associated with an increase in emissions, as indicated by the positive ethanol coefficients in most models, both for running and start emissions.

- Other factors being equal, increasing volatility is associated with reductions in (exhaust) emissions, as indicated by generally negative coefficients for RVP (and generally positive coefficients for T50).
- In relative terms, fuel effects are generally more pronounced for start than for running emissions, as indicated by the fact that in most cases, the coefficients for Bag 1 models are larger than their counterparts for Bag 2 models, with exceptions for the ethanol coefficients for NO_x and the aromatics coefficients for CH_4. If we assume that we can validly make direct comparisons between coefficients between Bag-1 and Bag-2 models, this result may suggest that the effects of fuel properties are more pronounced during engine starts than during running operation. One interpretation might be that fuel effects could be damped by efficient operation of the catalyst after the engine comes up to temperature.

It is important to consider the applicability and representativeness of the results. As noted above the vehicle sample comprises a judgment sample of high-sales models from major manufacturers in model year 2008. In terms of standards, the vehicles represent the emissions standards that are most prevalent for light-duty vehicles, including Bins 3 and 5 (or equivalent LEV and ULEV standards under LEV-II), as well as a single Bin 8. The selection of makes and models does not qualify as a random sample, as limitations in the size of the study precluded drawing a reasonably sized random sample of makes and models. Nonetheless, given the size of the sample, it is likely that a well-designed judgment sample can perform as well as a random sample.

Results for Toxic Compounds

Summary results for all compounds representing “cold-start” (Bag 1) and “hot-running” (Bag 2) emissions are presented below in Tables ES-3 and ES-4. Qualitative summaries of the direction and size of the coefficients are also shown in Figures ES-3 and ES-4.

Table ES-3. Models representing "Cold-start" Emissions for Selected Air Toxics.

Model term	Notation	Compound						
		Acetaldehyde[1]	Formaldehyde[1]	Acrolein[1]	Ethanol[1]	Benzene[2]	1,3-Butadiene[2]	Ethane[2]
Intercept	*Intercept*	-5.2323	-5.9771	-7.9338	-4.9080	-4.1029	-5.8371	-4.3079
etOH	Z_e	0.81449	0.2299	0.2476	1.4627	-0.00468	-0.01729	0.1204
Arom	Z_a	0.03483	0.02822	0.1122		0.4056	0.02673	-0.1728
RVP	Z_r	-0.04170	-0.04718	-0.06450	-0.06054			
T50	Z_5	0.08670	0.1672	0.1880	0.07029	0.04242	0.1247	0.2169
T90	Z_9	0.03801	0.1302	0.2489	-0.09923	0.01133	0.1004	0.09531
etOH × etOH	ZZ_{ee}	-0.1669		-0.08310	-0.4970			
T50 × T50	ZZ_{55}	0.06665	0.05262		0.1108			
etOH × Arom	ZZ_{ea}	0.01840	0.01651					
etOH × RVP	ZZ_{er}	0.02194						
etOH × T50	ZZ_{e5}		-0.01627	-0.1186				
etOH × T90	ZZ_{e9}		0.02004	0.04617				
Arom × RVP	ZZ_{ar}							
Arom × T50	ZZ_{a5}							
Arom × T90	ZZ_{a9}							
T50 × T90	ZZ_{59}	0.03959	0.03489	0.05986				
RVP × T90	ZZ_{r9}							
Vehicle	σ^2_{veh}	0.1149	0.3358	0.1032	0.1283	0.2741	0.2192	0.1407
residual	σ^2_{ε}	0.0885	0.1407	0.3629	0.5730	0.1873	0.1089	0.04970

[1] Reduced models fit under the full design, including 15 vehicles measured on 27 fuels.
[2] Full models fit under the reduced design, including 15 vehicles measured on 11 fuels. Note that these models do not include a linear term for RVP, and do not include any 2nd –order terms.

Table ES-4. Models representing "Hot-Running" Emissions for Selected Air Toxics[1].

Model term	Notation	Compound						
		Acetaldehyde	Formaldehyde	Acrolein	Ethanol	Benzene	1,3-Butadiene	Ethane
Intercept	*Intercept*	-9.4189	-8.6574	NO MODEL	-9.3072	NO MODEL	NO MODEL	-7.7241
etOH	Z_e	0.1520	0.08456		0.9233			0.07345
Arom	Z_a	0.07991	0.01575		-0.3772			-0.1260
T50	Z_5	-0.02997	0.01863		-0.01910			0.1815
T90	Z_9	-0.07836	-0.08138		-0.3017			0.1322
Vehicle	σ^2_{veh}	0.05654	0.08205		0.3707			2.6785
residual	σ^2_{ε}	0.3814	0.3762		1.0889			0.1458

[1] Full models fit under the "reduced design," including 5 vehicles measured on 11 fuels.

Figure ES-3. Qualitative Summary of the Sign and Magnitude of Linear-Effects Coefficients for "Cold-Start" Emissions (Bag 1) (Note: This figure does not attempt to represent interaction terms).

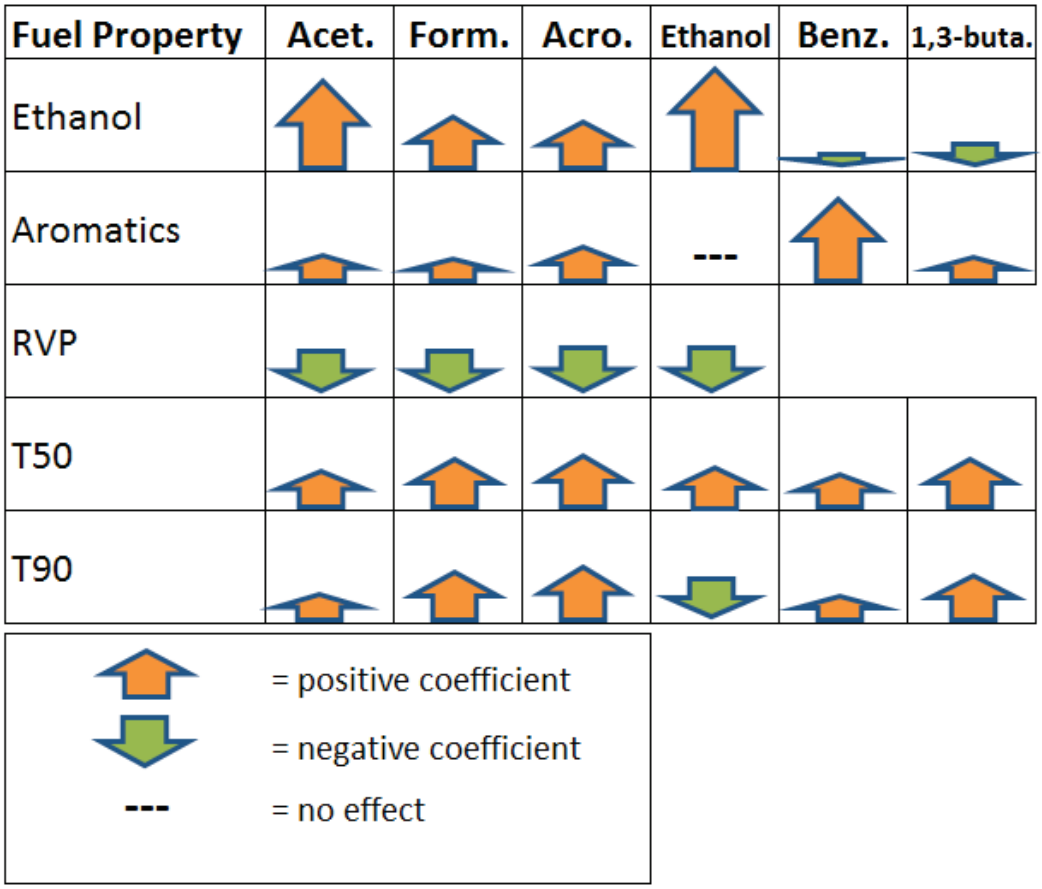
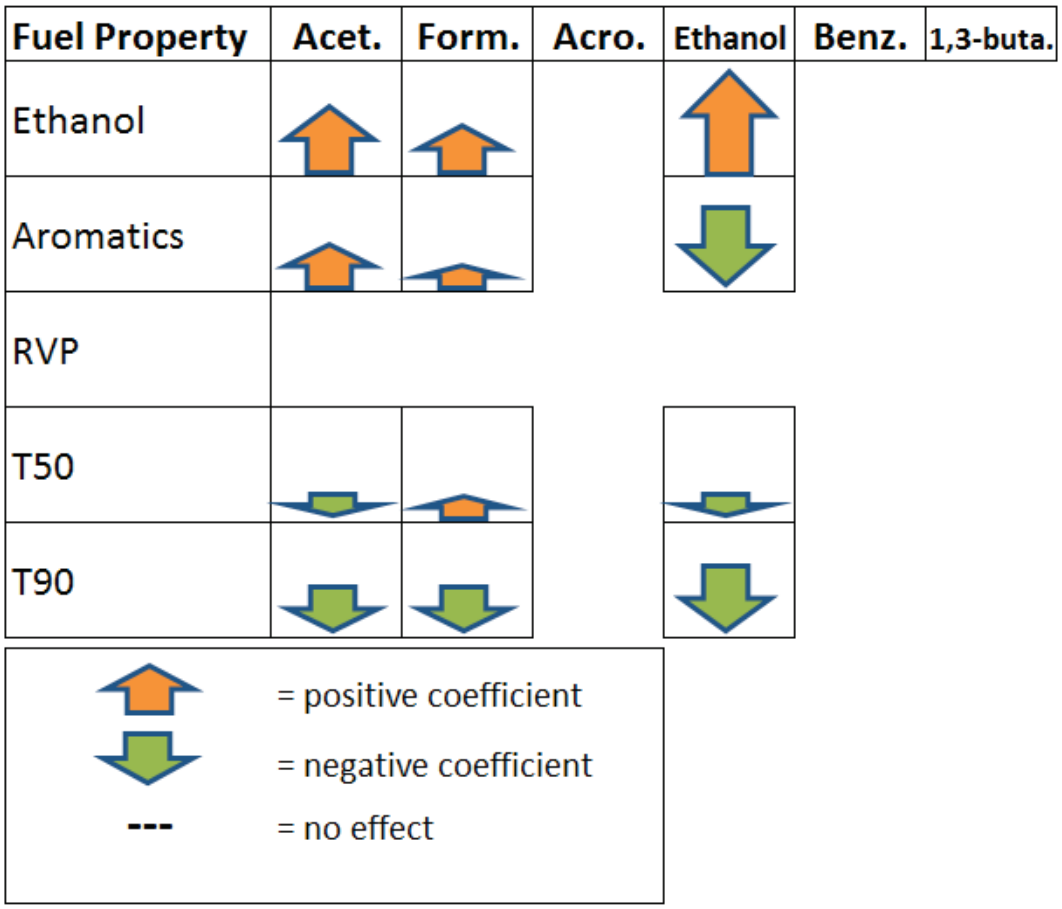


Figure ES-4. Qualitative Summary of the Sign and Magnitude of Linear-Effects Coefficients for "Hot-Running" Emissions (Bag 2) (Note: No models were fit for acrolein, benzene or 1,3-butadiene. This figure does not attempt to represent interaction terms).



The model results reflect the study design applied to each compound as well as the underlying physico-chemical processes. The reduced model structures are more complex for those compounds fit with the full design, specifically start emissions for the aldehydes, acrolein and ethanol (Table ES-3). These models are discussed in more detail below.

Ethanol. The ethanol coefficients are positive and large for the aldehydes, acrolein and ethanol. For acetaldehyde and ethanol, the ethanol effects are clearly dominant. These results are not surprising, given the structural affinity between acetaldehyde and ethanol, and that the strongest indicator of ethanol in the exhaust is ethanol in the fuel. For formaldehyde and acrolein, the ethanol coefficients are important but not as dominant. Neither benzene nor 1,3-butadiene retain ethanol coefficients in their reduced models. All compounds except formaldehyde retain large and negative etOH×etOH quadratic terms, which are clearly required to fit the downward curvature in the logarithmic trends.

Aromatics. In contrast to ethanol, the aromatics coefficients are small for the aldehydes, although several times stronger for acrolein. Ethanol does not retain an aromatics term in its reduced

model. Not surprisingly, the aromatics coefficient for benzene is large (Note that fuel benzene is also a strong predictor of exhaust benzene, but was not a target study parameter). The two aldehydes retain small but significant reinforcement interactions between aromatics and ethanol.

RVP. The sign and size of RVP coefficients are similar for all four compounds fit under the full design (but absent for those fit under the reduced design). As with the RVP terms in the models for aggregated hydrocarbons (THC, NMHC and NMOG), the signs of the RVP linear effects are negative and similar in size to those for the aggregate HC (-0.04 to -0.06). The interaction between ethanol and RVP is retained only in the acetaldehyde model, in which it is positive and small.

T50. For the four compounds fit under the full design, linear-effect coefficients for T50 are positive. However, the pattern in the size of the coefficients mirrors that for ethanol, in that the two compounds with largest ethanol coefficients (acetaldehyde and ethanol) have smaller T50 coefficients than formaldehyde and acrolein, which have T50 coefficients about twice as large. These results may reflect similarities in structure between the two pairs of compounds, or similarities in formation processes during combustion. In addition to large linear coefficients, formaldehyde and acrolein have small interference interactions between T50 and ethanol.

T90. More so than for the other properties, linear coefficients for T90 differ among the compounds fit under the full design. The coefficients for acetaldehyde, formaldehyde and acrolein are positive, but increasing, respectively, with the values for formaldehyde and acrolein approximately 3 and 8 times larger than that for acetaldehyde. In contrast, the coefficient for ethanol is negative, suggesting reduced ethanol emissions for less volatile fuels. In addition to large linear effects for ethanol and T90, formaldehyde and acrolein have small reinforcement interactions between these properties.

The structures for reduced models are much simpler for benzene, 1,3-butadiene and ethane, reflecting the limits imposed by the reduced design. It is clear that in model fitting for these compounds that only strong effects appear significant and are hence retained in the reduced models.

Corresponding sets of coefficients for the hot-running models are shown in Table ES-4. As with three of the cold-start models, these models are simpler, having been fit to a smaller data set (fewer vehicles and fuels). For these emissions, it was not possible to fit RVP effects for any compound, nor did we attempt to fit quadratic or interaction terms. For acrolein, benzene and 1,3-butadiene, no model fitting was attempted, given that large numbers of measurements for these compounds were lower than background levels. Intercepts are much lower than for cold start, showing the much lower emission levels during hot-running operation. The relative sizes of the ethanol effects on the aldehydes and ethanol emissions for running are similar to those for starts, i.e., ethanol emissions are affected most, followed by acetaldehyde and formaldehyde.

1 Introduction

1.1 Mandate and Scope

Since the early 1990's, a wealth of data has been collected, demonstrating that the properties of gasoline fuels, including aromatics, olefins, oxygenates, vapor pressure and distillation parameters had measurable effects on exhaust emissions from cars and trucks. Since that time, vehicle technologies have changed substantially and increasingly stringent emissions standards have been implemented. The net result of these two factors has been marked reductions in exhaust emissions from motor vehicles.

In model year 2004, cars and light trucks certified to Federal Tier 2 emissions standards (or their equivalents under California LEV-II) entered the market, with phase-in of the standards slated for completion by MY 2010. By 2017, we project that 70 percent of the car and light truck fleet will be comprised of Tier 2 vehicles, accounting for 80 percent of total vehicle miles travelled (VMT). Uncertainty remained as to whether vehicles employing improved technologies and certified to the new standards would respond to fuel property changes in ways similar to vehicles employing older technologies and certified to less stringent standards.

Prior fuel-effects models, such as the EPA Predictive Model and the Complex Model[1], were developed using data representing 1990s-technology vehicles meeting the Tier 0 and Tier 1 emission standards, levels an order of magnitude higher than current (Tier 2-compliant) vehicles[2]. With the fleet turning over to much lower-emitting vehicles, the Agency and stakeholders were interested in generating a coherent body of updated fuel-effects data, to provide the basis for generation of updated fuel effects models representing the gasoline vehicle fleet at the time of the study. In addition, in the Energy Policy Act of 2005 (EPAct), Congress required EPA to conduct the necessary research and develop updated models.

To carry out this effort, EPA entered a partnership with the Department of Energy (DOE) and the Coordinating Research Council (CRC) to undertake the largest fuels research program conducted since the Auto/Oil program in the early 1990s[3]. The program is aimed specifically at understanding the effects of fuel property changes on regulated and selected unregulated exhaust emissions from later technology vehicles certified to Tier 2 standards. The resulting study was dubbed the "EPAct/V2/E-89" program, with the three components of the label representing the designations of the study by the three partners, EPA, DOE and CRC, respectively.

The program was conducted in three phases. Phases 1 and 2 were pilot efforts involving measurements on 19 light-duty cars and trucks on three fuels, at two temperatures[4]. This work was completed at Southwest Research Institute between September 2007 and January 2009. The preliminary efforts in Phases 1 and 2 laid the groundwork for design of a full-scale research program. The full-scale program, involving the incorporation of experimental design, is designated as Phase 3.

This report describes the analysis of the dataset collected in Phase 3 of the EPAct/V2/E-89 program, conducted at Southwest Research Institute in San Antonio, Texas. A separate report describing the program design and data collection activities is available[5], but an overview is provided below.

The analysis process involved ongoing consultation and collaboration among EPA, DOE and CRC staff and contractors. However, it should be noted that this report describes analyses performed and conclusions reached by EPA independently of its partners, except where noted.

1.2 Development of the Fuel Matrix

To allow estimation of selected fuel effects across their respective ranges, a fuel matrix was developed to represent variation in five fuel properties: ethanol volume, aromatic content, RVP, T50 and T90. These five parameters were selected based on previous studies as having potential to affect exhaust emissions[c]. Prior studies also showed olefin content as playing an important role. However, funding limitations precluded expanding the fuel matrix to include a sixth parameter.

Some fuel parameters have nonlinear impacts on some emissions. To capture this behavior, three or more treatment levels of a given parameter must be included in the study design. Statistical models of data from prior studies suggested that T50, T90, and ethanol content may have nonlinear impacts on emissions. With support from DOE, the fuel matrix included four levels of ethanol (0, 10, 15, and 20 percent by volume). In addition to the potential for nonlinear emission impacts of T50 on emissions, five levels of T50 were also chosen to allow detailed characterization of its relations with ethanol. Finally, due to concerns over potential nonlinear effects of T90, CRC contributed additional funding to add fuels representing a third level of this parameter.[d] The remaining two fuel properties, aromatic content and RVP, were measured at two levels each.

A critical point about the design of the program is that the properties of the test fuel are assigned so as to span the boundaries of in-use fuel properties. This approach is designed specifically to provide a basis for the development of statistical models capable of predicting emissions for the majority of in-use fuels. The parameter ranges to be covered for T50, T90, aromatic content, and RVP were selected to represent the range of in-use fuels based on a review of the Alliance of Automobile Manufacturers' 2006 North American Fuel Survey. As the emissions tests were to be performed at a nominal temperature of 75°F, summer survey data was used. Test fuel parameter ranges were originally drafted to span roughly the 5th to 95th percentiles of survey

[c] Sulfur also affects exhaust emissions, but due to its impact on vehicles' catalysts, it is necessary to assess the effects of sulfur separately from those of other fuel properties.

[d] The intermediate level of T90 occurs along one edge of the fuel domain in Phase 3. Statistical analysis of nonlinear T90 effects was intended to include a fuel used in Phase 1 of the program as an additional source of data for the intermediate T90 level.

results for U.S. gasoline, though some test fuel parameters were adjusted after the actual fuel-blending process began. An intermediate level of T50 in E0 fuels was selected to coincide with the high level of T50 in E10 fuels. Similarly, an intermediate level of T50 in E10 fuels was selected to coincide with the low level of T50 in E0 fuels.

For E15 and E20 fuels, aromatics, RVP and T90 ranges selected for E0 and E10 fuels were applied. A single level of T50 was selected for E20 blends based on the information obtained from a report compiled by the Coordinating Research Council[6], as well as petroleum industry sources which indicated that it was largely independent of the hydrocarbon fraction of the fuel and would not deviate more than several degrees from 160°F due to the presence of a large fraction of ethanol.[e] At the time this fuel matrix was designed, no information was available on distillation properties of E15 fuels. Two levels of T50 were selected for the E15 fuels, the low level equal to the lowest T50 assumed for E10 fuels and the high level being a linear interpolation between the highest T50 of E10 fuels and the sole T50 level of E20 fuels.

1.3 Selection of the Vehicle Sample

An initial sample of 19 vehicles was chosen with the intent of representing the latest-technology light-duty vehicles sold at the time the program was launched (model year 2008). In terms of regulatory standards, the sample was to conform on average to Tier-2 Bin-5 exhaust levels and employ a variety of emission control technologies, realized by including a range of vehicle sizes and manufacturers.

Engine family sales data obtained from EPA certification and Wards databases was analyzed to generate a list of high-sales vehicles as candidates for inclusion.[f] Grouping sales data by engine family allowed additional transparency and flexibility in choosing test vehicles that represent a wider group than one specific make and model. No additional criteria were used to select the individual test vehicles for lease.

Due to budget constraints, the sample was reduced from 19 vehicles for the Phase-3 program. A power analysis was performed using data from 15 vehicles retained from Phase 1, and results suggested a power in the range of 0.7-0.8 for detecting a 25 percent relative difference at a confidence level of 0.05. (During analysis, relaxing the confidence level to 0.10 effectively increases power for the effect of the same size, and increases power for smaller effects.

---

[e] As ethanol blend level moves beyond 10 vol%, T50 becomes increasingly correlated (inversely) with ethanol content. At E15, the two can be manipulated independently with some effort within a relatively limited range. By E20, the behavior of the center of the distillation curve (where T50 lies) is dominated by ethanol's boiling point, and thus T50 cannot be moved outside a narrow range around 165°F. Thus, T50 and ethanol should only be understood to be independently blended parameters at E10 and below.

[f] Engine family (or "test group") is a term used in manufacturing and certification to describe a combination of a base engine and after-treatment system that may be used in several vehicle makes and models offered by a manufacturer.

Statistical analysis of Phase 1 data found significant fuel effects smaller than 25 percent, so assigned relative difference shouldn't be understood as a lower limit of detectable effects, but rather as a screen for the largest effect that is unlikely to be missed at the assigned power level.)

Reduction of the sample required choosing four vehicles to eliminate. Primary considerations in this process included retaining high-sales engine families, a balance of vehicle and engine sizes, and maintaining representation of all manufacturers originally included in order to represent a range of technologies and emission control strategies. There was also consideration of the fact that changes in the sample could shift the average program results. To explore this issue, all nineteen vehicles were ranked according to their NO_X and NMHC sensitivity to fuel ethanol level based on the Phase 1 data, with the intent of avoiding removal of several vehicles with similar emissions behavior (though the two pollutants could provide conflicting direction). In the end, a set of 15 vehicles were used to generate the full dataset over the 27 test fuels.

Table 1. Candidate Vehicles for the Phase-3 EPAct Program; all vehicles in MY2008 (the four highlighted Vehicles not included in the Phase 3 Vehicle Sample).

Make	Brand	Model	Program ID	Engine Size	Engine Family	Tier 2 Bin	LEVII Std	Phase 3 Starting odometer
GM	Chevrolet	Cobalt	CCOB	2.2L I4	8GMXV02.4025	5	NA	4,841
GM	Chevrolet	Impala FFV	CIMP	3.5L V6	8GMXV03.9052	5	L2	5,048[1]
GM	Saturn	Outlook	SOUT	3.6L V6	8GMXT03.6151	5	L2	5,212[1]
GM	Chevrolet	Silverado FFV	CSIL	5.3L V8	8GMXT05.3373	5	NA	5,347[2]
Toyota	Toyota	Corolla	TCOR	1.8L I4	8TYXV01.8BEA	5	U2	5,019[1]
Toyota	Toyota	Camry	TCAM	2.4L I4	8TYXV02.4BEA	5	U2	4,974[2]
Toyota	Toyota	Sienna	TSIE	3.5L V6	8TYXT03.5BEM	5	U2	4,997
Toyota	Toyota	Tundra	TTUN	4.0L V6	8TYXT04.0AES	5		
Ford	Ford	Focus	FFOC	2.0L I4	8FMXV02.0VD4	4	U2	5,150[1,2]
Ford	Ford	Taurus	FTAU	3.5L V6	8FMXV03.5VEP	5		
Ford	Ford	Explorer	FEXP	4.0L V6	8FMXT04.03DB	4	NA	6,799[3]
Ford	Ford	F150 FFV	F150	5.4L V8	8FMXT05.44HF	8	NA	5,523[1]
Chrysler	Dodge	Caliber	DCAL	2.4L I4	8CRXB02.4MEO	5	NA	4,959
Chrysler	Dodge	Caravan FFV	DCAR	3.3L V6	8CRXT03.3NEP	8		5,282[4]
Chrysler	Jeep	Liberty	JLIB	3.7L V6	8CRXT03.7NE0	5	NA	4,785
Honda	Honda	Civic	HCIV	1.8L I4	8HNXV01.8LKR	5	U2	4,765
Honda	Honda	Accord	HACC	2.4L I4	8HNXV02.4TKR	5		
Honda	Honda	Odyssey	HODY	3.5L V6	8HNXT03.54KR	5	U2	4,850
Nissan	Nissan	Altima	NALT	2.5L I4	8NSXV02.5G5A	5	L2	5,211[2]

[1] These vehicles added to the Phase 3 sample at a later date. Prior to inclusion, they received on-road mileage accumulation every other week.
[2] These vehicles included in an FTP interim program conducted between Phases 1 and 2.
[3] During Phase 1, the initial 4,000 miles of vehicle break-in was conducted with crankcase lubricant of the wrong viscosity grade. An additional 2,000-mile break-in was conducted with the correct lubricant.
[4] This vehicle was measured only on E85 fuel.

1.4 Study Execution

In September 2007, Southwest Research Institute (SwRI) began work at their facilities in San Antonio, Texas. By January 2009, SwRI had completed the pilot phases of the program (referred

to as Phases 1 and 2). These phases involved testing of the 19 light duty cars and trucks from Table 2-3 on three fuels, at two temperatures.

In March 2009, SwRI began Phase 3 of the program (also referred to as the full program) This report covers work conducted for Phase 3, also known as EPAct/V2/E-89, which involved the testing of the 15 vehicles listed in Table 1 (not shaded) on the set of 27 test fuels.[g] Phase 3 data collection was completed in June 2010.

Emissions measured include carbon dioxide (CO_2), carbon monoxide (CO), total hydrocarbons (THC), methane (CH_4), non-methane hydrocarbons (NMHC), oxides of nitrogen (NO_x), and particulate matter ($PM_{2.5}$). Emissions were measured on the LA92 test cycle at a nominal temperature of 75°F. In addition, hydrocarbons were speciated for subsets of vehicles, allowing calculation of derived parameters such as non-methane organic gases, as well as independent analyses of specific compounds including acetaldehyde, formaldehyde, acrolein, benzene and 1,3-butadiene.

1.5 Overview of the Report

This report describes the study design, dataset construction and subsequent analyses of data obtained in the Phase-3 dataset.

Chapter 2 describes the design of the Phase 3 program. In addition to the considerations discussed in 1.2 above, development of the fuel matrix involved the application of optimal experimental design to develop a matrix allowing for estimation of targeted fuel effects with the maximum attainable precision within budgetary and technical constraints. The experimental design entailed an iterative process that involved balancing the study goals of the three partners, (EPA, DOE and CRC) with technical limitations of fuel blending, in the context of experimental design. This process is summarized in 2.1.

Section 2.2 identifies the specific sets of vehicles and fuels for which speciation of hydrocarbon measurements was performed.

Section 2.3 describes an assessment of the degree of correlation among the fuel parameters in the dataset. The study design effectively neutralized correlations among the fuel parameters themselves, as intended. However, when quadratic or interactive terms are constructed from the linear terms, some strong correlations result. As strong correlation among predictors can adversely affect statistical models, we applied analytical techniques to neutralize the additional correlations, allowing the analysis of fuel effects to proceed.

Chapter 3 describes specific topics involved in the construction of the dataset. One important question, discussed in 3.1, was whether to base analyses of fuel effects on aggregate “Bag”

[g] Phase 3 also included testing of four flexible-fuel vehicles (FFVs) from the 15-vehicle sample on an E85 fuel. Analysis of results obtained on the E85 fuel is not discussed in this document.

measurements, on summed continuous measurements, or on some combination of the two. The underlying issue concerned whether aggregate measurements became unreliable at the low end of the measurement range, and whether the use or substitution of summed continuous measurements would represent an improvement in data quality. After some consideration, study participants elected to rely on the aggregate data, while applying appropriate techniques to address the resulting "censoring" of the data at low end of the range of values.

A second issue was raised by the fact that only subsets of vehicles and fuels were speciated in Bags 2 and 3. An implication of the situation was that the advantages of the full study design would not be available in analysis of speciation-dependent parameters such as NMOG and NMHC for the hot-running portions of the test. Given the importance of these parameters in analysis and interpretation of the results, we elected to statistically impute missing measurements from a related measure, "NMHC as measured by FID," ($NMHC_{FID}$), to which both NMOG and NMHC are very strongly correlated. These analyses are described in 3.2.

Prior to analysis and modeling, it is important to gain familiarity with the datasets. To achieve this aim, the data was plotted in raw and aggregate forms. To assess the existence of "linear" or "main effects," we averaged and plotted the data by each fuel property and by fuel. These plots indicated whether specific fuel effects seemed evident when the data was averaged across the remaining fuel properties. As an initial indication of the possibility of interactions between fuel properties, we constructed "conditional effects" plots by averaging the data by two selected fuel properties, repeating this step for multiple pairs of properties. Plotting and review for three subsets of data: NO_x (Bag 1), NO_x (Bag 2) and PM (Bag 1) were selected for purposes of illustration in this report. Similar plots for the remaining compounds are presented in Appendices G-Q.

Chapter 5 describes initial modeling performed for the purposes of data exploration and influence analysis. Influence analysis was used to identify influential observations and vehicles for detailed review. This chapter also describes methods and criteria adopted for model fitting, including issues such as treatment of outlying observations and the existence of "censoring" in the data, which addresses loss of observations at the low end of the range due to limitations in emissions measurement techniques.

Chapter 6 discusses measurement issues at the low end of the range of emissions highlighted by the influence analysis for vehicles. Specific issues included the validity of sample measurements falling into the range of background measurements, and the apparent drift of some measurements during the course of such a lengthy project.

Chapter 7 covers a second round of model fitting, incorporating the findings of the analyses in Chapters 5 and 6. This chapter applies the methods described in Chapter 5 to the development of a set of "reduced models" which include subsets of terms contributing to the fit of the models to the data, but excluding terms not found to significantly improve fit. This chapter also illustrates

a more detailed review of models fit to the data, illustrated using the examples of NO_x (Bag 1) and CO (Bag 1).

Chapter 8 describes analyses for selected speciated hydrocarbons. These compounds included aldehydes (acetaldehyde, formaldehyde and acrolein), ethanol, benzene, 1,3-butadiene and ethane. The methods used were very similar to those used for modeling the other emissions, with modifications to address issues of study design and measurement limitations specific to these compounds.

Chapter 9 summarizes results of the model fitting and analysis based on reduced models presented in Chapters 7 and 8.

2 Study Design

The design and implementation of the study, including the aspects of fuel blending, measurement methods and logistics are described in a separate report[5]. This section focuses on the experimental design of the fuel matrix.

2.1 Design Optimization

The EPAct Phase-3 program was conducted as a controlled experiment, for which the plan was to analyze results in terms of fuel properties, which are abstracted from actual fuels and treated as continuous numeric variables. Five fuel parameters were selected as experimental factors, specifically: Ethanol content (%), Aromatics content (%), Reid Vapor Pressure (RVP, psi)[h], and two distillation parameters, T50 (°F) and T90 (°F). The first two parameters represent the chemical composition of the fuels, and the remaining three represent commonly measured bulk physical parameters.

It is well known for fuel properties to be moderately to strongly correlated. This tendency stems from the fact that it is impossible to modify one factor without also affecting one or more of the others. As the goal is to enable analysis of fuel effects as though independent, and as statistical models assume independence of factors and can be adversely affected by strong correlations among factors, it is necessary to address these correlations in design and analysis. An important implication for experimental design is that orthogonal factorial designs (full or partial) cannot be constructed.

In such cases, which are not unusual in real applications, it is common practice to construct "optimal" designs, generated by iterative computer algorithms and based on specific criteria[7]. It is important to recognize that optimized designs are not unique – algorithms produce multiple solutions to specific problems. It is also important to remember that in optimized designs the

[h] This parameter was measured as Dry Vapor Pressure Equivalent (DVPE), but for simplicity and consistency, we will refer to it as RVP, which is numerically equivalent.

parameters are approximately rather than fully orthogonal and that effects can be somewhat correlated rather than uncorrelated.

In optimizing designs, it is necessary to specify a model to be analyzed using the study results, as well as a criterion for evaluating the "efficiency" of the optimized design, relative to an orthogonal fractional factorial. The criterion selected for design development for this study is "*G* efficiency". This parameter attempts to minimize the maximum standard error for prediction over the design points for the specified model. Efficiency is expressed as a percent, under the assumption that a fractional factorial represents 100%.

For the five fuel parameters selected, design points were constructed based on the levels shown in Table 2. A full factorial, based on this design, would include $4\times2^2\times5\times3 = 240$ fuels. Given the level of effort and expense involved in vehicle emissions measurement, and limitations in resources, it was not practical to perform measurements on this number of fuels. In addition, due to relationships among fuel parameters, many fuels in the full factorial design either do not exist or are not feasible to blend.

Table 2. Levels assigned to Experimental Factors (Fuel parameters) for the Phase-3 EPAct program.

Factor	No. Levels	Levels		
		Low	Middle	High
Ethanol (%)	4	0	10, 15	20
Aromatics	2	15		35
RVP (psi)	2	7		10
T50 (°F)	5	150	165, 190, 220	240
T90 (°F)	3	300	325	340

The final design is the result of an iterative process involving balancing among research goals, fuel blending feasibility, and experimental design. The design was developed through a series of steps, described below:

Step 1: An initial design was optimized for a domain covering an ethanol range from 0-10%. Out of a possible total of 64 possible fuels (in a full factorial), 20 were dropped due to issues with blending feasibility. Of the 44 remaining fuels, the optimization gave a reduced set of 16 fuels, with *G*-efficiency estimated at 73%. In this step, the response model input to the optimization process includes 10 terms and is shown in Equation 1.

$$Y = \beta_0 + \beta_1 \text{etOH} + \beta_2 \text{Arom} + \beta_3 \text{RVP} + \beta_4 \text{T50} + \beta_5 \text{T90} + \beta_6 \text{T50}^2 + \beta_7 \text{etOH}\times\text{Arom} + \beta_8 \text{etOH}\times\text{RVP} + \beta_9 \text{etOH}\times\text{T50} + \beta_{10} \text{etOH}\times\text{T90} + \varepsilon \qquad \textbf{Equation 1}$$

Step 2: The initial design was augmented to include a second domain covering an ethanol range from 10-20%, while keeping the original 16-fuel design intact, and adding a quadratic term for ethanol. At this point, we imposed additional design constraints: (1) that nine additional fuels be added, to give a total of 25 fuels, and (2) that three of the nine fuels have 15% ethanol, and that an additional three have 20% ethanol. To meet these constraints, it was not possible to assess this modification through an optimization algorithm, as in the previous step. Rather, all sets of fuels meeting these criteria were added to the design, and the efficiency evaluated for each whole 25-fuel design. The design selected had a *G*-efficiency of 69%.

$$Y = \beta_0 + \beta_1 \text{etOH} + \beta_2 \text{Arom} + \beta_3 \text{RVP} + \beta_4 \text{T50} + \beta_5 \text{T90} + \beta_6 \text{T50}^2 + \beta_{11} \text{etOH}^2 \beta_7 \text{etOH} \times \text{Arom} + \beta_8 \text{etOH} \times \text{RVP} + \beta_9 \text{etOH} \times \text{T50} + \beta_{10} \text{etOH} \times \text{T90} + \varepsilon$$

Equation 2

Step 3: a third quadratic term was added to the design model (T90×T90). Three fuels with pre-specified parameters were added, plus two additional fuels to maximize efficiency, to give a total of 30 fuels.

Step 4: As a final step, three fuels were removed from the 30-fuel matrix, to give a final total of 27 fuels. In addition, the exact fuel parameter levels assigned to certain fuels were modified to improve blending feasibility. The *G*-efficiency of this modified design was evaluated at 51.6%. This final fuel set was adopted for the Phase-3 EPAct program and is shown in Table 3. Note that the values of the fuel parameters in the table are nominal target values, rather than actual measured values, which varied slightly from the targets. Actual measured values, which were used in the analysis, are shown in Table 4.

Table 3. Nominal Target parameters for fuels in the Phase-3 EPAct program.

Fuel[1]	12-fuel subset[2]	etOH (%)	Aromatics (%)	RVP (psi)[3]	T50 (°F)	T90 (°F)
1		10	15	10	150	300
2		0	15	10	240	340
3	●	10	15	7	220	300
4	●	10	15	10	220	340
5		0	35	7	240	300
6	●	10	15	7	190	340
7	●	0	15	7	190	300
8		0	15	10	220	300
9		0	35	10	190	340
10	●	10	35	7	220	340
11		10	35	10	190	300
12		10	35	10	150	340
13	●	0	35	7	220	340
14	●	0	15	7	190	340
15		0	35	10	190	300
16		10	35	7	220	300
20		20	15	7	165	300
21	●	20	35	7	165	300
22		20	15	10	165	300
23	●	20	15	7	165	340
24		20	15	10	165	340
25		20	35	10	165	340
26		15	35	10	165	340
27	●	15	15	7	220	340
28	●	15	35	7	220	300
30		10	35	10	150	325
31	●	20	35	7	165	325

[1] Note that numbering of fuels is not entirely sequential throughout.
[2] Speciation performed for these fuels differs from that performed for the fuels not in the subset. See Table 5.
[3] This parameter was measured as "DVPE," but for simplicity, will be referred to as "RVP" in this document.

Table 4. Measured Parameters for Fuels in the Phase-3 EPAct Program.

Fuel[1]	etOH (%)	Aromatics (%)	RVP (psi)[2]	T50 (°F)	T90 (°F)
1	10.03	15.4	10.07	148.9	300.2
2	0	14.1	10.2	236.7	340.1
3	10.36	15.0	6.93	217.5	295.9
4	9.94	15.5	10.01	221.9	337.5
5	0	34.7	6.95	237.0	300.0
6	10.56	15.0	7.24	188.5	340.4
7	0	17.0	7.15	193.1	298.4
8	0	15.7	10.2	221.1	303.1
9	0	35.8	10.30	192.8	341.8
10	9.82	34.0	7.11	217.1	340.2
11	10.30	35.0	9.93	189.3	298.6
12	9.83	34.8	10.13	152.2	339.8
13	0	34.1	6.92	222.5	337.9
14	0	16.9	7.14	192.8	338.5
15	0	35.3	10.23	189.7	299.4
16	10.76	35.6	7.12	218.8	300.6
20	20.31	15.2	6.70	162.7	298.7
21	21.14	35.5	7.06	167.6	305.0
22	20.51	15.0	10.21	163.2	297.3
23	20.32	15.9	6.84	162.5	338.2
24	20.51	15.3	10.12	165.1	338.1
25	20.03	35.2	10.16	166.9	337.9
26	15.24	35.6	10.21	160.3	338.7
27	14.91	14.9	6.97	221.5	340.3
28	14.98	34.5	6.87	216.6	298.8
30	9.81	35.5	10.23	152.9	323.8
31	20.11	35.5	6.98	167.3	325.2

[1] Note that numbering of fuels is not entirely sequential throughout.

[2] This parameter was measured as "DVPE," but for simplicity, will be referred to as "RVP" in this document.

2.2 Measurements

2.2.1 Regulated Emissions, Total Hydrocarbons and Methane

The emissions measured on all vehicles and test fuels included carbon dioxide (CO_2), carbon monoxide (CO), total hydrocarbons (THC), methane (CH_4), oxides of nitrogen (NO_x), and particulate matter ($PM_{2.5}$). Measurements were conducted on the LA92 cycle at a nominal temperature of 75%. Measurement methods are discussed in detail in the testing report[5]. To allow calculation of derived measurements such as non-methane hydrocarbons (NMHC) and non-methane organic gases (NMOG), hydrocarbons were speciated for subsets of vehicles and fuels, as described below.

2.2.2 Hydrocarbon speciation

Due to the additional time and expense required, speciation was applied to results from a subset of fuels. For the selected fuels, measurements from a subset of five vehicles were speciated. Alcohols and carbonyls were speciated during bag 1 for all tests (all vehicles, fuels and replicates). For the first replicate test, C_1-C_{12} hydrocarbons were speciated for all vehicles while operated on a subset of twelve fuels which included Fuels 3, 4, 6, 7, 10, 13, 14, 21, 23, 27, 28, and 31. This subset was selected to provide, as nearly as possible, useful comparisons between differing levels of ethanol, aromatics, T50 and T90. In addition, all types of speciation were conducted for Bags 2 and 3 on a subset of five vehicles over the 12-fuel subset. The vehicle subset included the Honda Civic, Toyota Corolla, Chevrolet Impala, Ford F150, and Chevrolet Silverado. These vehicles were selected to represent the range of sizes and technologies present in the full vehicle sample. Table 5 summarizes the speciation schedule by vehicle. For additional detail on speciation, see the Testing Report[5].

Table 5. Speciation Schedule by Vehicle, Fuel, Bag and Replicate.

Vehicle Set	Speciation Type	Replicate 1		Replicate 2+	
		Bag 1	Bags 2-3	Bag 1	Bags 2-3
Impala, Silverado, F150, Civic, Corolla	Alcohols, Carbonyls	All fuels	12-fuel subset[1]	All fuels	-
	Hydrocarbons	12-fuel subset[1]	12-fuel subset[1]	-	-
Remaining 10 vehicles	Alcohols, Carbonyls	All fuels	-	All fuels	-
	Hydrocarbons	12-fuel subset[1]	-	-	-

[1] See Table 3.

2.3 Correlations Among Fuel Parameters

As mentioned, the purpose of this study is to attempt to relate emissions to changes in fuel parameters, treated as continuous variables, using multiple regression as the analysis technique. Thus, the "fuel-parameter matrix" is a set of properties abstracted from the set of fuels on which measurements were performed. However, it is commonly observed that fuel properties tend to be correlated. In addition, interaction terms often show strong correlations with the linear-effects terms from which they were constructed. Such correlations can be an issue in analysis of the

data as they result in collinearities among predictors, with potentially adverse effects on models fit with collinear terms, such as reduced precision in estimation of model coefficients. In extreme cases, coefficients can change substantially as additional terms are added to or removed from models, or can be of the wrong magnitude or sign.

The first step taken to neutralize correlations among fuel parameters was the design and optimization of the fuel-parameter matrix itself. A goal of the design process is to approximate, as closely as possible, the efficiency of a full factorial design in which all model terms would be independent and orthogonal. Taking the optimized full matrix as a starting point, we assessed the potential for issues related to collinearity by constructing all candidate 2nd order terms and then compiling a correlation matrix (R) of model terms. Results showed that a number of strong correlations persist among model terms in the optimized parameter matrix, in which "strong" is defined as $|R| > 0.50$, as shown in Table 6. In addition to correlations of moderate strength ($0.50 < R < 0.75$), a number of very strong correlations are apparent ($R > 0.90$). Not unexpectedly, these results appear in quadratic terms, e.g., etOH×etOH and T50×T50 are highly correlated to etOH and T50, with $R = 0.95$ and 1.00, respectively. Additionally, strong correlations appear among the several 2nd order interaction terms. For example, etOH is strongly correlated with all four of its interactions with the other fuel parameters.

In addition to correlations among the linear effects and interactions, and correlations among interactions, we can see one fairly strong correlation among the linear effects, specifically, between etOH and T50 ($R = -0.57$). This residual correlation reflects the physical relationship between ethanol content, reflecting the hydrocarbon content of the fuel, and the T50, which as a bulk property, is strongly influenced by ethanol, which tends to increase the volatility of the fuel. Specifically, it is not possible to fully "orthogonalize" a fuel matrix between etOH and T50 for etOH levels much above 10%, as fuels with high ethanol and high T50 cannot be blended. In the full parameter-matrix, all fuels with 20% ethanol have T50 in the neighborhood of 165 °F. The relationship between etOH and T50 is shown in Figure 1. For comparison, "rectangular" relationships between ethanol and aromatics and RVP are shown in Figure 2 and Figure 3. These plots show the greater level of success the optimization process achieved in "orthogonalizing" relationships between etOH, aromatics and RVP, as opposed to T50. Additional two-way plots of the fuel properties are presented in Appendix A.

Table 6. Correlation matrix for linear-effect and interaction terms in the design fuel-parameter matrix.

	Ethanol	Arom	RVP	T50	T90	etOH×etOH	T50×T50	T90×T90	etOH×T50	etOH×T90	etOH×Arom	etOH×RVP
Ethanol	1.00	-0.04	-0.15	**-0.57**	-0.01	**0.95**	**-0.56**	-0.01	**0.98**	**1.00**	**0.84**	**0.96**
Arom		1.00	0.05	-0.10	-0.02	-0.07	-0.10	-0.02	-0.04	-0.04	0.40	-0.03
RVP			1.00	-0.26	0.13	-0.15	-0.25	0.13	-0.22	-0.13	-0.12	0.08
T50				1.00	-0.02	**-0.57**	**1.00**	-0.02	-0.43	**-0.57**	-0.50	**-0.63**
T90					1.00	-0.01	-0.02	**1.00**	-0.01	0.06	-0.01	0.04
etOH×etOH						1.00	**-0.57**	-0.02	**0.90**	**0.95**	**0.78**	**0.91**
T50×T50							1.00	-0.02	-0.42	**-0.56**	**-0.50**	**-0.62**
T90×T90								1.00	-0.02	0.06	-0.01	0.04
etOH×T50									1.00	**0.98**	**0.82**	**0.92**
etOH×T90										1.00	**0.83**	**0.96**
etOH×Arom											1.00	**0.80**
etOH×RVP												1.00

Figure 1. Values of T50 vs. Ethanol for design points in the full fuel-parameter matrix.

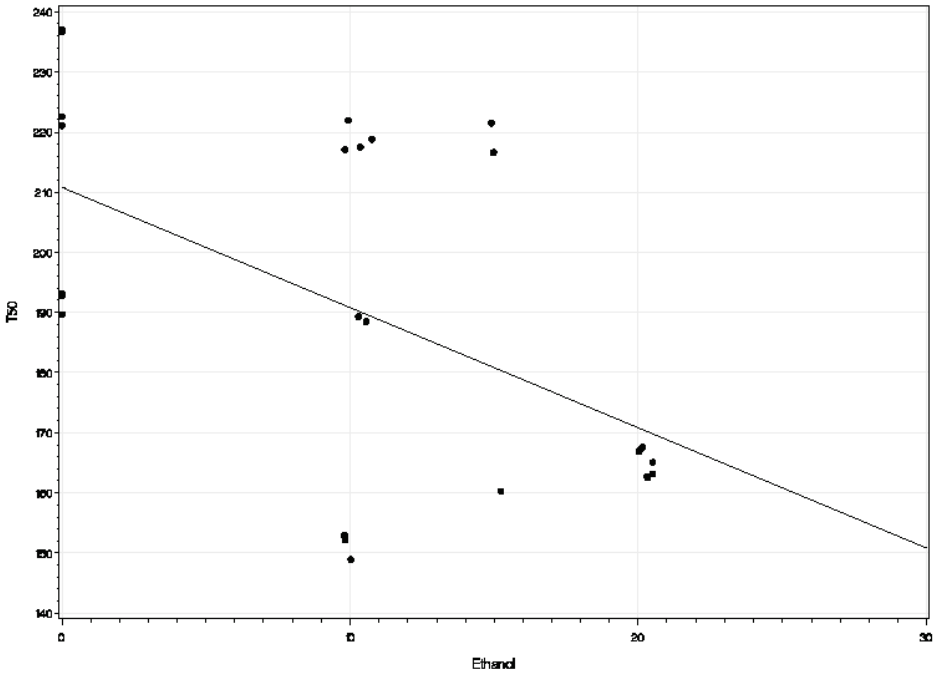


Figure 2. Values of Aromatics vs. Ethanol for design points in the full parameter-matrix.

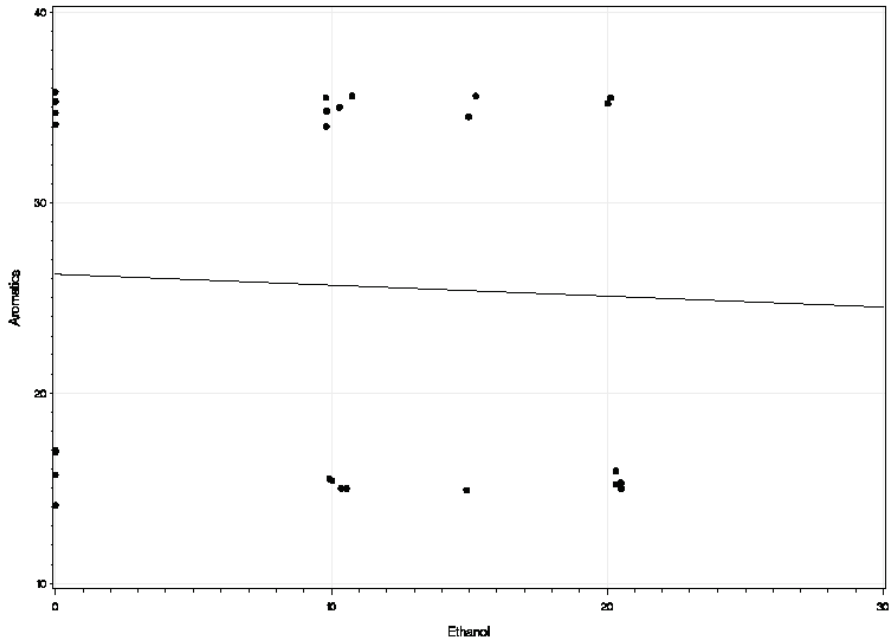
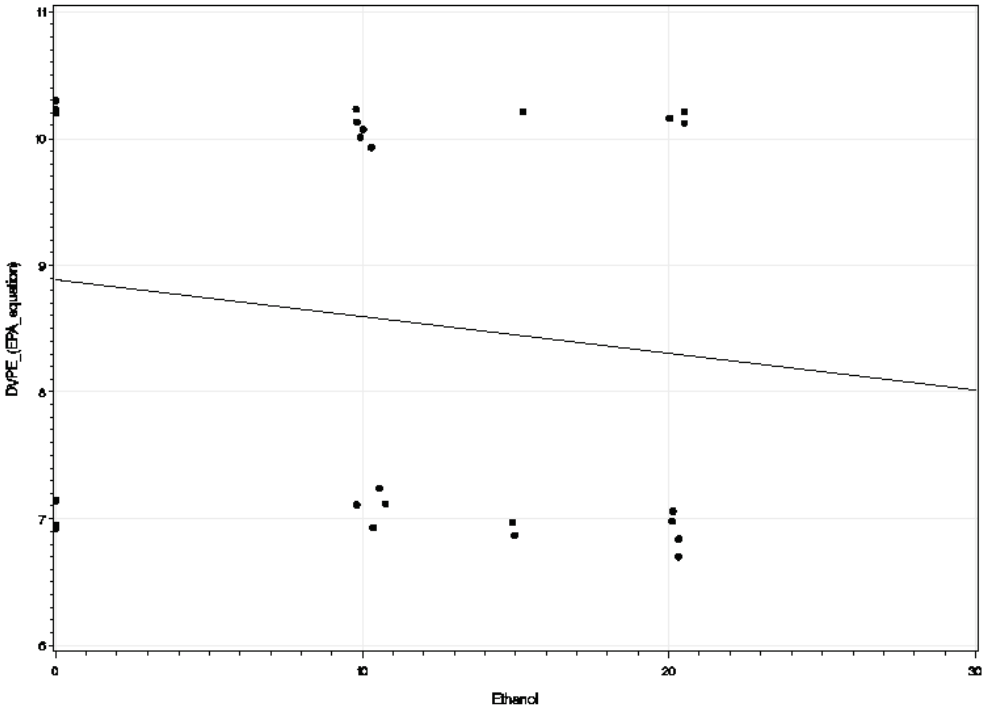


Figure 3. Values of RVP vs. Ethanol for design points in the full parameter-matrix.



Since it does not fit the goals of the study to resolve collinearities by removing terms from the models, it was necessary to find an additional means to neutralize the correlations.

2.3.1 Standardization of Fuel Parameters

As a preliminary step to modeling, an analytic step to neutralize remaining correlations in the fuel-parameter matrix is to center and scale the fuel parameters, a process known as "standardization." Standardization is commonly used for this purpose in analysis of experiments[8]. Standardization simply involves first "centering" the measured fuel properties by subtracting their means, and then "scaling" by then dividing the centered values by their respective standard deviations, as shown in Equation 3, with statistics for the fuel properties presented in Table 7. The result is a "*Z* score," representing a "standard normal distribution" with a mean of 0.0 and a standard deviation of 1.0.

$$Z_i = \frac{x_i - \bar{x}}{s} \qquad \textbf{Equation 3}$$

2.3.1.1 One-Stage Standardization

In "one-stage" standardization, the linear effects terms are centered and scaled, as shown above in Equation 3. Interaction terms are constructed directly from the linear effects and then standardized. Using ethanol as an example, the standardization of the linear-effect term was performed as shown in Equation 4.

$$Z_{\text{etOH}} = \frac{x_{\text{etOH}} - \bar{x}_{\text{etOH}}}{s_{\text{etOH}}} \qquad \textbf{Equation 4}$$

Using the etOH×Aromatics interaction term as an example, the standardized values were constructed as shown in Equation 5, where $x_{\text{etOH}\times\text{Arom}}$ and $s_{\text{etOH}\times\text{Arom}}$ are the mean and standard deviation of $x_{\text{etOH}}x_{\text{Arom}}$, respectively.

$$Z_{\text{etOH}\times\text{Arom}} = \frac{x_{\text{etOH}}x_{\text{Arom}} - \bar{x}_{\text{etOH}\times\text{Arom}}}{s_{\text{etOH}\times\text{Arom}}} \qquad \textbf{Equation 5}$$

After performing the "one-stage" standardization for linear and interaction terms, we reevaluated correlations among the standardized fuel-parameter matrix (see Table 8). Results show that this step did not neutralize strong correlations, particularly between linear effects and interactions, or among interactions. In fact, the pattern of correlations is identical to that in the unstandardized matrix (Table 6).

2.3.1.2 Two-stage standardization

As one-stage standardization did not neutralize correlations among model terms, we applied a second stage of standardization to the 2nd order terms[9] . This step was conducted by constructing 2nd order terms, not from the measurements themselves, but rather from the standardized values for the linear effects. Using the etOH×etOH term as an example, the two-stage standardized value, denoted by ZZ_{etOH2} , was calculated as

$$ZZ_{\text{etOH}^2} = \frac{Z^2_{\text{etOH}} - m_{Z^2_{\text{etOH}}}}{s_{Z^2_{\text{etOH}}}} \qquad \textbf{Equation 6}$$

where the square of the standardized etOH term (Z^2_{etOH}) is centered by subtracting its mean m and standard deviation s (Equation 6). Similarly, the two-stage value for the etOH×Aromatics interaction, $ZZ_{\text{etOH}\times\text{Arom}}$ was calculated as shown in Equation 7.

$$ZZ_{\text{etOH}\times\text{Arom}} = \frac{Z_{\text{etOH}} Z_{\text{Arom}} - m_{Z_{\text{etOH}} Z_{\text{Arom}}}}{s_{Z_{\text{etOH}} Z_{\text{Arom}}}} \qquad \textbf{Equation 7}$$

Following calculation of the two-stage values, we evaluated correlations in the standardized fuel-parameter matrix when one-stage standardized values (*Z)* were used for linear-effect terms and two-stage standardized values (*ZZ)* were used for quadratic and interaction terms. Table 9 shows that the combination of one- and two-stage standardization neutralizes the remaining correlations, with the exception of remaining correlation between the linear effects for ethanol and T50, as previously described. Several moderate correlations remain among 2nd order terms, but no correlations with $R > 0.70$ remain.

Table 7. Means and Standard deviations for Fuel Properties, based on Fuel Matrices for the Full and Reduced Designs.

Model Term	Full Design[1]		Reduced Design[2]	
	Mean	Standard deviation	Mean	Standard Deviation
Ethanol (%)	10.3137	7.87956	11.0182	8.05925
Aromatics (%)	25.6296	10.0154	24.3909	9.92426
RVP (psi)	8.5178	1.61137		
T50 (°F)	190.611	28.5791	197.000	23.4536
T90 (°F)	320.533	19.4801	323.527	19.6015
etOH × etOH	0.962963	0.802769		
T50 × T50	0.962963	0.739766		
etOH × Arom	-0.03674	0.978461		
etOH × RVP	-0.0992352	0.999615		
etOH × T50	-0.541342	0.769153		
etOH × T90	0.0163277	0.972825		
Arom × RVP	0.043792	0.984096		
Arom × T50	-0.068030	0.991737		
Arom × T90	-0.0062526	0.983536		
T90 × T90	0.962963	0.346951		
T50 × T90	-0.036304	0.960011		
RVP × T90	0.126761	0.972829		

[1] Applies to models fit with data for 15 vehicles measured on 27 fuels.
[2] Applies to models fit with data for 5 or 15 vehicles measured on 11 fuels. See Chapter 9.

Table 8. Correlation matrix for one-stage standardized linear-effect and interaction terms in the optimal-design fuel-parameter matrix.

	Ethanol	Arom	RVP	T50	T90	etOH×etOH	T50×T50	T90×T90	etOH×T50	etOH×T90	etOH×Arom	etOH×RVP
Ethanol	1.00	-0.04	-0.15	-0.57	-0.01	0.95	-0.56	-0.01	0.98	1.00	0.84	0.96
Arom		1.00	0.05	-0.10	-0.02	-0.07	-0.10	-0.02	-0.04	-0.04	0.40	-0.03
RVP			1.00	-0.26	0.13	-0.15	-0.25	0.13	-0.22	-0.13	-0.12	0.08
T50				1.00	-0.02	-0.57	1.00	-0.02	-0.43	-0.57	-0.50	-0.63
T90					1.00	-0.01	-0.02	1.00	-0.01	0.06	-0.01	0.04
etOH×etOH						1.00	-0.57	-0.02	0.90	0.95	0.78	0.91
T50×T50							1.00	-0.02	-0.42	-0.56	-0.50	-0.62
T90×T90								1.00	-0.02	0.06	-0.01	0.04
etOH×T50									1.00	0.98	0.82	0.92
etOH×T90										1.00	0.83	0.96
etOH×Arom											1.00	0.80
etOH×RVP												1.00

Table 9. Correlation matrix for Standardized linear-effect (one-stage) and interaction (two-stage) terms in the full-design fuel-parameter matrix.

	Ethanol	Arom	RVP	T50	T90	etOH×etOH	T50×T50	T90×T90	etOH×T50	etOH×T90	etOH×Arom	etOH×RVP
Ethanol	1.00	-0.04	-0.15	-0.57	-0.01	-0.03	-0.01	-0.18	-0.08	-0.02	-0.08	-0.01
Arom		1.00	0.05	-0.10	-0.02	-0.11	-0.10	-0.23	0.15	0.02	0.03	-0.02
RVP			1.00	-0.26	0.13	-0.01	0.16	-0.07	-0.06	0.09	-0.02	0.02
T50				1.00	-0.02	-0.08	0.12	0.26	-0.01	-0.02	0.12	-0.05
T90					1.00	-0.03	0.05	-0.52	-0.03	-0.07	0.02	0.08
etOH×etOH						1.00	-0.12	-0.07	-0.68	-0.02	-0.09	-0.07
T50×T50							1.00	-0.18	-0.44	-0.02	0.06	0.05
T90×T90								1.00	0.18	-0.12	-0.31	0.06
etOH×T50									1.00	0.04	-0.01	-0.04
etOH×T90										1.00	0.01	0.14
etOH×Arom											1.00	-0.05
etOH×RVP												1.00

3 Dataset Construction

3.1 Selection of Aggregate (Bag) vs. Continuous Data

Both dilute-bag and raw continuous (second-by-second) emission results were generated for THC, CH_4, CO, NO_x and CO_2 for all tests in this program. The dilute-bag method has been optimized over decades of use and is considered the "gold standard" for light-duty vehicle emission measurements. In this constant-volume sampling system, the vehicle exhaust is mixed with a large amount of filtered dilution air, and a small portion of this stream is continuously withdrawn to fill a sealed bag over the course of a test cycle. The total flowrate of exhaust plus dilution air is held constant by a critical flow venturi, and the bag fill rate is held constant by a pump and flowmeter, such that the concentration of emissions in the bag is a time-weighted average for the cycle. At the end of the cycle, the contents of the bag are flowed through an analyzer to determine the pollutant concentration. The primary advantage of this method is that it relies on well-understood physical phenomena for controlling flowrates and dilution of exhaust gases during collection of the sample, which obviates the need for continuous monitoring and/or adjustment of these parameters. Its primary disadvantage is that the overall dilution ratio of background air to exhaust must be fixed for an entire test, and is set relatively high to avoid condensation of water vapor within the system during periods of high exhaust flow. Emission rates of species of interest continue to decline with more stringent regulations and better control technology, however, the emission rate of water vapor, and thus sampling dilution ratio, is a function of fuel economy, which hasn't changed much over time. As a result, for very low-emitting vehicles the concentration of pollutants of interest that must be quantified in the dilute bag may be indistinguishable from the level in the background (dilution) air, given the limitations of the instruments and test procedures. When assessing the effects of fuel properties on very low emissions concentrations, the challenge is even greater.

In the raw continuous system, a small portion of the vehicle exhaust is continuously withdrawn and flowed through an analyzer, which generates a continuously-varying concentration measurement with a short lag that is a function of the system's internal dimensions and sample flowrate. This method does not perform any dilution, alleviating the issue of needing to quantify very low concentrations (and there is no background measurement to subtract). However, it does require continuous (>10 Hz) measurement of exhaust flow rate and precise coordination of the time series with the emission concentration in order to produce an accurate time-integrated emission result for a test cycle. These latter processes introduce varying amounts of error into an integrated continuous measurement. While there is very good agreement in general, it is not uncommon for the overall result to vary from the dilute bag measurement made on the same test especially at low emission levels.

Plots and regressions run between the dilute bag and integrated continuous datasets generally showed very good agreement for the vast majority of results. Plotting the lowest 10% of data sorted by magnitude, however, showed notable differences in trend as measurements approached

zero. The integrated continuous results were suggestive of a "floor" while the dilute bag results maintained a more linear trend toward zero (with increasing scatter) until censoring occurred at a very low level. Figure 4 and Figure 5 depict the behavior of NO_x in Bags 1 and 2, respectively.

Figure 4. Bag 1 NO_x: Integrated-Continuous vs. Dilute-Bag Measurements.

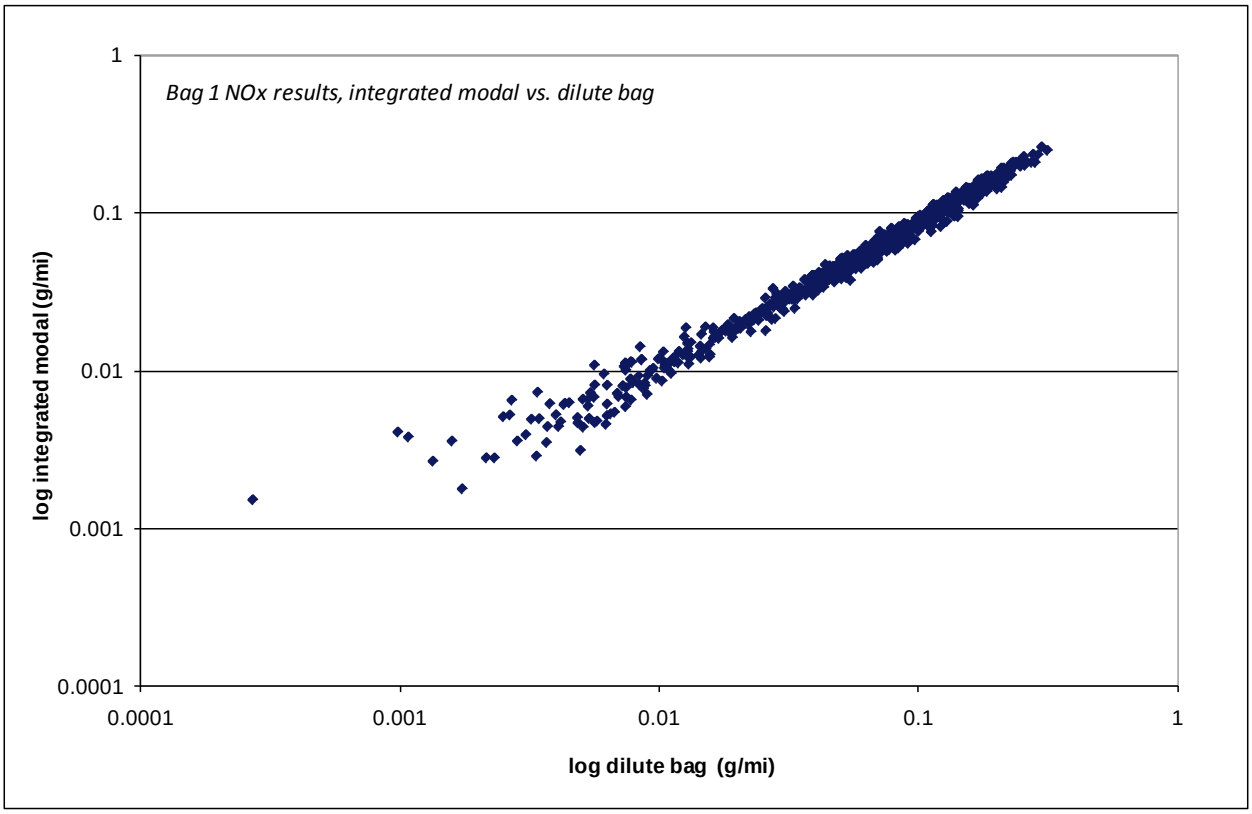
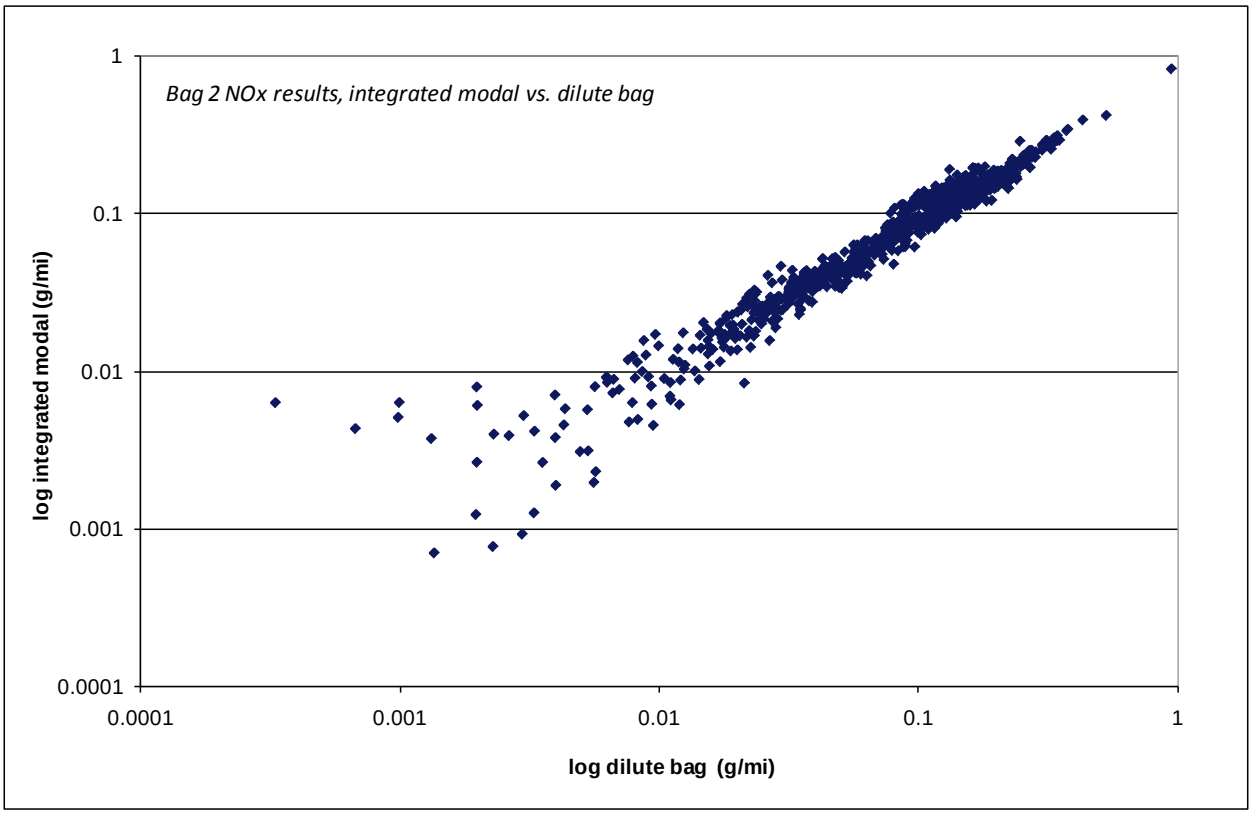


Figure 5. Bag 2 NO_x: Integrated-Continuous vs. Dilute-Bag Measurements.



As discussed further in Section 5, most datasets generated in this program contain results quantified as zero, due to the sample concentration being either below the detection limit of the analyzer, or quantified as being less than or equal to the measured background concentration. Zero values resulting from the first case represent low-end censoring, a statistical term meaning that there are one or more nonzero levels below which no emissions are captured due to limitations in the measurement process. If there are many censored values, they can introduce a bias into the analysis because they are unlikely to be truly zero; it is known that some emissions were produced at some point during the test cycle, but they were simply too small to measure with the given method. In the second case, where the sample concentration appears to be less than or equal to the background, it is likely in most cases due to the sample concentration being indistinguishable from the background due to measurement variability or error (see Section 6 for further discussion of measurement error).[i] This situation is evidenced by looking at the time

[i] This assessment of the situation neglects the possibility that the vehicle actually consumes or destroys a given pollutant species during parts of the test cycle, resulting in periods of "negative emissions", such that the average emission level over a test is truly zero. While such situations may occur over a limited period for some emissions in a highly polluted environment, e.g., PM or NMHC in congested traffic, it is highly unlikely in an emission test cell.

series for raw continuous measurements, which show emissions being produced at various points during all tests, even those having zero measurements as the net dilute bag result.

Given this fact, it is possible to use an integrated continuous result as an estimate to replace a zero bag result. The decision to do so would presume the continuous measurement has lower relative measurement error than the dilute bag measurement at very low emission levels; however, a nonzero result doesn't necessarily ensure that this condition holds. All measurement methods have error associated with them, and given the history and widespread use of the dilute bag method, its error levels and limitations are relatively well understood. Conversely, there is less history with continuous measurement methods to give confidence that those results are an improvement over dilute bag measurements (even when results were zero). For these reasons, the decision was made not to replace missing in the dilute bag dataset with integrated continuous measurements. Rather, these measurements were treated as "censored." Modeling approaches adopted to address the presence of "censoring" are described in 5.3 (page 96).

3.2 Imputation of Speciated Hydrocarbons (NMOG, NMHC)

Due to the speciation schedule described in Section 2.2, most tests in the dataset do not have alcohol and carbonyl measurements for bags 2-3.[j] As NMOG and NMHC are calculated emission results that use speciation data, they could not be computed for the portions of the dataset without speciation. A conservative approach to modeling these emissions might only use the small subset with speciation (71 tests). As a result, models fit for NMOG and NMHC would not have the advantage of the full optimized study design, as do the models for the other emissions (THC, NO_x, etc.). This outcome would impose a severe limitation, and is unsatisfactory, considering the importance of NMOG and NMHC.

However, it is possible to compensate for the limited level of speciation by drawing on an alternate measure of hydrocarbon emissions obtained from the flame ionization detector (FID). The alternate measure "NMHC as measured by FID" ($NMHC_{FID}$), was collected for the entire dataset, and is very tightly correlated with both NMOG and "true" NMHC. It is thus possible to estimate NMOG and NMHC results for tests without speciation by using correlations generated from those with speciation. This technique essentially estimates the offset between the response of the FID and the fully characterized emission stream, due to the incomplete measurement of oxygenates by the FID. For NMOG, this estimated value is typically between 2-20% higher than the $NMHC_{FID}$ measurement, depending on emission bag and fuel ethanol level.

To accomplish this step, we investigated the relationship between NMOG and NMHC with $NMHC_{FID}$, by bag. Based on strong correlations between these species, we developed statistical

[j] Additionally, a small number of tests (thirty), which had originally been voided for having incomplete speciation data for bag 1, were included in the main dataset as "salvaged" or "makeup" tests because they could provide valid results for other emissions such as NOx and CO. It was thus useful to impute NMOG and NMHC results for bag 1 from these tests as well.

models to impute NMOG and NMHC from corresponding $NMHC_{FID}$ measurements. The results of these analyses are presented below.[k] The results of the current project are very similar to those obtained in work performed at Oak Ridge National Laboratory that related NMOG to NMHC for selected ethanol blends on the FTP cycle[10].

In application of the models described in 3.2.1 and 3.2.2 were applied deterministically, i.e, the imputed values represent means at different levels of the predictors, rather than individual measurements. In taking this step, the random error, or "scatter" around the predicted means was neglected. Due to the very tight model fits, we expect that the degree of associated error in imputation is small, and did not substantially affect the results.

3.2.1 Imputation of NMOG

Scatterplots of NMOG vs. $NMHC_{FID}$ for Bag 1 are shown in Figure 6. A tight linear relationship is apparent between the two sets of measurements. Moreover, distinct trends are visible for each ethanol level, from lowest (E0) to highest (E20), which are even more conspicuous on the logarithmic plot. On this basis, we fit linear models for NMOG and NMHC in terms of $NMHC_{FID}$. The models were fit as least-squares regressions, with the ethanol levels set as indicator, or "dummy" variables.

Model results for Bag 1 NMOG are shown in Table 10. The intercept for the reference category (E0) is not significantly different from 0.0, suggesting that NMOG is 0 when $NMHC_{FID}$ is 0. However, the intercepts for the E10, E15 and E20 levels are positive and significantly different from 0.0, suggesting that at higher ethanol levels, some fraction of NMOG exists independently of $NMHC_{FID}$. This outcome is most likely a result of a combination of the presence of undetected formaldehyde (no FID response) and the increasing levels of unaccounted FID response due to ethanol and acetaldehyde emissions (which are partially detected by the FID) with increasing ethanol blends. The slope term for E0 indicates that NMOG is numerically very close to $NMHC_{FID}$, or approximately 0.9% higher. At ethanol levels of 10% or higher, the differences in the slope terms increase relative to E0, with the exception that the slope increment for E15 is slightly lower than that for E10. However, this difference is small, and not statistically significant at the 95% confidence level (two-tailed $p = 0.92$). All terms in this model were retained, because they gave an increase in overall fit, along with a more symmetric distribution of residuals around 0.0, than an alternative model with only intercepts for each ethanol level, but uniform slopes.

[k] The individual data files produced by SwRI do not contain any imputed values; both NMOG and NMHC values reported there for Bags 2-3 were simply set equal to $NMHC_{FID}$ when there were no speciation data. The summary database file produced by EPA contains a separate sheet with estimated NMOG results, showing imputed values inserted where they could not be rigorously computed. This sheet also contains another emission item called NM_FIDHC (equal to $NMHC_{FID}$) computed by EPA for each test during post-processing of the data. This value served as the input to the models shown in Equation 8 through Equation 11, and can also be used to produce plots or other analyses of the correlations and modeling.

Figure 7 shows scatterplots of NMOG vs. $NMHC_{FID}$ for Bag 2. The plot shows a tight linear trend, but no apparent sub-trends for ethanol levels. Model results for Bag 2 NMOG are shown in Table 11. Unlike Bag 1, the reference intercept is statistically different from 0.0, although very small. The model structure is simpler than that for Bag 1, in that intercepts were fit for each ethanol level, but a single slope was used for all ethanol levels. A more complex structure with individual slopes by ethanol level was fit, but the additional complexity did not improve the fit, nor were the additional slope effects significant. The slope term is again positive and highly significant, but very small (NMOG is 0.31% higher than $NMHC_{FID}$). In fitting this model, a single influential and outlying observation was deleted (studentized-deleted residual > 5.0).

Scatterplots of NMOG vs. $NMHC_{FID}$ for Bag 3 are presented in Figure 8. The plot shows a tight linear trend, but no apparent subtrends for ethanol levels. The picture is very similar to that for Bag 2, except that several points fall well off the main trend. Model results for Bag 3 are shown in Table 12. The model structure is identical to that in Bag 2, with intercepts by ethanol level with a uniform slope. The reference intercept for E0 is not significantly different from 0.0, the intercept for E10 is marginally significant, and that for E20 is significant, while that for E15 is insignificant. The separate intercepts were retained to retain the significant effect for E20. One outlier with a large studentized–deleted residual was removed before fitting this model.

The model for Bag 1 NMOG was applied using Equation 8

$$y_{NMOG} = \beta_0 + \beta_{10} i_{10} + \beta_{15} i_{15} + \beta_{20} i_{20} + (\gamma_0 + \gamma_{10} i_{10} + \gamma_{15} i_{15} + \gamma_{20} i_{20}) x_{NMHC} \qquad \textbf{Equation 8}$$

where x_{NMHC} is a measured value of $NMHC_{FID}$, y_{NMOG} is the predicted mean NMOG level for a given measurement of x_{NMHC}, β_0 is the reference intercept for E0 fuels, i_{10}, i_{15} and i_{20} are indicator or dummy variables for E10, E15 and E20 fuels, (e.g., $i_{10} = 1$ where ethanol level = 10%, 0 elsewhere), β_{10}, β_{15} and β_{20} are offsets to the reference intercept for E10, E15 and E20 fuels, γ_0 is the reference intercept for E0 fuels, and γ_{10}, γ_{15} and γ_{20} are offsets to the reference slope for E10, E15 and E20 fuels, respectively. Thus, for E0 fuels, where i_{10}, i_{15} and i_{20} all equal 0, Equation 8 reduces to Equation 9.

$$y_{NMOG} = \beta_0 + \gamma_0 x_{NMHC} \qquad \textbf{Equation 9}$$

Similarly, for E10 fuels, where $i_{10} = 1$ and i_{15} and i_{20} both = 0, Equation 8 reduces to

$$y_{NMOG} = \beta_0 + \beta_{10} + (\gamma_0 + \gamma_{10}) x_{NMHC} \qquad \textbf{Equation 10}$$

For the bag 2 and 3 models, the counterpart to Equation 8 is Equation 11, which simplifies for blends other than E0 similarly to Equation 8, except that the slope term is always as in Equation 9.

$$y_{NMOG} = \beta_0 + \beta_{10} i_{10} + \beta_{15} i_{15} + \beta_{20} i_{20} + \gamma_0 x_{NMHC} \qquad \textbf{Equation 11}$$

During examination of measurements below 0.001 g/mi in scatterplots and regression fits, it was noted that there was higher variability of NMOG measurements at very low emission levels. This behavior seemed to be especially prominent below 0.0001 g/mi $NMHC_{FID}$, and is likely due to the fact that the FID and wet-chemistry methods have different sources and magnitudes of measurement variability, an issue not resolvable through statistical methods. Thus, a decision was made to substitute a zero result for NMOG in cases where $NMHC_{FID}$ values were <0.0001 g/mi, rather than perform the imputation. This outcome occurred in 44 bag-2 tests and 119 bag-3 tests.

Additional detailed information on these analyses is presented in Appendices B.1 – B.3.

Figure 6. Scatterplots of NMOG and ln(NMOG) vs $NMHC_{FID}$ and ln($NMHC_{FID}$), for BAG 1, by ethanol level, showing results for five Vehicles with Speciated Measurements.

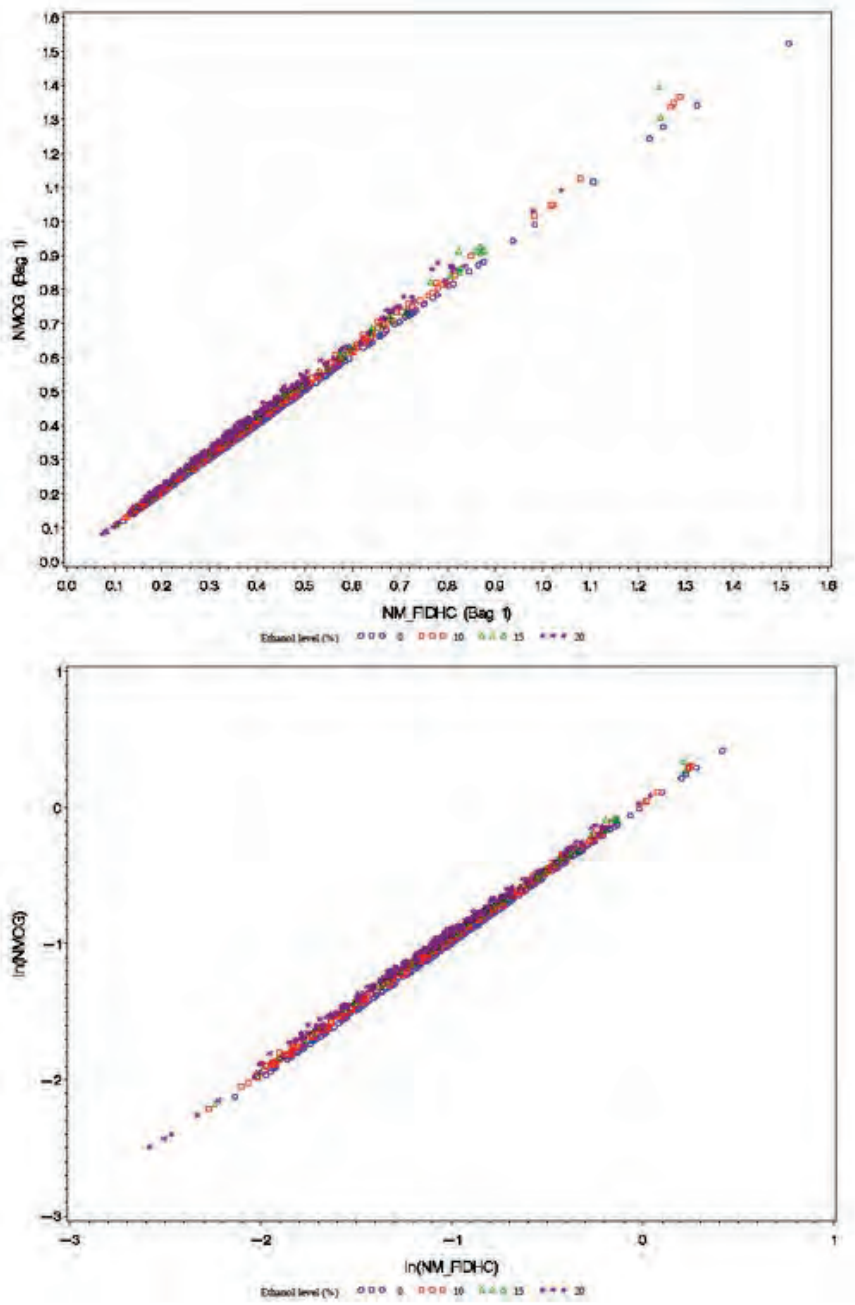


Figure 7. Scatterplots of NMOG and ln(NMOG) vs $NMHC_{FID}$ and ln($NMHC_{FID}$), for BAG 2, by ethanol level, showing results for five Vehicles with Speciated Measurements.

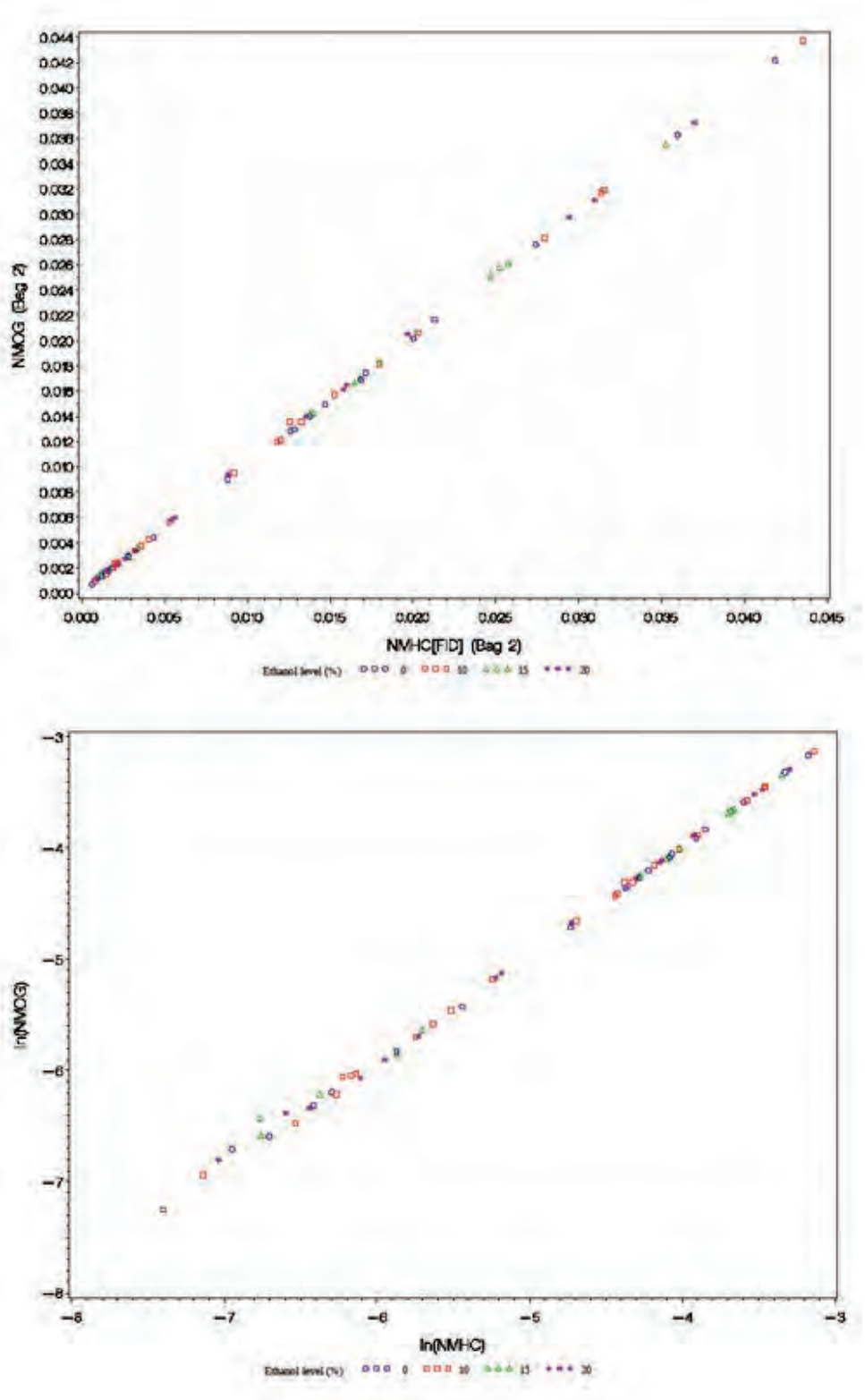


Figure 8. Scatterplots of NMOG and ln(NMOG) vs $NMHC_{FID}$ and ln($NMHC_{FID}$), for BAG 3, by ethanol level, showing results for five Vehicles with Speciated Measurements.

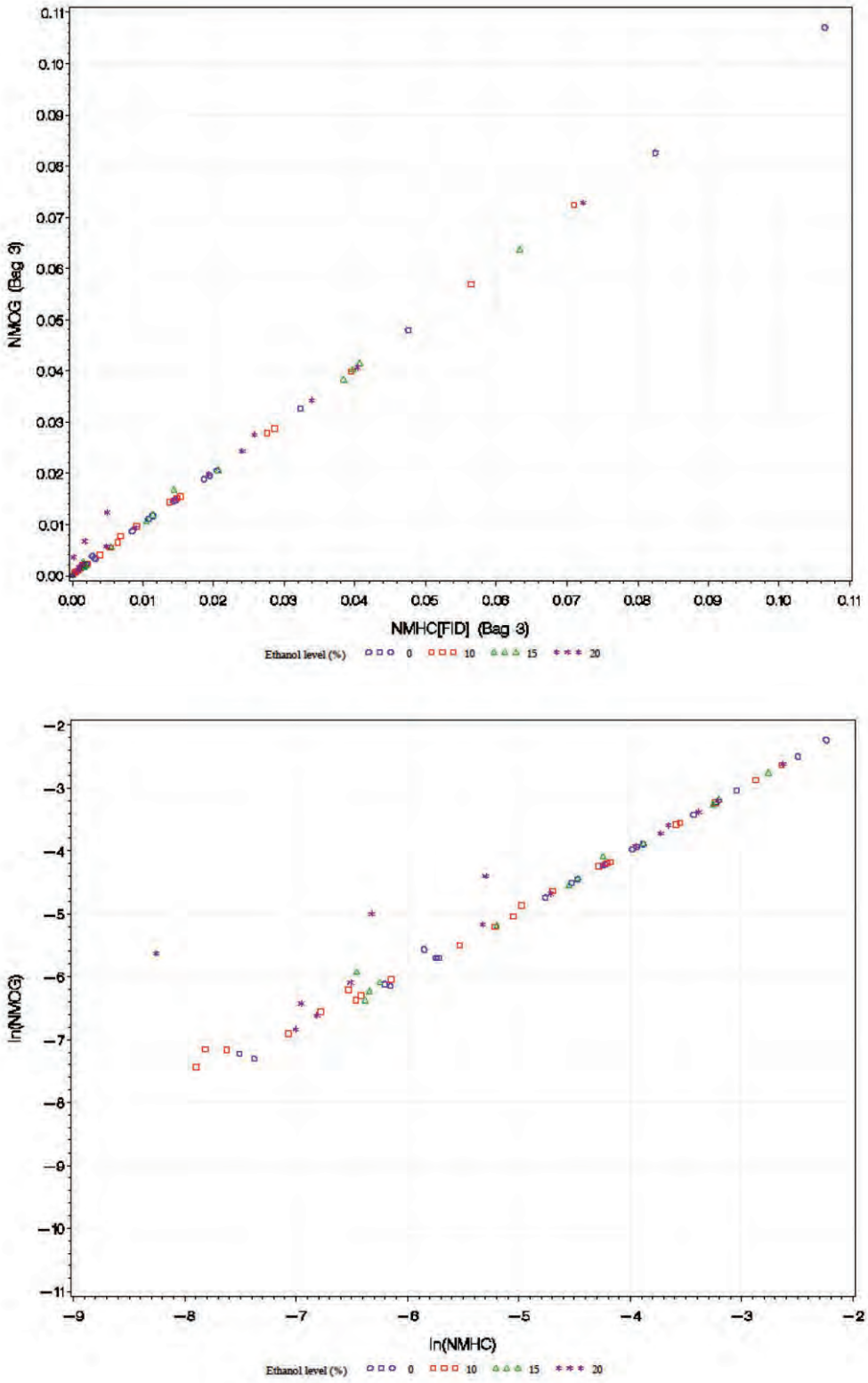


Table 10. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 1 NMOG (dependent variable = NMOG (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	0.000438657	0.00077499	0.57	0.5715
Slope (E0)	0.009240268	0.00166026	608	<0.0001
Intercept Increment (E10)	0.002882016	0.00107999	2.67	0.0078
Slope Increment (E10)	0.030289819	0.00236797	13.0	<0.0001
Intercept Increment (E15)	0.010042668	0.00167492	6.00	<0.0001
Slope Increment (E15)	0.030289819	0.00336111	9.01	<0.0001
Intercept Increment (E20)	0.010298206	0.00120701	8.53	<0.0001
Slope Increment (E20)	0.048048225	0.00285922	16.8	<0.0001
Fit Information: d.f.[model] = 7, d.f.[error] = 911, d.f.[total] = 918. *F* value = 157,316, Pr > *F* = <0.0001. R^2 = 0.999173, Root MSE = 0.005771.				

Table 11. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 2 NMOG (dependent variable = NMOG (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	0.000156975	0.00002878	5.45	<0.0001
Slope	1.003147342	0.00111376	900.68	<0.0001
Intercept Increment (E10)	0.000057909	0.00003388	1.71	0.092
Intercept Increment (E15)	0.000123063	0.00003767	3.27	0.0018
Intercept Increment (E20)	0.000087647	0.00003699	2.37	0.021
Fit Information: d.f.[model] = 4, d.f.[error] = 64, d.f.[total] = 68. *F* value = 204,200, Pr > *F* = <0.0001. R^2 = 0.999922, Root MSE = 0.000107.				

Table 12. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 3 NMOG (dependent variable = NMOG (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	0.000169016	0.00013708	1.23	0.22
Slope	1.002069357	0.00290768	345	<0.0001
Intercept Increment (E10)	0.000153740	0.0016850	0.91	0.37
Intercept Increment (E15)	0.00285066	0.00018829	1.51	0.14
Intercept Increment (E20)	0.000492911	0.00018850	2.61	0.011
Fit Information: d.f.[model] = 4, d.f.[error] = 64, d.f.[total] = 68. *F* value = 30,136, Pr > *F* = <0.0001. R^2 = 0.999469, Root MSE = 0.000534.				

3.2.2 Imputation of NMHC

As with NMOG, we first look to the plots of the data to investigate the feasibility of imputation and inform the process of model development. Inspecting the plots for Bag 1 (Figure 9), we see a pattern similar to that for NMOG, but in reverse. In the case of NMHC, distinct trends are visible for each ethanol level, but trends for higher ethanol levels are situated lower than for lower ethanol levels. In the top plot (linear scale) we see that the slopes also differ by ethanol level, with the sub-trends arrayed in a fan. In the logarithmic plot, however, the slopes appear parallel.

Following the visual review, we fit regression models, as described above for NMOG. Model fitting results for Bag 1 NMHC are shown in Table 13. As Figure 9 suggests, the best fit model includes distinct intercepts and slopes by ethanol level. The main intercept (representing 0% ethanol) is not significantly different from zero, and the main slope (again representing 0% ethanol) is very close to 1.0, although significantly different, displaying the very strong correlation between NMHC and $NMHC_{FID}$. The intercept and slope increments for 10%, 15% and 20% ethanol levels are all negative, as suggested in the plot, and increase in size with increasing ethanol, also as suggested in the plot. All these terms are significant, but with the degree of significance increasing with ethanol level. A likely explanation is that the more ethanol there is in the fuel, the more of the exhaust stream is comprised of oxygenated species. This means that the FID measurement of HCs deviates further and further from the true mass of emissions as determined by chemical speciation. As this difference increases, it is easier for the statistical tests to resolve in the midst of measurement variability present. In developing the best fit model, seven outlying measurements with high studentized-deleted residuals were removed ($|r_{-i}| \geq 3.5$). The model shown in Table 13 was applied using Equation 8. However, for Bag 1 it was only necessary to impute a relatively small number of NMHC measurements (< 50). For the vast majority of cases, actual measurements were available.

In the scatterplot for Bag-2 results (Figure 10), it is more difficult to visually assess the form of the best fitting model. Viewing the linear-scale plot (top), the measurements appear to sit on a single trend, very close to but not coinciding with a one-to-one trend. The set of available data is smaller (five vehicles), and a distinction of trends by ethanol level is not obvious. The same is true of the logarithmic plot (bottom), although is it apparent that the degree of relative scatter around the trend is greater at the low end than at the high end.

The model structure fit for Bag 2 is identical to that for Bag-2 NMOG; distinct intercepts for each ethanol level, but with a uniform slope across all levels (Table 14). The main intercept is negative and very small, but it is statistically significant, suggesting the possibility of a small but real offset between running NMHC and $NMHC_{FID}$ for the E0 blends measured. Based on these results and others, it appears there is a certain level of production of oxygenated species (primarily formaldehyde, around a few percent by mass), which is a baseline resulting from

combustion of any gasoline, regardless of oxygenate content. This appears as an offset for E0, which then grows as ethanol is added and emissions of ethanol and acetaldehyde increase. The intercept increments for the ethanol blends are negative and generally increasing (in absolute value) with ethanol level. One apparent anomaly is that the offset for 15% ethanol is slightly larger (more negative) than that for 20% ethanol. Both offsets are significantly different from zero, but almost certainly not significantly different from each other. If we discount the results at 15% ethanol due to the small number of fuels at that level and their combined properties, the pattern appears consistent, if not necessarily statistically significant. The model for Bag 2 was applied to impute NMHC measurements using Equation 11.

The visual impression for Bag 3 (Figure 11) is very similar to Bag 2, except that a small number of measurements sitting well off the main trend are more conspicuous, particularly in the logarithmic plot. The characteristics of the model fit are generally similar to Bag 2, with two exceptions (Table 15). A first exception is that the main intercept is not significantly different from zero, making the Bag-3 model more similar to the Bag-1 model in this respect. The main slope is again significant and close to 1.0, but not as close as the slope for Bag 2, although closer than that for Bag 1. The intercept increments are all negative; those for 10 and 15% ethanol are not significant, although the increment for 20% ethanol is highly significant. It is also an order of magnitude larger than the increments for the other two ethanol levels, ostensibly reflecting the greater variability of the measurements for 20% ethanol, even after removing two outlying measurements. As with Bag 2, the Bag 3 model was applied for imputation using Equation 11.

Additional information on these analyses is presented in Appendices C.1-C.3.

Figure 9. Scatterplots of NMHC and ln(NMHC) vs $NMHC_{FID}$ and $ln(NMHC_{FID})$, for BAG 1, by ethanol level. showing results for fifteen Vehicles with Speciated Measurements.

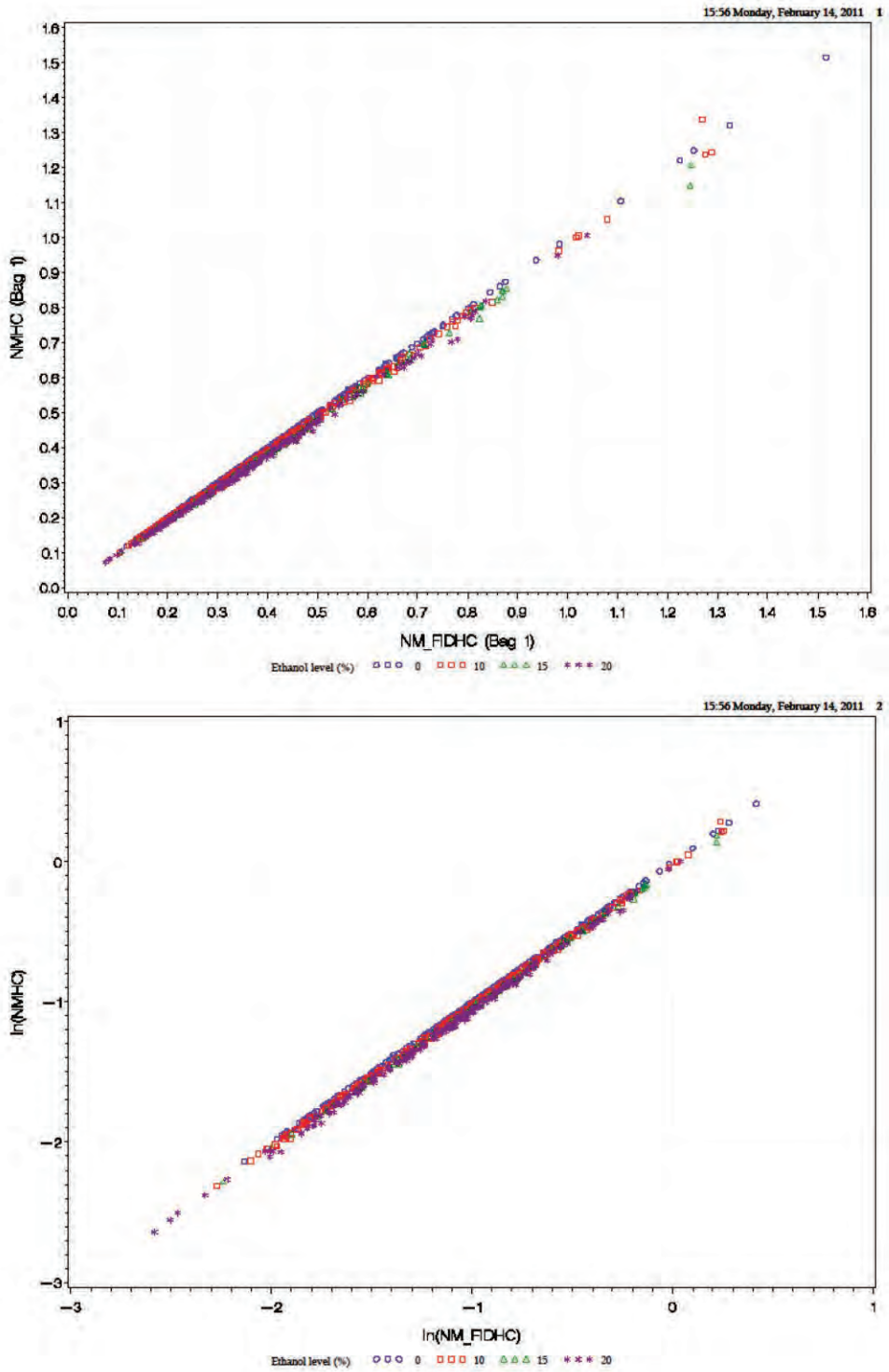


Figure 10. Scatterplots of NMHC and ln(NMHC) vs $NMHC_{FID}$ and ln($NMHC_{FID}$), for BAG 2, by ethanol level, showing results for five Vehicles with Speciated Measurements.

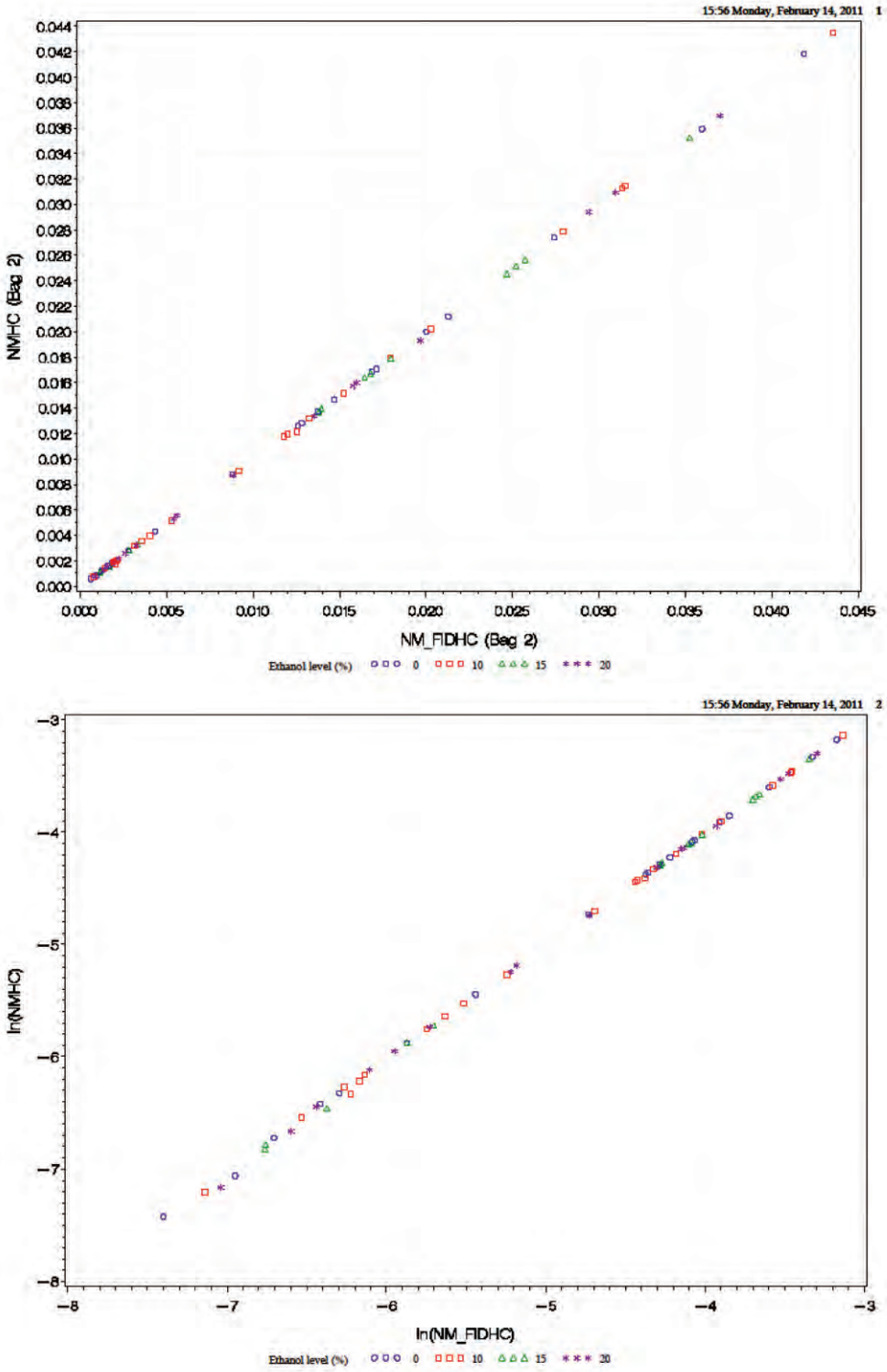


Figure 11. Scatterplots of NMHC and ln(NMHC) vs $NMHC_{FID}$ and ln($NMHC_{FID}$), for BAG 3, by ethanol level, showing results for five Vehicles with Speciated Measurements.

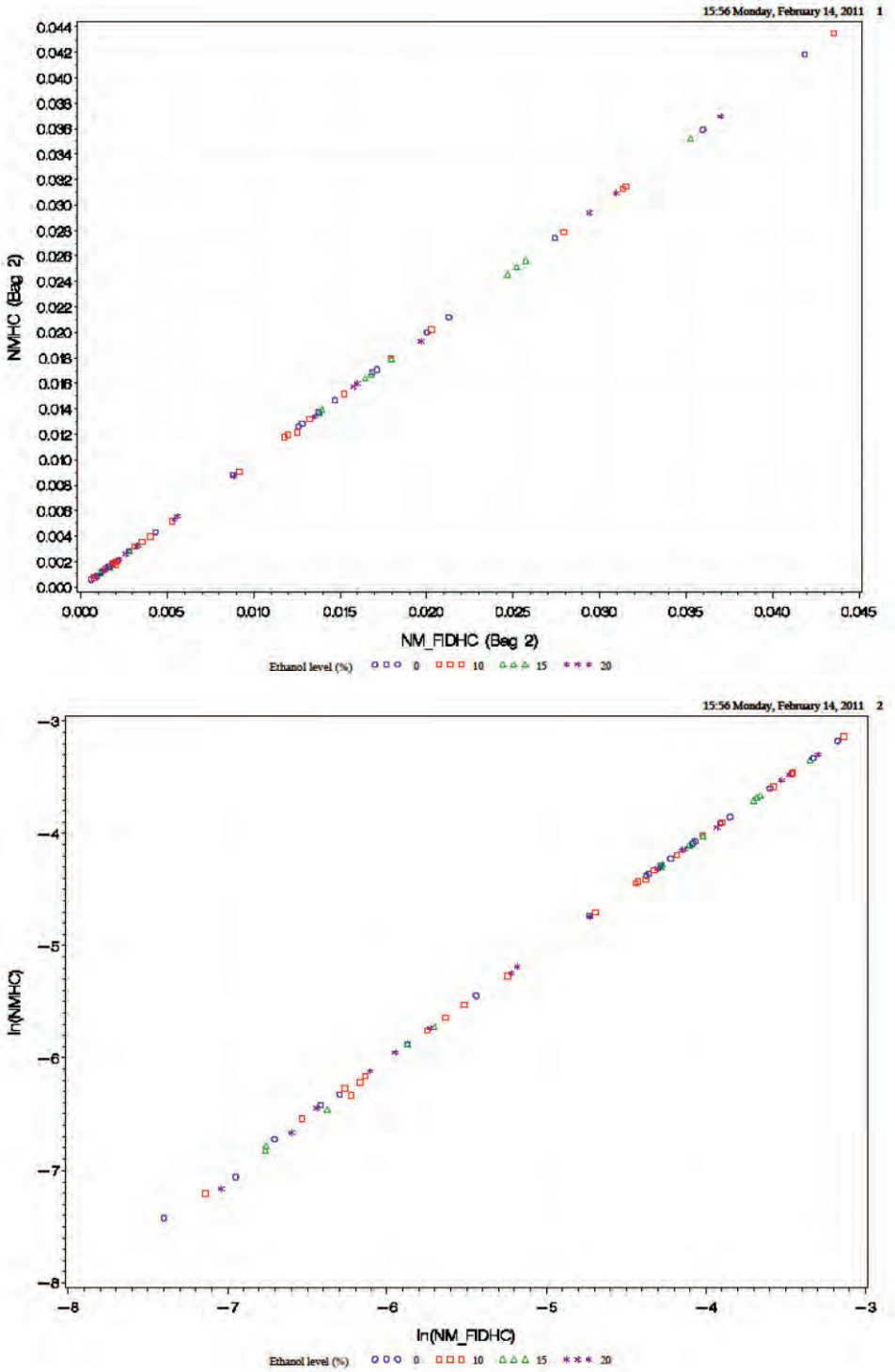


Table 13. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 1 NMHC (dependent variable = NMHC (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	-0.0002943088	0.00055230	-0.53	0.59
Slope (E0)	0.9987517260	0.00118319	844.12	<0.0001
Intercept Increment (E10)	-0.0019509760	0.00078811	-2.48	0.014
Slope Increment (E10)	-0.0209384772	0.00175580	-11.93	<0.0001
Intercept Increment (E15)	-0.0051528497	0.00118670	-4.34	<0.0001
Slope Increment (E15)	-0.0246110233	0.00236599	-10.40	<0.0001
Intercept Increment (E20)	-0.0064657095	0.00085103	-7.60	<0.0001
Slope Increment (E20)	-0.0333647556	0.00199484	-16.73	<0.0001
Fit Information: d.f.[model] = 7, d.f.[error] = 911, d.f.[total] = 918. *F* value = 276,173, Pr > *F* = <0.0001. R^2 = 0.999529, Root MSE = 0.004112.				

Table 14. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 2 NMHC (dependent variable = NMHC (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	-0.000031742	0.00001232	-2.58	0.012
Slope	1.000457255	0.00047680	2098.27	<0.0001
Intercept Increment (E10)	-0.000026796	0.00001450	-1.85	0.069
Intercept Increment (E15)	-0.000039842	0.00001613	-2.47	0.016
Intercept Increment (E20)	-0.000036134	0.00001584	-2.28	0.026
Fit Information: d.f.[model] = 4, d.f.[error] = 64, d.f.[total] = 68. *F* value = 1,108,270, Pr > *F* = <0.0001. R^2 = 0.999986, Root MSE = 0.000046.				

Table 15. Model Coefficients, Tests of Effect, and Goodness-of-fit Parameters for BAG 3 NMHC (dependent variable = NMHC (g/mi), independent variable = $NMHC_{FID}$ (g/mi)).

Parameter	Estimate	Standard Error	*t*-value	Pr > |*t*|
Intercept (E0)	-0.000051391	0.00006850	-0.75	0.46
Slope	1.000231724	0.00144955	690.03	<0.0001
Intercept Increment (E10)	-0.000038854	0.00008424	-0.46	0.65
Intercept Increment (E15)	-0.000021166	0.00009618	-0.22	0.83
Intercept Increment (E20)	-0.000292149	0.00009253	-3.16	0.0024
Fit Information: d.f.[model] = 4, d.f.[error] = 64, d.f.[total] = 68. *F* value = 121,075, Pr > *F* = <0.0001. R^2 = 0.999868, Root MSE = 0.000267.				

4 Data Review

The data collected in this study are difficult to visualize in that they encompass variation of emissions in the five-dimensional fuel-parameter space. Due to human limitations, it is practical to view the data in only two dimensions at a time, in some cases including multiple series to represent levels in a third dimension. Despite the risk of misinterpreting visual portrayals that may oversimplify the actual emissions behavior in all dimensions, it is valuable to review the data visually before model development.

At the outset, it is helpful to get an overview of the raw results, sorted by vehicle and fuel, which gives an initial impression of variability among vehicles and fuels, as well as within vehicles. This view also gives an initial impression of vehicles or observations that may prove influential. In addition, we averaged and plotted the data to check for evidence of "main effects," or "linear effects," i.e., trends in emissions across all levels of a single fuel parameter. We constructed these views by averaging the data by the levels of one fuel parameter and by vehicle, across all levels of the remaining four parameters, repeating the process for each fuel parameter in turn. We took this step for the emission results themselves (i.e., in "linear space"), as well as for natural-log transforms of the data (i.e., "log space"). We made a point of examining the log-transformed results, as the statistical models were developed using the transforms, rather than the raw results.

The study design anticipates the possibility that the response of emissions to changes in multiple fuel parameters may involve several 2-way interactions, which suggests that limiting our examination to "linear effects" may be simplistic. To examine 2-way emissions responses, we also averaged and plotted the data by two fuel parameters simultaneously to examine potential "conditional" or interaction effects, or how the effect of each fuel parameter varied with the levels of the other parameters.

Below, we illustrate these concepts for three sets of results: Bag 1 NO_x, Bag 2 NO_x, and Bag 1 PM. For these compounds, as for others, the ethanol ×T50 interaction gives an example of two interrelated variables and the importance of supplementing "linear-effects" plots with "interaction" plots.

The plots presented and described below, as well as additional plots not shown in this document are presented in Appendices G – J.

4.1 NO_x (Bag 1)

Figure 12 shows the set of observations for Bag 1 NO_x, with the data portrayed as the common logarithm of the measurements (base 10). Across all fuels, the range of variability differs by vehicle. For several vehicles, the range of variability over all fuels spans about one order of magnitude (e.g., Civic, Corolla, Odyssey). The two cleanest vehicles (Focus, Sienna) are also

the most variable, spanning over an order of magnitude. Variability for the remaining vehicles spans half an order of magnitude or less.

4.1.1 Linear Effects

A linear-effects plot for ethanol is shown in Figure 13, which suggests that an ethanol effect is visible when the data are averaged across the other four fuel properties. Trends for individual vehicles show a general increase in NO_x with increasing ethanol, with some exceptions. For example, the Altima and Odyssey show generally declining trends, and several vehicles are lower at 20% ethanol than at 10%, including the Impala, the Outlook and Caliber. The view of $\ln(NO_x)$ is similar, except that it may be suggestive of a shift in trends by vehicle, with vehicles having lower $\ln(NO_x)$ showing stronger ethanol increases than those with higher levels.

In a similar plot for aromatics (Figure 14), an increase in NO_x with increasing aromatics is clearly evident for all vehicles. These results display the common "fan" effect, with trends for vehicles with higher emissions generally steeper than for vehicles with lower emissions. This pattern is consistent with the plot for $\ln(NO_x)$, which shows generally uniform trends among vehicles.

When viewing data averaged by RVP and vehicle (Figure 15), the results suggest no net effect, although individual vehicles show gentle positive or negative trends. This view leaves the impression that RVP may not prove to be an important predictor for Bag 1 NO_x, although it is possible that RVP effects may be masked by other fuel effects.

The linear-effects plot for T50 (Figure 16) is not clearly suggestive of an effect, whether positive or negative. Some patterns within the data raise questions. Some vehicles show a general positive trend, others a general negative trend. A curious feature is that the NO_x level for some vehicles is higher at 165 than at 150 or 190. When viewed as a whole, the view for $\ln(NO_x)$ seems suggestive of a slight negative effect. However, this conclusion would be spurious. The observed pattern stems from the relationship between ethanol and T50, both in fact as well as in the fuel-parameter matrix (See Figure 1). What the linear-effects view obscures is that the data points with lower T50 have higher ethanol and vice versa, with some overlap in the center of the range. Thus, what purports to give a picture of the linear effect of T50 is confounded by the somewhat stronger effect of ethanol. Thus, the case of T50 is an example of case where the linear effects view is misleading.

The corresponding view for T90 (Figure 17) shows mixed results, possibly suggesting positive trends for vehicles with lower NO_x and vice versa. A conspicuous feature is that NO_x at the 325° level tends to be higher than at 340°. A superficial conclusion could be that a quadratic fit for T90 might be appropriate. However, it is important to remember that the points at 325° may not deserve as much weight as those at the other two levels, as only two fuels were assigned T90 levels of 325°. The other properties of these fuels may also be important. One had ethanol at 10% and the other at 20%. T50 for both fuels was correspondingly low (<170°). However, both

had aromatics at the higher level (35%), which may best account for the apparent curvature in the trend.

4.1.2 Interactions

It is necessary to go a step further, and look at "interaction" or "conditional effects" plots, starting with the interaction of ethanol and T50, which deserves special attention due to the remaining correlation between these two properties, which cannot be neutralized by the fuel-matrix design. To construct the plot, we average the data by levels of etOH and T50, across the levels of the other properties, as well as across vehicles. Note that in taking this step, we average by the target fuel properties for each fuel, not the actual measured properties, which have some variability. In plotting the set of means, we can construct two views. We can plot the averages vs. ethanol, with a separate series for each T50 level (etOH×T50), and we can plot averages vs. T50, with a separate series for each ethanol level (T50×etOH).

In the plot for etOH×T50 (Figure 18), the view seems to indicate an upward trend from E0 through E20, but with some downward curvature above E10. At first glance, the picture seems to suggest the existence of negative quadratic trend for ethanol. However, in interpreting this plot, we may need to discount what we think we see. Specifically, we need to consider that not all points on the plot represent equal numbers of fuels, or are equally balanced in terms of the remaining three fuel properties, for which reason it is not certain that all points make equal contributions to the overall trend. For example, the leftmost point in the green trend (etOH=15%, T50=165°), represents a single fuel, which also has the higher aromatics level (35%). Similarly, the rightmost point on the red trend (etOH=10%, T50=190°), represents two fuels with one point having high aromatics and low T90, and the other low aromatics and high T90. Taking these factors into account, recognizing the probable positive effects of aromatics and T90 on NO_x, leads us to the tentative conclusion that it is possible, but not certain, that models may fit a quadratic trend to these data.

Plotting the data averaged by T50, with series for different ethanol levels (T50×etOH), gives another view (Figure 19). At first glance, the trend appears to "zig-zag," from low to high. As with the previous view, however, several points creating this apparent visual effect probably do not make contributions to the main trend as strong as other sets of points that are not as far off the main trend. For example, the left-most point on the red trend (T50=165°, etOH=15%), represents a single fuel with high aromatics, and appears conspicuous as in the previous view (leftmost green). The center point in the green trend (T50=190, etOH=10%) represents the same two fuels as the rightmost red point in the previous plot. The center point in the black trend also remains conspicuous in this view (T50=220, etOH=0%). This point represents two fuels (8,13) which are, in concept, balanced with respect to aromatics and T90, in that fuel 8 has low aromatics and T90, and fuel 13 has high aromatics and T90. Accounting for these points leads to tentative conclusion that the effect for T50 may be positive, but not large.

The interaction between ethanol and aromatics is more straightforward. First, the fuel-parameter designs had better success in balancing the numbers of fuels with respect to these two parameters. Averaging four ethanol levels and two aromatics levels gives eight means. In the etOH×aromatics plot (Figure 20), the eight means are arranged in two series of four points. The impression is that the trend for the lower aromatics level (black) has a somewhat steeper slope than the trend for the higher aromatics level. In the aromatics×etOH view (Figure 21) the eight means give four series with 2 points in each. In this view, the trend for the lowest ethanol level appears to have a steeper slope than those for the remaining three ethanol levels, which have approximately equal slopes. Taken together, these pictures are suggestive of a negative, or "interference" interaction between ethanol and aromatics.

Figure 12. Common logarithm of Bag 1 NO_x, by Vehicle and Fuel. The second view highlights "censored" measurements.

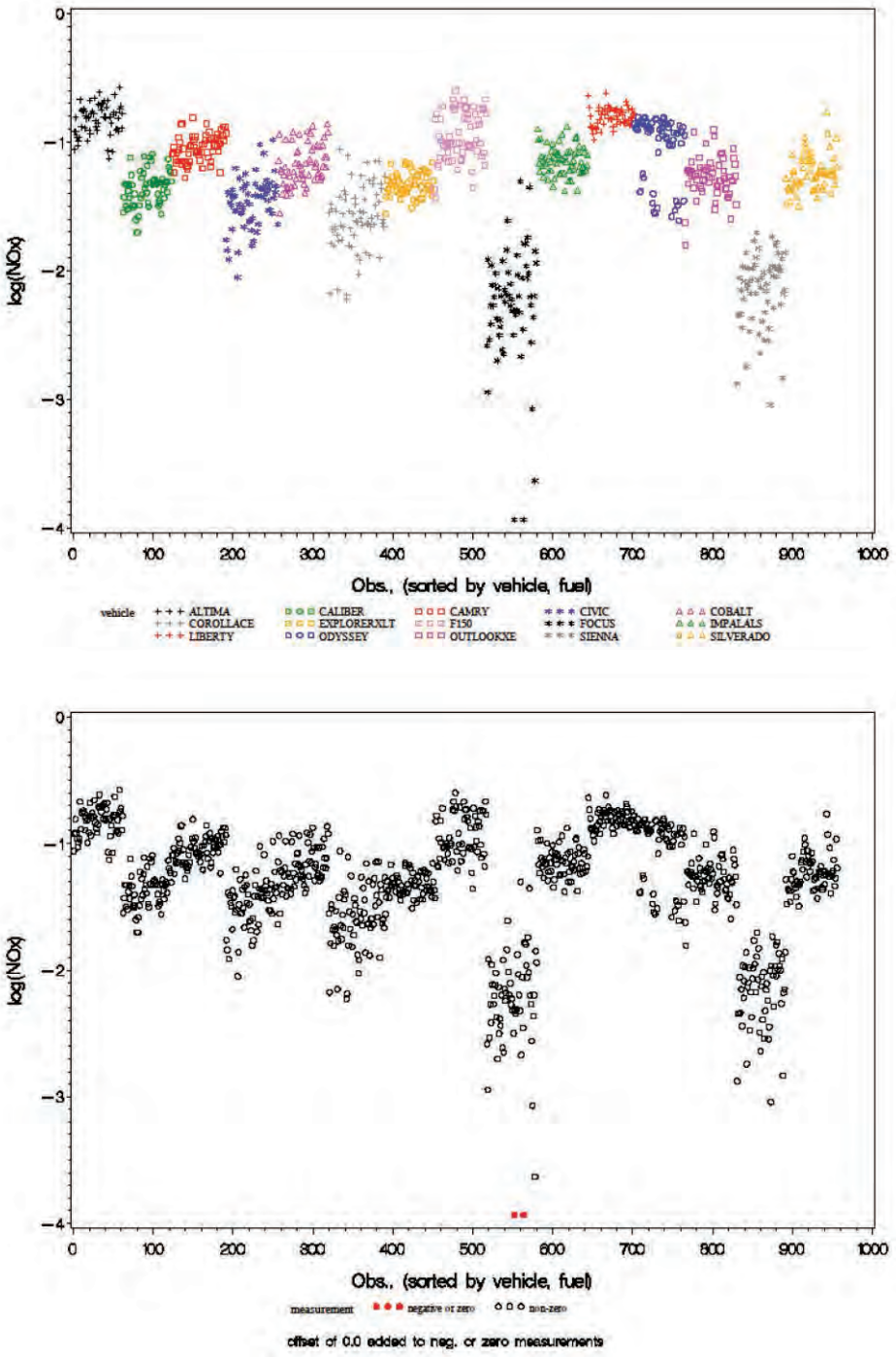


Figure 13. Linear-effects plot for Ethanol on Bag-1 NO_x: NO_x and ln(NO_x) vs. Ethanol Level, with data averaged by four Ethanol levels and by Vehicle.

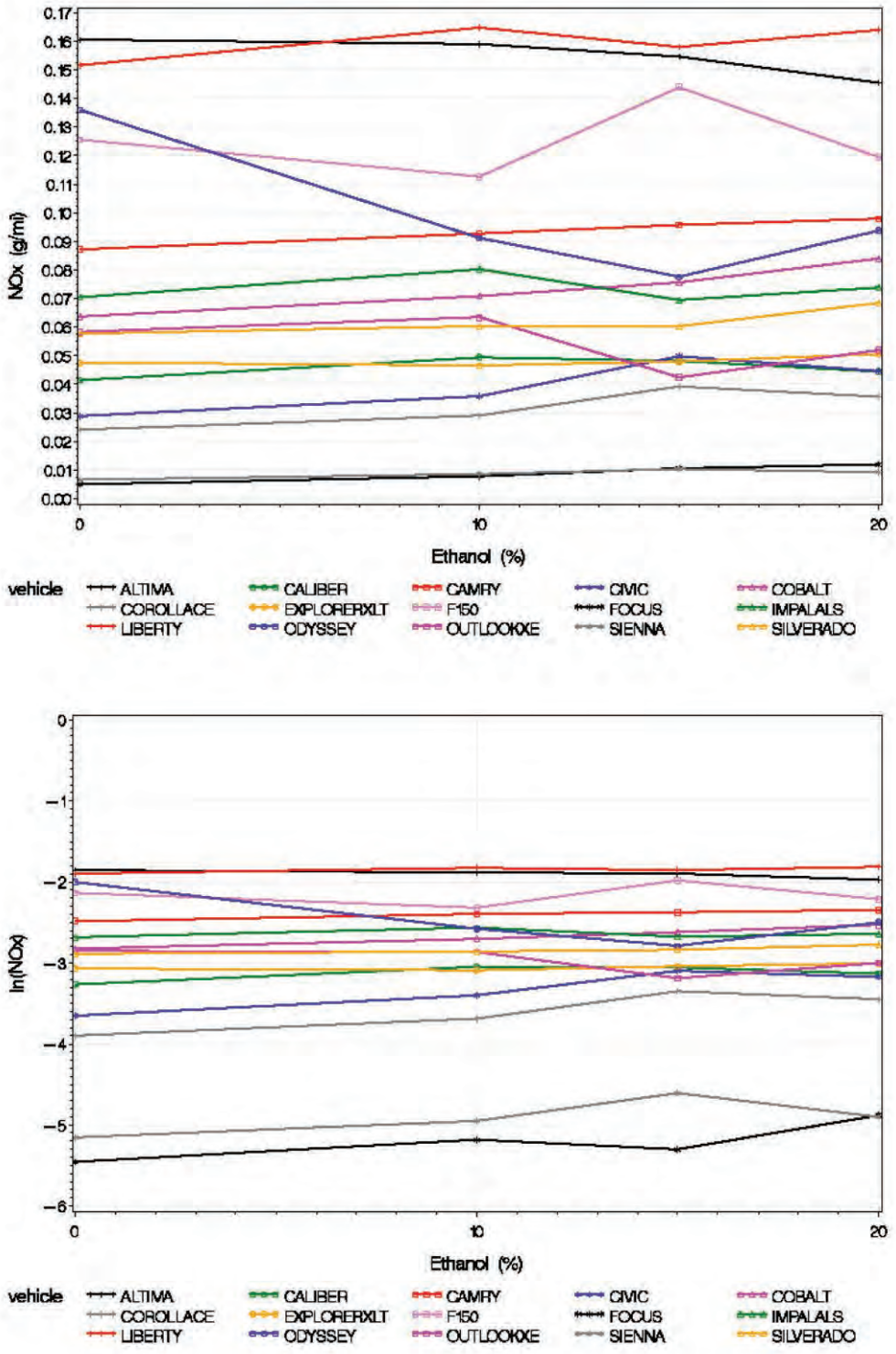
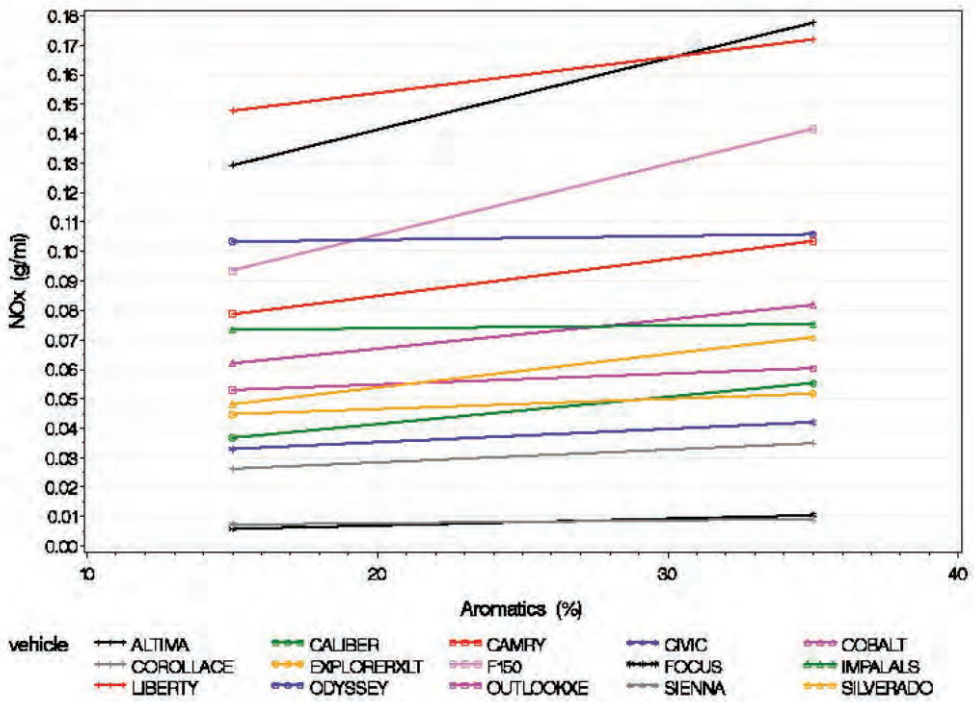


Figure 14. Linear-effects plot for Aromatics on Bag-1 NO_x: NO_x and ln(NO_x) vs. Aromatics Level, with data averaged by two Aromatics levels and by Vehicle.



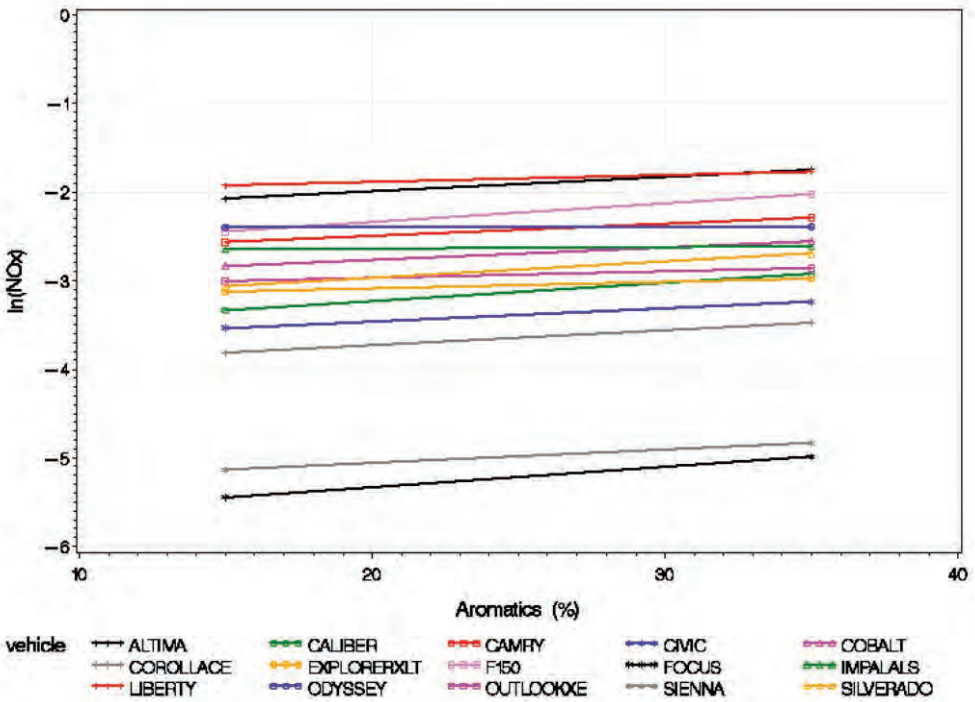


Figure 15. Linear-effects plot for RVP on Bag-1 NO_x: NO_x and ln(NO_x) vs. RVP Level, with data averaged by two RVP levels and by Vehicle.

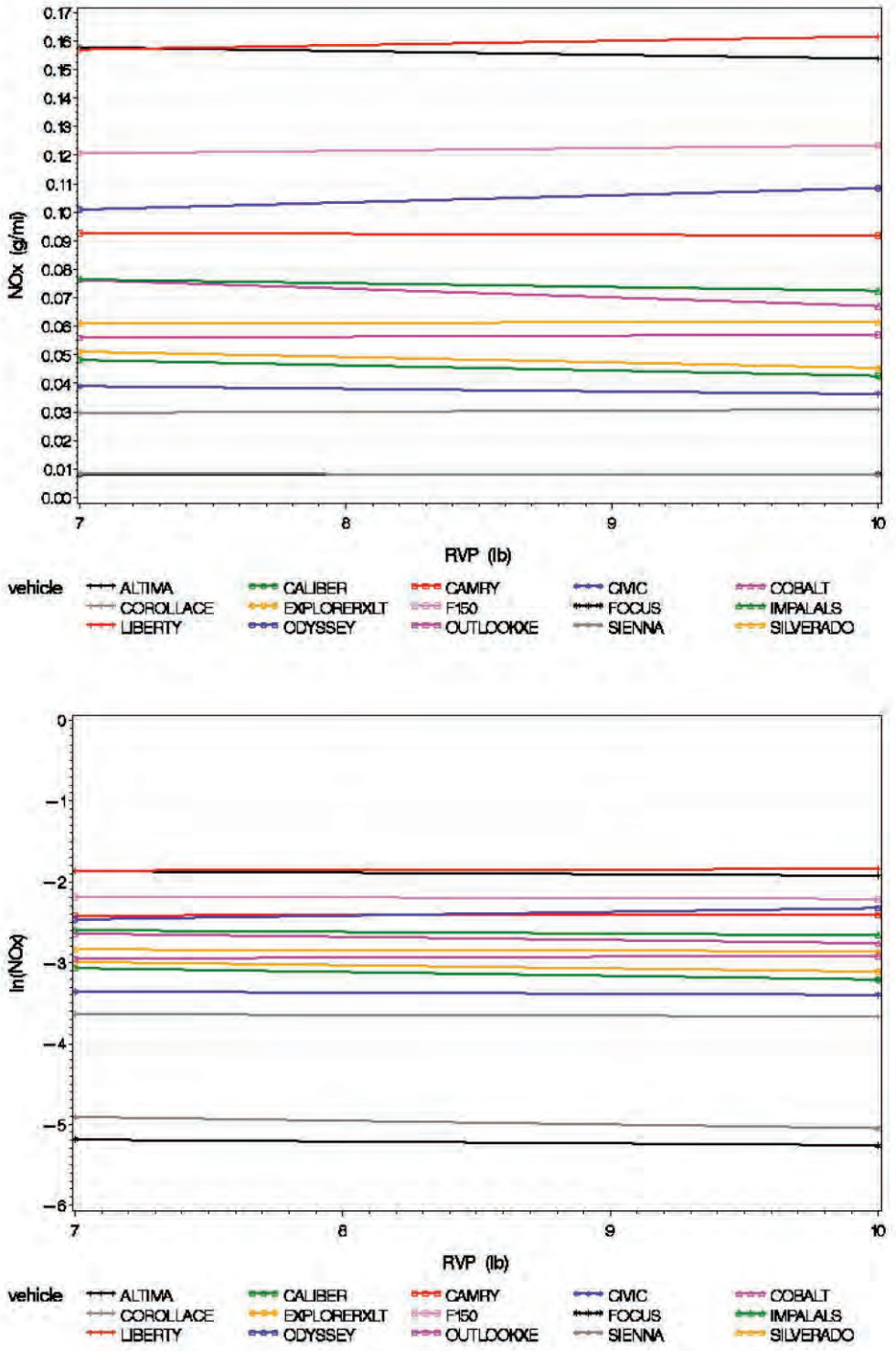


Figure 16. Linear-effects plot for T50 on Bag-1 NO_x: NO_x and ln(NO_x) vs. T50 Level, with data averaged by five T50 levels and by Vehicle.

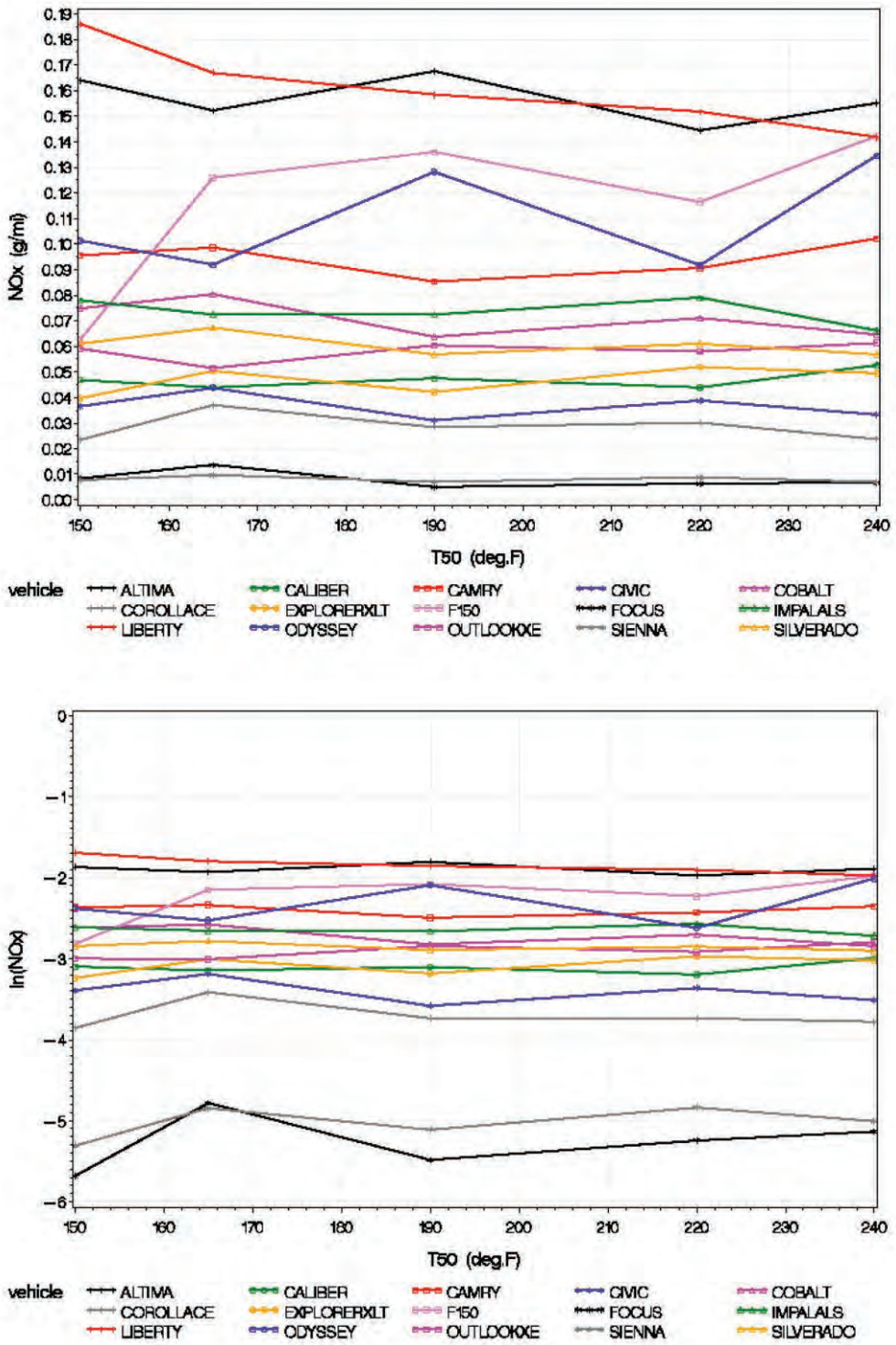


Figure 17. Linear-effects plot for T90 on Bag-1 NO_x: NO_x and ln(NO_x) vs. T90 Level, with data averaged by three T90 levels and by Vehicle.

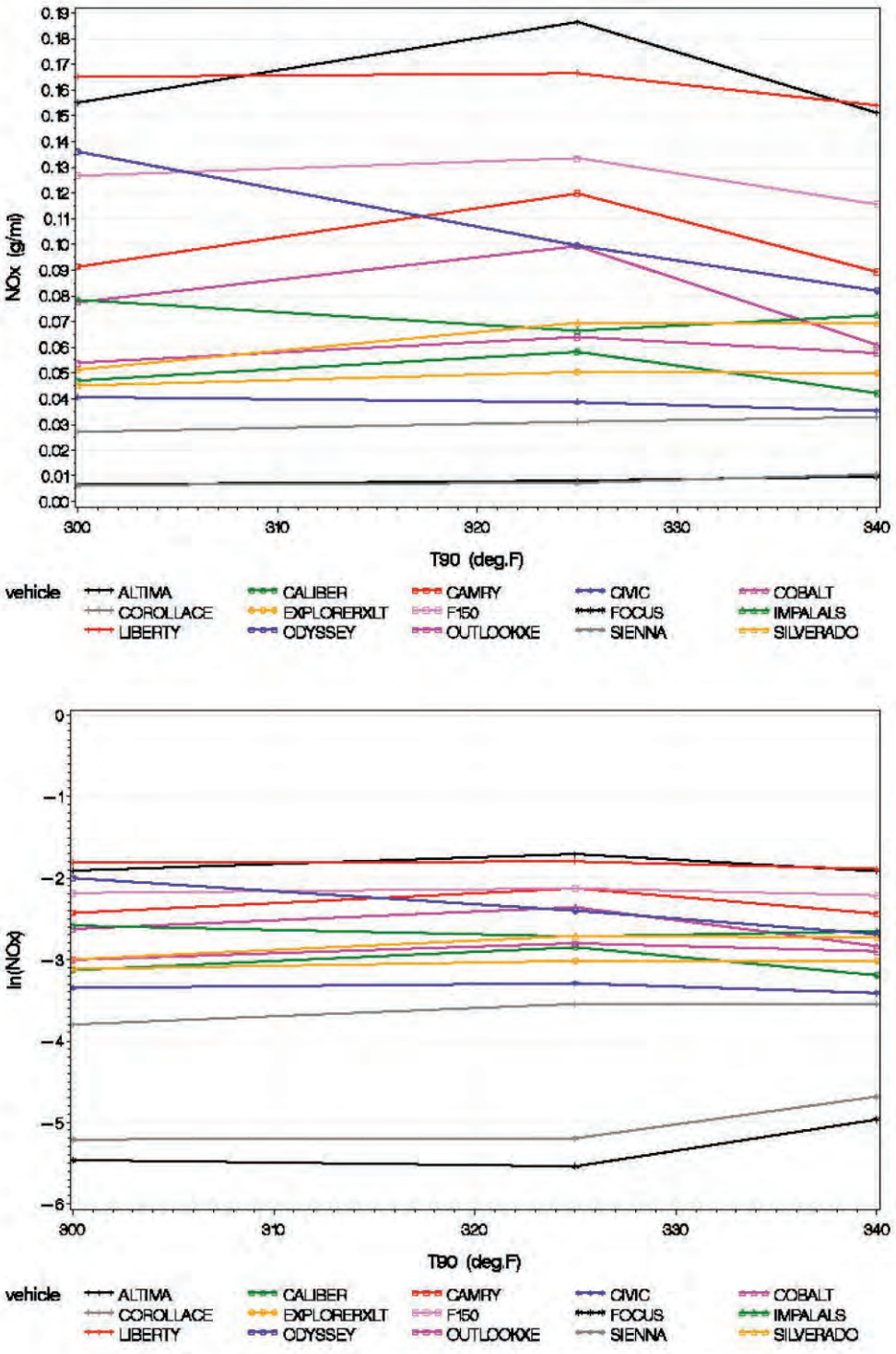


Figure 18. Interaction plot for EtOH×T50: Data are averaged by Ethanol and T50 levels, and plotted vs. Ethanol level, with a separate series for each T50 level. Note that not all points on the plot represent the same number of fuels.

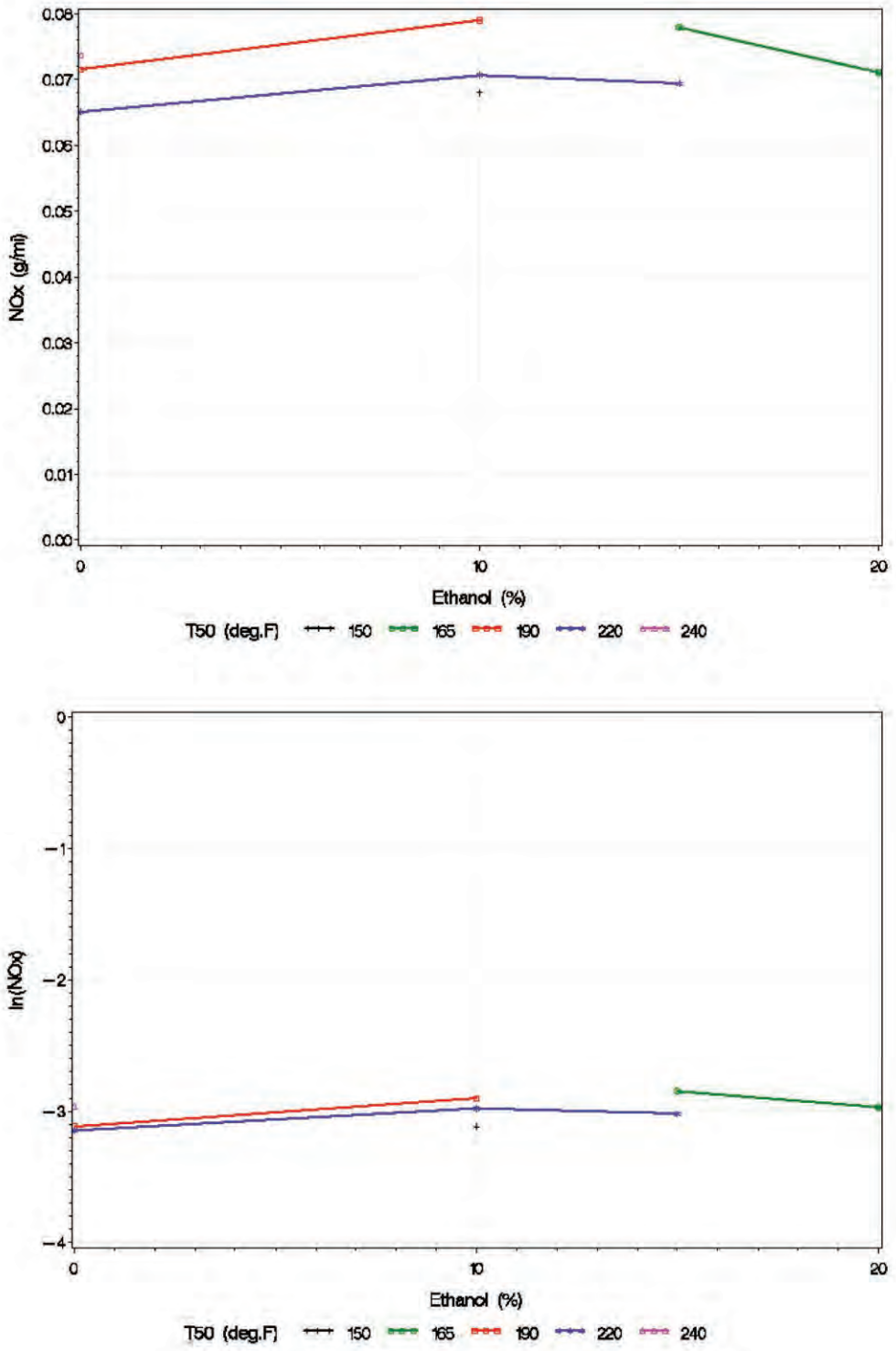


Figure 19. Interaction plot for T50×etOH: Data are averaged by T50 and etOH levels, and plotted vs. T50 level, with a separate series for each ethanol level. Note that not all points on the plot represent the same number of fuels.

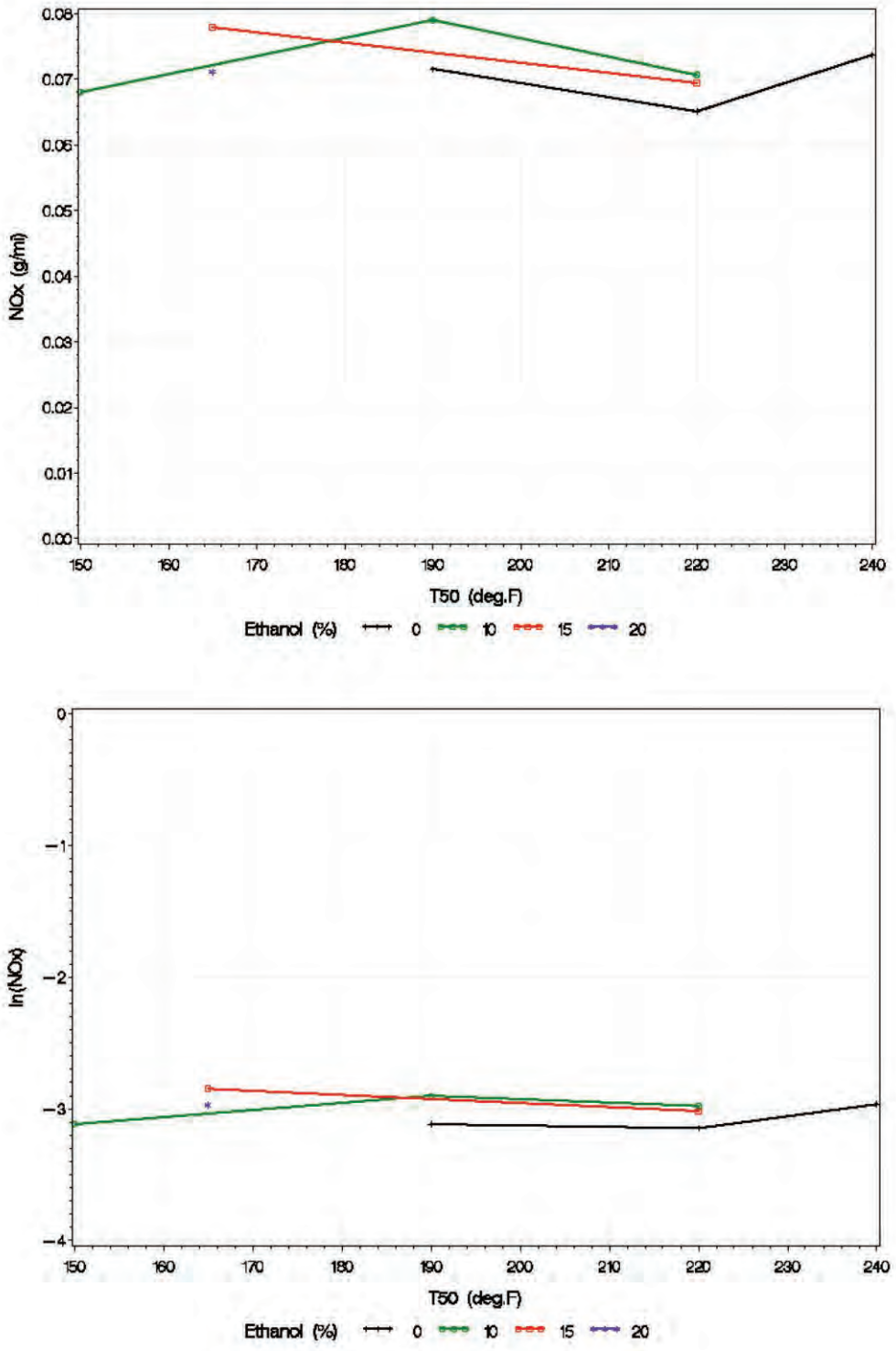


Figure 20. Interaction plot for etOH×Aromatics: Data are averaged by etOH and Aromatics levels, and plotted vs. Ethanol level, with a separate series for each Aromatics level. Note that not all points on the plot represent the same number of fuels.

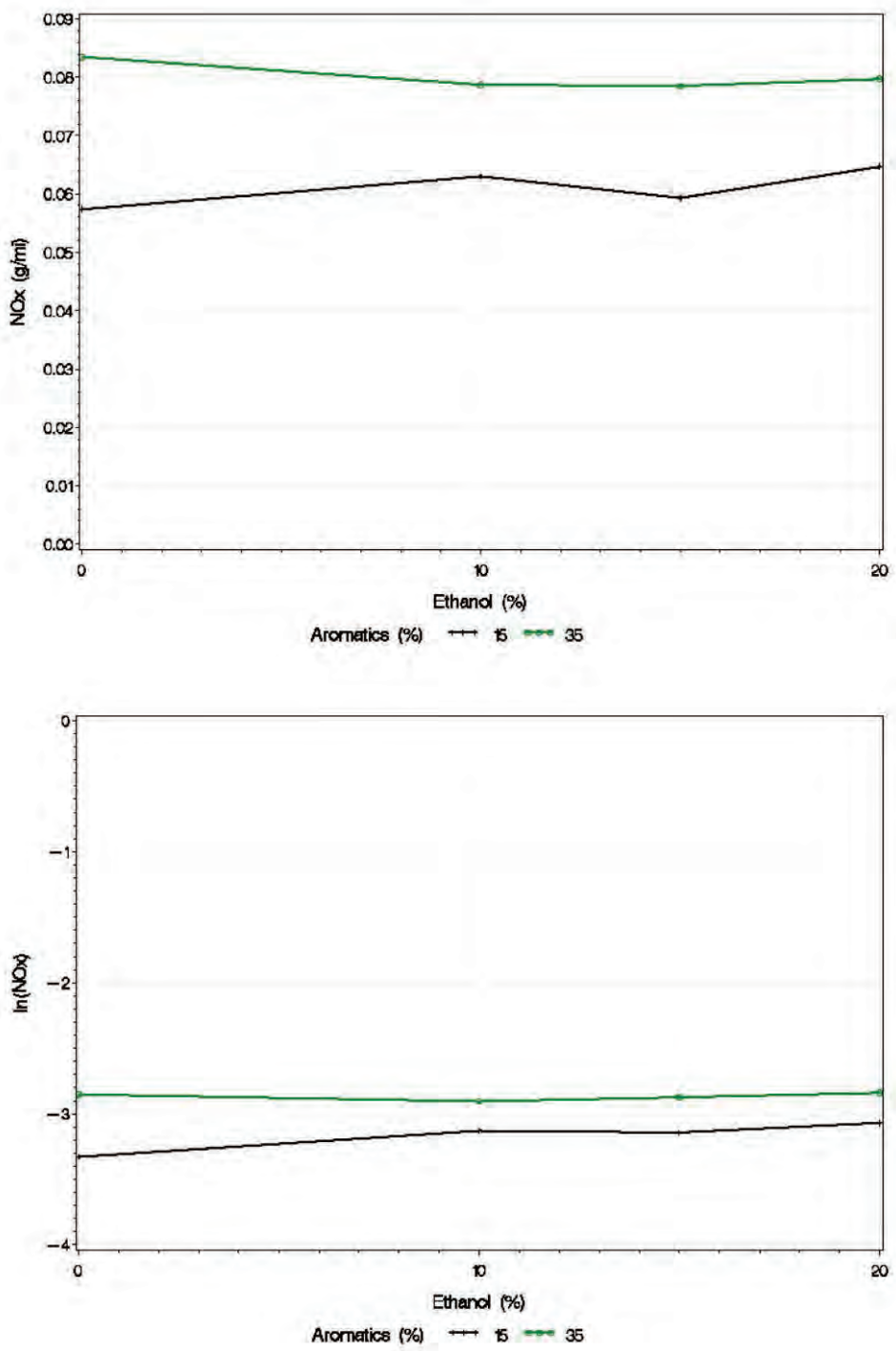
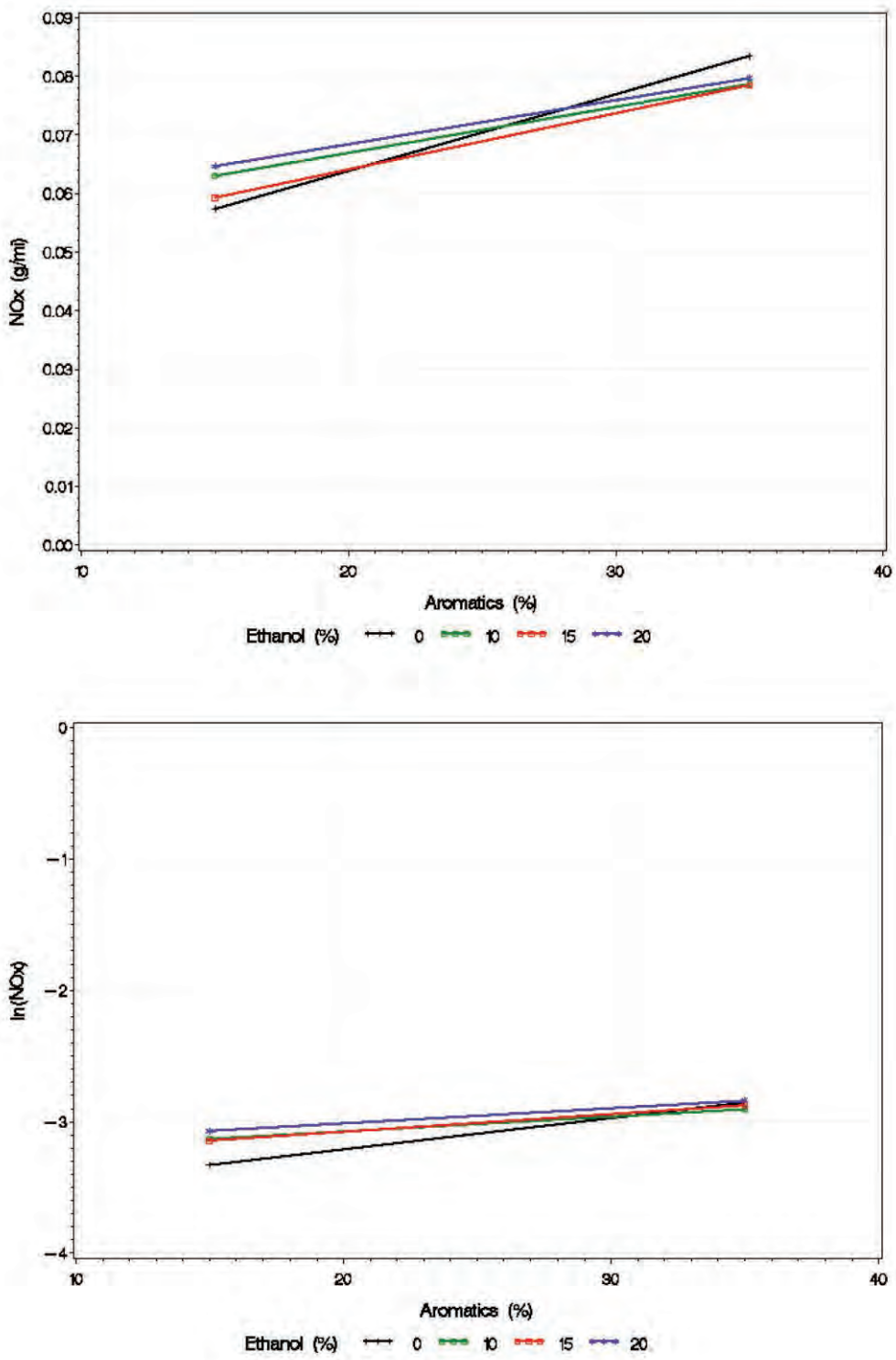


Figure 21. Interaction plot for etOH×Aromatics: Data are averaged by etOH and Aromatics levels, and plotted vs. Ethanol level, with a separate series for each Aromatics level. Note that not all points on the plot represent the same number of fuels.



4.2 NO_x (Bag 2)

Overall, the patterns observed for NO_x in Bag 2 are quite similar to those for Bag 1. The view of all data, sorted by vehicle and fuel, looks similar to Bag 1, except that, as expected, Bag-2 values are roughly one order of magnitude lower than values in Bag 1 (Figure 22). Overall, the data in both Bags span about the same range, 2.5 orders of magnitude when including "low" vehicles, i.e., Focus and Sienna for Bag 1, and Cobalt for Bag 2, or 1.5 orders of magnitude, excluding the "low" vehicles. As in Bag 1, variability for most individual vehicles spans one third to half an order of magnitude. Also as in Bag 1, there are several censored measurements for the vehicle with the lowest measurements (Cobalt).

4.2.1 Linear Effects

The linear-effects plot for ethanol shows a general increasing trend with increasing ethanol (Figure 23). An apparent anomaly is that several vehicles show higher NO_x at the 15% level than at 10% or 20% levels, ostensibly due to the high aromatics or high T90 levels for these fuels. Several vehicles show no apparent trend. Two vehicles, Civic and Odyssey, track very closely, and show generally steeper trends than the other vehicles. In the plot of $\ln(NO_x)$, the general NO_x increase is apparent, with three exceptions, the Corolla, the Focus and the Cobalt.

The plot for Bag-2 aromatics is quite similar to its counterpart for Bag 1 (Figure 24). Most vehicles show an increase with aromatics, but the trends for the vehicles with higher emissions tend to be steeper. In the logarithmic plot, the increase is apparent, although small, except for the Cobalt, which shows a noticeable decrease.

As in Bag 1, the trends across the two RVP levels vary by vehicle, with no apparent overall pattern across vehicles (Figure 25). The effects for T50 and T90 are similar in that the trends across the five T50 levels and three T90 levels are similar, with no overall trend apparent, for the same reasons noted above for Bag 1 (Figure 26, Figure 27). With respect to T50 in particular, we avoid drawing conclusions from the Linear Effects plot, pending review of the interaction plots.

4.2.2 Interactions

With respect to ethanol and T50, the plot of etOH by T50 (Figure 28) shows a pattern quite similar to that in Bag 1. Overall the pattern suggests a small positive ethanol effect. The view of T50×etOH for Bag 2 is also similar, but not identical to its counterpart for Bag 1 (Figure 29). On the whole, it is not clear that this plot (particularly the logarithmic view), suggests an evident effect for T50.

The interaction plots for ethanol and aromatics are suggestive of a negative or "interference" interaction between these two properties. In the etOH×Arom plot (Figure 30), the slope for the low aromatics level appears steeper than for the high aromatics level. Similarly, in the Arom × etOH plot (Figure 31), the slope for the lowest ethanol level appears to be steeper than those for

the higher levels. It is not clear, though, that the apparent effect will be considered significant when models are fit.

Figure 22. Common logarithm of Bag-2 NO_x, by Vehicle and Fuel. The second view highlights "censored" measurements.

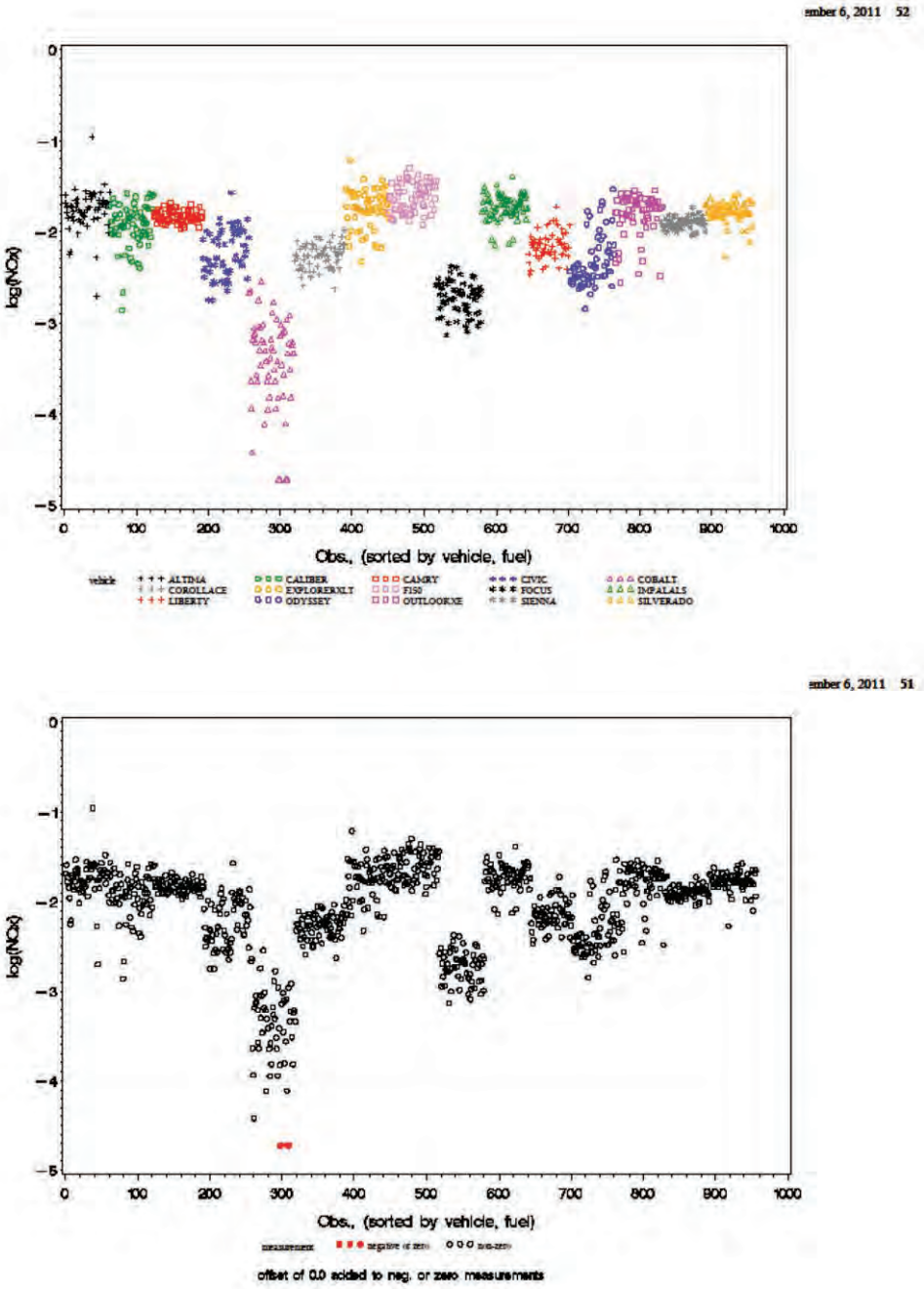


Figure 23. Linear-effects plot for Ethanol on Bag-2 NO_x: NO_x and ln(NO_x) vs. Ethanol Level, with data averaged by four Ethanol levels and by Vehicle.

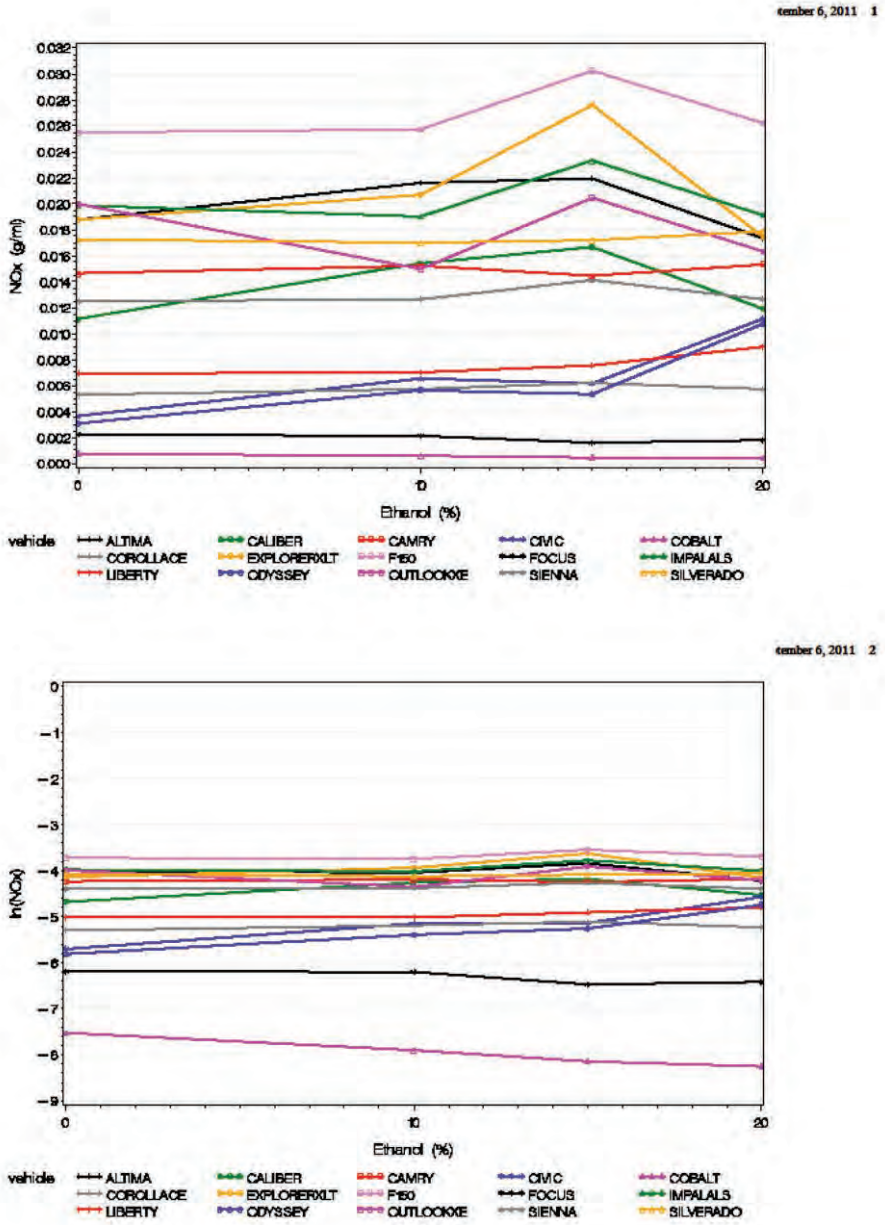


Figure 24. Linear-effects plot for Aromatics on Bag-2 NO_x: NO_x and ln(NO_x) vs. Aromatics Level, with data averaged by two Aromatics levels and by Vehicle.

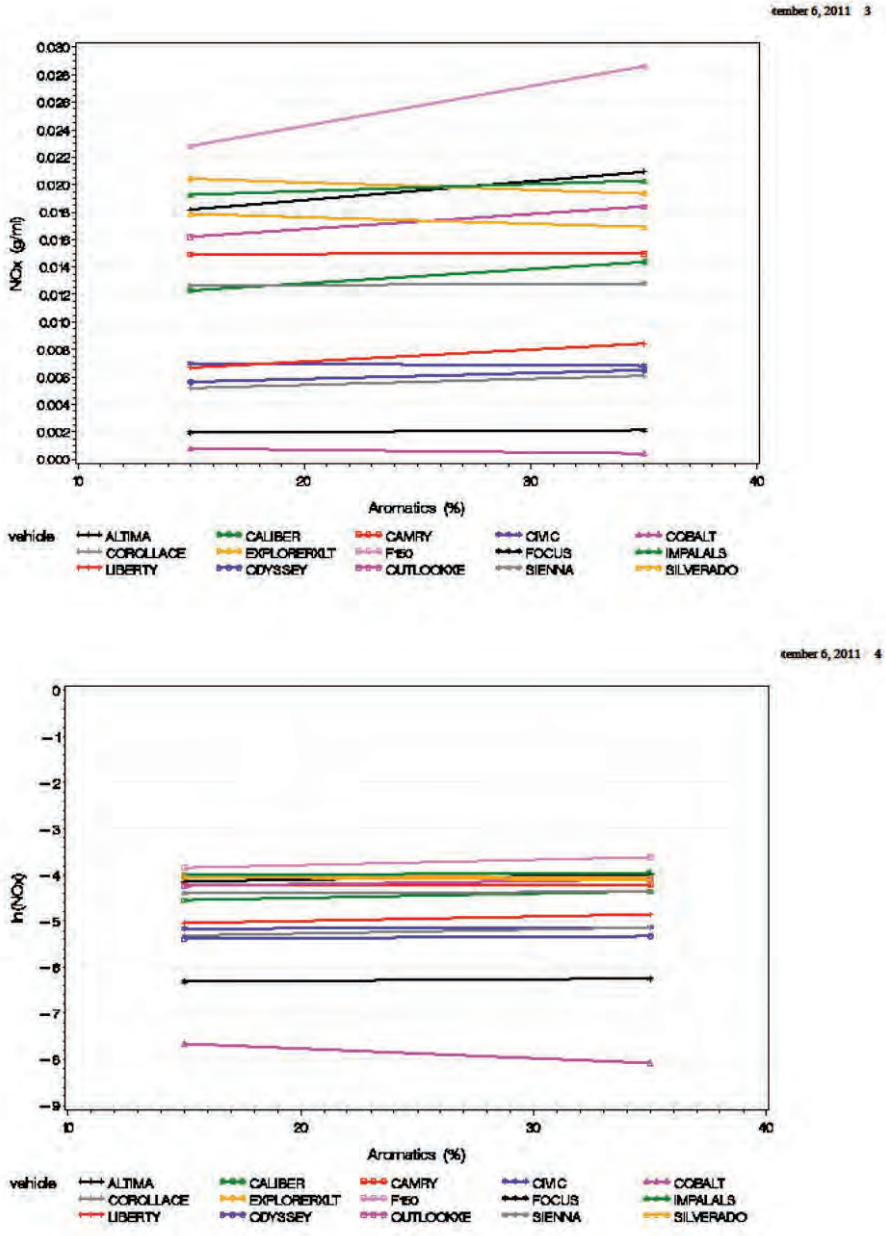


Figure 25. Linear-effects plot for RVP on Bag-2 NO_x: NO_x and ln(NO_x) vs. RVP Level, with data averaged by two RVP levels and by Vehicle.

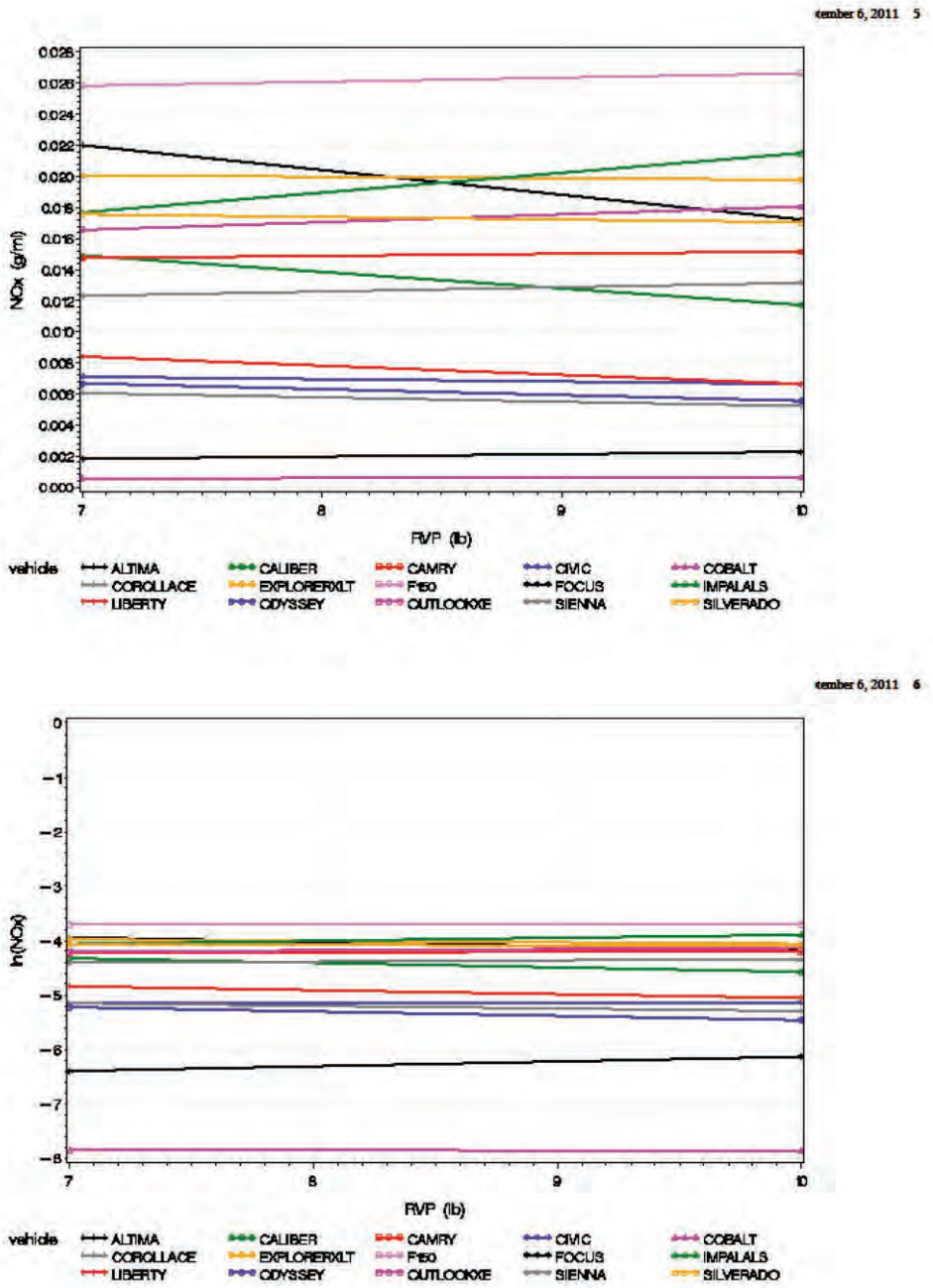


Figure 26. Linear-effects plot for T50 on Bag-2 NO_x: NO_x and ln(NO_x) vs. T50 Level, with data averaged by five T50 levels and by Vehicle.

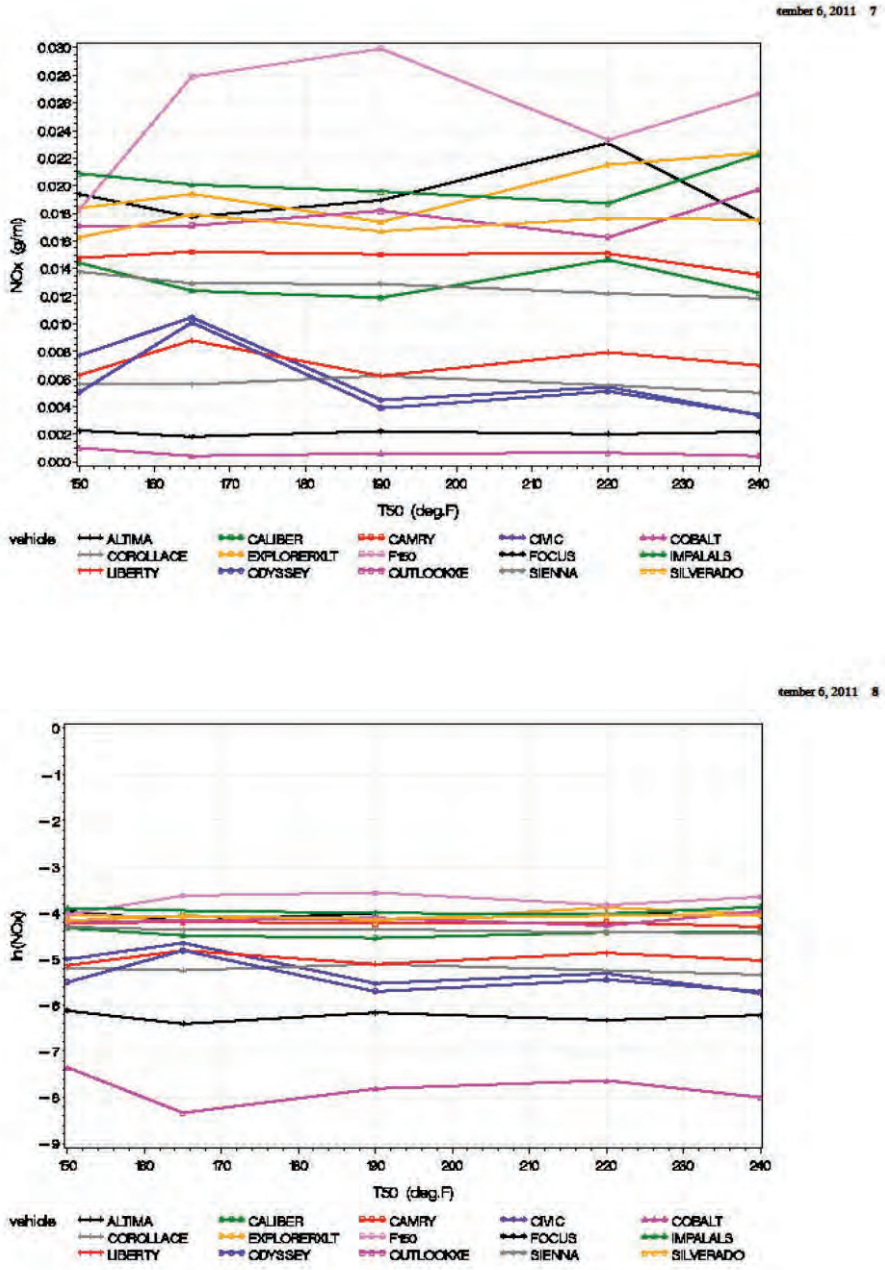


Figure 27. Linear-effects plot for T90 on Bag-2 NO_x: NO_x and ln(NO_x) vs. T90 Level, with data averaged by three T90 levels and by Vehicle.

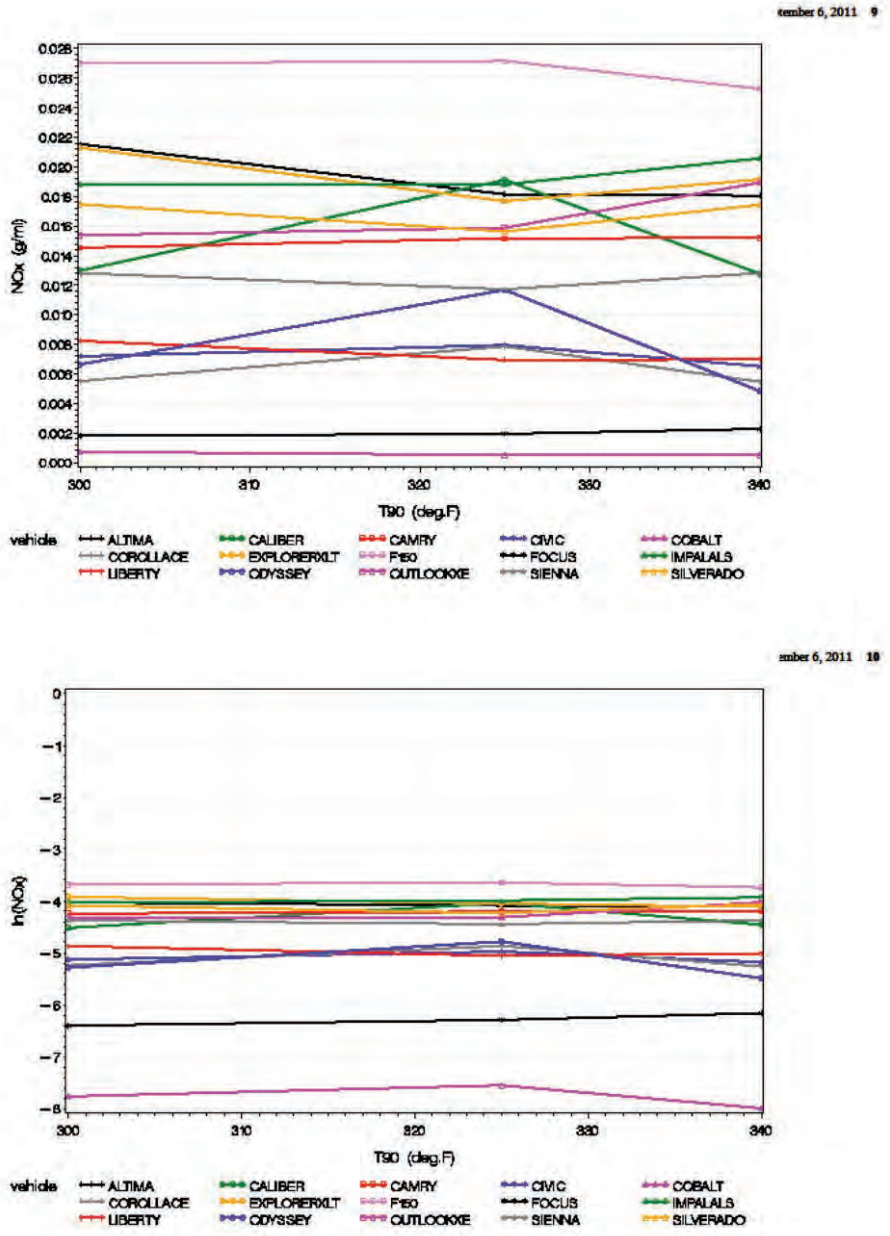


Figure 28. Bag-2 NO_x: Interaction plot for EtOH×T50. Data are averaged by Ethanol and T50 levels, and plotted vs. Ethanol level, with a separate series for each T50 level. Note that not all points on the plot represent the same number of fuels.

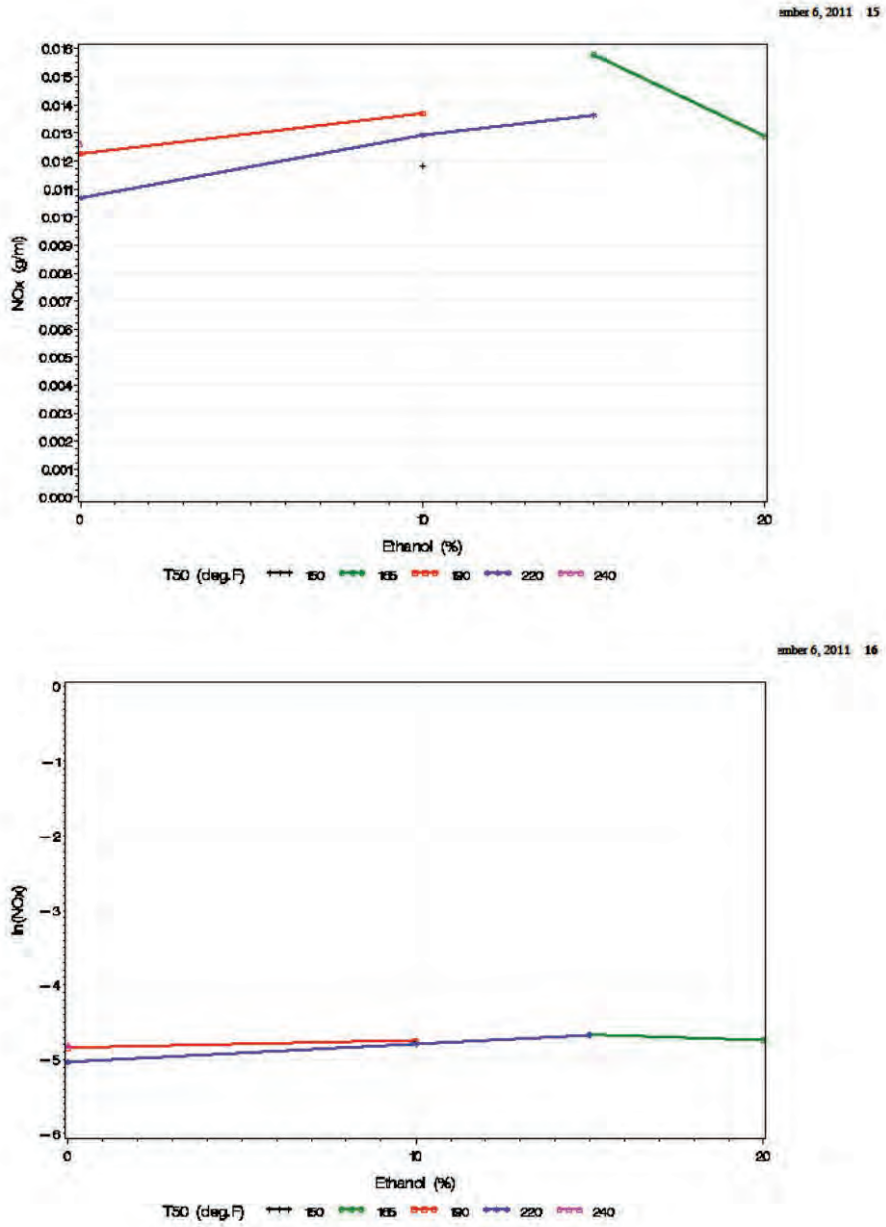


Figure 29. Bag-2 NO_x: Interaction plot for T50×etOH. Data are averaged by T50 and ethanol levels, and plotted vs.T50 level, with a separate series for each ethanol level. Note that not all points on the plot represent the same number of fuels.

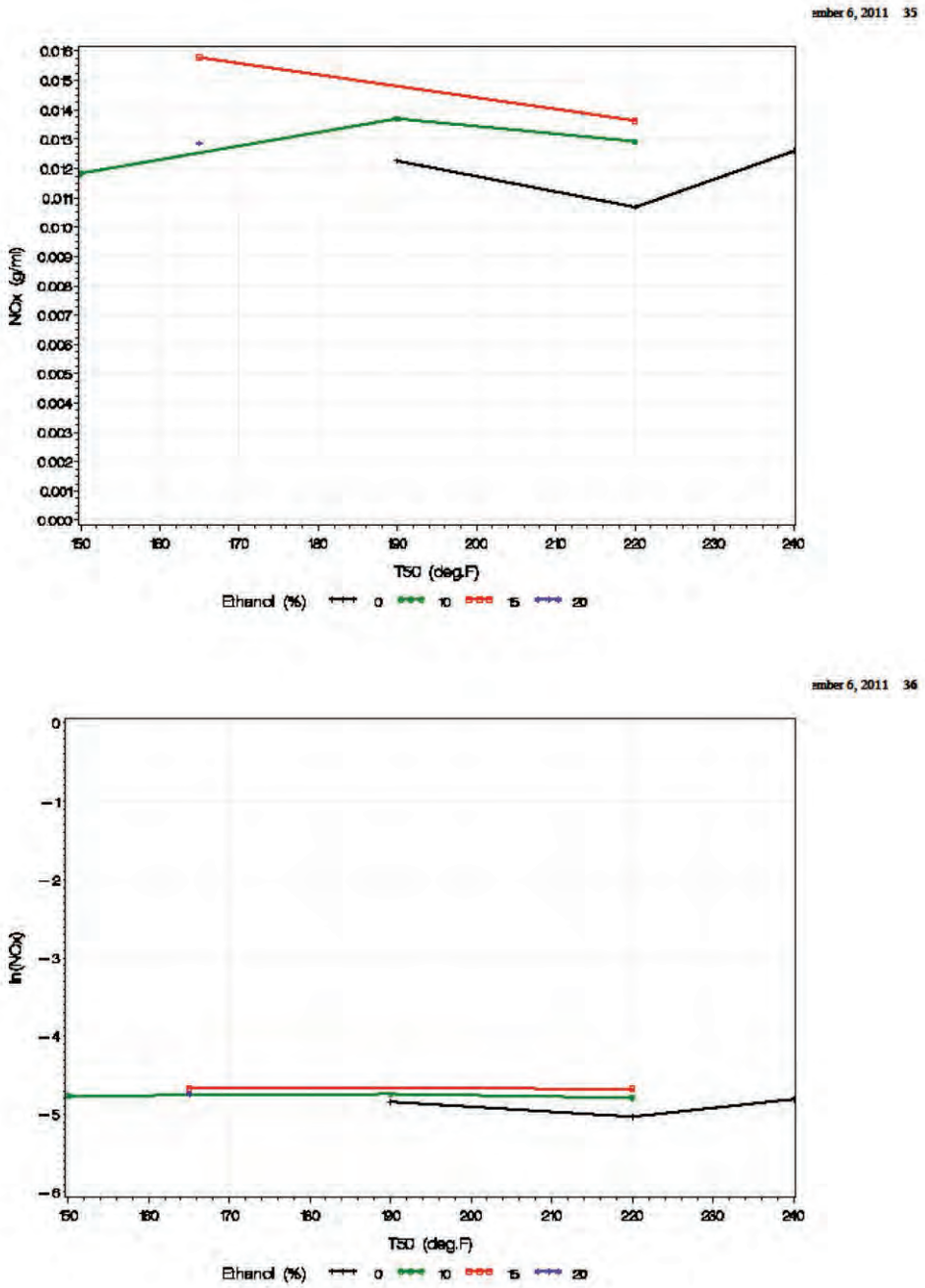


Figure 30. Bag-2 NO_x: Interaction plot for EtOH×Aromatics. Data are averaged by Ethanol and Aromatics levels, and plotted vs. Ethanol level, with a separate series for each Aromatics level. Note that not all points on the plot represent the same number of fuels.

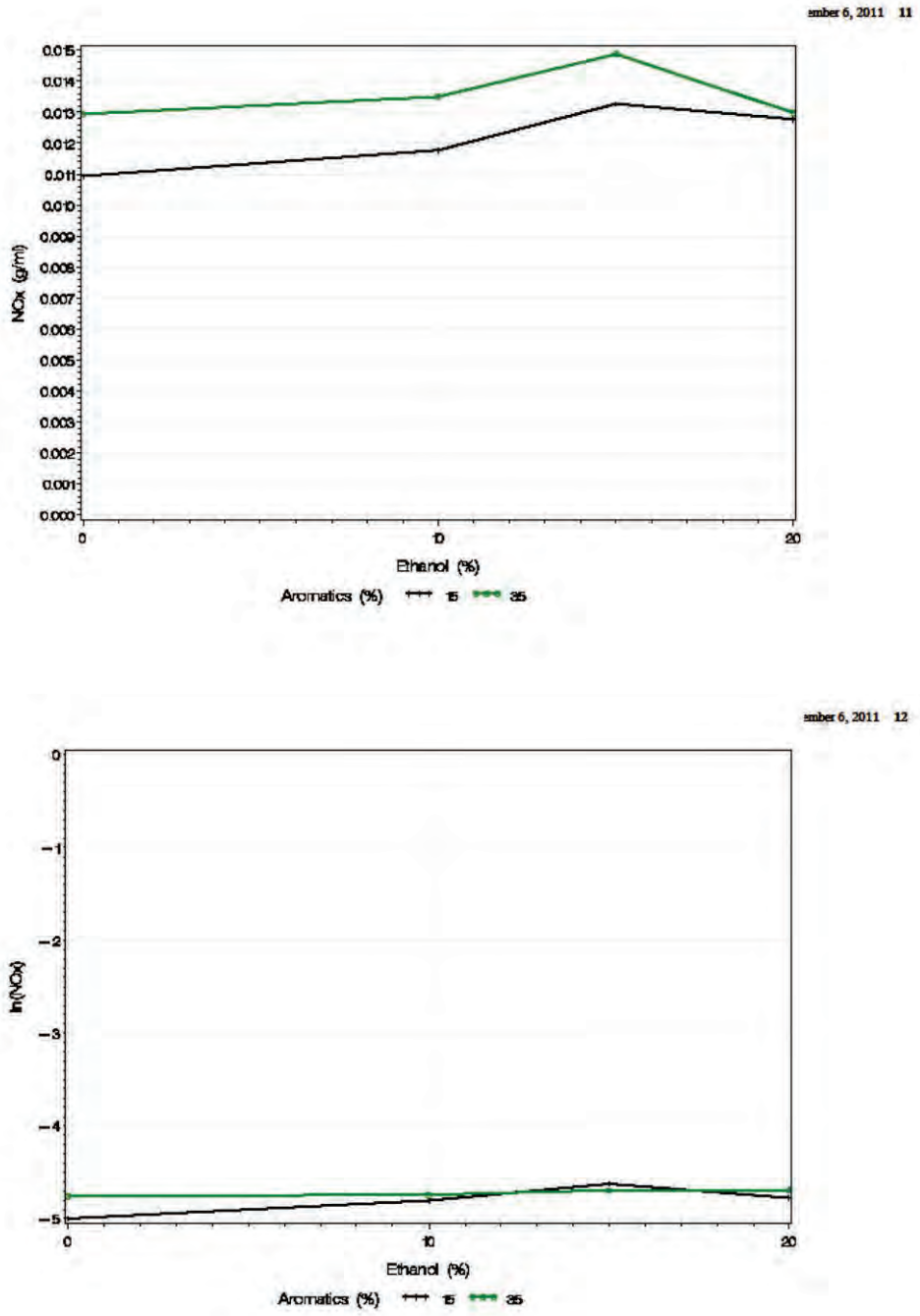
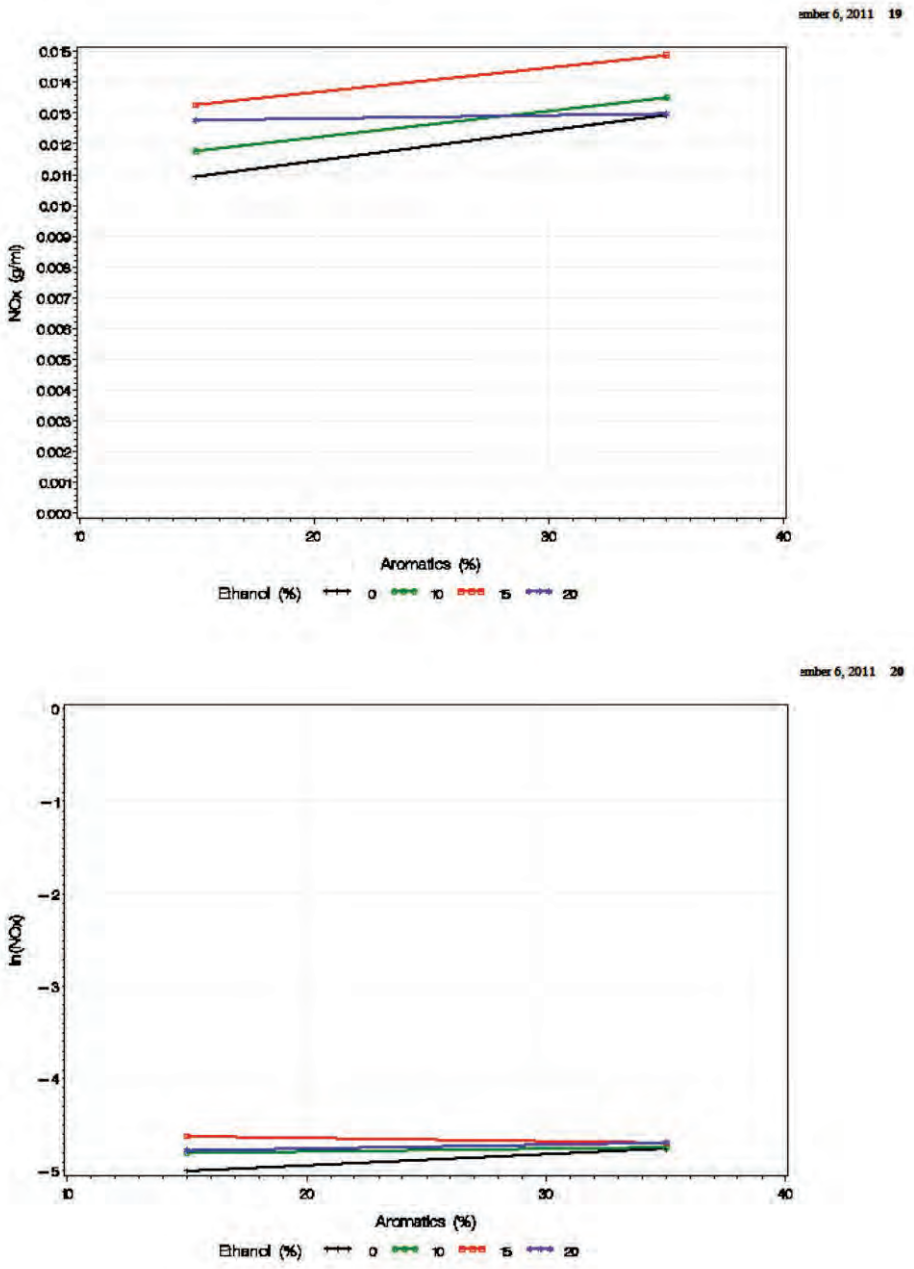


Figure 31. Bag-2 NO_x: Interaction plot for Aromatics×etOH. Data are averaged by Aromatics and ethanol levels, and plotted vs. Aromatics level, with a separate series for each ethanol level. Note that not all points on the plot represent the same number of fuels.



4.3 Particulate Matter (PM, Bag 1)

As with NO_x, we begin with a view of all results, by vehicle and fuel, shown in Figure 32. As before, this view presents the data as common logarithms, and includes censored values and two apparent outlying measurements. The variability within vehicles is about 1-1.25 orders of magnitude. Interestingly, the variability between vehicles is not large. A striking exception to this pattern is the Liberty, which is considerably higher than the other vehicles. This vehicle also has a very high, apparently outlying observation. While not as extreme, the Corolla also has a single measurement that appears high relative to the remaining measurements on the same vehicle. About a third of the vehicles have measurements that appear quite low in comparison to the remaining measurements on their respective vehicles. Also, for this set of results, 45 out of 913 measurements are censored, shown in red at the bottom of the figure.

Note that the linear-effect and interaction plots presented below were generated after excluding the two outlying observations and all censored measurements. Both outliers and censored values can affect the apparent patterns in the resulting means, and in any case, it is not possible to reflect censored values in logarithmic plots.

4.3.1 Linear Effects

The Linear Effects plot for ethanol shows some mixed results (Figure 33), but with an apparent increase from 0% to 10% ethanol, followed by a leveling or decline at higher ethanol levels. It appears possible that models may fit a quadratic as well as a linear term for ethanol.

The plot for aromatics shows a pronounced aromatics effect with the results for different vehicles arrayed in a "fan" (Figure 34). The vehicle with highest PM (Liberty) has steeper trend than the other vehicles in the linear plot (top). In the logarithmic plot, most vehicles have similar slopes (accounting for some degree of variability). It is interesting to note that the logarithmic trend for the Liberty is similar to those for the other vehicles, although its emissions are a full order of magnitude higher. The results lend to some support to the assumption that the effect of aromatics (and other properties) on PM (and other emissions) can be expressed multiplicatively and is similar across vehicles, even across widely ranging emission levels.

As with NO_x, the linear-effects plot for RVP shows mixed results, with increases for some vehicles but decreases for others (Figure 35). This plot is suggestive of no overall effect but it is difficult to draw a conclusion, without having viewed interaction plots.

The Linear Effects plot for T50 also appears suggestive of no overall effect, or perhaps a slight negative effect (Figure 36). But this impression is almost certainly spurious, for the same reasons as described for Bag 1 NO_x, above. It is necessary to defer conclusions pending review of interaction plots model results.

The Linear Effects plot for T90 is clearly suggestive of an overall positive effect, when considering all vehicles (Figure 37). An apparent anomaly is that the data points at 325° are higher than at the other two levels. This outcome can be ostensibly attributed to both fuels at this T90 level having high aromatics, with one of the fuels having 20% ethanol and the other 10%.

4.3.2 Interactions

The view of etOH × T50 is not clearly suggestive of an interaction (Figure 38), although the trend for 165° (green) appears anomalous. However, we can probably discount the results at this temperature. The leftmost green point represents a single fuel, with both high aromatics and high T90, which may be expected to give elevated PM, as we have seen. The point on the right represents an additional five fuels, but appears more in line with the main trend. The right-hand point of blue trend (220°) represents two fuels, balanced in aromatics and T90, but with low RVP. In the complementary plot of T50 × etOH (Figure 39), the trend for 15% ethanol (red) looks anomalous. The left-hand point represents the same fuel as the left-hand point in the green trend in the previous plot. Aside from this one point, the behavior of the trends for 0% and 10% ethanol (green and black) may suggest a positive quadratic curvature with respect to T50, but does not obviously suggest an interaction between ethanol and T50.

In contrast, the plots for ethanol and aromatics are suggestive of a positive or "reinforcement" interaction. In the etOH × arom view (Figure 40), the trend for the higher aromatics level (green) appears steeper than for the lower aromatics level (black). Similarly, in the arom × etOH view (Figure 41) the trends with respect to aromatics increase in steepness with ethanol level, with (typical) exception of the trend for 15%. As before, we can discount this result; the left-hand point in the 15% trend represents a single fuel, with low aromatics but high T90, and the right-hand point end only 2 fuels. Of course, any conclusions based on the visual review must remain tentative pending the generation of modeling results.

The plot of aromatics ×T90 may suggest a reinforcement interaction, as the trend for 340 is steeper than that at 300 (Figure 42). Both fuels with T90 at 325 have 35% aromatics and so do not help confirm or rule out an interaction. The view of T90 × Aromatics gives a similar picture (Figure 43), with the trend with respect to T90 steeper for high aromatics (green) than for low aromatics (black).

Figure 32. Common logarithms of Particulate Measurements for Bag 1. The bottom view highlights censored measurements.

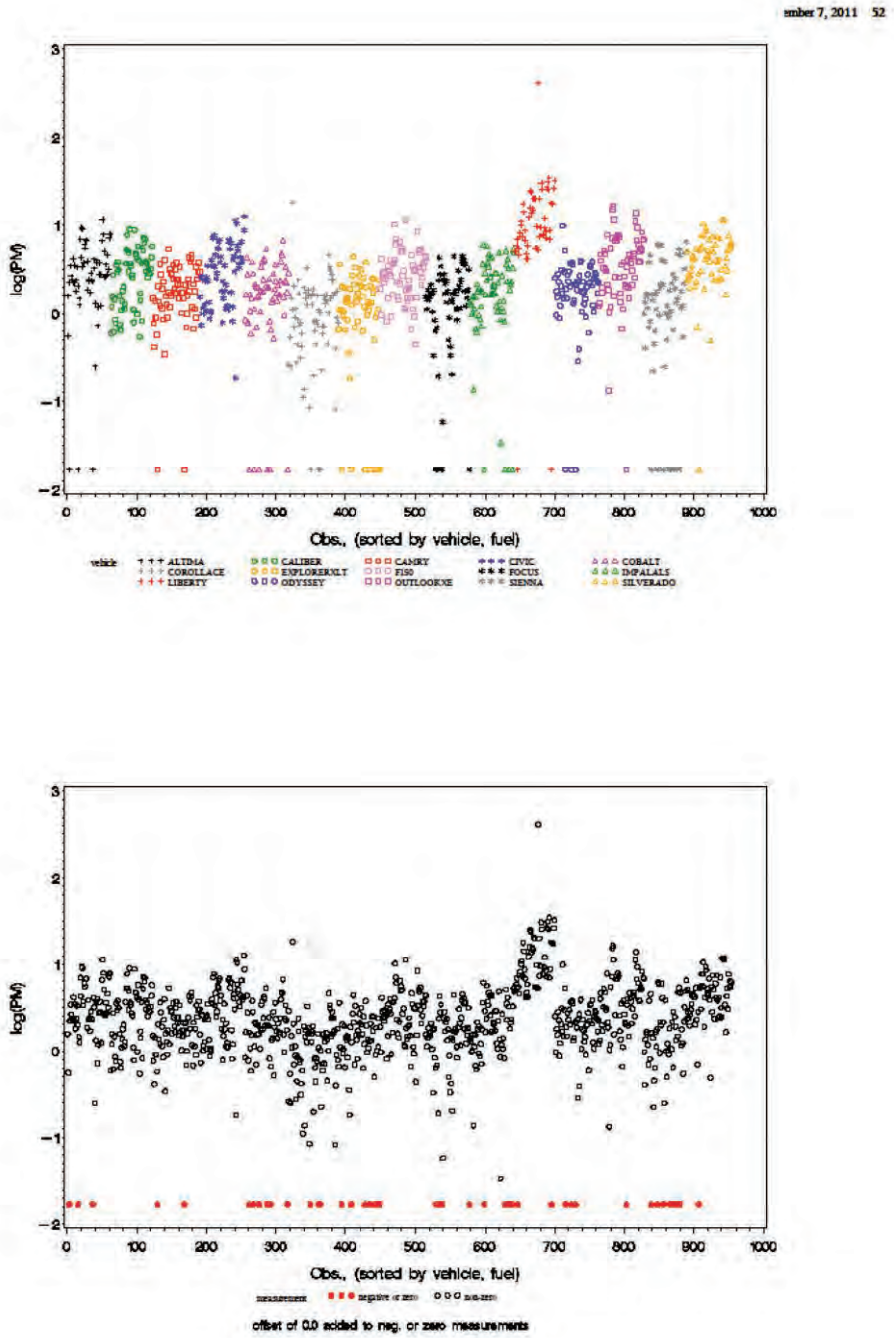


Figure 33. Linear-effects plot for Ethanol on Bag-1 PM: PM and ln(PM) vs. Ethanol Level, with data averaged by four Ethanol levels and by Vehicle.

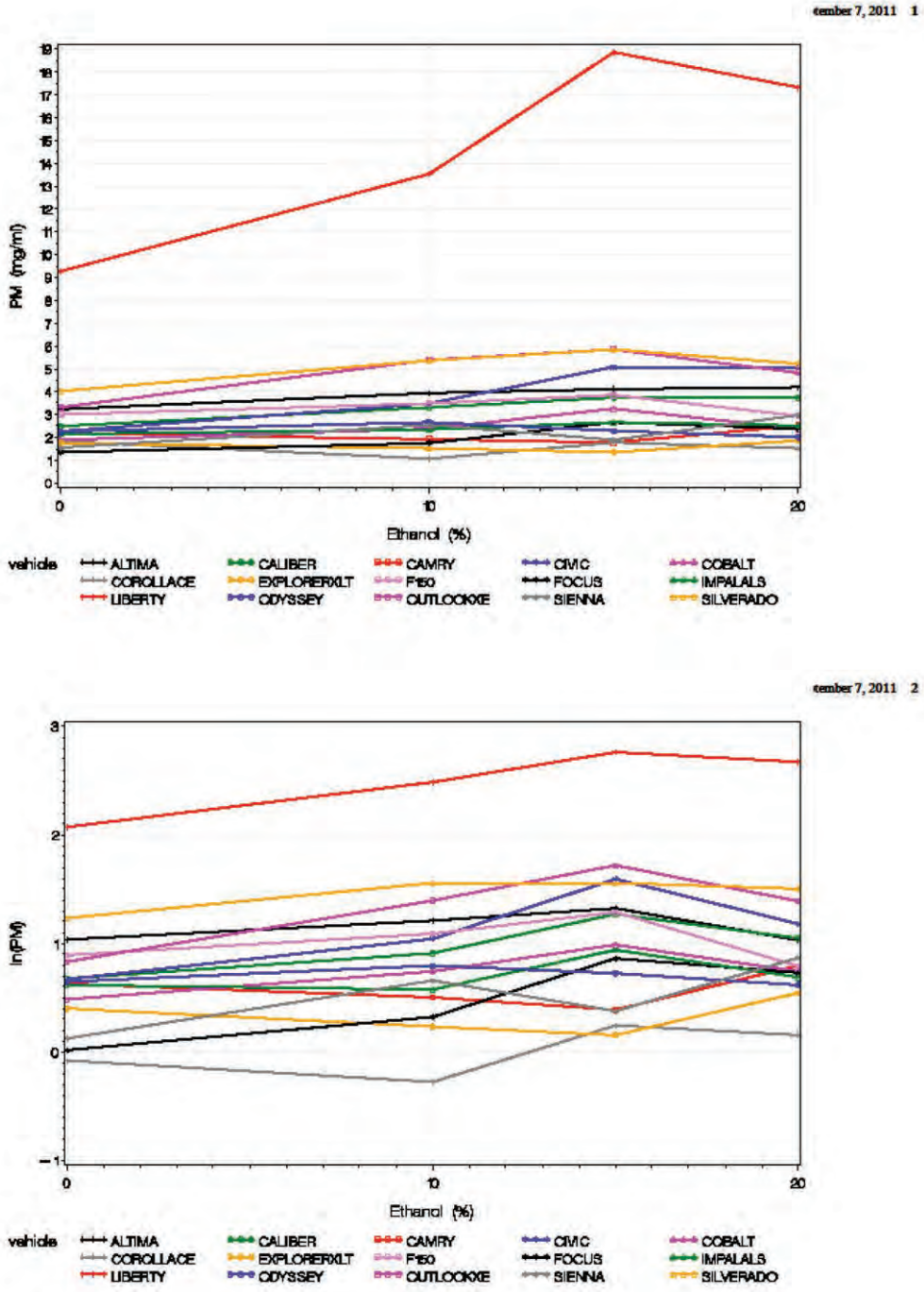
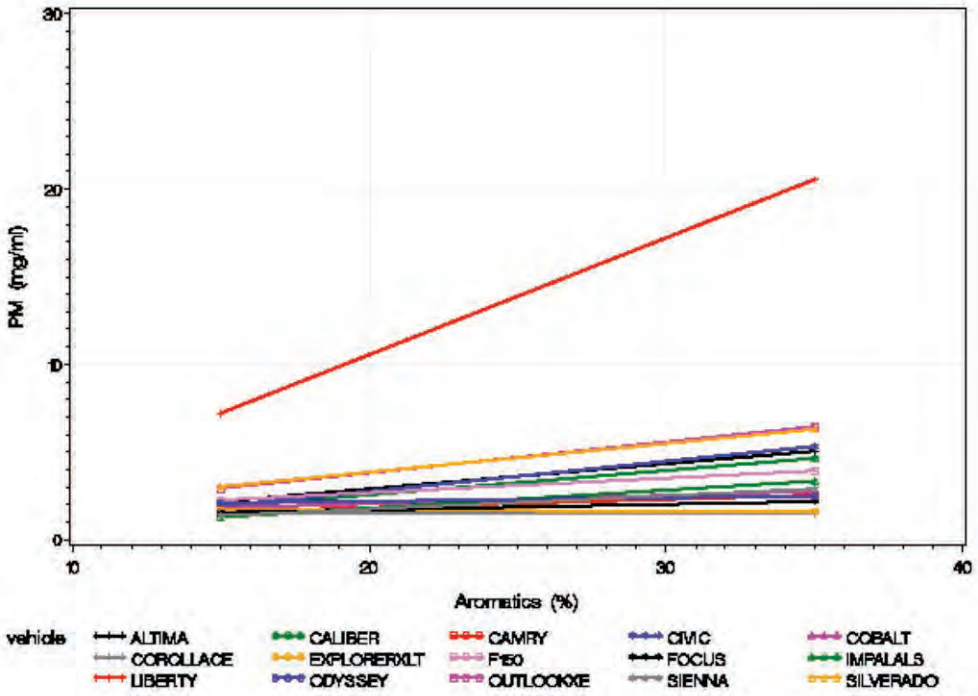


Figure 34. Linear-effects plot for Aromatics on Bag-1 PM: PM and ln(PM) vs. Aromatics Level, with data averaged by two Aromatics levels and by Vehicle.



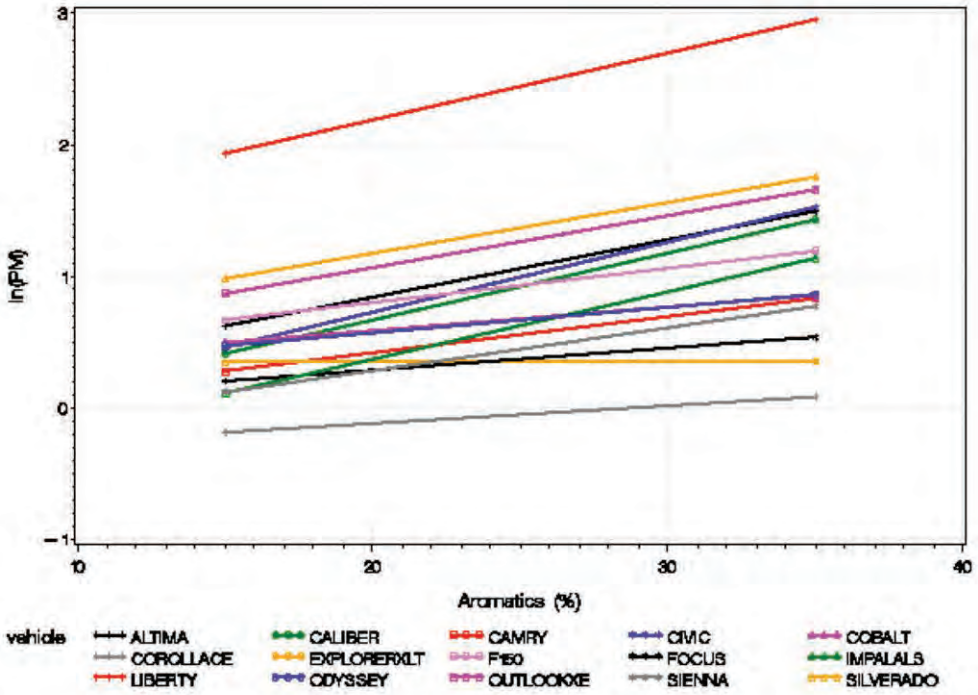


Figure 35. Linear-effects plot for RVP on Bag-1 PM: PM and ln(PM) vs.RVP Level, with data averaged by two RVP levels and by Vehicle.

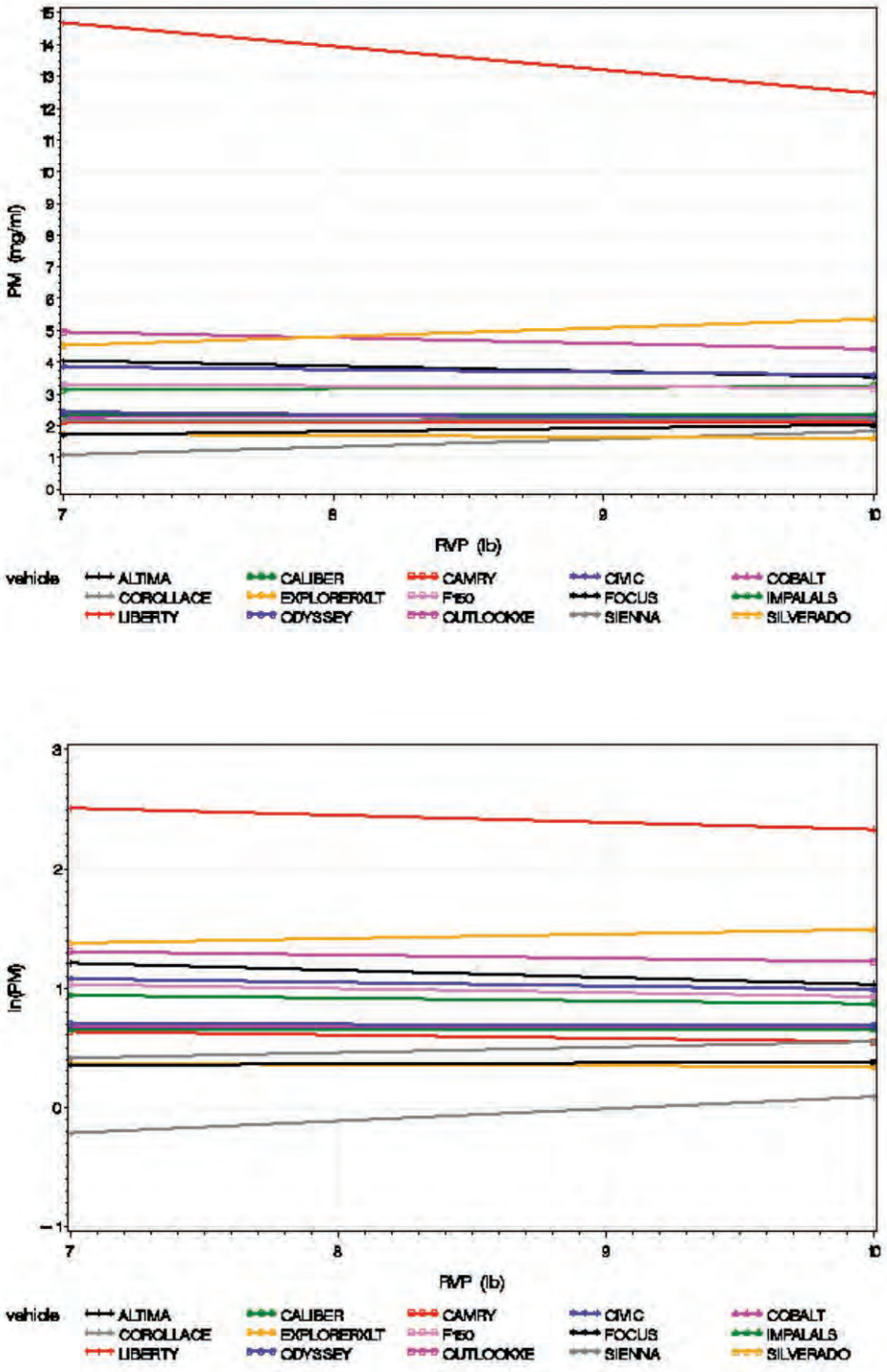


Figure 36. Linear-effects plot for T50 on Bag-1 PM: PM and ln(PM) vs.T50 Level, with data averaged by five T50 levels and by Vehicle.

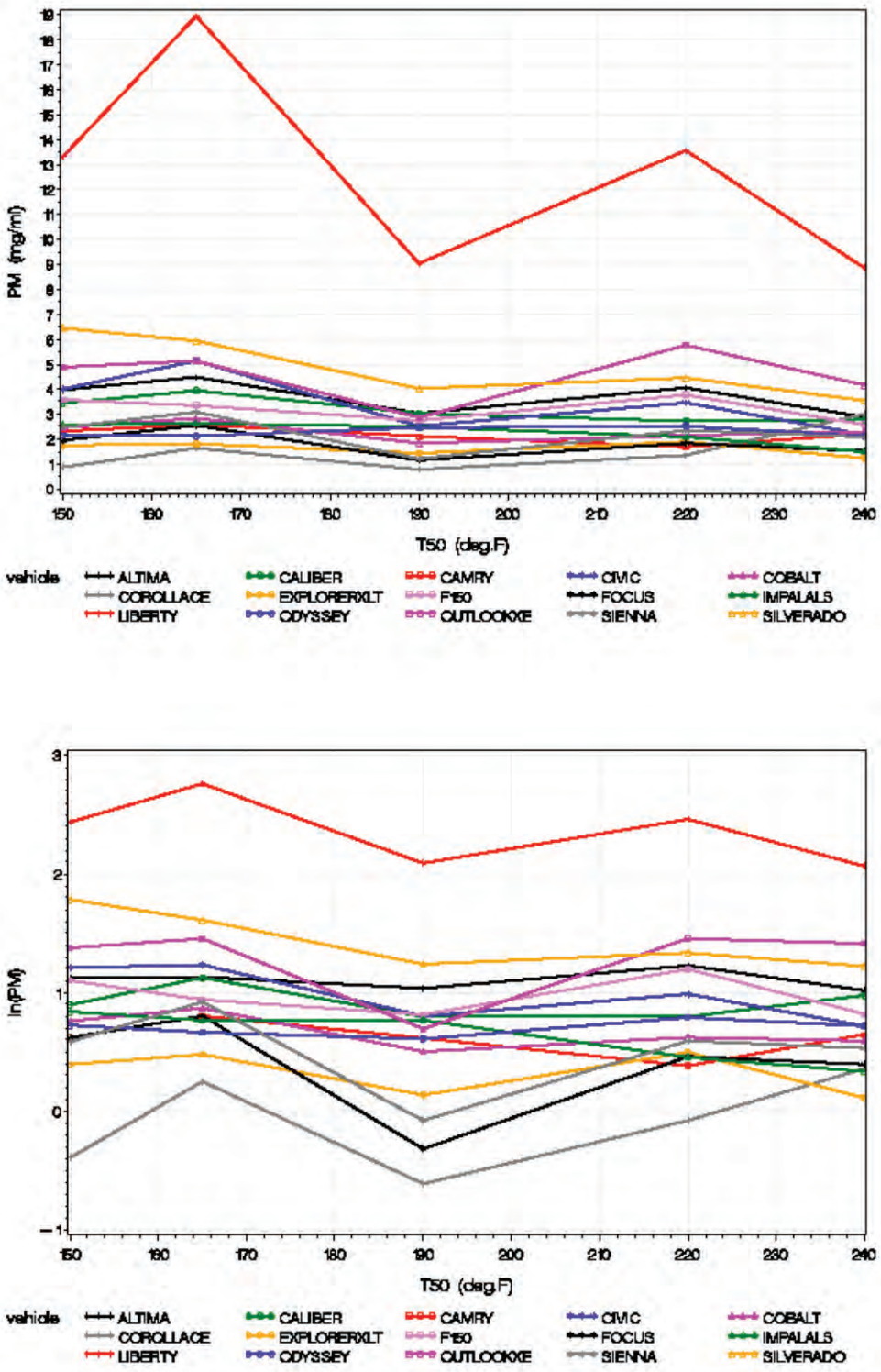
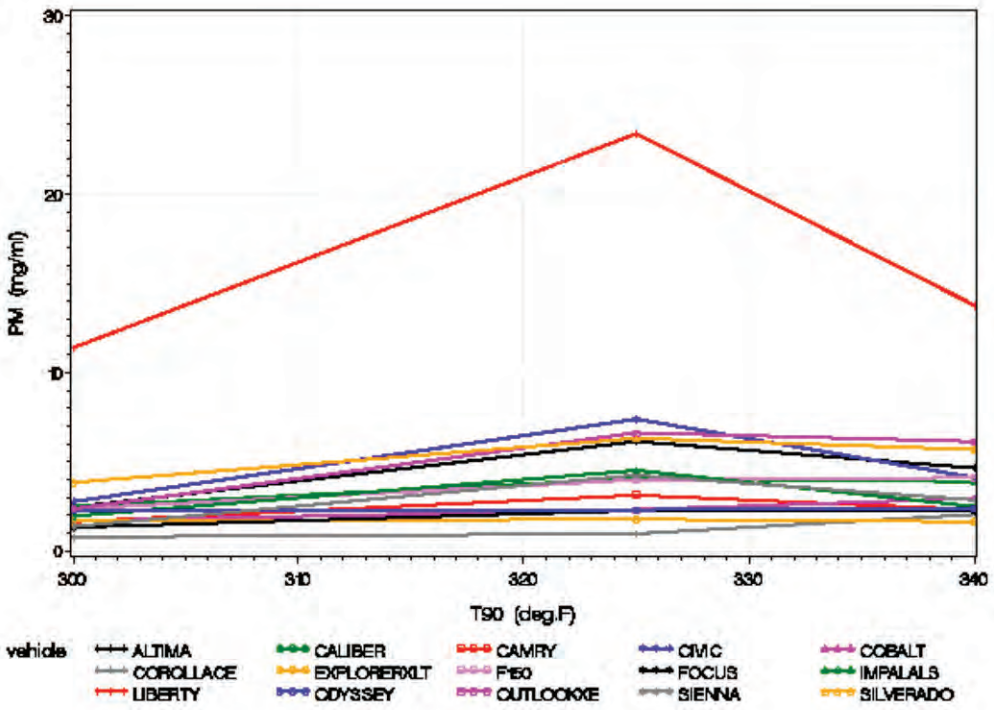


Figure 37. Linear-effects plot for T90 on Bag-1 PM: PM and ln(PM) vs.T90 Level, with data averaged by three T90 levels and by Vehicle.



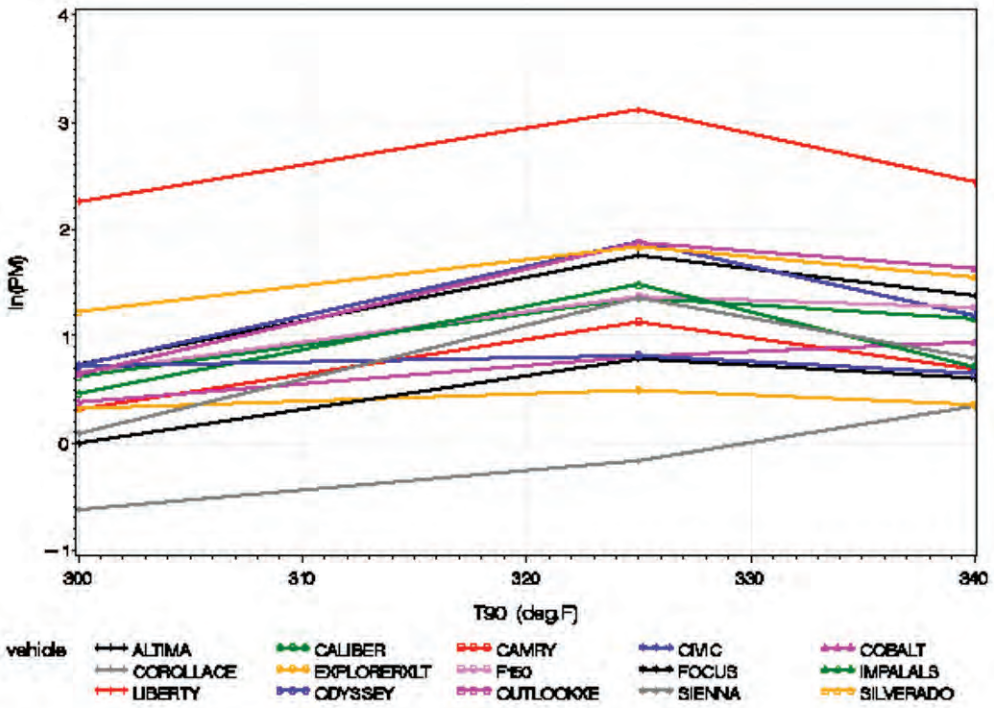


Figure 38. Bag-1 PM: Interaction plot for etOH×T50. Data are averaged by Ethanol and T50 levels, and plotted vs. Ethanol level, with a separate series for each T50 level. Note that not all points on the plot represent the same number of fuels.

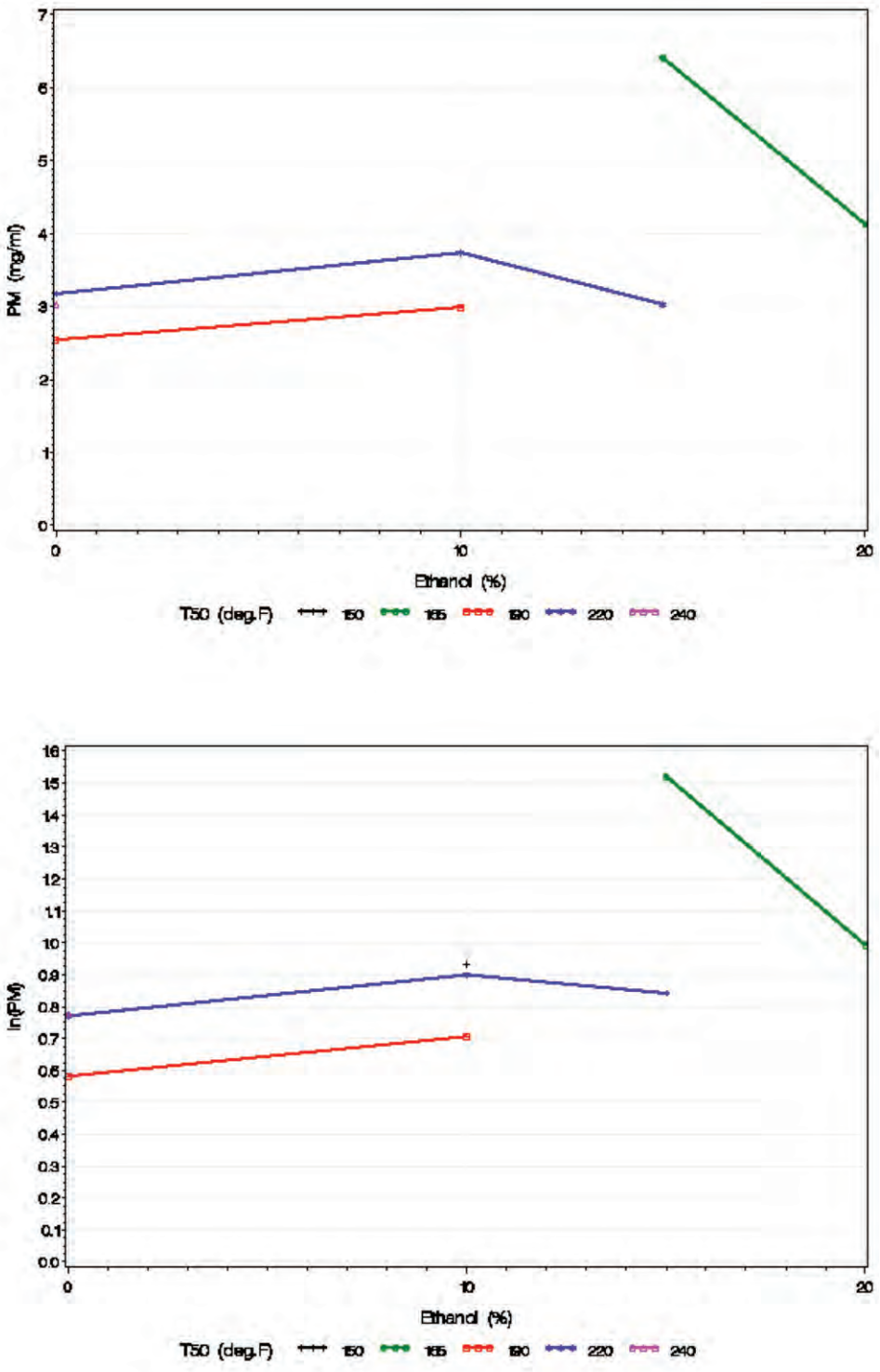


Figure 39. Bag-1 PM: Interaction plot for T50×Ethanol. Data are averaged by T50 and Ethanol levels, and plotted vs. T50 level, with a separate series for each Ethanol level. Note that not all points on the plot represent the same number of fuels.

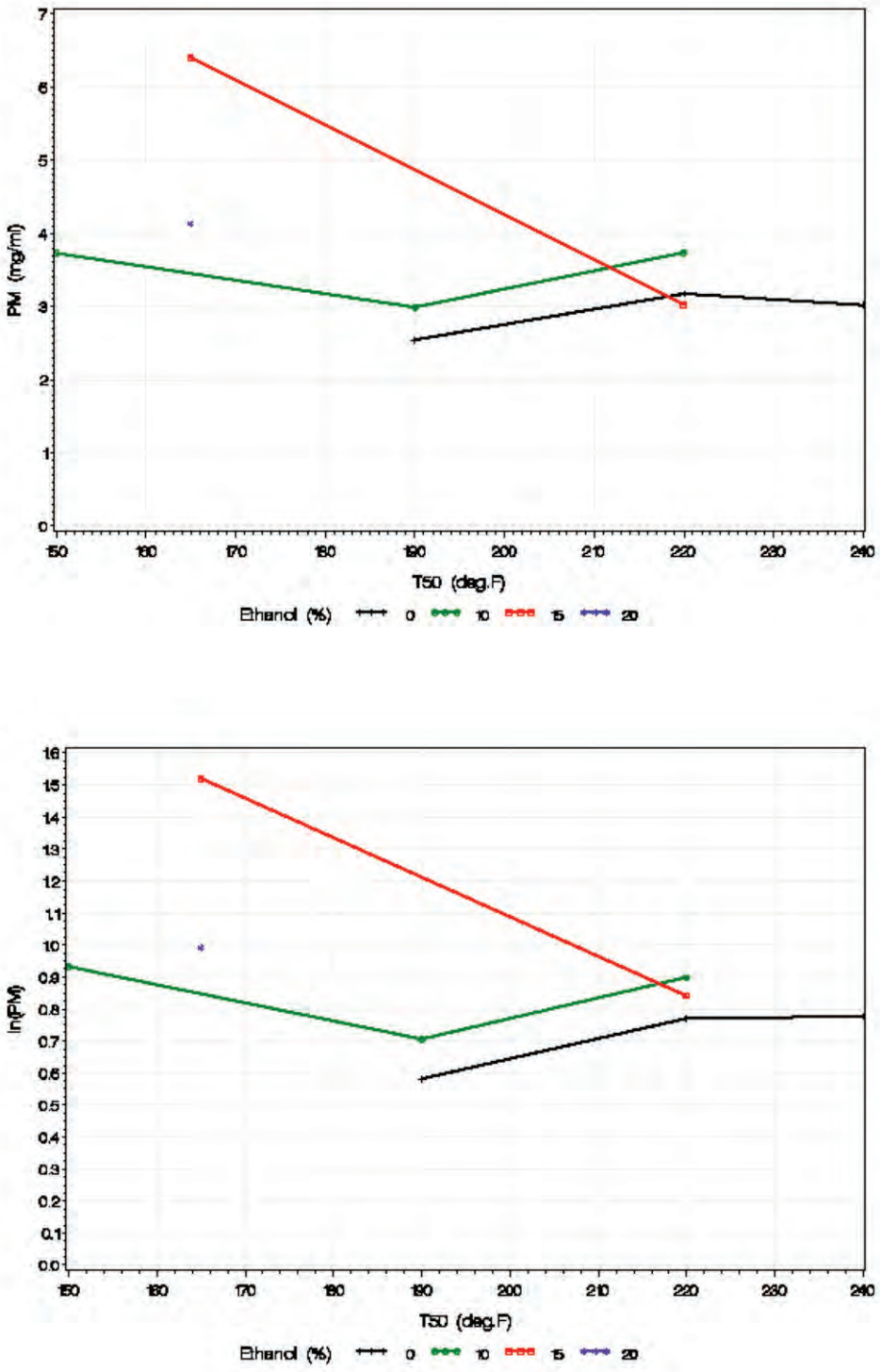


Figure 40. Bag-1 PM: Interaction plot for Ethanol × Aromatics. Data are averaged by Ethanol and Aromatics levels, and plotted vs. Ethanol level, with a separate series for each Aromatics level. Note that not all points on the plot represent the same number of fuels.

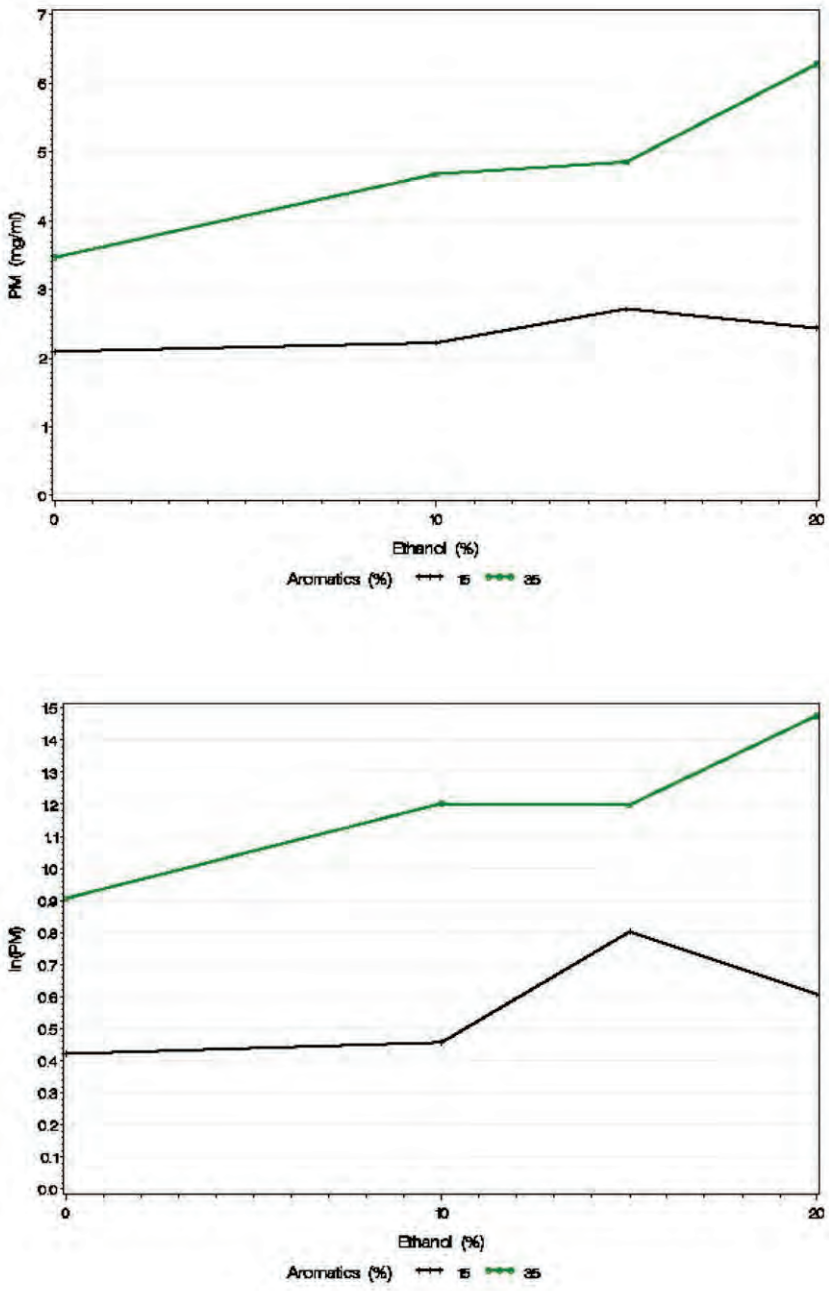


Figure 41. Bag-1 PM: Interaction plot for Aromatics × Ethanol. Data are averaged by Aromatics and Ethanol levels, and plotted vs. Aromatics level, with a separate series for each Ethanol level. Note that not all points on the plot represent the same number of fuels.

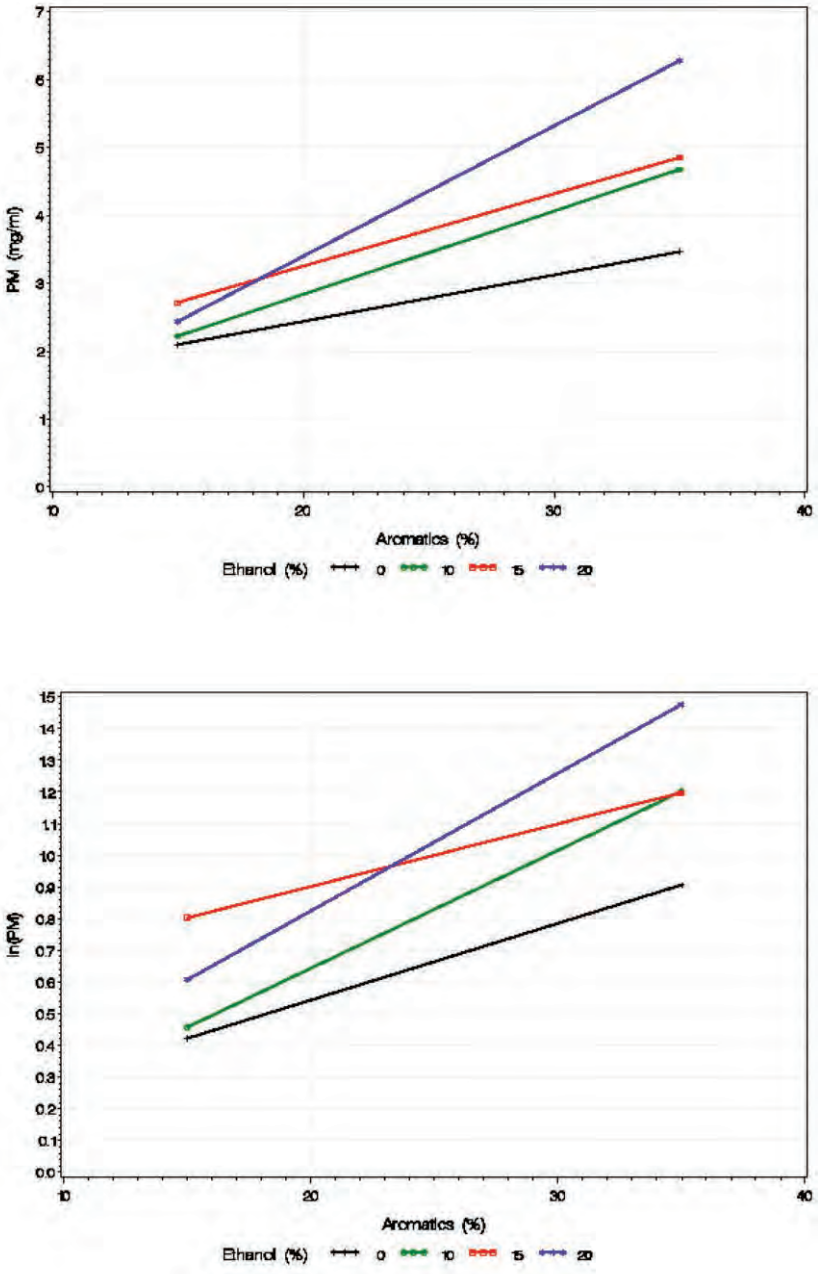


Figure 42. Bag-1 PM: Interaction plot for Aromatics × T90. Data are averaged by Aromatics and T90 levels, and plotted vs. Aromatics level, with a separate series for each T90 level. Note that not all points on the plot represent the same number of fuels.

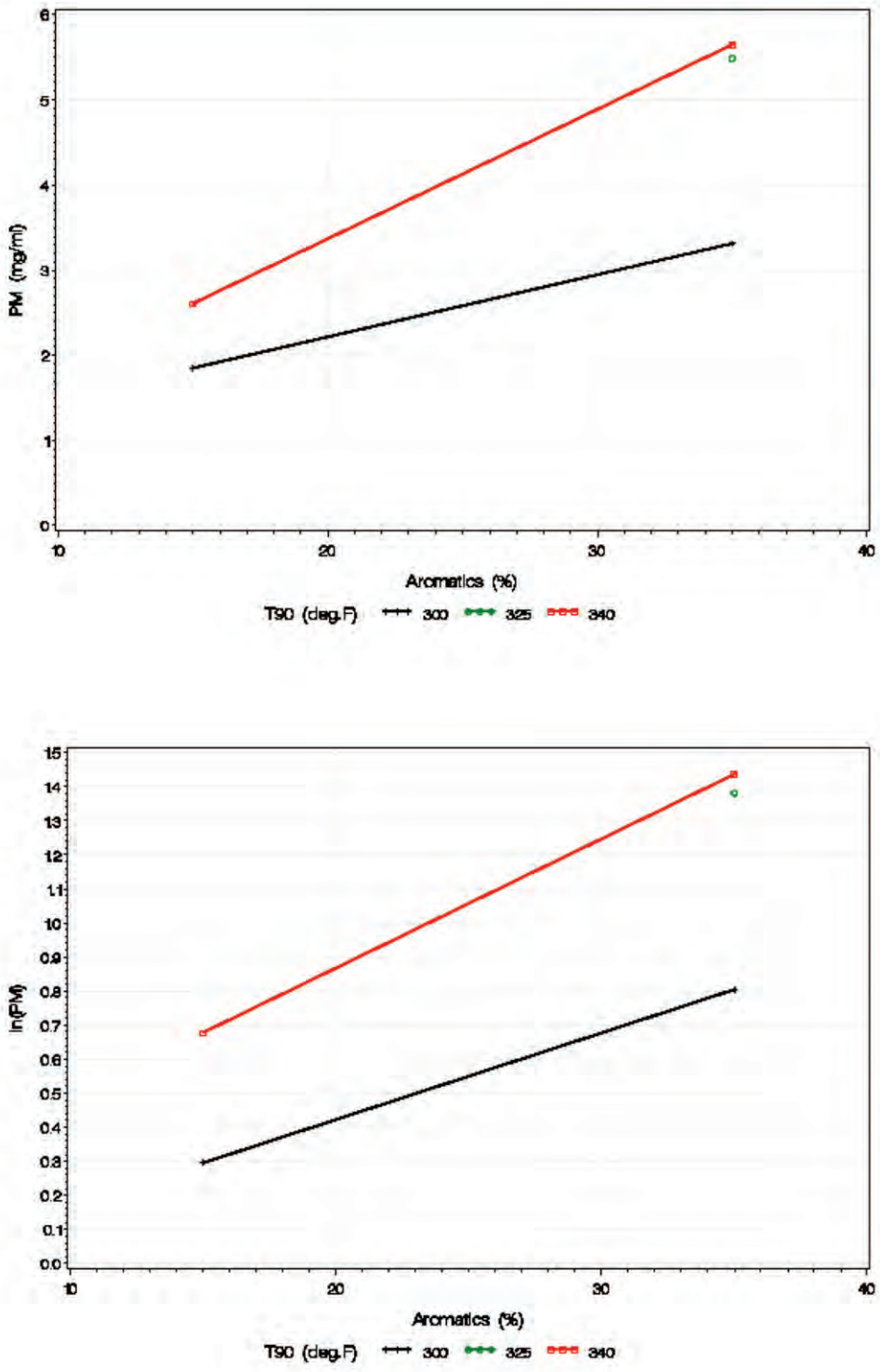
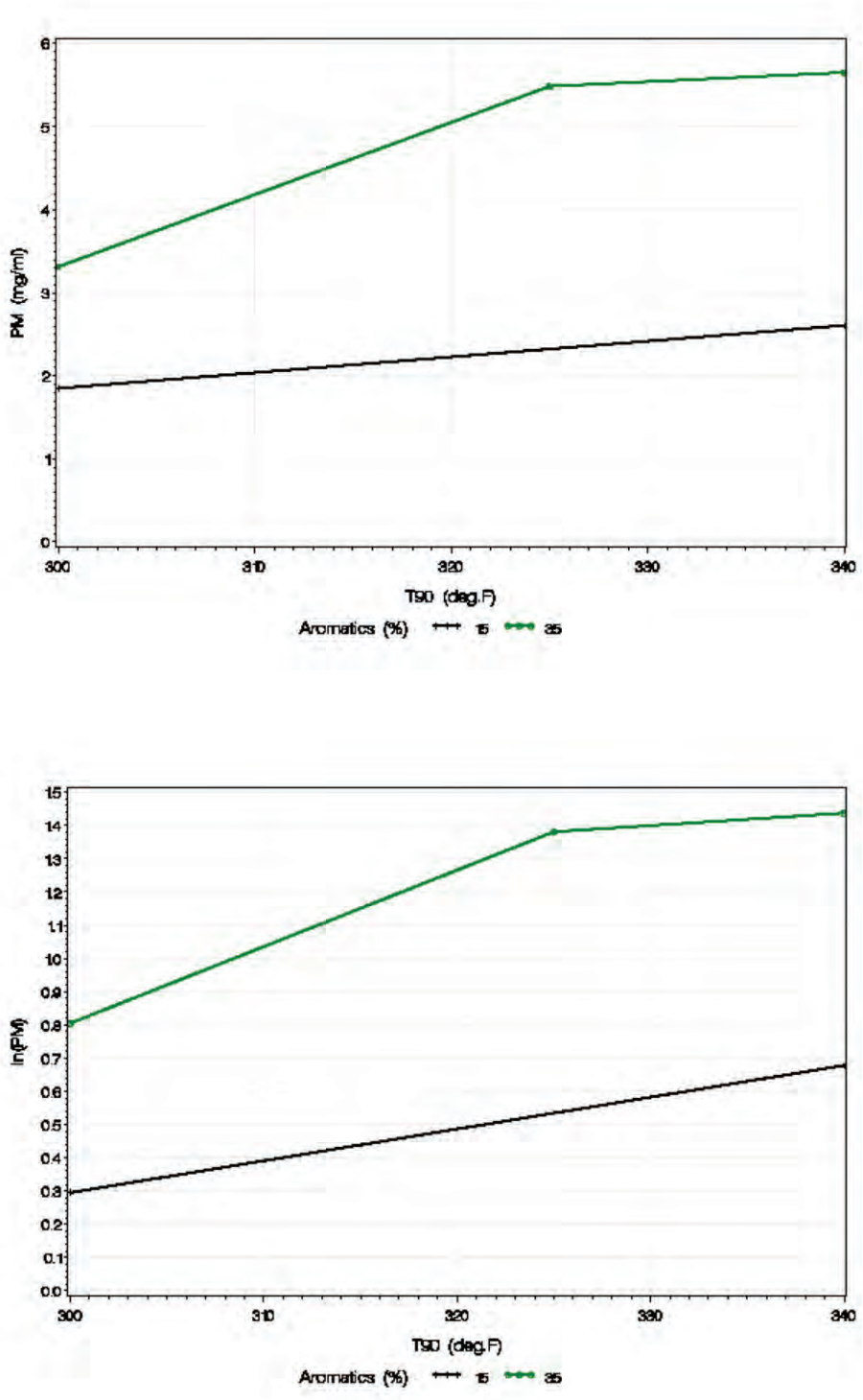


Figure 43. Bag-1 PM: Interaction plot for T90 × Aromatics. Data are averaged by T90 and Aromatics levels, and plotted vs. T90 level, with a separate series for each Aromatics level. Note that not all points on the plot represent the same number of fuels.



5 Preliminary Modeling

Following construction of the dataset, we conducted an initial round of modeling, primarily for purposes of outlier detection and influence analysis.

5.1 Assumptions

In the course of the analysis, we fit statistical models to emissions results for selected species, including THC, NMOG, NMHC, CH_4, CO, NO_x and PM. The emissions data represent cycle aggregates (i.e., Bag measurements) for the phases of the LA92 cycle, as described above. To allow for potential differences in behavior between start and running emissions processes, we analyzed the data separately by test phase (Bag).

Quantitation. For some emissions in some bags, subsets of observations fell below the limits of quantification (LOQ) for the measurement techniques used. In practical terms, this outcome means that the sample measurements, ostensibly from the tailpipe, were lower than background levels in the ambient air. When this result occurred, the measurements were set to zero. Note that for NMOG and NMHC, the situation is more complex in that these emissions represent quantities calculated from sets of speciated hydrocarbons, as described above in Section 2.2. In any case, values for either of these two emissions can also be considered missing if the calculation cannot be performed for any reason, such as the absence of a key constituent.

Censoring. For purposes of analysis, we treated these measurements as "censored." Specifically, we refer to affected datasets as "left-censored," because the lower, or "left tail" of the distribution was censored by limitations in our ability to quantify very small pollutant concentrations in the exhaust sample. We assume that a very small but positive measurement existed but was not captured and quantified. Assigning a value of zero to these observations is an example of a common approach to censoring of observations, known as "substitution." In this approach, a small but fixed quantity is substituted for the censored observations. Values used for substitution include zero, as mentioned, or small but positive quantities such as the smallest observation, a multiple of the smallest observation, the limit of quantitation (LOQ) or half the limit of quantitation (LOQ/2)[11]. At different stages of the analysis, we addressed censoring in different ways.

The degree of censoring varied widely by emission and bag, as shown in Table 16, although some patterns were observed, which can be related to the characteristics of the test cycle. The cold-start phase (Bag 1) of the LA92 is shorter than its counterpart in the FTP. Nonetheless, the presence of the cold-start increment provided generally larger measurable masses, hence the fraction of censored values is lowest in Bag 1. The hot-running phase, Bag 2 is longer than the preceding start bag, as well as the hot-running bag of the FTP. In addition, it is more aggressive, containing some transient operation with hard acceleration. These characteristics, despite the fact that the catalyst was lit off, provided sufficient measurable mass that the censoring rates in bag 2 were relatively low. In contrast, Bag 3 represents a hot start condition, involving a repeat of Bag

1 for a conditioned engine and control system. Thus, the final bag presented the greatest challenge to the measurement techniques employed, as shown through the highest censoring rates and the highest levels of model uncertainty and random error relative to the other two bags.

Table 16. Numbers of Censored Measurements, by Emission and Bag (n_{total} = 956 measurements).

Emission	Involves speciation	Test Phase		
		Bag 1	Bag 2	Bag 3
THC	NO	0	3	2
NMOG	YES	0	44	119
NMHC	YES	0	44	119
CH_4	NO	0	0	0
NO_x	NO	2	4	25
CO	NO	0	0	0
PM	NO	45	47	82

Transformation of Emissions Data. In all models, the response variable is always the natural log transformation of the emissions result. This step takes advantage of the tendency of regulated emissions to follow approximately log-normal distributions. In addition, this transformation is a standard approach to normalizing the distributions of residuals and stabilizing their variance across ranges of fuel properties[2]. Two additional justifications apply to use of the transformation. First, in interpretation of results, effects of fuel changes are expressed as ratios or percentages, calculations which are multiplicative on un-transformed data but very conveniently translated to additive operations on transformed data (i.e., differences of logarithms). Second, after reverse transformation, no fitted response can give a negative result, whether appropriately or inappropriately extrapolated[12].

Standardization. In all models, the independent or predictor variables are always standardized as described above (one- and two-stage). One-stage standardization was applied to linear or linear-effects terms (Equation 5), and two-stage standardization to quadratic or interaction terms (Equation 6, Equation 7).

Design Model. As described above, the study design was optimized with respect to a model including 11 terms (12 with the intercept), including five linear-effect terms, two quadratic terms, and four interaction terms between ethanol and the other four fuel properties (see Equation 2, page 21).

Extended Models. During modeling, six additional terms were included, to explore the possibility that additional effects could be estimated using this dataset. Throughout this document, models including all possible terms will be referred to as "full" models. Models including all 17 terms (excluding the intercept) will be referred to as "17-term extended models," and those with the 11 optimized terms will be referred to as the "11-term full model," or the "design model." All 17 terms and the notation used to identify them are shown in Table 17.

Table 17. Description and notation for parameters included in model fitting.

Fuel Parameter	Model term	In optimized design	Notation	Standardization[1]
Ethanol content (%)	etOH	YES	Z_e	One-stage
Aromatics content (%)	Arom	YES	Z_a	One-stage
RVP (psi)	RVP	YES	Z_r	One-stage
T50(°F)	T50	YES	Z_5	One-stage
T90 (°F)	T90	YES	Z_9	One-stage
	etOH × etOH	YES	ZZ_{ee}	Two-stage
	T50 × T50	YES	ZZ_{55}	Two-stage
	etOH × Arom	YES	ZZ_{ea}	Two-stage
	etOH × T50	YES	ZZ_{e5}	Two-stage
	etOH × T90	YES	ZZ_{e9}	Two-stage
	etOH × RVP	YES	ZZ_{er}	Two-stage
	Arom × T50	NO	ZZ_{a5}	Two-stage
	Arom × T90	NO	ZZ_{a9}	Two-stage
	T90 × T90	NO	ZZ_{99}	Two-stage
	T50 × T90	NO	ZZ_{59}	Two-stage
	Arom × RVP	NO	ZZ_{ar}	Two-stage
	RVP × T90	NO	ZZ_{r9}	Two-stage

[1] For one-stage standardization, see Equation 5, for two-stage standardization, see Equation 6 and Equation 7.

5.2 Identification of Influential Observations

We fit an initial model for each Bag and emission to allow identification of influential observations. For this purpose, we used 17-term extended full models. In this initial step, censored measurements were replaced with the minimum positive value measured for the emission and Bag.

All models were fit as mixed models. The standardized fuel properties were treated as continuous numeric variables and assigned as fixed factors; each vehicle was treated as a class variable and assigned as a random factor. We fit the models as "random coefficients models," in which the random effect is a random intercept fit for each vehicle. However, we did not attempt to fit random slope coefficients for individual vehicles.

Random slopes by vehicle would effectively comprise an interaction between vehicle and fuel. However, the study design and analysis approach do not allow for fitting such interactions. At the outset, it is clearly logical to fit random intercepts by vehicle, given that "vehicle" is the sampling unit for the study, for which reason we treat it as the blocking variable with respect to the fuel parameters, i.e., the "treatments" in the experiment. However, similar conditions with respect to "fuel" do not apply. As mentioned, we are not analyzing emissions in relation to the actual fuels used, but rather to fuel properties abstracted from the fuels and assumed to be

effectively independent (after design optimization and standardization). Thus, it is inappropriate to include the carefully designed and optimized fuel property matrix in the random component of the model because it in no way represents a random sample of available fuels, nor even a reasonable judgment sample spanning common combinations of fuel properties. Finally, the purpose of the study is to assess changes in mean emissions with respect to changes in fuel properties, which makes the fuel properties "fixed," by definition. Similarly, the various vehicles contribute much variability, which we isolate by treating vehicles as "blocks," which allows us to estimate the between-vehicle variability, without impairing our ability to estimate the effect of "treatment," i.e., the fuel parameter effects.

We fit the models using the MIXED procedure in SAS 9.2© using "maximum likelihood" as the solution method to allow comparison of fit among nested models with different numbers of parameters. No covariance structures were modeled using the REPEATED statement, reflecting an assumption that the variance of residuals is uniform throughout.

As a measure of influence, we calculated the externally studentized or "studentized-deleted" residual (r_{-i}). Observations with externally-studentized residuals greater than or equal to 3.5 in absolute value ($| r_{-i} | \geq 3.5$) were flagged for further evaluation. Numbers of influential measurements are summarized in Table 18, by emission and bag. Individual influential measurements are listed in Appendix G.

Table 18. Counts of Influential Measurements (out of a total of 956), by Emission and Bag (with "influential" defined as having a studentized-deleted residual ≥ 3.5 or ≤ -3.5).

Emission	Test Phase		
	Bag 1	Bag 2	Bag 3
CO	0	0	0
THC	0	0	0
NMOG	0	0	0
NMHC	0	0	0
CH_4	0	0	0
NO_x	3	7	0
PM	1	1	0

Measurements identified as influential were reviewed in detail on a case-by-case basis. EPA and its collaborators in the study agreed that measurements could be considered for removal from the dataset if (1) a physical reason were found that plausibly indicated that the data points could be invalid, or (2) the study participants reached a consensus that specific data points were sufficiently unusual or problematic to justify their removal. Based on these criteria, Two measurements were removed, both for particulate matter (PM). One measurement was removed

in Bag 1 and another in Bag 2. An additional measurement in Bag 3 was removed, even though it was not flagged as influential.

In Bag 1, the run selected for removal (run 6247) has a value of 413 mg/mi (corresponding to a studentized-deleted residual of 4.45) whereas all remaining measurements on the same vehicle (Liberty) have values less than 35 mg/mi. Further, the replicate measurement (run 6259) on the same fuel (16) has a value of 20.0 mg/mi, which is lower than that for run 6247 by a factor of 20.7. See Figure 44 below.

Similarly, in Bag 2, the run selected for removal (Run 5284) has a value of ~110 mg/mi (studentized deleted residual = 4.27) whereas all remaining measurements on the same vehicle (Explorer) have values of 0.50 or less. The selected run exceeds all other measurements on all other fuels by a factor of 220. See Figure 45.

Finally, in Bag 3, the run selected for removal (Run 6281) has a value of ~62 mg/mi, whereas the remaining measurements on the same vehicle (Explorer) have values of 5 mg/mi or less. See Figure 46. Despite its size, this point was not flagged as influential.

Figure 44. PM (Bag 1): Measurements for the Jeep Liberty, by fuel. The measurement identified as influential and selected for removal (Run 6247, on fuel 16) has an exceptionally high value (413 mg/mi).

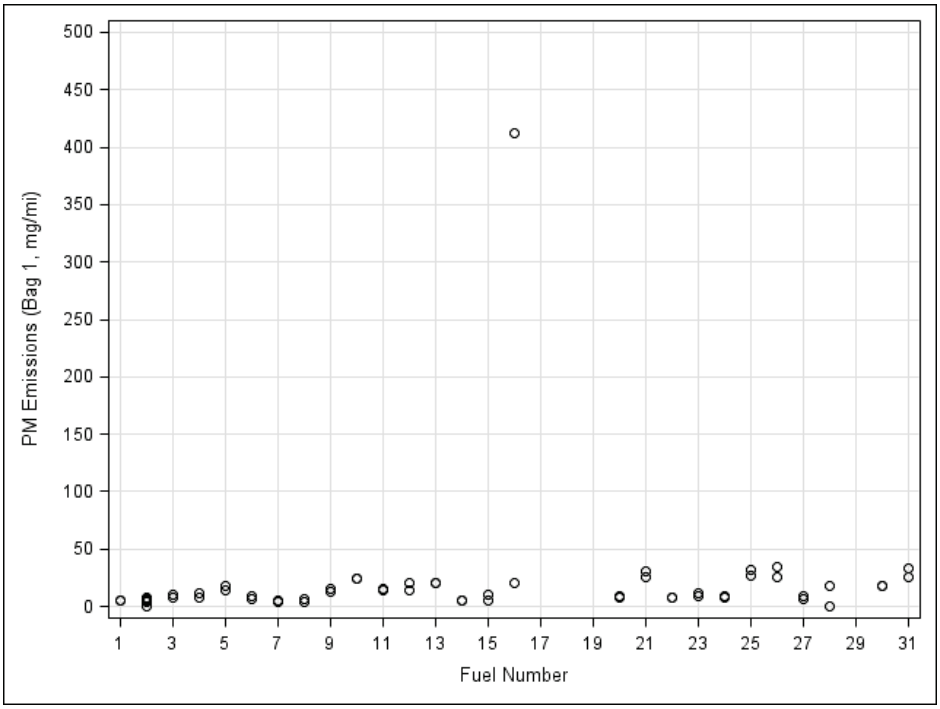


Figure 45. PM (Bag 2): Measurements for the Ford Explorer, by fuel. The measurement identified as influential and selected for removal (Run 5284, on fuel 27) has an exceptionally high value (~110 mg/mi).

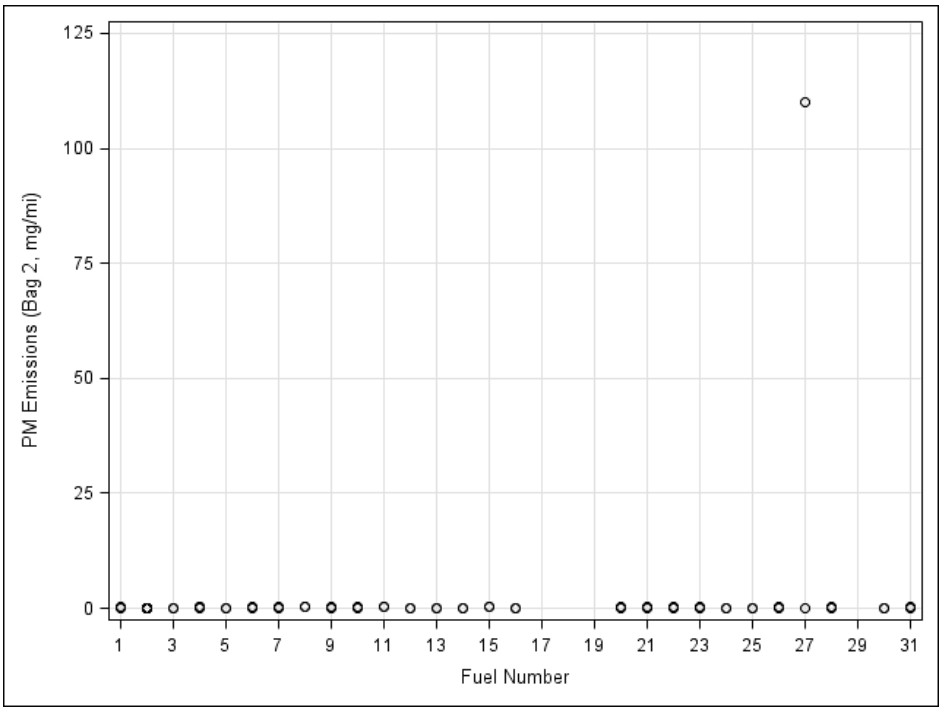
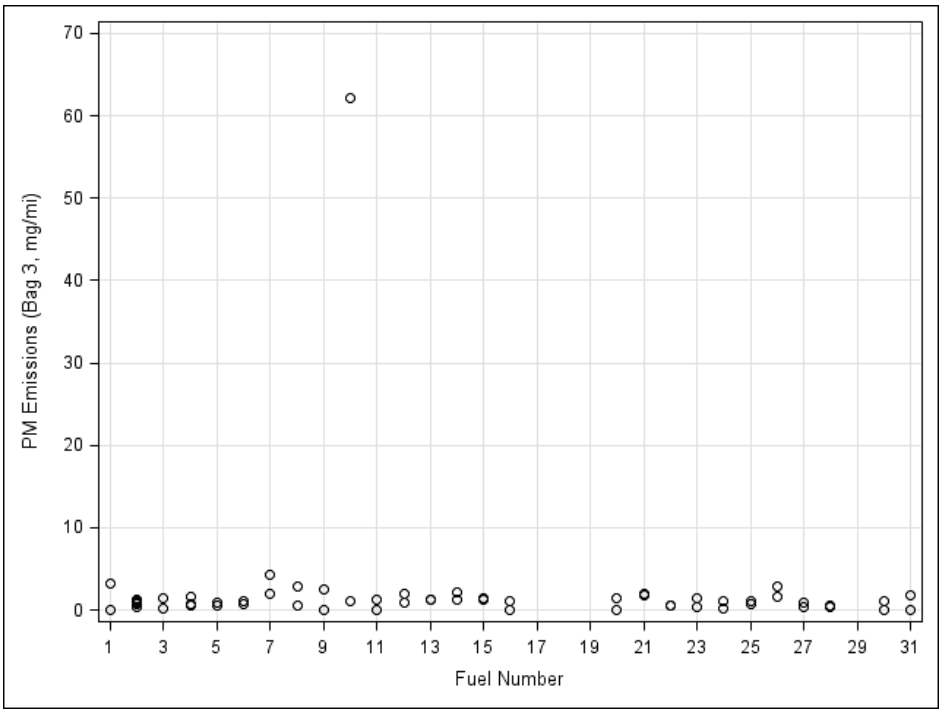


Figure 46. PM (Bag 3): Measurements for the Ford Explorer, by fuel. The measurement identified as influential and selected for removal (Run 6281, on fuel 10) has an exceptionally high value (~62 mg/mi).



5.3 Reduced Models

With the full models as a starting point, the next step was to begin the process of fitting "reduced" models, defined as models containing fewer terms than the full models. The reasons for this approach are several. A primary goal is to identify models including subsets of terms limited to those shown to be the most meaningful or useful in explaining and predicting the emissions measured. The full sets of terms in the optimized design include terms anticipated to be meaningful for any of the emissions to be measured. However, it was not anticipated that all the terms included would necessarily be meaningful for all emissions in all bags. A closely related goal is to develop models that would be, to the extent possible, explicable in terms of knowledge of the relevant physical and chemical processes. Parsimonious models are preferred over full models for this purpose, as their simpler structure makes their behavior easier to assess and explain. Finally, with respect to explicability, it is much preferred to minimize the potential for overfitting, which could reduce the generality of models selected for prediction. To guide the process, we adopted several assumptions, described below.

Hierarchy. At the outset, we imposed a requirement that the principle of "hierarchy" be maintained during the fitting process. Briefly stated, the principle requires that if any two-way interaction term A×B is retained in a model, both linear terms A and B must also be included, whether or not the linear terms appear highly significant taken alone. The maintenance of hierarchy serves to ensure that the reduced model(s) obtained are interpretable. The retention of an interaction and its linear terms is interpreted to mean that linear relationships exist between the response variable and both linear terms (A and B), and that the interaction (A×B) describes how these relationships vary with differing levels of the two factors involved. For example, if the slope for A×B is positive and significant, it suggests that the slope for A is steeper at a high level of B than at a low level of B, and vice versa, or that increasing the level of B *reinforces* the effect of A, and vice versa. Conversely, a negative and significant slope for A×B suggests that the slope for A is steeper at a low level of B than at a high level, suggesting that increasing the level of B may dampen or *interfere with* the effect of A. In either case, the interaction acts as a modifier or refinement to the underlying linear relationships, but for this model structure to remain intelligible, both linear effects must be retained, if the interaction is retained.

Removal of Outliers. Before the outset of model fitting, influential observations identified as problematic outliers were removed. As described above, this step affected only PM models (Bags 1-3).

Left-censoring. After deciding that censoring was too important an issue to neglect, the study participants adopted a consensus on how to address it in model fitting. For minimal levels of censoring, defined as five or fewer censored measurements ($n_{\text{censored}} \leq 5$), we elected to substitute the minimum positive measured value for the missing measurements. After substitution we fit mixed models as described above.

For higher levels of censoring, defined as more than five censored measurements ($n_{censored} > 5$) in a given Bag for a given emission, we elected to not rely on substitution, for two main reasons. The first is that substitution is known to introduce bias in the estimation of summary statistics for censored distributions[13]. Secondly, given that the purpose of this analysis is not to estimate means and variances for single censored distributions, but rather to develop multidimensional models to represent the response of emissions to simultaneous changes in multiple fuel parameters, substitution of uniform values for multiple missing measurements could introduce bias by obviating potential relationships between emissions and fuel properties. In cases of severe censoring, we adopted another approach commonly applied to left-censored datasets, known as "censored normal regression" or "Tobit regression." We describe this approach in greater detail below.

Removal of T90×T90 term. For development of reduced models, we removed the interaction from the full model. Our rationale for this step is that the two fuels containing T90 at the middle level (325) lie along one edge of the fuel parameter space, and thus lack sufficient depth across the modeling space to allow for adequate estimation of this term. Thus, in development of reduced models, we will refer to the "*16-term extended model*," as opposed to the "*17-term extended model*" as used in 5.1 and 5.2.

5.3.1 Minimal Censoring (Mixed Models)

For cases involving minimal censoring, we fit mixed models as described above.

We employed two model fitting approaches (1) Backwards elimination, and (2) fitting all possible models. In both approaches, the goodness of fit of various models was assessed using the Bayesian Information Criterion (BIC), as calculated by the MIXED procedure. In this formulation of the parameter, a reduction in the BIC indicates an improvement in fit. Thus, the goal in model fitting is to identify the model(s) giving the minimum value of BIC. The MIXED procedure was run using "maximum likelihood" (ML) as the solution method, to allow comparisons of fit between models with different sets of fixed effects.

Backwards Elimination (based on BIC). This approach was applied as follows. The full model was fit, including that maximum number of terms (p) and its BIC recorded (BIC{p}). The Type-III tests of effect (t-tests or χ^2-tests) for the full model were reviewed, and one or more parameters with insignificant tests at the 10% confidence level ($\alpha = 0.10$) were identified. The model was refit after dropping the k_1 selected parameters, and its BIC recorded. If the BIC for the reduced model (BIC{p-k_1}) was less than that for the full model, the reduced model was retained as the current "best-fit" model for the first step. After reviewing the tests of effect for the current reduced model, additional k_2 parameters could be selected for evaluation. The reduced model with p-k_1-k_2 parameters was fit, and retained as the "best-fit" if BIC{p-k_1-k_2} < BIC{p-k_1}. These steps were repeated until no additional terms could be removed without increasing the BIC. Often, but not always, the tests of effect for remaining terms in the last

model fit had p-values < 0.10. In some cases, however, linear terms with high p-values would be retained to maintain hierarchy. Note that throughout, no terms were dropped based solely on the tests of effect, but rather only on the goodness-of-fit. The α-level was set somewhat higher than the commonly used 5% level, to reduce the potential for type-II error in dropping terms, i.e., erroneously dropping terms that help explain the results or improve model fit.

Fitting all possible Models. This approach involves fitting all possible combinations of the p parameters in the full model. A total of 16 parameters gives a total of $2^{16} = 65{,}536$ possible models. With 11 parameters the corresponding total would be $2^{11} = 2{,}048$ possible models. However, in both cases, very high fractions (~95%) of the totals do not respect hierarchy, and thus are not eligible for further consideration.

In this process, the models were again run as mixed models, with algorithms written specifically to submit every possible candidate to the MIXED procedure. At the completion of the process a selected number of models with the best fits (lowest BIC), after excluding models without hierarchy, were listed and ranked. The model with the lowest BIC typically matched the best fit model from the backwards elimination process.

5.3.2 Severe Censoring (Tobit Regression)

For compounds and bags with high levels of censoring, we fit "censored normal regression," or "Tobit" models, a technique commonly used for left-censored data[14,15]. We fit the models using the LIFEREG procedure in SAS 9.2, as applied for cases of left censoring.

As with the mixed models, the procedure solves for the model parameters using maximum likelihood estimation. However, the Tobit approach does not attempt to estimate the missing values. Rather, the formulation of the maximum likelihood function (L) is modified so as to compensate for the absence of the censored values and to estimate values for the model coefficients accordingly[13]. In the Tobit model, each measurement is represented by its probability density (standard normal), given an assumed set of parameters, and each censored value is represented by the cumulative probability that the value would be less than the effective censoring level.

As the LIFEREG procedure is not able to handle random factors, it was necessary to enter each vehicle as a fixed factor, represented as an indicator or "dummy" variable. Thus, the model outputs an intercept for each vehicle, and an estimate of random error variance. It does not estimate a grand intercept for all vehicles, nor a component of variance representing the between vehicle-variability (the variance of the random intercepts). These steps were performed manually.

The procedure outputs the log likelihood ($\ln L$) as a goodness-of-fit parameter. It does not output an estimate of the BIC, but the BIC is readily calculated from L, the number of model terms p and the total number of (non-missing) observations n, as

$$\mathrm{BIC} = -2\ln L + p\ln n \qquad \textbf{Equation 12}$$

In model-fitting, we started with the full model, and proceeded by backwards elimination. In each step, one or more parameters were removed, and the model was refit. As with the mixed models, models with one or more terms removed are referred to as "reduced models." Models were selected for removal based on the *p*-value for their respective χ^2-test of effect ($p > 0.10$), starting with parameters with the highest *p*-values.

At each step, each successive reduced model was tested for goodness-of-fit against the preceding reduced model. At each step, if the current reduced model was not a significantly poorer fit than either the full model or its predecessor, it was accepted as the current "best fit." To interpret the goodness-of-fit test, the current reduced model was considered a poorer fit than the full or its predecessor if the *p*-value for the likelihood ratio test was < 0.10. The process was repeated until the current reduced model was a significantly poorer fit than its predecessor.

In performing the likelihood ratio tests, it was necessary that the two models included in the test be "nested," i.e., that both models have all terms in common except the subset of terms whose inclusion is the subject of the test. This condition always applied, in that all reduced models were nested within the full model, and each reduced model was nested within the preceding reduced model.

For a specific test, the model with more parameters is designated as the "reference" model, and the model with fewer parameters as the "nested" model. The test was fit in standard fashion, using the log-likelihood statistics output as the primary fit statistics for the models (all models were fit by maximum likelihood estimation) as shown in Equation 13. The test statistic is calculated as the difference in the -2log-likelihood between the nested and reference models, and which is assumed to be distributed as a χ^2 statistic with *d* degrees of freedom, where *d* is the difference in the numbers of parameters between the two models ($p_{\mathrm{ref}} - p_{\mathrm{nested}}$).

$$\chi^2_{test} = -2\ln\left(\frac{L_{\mathrm{nested}}}{L_{\mathrm{reference}}}\right) = -2\ln L_{\mathrm{nested}} - \left(-2\ln L_{\mathrm{reference}}\right) \quad \sim \chi^2_d \qquad \textbf{Equation 13}$$

The test is considered significant if the *p*-value was less than 0.10. The process is repeated until a significant result is obtained or if the remaining fixed effects were significant (p-value < 0.10). The set of terms remaining following the final step was retained as the "best fit" reduced model.

5.4 Initial Modeling: Summary

All models are based on the 16-parameter full model, as shown in Equation 14. All terms shown in Table 17 above were included at the outset of model fitting, with the exception of T90×T90. Results for preliminary models are summarized and presented in Appendix E.

Using the notation for standardized predictors, also shown in Table 17, the model is also expressed in Equation 15.

$$\begin{aligned}\ln Y = \beta_0 + \\ &\beta_1 \cdot \text{etOH} + \beta_2 \cdot \text{Arom} + \beta_3 \cdot \text{RVP} + \beta_4 \cdot \text{T50} + \beta_5 \cdot \text{T90} + \\ &\beta_6 \cdot \text{T50}^2 + \beta_7 \cdot \text{etOH}^2 \\ &\beta_8 \cdot \text{etOH} \times \text{Arom} + \beta_9 \cdot \text{etOH} \times \text{RVP} + \beta_{10} \cdot \text{etOH} \times \text{T50} + \beta_{11} \cdot \text{etOH} \times \text{T90} + \\ &\beta_{12} \cdot \text{Arom} \times \text{T50} + \beta_{13} \cdot \text{Arom} \times \text{T90} + \beta_{14} \cdot \text{T50} \times \text{T90} + \beta_{15} \cdot \text{Arom} \times \text{RVP} + \beta_{16} \cdot \text{RVP} \times \text{T90} + \\ &\varepsilon\end{aligned} \qquad \textbf{Equation 14}$$

$$\begin{aligned}\ln Y = \beta_0 + \\ &\beta_1 Z_e + \beta_2 Z_a + \beta_3 Z_r + \beta_4 Z_5 + \beta_5 Z_9 + \\ &\beta_6 ZZ_{55} + \beta_7 ZZ_{ee} \\ &\beta_8 ZZ_{ea} + \beta_9 ZZ_{er} + \beta_{10} ZZ_{e5} + \beta_{11} ZZ_{e9} + \\ &\beta_{12} ZZ_{a5} + \beta_{13} ZZ_{a9} + \beta_{14} ZZ_{59} + \beta_{15} ZZ_{ar} + \beta_{16} ZZ_{r9} + \\ &\varepsilon\end{aligned} \qquad \textbf{Equation 15}$$

5.5 Initial Modeling: Influence Analysis

A parallel analysis of these data, performed under contract to the DOE, employed methods and approaches similar to those described so far in this chapter (chapter 5). This analysis identified influential observations using an approach very similar to that described above in 5.2. The approach to analysis of censored measurements, as described in 7.2.2.5, was also adopted based on guidance from the author of the DOE research[1]. Thus, the criteria for applying mixed models or Tobit regression is the same in both analyses. Reduced models were also identified in the DOE analysis, emphasizing the "fitting all models" approach using mixed models when censoring was minimal, and reporting lists of terms making up potential good candidate sets. When censoring was severe the DOE analysis also applied Tobit regression, with models fit by backwards elimination, as described above. On the whole, the "best-fitting" reduced models reported for the DOE analysis are similar to those reported above in 5.4. However, due to limitations in time and resources, the DOE research did not extend into the additional influence analyses described below.

[1] Professor Richard F. Gunst, Southern Methodist University, Department of Statistical Science.

Following the completion of the initial model fitting, the reduced models were used for purposes of influence diagnostics. In this round, however, rather than identifying individual influential observations, the goal was to assess the influence of vehicles, which served as the sampling units in the project. A broader aim was to assess the extent to which inclusion or exclusion of particular vehicles might affect the model fits.

To achieve this step, we ran additional models, using the sets of terms retained in the initial best fits. These models were run as mixed models in all cases, even when the initial reduced model was obtained using a Tobit model. This change in procedure was adopted as the LIFEREG procedure used to fit the Tobit models lacks the automated diagnostic features available in the MIXED procedure. Thus, in running these models, censored values were replaced with the minimum positive measured value in each bag for each emission.

The influence of each vehicle was assessed using several diagnostic parameters. These included the "restricted likelihood distance" (RLD), the "multivariate DFFITS" (MDFFITS), the "covariance ratio" (CovRatio), the "PRESS statistic" (PRESS) and the "multivariate DFFITS for the covariance parameters" (MDFFITSCP). Each of these parameters assesses the influence of each vehicle with respect to the various aspects of the model-fitting process[16].

At the outset, the RLD is a measure of the influence of each vehicle on the overall fit of the model. It measures the change in the value of the log-likelihood for the entire dataset, using sets of parameters generated with and without the subset of data under consideration. The MDFFITS measures the influence of vehicles on the values of the coefficients of the fixed parameters, i.e., the parameter vector. It calculates the change in the set of coefficients, relative to their uncertainty. Similarly, the MDFFITSCP estimates the influence of vehicles on the values of the covariance parameters, or the variance of the random vehicle intercepts in this analysis. As opposed to the values of the coefficients themselves, the covRatio measures the influence of vehicles on the precision of the estimates, and is calculated as the ratio of the determinants of the covariance matrix of the parameters (covβ). The influence of vehicles on the models' predicted values is assessed using the PRESS statistic (Predicted Residual Sum of Squares). For each observation, the PRESS residual is calculated as the difference between each observation and the predicted marginal mean, estimated without the vehicle in question. The marginal mean is the prediction obtained using only the fixed parameters in the model (grand intercept and fuel-parameter effects), but excluding the random parameters (individual vehicle intercepts). For a vehicle, the PRESS statistic is the sum of squared PRESS residuals for all observations on that vehicle. For the RLD, MDFFITS and PRESS statistics, large values for a vehicle, in relation to those for other vehicles indicate high influence. For the covRatio statistic, a value of 1.0 indicates "no influence," and a value much larger or smaller than 1.0 indicates a high degree of influence.

Table 19 shows influence results for NO_x (Bag 1). The Ford Focus (FFOC) stands out as highly influential based on all five statistics. Its strong influence on the overall fit is indicated by its

high RLD, followed by the Odyssey. On the basis of MDFFITS however, the Odyssey is the most influential, with the Focus as a very close second, indicating that both these vehicles strongly influence the values of the coefficients. For the covRatio, a number of vehicles, including the Focus, have values greater or larger than 1.0, where a value of 1.0 is a benchmark of "low influence." The values of the covRatio, suggest that the vehicles strongly affect the precision of the fixed-effect parameters. For the PRESS statistic, the Focus has the highest value by a wide margin, but the second place is held by the Sienna rather than the Odyssey. For the MDFFITSCP statistic, the Focus is clearly most influential, but with the Odyssey much less so.

Review of the dataset clarifies these results to some degree (Figure 12, page 53). The emissions for the Focus are a full order of magnitude lower than those for all other vehicles, with the exception of the Sienna, which is also somewhat influential. However, neither the Odyssey nor the Sienna is nearly as influential as the Focus. Interestingly, the emissions for the Odyssey are among the highest for all the vehicles. However, the Odyssey stands out in that a subgroup of its measurements are markedly lower than the main group, substantially increasing the variability of its measurements overall.

Table 19. NO_x (Bag 1): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	1.309	0.0673	3.608	16.17	0.3446
Impala	64	1.366	0.0460	4.154	16.39	0.6556
Silverado	65	1.665	0.0504	4.605	6.85	0.8674
Caliber	61	1.621	0.0479	4.342	4.98	0.8584
F150	66	1.619	0.0936	3.535	57.69	0.2676
Explorer	61	1.784	0.0343	4.847	2.72	1.1777
Focus	63	59.671	0.5843	0.039	405.82	59.6382
Civic	65	1.105	0.0748	2.896	21.09	0.052
Odyssey	63	12.131	0.5982	0.876	55.51	3.7042
Liberty	58	1.823	0.0497	4.001	94.65	1.0302
Altima	62	1.658	0.0379	4.187	91.64	1.0217
Outlook	65	1.092	0.0476	3.784	8.15	0.3904
Camry	68	2.067	0.0216	5.737	31.17	1.5672
Corolla	71	2.3	0.1148	1.783	52.85	0.6622
Sienna	61	5.825	0.2170	0.797	292.27	2.831

Table 20 shows results for NO_x (Bag 2). In this case the Cobalt stands out as most influential by all measures. Identifying the second- and third-most influential vehicles is more ambiguous, but the Civic and Odyssey have RLD values slightly higher than the remaining vehicles and MDFFITS values considerably higher than the remaining vehicles. In bag 2, the set of measurements for the Cobalt is more variable than for any other vehicle, with measurements on

different fuels spanning roughly two orders of magnitude; no other vehicle is nearly as variable (Figure 22, page 65). Again, the Odyssey stands out somewhat in that it has a noticeable subset of measurements that stand out from the majority, although in contrast to Bag 1, the subset is higher rather than lower than the main group.

Table 20. NO_x (Bag 2): Selected Influence Statistics, by Vehicle.

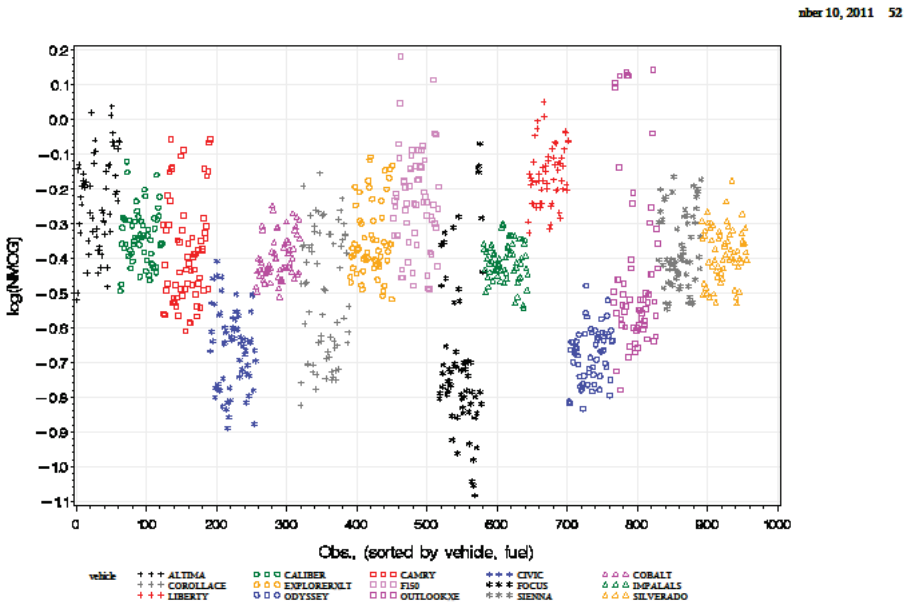
Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	60.037	1.8257	0.289	747.50	65.8804
Impala	64	0.788	0.0193	1.374	55.92	0.6593
Silverado	65	1.318	0.0393	1.424	41.46	1.0744
Caliber	61	0.391	0.0386	1.251	27.75	0.2519
F150	66	0.863	0.0742	1.317	97.78	0.5666
Explorer	61	0.134	0.0389	1.273	56.08	0.008
Focus	63	0.75	0.1631	1.184	169.42	0.2108
Civic	65	2.919	0.8214	1.252	31.15	0.2712
Odyssey	63	3.612	0.8051	1.181	48.70	0.9911
Liberty	58	0.674	0.0317	1.402	8.34	0.5237
Altima	62	0.281	0.0769	1.255	54.90	0.031
Outlook	65	0.406	0.0808	1.254	47.91	0.1363
Camry	68	2.14	0.0105	1.490	27.68	1.8705
Corolla	71	1.249	0.0294	1.470	21.97	1.0328
Sienna	61	1.698	0.0084	1.470	13.63	1.5037

Table 21 shows the results for NMOG (Bag 1). For this compound, two vehicles, the Focus and Outlook, are roughly tied as most influential by three of five criteria, with the Altima, Corolla, and F150 following in 3rd to 5th places, respectively. These two vehicles stand out in terms of their influence on the overall fit, the fixed-model coefficients and precision of the estimates. Examining individual measurements by vehicle, as shown in Figure 47, shows that these two vehicles are distinguished in the variability and distribution of their observations. The Focus shows order-of-magnitude variability, largely due the fact that its measurements form two distinct groups. One group, containing approximately 25% of its measurements is in the same range as most other vehicles. However, a second group sits lower than emissions for all but three of the 15 vehicles, with the lowest values in the entire dataset coming from the Focus. The Outlook, by contrast, is distinguished in that while most of its measurements are in the same range as most of the vehicles, it has a subset of measurements that are 0.3-0.4 orders of magnitude higher than the remainder, with this vehicle accounting for the highest measurements in the dataset, with the exception of the F150. It appears that these patterns may be related to the relatively high influence of these vehicles. As with NO_x, the covariance ratio shows that the two vehicles having the most influence on the values of the coefficients have the least influence on their precision.

Table 21. NMOG (Bag 1): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	1.938	0.0302	4.062	0.96	1.4196
Impala	64	2.605	0.1825	3.094	2.98	0.4117
Silverado	65	2.545	0.1836	3.058	3.53	0.3447
Caliber	61	1.645	0.0168	3.806	2.89	1.3058
F150	66	4.741	0.3845	1.860	18.53	0.1086
Explorer	61	0.881	0.0681	2.493	5.34	0.0735
Focus	63	14.897	0.5206	0.475	40.86	8.9811
Civic	65	3.822	0.3071	2.126	27.61	0.1529
Odyssey	63	2.354	0.0783	2.998	27.69	1.3193
Liberty	58	2.088	0.0787	2.819	23.51	1.063
Altima	62	4.575	0.3822	1.971	18.04	0.0126
Outlook	65	19.239	0.6169	0.485	15.17	12.4328
Camry	68	0.747	0.0574	2.203	6.22	0.0629
Corolla	71	5.984	0.2777	1.165	13.73	2.5616
Sienna	61	1.084	0.0269	3.324	2.68	0.7003

Figure 47. NMOG (Bag 1): Common logarithm of Measurements, by Vehicle and Fuel.



For Bag-2 NMOG, influence statistics are shown in Table 22. These results show the Sienna to be most influential by all measures except the PRESS statistic, with the Odyssey in second place for all measures except the PRESS statistic. The F150 has the highest PRESS statistic, which in

this case reflects that fact that its measurements are highest of all the vehicles, as shown in Figure 48. In contrast, the Sienna and Outlook are conspicuous in that they are most variable, have the lowest measurements and contribute all censored measurements for this dataset. Note also that the measurements for these vehicles are imputed from $NMHC_{FID}$, as previously described in 3.2.1 and shown in Figure 49.

Table 22. NMOG (Bag 2): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	0.545	0.0560	1.911	65.32	0.0063
Impala	64	1.223	0.0597	2.525	83.66	0.5836
Silverado	65	1.885	0.0591	2.772	137.40	1.181
Caliber	61	0.761	0.0035	2.641	6.43	0.6627
F150	66	2.499	0.0974	2.281	340.19	1.4651
Explorer	61	1.092	0.0678	2.512	16.64	0.4047
Focus	63	1.361	0.1360	2.094	24.11	0.0477
Civic	65	0.366	0.0293	1.770	62.13	0.0766
Odyssey	63	17.019	0.1346	0.371	330.09	16.5429
Liberty	58	0.797	0.0339	2.471	7.14	0.4293
Altima	62	1.813	0.0814	2.727	60.60	0.9221
Outlook	65	1.885	0.0428	2.927	102.54	1.3234
Camry	68	3.009	0.2151	1.380	90.73	0.891
Corolla	71	1.458	0.0272	3.219	18.02	1.0684
Sienna	61	33.597	0.8614	0.281	222.50	27.4382

Figure 48. NMOG (Bag 2): Common Logarithms of Measurements, by Vehicle and Fuel.

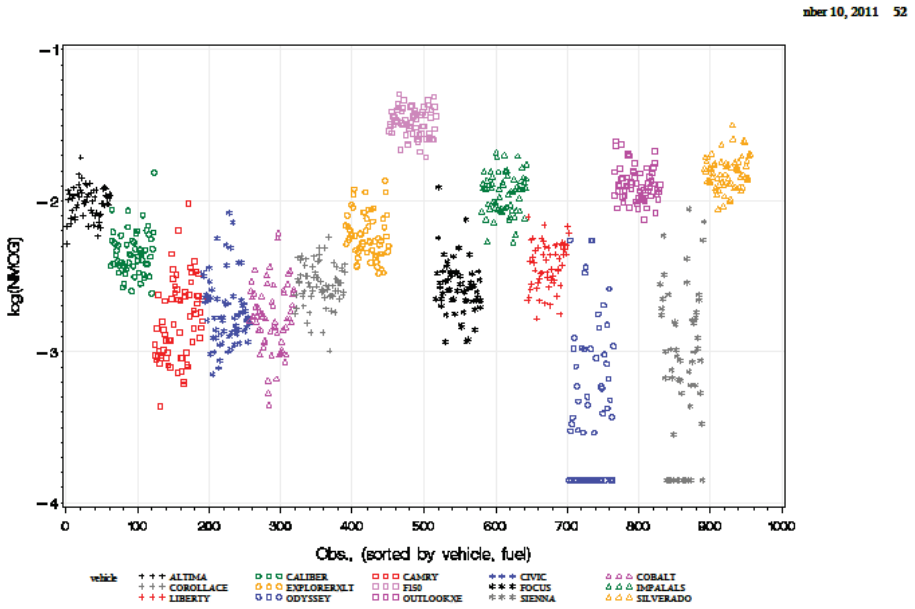
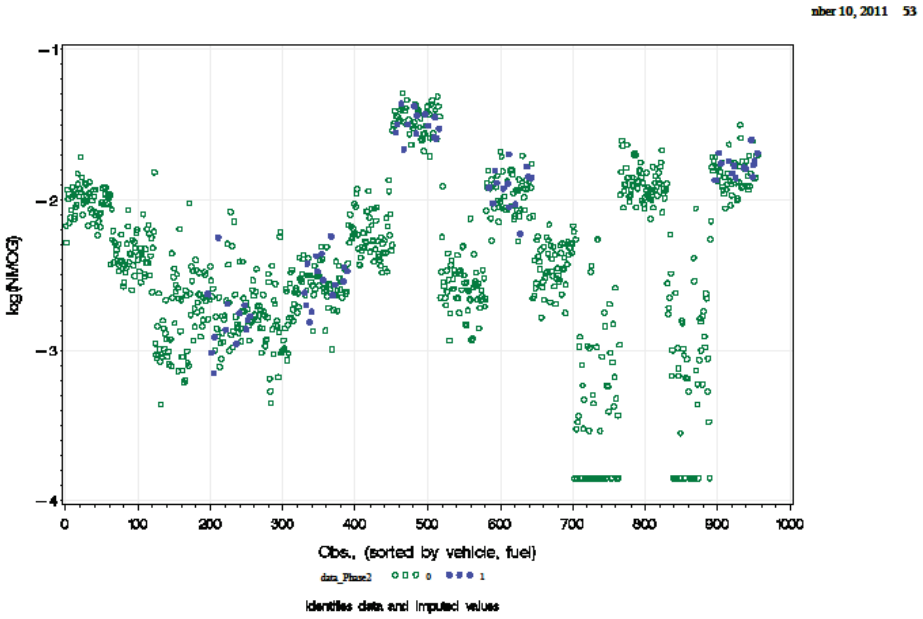


Figure 49. NMOG (Bag 2): Common logarithms of Observations by Vehicle and Fuel, delineating speciated measurements (blue) and values imputed from $NMHC_{FID}$ (green).



Influence statistics for Bag 1 PM are shown in Table 23. The vehicles with the greatest influence overall are the Liberty, Sienna, Explorer and Focus. However, in contrast to other emissions and Bags, the range in the various measures is not as wide. Thus no individual vehicle stands out dramatically with respect to the others. The Liberty has a markedly higher PRESS statistic, but this value merely indicates that its emissions are highest among a pool of vehicles with low between-vehicle variability compared to other emissions (Figure 32, page 77).

For PM (Bag 2), patterns in the influence measures are broadly similar to those in Bag 1 in that no vehicles stand out dramatically in respect to model fit, parameter values or their precision (See Table 24). As with the other emissions, the vehicle with the highest measurements has a conspicuously high PRESS statistic.

Table 23. PM (Bag 1): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	0.8448	0.0640	1.369	96.21	0.41179
Impala	64	1.1711	0.1023	1.348	102.55	0.48901
Silverado	65	0.819	0.0668	1.744	71.50	0.35215
Caliber	60	1.24	0.0159	1.995	21.41	1.03185
F150	66	1.4635	0.0190	2.128	24.63	1.20473
Explorer	60	2.6696	0.1348	1.069	152.49	1.76658
Focus	63	1.6378	0.1889	1.309	113.75	0.41219
Civic	65	1.421	0.0420	2.070	29.33	1.03774
Odyssey	63	1.1893	0.1588	1.474	82.89	0.15175
Liberty	57	3.6971	0.2224	0.816	231.28	2.9276
Altima	62	0.4994	0.0740	1.587	67.64	0.02499
Outlook	65	1.3185	0.1104	1.853	53.70	0.55396
Camry	68	0.7271	0.0324	1.952	43.38	0.46756
Corolla	71	1.5608	0.2300	1.465	135.61	0.0428
Sienna	61	3.1149	0.0869	1.032	153.88	2.51621

Table 24. PM (Bag 2): Selected Influence Measures, by Vehicle.

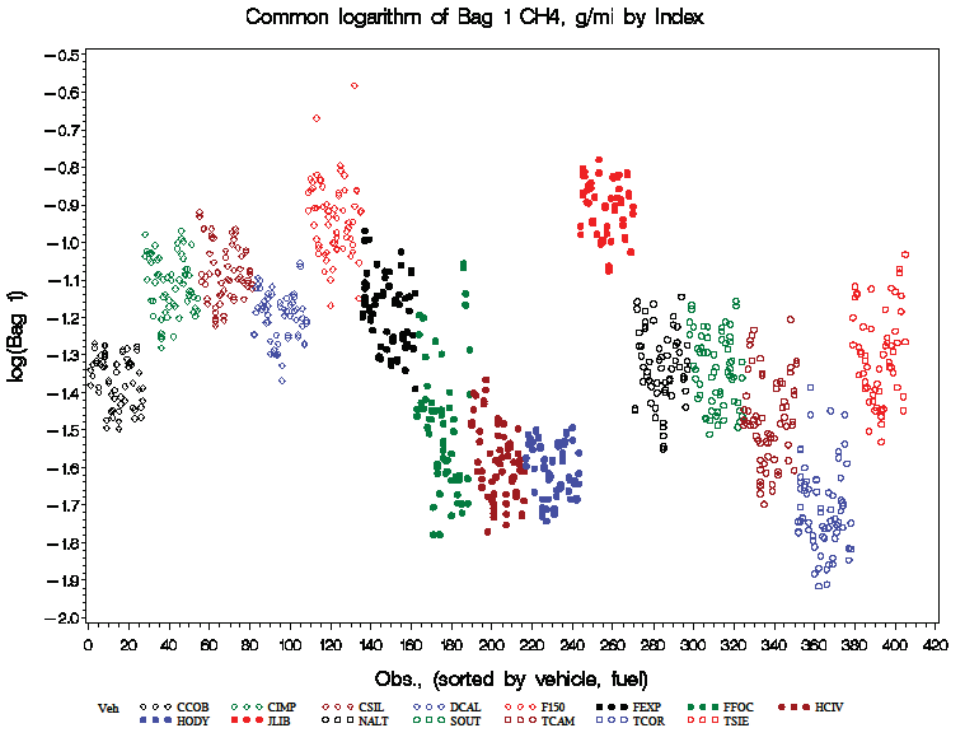
Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	0.3003	0.0262	1.280	126.44	0.17797
Impala	64	1.0491	0.1604	1.048	265.14	0.39073
Silverado	65	0.8582	0.0989	1.355	142.87	0.39695
Caliber	60	3.7201	0.5356	1.177	138.80	1.37541
F150	66	0.671	0.0516	1.545	53.98	0.41104
Explorer	60	1.4893	0.1711	1.126	172.71	0.7334
Focus	63	0.4502	0.0438	1.483	67.71	0.2415
Civic	65	0.7138	0.0158	1.578	40.17	0.58636
Odyssey	63	0.5833	0.0050	1.565	38.64	0.51223
Liberty	57	2.4291	0.3058	0.872	314.47	1.22254
Altima	62	0.7088	0.1595	1.373	87.19	0.02283
Outlook	65	0.353	0.0733	1.434	75.24	0.03507
Camry	68	1.1214	0.0241	1.632	38.98	0.91643
Corolla	71	1.3307	0.1663	1.179	201.48	0.57426
Sienna	61	1.6511	0.2862	1.236	132.14	0.39142

Influence diagnostics for CH_4 (Bag 1) are shown in Table 25. Three of the five measures, except the CovRatio and PRESS, show the Focus to be the most influential vehicle. Of all the vehicles, the Focus shows the strongest influence on the overall fit and the values of the coefficients. With respect to the precision of the coefficients, the Focus is in the four most influential vehicles, although its influence opposite in direction. The Corolla, F150 and Liberty have the largest values of the PRESS statistic. Review of the data points in Figure 50 shows that the Corolla has the lowest values in the dataset, and the other two vehicles the maximum values. Thus, these results for the PRESS are expected. The values for the Focus are at the lower end of the range, but this vehicle is on par with two others that are not nearly as influential (Civic and Odyssey). However, the Focus has a subset of points that sit considerably higher than the majority, which may account for its influence.

Table 25. CH_4 (Bag 1): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	2.032	0.037	6.464	1.816	1.286
Impala	64	1.941	0.091	4.422	16.408	0.431
Silverado	65	3.112	0.187	3.463	22.734	0.082
Caliber	61	1.692	0.027	5.942	4.436	1.130
F150	66	4.786	0.266	1.878	60.582	0.359
Explorer	61	2.329	0.130	2.125	7.301	0.238
Focus	63	28.436	0.640	0.197	19.503	19.340
Civic	65	3.488	0.215	3.009	31.095	0.017
Odyssey	63	2.112	0.054	5.446	33.478	1.124
Liberty	58	1.984	0.061	4.056	55.542	0.941
Altima	62	3.732	0.223	3.408	1.742	0.081
Outlook	65	1.859	0.098	4.302	1.751	0.260
Camry	68	1.982	0.081	1.845	13.167	0.649
Corolla	71	2.857	0.121	1.594	75.958	0.860
Sienna	61	3.440	0.192	2.059	2.474	0.302

Figure 50. CH_4 (Bag 1): Common logarithm of Measurements, by Vehicle and Fuel.

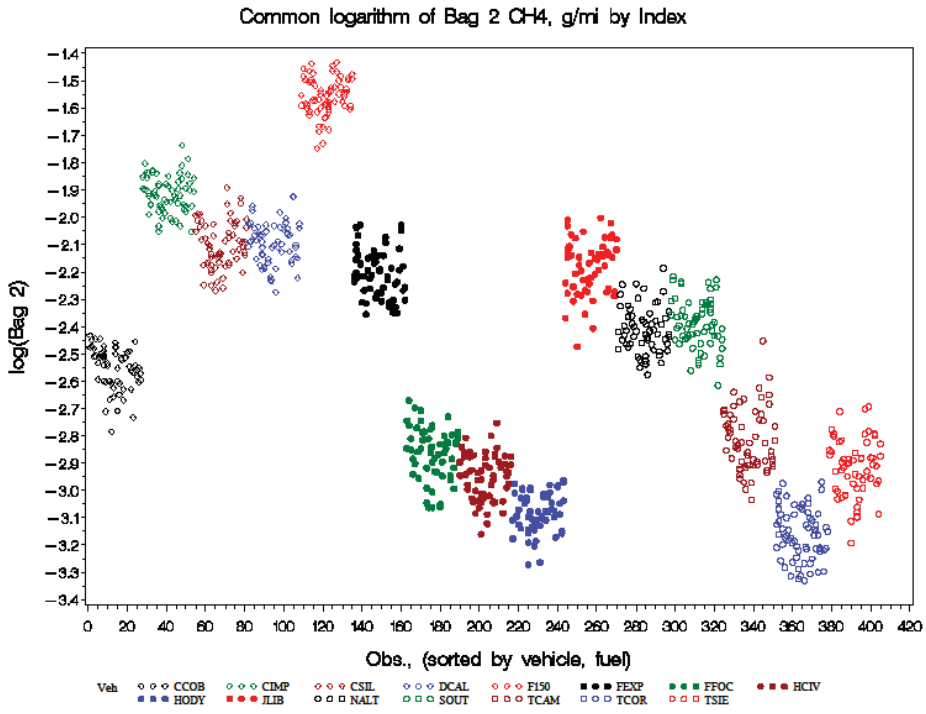


For CH_4 (Bag 2), influence measures are shown in Table 26. Again, the Focus is most influential by all measures except the PRESS and covRatio. However, the differences between the Focus and the other vehicles are not nearly as wide as in Bag 1. Viewing the Focus in the context of the other vehicles, as shown in Figure 51, shows the results for this vehicle to be unremarkable. Its measurements are on par with those for five other vehicles, and are not unusually variable. The reasons for this vehicle's influence are thus unclear based on an initial review. Again, the vehicles highest (Liberty) and lowest measurements (Odyssey, Corolla) have the largest PRESS values.

Table 26. CH_4 (Bag 2): Selected Influence Statistics, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	1.317	0.0896	2.6975	3.0700	0.3057
Impala	64	0.873	0.0613	2.3124	124.7140	0.1939
Silverado	65	1.280	0.0970	2.5374	59.9610	0.2222
Caliber	61	1.250	0.0545	2.7950	55.1470	0.6034
F150	66	2.687	0.1901	1.6989	349.0660	0.6997
Explorer	61	3.043	0.1984	1.3917	28.3030	0.8980
Focus	63	5.349	0.2505	1.0362	59.1910	2.6246
Civic	65	0.724	0.0514	2.3687	86.9690	0.1556
Odyssey	63	1.066	0.0528	2.4949	138.3610	0.4563
Liberty	58	1.851	0.1495	1.6257	32.1670	0.2484
Altima	62	0.822	0.0733	2.0519	3.6390	0.0319
Outlook	65	1.227	0.1035	2.4265	4.2500	0.1068
Camry	68	2.651	0.2067	1.6445	49.1980	0.4080
Corolla	71	1.313	0.1115	1.8295	204.3760	0.0995
Sienna	61	1.242	0.1037	1.6813	72.4840	0.1304

Figure 51. CH_4 (Bag 2): Common Logarithm of Measurements, by Vehicle and Fuel.

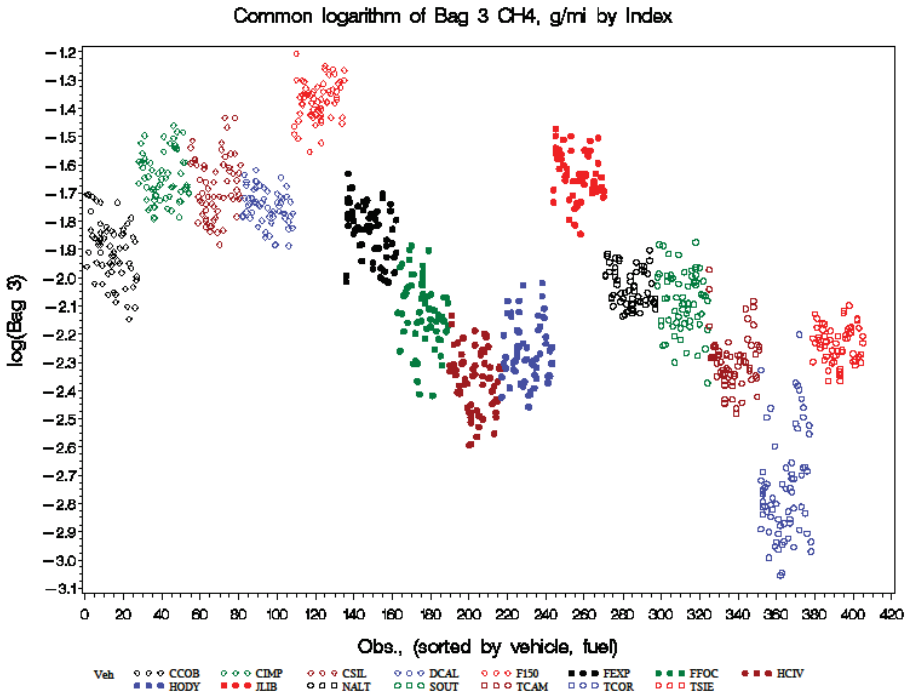


For CH_4 (Bag 3), influence measures are shown in Table 27. In this case, the Corolla is conspicuous with respect to all five measures. As the vehicle with the lowest measurements, as shown in Figure 52, it has a large PRESS, followed by the F150, which has the highest measurements.

Table 27. CH_4 (Bag 3): Selected Influence Measures, by Vehicle.

Vehicle	*n*	RLD	MDFFITS	CovRatio	PRESS	MDFFITSCP
Cobalt	63	0.637	0.0919	1.5736	6.0930	0.0385
Impala	64	0.786	0.0306	1.7951	49.8830	0.5357
Silverado	65	1.049	0.1372	1.6826	44.5850	0.1494
Caliber	61	1.175	0.0485	1.9223	23.3200	0.7871
F150	66	2.129	0.2044	1.4788	163.2620	0.7787
Explorer	61	1.449	0.2209	1.5706	12.0210	0.0314
Focus	63	3.922	0.3757	1.2113	10.7820	1.4645
Civic	65	1.906	0.2950	1.5508	55.2680	0.0139
Odyssey	63	4.417	0.6303	1.3534	30.8480	0.3214
Liberty	58	1.055	0.0921	1.7339	46.5800	0.4262
Altima	62	1.378	0.0174	2.0765	1.0980	1.1424
Outlook	65	0.794	0.1099	1.5330	6.5010	0.0809
Camry	68	1.103	0.1504	1.7131	38.4500	0.1199
Corolla	71	36.655	1.0974	0.3455	257.1930	32.1543
Sienna	61	1.526	0.0399	2.0142	21.4020	1.1480

Figure 52. CH_4 (Bag 3): Common Logarithms of Measurements, by Vehicle and Fuel.



As an outcome of the influence analyses, we performed additional detailed review of the measurements for the vehicles identified as highly influential, summarized in Table 28. The review and its results are discussed below in Section 6.

Table 28. Vehicles Selected for detailed review for subsets of NO_x, NMOG and CH_4 results.

Vehicle	NO_x		NMOG		CH_4		
	Bag 1	Bag 2	Bag 1	Bag 2	Bag 1	Bag 2	Bag 3
Cobalt		●					
Impala							
Silverado							
Caliber							
F150							
Explorer							
Focus	●						
Civic							
Odyssey				●			
Liberty							
Altima							
Outlook							
Camry							
Corolla							●
Sienna				●			

6 Measurement Issues

6.1 Data Quality at Very Low Emission Levels

Looking at the results of the influence analysis alongside plots of the measured values revealed that the most influential vehicles tended to have a large number of very low measurements. These observations led to a closer examination of measurement error in the dataset.

All measurement processes have error associated with them as a result of the physics of mixing and sampling from a gas stream or noise in electronic components such as optoelectronic detectors or signal amplifiers. This means that repeated measurements taken under identical process conditions would produce a range of results, their average being the true response of the instrument and the range around it representing the measurement variability. The dilute bag method requires measurement of concentrations in both the sample and background (ambient or dilution air) bags, followed by a calculation that subtracts the background from the sample after a dilution factor correction. Thus, the net result contains a linear combination (sum) of variability from two measurements. For the analyzers used in this test program, this variability is generally of a fixed size in terms of concentration, resulting in a relative error that increases as the concentration being measured decreases. This fact is important to consider when using measurements to produce relative difference models.

Results such as NMOG and PM mass involve more complex measurement processes, making sources of error harder to quantify. NMOG is a calculated result, comprised of measurements of many different aggregated and individual species made using different types of instruments. PM mass is measured by technicians who must carefully handle and weigh filters, presenting opportunities for other types of errors and biases.

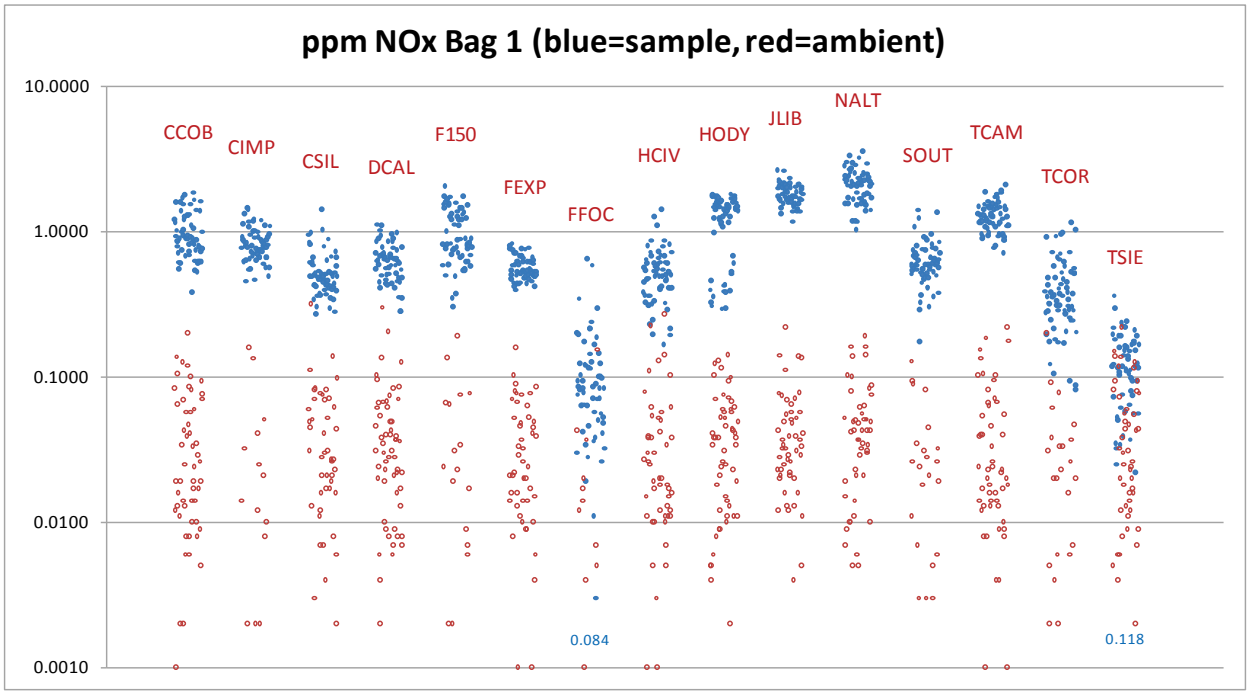
6.1.1 NO_x

Southwest Research Institute reported that the site NO_x analyzer used for this program was expected to have an inherent noise level of about ±10 ppb based on earlier experiments involving repeat measurement of very low concentrations. Two separate but related issues exist to varying degrees within the NO_x dataset. One is the issue of very low sample measurements for some vehicles, resulting in sufficient inherent noise to cause a large relative measurement error. Another is the overlapping ranges of the sample and background measurements for these same vehicles, such that their net results may be smaller than the measurement variability. If this program were simply trying to quantify the magnitude of NO_x emissions from such vehicles, this level of error may be acceptable. However, since we are looking for meaningful differences in emissions between fuels, this large relative error is particularly problematic.

Figure 53 shows bag 1 background and sample NO_x concentrations, with the medians for the two lowest-emitting vehicles (Focus, Sienna) shown in blue text below their points. A measurement

variability of ±10 ppb represents about 10% of the sample medians for these two vehicles. Moreover, performing a subtraction of background would double this margin, for a relative error in the net result of more than 20% for these two vehicles, because the measurement variability applies to both the sample and background measurements. This figure is of the same magnitude as the fuel effect the program was designed to detect. For the other vehicles the relative error is generally less than 5%.

Figure 53. Bag 1 Sample and Background NO_x Concentrations by Vehicle (median values in blue).



Also noticeable for the two lowest-emitting vehicles is the substantial overlap of the range of sample and background measurements, with differences in some cases being less than the measurement variability. This condition would be expected to give a zero result as discussed in Section 3.1. Meanwhile, for the other vehicles the range of sample measurements is separated by a few hundred ppb or more, generally giving a net result that is discernible from measurement variability.

As described in Section 5.5, vehicles having strong influence on model fitting were identified. For bag 1 NO_x, results show that a highly influential vehicle, the Focus, also has sample measurements in the range of background levels. This suggests it likely has higher measurement noise than data from the other vehicles, and thus many measurements may not be reliably distinguishable from background levels. Since each vehicle is subjected to the same number of tests on the same fuel set, under the same test conditions, a vehicle can reasonably be considered an integral sampling unit. Thus, the decision was made to remove the measurements on the

Focus from subsequent Bag-1 modeling analysis. The data for the Sienna, though similar in their range of sample and background measurements, were not found to be exceptionally influential to model-fitting and therefore were not removed from the dataset.

Figure 54 shows background and sample NO_x concentrations for Bags 2 with medians for the two lowest-emitting vehicles (Cobalt, Focus) shown below their data in blue text. The measurement variability of ±10 ppb represents more than 20% of the median sample measurement for the Cobalt, and more than 10% for the Focus. In addition, subtraction of background results in the majority of Cobalt net results being smaller than the variability.

As described above, the influence analysis identified the Cobalt as very strongly influential in model fitting (Table 20). This graphical analysis suggests that the Cobalt measurements are, like those for the Focus in Bag 1, affected by a high level of measurement uncertainty, and may not be reliably distinguished from background levels. Thus, based on judgment parallel to that applied for bag 1, the subset of measurements for the Cobalt were removed from the dataset used for subsequent modeling analyses.

Figure 55 shows Bag 3 NO_x data. In this case, emission levels for most vehicles are lower than in Bag 2, and show even more overlap with background measurements. This dataset as a whole is expected to contain a large amount of measurement error, and models produced from it would be of questionable value.

Figure 54. Bag 2 Sample and Background NO_x Concentrations by Vehicle.

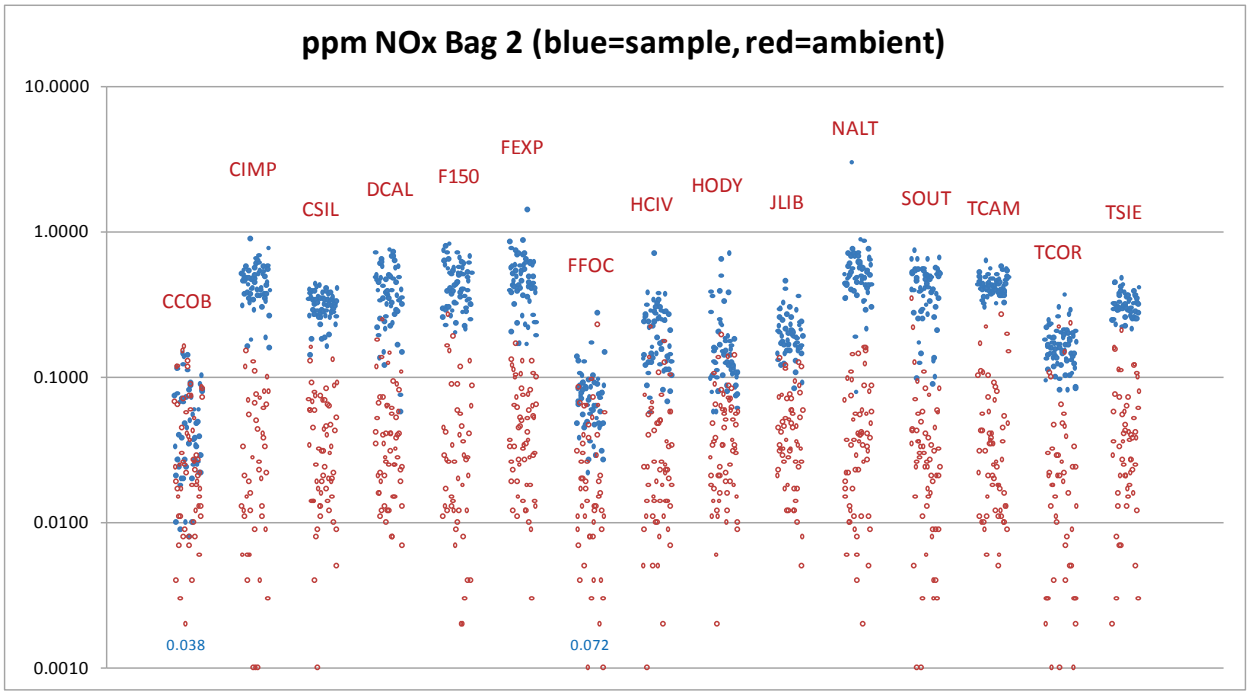
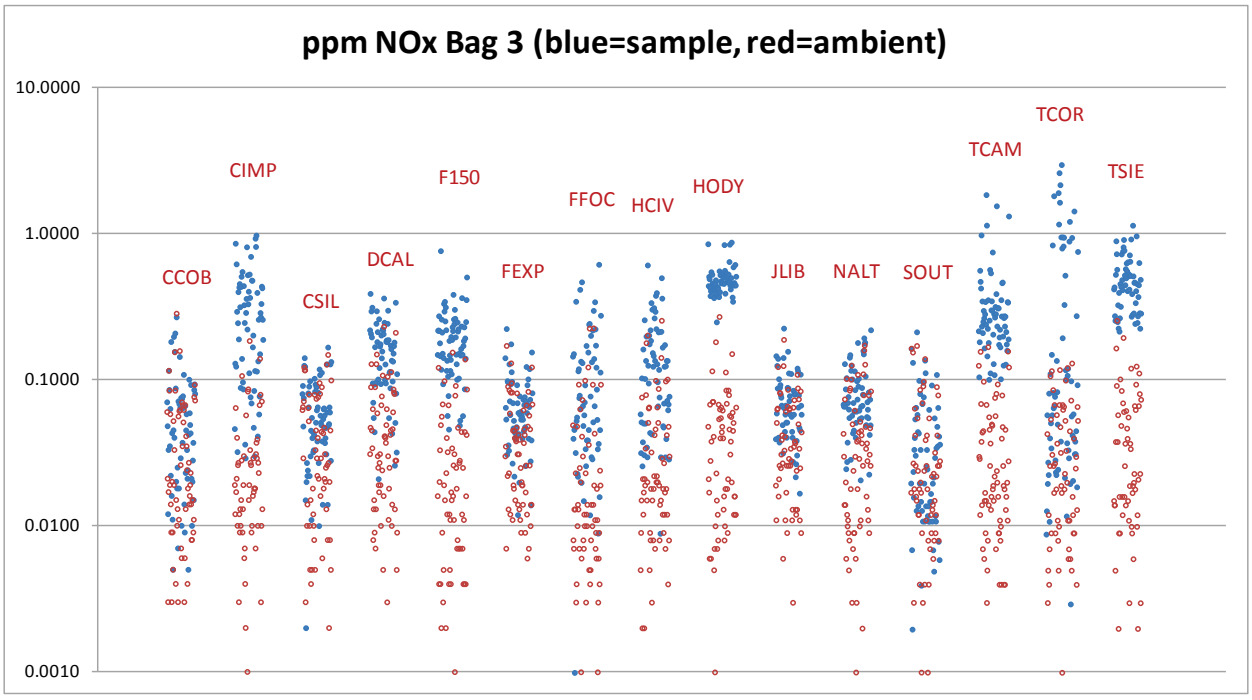


Figure 55. Bag 3 Sample and Background NO_x Concentrations by Vehicle.



6.1.2 Particulate Matter

Unlike most gaseous pollutants, measurement of particulate matter (PM) mass involves handling filter media, including pre- and post-weight measurements done, in the case of this program, by a human operator using a microbalance in a clean room. Prior to a test, new filters are acclimated for a period of at least 12 hours in the clean room, and each is marked with a serial number then weighed three times to produce an average pre-test weight. The filters are placed into clean containers, transported to the test cell, and inserted into the sampling apparatus. After the test is conducted, the filters are removed, placed back into the clean containers and returned to the clean room. The acclimation and weighing procedures are repeated as before to produce an average post-test weight. The difference (in micrograms) is referred to as the filter weight gain, which is then combined with dilution and sample flow data to produce a PM emission rate for the test period (by bag). In this program, dilution air was HEPA-filtered and presumed to be free of PM, so there was no background filter sample collected for later subtraction as is typical with other emissions. Thus, the net PM result is generated from the clean and dirty weights of a single sample filter.

Particulate emission rates from vehicles are highly sensitive to driver behavior, and thus significant variability is expected in the measured parameter itself. One source of variability in the measurement process is unintended gain or loss of material during filter handling. Material gain may occur from accidental exposure to ambient dust or stray particles released from the sampling apparatus, while loss may occur from exposure to elevated temperatures or air currents. Two measurement process parameters, filter temperature and sample flow velocity at the filter face, were captured during each test as part of data quality assurance.[m] Other measurement artifacts may occur due to insufficient neutralization of static charge on the filter, as well as errors in accounting for temperature, humidity, and barometric pressure during filter storage and weighing. The process of averaging triplicate weights attempts to mitigate some of these.

Discussion with EPA staff experienced with PM measurement suggests that for the data as collected in this program, where typical certification procedures and other best practices were generally followed, a variability of approximately ±1 μg should be assumed for all filter weights.[n] Considering that the net PM result is calculated by subtracting two filter weights (dirty minus clean), it should be understood to have a variability range of ±2 μg, as the measurement error applies to both weights. Therefore, a net weight gain of 10 μg would have a relative error of 20% associated with it, a figure of the same order of magnitude as the fuel effects we are attempting to capture.

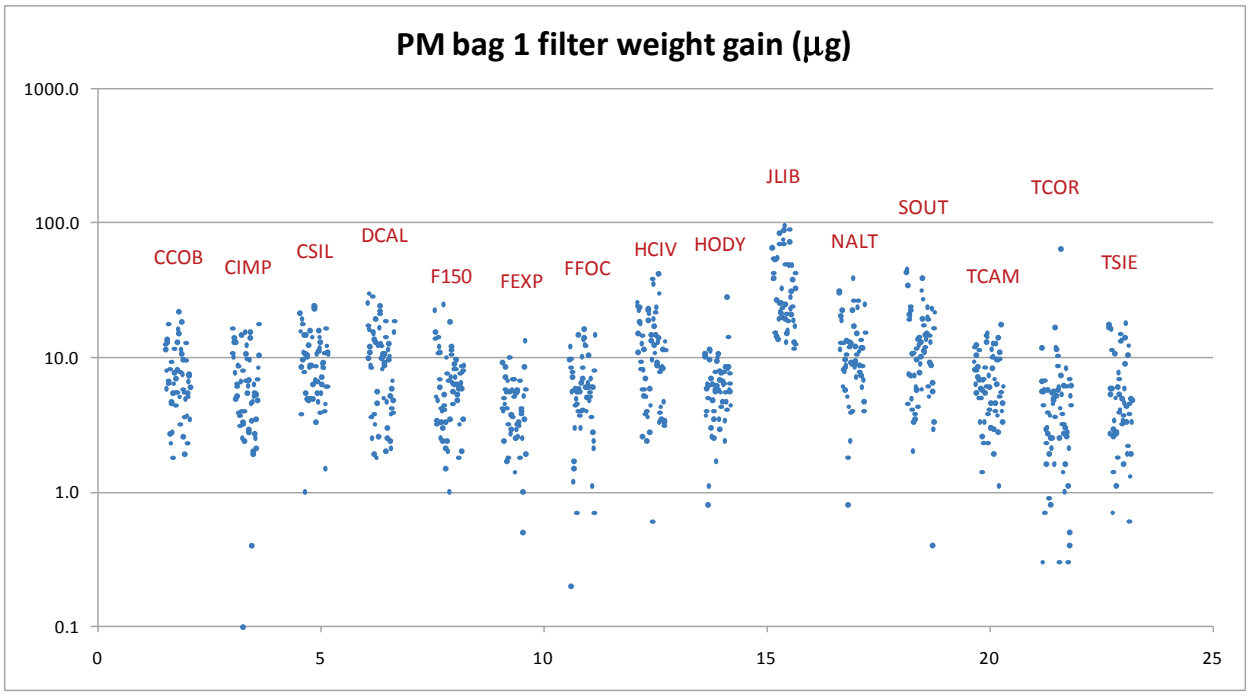
---

[m] These data are available in the "Output-QA" sheet of the individual test files.

[n] See, for example, Figure 2 in Chase, R., Duszkiewicz, G., Lewis, D., and Podsiadlik, D., "Reducing PM Measurement Variability by Controlling Static Charge," SAE Technical Paper 2005-01-0193, 2005, doi:10.4271/2005-01-0193.

Filter weight gain data for bag 1 spans a range of 1 to 100 µg as depicted in Figure 56. The majority of results are below 10 µg and thus contain substantial error relative to the fuel effects being investigated. Across the dataset, most vehicles' PM results fall into a range of one order of magnitude over a range of 2.0-20 µg (with the possible exception of the Liberty). Looking at the figure, no vehicles stand out as having a large number of very low measurements. Additionally, no particularly influential vehicles were found in the influence analysis (Table 23) . Thus, while there were concerns that the model may contain artifacts due to the uncertainty in the data, no vehicle-specific subsets of data were more suspect than others. Therefore, all bag 1 PM data were retained for subsequent model fitting.

Figure 56. Bag 1 Particulate Mass Filter Weight Gain by Vehicle.



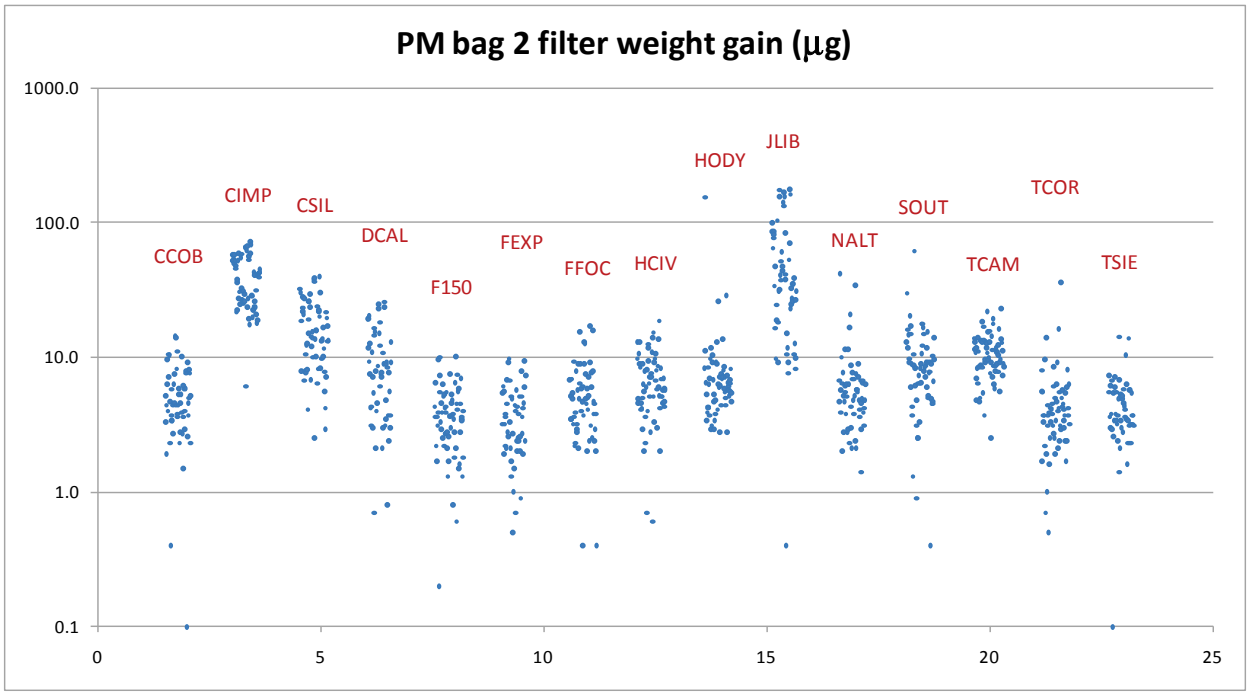
Filter weight gain data for bag 2 spans a range of about 1 to 200 µg as depicted in

Figure 57. In general, these results are very similar in magnitude and range to those produced in bag 1. Influence analysis did not point to any particularly influential vehicles, and again, dataset manipulation was not expected to be useful in reducing the presence of variability or other artifacts in modeling.

While bag 1 PM is largely a result of over-fueling during the first few seconds of cold engine start-up, bag 2 of the LA92 test cycle contains some high-acceleration portions, one of which (around 865 seconds elapsed test time) requires the driver to briefly use wide-open-throttle on some vehicles, potentially bringing the vehicle into open-loop or enrichment fuel control mode

for a few seconds. This operation mode is known to have greatly increased emission rates of PM, which may account for the relatively high PM masses shown here for a portion of the test that is fully warmed-up and would otherwise have the vehicle's emission controls working at high efficiency.

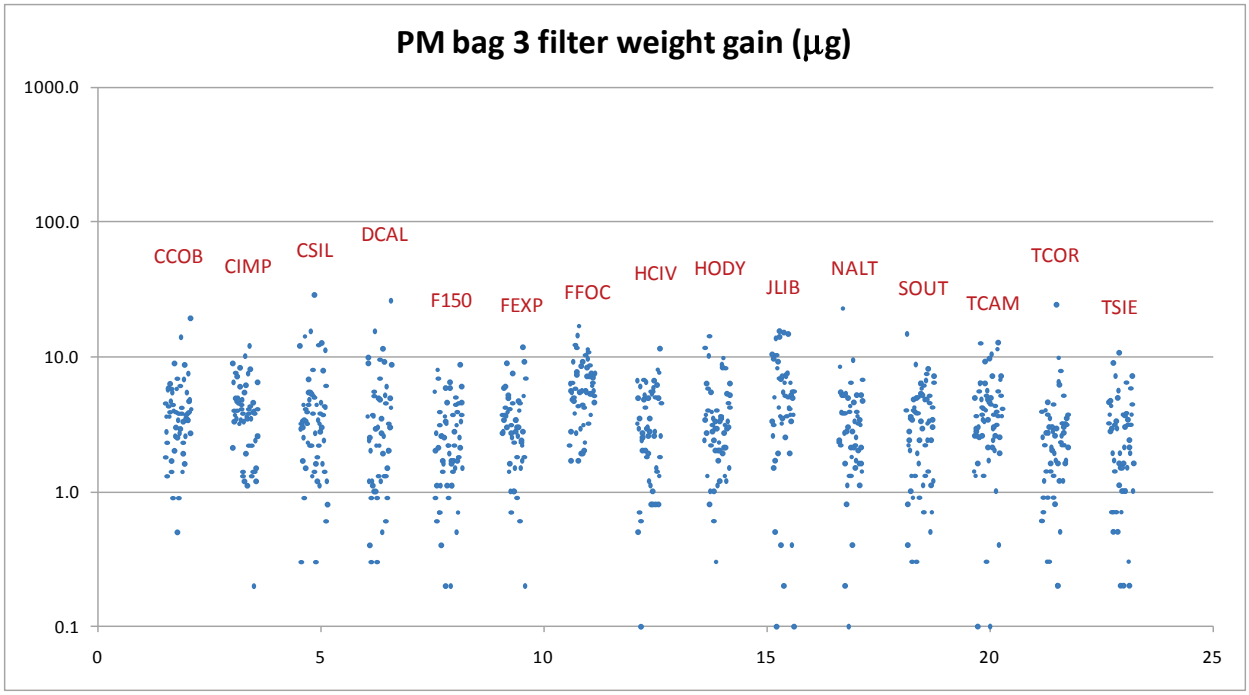
Figure 57. Bag 2 Particulate Mass Filter Weight Gain by Vehicle.



Filter weight gain data for bag 3 is shown in Figure 58, with lower median magnitudes and spanning a smaller range than in bags 1-2. Nearly all the data here are below 10 µg, with the majority of masses below 5 µg. PM emissions in bag 3 are small because the hot restart has much less over-fueling than the cold start seen in bag 1, and the driving schedule is much milder here than in bag 2.

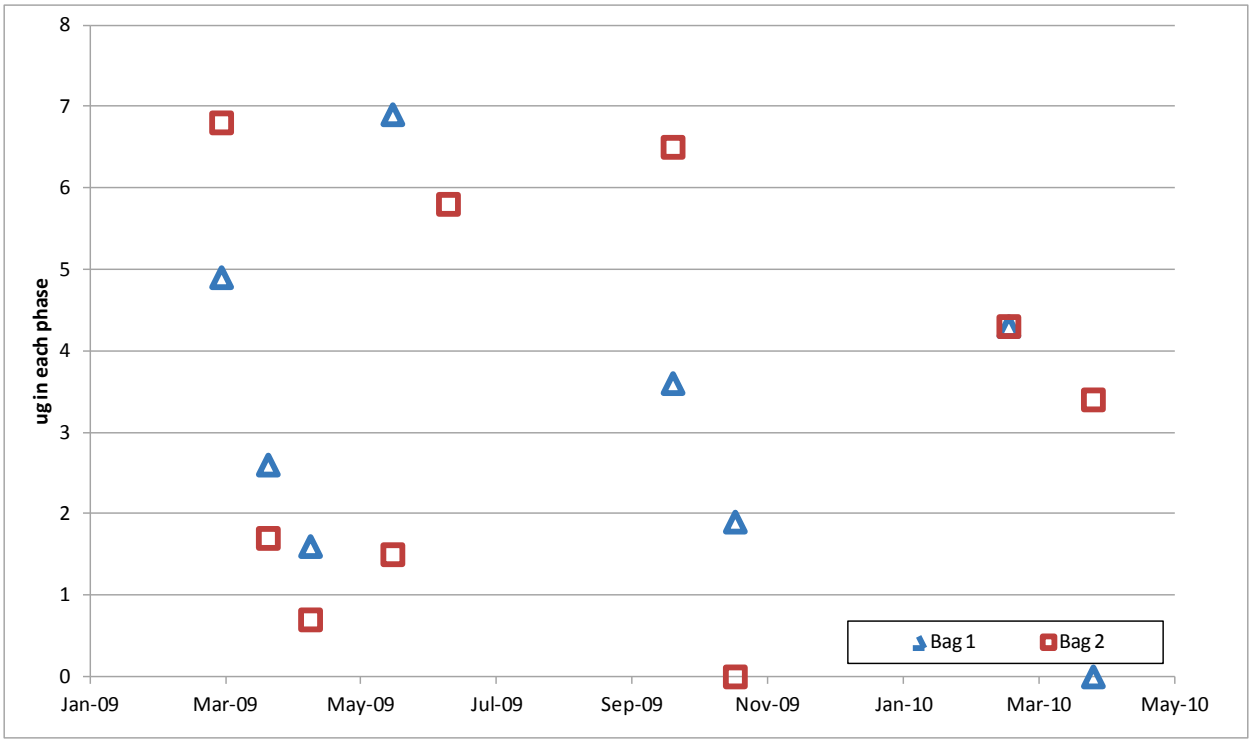
Influence analysis was again unable to discern any subsets of the data that were likely to contain more measurement noise than others. In this case, measurement variability is considered large relative to the fuel effect being investigated and modeling of the data should be done with caution. Indeed, model fits presented in Section 7 show that magnitude and sign of coefficients are highly variable depending on which terms are included in the model, behavior that is consistent with a dataset containing substantial uncertainty.

Figure 58. Bag 3 Particulate Mass Filter Weight Gain by Vehicle.



Tunnel blanks were performed periodically throughout the data collection period as a screen for significant contamination or other measurement issues. Results are shown in Figure 59. This procedure is done by setting up and running the dilution and sampling systems as if an emission test were occurring, but without collecting any vehicle exhaust (the intake is typically capped). The results of this procedure show background levels of all the pollutants at the time the blank is being run, combined with any capture or release of material from internal surfaces of the sampling equipment. Additionally, for pollutants like PM and speciated compounds, the results include any contamination or variability introduced during handling and analysis of sample media. Generally speaking, tunnel blanks are of limited utility in assessing the contribution of material released from sampling system surfaces during an actual test, since the blank test procedure does not produce the same range of temperature and humidity conditions within the system.

Figure 59. Particulate Mass Tunnel Blank Filter Loading for Bags 1-2.



The PM blank results may be of some interest here because PM data were not background corrected. Blank filter weight gains for bags 1-2 varied within a range of approximately 0-7 µg. While these results suggest that background levels may have contributed to any given measurement, they do not support the conclusion that there is any recurring or systematic bias due to high background or tunnel contamination. Therefore, in addition to the fuel effect information and test-to-test variability in vehicle operation, we would expect the PM dataset to contain additional variability due to inclusion of variable background and media handling components. All these sources of variability are accommodated by the statistical procedures used here.

6.1.3 NMOG and NMHC

Values for NMOG (and NMHC as defined in this program) are computed from seven results for each bag: THC (measured by FID), methane, methanol, ethanol, 2-propanol, formaldehyde, and acetaldehyde. The results for the five oxygenated species are each comprised of a linear combination of nine different measurements, including primary and secondary samplers in series for both background and sample, and a running average of five days of media blanks. There is further statistical complexity in assessing aggregate measurement error due to the fact that the

emission level of the individual species is not independent from the THC measurement (of which they are a subset).

The speciated carbonyl and alcohol results were subjected to limits of quantitation (LOQs) by Southwest Research Institute before being used in the NMOG calculations. These LOQs were set by tracking the behavior of media blanks (the cartridges used to make the measurements). Background and sample results were determined to be different from zero only if the blank-corrected measurement was greater than three times the standard deviation of a running five-day set of media blanks. The final result was then reported as nonzero if the sample exceeded background.[o] This process was developed to accommodate the fact that the concentration of species of interest in dilute exhaust was often of similar magnitude to levels present as low-level contamination in the measurement media. However, this additional rigor resulted in a relatively conservative standard for quantification, such that nonzero mass emissions should contain relatively low measurement error.

The THC results are produced by FID analyzers in the test cell, and have measurement errors on the order of 10 ppb. In the range of data being measured (background levels around 2 ppm, samples in Bag 1 >10 ppm and in Bag 2 around 3 ppm), this represents less than 1% measurement error. Figure 60, Figure 61, and Figure 62 show sample and ambient THC measurements for bags 1, 2, and 3, respectively. Figure 63 and Figure 64 show net measurements as a percentage of sample for bags 2 and 3, as an attempt to understand the magnitude of the nets relative to the measurement error where nets are small. Several vehicles show a portion of tests with net concentrations less than 10% of sample measurement, which would put them in the same order of magnitude as the estimated error of measurement. Two vehicles (Odyssey, Sienna) have the majority of points below those of other vehicles.

The methane results are similar in magnitude and relationship to background as THC, though methane generally shows less cold start effect than total hydrocarbons because it is not a fuel component and therefore is not directly emitted during initial start-up. Measurement error for methane is expected to be similar to that for THC. Plots of ambient and sample for bags 1-3 are shown in Figure 65, Figure 66, and Figure 67, followed by examination of nets as percentage of sample for bags 2 and 3 in Figure 68 and Figure 69. In this case, six vehicles have most or all of their nets falling under 10% of sample concentrations. For the two vehicles mentioned above (Odyssey, Sienna), measurement error may again play a significant role in the variability of the net measurement.

While a number of vehicles have measurement error concerns, influence analysis for Bag 2 NMOG identifies the Odyssey and Sienna as highly influential vehicles in terms of overall fit, effect on fixed-model coefficients and precision of fixed-model coefficients (Table 22, page 105). Subsequent review of the measurements as shown in Figure 63 and Figure 68 again

[o] Details on this LOQ method are described in Appendix L of the EPAct/V2/E-89 testing report.

suggest that high measurement variability impairs our ability to discern tailpipe emissions from background levels. Thus, data from these two vehicles were removed from subsequent bag 2 NMOG models.

Figure 60. Bag 1 Total Hydrocarbon Sample and Ambient Concentrations by Vehicle (ppmC).

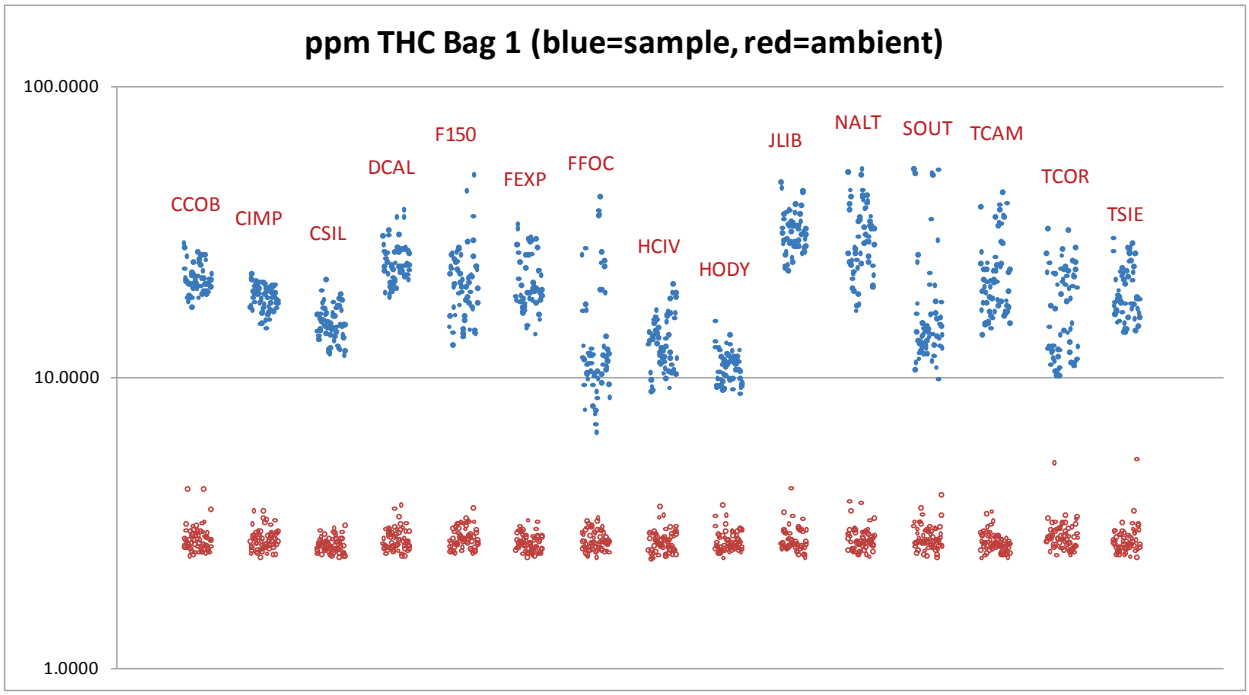


Figure 61. Bag 2 Total Hydrocarbon Sample and Ambient Concentrations by Vehicle (ppmC).

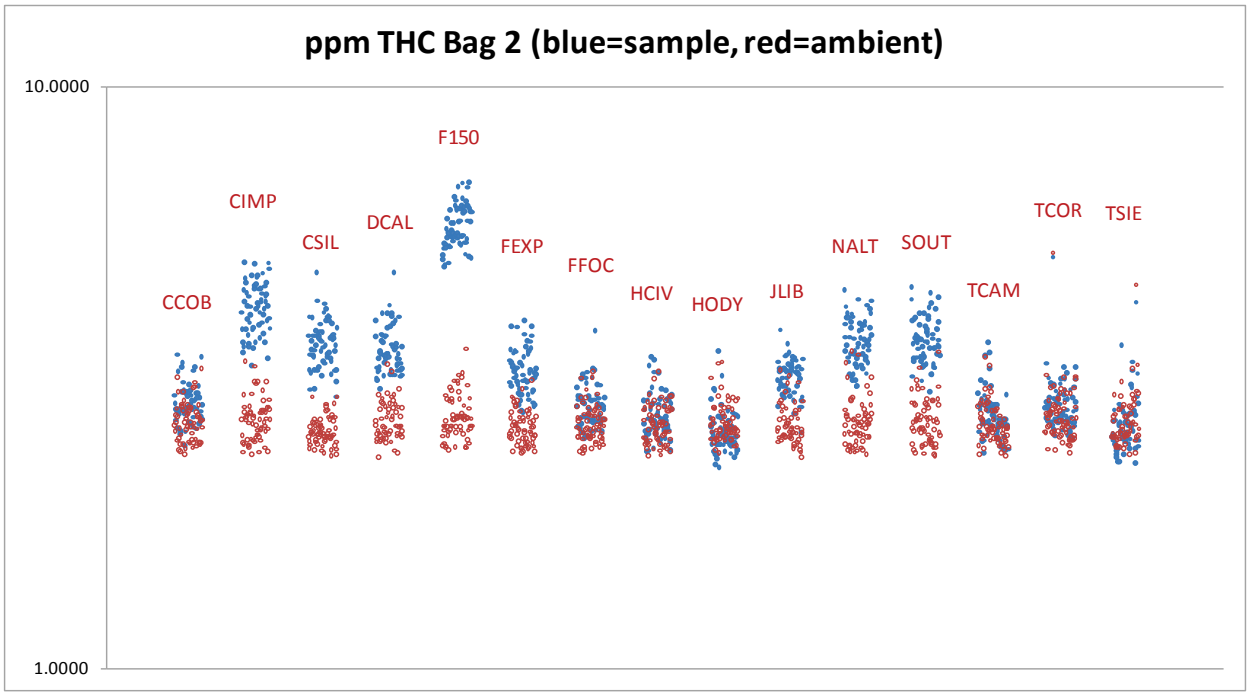


Figure 62. Bag 3 Total Hydrocarbon Sample and Ambient Concentrations by Vehicle (ppmC).

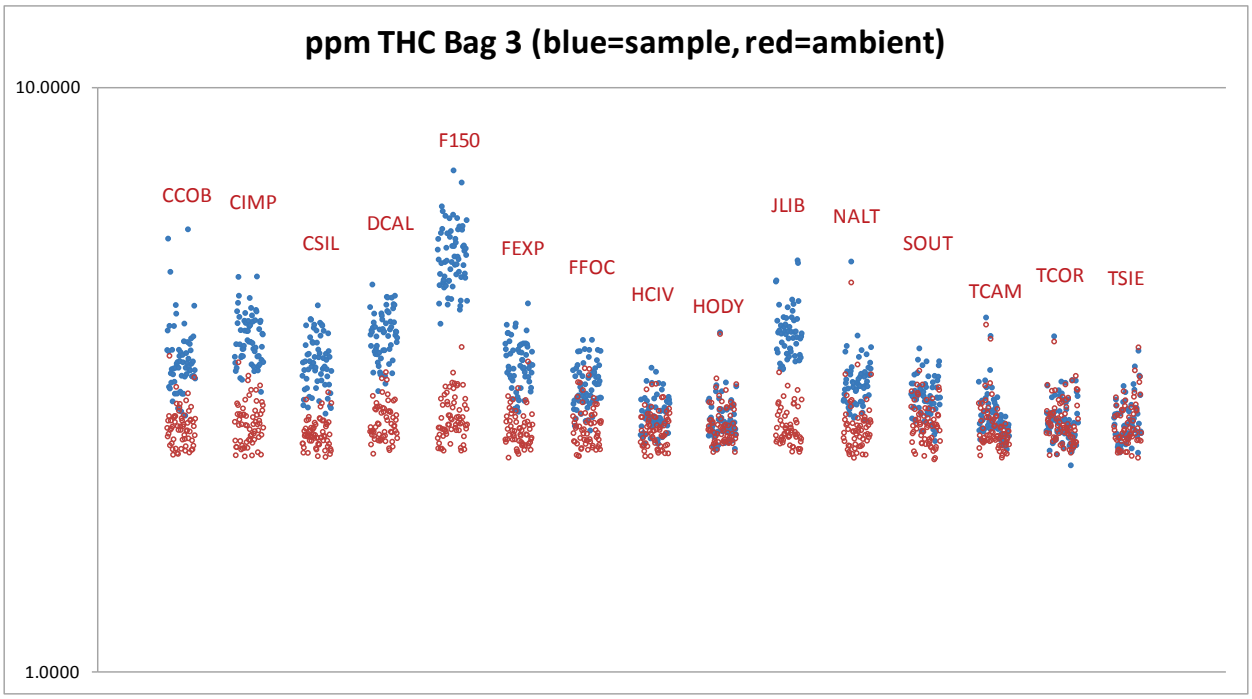


Figure 63. Bag 2 Total Hydrocarbon as Percent of Sample Measurement.

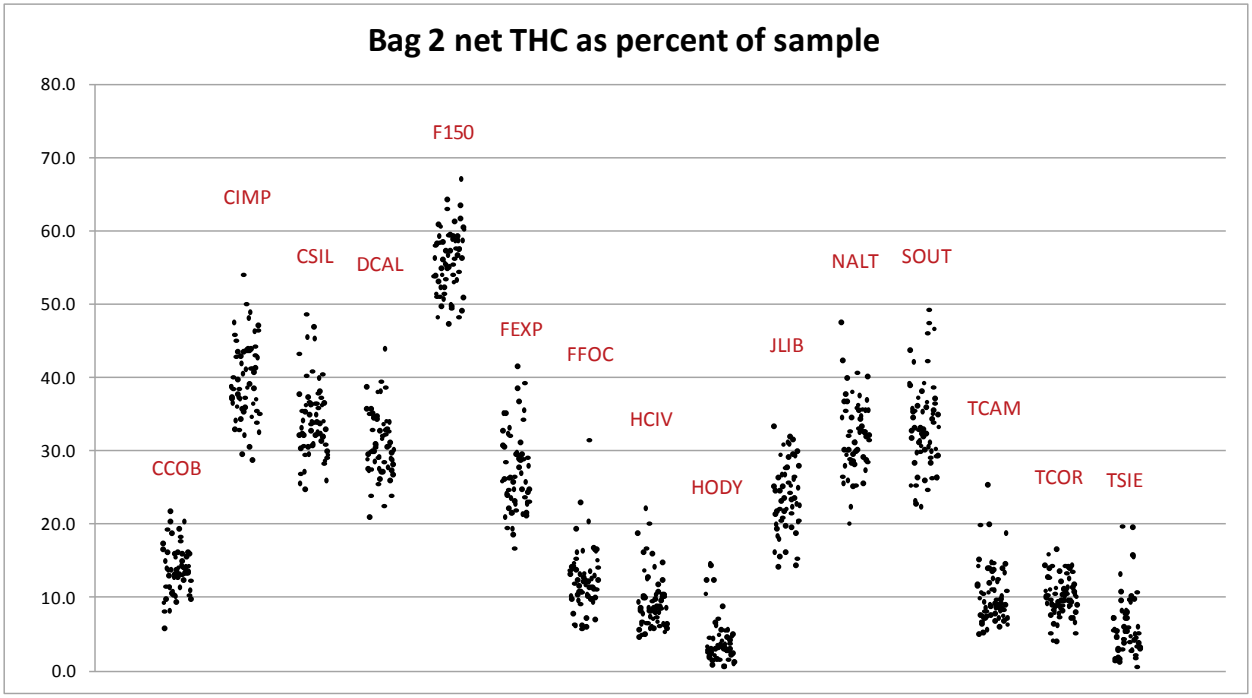


Figure 64. Bag 3 Total Hydrocarbon as Percent of Sample Measurement.

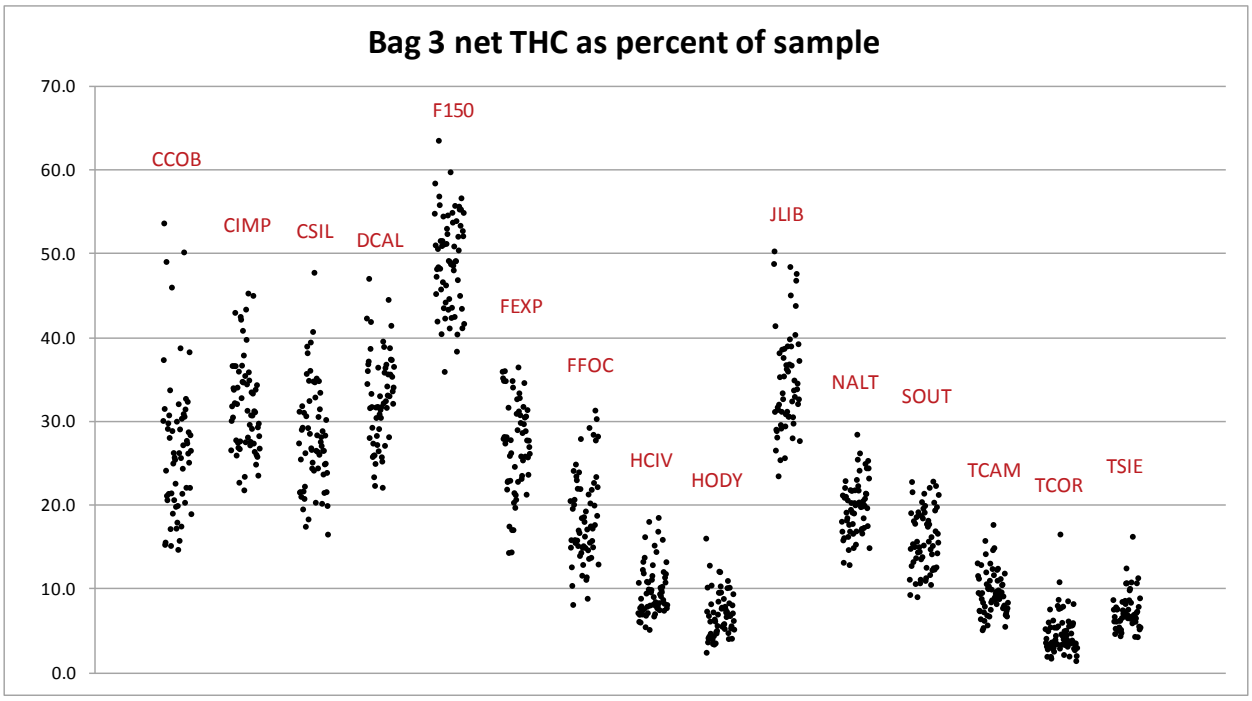


Figure 65. Bag 1 Methane Sample and Ambient Concentrations by Vehicle (ppm).

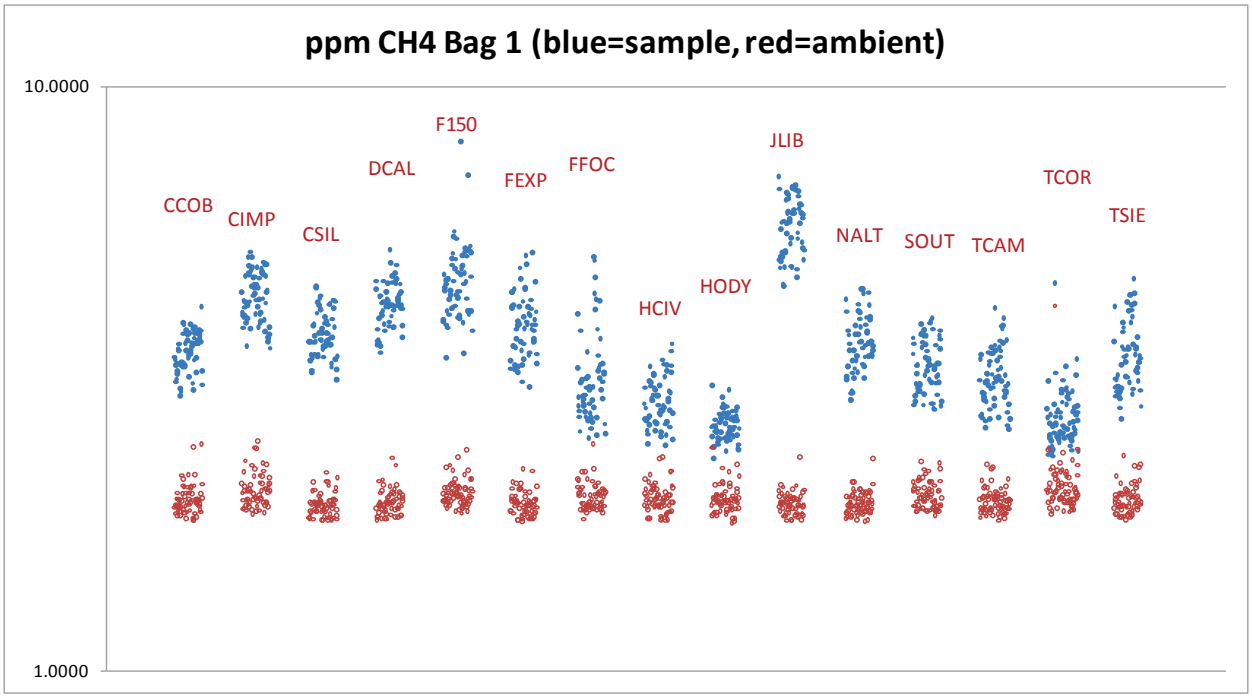


Figure 66. Bag 2 Methane Sample and Ambient Concentrations by Vehicle (ppm).

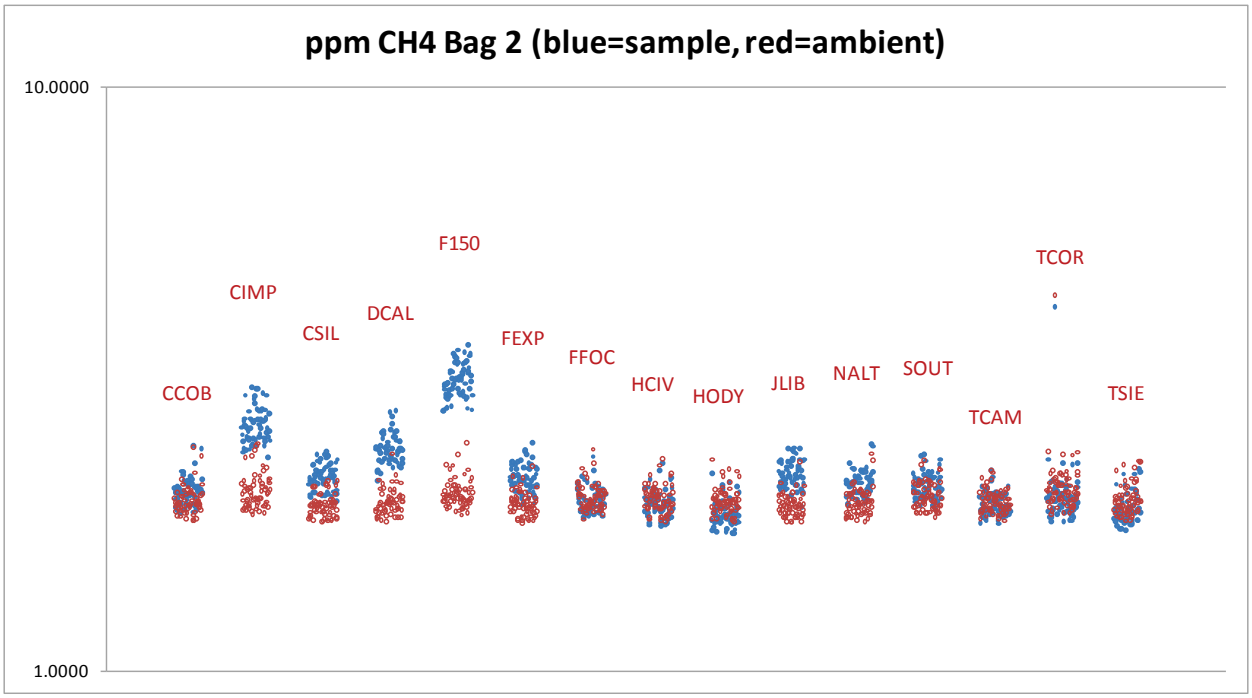


Figure 67. Bag 3 Methane Sample and Ambient Concentrations by Vehicle (ppm).

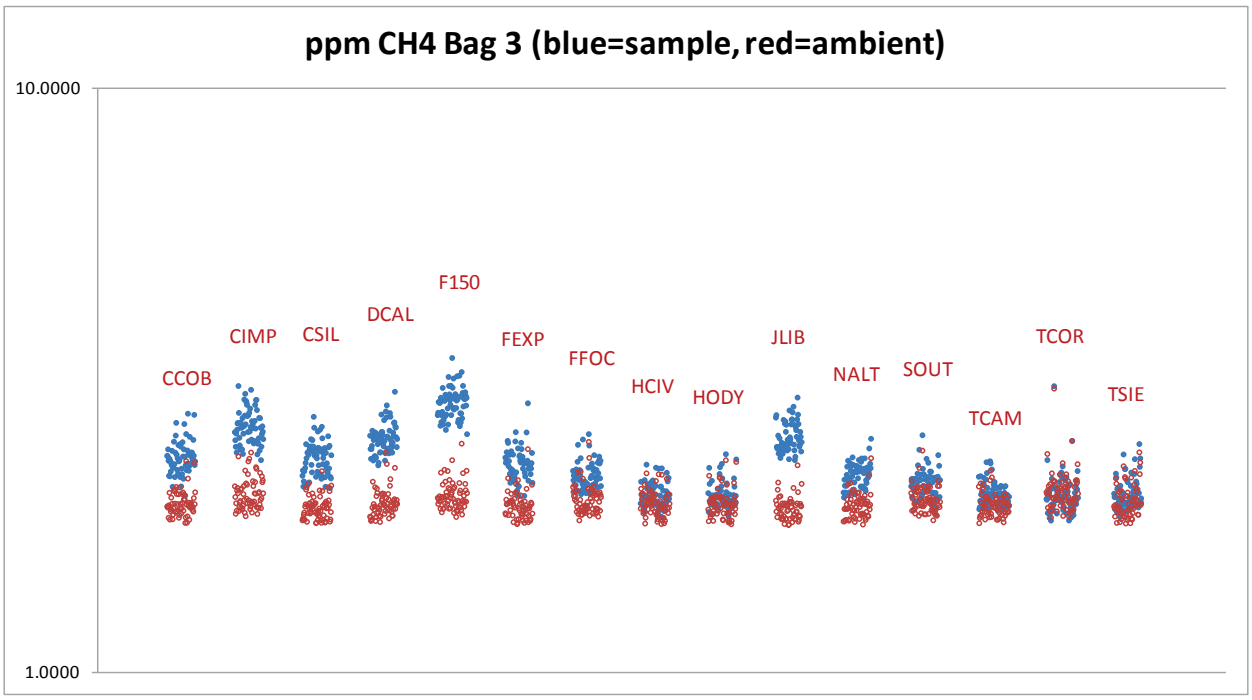


Figure 68. Bag 2 Methane as Percent of Sample Measurement.

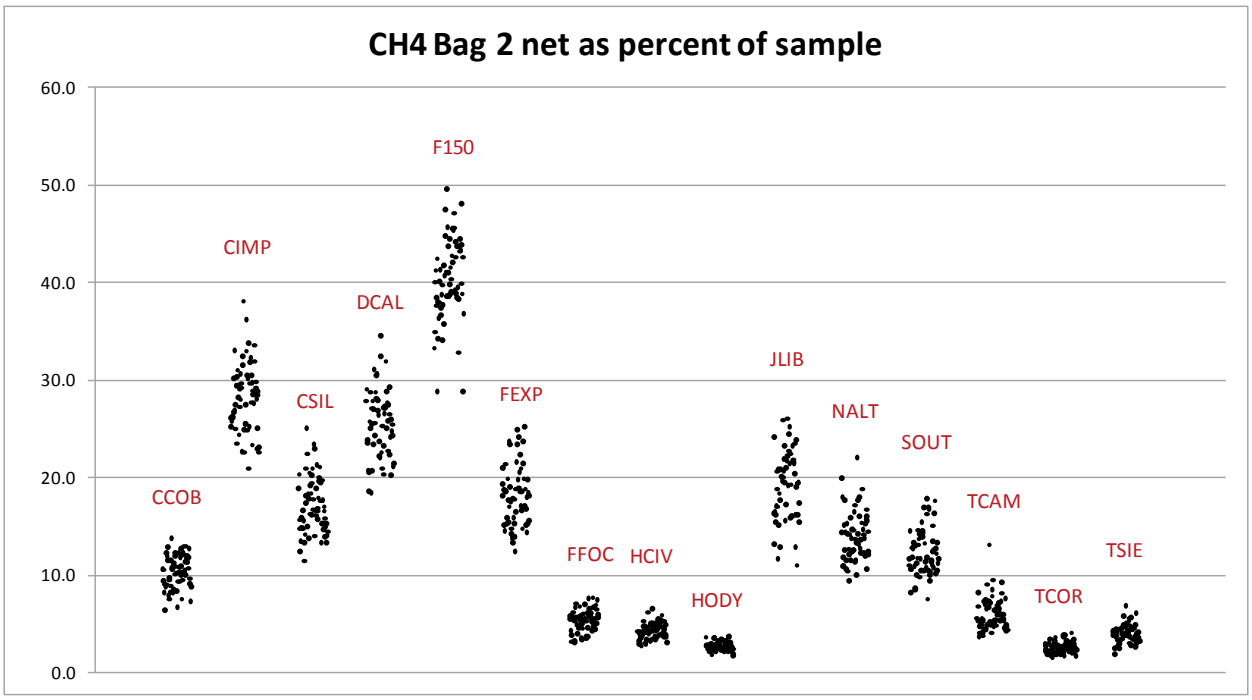
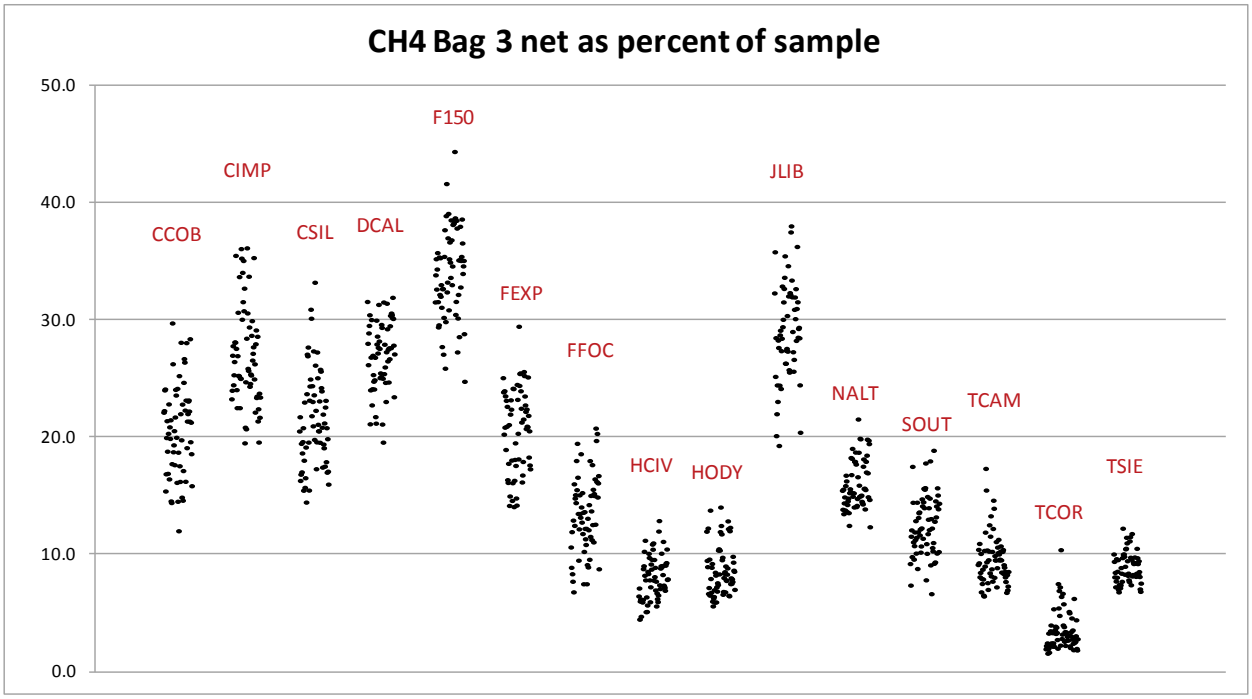


Figure 69. Bag 3 Methane as Percent of Sample Measurement.



6.2 Analyzer Drift

During investigation of influential vehicles in the NO_x dataset, concentration data were examined for both sample and background measurements. In addition to the very low emission levels seen for two vehicles, it was also noted that a large proportion of background concentrations for several vehicles were reported as zero. Further investigation revealed that these vehicles were those added to the test fleet later in the program, such that their testing had taken place after August 2009. The timeline of addition of vehicles and fuels to the test scheduling is shown in Table 29.[p] Since it appeared this issue may have been related to testing procedures or conditions, the analyzer zero check results (available in the QA sheet of the individual test files) were plotted chronologically along with the background measurements.[q] These data are shown in Figure 70

[p] The order in which vehicle-fuel combinations were tested in this program was randomized to minimize effects of any systematic bias on the subsequent analyses. However, due to funding limitations early in the program, not all vehicles began testing at the same time. As funding was received, vehicles were added to the randomization schedule.

[q] Before each test the NOx analyzer performs an automated calibration process whereby the "zero" and "span" levels are set by flowing gas streams of known concentrations through the analyzer. Zero and span checks are subsequently made after each bag is analyzed to verify that the calibration held.

and Figure 71. This presentation shows that the zero point on the NO_x analyzer had been experiencing an offset of seemingly random magnitude during each test beginning in August 2009 (affecting both sample and background). The offset values were distributed over a range of about 300 ppb, mostly negative for bag 1 and mostly positive for bags 2-3. As there was no discernible pattern of offset behavior among bags within a test, or tests within a day, it wasn't clear that the zero check measurement taken at the end of each bag was representative of the measurement process immediately preceding it. Thus, no drift corrections could be applied to the dataset. More details of this analysis are available in Appendix F.

Table 29. Timeline of Addition of Fuels and Vehicles to the Test Scheduling.

Phase 3 Week	Fuels Added	Vehicles[a] Added	Vehicle/Fuel Assignments
Week 1	2, 7, 8, 9 and 15	CCOB, TCAM, FEXP, DCAL, HODY	EPA
Week 2	None	CSIL, TSIE, DLIB, HCIV, NALT	
Week 3	None	None	
Week 4	1, 12, 13	None	
Week 5	None	None	
Week 6	22, 24	None	
Week 7	None	None	
Week 8	3, 4, 5, 11, 14, 16, 20, 21, 23, 30	None	
Week 9	None	None	
Week 10	None	None	Randomized for rest of program, except for E85
Week 12	6, 10, 25, 26, 27, 28, 31	None	
Week 25	None	FFOC, SOUT	
Week 37	None	CIMP, F150, TCOR	
Week 55	29 (E85)	DCAR	Last fuel tested
Week 60	End of Phase 3 testing		

[a] - Vehicle designations are explained in Section 2.

Figure 70. Bag 1 NO_x Background and Zero Check by Date.

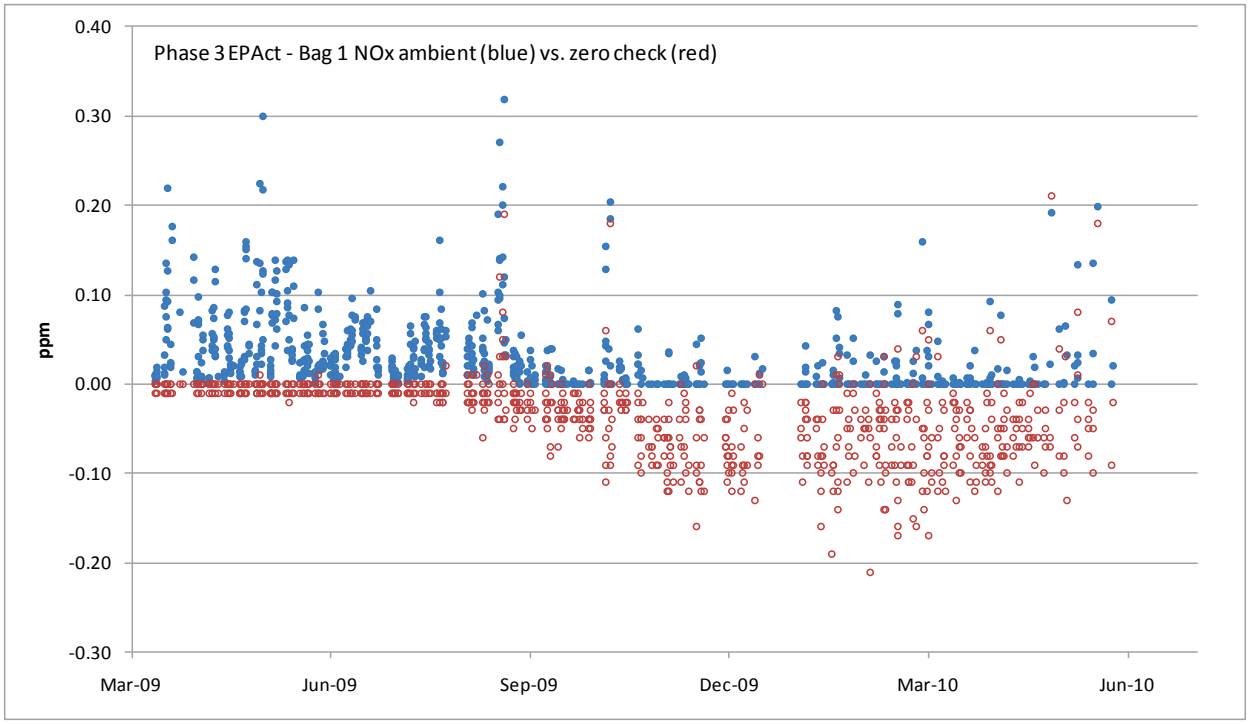
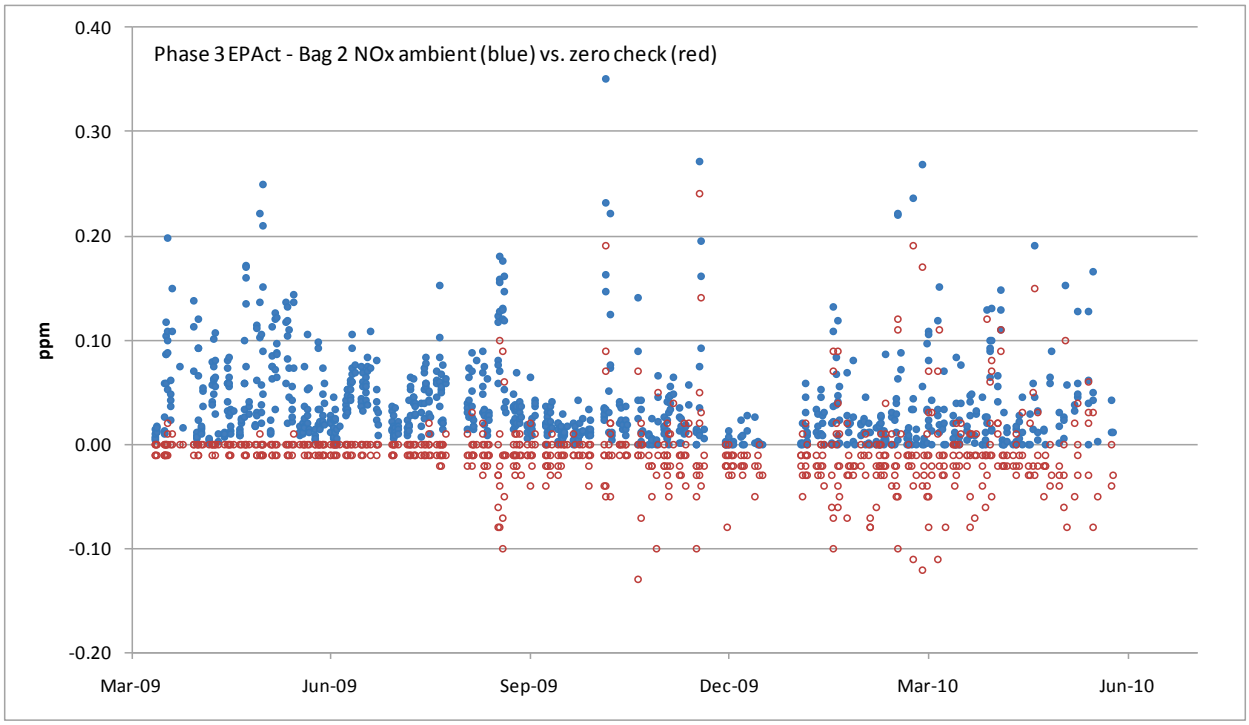


Figure 71. Bag 2 NO_x Background and Zero Check by Date.



The root cause of the offset was never fully determined, but after some troubleshooting with SwRI staff, the most likely explanation was intermittent malfunction or failure of a component in the analyzer. Immediately following the completion of testing, and before the discovery of this zero offset issue, the analyzer underwent scheduled maintenance that included replacement and recalibration of several components around the reaction cell and related measurement electronics. Thus, follow-up testing to further characterize the malfunction behavior was not possible.

As a result, each NO_x result has a random noise component included, which is more significant on a relative basis for the lowest-emitting vehicles, adding to the measurement concerns discussed in Section 6.11.

7 Final Modeling

7.1 Design Efficiency for Extended Models.

The optimized design for the fuel matrix and its development was previously described above (Section 2.1, page 19). As mentioned, the design was optimized for a model including eleven terms, with a resulting *G*-efficiency of 51.6%. This level of efficiency is considered to be at the low end of the range considered adequate for efficient estimation of effects[17].

For purposes of analysis, additional parameters were allowed into the full model, to explore the possibility that important effects might not have been included in the design model. All effects considered for inclusion are listed above in Table 17, here reproduced as Table 30.

As an additional step, we retrospectively evaluated the design efficiency of the extended model (design terms plus additional terms). In addition to the 11-term design model, we evaluated the efficiency for two additional models, one of which includes 17 terms, and a second that includes 16 terms, after omitting the T90×T90 term. In Table 30, these models are denoted as the "11-term," "17-term" and "16-term" models, as mentioned in 5.1and 5.3, respectively. The *G*-efficiencies of the three models are shown in Table 31. Results show that inclusion of the additional six or five terms reduces the effective *G*-efficiency to values between 20 and 23%.

In terms of design, this outcome implies that had some or all of the additional terms been included in the design model, the fuel matrix would have required additional fuels to achieve an adequate *G*-efficiency. In terms of analysis, this result does not necessarily or absolutely preclude the additional terms as candidates for inclusion. However, it does suggest that caution is required in the fitting of the additional terms, as the additional terms may be estimated less precisely than the design terms. As an initial precaution, it is valuable to review whether any of the additional terms are highly correlated with each other, or with any of the design terms, expanding the analysis shown earlier for the design terms (Table 6 - Table 9).

A correlation matrix for standardized model terms, incorporating one-stage standardization for linear terms and two-stage standardization for quadratic and interaction terms is shown in Table 32. The table shows that no high correlations exist among the additional terms, or between additional terms and design terms, with two exceptions. Thus, with appropriate cautions and qualifications, it is possible to consider candidate terms from among the additional terms for inclusion in reduced models.

It appears that fitting models based on the extended full model may yield important insights. However, given the fact that design efficiency for the extended full model is low, it is not certain that reduced models based on the extended model will be retained for use in prediction or description, following additional review and interpretation.

Table 30. Description and notation for parameters included in model fitting

Term No.	Fuel Parameter	Model term	In optimized design?	11-term	17-term	16-term
1	Ethanol content (%)	etOH	YES	●	●	●
2	Aromatics content (%)	Arom	YES	●	●	●
3	RVP (psi)	RVP	YES	●	●	●
4	T50(°F)	T50	YES	●	●	●
5	T90 (°F)	T90	YES	●	●	●
6		etOH × etOH	YES	●	●	●
7		T50 × T50	YES	●	●	●
8		etOH × Arom	YES	●	●	●
9		etOH × T50	YES	●	●	●
10		etOH × T90	YES	●	●	●
11		etOH × RVP	YES	●	●	●
12		Arom × T50	NO		●	●
13		Arom × T90	NO		●	●
14		T90 × T90	NO		●	
15		T50 × T90	NO		●	●
16		Arom × RVP	NO		●	●
17		RVP × T90	NO		●	●

Table 31. Design Efficiencies for Sets of Candidate Model terms.

Model	*G*-efficiency (%)
11-term	51.6
17-term	20.6
16-term	22.3

Table 32. Correlation matrix for Standardized linear-effect (one-stage) and interaction (two-stage) terms in the full-design fuel-parameter matrix.

	Ethanol	Arom	RVP	T50	T90	Benz	Olefins	etOH×etOH	T50×T50	T90×T90	T50×T90	etOH×T50	Arom×T50	etOH×T90	RVP×T90	Arom×T90	etOH×Arom	etOH×RVP	Arom×RVP
Ethanol	1.00	-0.04	-0.15	-0.57	-0.01	0.75	0.08	-0.03	-0.01	-0.18	-0.02	-0.08	0.12	-0.02	0.09	0.02	-0.08	-0.01	-0.02
Arom		1.00	0.05	-0.10	-0.02	-0.07	-0.54	-0.11	-0.10	-0.23	-0.31	0.15	0.00	0.02	0.15	0.01	0.03	-0.02	0.05
RVP			1.00	-0.26	0.13	-0.15	-0.36	-0.01	0.16	-0.07	0.06	-0.06	-0.34	0.09	-0.02	0.15	-0.02	0.02	0.01
T50				1.00	-0.02	-0.60	-0.16	-0.08	0.12	0.26	0.04	-0.01	-0.08	-0.02	0.06	-0.30	0.12	-0.05	-0.34
T90					1.00	0.00	-0.06	-0.03	0.05	-0.52	0.10	-0.03	-0.30	-0.07	-0.02	-0.08	0.02	0.08	0.15
Benz						1.00	0.39	-0.18	-0.13	-0.05	-0.17	0.10	0.15	0.07	0.15	0.05	-0.06	0.04	0.23
Olefins							1.00	-0.04	-0.28	0.21	0.02	0.21	0.14	0.04	-0.18	-0.35	0.02	0.10	0.30
etOH×etOH								1.00	-0.12	-0.07	0.07	-0.68	0.14	-0.02	0.02	0.04	-0.09	-0.07	-0.12
T50×T50									1.00	-0.18	0.07	-0.44	-0.11	-0.02	0.21	0.03	0.06	0.05	-0.42
T90×T90										1.00	0.02	0.18	0.37	-0.12	0.10	0.15	-0.31	0.06	0.14
T50×T90											1.00	-0.05	0.01	-0.53	-0.31	0.04	-0.04	-0.16	-0.20
etOH×T50												1.00	-0.06	0.04	-0.10	-0.06	-0.01	-0.04	0.40
Arom×T50													1.00	-0.04	-0.18	0.03	-0.59	0.27	-0.19
etOH×T90														1.00	-0.13	-0.16	0.01	0.14	0.14
RVP×T90															1.00	0.10	0.14	-0.02	0.02
Arom×T90																1.00	-0.04	0.13	0.15
etOH×Arom																	1.00	-0.05	-0.20
etOH×RVP																		1.00	-0.09
Arom×RVP																			1.00

7.2 Fitting Reduced Models

7.2.1 Guiding Assumptions

Incorporating results from analyses performed up to this point, we performed a final round of model fitting. The goal of the process was to identify reasonable sets of reduced models, both to describe and represent relevant processes, and for use in prediction. At the outset, we express a preference for reduced over full models, for several reasons.

1. The candidate terms in the design model were selected because we anticipated that they could be important for one or more of the measured emissions. However, we did not anticipate that all fuel properties would be important for all the compounds selected for measurement.
2. The descriptive or representational aspects of the models are very important. Insofar as possible, it is highly desirable that the models selected be intelligible and interpretable in terms of physical and chemical processes. It is also important that the terms in the models be limited to effects that describe important processes affecting emissions. Accordingly, we place emphasis on inclusion of parameters that can be precisely estimated.
3. Consistent with (2), it is important to avoid overfitting of models. We prefer to avoid inclusion of parameters that may represent study artifacts or random noise, unique to this study that might not be replicated in other studies or in real-world emissions.

For these reasons, the models reported in this section are as parsimonious as the data and subject-matter knowledge allow. That is, they include the minimum number of terms needed to avoid a decline in model fit to the Phase-3 data.

In addition, this round of fitting incorporated the outcomes of analyses previously described. We took the following specific steps:

1. We excluded the T90×T90 term, as mentioned in 5.3 (page 96). Upon consideration, we concluded that the fuel matrix was not adequate to estimate this term. Only two fuels (30 and 31) were assigned the middle T90 value (325 °F) needed to model the quadratic term. In addition, the remaining fuel properties for these two fuels were not well balanced across the parameter space. Both fuels had high aromatics and low T50 (< 165 °F). With respect to ethanol and RVP, one fuel has middle and high values, and the other high and low values.
2. We dropped selected influential observations (5.2). As mentioned, this step affected only PM (Bag 1).
3. Following influence analysis for vehicles (5.5), we dropped selected vehicles in selected analyses (Section 6). We took this step based on conclusions that subsets

of measurements for these vehicles were affected by measurement issues that called their validity into question.

In the two subsections below, we report results based on both the 16-term full model as well as the 11-term design model.

7.2.2 Methods

In the final round of model fitting, we followed a series of steps:

1) Fit all possible models.
2) Identify a set of leading candidates
3) Construct a "superset" of terms from the set of leading candidates
4) Identify the best fit model, using the "superset" model as a starting point.

These steps were performed for the gaseous emissions: THC, NMOG, NMHC, CO, NO_x and CH_4. As before, the modeling was performed separately for each of bags 1, 2 and 3. These steps were not performed for subsets of data affected by "severe censoring," as defined in 5.3.2 (page 98). Below, we describe each of these steps.

Fit all possible models: We fit all possible models respecting hierarchy, based on a designated full model, either the 11-term design model (Equation 16) or the 16-term extended model (Equation 17) (See also Table 30, page 135). The models were fit using the SAS MIXED Procedure, as previously described, and models were ranked on the basis of the Bayesian Information Criterion (BIC).

$$\begin{aligned}\ln Y = \beta_0 + \\ &\beta_1 Z_e + \beta_2 Z_a + \beta_3 Z_r + \beta_4 Z_5 + \beta_5 Z_9 + \\ &\beta_6 ZZ_{ee} + \beta_7 ZZ_{55} + \\ &\beta_8 ZZ_{ea} + \beta_9 ZZ_{er} + \beta_{10} ZZ_{e5} + \beta_{11} ZZ_{e9} + \\ &\varepsilon\end{aligned}$$

Equation 16

$$\begin{aligned}\ln Y = \beta_0 + \\ &\beta_1 Z_e + \beta_2 Z_a + \beta_3 Z_r + \beta_4 Z_5 + \beta_5 Z_9 + \\ &\beta_6 ZZ_{55} + \beta_7 ZZ_{ee} \\ &\beta_8 ZZ_{ea} + \beta_9 ZZ_{er} + \beta_{10} ZZ_{e5} + \beta_{11} ZZ_{e9} + \\ &\beta_{12} ZZ_{a5} + \beta_{13} ZZ_{a9} + \beta_{14} ZZ_{59} + \beta_{15} ZZ_{ar} + \beta_{16} ZZ_{r9} + \\ &\varepsilon\end{aligned}$$

Equation 17

Identify a subset of leading candidates: Based on inspection of a plot of BIC vs. the number of terms in each model, *p* (including the intercept), we identified a subset of the top 5 to 10 candidate reduced models.

Construct superset of terms: We constructed a set of terms that included all terms in any of the models in the set of candidates identified in the previous step. We designated this set of terms as the "superset" and used it as a *de facto* full model in subsequent modeling.

Identify a "best fit" model: Starting with the superset, we performed model fitting by backwards elimination. We used likelihood-ratio tests to evaluate goodness of fit, at a 90% confidence level ($\alpha = 0.10$). As before, we dropped terms or groups of terms if the *p*-value for the test was greater than 0.10.

The process will be illustrated using NO_x (Bag 1) and CO (Bag 1). Following these examples, the model-fitting process using Tobit regression will be illustrated using PM (Bag 1) as an example.

7.2.2.1 NO_x (Bag 1): Model fitting based on the 11-term Design Model

Inclusion of 11 terms in the "full" model gives a total of $2^{11} = 2{,}048$ possible models. However, of these, a total of only 294 models (14.5%) respect hierarchy. Results for these models are shown in Figure 72. The plot shows a measure of goodness-of-fit (BIC) versus the number of terms in the model, including the intercept (*p*). This view shows the set of models forming two major groups differing in terms of BIC. In this case, the model with minimum BIC is not clearly visible at this scale.

Figure 73 shows a close-up view of the same results, focusing on the major group with lowest BIC. This view shows that the major group is composed of an additional 3-4 subgroups. At this scale the model with minimal BIC, having five terms ($p = 5$) is apparent.

To get a closer view of the subset of models with lowest BIC, Figure 74 shows a close-up focusing on the lowest edge of the lowest main group of models. The "best" model is clearly visible, with two models fairly close in second and third places. In this case, we selected the six best-fitting candidates (BIC < 913) to construct the superset for subsequent model fitting. Table 38 shows the *p*, BIC and specific terms included in the 35 best-fitting models out of 294.

Final model fitting began with the 6 terms included in the "superset model" shown in Table 34. In this case, only one additional reduced model was fit. Two terms, with *p*-values substantially higher than 0.10, were selected for removal. The *p*-value for the associated likelihood ratio test was also well above the designated α level. Accordingly, the reduced model was retained as the "best fit" model. Note that the model selected by the goodness of fit testing also has the lowest BIC, as shown in Table 33.

Figure 72. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 11 terms in the design model.

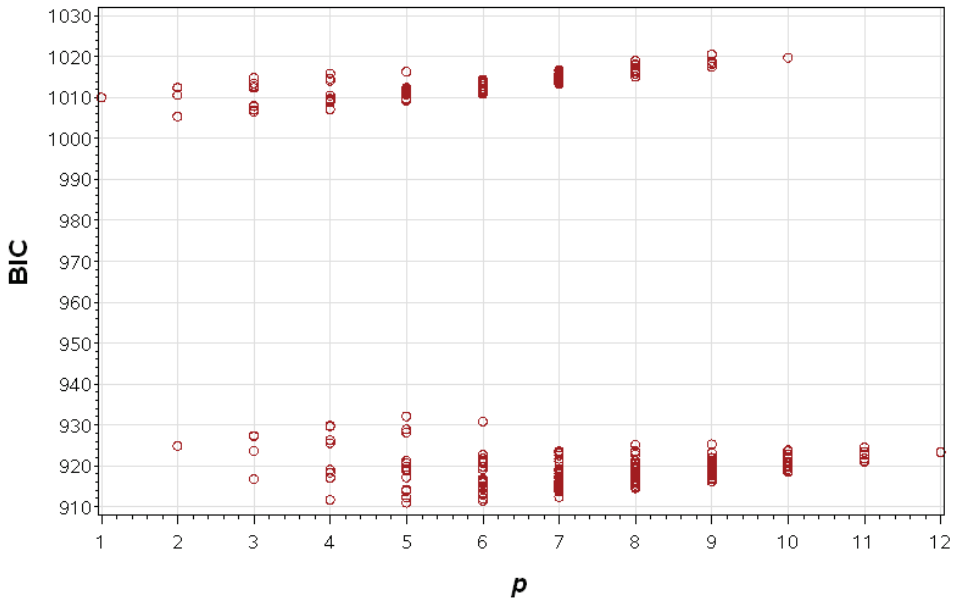


Figure 73. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy (CLOSE-UP of Figure 72).

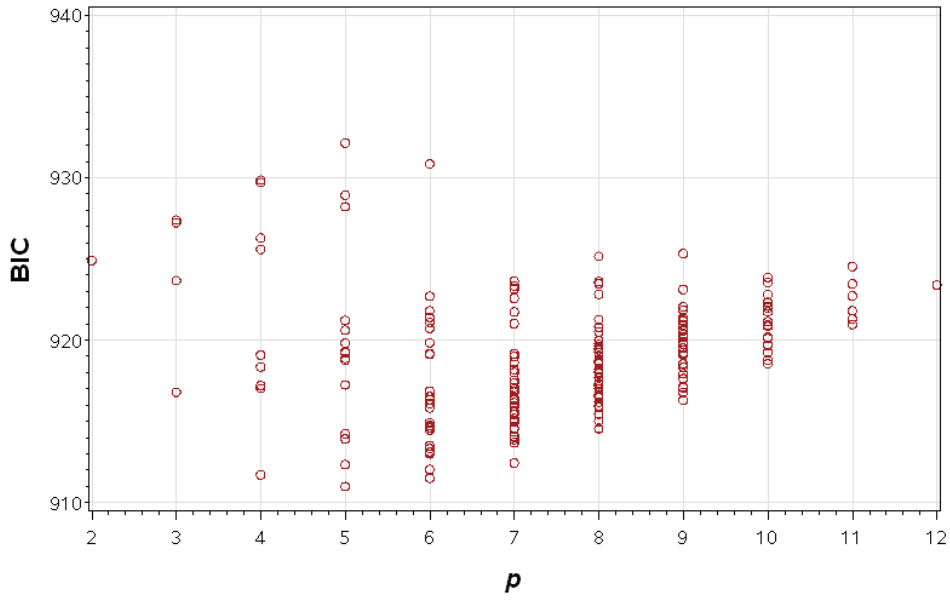


Figure 74. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy (CLOSEUP of Figure 73).

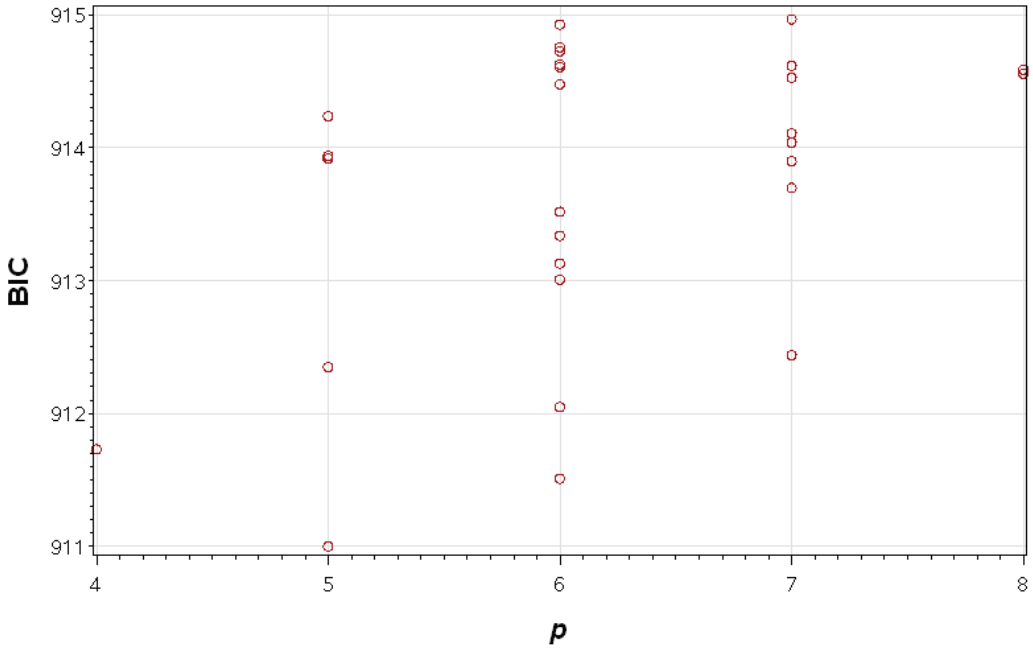


Table 33. NO_x (Bag 1): Number of terms (p), Goodness-of-fit (BIC) and terms included in the 35 best-fitting candidate models (out of a total of 294 possible models with hierarchy). (Terms included in models ranked 1-6 comprise the "superset" for final model-fitting).

Rank	p	BIC	Design Terms										
			etOH	Arom	RVP	T50	T90	etOH × etOH	T50 × T50	etOH × Arom	etOH × RVP	etOH × T50	etOH × T90
1	5	911.00	•	•		•				•			
2	6	911.51	•	•		•				•		•	
3	4	911.73	•	•		•							
4	6	912.05	•	•		•		•		•			
5	5	912.35	•	•		•		•					
6	5	912.35	•	•		•						•	
7	7	912.44	•	•		•			•	•		•	
8	6	913.01	•	•		•			•			•	
9	6	913.13	•	•		•	•			•			
10	6	913.34	•	•		•			•	•			
11	6	913.52	•	•	•	•				•			
12	7	913.70	•	•		•	•			•		•	
13	7	913.90	•	•	•	•				•		•	
14	5	913.92	•	•		•	•						
15	5	913.94	•	•		•			•				
16	7	914.04	•	•		•		•		•		•	
17	7	914.11	•	•		•	•	•		•			
18	5	914.24	•	•	•	•							
19	6	914.48	•	•		•	•	•					
20	7	914.53	•	•		•		•	•	•			
21	8	914.56	•	•		•	•		•	•		•	
22	8	914.59	•	•		•		•	•	•		•	
23	6	914.61	•	•		•	•					•	
24	7	914.62	•	•	•	•		•		•			
25	6	914.63	•	•		•		•				•	
26	6	914.73	•	•	•	•						•	
27	6	914.76	•	•		•		•	•				
28	6	914.93	•	•	•	•		•					
29	7	914.97	•	•		•	•			•			•
30	8	915.02	•	•	•	•			•	•		•	
31	7	915.19	•	•		•	•		•			•	
32	7	915.43	•	•		•	•		•	•			
33	8	915.45	•	•		•	•			•		•	•
34	7	915.47	•	•		•		•	•			•	
35	7	915.57	•	•	•	•	•			•			

Table 34. Models fit for NO_x (Bag 1): (all models include an intercept term).

Model Term	Notation	Model	
		Superset	SM2[1]
etOH	Z_e	●	●
Arom	Z_a	●	●
RVP	Z_r	---	---
T50	Z_5	●	●
T90	Z_9	---	---
etOH × etOH	ZZ_{ee}	●	×
T50 × T50	ZZ_{55}	---	---
etOH × Arom	ZZ_{ea}	●	●
etOH × RVP	ZZ_{er}	---	---
etOH × T50	ZZ_{e5}	●	×
etOH × T90	ZZ_{e9}	---	---
[1] denotes "Superset minus 2."			

Table 35. NO_x (Bag 1): Model Fitting History, starting with the 7-term Superset model.

Fit Parameters				*Test with respect to Previous Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.	*d*	Pr>χ^2
Superset	7	890.288	914.039			
SM2[2]	5	892.527	911.001	2.240	2	0.326

[1] A lower value indicates a better fit.
[2] Best fit with respect to the 11-term design model.

Table 36. NO_x (Bag 1): Coefficients and Tests of Effect for the Superset and Reduced Models, with respect to the 11-term design model.

Effect	*Full Model (Superset)*					*Reduced Model (SM2)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*
Intercept	-2.8598	0.2061	14	-13.87	<0.0001	-2.8594	0.2061	14	-13.87	<0.0001
Z_e	0.06830	0.01688	879	4.05	<0.0001	0.06750	0.01568	879	4.30	<0.0001
Z_a	0.1368	0.01333	879	10.27	<0.0001	0.1339	0.01320	879	10.15	<0.0001
Z_r	---	---	---	---	---	---	---	---	---	---
Z_5	0.04678	0.01688	879	2.77	0.0057	0.04783	0.01619	879	2.95	0.0032
Z_9	---	---	---	---	---	---	---	---	---	---
ZZ_{ee}	0.00634	0.01899	879	0.334	0.74	---	---	---	---	---
ZZ_{55}	---	---	---	---	---	---	---	---	---	---
ZZ_{ea}	-0.02343	0.01302	879	-1.80	0.072	-0.02369	0.01290	879	-1.84	0.067
ZZ_{er}	---	---	---	---	---	---	---	---	---	---
ZZ_{e5}	-0.01495	0.01857	879	-0.805	0.42	---	---	---	---	---
ZZ_{e9}	---	---	---	---	---	---	---	---	---	---
σ^2_{veh}	0.5926					0.5925				
σ^2_{ε}	0.1454					0.1458				

7.2.2.2 NO_x (Bag 1): Model Fitting starting with the 16-term Extended Model

Inclusion of 16 terms in the "full" model gives a total of 2^{16} = 65,536 possible models. However, of these, a total of only 2,964 models (4.5%) respect hierarchy. Results for these models are shown in Figure 75. The plot shows a measure of goodness-of-fit (BIC) versus the number of terms in the model, including the intercept (p). The view shows the set of models forming two major groups differing in terms of BIC, similar to those for the design model. However, in this case, the model with minimum BIC, having eight terms (p = 8) is visible even when the results are viewed at this scale.

Nonetheless, Figure 76 shows a close-up view of the same results, focusing on the major group with lowest BIC. This view shows that the major group is composed of an additional two subgroups. Of course, the model with minimum BIC is clearly visible.

To get a closer view of the subset of models with lowest BIC, Figure 77 shows a close-up focusing on the lowest edge of the lowest main group of models. The "winning" model is obvious, as the only model with BIC < 907, with no "close" ties. Nonetheless, we conservatively selected the six best-fitting candidates to construct the superset for subsequent model fitting. Table 37 shows the p, BIC and specific terms included in the 35 best fitting models out of 2,964.

Final model fitting began with the 9 terms included in the "superset model" shown in Table 38. In this case, two reduced models were fit. In the first step, two terms with high p-values (T50 and etOH×T50) were selected for removal. The p-value for the associated likelihood ratio test was well above the designated α level. Accordingly, these two terms were dropped. In the second step, four additional terms, including RVP, T90 and their interaction terms arom×T90 and RVP× T90 were tested, due to the high p-values for both linear terms. For this test, the result was highly significant, suggesting that this set of terms to contribute to model fit. The insignificant linear terms were thus retained, as well as the significant interaction terms, to maintain hierarchy. Based on these results, the first reduced model (SM2) was retained as the "best fit" model. Note that the model selected by the goodness of fit testing also has the lowest BIC, as shown in Table 39. Coefficients and Type-III tests of effect for the superset and reduced models are shown in Table 40.

Figure 75. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model.

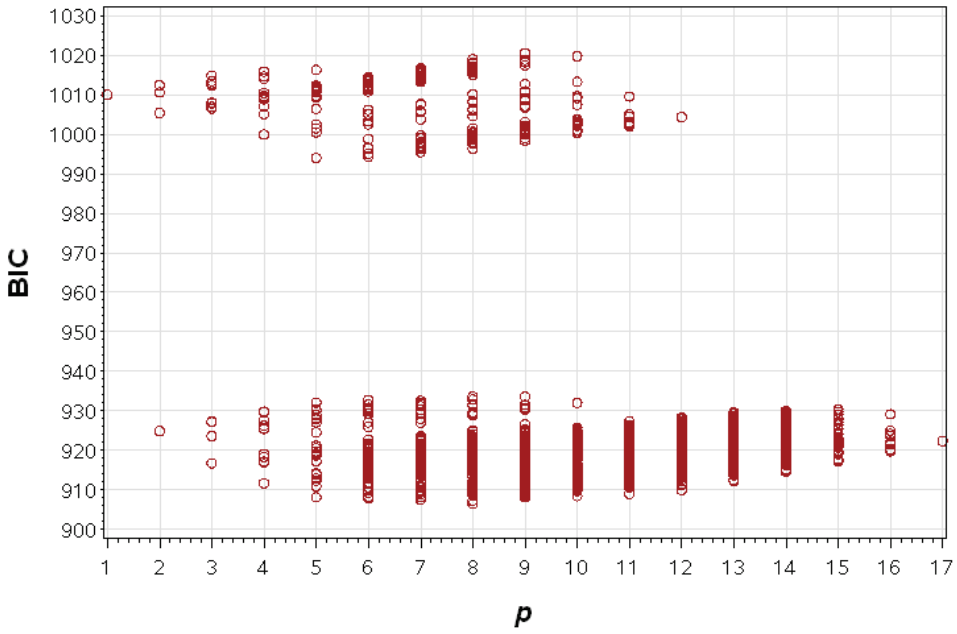


Figure 76. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model (CLOSE-UP of Figure 75).

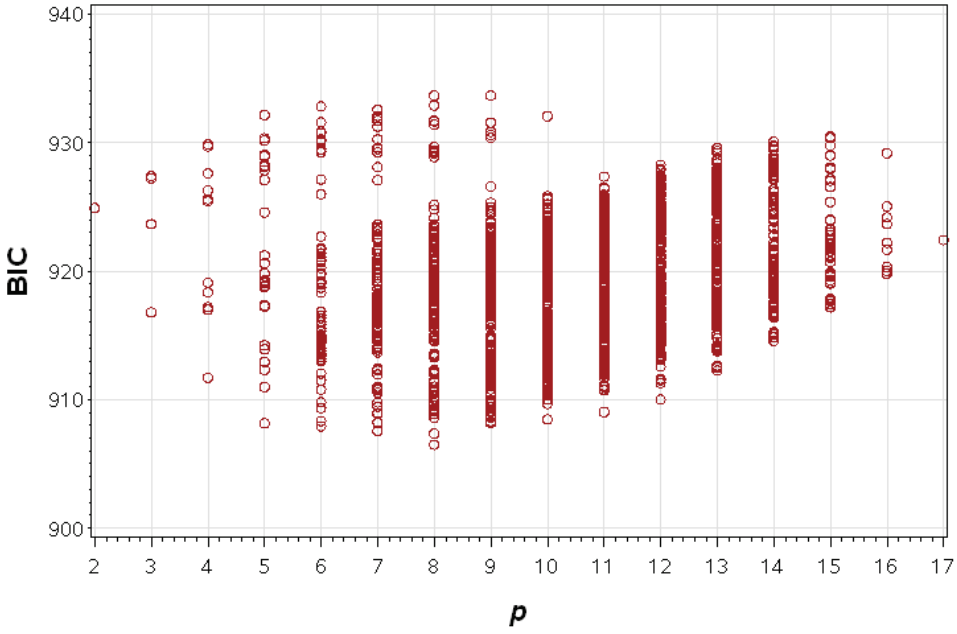


Figure 77. NO_x (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model (CLOSE-UP of Figure 76).

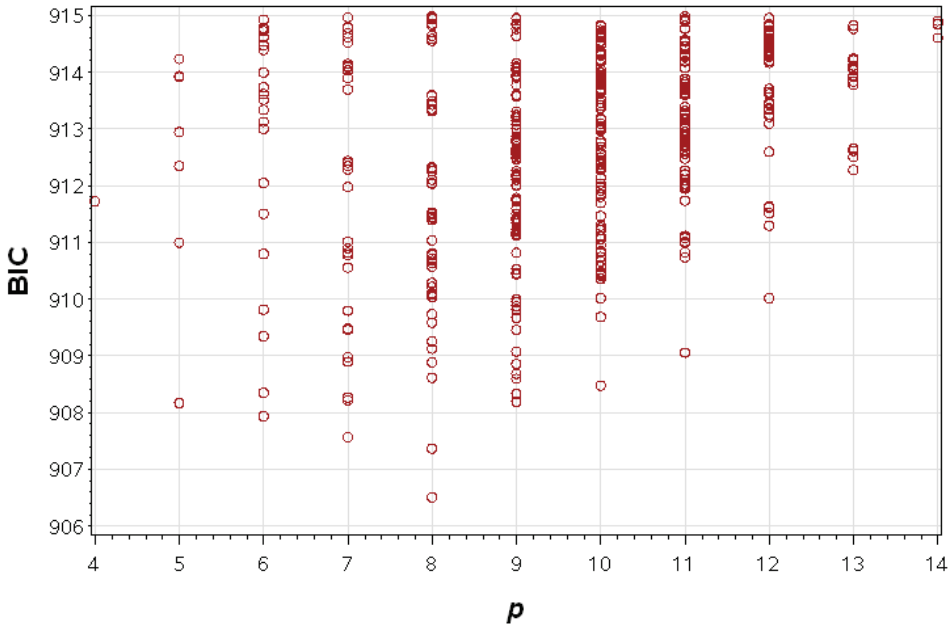


Table 37. NO_x (Bag 1): Number of terms (p), Goodness-of-fit (BIC) and terms included in the 35 best-fitting candidate models (out of a total of 2,964 possible models with hierarchy). (Terms included in models ranked 1-6 comprise the "superset" for final model-fitting).

Rank	p	BIC	Design Terms											Extended Terms				
			etOH	Arom	RVP	T50	T90	etOH × etOH	T50 × T50	etOH × Arom	etOH × RVP	etOH × T50	etOH × T90	Arom × RVP	Arom × T50	Arom × T90	T50 × T90	RVP × T90
1	8	906.51	•	•	•		•			•						•		•
2	8	907.37	•	•		•	•			•		•				•		
3	7	907.57	•	•		•	•			•						•		
4	6	907.94	•	•			•			•						•		
5	5	908.17	•	•			•									•		
6	9	908.19	•	•	•	•	•			•						•		•
7	7	908.22	•	•	•		•									•		•
8	7	908.27	•	•		•	•					•				•		
9	9	908.33	•	•	•		•	•		•						•		•
10	9	908.34	•	•	•		•			•				•		•		•
11	6	908.35	•	•		•	•									•		
12	10	908.48	•	•	•	•	•			•		•				•		•
13	9	908.60	•	•		•	•		•	•		•				•		
14	8	908.62	•	•		•	•	•		•						•		
15	9	908.69	•	•		•	•			•		•				•	•	
16	9	908.86	•	•	•		•			•			•			•		•
17	8	908.89	•	•		•	•			•						•	•	
18	7	908.90	•	•		•	•								•	•		
19	7	908.98	•	•		•	•	•								•		
20	11	909.06	•	•	•	•	•		•	•		•				•		•
21	9	909.08	•	•	•		•			•	•					•		•
22	8	909.13	•	•		•	•					•			•	•		
23	8	909.26	•	•		•	•		•			•				•		
24	6	909.35	•	•			•	•								•		
25	9	909.46	•	•	•	•	•								•	•		•
26	7	909.46	•	•	•		•			•						•		
27	7	909.49	•	•			•	•		•						•		
28	8	909.59	•	•	•		•	•								•		•
29	9	909.67	•	•	•	•	•			•		•				•		
30	10	909.69	•	•	•	•	•	•		•						•		•
31	8	909.74	•	•		•	•					•				•	•	
32	7	909.80	•	•		•	•									•	•	
33	9	909.82	•	•		•	•	•		•						•	•	
34	6	909.82	•	•	•		•									•		
35	9	909.88	•	•		•	•			•		•	•			•		

Table 38. Models fit for NO_x (Bag 1): (all models include an intercept term).

Model Term	Notation	Model		
		Superset	SM2[1]	SM6
etOH	Z_e	●	●	●
Arom	Z_a	●	●	●
RVP	Z_r	●	●	×
T50	Z_5	●	×	
T90	Z_9	●	●	×
etOH × etOH	ZZ_{ee}			
T50 × T50	ZZ_{55}			
etOH × Arom	ZZ_{ea}	●	●	●
etOH × RVP	ZZ_{er}			
etOH × T50	ZZ_{e5}	●	×	
etOH × T90	ZZ_{e9}			
Arom × RVP	ZZ_{ar}			
Arom × T50	ZZ_{a5}			
Arom × T90	ZZ_{a9}	●	●	×
T50 × T90	ZZ_{59}			
RVP × T90	ZZ_{r9}	●	●	×
[1] denotes "Superset minus 2, etc."				

Table 39. NO_x (Bag 1): Model fitting history, starting with the 9-term superset model.

Fit Parameters				*Test with respect to Previous Model*		
Model	p	-2ln$L$	BIC[1]	Dev.	d	Pr>χ^2
Superset	10	876.813	908.482			
SM2[2]	8	880.118	906.509	3.305	2	0.192
SM6	4	901.210	917.044	21.09	4	0.0003

[1] A lower value indicates a better fit.
[2] Best fit with respect to the 16-term extended model.

Table 40. NO_x (Bag 1): Coefficients and Tests of Effect for the Superset and Reduced Models, with respect to the 16-term extended model.

Effect	*Full Model (superset)*					*Reduced Model (SM2)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*
Intercept	-2.8603	0.02063	14	-13.87	<0.0001	-2.8602	0.2064	14	-13.86	<0.0001
Z_e	0.04635	0.01749	879	2.65	0.0082	0.03718	0.01303	879	2.85	0.0044
Z_a	0.1313	0.01350	879	9.73	<0.0001	0.1258	0.01316	879	9.56	<0.0001
Z_r	-0.01084	0.01469	879	-0.74	0.46	-0.01452	0.01341	879	-1.08	0.28
Z_5	0.01472	0.01935	879	0.76	0.45					
Z_9	0.004922	0.01314	879	0.37	0.71	0.005211	0.01316	879	0.40	0.69
ZZ_{ee}	---	---	---	---	---	---	---	---	---	---
ZZ_{55}	---	---	---	---	---	---	---	---	---	---
ZZ_{ea}	-0.02774	0.01292	879	-2.15	0.032	-0.02699	0.01292	879	-2.09	0.037
ZZ_{er}	---	---	---	---	---	---	---	---	---	---
ZZ_{e5}	-0.02040	0.01331	879	-1.53	0.13					
ZZ_{e9}	---	---	---	---	---	---	---	---	---	---
ZZ_{ar}	---	---	---	---	---	---	---	---	---	---
ZZ_{a5}	---	---	---	---	---	---	---	---	---	---
ZZ_{a9}	-0.04640	0.01415	879	-3.28	0.0011	-0.04990	0.01324	879	-3.77	0.00018
ZZ_{59}	---	---	---	---	---	---	---	---	---	---
ZZ_{r9}	0.02677	0.01367	879	1.96	0.051	0.03160	0.01334	879	2.37	0.018
σ^2_{veh}	0.5934					0.5939				
σ^2_{ε}	0.1432					0.1437				

7.2.2.3 CO (Bag 1): Model Fitting based on the 11-term Design Model

Inclusion of 11 terms in the "full" model gives a total of 2^{11} = 2,048 possible models. However, of these, a total of only 294 models (14.5%) respect hierarchy. As with NO_x, Figure 78 shows a measure of goodness-of-fit (BIC) versus the number of terms in the model, including the intercept (p). This view shows four major groups of models differing in terms of BIC. In this case, the model with minimum BIC is not clearly visible at this scale.

Figure 79 shows a close-up view of the same results, focusing on the major group with lowest BIC. This view shows that the major group is composed of an additional four subgroups. At this scale the model with minimal BIC, having five terms (p = 9) is apparent, although not strikingly so.

To get a closer view of the subset of models with lowest BIC, Figure 80 shows a close-up focusing on the lowest edge of the lowest main group of models. The "best" model is clearly visible, with five additional models forming a second tier. In this case, we selected the six best-fitting candidates (BIC < 324) to construct the superset for subsequent model fitting. Table 41 shows the p, BIC and specific terms included in the 35 best fitting models out of 294.

Final model fitting began with the 6 terms included in the "superset model." In this case, only several reduced models were fit. In the first trial, a single term (etOH×T90) was tested for removal based on its high p-value (0.66). As the test of fit gave a similar result, this term was dropped. A second trial tested RVP and its interaction, etOH×RVP (SM3a), as the *p*-value for the linear effect is well over the critical value and that for the interaction is close to the critical value ($p = 0.091$). With the result of this test giving a significant value, both the linear and interaction terms were retained, to maintain hierarchy. Similarly, a third trial (SM3b), tested the removal of aromatics and its interaction, etOH×Arom. As with the test for RVP, the test of fit returned a highly significant value, with the result that both terms were also retained. A fourth and final trial (SM5) tested the removal of both aromatics and RVP and their two interactions with ethanol. As in the two previous trials, this test also returned a significant result ($p = 0.00082$). Accordingly, the initial model, "SM1" was selected as the "best fit." Note that the model selected by the goodness of fit testing has the second lowest, rather than the lowest BIC, as shown in Table 41. Were the BIC used as the sole criterion, "SM3a", rather than "SM1" would have been selected as the "best fit."

Figure 78. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 11 terms in the design model.

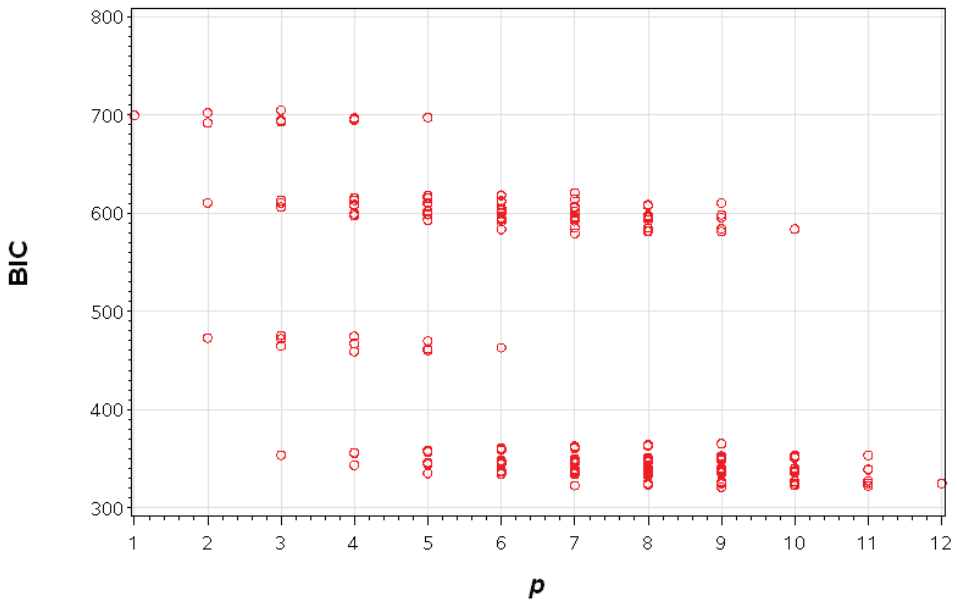


Figure 79. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (*p*) for all models respecting hierarchy, selected from the 11 terms in the design model (CLOSEUP of Figure 78).

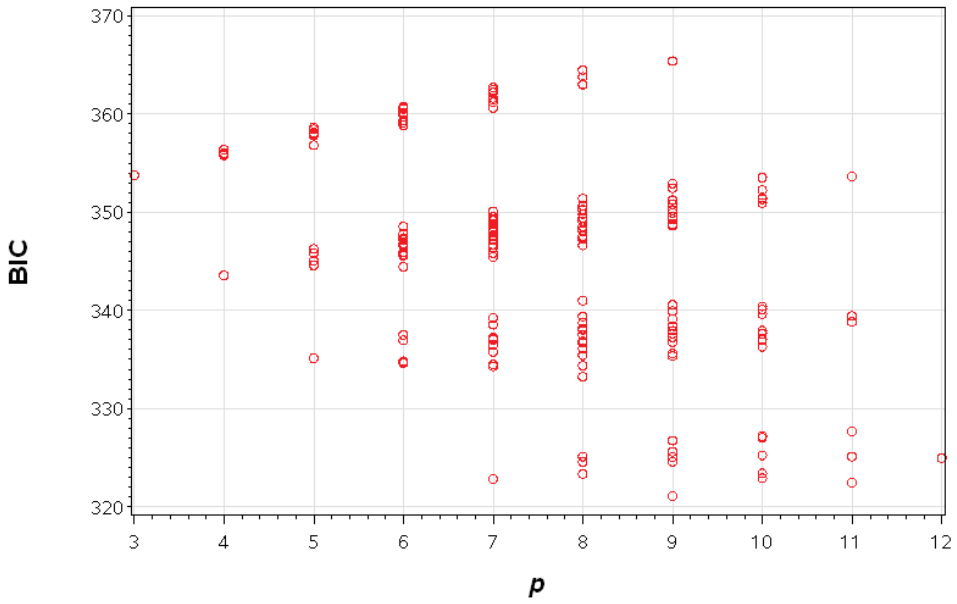


Figure 80. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (*p*) for all models respecting hierarchy, selected from the 11 terms in the design model (CLOSEUP of Figure 79).

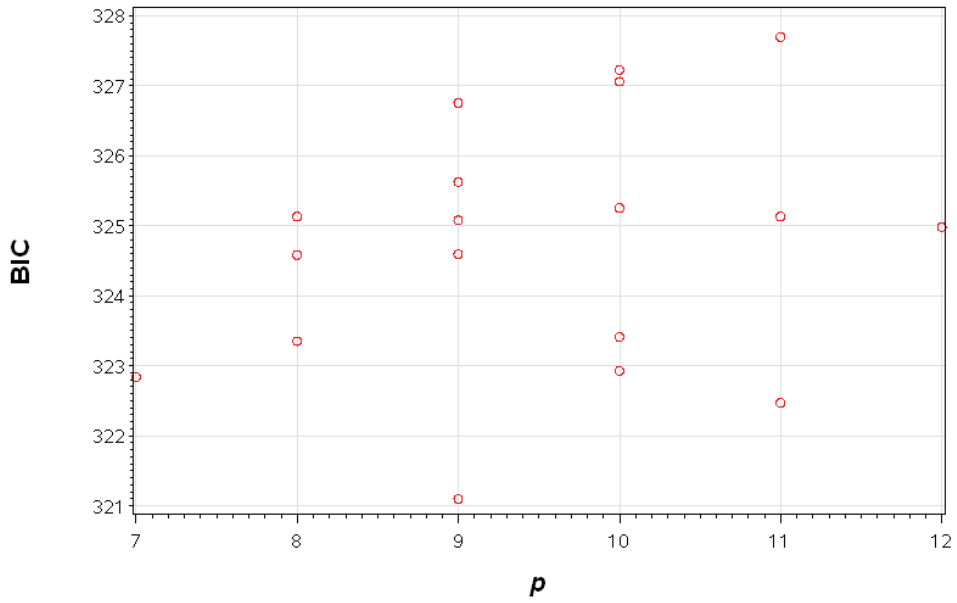


Table 41. CO (Bag 1): Number of terms (*p*), Goodness-of-fit (BIC) and terms included in the 35 best-fitting candidate models (out of a total of 294 possible models with hierarchy). (Terms included in models ranked 1-6 comprise the "superset" for final model-fitting).

Rank	*p*	BIC	Design Terms										
			etOH	Arom	RVP	T50	T90	etOH × etOH	T50 × T50	etOH × Arom	etOH × RVP	etOH × T50	etOH × T90
1	9	321.10	•	•		•	•	•	•	•		•	
2	11	322.48	•	•	•	•	•	•	•	•	•	•	
3	7	322.84	•			•	•	•	•			•	
4	10	322.93	•	•	•	•	•	•	•	•		•	
5	8	323.36	•	•		•	•	•	•			•	
6	10	323.42	•	•		•	•	•	•	•		•	•
7	8	324.59	•		•	•	•	•	•			•	
8	9	324.60	•		•	•	•	•	•		•	•	
9	12	324.99	•	•	•	•	•	•	•	•	•	•	•
10	9	325.08	•	•	•	•	•	•	•			•	
11	8	325.14	•			•	•	•	•			•	•
12	11	325.14	•	•	•	•	•	•	•	•		•	•
13	10	325.26	•	•	•	•	•	•	•		•	•	
14	9	325.63	•	•		•	•	•	•			•	•
15	9	326.76	•		•	•	•	•	•			•	•
16	10	327.06	•		•	•	•	•	•		•	•	•
17	10	327.23	•	•	•	•	•	•	•			•	•
18	11	327.70	•	•	•	•	•	•	•		•	•	•
19	8	333.26	•	•		•	•	•		•		•	
20	7	334.32	•	•		•	•			•		•	
21	8	334.38	•	•		•	•		•	•		•	
22	7	334.50	•	•		•	•		•			•	
23	7	334.54	•	•		•	•	•				•	
24	6	334.68	•	•		•	•					•	
25	6	334.76	•			•	•	•				•	
26	6	334.85	•			•	•		•			•	
27	5	335.15	•			•	•					•	
28	9	335.38	•	•	•	•	•		•	•		•	
29	8	335.45	•	•	•	•	•		•			•	
30	9	335.63	•	•		•	•	•		•		•	•
31	9	335.66	•	•	•	•	•	•		•		•	
32	7	335.79	•		•	•	•		•			•	
33	8	336.14	•	•	•	•	•			•		•	
34	10	336.32	•	•	•	•	•	•		•	•	•	
35	7	336.50	•	•	•	•	•					•	

Table 42. Models fit for CO (Bag 1): (all models include an intercept term).

Model Term	Notation	Model				
		Superset	SM1[1]	SM3a	SM3b[2]	SM5[3]
etOH	Z_e	●	●	●	●	●
Arom	Z_a	●	●	●	×	×
RVP	Z_r	●	●	×	●	×
T50	Z_5	●	●	●	●	●
T90	Z_9	●	●	●	●	●
etOH × etOH	ZZ_{ee}	●	●	●	●	●
T50 × T50	ZZ_{55}	●	●	●	●	●
etOH × Arom	ZZ_{ea}	●	●	●	×	×
etOH × RVP	ZZ_{er}	●	●	×	●	×
etOH × T50	ZZ_{e5}	●	●	●	●	●
etOH × T90	ZZ_{e9}	●	×			

[1] denotes "Superset minus 1."
[2] Not nested within SM3a; test with respect to SM1.
[3] Not nested within SM3b or SM3a; test with respect to SM1.

Table 43. CO (Bag 1): Model Fitting History, starting with the 11-term Superset model.

Fit Parameters				*Test with respect to Previous Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.	*d*	Pr>χ^2
Superset	12	287.074	324.986			
SM1[2]	11	287.271	322.475	0.1970	1	0.657
SM3a	9	291.311	321.010	4.041	2	0.0444
SM3b[3]	9	294.809	324.597	7.538	2	0.0060
SM5[3]	7	298.471	322.843	11.200	4	0.00082

[1] A lower value indicates a better fit.
[2] Best fit with respect to the 11-term design model.
[3] Test with respect to SM1.

Table 44. CO (Bag 1): Coefficients and Tests of Effect for the Superset and Reduced Models, with respect to the 11-term design model.

Effect	*Full Model (Superset)*					*Reduced Model (SM1)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*
Intercept	1.3467	0.1618	15	8.32	<0.0001	1.3466	0.1619	15	8.32	<0.0001
Z_e	-0.1051	0.01305	941	-8.06	<0.0001	-0.1049	0.01304	941	-8.05	<0.0001
Z_a	-0.01248	0.009092	941	-1.37	0.170	-0.01242	0.009092	941	-1.37	0.172
Z_r	-0.0081	0.01038	941	-0.780	0.436	-0.00762	0.01033	941	-0.737	0.461
Z_5	-0.03285	0.01310	941	-2.51	0.0123	-0.03273	0.01310	941	-2.50	0.0126
Z_9	-0.1565	0.009095	941	-17.20	<0.0001	-0.1571	0.008992	941	-17.47	<0.0001
ZZ_{ee}	0.07290	0.01751	941	4.16	<0.0001	0.07304	0.01750	941	4.17	<0.0001
ZZ_{55}	0.05362	0.01311	941	4.09	<0.0001	0.05358	0.01311	941	4.09	<0.0001
ZZ_{ea}	0.02074	0.008894	941	2.33	0.0199	0.02086	0.008891	941	2.35	0.0192
ZZ_{er}	0.01535	0.009073	941	1.69	0.0911	0.01596	0.008967	941	1.78	0.0753
ZZ_{e5}	0.1062	0.01879	941	5.65	<0.0001	0.1064	0.01878	941	5.67	<0.0001
ZZ_{e9}	0.003963	0.008928	941	0.444	0.657			941		
σ^2_{veh}	0.3917					0.3920				
σ^2_{ε}	0.07212					0.07214				

7.2.2.4 CO (Bag 1): Model Fitting based on the 16-term Extended Model.

Inclusion of 16 terms in the "full" model gives a total of 2^{16} = 65,536 possible models. However, of these, a total of only 2,964 models (4.5%) respect hierarchy. Results for these models are shown in Figure 81. The plot shows a measure of goodness-of-fit (BIC) versus the number of terms in the model, including the intercept (p). This view shows four major groups of models differing in terms of BIC, similar to those for the design model. As with the design model, the model with minimum BIC is not clearly visible.

Figure 82 shows a close-up view of the same results, focusing on the major group with lowest BIC. This view shows that the major group is composed of five additional subgroups, differing greatly in size. Within groups there is a general trend of decline in fit with increasing numbers of terms. However, the model with minimal BIC in each group tends to have more terms as the BIC range for each group declines. At this scale the model with minimal BIC, having thirteen terms (p = 13) is obvious, as occupying the low end of a trend composed of models having 13-17 terms. The extended model (p = 17) as also a member of this group. The BIC for the extended model is substantially higher than that for the "best fit," but it is also clear that the full extended model has a better fit than the majority of models with fewer terms.

To get a closer view of the subset of models with lowest BIC, Figure 83 shows a close-up focusing on the lowest edge of the lowest main group of models. The "winning" model is clearly visible (BIC=297), with two full units between the "best fit" and the 2nd ranked model (BIC~299). We selected the five best-fitting candidates (BIC < 300) to construct the superset for

subsequent model fitting. Table 45 shows the *p*, BIC and specific terms included in the 35 best top-ranking models.

Final model fitting began with all 16 terms included in the "superset model." Three reduced models were fit. In the first trial, four terms were tested for removal based on high *p*-values for effect (0.24-0.83). As expected, the test of fit indicated that these terms did not contribute significantly to fit, and they were accordingly dropped. In the initial model, however, three linear terms, aromatics, RVP and T50 had high p-values. Thus, for thoroughness, a second trial tested aromatics and four interaction terms. Not unexpectedly, the result of this test was highly significant, and the aromatics linear term was retained on the strength of its interactions. Similarly, the third trial tested the removal of T50, its quadratic term and two interactions, with the result that the T50 linear term was also retained on the strength of its 2nd order terms. On this basis, the initial model, "SM4" was selected as the "best fit," as shown in Table 47. Note that the model selected by the goodness of fit testing also has the lowest BIC of all possible models, as shown in Table 45. For the superset and "best fit" models, coefficients and tests of effect are shown in Table 48.

Figure 81. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model.

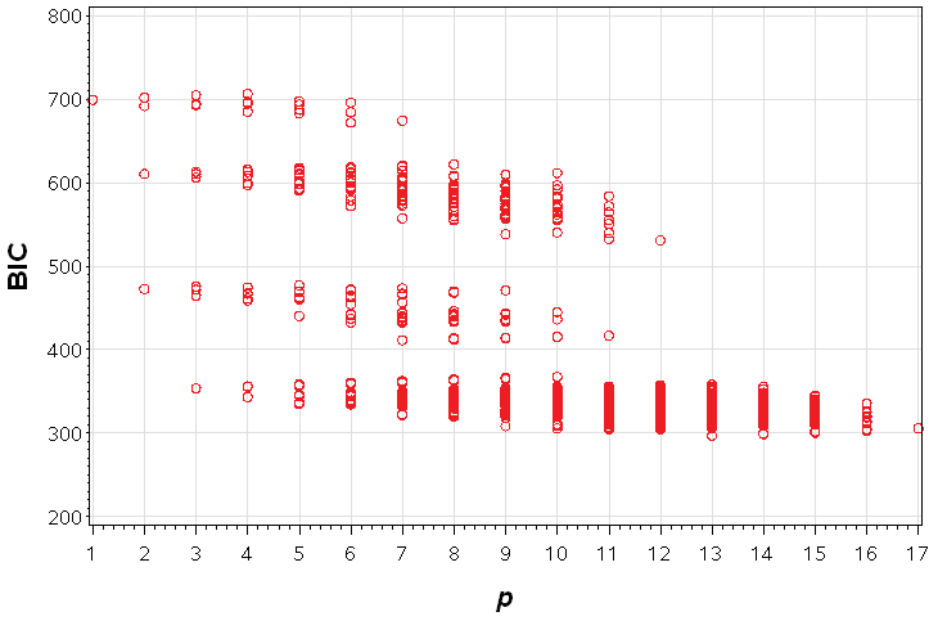


Figure 82. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model (CLOSE-UP of Figure 81).

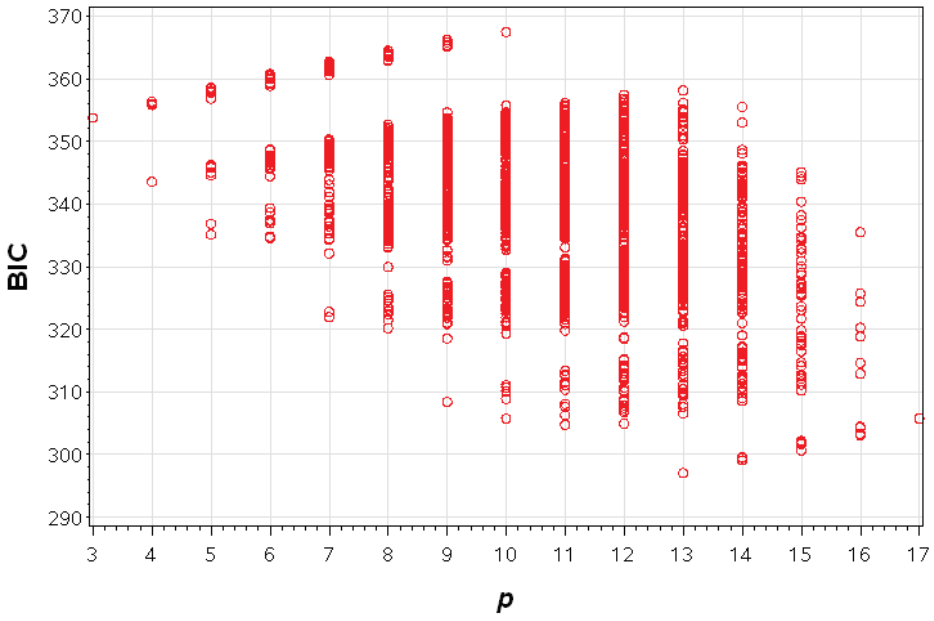


Figure 83. CO (Bag 1): Bayesian Information Criterion (BIC) vs. Number of terms (p) for all models respecting hierarchy, selected from the 16 terms in the extended model (CLOSE-UP of Figure 82).

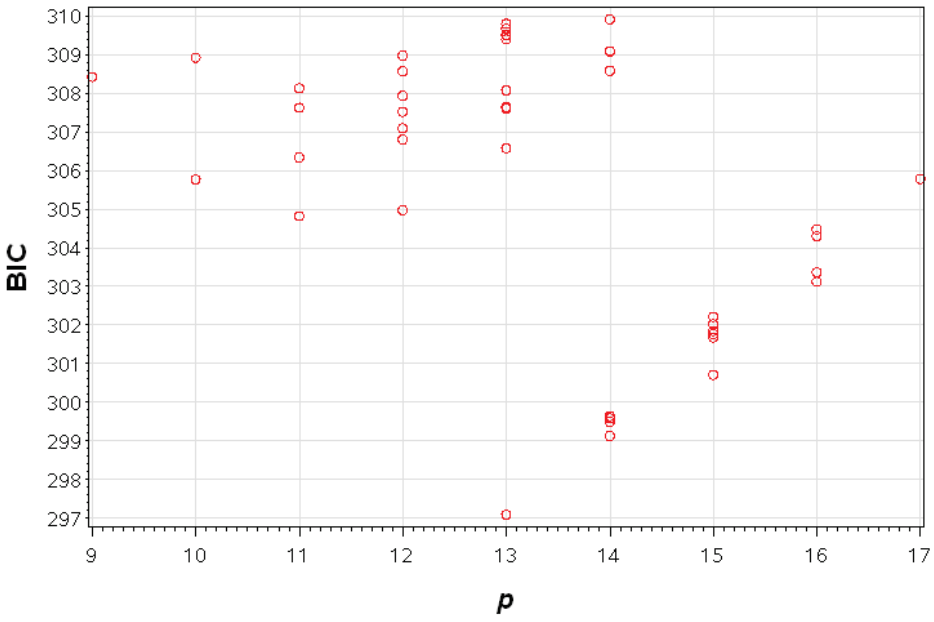


Table 45. CO (Bag 1): Number of terms (*p*), Goodness-of-fit (BIC) and terms included in the 35 best-fitting candidate models (out of a total of 2,964 possible models with hierarchy). (Terms included in models ranked 1-5 comprise the "superset" for final model-fitting).

Rank	*p*	BIC	Design Terms											Extended Terms				
			etOH	Arom	RVP	T50	T90	etOH × etOH	T50 × T50	etOH × Arom	etOH × RVP	etOH × T50	etOH × T90	Arom × RVP	Arom × T50	Arom × T90	T50 × T90	RVP × T90
1	13	297.09	•	•	•	•	•	•	•	•		•		•	•	•		
2	14	299.13	•	•	•	•	•	•	•	•		•	•	•	•	•		
3	14	299.50	•	•	•	•	•	•	•	•	•	•		•	•	•		
4	14	299.59	•	•	•	•	•	•	•	•		•		•	•	•	•	
5	14	299.65	•	•	•	•	•	•	•	•		•		•	•	•		•
6	15	300.72	•	•	•	•	•	•	•	•		•	•	•	•	•	•	
7	15	301.69	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
8	15	301.77	•	•	•	•	•	•	•	•		•	•	•	•	•		•
9	15	301.85	•	•	•	•	•	•	•	•	•	•		•	•	•	•	
10	15	302.03	•	•	•	•	•	•	•	•	•	•		•	•	•		•
11	15	302.22	•	•	•	•	•	•	•	•		•		•	•	•	•	•
12	16	303.13	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
13	16	303.37	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•
14	16	304.31	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•
15	16	304.49	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•
16	11	304.83	•	•		•	•	•	•	•		•			•	•		
17	12	304.98	•	•		•	•	•	•	•		•	•		•	•		
18	10	305.77	•	•		•	•	•	•	•		•				•		
19	17	305.79	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
20	11	306.35	•	•		•	•	•	•	•		•	•			•		
21	13	306.59	•	•	•	•	•	•	•	•	•	•		•		•		
22	12	306.81	•	•	•	•	•	•	•	•		•		•		•		
23	12	307.10	•	•		•	•	•	•	•		•			•	•	•	
24	12	307.53	•	•	•	•	•	•	•	•		•			•	•		
25	13	307.62	•	•		•	•	•	•	•		•	•		•	•	•	
26	11	307.63	•	•	•	•	•	•	•	•		•				•		
27	13	307.65	•	•	•	•	•	•	•	•		•	•		•	•		
28	12	307.95	•	•	•	•	•	•	•	•		•	•			•		
29	13	308.09	•	•	•	•	•	•	•	•		•	•	•		•		
30	11	308.14	•	•		•	•	•	•	•		•				•	•	
31	9	308.43	•	•		•	•	•	•			•				•		
32	12	308.58	•	•	•	•	•	•	•	•	•	•				•		
33	14	308.60	•	•	•	•	•	•	•	•	•	•	•	•		•		
34	10	308.93	•	•		•	•	•	•			•	•			•		
35	12	308.98	•	•		•	•	•	•	•		•	•			•	•	

Table 46. Models fit for CO (Bag 1): (all models include an intercept term).

Model Term	Notation	Model			
		Superset	SM4[1]	SM9	SM8[2]
etOH	Z_e	●	●	●	●
Arom	Z_a	●	●	×	●
RVP	Z_r	●	●	●	●
T50	Z_5	●	●	●	×
T90	Z_9	●	●	●	●
etOH × etOH	ZZ_{ee}	●	●	●	●
T50 × T50	ZZ_{55}	●	●	●	×
etOH × Arom	ZZ_{ea}	●	●	×	●
etOH × RVP	ZZ_{er}	●	×		
etOH × T50	ZZ_{e5}	●	●	●	×
etOH × T90	ZZ_{e9}	●	×		
Arom × RVP	ZZ_{ar}	●	●	×	●
Arom × T50	ZZ_{a5}	●	●	×	×
Arom × T90	ZZ_{a9}	●	●	×	●
T50 × T90	ZZ_{59}	●	×		
RVP × T90	ZZ_{r9}	●	×		

[1] Denotes "Superset minus 4, etc."
[2] Formed by dropping aromatics and its interactions.
[3] Formed by dropping T50 and its remaining 2nd order terms; Nested within SM4, not SM9.

Table 47. CO (Bag 1): Model Fitting History, starting with the 16-term Superset model.

Fit Parameters				*Test with respect to Previous Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.	*d*	Pr>χ^2
Superset	17	254.340	305.793			
SM4[2]	13	256.473	297.094	2.134	4	0.711
SM9	8	297.508	324.588	41.034	5	0.000000092
SM8[3]	9	302.800	332.589	46.327	4	0.0000000021

[1] A lower value indicates a better fit.
[2] Best fit with respect to the 16-term design model.
[3] Test with respect to SM4.

Table 48. CO (Bag 1): Coefficients and Tests of Effect for the Superset and Reduced Models, with respect to the 16-term extended model.

Effect	*Full Model (superset)*					*Reduced Model (SM1)*				
	Estimate	Std. Err.	d.f.	*t* -value	Pr>*t*	Estimate	Std. Err.	d.f.	*t* -value	Pr>*t*
Intercept	1.3476	0.1620	15	8.32	0.000001	1.3472	0.1620	15	8.32	0.000001
Z_e	-0.07050	0.01548	941	-4.55	0.000006	-0.06967	0.01462	941	-4.76	0.000002
Z_a	-0.007545	0.009715	941	-0.777	0.44	-0.01156	0.008946	941	-1.29	0.20
Z_r	0.01763	0.01307	941	1.35	0.18	0.02063	0.01273	941	1.62	0.11
Z_5	0.02091	0.01725	941	1.21	0.23	0.02160	0.01652	941	1.31	0.19
Z_9	-0.1477	0.01001	941	-14.8	0.000000	-0.1469	0.009473	941	-15.5	0.000000
ZZ_{ee}	0.08386	0.01805	941	4.65	0.000004	0.08535	0.01725	941	4.95	0.000001
ZZ_{55}	0.07051	0.01473	941	4.79	0.000002	0.07222	0.01382	941	5.23	0.000000
ZZ_{ea}	0.06085	0.01460	941	4.17	0.000034	0.06244	0.01410	941	4.43	0.000011
ZZ_{er}	0.005291	0.009901	941	0.534	0.59					
ZZ_{e5}	0.1095	0.01906	941	5.74	0.000000	0.1117	0.01857	941	6.02	0.000000
ZZ_{e9}	0.01387	0.01171	941	1.18	0.24					
ZZ_{ar}	0.05398	0.01582	941	3.41	0.00067	0.05370	0.01476	941	3.64	0.00029
ZZ_{a5}	0.05549	0.01763	941	3.15	0.0017	0.05859	0.01657	941	3.54	0.00042
ZZ_{a9}	0.03949	0.009900	941	3.99	0.000072	0.03861	0.009502	941	4.06	0.000052
ZZ_{59}	0.01456	0.01317	941	1.11	0.27					
ZZ_{r9}	0.00238	0.01119	941	0.21	0.83					
σ^2_{veh}	0.3924					0.3926				
σ^2_{ε}	0.06966					0.06981				

7.2.2.5 PM (Bag 1): Model fitting based on the 11-term Design Model (Tobit Regression)

In the three tables below, we present the results of a Tobit model-fitting process, using Bag 1 PM as an example. Each of the reduced models is identified by the number of terms having been removed from the full model, i.e., "FM5" is read as "*Full minus 5*," etc.

Table 49 shows the set of reduced models fit, starting with a full model including the 11 terms in the design model. Five reduced models were fit, removing two to six terms, with each successive model nested within its predecessor. For each, the terms removed from the previous model are indicated by the "×" symbol.

Table 50 illustrates the model-fitting results for this process. The left-most block in the table "*Fit Parameters*" shows the number of terms in each model *p-1*, excluding the intercept. It then shows the -2 log likelihood statistics (-2ln*L*) for each model.

The rightmost block shows the results of goodness of fit tests for each model and its preceding reduced model, i.e., FM6 is tested against FM5, and FM7 against FM6, etc. The first column shows the "deviation," which is equal to the test statistic χ^2_{test} as shown in Equation 13, between each reduced model and its predecessor. The second column shows the difference in the numbers of terms between the two models *d*, and the third shows the p-value for the χ^2 test with *d* degrees of freedom. A value greater than the assigned level (0.10) indicates an insignificant result, i.e., we retain the null hypothesis that there is no significant difference in fit between the two models. Stated differently, the additional terms in the full model do not give a corresponding improvement in fit.

The tests are continued until either the test result is significant ($p < 0.10$), or the tests of effect for the remaining terms are all significant (also $p < 0.10$). In the test between FM5 and FM4, the result is almost significant at the 0.10 level, but we nonetheless proceed to drop the etOH×T90 term based on our established criterion. In the final step, we test the removal of the RVP term (FM6 vs. FM5); the result is insignificant ($p = 0.625$), suggesting that this term does not improve the fit. Thus, we drop RVP and accept the resulting five-term model (FM6) as the "best fit." The linear term for T50 is insignificant, but we retain it due to the significance of its quadratic term.

Table 51 shows coefficients, standard errors and tests of effect for the best-fit model, with results for the full model included for reference. Note that the values of coefficients for terms in common for both models are similar, but not identical. The standard errors, on the other hand, are generally somewhat lower in the reduced model than for corresponding terms in the full model, indicating that the terms are estimated with greater precision in the reduced model. The improvement in precision corresponds to a reduction in the uncertainty (standard error) relative to the mean (coefficient), which is also reflected in substantial increases in the test statistics and corresponding reductions in the *p*-values (higher significance). For example, the ethanol coefficient in the full model is fairly large and its *p*-value quite low. However, in the reduced model, its standard error drops by 20%, and its *p*-value decreases by an order of magnitude or more.

Table 49. Models fit for PM (Bag 1): (Grand Intercept not fit by Tobit models).

Model Term	Notation	Model					
		Full	FM2[1]	FM3	FM4	FM5	FM6
etOH	Z_e	●	●	●	●	●	●
Arom	Z_a	●	●	●	●	●	●
RVP	Z_r	●	●	●	●	●	×
T50	Z_5	●	●	●	●	●	●
T90	Z_9	●	●	●	●	●	●
etOH × etOH	ZZ_{ee}	●	×				
T50 × T50	ZZ_{55}	●	●	●	●	●	●
etOH × Arom	ZZ_{ea}	●	●	●	×		
etOH × RVP	ZZ_{er}	●	●	×			
etOH × T50	ZZ_{e5}	●	×				
etOH × T90	ZZ_{e9}	●	●	●	●	×	

[1] Denotes "Full minus 2," etc.

Table 50. PM (Bag 1): Model fitting history, starting with the 11-term design model.

Fit Parameters			*Test with respect to Previous Model*		
Model	p-1[1]	-2ln$L$	Dev.	d	Pr>χ^2
Full Model	11	2789.671			
FM2	9	2790.118	0.447	2	0.800
FM3	8	2791.074	0.956	1	0.328
FM4	7	2793.414	2.34	1	0.126
FM5	6	2796.069	2.655	1	0.103
FM6[2]	5	2796.308	0.239	1	0.625

[1] Total number of terms excludes the intercept, as the Tobit models do not fit intercepts.
[2] Best fit with respect to the 11-term design model.

Table 51. PM (Bag 1): Coefficients and Tests of Effect for the Full and Reduced Models, with respect to the 11-term design model.

Effect	*Full Model*					*Reduced Model (FM6)*				
	Estimate	Std. Err.	d.f.	χ^2- value	Pr>χ^2	Estimate	Std. Err.	d.f.	χ^2- value	Pr>χ^2
Intercept[1]						0.6559				
Z_e	0.1365	0.05030	1	7.35	0.0067	0.1582	0.04130	1	14.7	0.00010
Z_a	0.3840	0.03510	1	119.96	<.0001	0.3833	0.03480	1	121	<.0001
Z_r	-0.0227	0.04000	1	0.32	0.57					
Z_5	0.0338	0.05050	1	0.45	0.50	0.0550	0.04310	1	1.63	0.20
Z_9	0.2965	0.03510	1	71.48	<.0001	0.2923	0.03440	1	72.2	<.0001
ZZ_{ee}	-0.0401	0.06750	1	0.35	0.55					
ZZ_{55}	0.0700	0.05050	1	1.92	0.166	0.0935	0.03420	1	7.46	0.0063
ZZ_{ea}	0.0508	0.03430	1	2.19	0.139					
ZZ_{er}	0.0295	0.03500	1	0.71	0.40					
ZZ_{e5}	-0.0482	0.07230	1	0.44	0.51					
ZZ_{e9}	0.0503	0.03440	1	2.14	0.14					
σ^2_{veh} [1]						0.4251				
σ^2_{ε}	1.0321					1.0359				

[1] Not fit by Tobit model; calculated manually from individual vehicle intercepts.

7.2.2.6 PM (Bag 1): Model Fitting based on the 16-term Extended Model (Tobit Regression)

Table 52 shows the set of reduced models fit, starting with a full model including the 17 terms in the extended model. Seven reduced models were fit, removing three to nine terms, with each successive model nested within its predecessor. For each, the terms removed from the previous model are indicated by the "×" symbol. The model fitting history is shown in Table 53. Coefficients and tests of effect are shown in Table 55. The final model (FM10) contains seven terms, six from the design model, and one from the extended model (arom×T90). Its coefficient is positive, and relatively small compared to its two linear effects, which are large and positive. Thus, this term acts as a small reinforcement interaction between aromatics and T90.

Table 52. Models fit for PM (Bag 1): (Grand Intercept not fit by Tobit models).

Model Term	Notation	Model							
		Full	FM3[1]	FM5	FM6	FM7	FM8	FM9	FM10
etOH	Z_e	●	●	●	●	●	●	●	●
Arom	Z_a	●	●	●	●	●	●	●	●
RVP	Z_r	●	●	●	●	●	●	●	×
T50	Z_5	●	●	●	●	●	●	●	●
T90	Z_9	●	●	●	●	●	●	●	●
etOH × etOH	ZZ_{ee}	●	×						
T50 × T50	ZZ_{55}	●	●	●	●	●	●	●	●
etOH × Arom	ZZ_{ea}	●	●	●	●	●	●	×	
etOH × RVP	ZZ_{er}	●	×						
etOH × T50	ZZ_{e5}	●	●	×					
etOH × T90	ZZ_{e9}	●	●	●	●	●	●	●	●
T90 × T90	ZZ_{99}	●	●	●	●	×			
Arom × RVP	ZZ_{ar}	●	●	●	×	●	●	●	●
Arom × T50	ZZ_{a5}	●	×						
Arom × T90	ZZ_{a9}	●	●	●	●	●	●	●	●
T50 × T90	ZZ_{59}	●	●	×					
RVP × T90	ZZ_{r9}	●	●	●	●	●	×		

[1] Indicates "Full minus 3," etc.

Table 53. PM (Bag 1): Model fitting history, starting with the 17-term extended model.

Fit Parameters			*Test with respect to Previous Model*		
Model	p-1[1]	-2ln$L$	Dev.	d	Pr>χ^2
Full Model	17	2804.455			
FM3	14	2804.778	0.323	3	0.956
FM5	12	2805.586	0.808	2	0.668
FM6	11	2806.743	1.157	1	0.282
FM7	10	2809.264	2.521	1	0.112
FM8	9	2811.736	2.472	1	0.116
FM9	8	2813.991	2.255	1	0.133
FM10[2]	7	2814.388	0.397	1	0.529

[1] Total number of terms excludes the intercept, as the Tobit models do not fit intercepts.
[2] Best fit with respect to the 17-term extended model.

Table 54. PM (Bag 1): Coefficients and Tests of Effect for the Full and Reduced Models, with respect to the 17-term extended model.

Effect	*Full Model*					*Reduced Model (FM10)*				
	Estimate	Std. Err.	d.f.	χ^2- value	Pr>χ^2	Estimate	Std. Err.	d.f.	χ^2- value	Pr>χ^2
Intercept[1]										
Z_e	0.1512	0.0628	1	5.80	0.016	0.1807	0.0425	1	18.0	<.0001
Z_a	0.4355	0.0428	1	104	<.0001	0.3792	0.0351	1	116.6	<.0001
Z_r	-0.0619	0.0585	1	1.12	0.2897					
Z_5	0.0047	0.0837	1	0.00000	0.96	0.1004	0.0470	1	4.6	0.0326
Z_9	0.3494	0.0459	1	58.0	<.0001	0.3034	0.0353	1	74.0	<.0001
ZZ_{ee}	-0.0213	0.0712	1	0.090	0.7653					
ZZ_{55}	0.0790	0.0582	1	1.84	0.17	0.0806	0.0346	1	5.4	0.0198
ZZ_{ea}	0.0682	0.0579	1	1.39	0.24					
ZZ_{er}	0.0119	0.0392	1	0.090	0.76					
ZZ_{e5}	-0.0452	0.0752	1	0.36	0.55					
ZZ_{e9}	0.1018	0.0478	1	4.53	0.033	0.0647	0.0349	1	3.4	0.0636
ZZ_{ar}	-0.0602	0.0731	1	0.68	0.41					
ZZ_{a5}	-0.0340	0.0788	1	0.19	0.67					
ZZ_{a9}	0.0551	0.0399	1	1.91	0.17	0.0665	0.0380	1	3.1	0.0804
ZZ_{59}	0.0389	0.0520	1	0.560	0.45					
ZZ_{r9}	-0.0558	0.0452	1	1.53	0.22					
σ^2_{veh} [1]										
σ^2_{ε}						1.0443				

[1] Not fit by Tobit model.

7.2.3 Coefficients for Reduced Models

The tables below present sets of coefficients for reduced models. The tables present models representing cold-start emissions (Bag 1), hot-running emissions (Bag 2) and hot-start emissions (Bag 3). Parameters for each model are shown in Table 55, including numbers of observations, numbers of vehicles included, and numbers of observations censored, missing or removed. In addition, the model indicates the model type. Sets of model coefficients for THC, NMOG, NMHC, CH_4, CO, NO_x and PM are presented in Table 56-Table 69. In reviewing the coefficients it is important to remember that the values presented represent "standardized coefficients," i.e., coefficients relating the change in the natural logarithm of emissions to standardized fuel properties (see 2.3.1, page 28). Generally, the coefficient represents the change in the logarithm of emissions associated with a change in the fuel property of one standard deviation, calculated with respect to the fuel matrix used in this program.

At this point it is important to note that the results reported below differ from those described above in 5.3, as well as from those reported for the DOE analysis[9]. As mentioned, we excluded the T90×T90 term, whereas the DOE analysis fit reduced models with and without it, and reported that differences in candidate models resulted in some cases. With respect to influential observations, the DOE analysis removed the same outlying measurements, based on the same evaluation. However, the model fitting reported in the DOE analysis does not reflect the influence analyses and subsequent evaluation reported in section 5.5 and Chapter 6. Consequently, the DOE analyses retain all 15 vehicles for all analyses, whereas we have dropped selected vehicles for specific models as described above. This difference in approaches does lead to differences in some model fits. Finally, when ranking candidate models, the DOE analysis applies the Mallow's C_p criterion, whereas we have applied the Bayesian Information Criterion (BIC). Ranking by these two criteria is expected to give similar but not necessarily identical results.

Table 55. Design and Modeling Parameters for Final Model-fitting.

Compound	Bag	n_{obs}	n_{veh}	$n_{censored}$	$n_{missing}$	$n_{removed}$	Model Type
CO	1	956	15	0	0	0	Mixed
	2	956	15	0	0	0	Mixed
	3	956	15	0	0	0	Mixed
NO_x	1	893	14	0	0	0	Mixed
	2	893	14	0	0	0	Mixed
	3	931	15	25	0	0	Tobit
PM	1	908	15	45	2	1	Tobit
	2	906	15	47	2	1	Tobit
	3	873	15	82	0	1	Tobit
THC	1	956	15	0	0	0	Mixed
	2	832	13	0	0	0	Mixed
	3	954	15	2	0	0	Mixed
NMOG	1	956	15	0	0	0	Mixed
	2	832	13	0	0	0	Mixed
	3	837	15	119	0	0	Tobit
NMHC	1	956	15	0	0	0	Mixed
	2	832	13	0	0	0	Mixed
	3	828	15	128	0	0	Tobit
CH_4	1	956	15	0	0	0	Mixed
	2	956	15	0	0	0	Mixed
	3	885	14	0	0	0	Mixed

Table 56. THC: Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3
Intercept	Intercept	-0.8664	-4.6533	-4.2300
etOH	Z_e	0.0548	0.0327	0.0079
Arom	Z_a	0.0676	-0.0195	-0.0612
RVP	Z_r	-0.0445	-0.0355	-0.0142
T50	Z_5	0.1288	0.0501	0.0360
T90	Z_9	0.0183	0.0514	0.0490
etOH × etOH	ZZ_{ee}	0.0436	---	
T50 × T50	ZZ_{55}	0.0736	0.0337	
etOH × Arom	ZZ_{ea}	0.0179		---
etOH × RVP	ZZ_{er}			0.0167
etOH × T50	ZZ_{e5}	0.0445	---	-0.0313
etOH × T90	ZZ_{e9}	0.0214		---
σ^2_{veh}		0.1325	0.8384	0.8860
σ^2_{ε}		0.06872	0.06717	0.09218

[1] Fit excluding the Odyssey and Sienna.

Table 57. THC: Reduced Models, based on the 16-term extended Model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3
Intercept	Intercept	-0.8658	-4.6543	-4.2300
etOH	Z_e	0.06793	0.03470	0.0079
Arom	Z_a	0.08344	-0.01968	-0.0612
RVP	Z_r	-0.04669	-0.02641	-0.0142
T50	Z_5	0.1490	0.05122	0.0360
T90	Z_9	0.01434	0.06077	0.0490
etOH × etOH	ZZ_{ee}	0.04065	---	
T50 × T50	ZZ_{55}	0.07555	0.03392	
etOH × Arom	ZZ_{ea}	0.02297	0.02892	---
etOH × T50	ZZ_{er}		-0.02184	0.0167
etOH × T90	ZZ_{e5}	0.03426	---	-0.0313
etOH × RVP	ZZ_{e9}	0.04782	---	---
Arom × RVP	ZZ_{ar}	0.02553		
Arom × T50	ZZ_{a5}		0.03112	
Arom × T90	ZZ_{a9}	0.02260		
T50 × T90	ZZ_{59}	0.05004		
RVP × T90	ZZ_{r9}			
σ^2_{veh}		0.1317	0.8395	0.8860
σ^2_{ε}		0.06687	0.6646	0.09218

[1] Fit excluding the Odyssey and Sienna.

Table 58. NMOG: Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3
Intercept	Intercept	-0.95209	-5.2360	-5.8449[2]
etOH	Z_e	0.080186	0.02673	0.0339
Arom	Z_a	0.087823	0.03634	
RVP	Z_r	-0.04224	-0.04786	-0.0572
T50	Z_5	0.134524	0.04915	0.0783
T90	Z_9		0.07252	0.1467
etOH × etOH	ZZ_{ee}	0.044316	---	
T50 × T50	ZZ_{55}	0.075786	0.05349	
etOH × Arom	ZZ_{ea}	0.016927	0.02171	
etOH × RVP	ZZ_{er}			0.0707
etOH × T50	ZZ_{e5}	0.04653	0.02586	-0.0728
etOH × T90	ZZ_{e9}			
σ^2_{veh}			0.8502	
σ^2_{ε}			0.1310	

[1] Fit excluding the Odyssey and Sienna. See 6.1.3.
[2] Not fit by the Tobit model; calculated manually from individual vehicle intercepts.

Table 59. NMOG: Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3[3]
Intercept	Intercept	-0.9513	-5.2369	-5.8449[2]
etOH	Z_e	0.0927	0.02947	0.0339
Arom	Z_a	0.1051	0.03540	
RVP	Z_r	-0.0483	-0.03415	-0.0572
T50	Z_5	0.1541	0.05338	0.0783
T90	Z_9	0.0112	0.08637	0.1467
etOH × etOH	ZZ_{ee}	0.0420	---	
T50 × T50	ZZ_{55}	0.0787	0.05652	
etOH × Arom	ZZ_{ea}	0.0217	0.04544	
etOH × T50	ZZ_{er}		-0.02795	0.0707
etOH × T90	ZZ_{e5}	0.0357	0.03080	-0.0728
etOH × RVP	ZZ_{e9}	0.0476	---	
Arom × RVP	ZZ_{ar}	0.0272	---	
Arom × T50	ZZ_{a5}		0.04528	
Arom × T90	ZZ_{a9}	0.0205		
T50 × T90	ZZ_{59}	0.0544	---	
RVP × T90	ZZ_{r9}		---	
σ^2_{veh}		0.1215	0.8507	
σ^2_{ε}		0.07335	0.1300	

[1] Fit excluding the Odyssey and Sienna. See 6.1.3.
[2] Not fit by the Tobit model; calculated manually from individual vehicle intercepts.
[3] Results identical to those for design model above.

Table 60. NMHC: Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3
Intercept	Intercept	-1.0315	-5.3253	
etOH	Z_e	0.03094		-0.05810
Arom	Z_a	0.09461	0.03987	
RVP	Z_r	-0.04568	-0.05881	-0.03130
T50	Z_5	0.13689	0.04548	0.1356
T90	Z_9	0.02160	0.08202	0.1546
etOH × etOH	ZZ_{ee}	0.04612	---	
T50 × T50	ZZ_{55}	0.07534	0.04774	
etOH × Arom	ZZ_{ea}	0.02045		
2etOH × RVP	ZZ_{er}			0.07730
etOH × T50	ZZ_{e5}	0.04729		
etOH × T90	ZZ_{e9}	0.02441	---	
σ^2_{veh}		0.1266	0.9691	
σ^2_{ε}		0.07624	0.1708	

[1] Fit excluding the Odyssey and Sienna. See 6.1.3.
[2] Not fit by the Tobit model; calculated manually from individual vehicle intercepts.

Table 61. NMHC: Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1	Bag 2[1]	Bag 3[3]
Intercept	Intercept	-1.0308	-5.3256	
etOH	Z_e	0.04439	0.01334	-0.05810
Arom	Z_a	0.11119	0.03743	
RVP	Z_r	-0.04765	-0.03909	-0.03130
T50	Z_5	0.15733	0.05566	0.1356
T90	Z_9	0.01682	0.09799	0.1546
etOH × etOH	ZZ_{ee}	0.04274	---	
T50 × T50	ZZ_{55}	0.07769	0.05909	
etOH × Arom	ZZ_{ea}	0.02594	0.05087	
etOH × T50	ZZ_{er}		-0.03153	0.07730
etOH × T90	ZZ_{e5}	0.03579	0.02865	
etOH × RVP	ZZ_{e9}	0.05125	---	
Arom × RVP	ZZ_{ar}	0.02820		
Arom × T50	ZZ_{a5}		0.05429	
Arom × T90	ZZ_{a9}	0.02068		
T50 × T90	ZZ_{59}	0.05225		
RVP × T90	ZZ_{r9}	---		
σ^2_{veh}		0.1257	0.9723	
σ^2_{ε}		0.07430	0.1684	

[1] Fit excluding the Odyssey and Sienna. See 6.1.3.
[2] Not fit by the Tobit model; calculated manually from individual vehicle intercepts.
[3] Results identical to those for design model above.

Table 62. CH_4 : Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2	Bag 3[1]
Intercept	Intercept	-3.0074	-5.7075	-4.4742
etOH	Z_e	0.06994	0.05860	0.02805
Arom	Z_a	-0.1053	-0.09836	-0.09578
RVP	Z_r	-0.03275	-0.02049	
T50	Z_5	0.07554	0.04394	0.03025
T90	Z_9		0.02575	0.01691
etOH × etOH	ZZ_{ee}	0.02844	---	
T50 × T50	ZZ_{55}	0.05170	0.01227	
etOH × Arom	ZZ_{ea}	0.02088	0.008769	0.01528
etOH × RVP	ZZ_{er}	0.01082		---
etOH × T50	ZZ_{e5}	0.03048	---	-0.02079
etOH × T90	ZZ_{e9}			
σ^2_{veh}		0.2855	1.1108	0.4477
σ^2_{ε}		0.03014	0.02518	0.03439

[1] Fit excluding the Corolla.

Table 63. CH_4 : Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1	Bag 2	Bag 3[1]
Intercept	Intercept	-3.0068	-5.7076	-4.4742
etOH	Z_e	0.08877	0.06076	0.02805
Arom	Z_a	-0.09816	-0.09211	-0.09578
RVP	Z_r	-0.02455	-0.02082	
T50	Z_5	0.1059	0.04477	0.03025
T90	Z_9	0.008573	0.02445	0.01691
etOH × etOH	ZZ_{ee}	0.03133	---	
T50 × T50	ZZ_{55}	0.05882	0.01398	
etOH × Arom	ZZ_{ea}	0.03977	0.01047	0.01528
etOH × T50	ZZ_{er}			---
etOH × T90	ZZ_{e5}	0.02883	---	-0.02079
etOH × RVP	ZZ_{e9}	0.02655	---	
Arom × RVP	ZZ_{ar}	0.02791	---	
Arom × T50	ZZ_{a5}	0.02585	---	
Arom × T90	ZZ_{a9}	0.03072	---	
T50 × T90	ZZ_{59}	0.02280	0.01374	
RVP × T90	ZZ_{r9}		-0.009921	
σ^2_{veh}		0.2853	1.1109	0.4477
σ^2_{ε}		0.02883	0.02486	0.03439

[1] Results identical to those for the design model above.

Table 64. CO: Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2	Bag 3
Intercept	Intercept	1.3466	-1.3893	-1.1409
etOH	Z_e	-0.1049		-0.0815
Arom	Z_a	-0.01242	0.0913	0.0719
RVP	Z_r	-0.00762	0.0299	0.0239
T50	Z_5	-0.03273	0.0261	
T90	Z_9	-0.1571	0.0440	0.0578
etOH × etOH	ZZ_{ee}	0.07304		---
T50 × T50	ZZ_{55}	0.05358		---
etOH × Arom	ZZ_{ea}	0.02086		
etOH × RVP	ZZ_{er}	0.01596		---
etOH × T50	ZZ_{e5}	0.1064		---
etOH × T90	ZZ_{e9}			
σ^2_{veh}		0.3920	1.9187	2.4412
σ^2_{ε}		0.07214	0.1256	0.1819

Table 65. CO: Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1	Bag 2	Bag 3[1]
Intercept	Intercept	1.3472	-1.3895	-1.1409
etOH	Z_e	-0.06967		-0.0815
Arom	Z_a	-0.01156	0.09800	0.0719
RVP	Z_r	0.02063	0.02839	0.0239
T50	Z_5	0.02160	0.02484	
T90	Z_9	-0.1469	0.04177	0.0578
etOH × etOH	ZZ_{ee}	0.08535		---
T50 × T50	ZZ_{55}	0.07222		---
etOH × Arom	ZZ_{ea}	0.06244	---	
etOH × T50	ZZ_{er}		---	---
etOH × T90	ZZ_{e5}	0.1117	---	---
etOH × RVP	ZZ_{e9}		---	
Arom × RVP	ZZ_{ar}	0.05370	---	
Arom × T50	ZZ_{a5}	0.05859	---	
Arom × T90	ZZ_{a9}	0.03861	---	
T50 × T90	ZZ_{59}		0.02177	
RVP × T90	ZZ_{r9}			
σ^2_{veh}		0.3926	1.9196	2.4412
σ^2_{ε}		0.06981	0.1252	0.1819

[1] Results identical to those for design model above.

Table 66. NO_x: Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1[1]	Bag 2[2]	Bag 3
Intercept	Intercept	-2.8594	-4.5692	
etOH	Z_e	0.06750	0.06299	
Arom	Z_a	0.1339	0.04407	
RVP	Z_r			
T50	Z_5	0.04783		
T90	Z_9			
etOH × etOH	ZZ_{ee}			
T50 × T50	ZZ_{55}		---	
etOH × Arom	ZZ_{ea}	-0.02369		
etOH × RVP	ZZ_{er}			
etOH × T50	ZZ_{e5}		---	
etOH × T90	ZZ_{e9}		---	
σ^2_{veh}		0.5925	0.4720	
σ^2_{ε}		0.1458	0.1836	

[1] Fit excluding the Ford Focus.
[2] Fit excluding the Chevrolet Cobalt.

Table 67. NO_x : Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1[1]	Bag 2[2]	Bag 3[3]
Intercept	Intercept	-2.8602	-4.5692	
etOH	Z_e	0.03718	0.06299	
Arom	Z_a	0.1258	0.04407	
RVP	Z_r	-0.01452		
T50	Z_5			
T90	Z_9	0.005211		
etOH × etOH	ZZ_{ee}			
T50 × T50	ZZ_{55}		---	
etOH × Arom	ZZ_{ea}	-0.02699		
etOH × T50	ZZ_{er}			
etOH × T90	ZZ_{e5}		---	
etOH × RVP	ZZ_{e9}		---	
Arom × RVP	ZZ_{ar}			
Arom × T50	ZZ_{a5}			
Arom × T90	ZZ_{a9}	-0.04990		
T50 × T90	ZZ_{59}			
RVP × T90	ZZ_{r9}	0.03160		
σ^2_{veh}		0.5939	0.4720	
σ^2_{ε}		0.1437	0.1836	

[1] Fit excluding the Ford Focus. [2] Fit excluding the Chevrolet Cobalt.
[3] Results identical to those for design model above.

Table 68. PM : Reduced Models, based on the 11-term design model.

Model term	Notation	Bag 1	Bag 2	Bag 3
Intercept	Intercept			
etOH	Z_e	0.1582	0.1126	0.0173
Arom	Z_a	0.3833	0.1662	0.0216
RVP	Z_r			-0.1098
T50	Z_5	0.0550		0.0167
T90	Z_9	0.2923	0.1072	
etOH × etOH	ZZ_{ee}			
T50 × T50	ZZ_{55}	0.0935		
etOH × Arom	ZZ_{ea}			0.1023
etOH × RVP	ZZ_{er}			
etOH × T50	ZZ_{e5}			-0.1218
etOH × T90	ZZ_{e9}			
σ^2_{veh}				
σ^2_{ε}				

Table 69. PM : Reduced Models, based on the 16-term extended model.

Model term	Notation	Bag 1	Bag 2	Bag 3
Intercept	Intercept	0.6529[1]	-1.3121[1]	
etOH	Z_e	0.1807	0.1158	0.0884
Arom	Z_a	0.3792	0.1988	0.0667
RVP	Z_r			0.1692
T50	Z_5	0.1004	-0.003700	-0.0224
T90	Z_9	0.3034	0.09530	0.085
etOH × etOH	ZZ_{ee}			
T50 × T50	ZZ_{55}	0.0806		
etOH × Arom	ZZ_{ea}			
etOH × T50	ZZ_{er}			
etOH × T90	ZZ_{e5}	0.0647		0.0637
etOH × RVP	ZZ_{e9}			
Arom × RVP	ZZ_{ar}			
Arom × T50	ZZ_{a5}			0.0894
Arom × T90	ZZ_{a9}	00665		
T50 × T90	ZZ_{59}		0.1059	
RVP × T90	ZZ_{r9}			
σ^2_{veh}				
σ^2_{ε}				

[1] Not fit by the Tobit model; calculated manually from individual vehicle intercepts.

7.3 Detailed Review and Interpretation

At this point, it is appropriate to give selected models and terms additional review and scrutiny, with particular focus on interaction terms. We illustrate this involved and intensive process through two examples: (1) NO_x (Bag 1) and CO (Bag 1).

7.3.1 Example 1: NO_x (Bag 1)

The reduced model based on the 11-term design model contains three linear terms and one interaction (etOH×Arom) (see Table 36, page 143). The corresponding model based on the 16-term extended model contains two additional interactions, arom×T90 and RVP×T90 (Table 40, page 149).

Review of the coefficients shows that all three interaction terms can be considered as "interference" terms. For the etOH×Arom and Arom×T90 terms this interpretation is straightforward because the linear effects are positive in sign while the interaction is negative. The RVP×T90 also qualifies as an interference because while the sign of the interaction itself is positive, one of the linear terms (RVP) has a negative sign, while the other (T90) has a positive sign, although they are very similar in magnitude. To review the interactions in detail, we can plot them in several ways.

We can begin with interaction, or conditional-effects plots. As before, we calculated them by averaging the data by target levels of the two fuel properties of interest, and across the levels of the remaining three fuel properties. These plots are similar to those in 4.1 (page 49), with some slight modifications. For effects involving ethanol, we omitted the 15% ethanol level and for T90, we omitted the 325°F level. The selected levels of these two properties represent small numbers of fuels, not well balanced across the remaining fuel properties. After omitting them, the underlying patterns detected by the models are simpler to illustrate.

Interaction plots for these three terms are shown in Figure 84. These plots display the "interference effects" described by these terms. For example, in subplots (a) and (c), trends in $\ln(NO_x)$ vs. ethanol and T90 levels are displayed by levels of aromatics. In both cases the trends at the lower aromatics level are positive and relatively steep ($|\Delta \ln NO_x| \sim 0.15$), whereas the trends at the higher aromatics level are negative and relatively less steep ($|\Delta \ln NO_x| \sim 0.05$). Subplots (b) and (d) show the complementary views, with $\ln(NO_x)$ vs. Aromatics by levels of Ethanol and T90. In these views the trends vs. ethanol and T90 are similar in that the steepness of the trends vs. aromatics declines from lower to higher levels of ethanol and T90. However, in contrast to (a) and (c), the strong aromatics effect is prominent in (b) and (d) ($|\Delta \ln NO_x| \sim 0.27$).

The pattern in subplots (e) and (f), showing the RVP×T90 interaction, differs in that the trends at the higher levels are positive, and those at the lower levels are negative. The size of the effects also alternates, in that in (e) the negative effect is stronger whereas in (f) the positive effect is

stronger, and vice versa for the positive effect in (e) and the negative effect in (f). The visual effect is that of a "mirror image" between (e) and (f), which does not apply between (a) and (b) or (c) and (d).

Figure 84 also shows linear effects, in which the data is averaged across the target levels of only one fuel parameter. Subplots (b) and (d) show the strong (and significant) aromatics effect, whereas (c), (e) and (f) show the relatively small (and insignificant) effects for RVP and T90. The graphic presentation, in conjunction with model fitting, makes it clear that RVP and T90 are retained in the model solely on the strength of their interactions, for the maintenance of hierarchy. For the etOH×Arom interaction, however, both underlying linear effects contribute to the goodness-of-fit without their interaction, although including the interaction does improve the fit.

Figure 84 shows interaction plots for an additional pair of parameters, etOH and T50. This term does not qualify as an "interaction" as no interaction term is retained in reduced models, whether fit with respect to the 16-term or 11-term full models. However, the position of the T50 term is unclear in that it is dropped during model fitting based on the 16-term model, but retained in model fitting based on the design model. In subplot (g) ln NO_x is portrayed with respect to ethanol, with different series by T50 level. Due to the non-orthogonal relation between ethanol and T50 in the design (Figure 1, page 27), the plot appears disjointed, with no complete series across the entire range of ethanol. Also, in this presentation, three T50 levels are paired with only one ethanol level (150°F, 10%; 165°F, 20%; and 240°F, 0%). The remaining two T50 levels, 190°F and 220°F, are both paired with 0% and 10% ethanol levels, with the series for 190°F higher than that for 220°F, apparently suggesting a negative effect for T50. The picture is complicated by the fact that the point for 240°F is higher than the 190°F series, and the point for 150°F is lower than the 220°F series. The complementary view in subplot (h) also shows apparent "local" decreases with increasing T50 between 190°F and 220°F, both for 0% and 10% ethanol. However, it also shows apparent local increases between 150°F and 190°F for 10% ethanol, and between 220°F and 240°F for 0% ethanol. The overall apparent trend in means over all T50 levels and across all ethanol levels is positive, which is reflected in the reduced model fit with respect to the design model, but which is absent from the reduced model fit with respect to the 16-term extended model.

The conclusions from the graphic presentation are confirmed by model fitting parameters shown in Table 70-Table 72. These tables represent the results of model fitting, if the 5-term linear effects model is treated as the full model. The results show that the fit is improved by removing the RVP and T90 linear terms, which is not a surprising outcome in view of Figure 84.

Another way of viewing the interactions is to average and plot the residuals of the linear effects model. Figure 85 shows the residuals averaged and plotted by target fuel properties. For all three interactions (subplots (a)-(f)), the results reflect the interference interactions with trends that cross each other, although the residual plots are "detrended", with the trends for individual factor

levels crossing the 0.0 line on the *y*-axis. For the etOH×T50 pair, the view is similar to that in Figure 84, with the residuals "detrended" around the 0.0 line, although the patterns of points underlying the subtrends do not change. However, the "mean" lines reflect the results of the linear effects model in that they are close to the 0.0 line, with the trend against ethanol (g) showing smaller deviations than the trend against T50 (h). However, within the "zig-zag" effect, the main points are apparently balanced around the no-trend line, consistent with retention of a positive T50 term in the linear effects model.

In Figure 86, the residuals are averaged and plotted against the two-stage standardized interaction terms (denoted by *ZZ*). In subplots (a) to (f), the existence of interactions is suggested by the trends across the zero line. Further, the direction and sign of the trends reflect the sign and magnitude of the interaction terms when they are included in the model (Table 36, Table 40). For the etOH×T50 pair (subplots (g) and (h)), despite the irregular patterns of the points when arranged in ethanol and T50 series, the residuals appear more or less balanced across the no trend line, suggesting why the models do not fit an interaction for these two parameters.

To further explore the effect of T50, it is possible to view a subset of data balanced on ethanol and aromatics levels, which are the remaining two properties with clear effects on NO_x. Figure 87 shows a subset of data for four fuels, averaged by T50 and vehicle. The fuels included are 1, 3, 4, and 6, which have target T50 levels of 150, 220, 220 and 190°F, respectively. These fuels have ethanol and aromatics levels of 10% and 15%, respectively, but have varying levels of RVP and T90 levels. The plot shows a separate trend for each vehicle, with apparently mixed results but no apparent overall trend. Patterns do, however, appear to vary by vehicle. For example, some vehicles, including the Liberty, Impala, Cobalt and Odyssey, show apparent negative trends, while others, including the Explorer, F150, Outlook and Civic, show apparent positive trends. One interpretation of these results could be that behavior with respect to T50 is vehicle specific, depending on factors such as vehicle design and calibration. Another possibility is that the effect of T50 in this presentation is masked by variation in RVP and T90 levels across these fuels. In any case, this review of results for the selected fuels does not clarify the relation between NO_x emissions and T50 for Bag 1.

Viewed in physical terms, it seems plausible that bag 1 NO_x as a function of T50 may be highly sensitive to some particulars of vehicle design or calibration. Engine design could be a factor, given that NO_x is primarily a function of combustion temperature, and different vehicles could have differences in design features such as the location of fuel injection in the intake port, rate of heat-up of the manifold and cylinder block, spark timing algorithms (retarding=hotter exhaust gases), etc., all of which could interact with the T50 parameter to produce differences in combustion temperature during the first 30 seconds of operation before the catalyst is active.

Figure 84. ln(NO_x) (Bag 1): Two-way Conditional Effects Plots for Four Pairs of Terms, viewed with respect to both Fuel Parameters : (a) Ethanol×Aromatics, (b) Aromatics×Ethanol, (c) T90×Aromatics, (d) Aromatics×T90, (e) RVP×T90, (f) T90×RVP (g) etOH×T50, (h) T50×etOH.

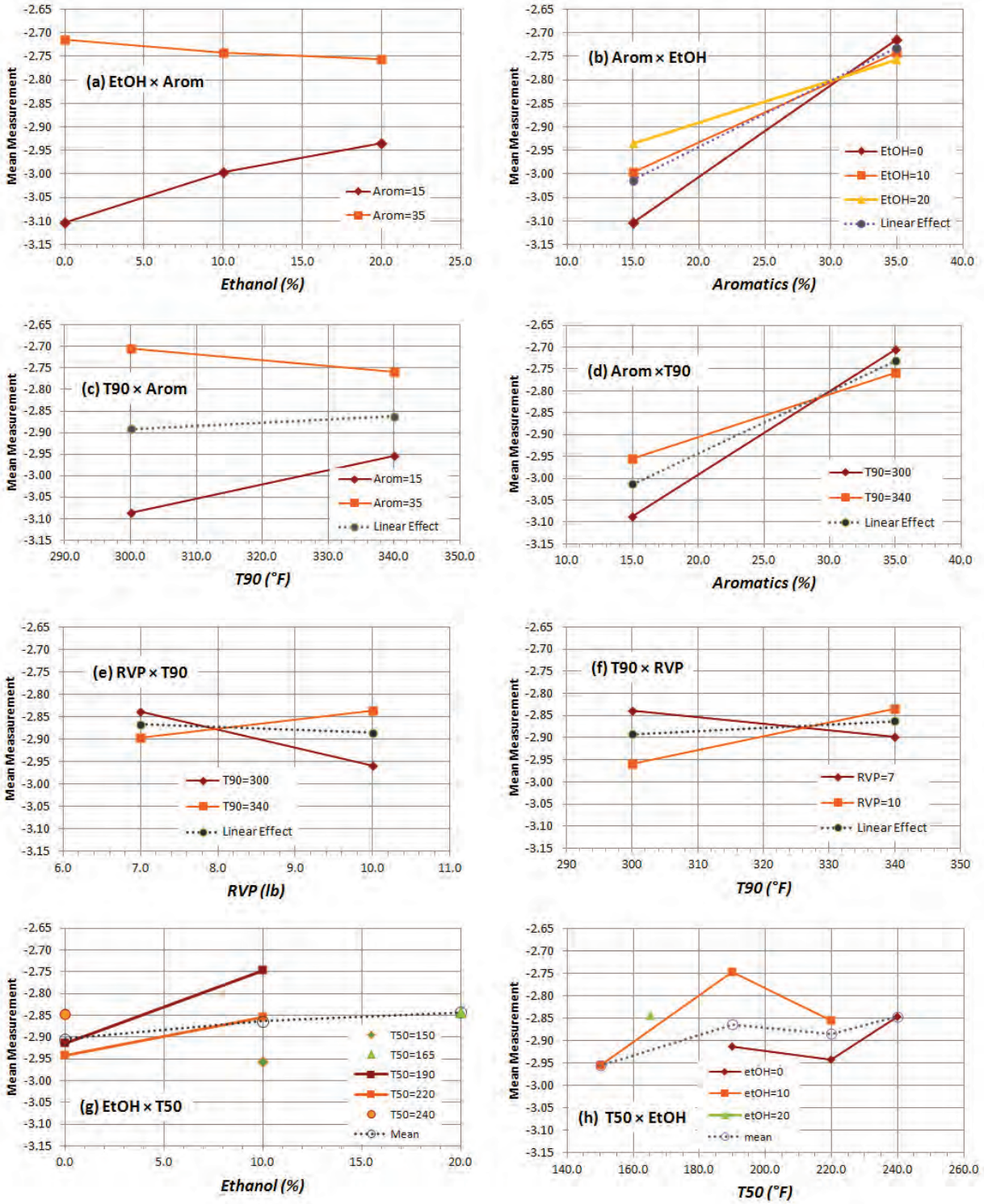


Table 70. Models fit for NO_x (Bag 1) (all models include an intercept term).

Model Term	Notation	Model		
		Linear Effects	LM1	LM2
etOH	Z_e	●	●	●
Arom	Z_a	●	●	●
RVP	Z_r	●	×	
T50	Z_5	●	●	●
T90	Z_9	●	●	×

Table 71. NO_x (Bag 1): Model Fitting History, starting with the 5-term linear effects model.

Fit Parameters				*Test with respect to Full Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.	*d*	Pr>χ^2
Linear Effects	6	895.2	916.4			
LM1	5	895.5	913.9	0.202	1	0.653
LM2[2]	4	895.9	911.7	0.444	1	0.505

[1] A lower value indicates a better fit.
[2] Best fit respect to the 5-term linear-effects model.

[1] The deviation is the difference in the -2loglik statistics for the nested and reference models, respectively, per Equation 13.

Table 72. NO_x (Bag 1): Coefficients and Tests of Effect for the Full and Best-Fit Models, with respect to the 5-term Linear Effects model.

Effect	*Full Model*					*Best-Fit Model (LM2)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*	Estimate	Std.Err.	d.f.	*t*-value	Pr> *t*
Intercept	-2.8616	0.2062	14.00	-13.88	<0.0001	-2.8616	0.2062	14.00	-13.88	<0.0001
Z_e	0.06477	0.01693	879	3.83	0.00014	0.06743	0.01567	879	4.30	<0.0001
Z_a	0.1322	0.01301	879	10.16	<0.0001	0.1322	0.01301	879	10.16	<0.0001
Z_r	-0.006509	0.01448	879	-0.45	0.65					
Z_5	0.04080	0.01738	879	2.35	0.019	0.04381	0.01573	879	2.78	0.0055
Z_9	0.009373	0.01298	879	0.72	0.47					
σ^2_{veh}	0.5929					0.5931				
σ^2_{ε}	0.1462					0.1463				

Figure 85. NO_x (Bag 1): Mean Residuals for the Linear Effects Model, vs. Target Fuel Properties for four pairs of terms: (a) Ethanol × Aromatics, (b) Aromatics ethanol, (c) Aromatics × T90, (d) T90×Aromatics, (e) RVP × T90, (f) T90 × RVP, (g) ethanol × T50, (h) T50 × etOH.

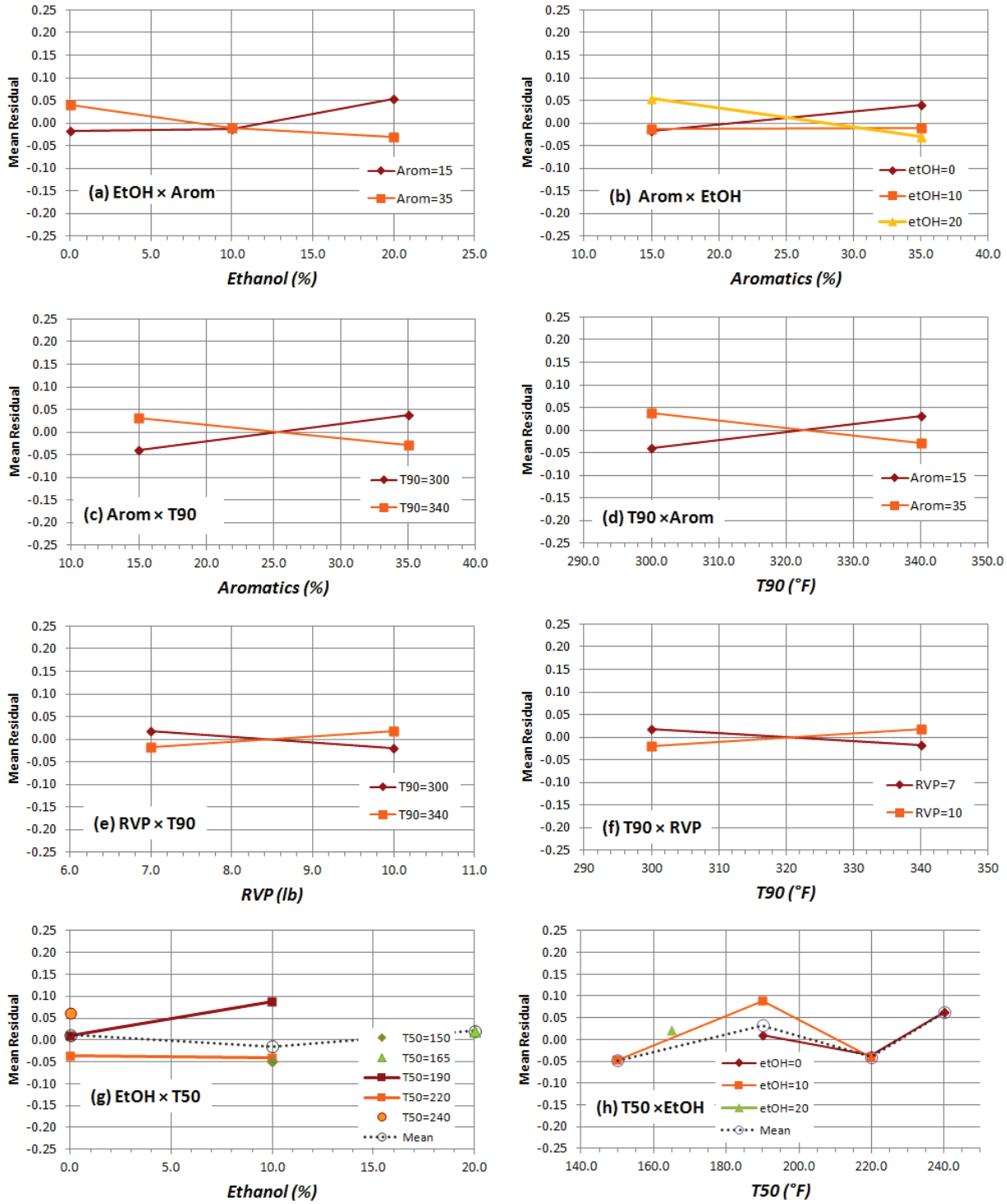


Figure 86. NO_x (Bag 1): Mean Residuals for the Linear-Effects Model, vs. Two-stage Standardized Predictors: (a) Ethanol × Aromatics (b) Aromatics × Ethanol, (c) Aromatics × T90, (d) T90 × Aromatics, (e) T90 × RVP, (f) RVP × T90, (g) ethanol × T50, (h) T50 × ethanol.

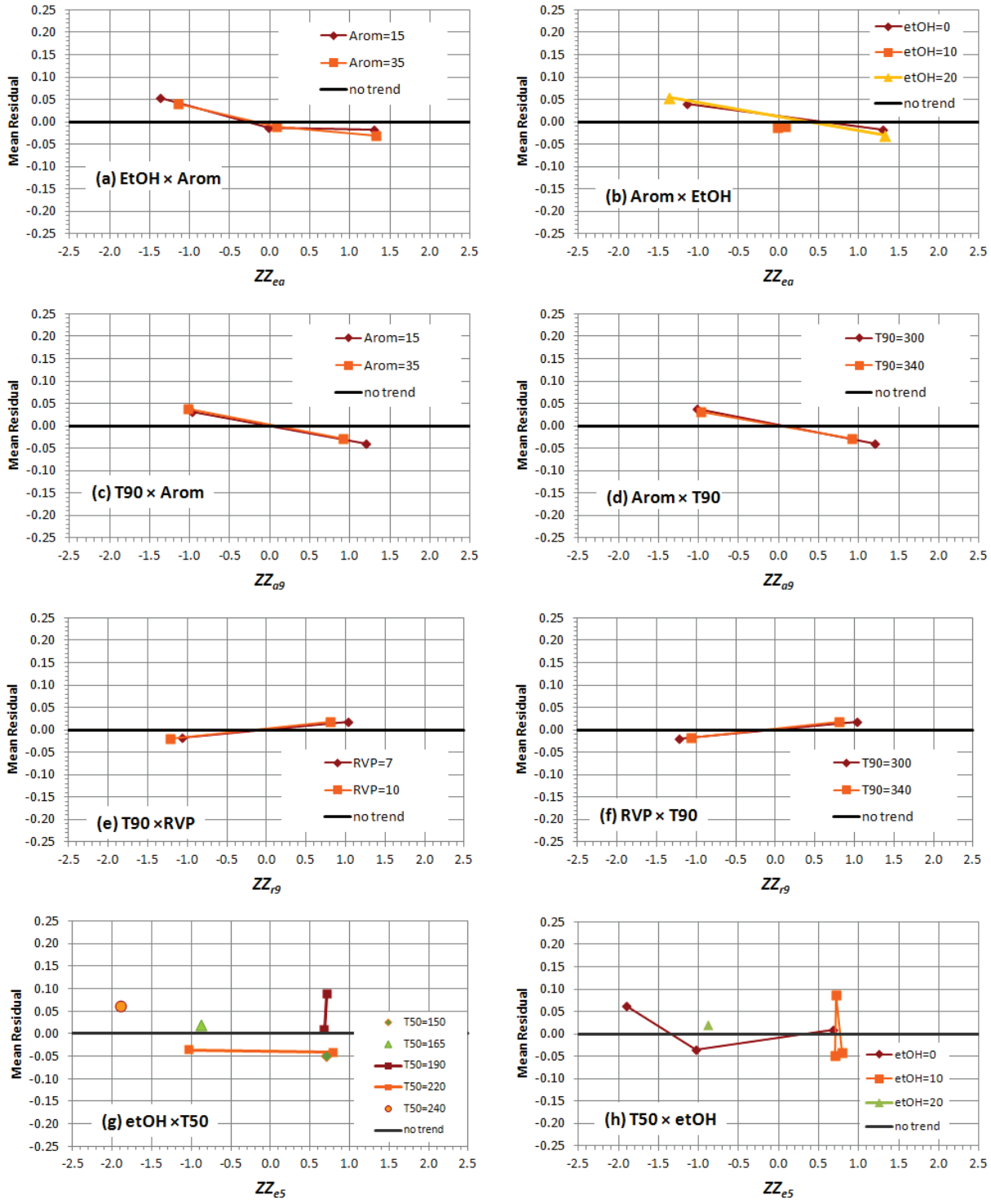
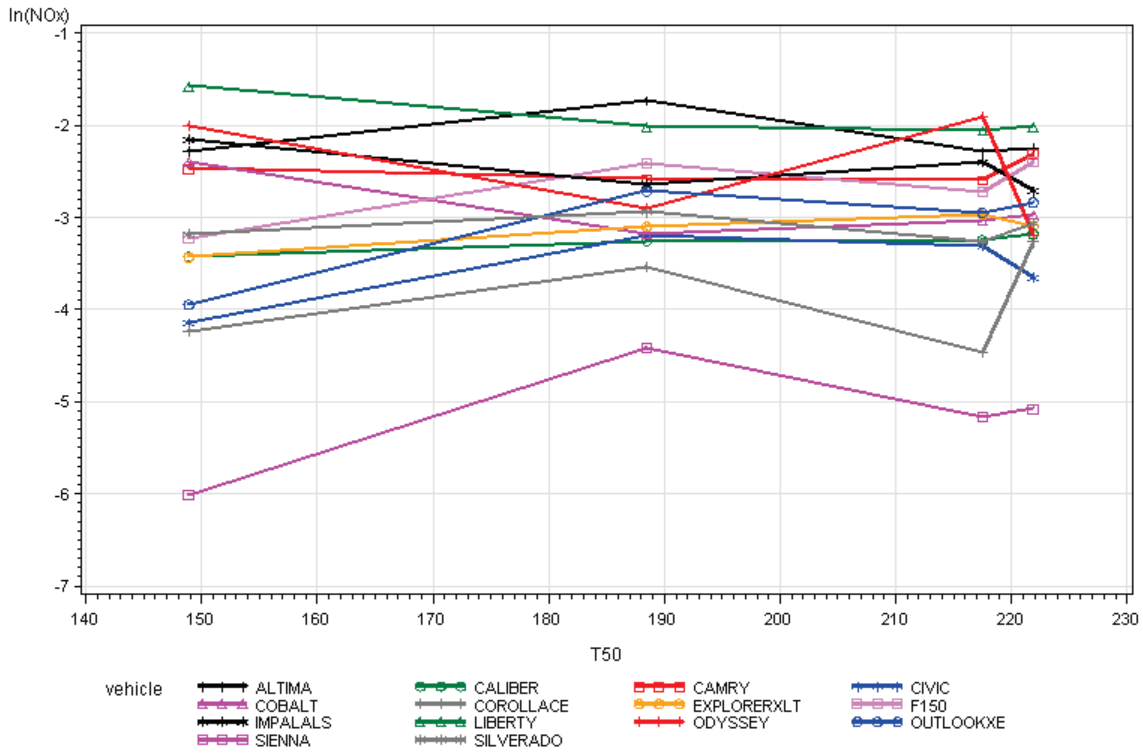


Figure 87. lnNO_x vs. T50 for 4 selected fuels: 1 (T50=150), 6 (T50=190) and 3, 4 (T50=220), by Vehicle. Ethanol = 10%



7.3.2 Example 2: CO (Bag 1)

The reduced model based on the design model retains two quadratic terms and three interaction terms: etOH×Arom, etOH×RVP and etOH×T50 (Table 44, page 154). The reduced model based on the extended model includes three additional interactions, Arom×RVP, Arom×T50 and Arom×T90 (Table 48, page 159).

Review of the coefficients shows that all five interaction terms can be considered as "interference" terms. For the etOH×Arom and Arom×T90 terms this interpretation is straightforward because the linear effects are negative in sign while the interaction is positive. As another example, the etOH×T50 term also qualifies as an interference because while the sign of the interaction itself is positive, one of the linear terms (ethanol) has a negative sign, while the other (T50) has a positive sign, although they differ in magnitude. To review the interactions in detail, we selected three terms for review, plus the quadratic term etOH×etOH.

We begin with interaction, or conditional-effects plots. As before, we calculated these plots by averaging the data by target levels of the two fuel properties of interest, and across the levels of

the remaining three fuel properties. These plots are similar to the ones viewed previously, with some slight modifications. For effects involving ethanol, we omitted the 15% ethanol level and for T90, we omitted the 325°F level. The selected levels of these two properties represent small numbers of fuels, not well balanced across the remaining fuel properties. After omitting them, the underlying patterns detected by the models are simpler to illustrate.

Interaction plots for the selected terms are shown in Figure 88. These plots display the "interference effects" described by these terms. For example, in subplot (a) the trends in lnCO are displayed by aromatics levels. The trends are close to each other and appear close to parallel. A strong interaction is not necessarily obvious in this presentation. Subplots (b) and (c) show trends vs. aromatics level, by levels of ethanol and T90, respectively. In (b), as in (a), a strong interaction is not visually apparent and the relatively strong ethanol effect is shown by the spaces between the trends. In (c), however, the interaction between aromatics and T90 is strongly suggested by the non-parallel slopes for the two T90 levels. In both (b) and (c) the almost level linear effects also portray the weak aromatics linear effect shown by the models. In contrast, (d) and (f) show a strong negative T90 effect, which is the single strongest effect in the model. It is interesting to note that while (d) and (f) appear somewhat similar visually, the model considers the arom×T90 interaction highly significant but the RVP×T90 interaction insignificant.

The conclusions from the graphic presentation are confirmed by model fitting parameters shown in Table 73-Table 75. These tables represent the results of model fitting, if the 5-term linear effects model is treated as the "full" model. The results show that the fit is improved by removing the Aromatic and RVP linear terms, which is not a surprising outcome after having viewed Figure 88.

Another way of viewing the interactions is fit models containing only linear terms, and to examine patterns in residuals averaged and plotted by fuel property levels. In this presentation, "large" deviations from the no-trend line (mean residual = 0.0) suggest the existence of interactions. As review of the plots will show, the phase-3 dataset is large enough to identify apparently "small" interaction effects. Review of presentations such as this can inform consideration of whether interaction effects are physically important as well as statistically significant (or insignificant).

Figure 89 shows the residuals averaged and plotted by target fuel properties. Note that this figure is scaled to allow direct comparison to Figure 88, in that the range shown on the y-axis is identical in both plots ($\Delta y = 0.70$). For the etOH×Arom effect (a), the trends cross and show a slight negative trend; the trend appears slight but is nonetheless identified by the model and rates as significant. In (b) the trends also cross for the RVP×T90 interaction, but the space between them is narrow and the blue trend (T90=300) is nearly coincident with the no-trend line. As mentioned, this interaction is not significant, and not retained in the reduced model. In the Arom×T90 interaction (c), the trends cross the no-trend line in a balanced fashion, so as to mask an overall trend; nonetheless the differences are large enough to rate as highly significant.

Finally, for the etOH×etOH quadratic term (e) the effect does not appear dramatic visually, but is nonetheless highly significant.

In Figure 90, the residuals are averaged and plotted against the two-stage standardized interaction terms (denoted by *ZZ*) used in model fitting. In subplots (a) and (c) the etOH×Arom and Arom×T90 interactions are clearly suggested by the positive trends across the no-trend line. The remaining plots for the RVP×T90 interaction and the etOH×etOH quadratic term appear similar visually, but as noted the former is not significant whereas the latter is. This result may follow from the fact that only one trend for RVP×T90 differs markedly from the no-trend line (T90=340).

Figure 88. ln(CO) (Bag 1): Two-way Conditional Effects Plots for Three Interactions and one Quadratic term, viewed with respect to Both fuel Parameters : (a) Ethanol×Aromatics, (b) Aromatics×Ethanol, (c) Aromatics×T90, (d) T90×Aromatics, (e) RVP×T90, (f) T90×RVP, (g) etOH×etOH.

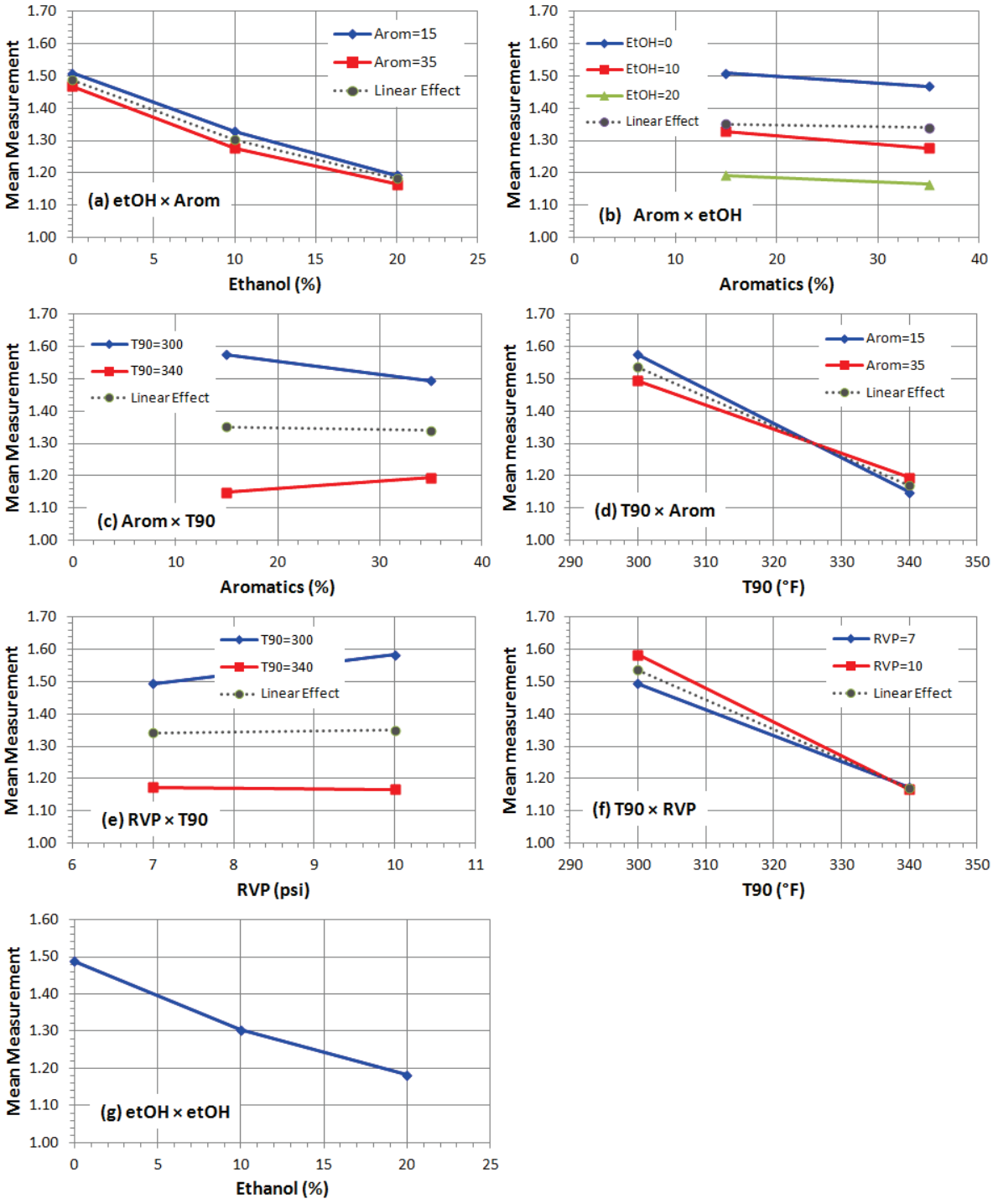


Table 73. Linear-Effects Models fit for CO (Bag 1) (all models include an intercept term).

Model Term	Notation	Model			
		Linear Effects	LM1	LM2	LM3
etOH	Z_e	●	●	●	●
Arom	Z_a	●	●	×	
RVP	Z_r	●	×		
T50	Z_5	●	●	●	×
T90	Z_9	●	●	●	●

Table 74. CO (Bag 1): Model Fitting History, starting with the 5-term linear effects model.

Fit Parameters				*Test with respect to Full Model*		
Model	p	-2ln$L$	BIC[1]	Dev.	d	Pr>χ^2
Linear Effects	6	323.94	345.60			
LM1	5	325.65	344.60	1.709	1	0.19
LM2[2]	4	327.34	343.58	1.691	1	0.19
LM3	3	340.25	353.79	12.915	1	0.0003

[1] A lower value indicates a better fit.
[2] Best fit respect to the 5-term linear-effects model.

[1] The deviation is the difference in the -2loglik statistics for the nested and reference models, respectively, per Equation 13.

Table 75. CO (Bag 1): Coefficients and Tests of Effect for the Full and Best-Fit Models, with respect to the 5-term Linear Effects model.

Effect	*Full Model*					*Best-Fit Model (LM2)*				
	Estimate	Std.Err.	d.f.	t-value	Pr> t	Estimate	Std.Err.	d.f.	t-value	Pr> t
Intercept	1.3479	0.1616	15	8.34	<0.0001	1.3479	0.1616	15	8.34	<0.0001
Z_e	-0.1315	0.1167	941	-11.27	<0.0001	-0.1243	0.01079	941	-11.52	<0.0001
Z_a	-0.01192	0.008995	941	-1.33	0.185					
Z_r	-0.01304	0.009967	941	-1.31	0.191					
Z_5	-0.04751	0.01198	941	-3.96	<0.0001	-0.03892	0.01079	941	-3.60	0.0003
Z_9	-0.1549	0.008970	941	-17.26	<0.0001	-0.1561	0.008913	941	-17.51	<0.0001
σ^2_{veh}	0.3906									
σ^2_{ε}	0.07501									

Figure 89. CO (Bag 1): Mean Residuals for the Linear Effects Model, vs. Target Fuel Properties for three Interactions: (a) Ethanol × Aromatics, (b) Aromatics × T90, and (c) RVP × T90, and the quadratic term etOH×etOH.

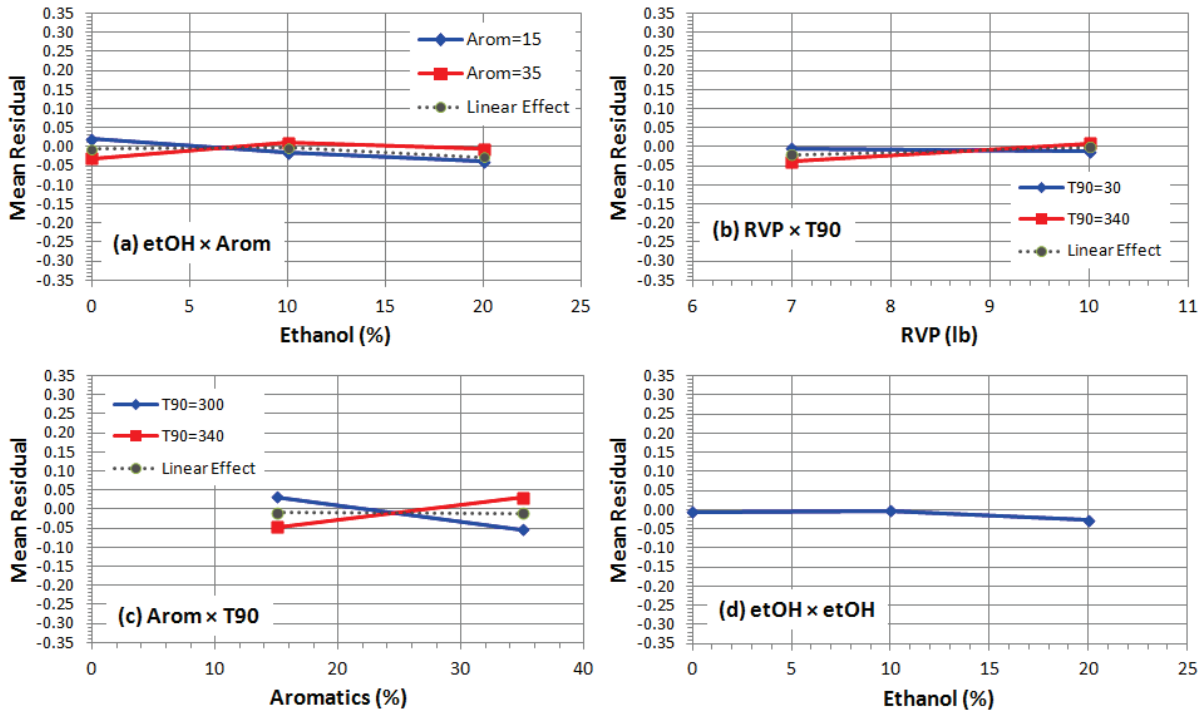
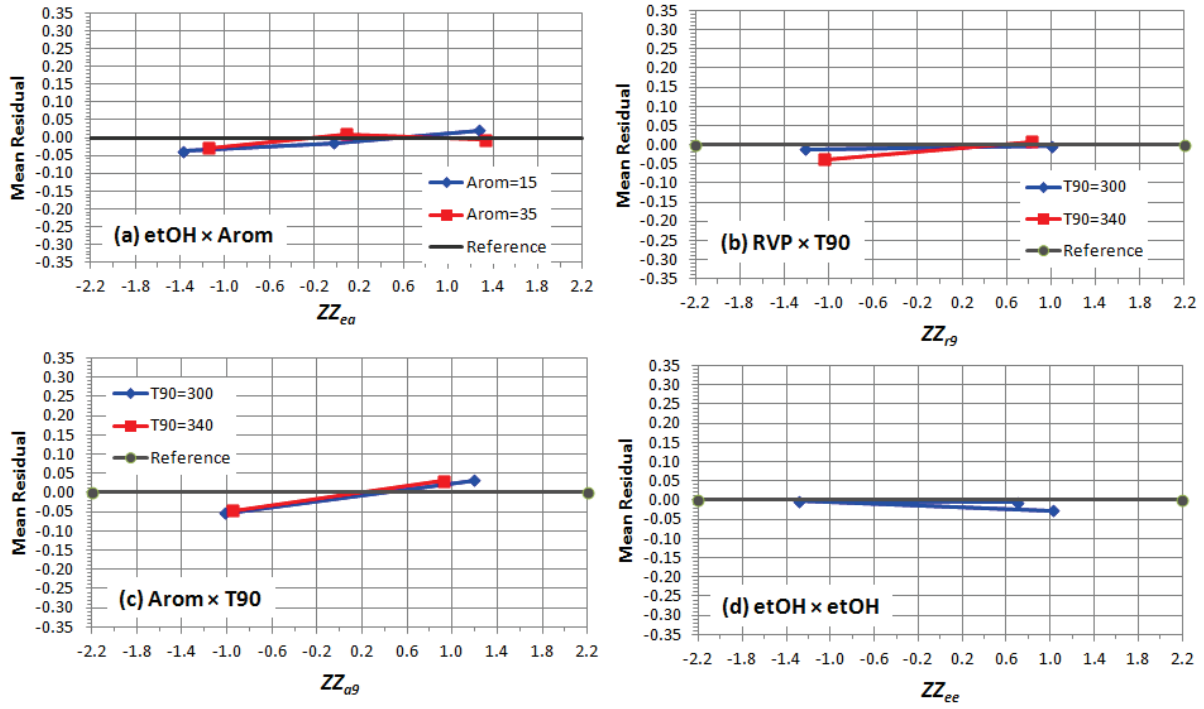


Figure 90. CO (Bag 1): Mean Residuals for the Linear-Effects Model, vs. Two-stage Standardized Predictors: (a) Ethanol × Aromatics interaction, (b) RVP × T90 Interaction (insignificant), (c) Arom × T90 Interaction, and (d) quadratic term etOH×etOH.



8 Analyses for Speciated Hydrocarbon Compounds and Air Toxics

8.1 Measurements

During the Phase-3 project, hydrocarbons were speciated for specific subsets of vehicles and fuels (See Table 5, page 20). The processes and methods for hydrocarbon speciation are described in greater detail in the testing report[5]. A subset of these compounds selected for statistical analysis include acetaldehyde, formaldehyde, acrolein, ethanol, benzene, 1,3-butadiene, and ethane.

This chapter describes the development of models relating the emissions of selected species to changes in fuel properties. The approaches and methods used are similar to those used for the regulated pollutants, except as specifically described. One difference is that for the speciated compounds, models were fit to Bag 1 and Bag 2 results only, representing "cold-start" and "hot-running" processes. No models were fit for Bag 3 results.

8.2 Parameters and Design Efficiency

Due to limitations in budget, the entire study design was not applied to speciated hydrocarbons, including those discussed in this chapter. For the speciated compounds, the volume of data collected varies by Bag, compound and vehicle. For some compounds, measurements for Bag 1 were taken for all vehicles over the entire fuel set, thus encompassing the entire study as designed, including replication. However, for the remaining compounds in Bag 1 and for all compounds in Bags 2 and 3, measurements were taken for a smaller number of vehicles over a reduced set of fuels, without replication. The combinations of fuels and vehicles included for each compound analyzed are summarized in Table 76.

Note that the level of effort applied to Bag 1 measurements is variable. Four compounds were measured under the "*full design*," i.e., 15 vehicles over 27 fuels, with replication. Two compounds were measured under the "*reduced design*," i.e., 15 vehicles over 11 fuels, without replication. A single compound was measured under the reduced design, i.e., on five vehicles over 11 fuels.

Throughout this chapter, the set of 27 fuels included in the full design will be denoted as the "full fuel-parameter matrix," as it includes all the fuel parameter points for which the design was optimized. Similarly, the set of 11 fuels included in the "reduced design" will be denoted as the "reduced fuel-parameter matrix," as it covers a set of fuel parameter points narrower than that for which the design was originally optimized. The study fuels included in the full and reduced fuel matrices are shown in Table 77.

Table 76. Features of the Study Design Applied to Speciated Compounds Selected for Analysis.

Compound	Bag 1			Bag 2		
	No. vehicles	No. Fuels	replication	No. vehicles	No. Fuels	replication
Acetaldehyde	15	27	YES	5	11	NO
Formaldehyde	15	27	YES	5	11	NO
Acrolein	15	27	YES	5	11	NO
Ethanol	15	27	YES	5	11	NO
Benzene	15	11	NO	5	11	NO
1,3-Butadiene	15	11	NO	5	11	NO
Ethane	5	11	NO	5	11	NO

Table 77. Nominal Target parameters for fuels in the Phase-3 EPAct program.

Fuel[1]	In Reduced Matrix	etOH (%)	Aromatics (%)	RVP (psi)[2]	T50 (°F)	T90 (°F)
1		10	15	10	150	300
2		0	15	10	240	340
3	YES	10	15	7	220	300
4	NO[3]	10	15	10	220	340
5		0	35	7	240	300
6	YES	10	15	7	190	340
7	YES	0	15	7	190	300
8		0	15	10	220	300
9		0	35	10	190	340
10	YES	10	35	7	220	340
11		10	35	10	190	300
12		10	35	10	150	340
13	YES	0	35	7	220	340
14	YES	0	15	7	190	340
15		0	35	10	190	300
16		10	35	7	220	300
20		20	15	7	165	300
21	YES	20	35	7	165	300
22		20	15	10	165	300
23	YES	20	15	7	165	340
24		20	15	10	165	340
25		20	35	10	165	340
26		15	35	10	165	340
27	YES	15	15	7	220	340
28	YES	15	35	7	220	300
30		10	35	10	150	325
31	YES	20	35	7	165	325

[1] Note that numbering of fuels is not entirely sequential throughout.
[2] This parameter was measured as "DVPE," but for simplicity, will be referred to as "RVP" in this document.
[3] Fuel 4 was originally considered for inclusion in the reduced fuel set. However, it was excluded as the only fuel with an RVP level of 10 psi. All other fuels in the reduced matrix have RVP values of 7 psi.

The efficiency of the full design was evaluated during the design of the Phase-3 program. It was evaluated for a model design that included eleven terms. The "*G*-efficiency" for the full design was estimated at 51.6% for the eleven design parameters, as previously described in 2.1 above (page 19).

However, in considering analysis of fuel effects under the reduced design, it was clear that implementation of the reduced fuel set represents an effective design change, and that its design efficiency would not necessarily be adequate for all the original 11 model terms. We therefore reevaluated the efficiency of the reduced design, to identify a set of fuel parameters for which it could estimate effects with reasonable efficiency. While application of results is beyond the scope of this report, this *de facto* design change warrants attention.

The model terms considered for the full and reduced designs are summarized in Table 78, with their respective design efficiencies. The reevaluation of the reduced design commenced with the five linear effects, evaluated over 12 fuels, including the 11 identified in Table 76, plus Fuel 4 (design 1a). Because Fuel 4 had an RVP level of 10 psi, whereas the remaining 11 fuels had values of 7 psi, a design with four linear effects, excluding RVP, was also evaluated, for sets of 12 and 11 fuels (designs 1b and 1c, respectively). As excluding Fuel 4 from the fuel set and dropping RVP as a model term gave a marked increase in *G*-efficiency, design 1c was selected as the full model for the reduced design. At this point, design efficiency for the inclusion of one quadratic term and three interaction terms was evaluated (designs 2-5). As shown in the table, the design efficiency drops sharply with inclusion of 2nd order terms, to less than 5% for designs four and five. Based on this analysis, we concluded that the reduced design would not give adequate design efficiency for evaluation of the 2nd order terms. Therefore, we selected design 1c for modeling of fuel effects with the reduced fuel-parameter matrix.

Table 78. Model terms and associated *G*-Efficiency for the full design and selected reduced designs.

Design	Fuels	Model terms											*G*-efficiency (%)
		etOH	Arom	RVP	T50	T90	etOH×etOH	T50×T50	etOH*Arom	etOH*RVP	etOH*T50	etOH*T90	
Full	27	●	●	●	●	●	●	●	●	●	●	●	51.6
1a	12	●	●	●	●	●							15.1
1b	12	●	●		●	●							48.8
1c[1]	11	●	●		●	●							58.3
2	11	●	●		●	●			●				21.1
3	11	●	●		●	●	●		●				17.1
4	11	●	●		●	●	●		●		●		2.8

5	11	•	•		•	•	•		•		•	•	3.0
[1] Design 1c was selected for fuel effects modeling with the reduced fuel set.													

8.3 Additional Terms for Modeling

A model including all possible candidate terms will be referred to as a "full" model. Accordingly, the full model including eleven terms listed in Table 78 will be referred to as the "11-term design model." In addition to the eleven design parameters for which the full parameter matrix was optimized, an additional six parameters were considered for inclusion during model fitting. A full model including these additional terms will be referred to as the "17-term full model." Similarly, the model including only the four linear effects (excluding RVP) will be referred to as the "4-term full model." Candidate terms considered for inclusion for the full and reduced designs are shown in Table 76. For compounds measured under the full design, model fitting started from the 17-term full models; for compounds measured under the reduced design, model fitting started from 4-term full models.

8.4 Correlations Among fuel parameters

In modeling the toxic compounds, the fuel parameters were standardized as discussed above in 2.3.1 (page 28).

8.5 Measurement Issues

Measurement of individual speciated hydrocarbons, including the toxics discussed in this document, involved measurement of small to very small quantities, particularly for the Bag 2 measurements. For measurements at the lower end of the range, questions exist as to whether the concentration of compounds measured at the tailpipe, i.e., "sample" measurements, exceeded measurable concentrations in ambient air, i.e., "background" measurements.

In addition, for some compounds, there was an additional issue concerning whether the measurement media used to collect mass from the sample measurements emitted measurable amounts of the compounds to be measured. Compounds affected by this issue include the carbonyls, acetaldehyde, formaldehyde and acrolein. These issues are discussed in greater detail in Appendix L to the testing report[5].

In the analysis of these compounds, the issue of censoring applied, as for the regulated emissions (see 5.1 (page 89), 5.3 (page 96) and 7.2.2.5 (page 159)). For some speciated compounds in Bag 2, censoring was severe enough to preclude the development of models.

Nonetheless, after some consideration, we elected to combine three approaches to censoring. The first is substitution (for minor censoring). The second is another classic approach to left-censoring (Tobit regression) which we applied in cases of more severe censoring. In addition, for the compounds affected by media contamination, we applied a modeling approach that allows

estimation of uncertain low-level measurements while acknowledging measurement uncertainty (for all measurements) and factoring it into the model-fitting process. This approach, the "estimated dependent variable model" (EDV) will be described below.

At the outset, to estimate measured values, and to identify censored values, we followed the process described here:

First, we assume that the true, but unknown, tailpipe measurements (Y_i) are confounded by media (k) and background (b) contamination, but can be represented as

$$\tilde{Y}_i = Y_i + b_i + k_i \qquad \textbf{Equation 18}$$

where the $\tilde{Y}_i$ are the apparent "contaminated" sample values. But because both k and b have been measured, we can attempt reasonable estimates of the true values $\hat{Y}_i$, as

$$\hat{Y}_i = \tilde{Y}_i - b_i - k_i \qquad \textbf{Equation 19}$$

These equations describe the concept. The implementation was somewhat more complicated, as described below.

It is necessary to emphasize that the measurements of background and media concentrations do not correspond with the sample measurements on the one-to-one basis. Both were measured on a basis corresponding to compound and day, rather than to fuel and vehicle (on specific days). Thus, the background measurements were obtained using sample media for each compound on a daily basis, and can be thus associated with sample measurements (by vehicle and fuel) on a chronological basis. Similarly, the estimates of media contamination were taken as 5-day moving averages of daily measurements from "media blanks" over the previous five days ($\bar{k}_5$), as described in Appendix L to the testing report, and depicted in Equation 20.

$$\bar{k}_5 = \frac{k_{i-5} + k_{i-4} + k_{i-3} + k_{i-2} + k_{i-1}}{5} \qquad \textbf{Equation 20}$$

In addition, some measurements were collected on two media in series, to account for the possibility that the first medium (Y_1) could become saturated before the end of the test, in which case the second medium (Y_2) would be present to collect the "overflow" mass from the first. Both sample ($\tilde{Y}_i$) and background (b_i) measurements were adjusted for media contamination in compilation of the dataset. For a process with two measurement media in series, this process can be represented as

$$\hat{Y}_i = \left(\tilde{Y}_1 - \bar{k}_5\right) - \left(b_1 - \bar{k}_5\right) \quad + \quad \left(\tilde{Y}_2 - \bar{k}_5\right) - \left(b_2 - \bar{k}_5\right) \qquad \textbf{Equation 21}$$

which simplifies as

$$\begin{aligned} \hat{Y}_i &= \tilde{Y}_1 - b_1 + \tilde{Y}_2 - b_2 \\ &= \tilde{Y}_1 + \tilde{Y}_2 - b_1 - b_2 \end{aligned} \qquad \textbf{Equation 22}$$

In addition to estimating the sample measurements, as shown above, we estimated two variances, the first being the variance of the 5-day moving average of the media blanks

($\hat{\sigma}_k^2$), and the second being a variance of random errors ($\hat{\sigma}_\varepsilon^2$).

With respect to the media-blank variance, we make three assumptions: (1) the variance of media blanks can vary over time, (2) media variance is not correlated with the fuel properties, and (3) media variance is not correlated with the random error. With respect to the random error, we also make three assumptions: (1) the variability of the random errors does not vary with time, (2) random error is not correlated with the fuel properties, and (3) random error is not serially auto-correlated.

With these assumptions in place, we estimated the variance of the mean five-day media contamination for each measurement day *i* as

$$\hat{\sigma}_k^2 = \frac{\sum_{j=1}^{5} \left(k_{i-j} - \bar{k}_5\right)^2}{5-1} \qquad \textbf{Equation 23}$$

where *j* ranges from 1 to 5, representing the media blank measurements over the previous five days. The application of these variances will be described below.

Equation 22 above applies to all compounds, whether or not they were affected by media contamination, as all measurements were potentially affected by background interference. After application of Equation 22, if the estimated value of $\hat{Y}_i$ was ≤ 0, the measurement was considered censored and assigned a missing value. Prior to model development, sets of measurements estimated as described above were plotted, with and without aggregation, as described in the following section. Table 79 shows the numbers of censored measurements by compound and Bag. Note that the numbers of total measurements are consistent with the numbers of fuels and vehicles shown in Table 76.

Table 79. Numbers of individual measurements, by compound and Bag, indicating the total number of measurements (n_{total}) and the number of censored measurements ($n_{censored}$), as calculated using Equation 22.

Compound	Bag 1		Bag 2	
	n_{total}	$n_{censored}$	n_{total}	$n_{censored}$
Acetaldehyde	913	0	63	1
Formaldehyde	913	0	63	1
Acrolein	913	23	63	34
Ethanol	913	193	63	24
Benzene	176	0	62	41
1,3-butadiene	62	0	62	42
Ethane	62	0	62	1

8.6 Review of Data

8.6.1 Linear Effects

The data collected in this study are difficult to visualize, in that they encompass variation of emissions in the five-dimensional fuel-parameter space. Due to human limitations, it is practical to view the data in only two dimensions at a time, in some cases including multiple series to represent levels in a third dimension. Despite the risk of misinterpreting graphic portrayals that may oversimplify the actual emissions behavior in all dimensions, it is valuable to review the data visually before model development.

At the outset, it is helpful to get an overview of the raw results, sorted by vehicle and fuel, which gives an initial impression of variability among vehicles and fuels, as well as within vehicles. This view also gives an initial impression of vehicles or observations that may prove influential. Figure 91(a) shows the set of observations for Bag 1 Acetaldehyde, with the data portrayed as common logarithm of the measurements (base 10). Across all fuels, the range of variability for most vehicles spans just over one order of magnitude. The set of measurements for one vehicle, the Focus, are low relative to the other vehicles. For another vehicle, the Outlook, the main body of measurements is in the same range as the other vehicles. However, for this vehicle, a subset of measurements is exceptionally high, relative to those from other vehicles. Figure 91(b) shows a similar depiction of the acrolein results for Bag 1, which depicts and highlights the censored measurements for this compound, which are represented by a uniform low value, to make them visible on the logarithmic plot. For acrolein, the variability both among and within vehicles is greater than for acetaldehyde, with the Focus and Outlook also having the lowest and highest measurements, respectively.

In addition, we averaged and plotted the data to check for evidence of linear effects, i.e., patterns of emissions across all levels of a single fuel parameter. We constructed these views by averaging the data by the levels of one fuel parameter and by vehicle, across all levels of the remaining four parameters, repeating the process for each fuel parameter in turn. We took this step for the emission results themselves (i.e., in "linear space"), as well as for natural-log transforms of the data (i.e., "ln space"). We made a point of examining the ln-transformed results, as the statistical models were developed using the transforms, rather than the raw results.

The study design anticipates the possibility that the response of emissions to changes in multiple fuel parameters may involve several 2-way interactions, which suggests that limiting our examination to "Linear Effects" may be simplistic. To examine 2-way emissions responses, we also averaged and plotted the data by two fuel parameters simultaneously to examine potential "conditional" or interaction effects, to examine how the effect of each fuel parameter varied with the levels of the other parameters.

Below, we illustrate these concepts using results for Acetaldehyde (Bag 1). For this compound, the ethanol×T50 interaction gives an excellent example of two interrelated variables and the importance of supplementing "linear-effects" plots with "interaction" plots.

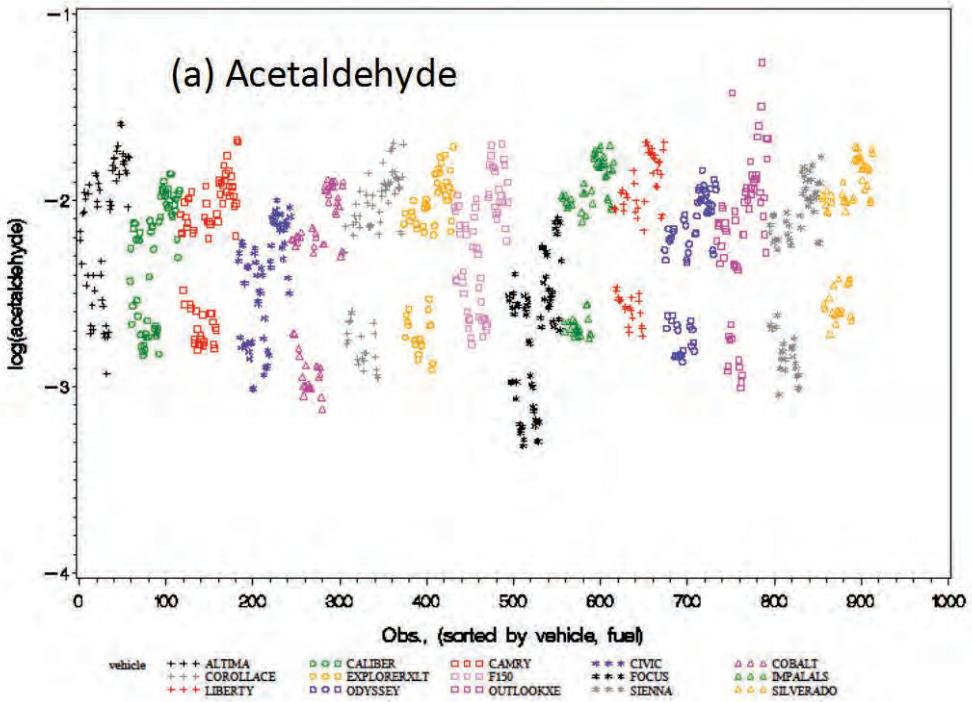
Figure 92 shows "linear-effects" plots for ethanol. By "linear effect" we mean the effect of one factor (ethanol) across all levels of all other factors. In this case we have averaged the data by the four ethanol levels and by vehicle. A strong ethanol effect is visible, which is not surprising for this compound, considering the structural affinity between acetaldehyde and ethanol. The variability among vehicles, also not surprisingly, is fairly wide, spanning about a factor of three at each ethanol level. Thus, absolute variability increases with ethanol level, but relative variability remains fairly stable. This pattern is confirmed by the trends in the transformed results. In logarithmic space, the trends for individual vehicles track closely, whereas in linear space they show a characteristic "fan" shape. These results suggest that the effect of ethanol on acetaldehyde emissions can be expressed multiplicatively, and that the multiplicative factor is similar across the selection of vehicles measured. The gradual down-sloping of the logarithmic trends between 10% and 20% ethanol also suggests that fitting a quadratic term for ethanol would be appropriate.

The linear effects plot for acetaldehyde in relation to T50 (Figure 93) illustrates why viewing the results of a multidimensional experiment in a single dimension can be misleading. The trend shows an increase in emissions from 153 °F to 165 °F, followed by a strong decrease at higher T50 levels. Closer examination and modeling show that this pattern is an example of "Simpson's Paradox," in which averaging emissions across all levels of the secondary variable, ethanol in this case, gives an apparent inversion of the actual trend(s). In this case, the apparent inversion is caused by the relationship between ethanol and T50 levels in the fuel matrix (Figure 1, page 27), compounded by the fact that the effect of ethanol is far stronger than the effect of T50.

These interrelationships are illustrated in the interaction plots, which show the data averaged by levels of two fuel parameters, but not by vehicle. The first two plots (Figure 94) show acetaldehyde vs. ethanol, by T50 level, in linear and logarithmic space. These views show a strong ethanol effect (in the trends) and a smaller T50 effect (spaces between trends). In addition, the shapes of the trends are very similar to those in the previous linear-effects plots. However, the reverse does not hold in the remaining two plots (Figure 95), which show acetaldehyde and ln(acetaldehyde) vs. T50 by ethanol level. These two plots show a very different picture from their corresponding linear-effects views. Specifically, these views show a slight increase in emissions with increasing T50 at each ethanol level, with the strong ethanol effect shown as wide gaps between the trends. It is also apparent why averaging across ethanol levels gives an apparent declining trend, as mentioned above. Review of these data show that it may not be possible to interpret the effect of one fuel parameter without accounting for the levels of one or more additional parameters.

Similar plots for the remaining compounds are presented in Appendices K to Q.

Figure 91. Common logarithms of (a) Acetaldehyde, and (b) Acrolein, by vehicle and fuel (Bag 1, full design).



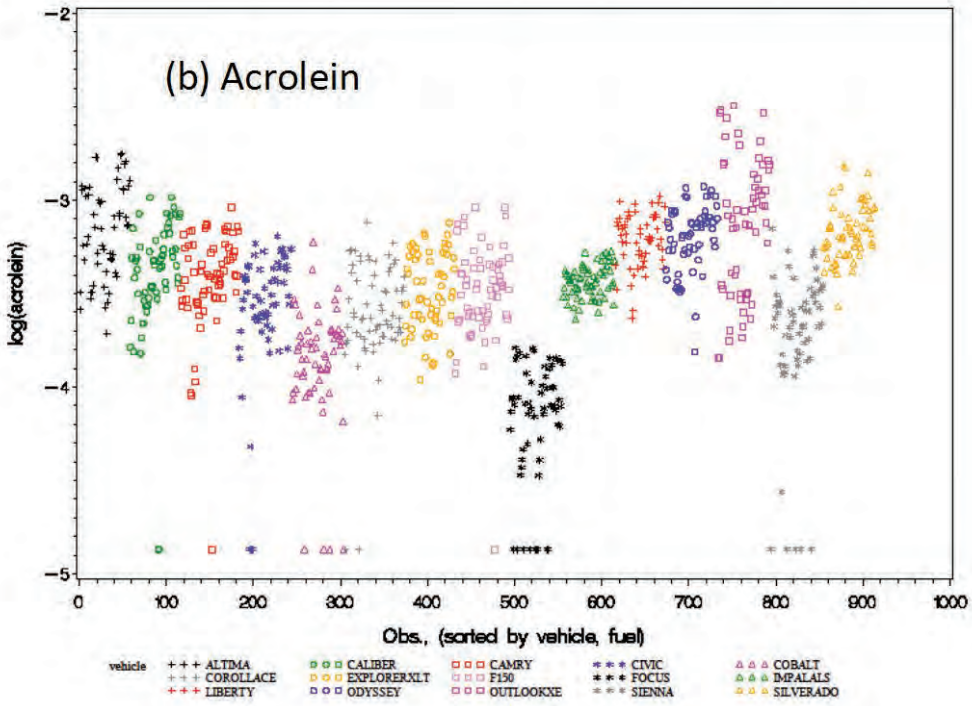
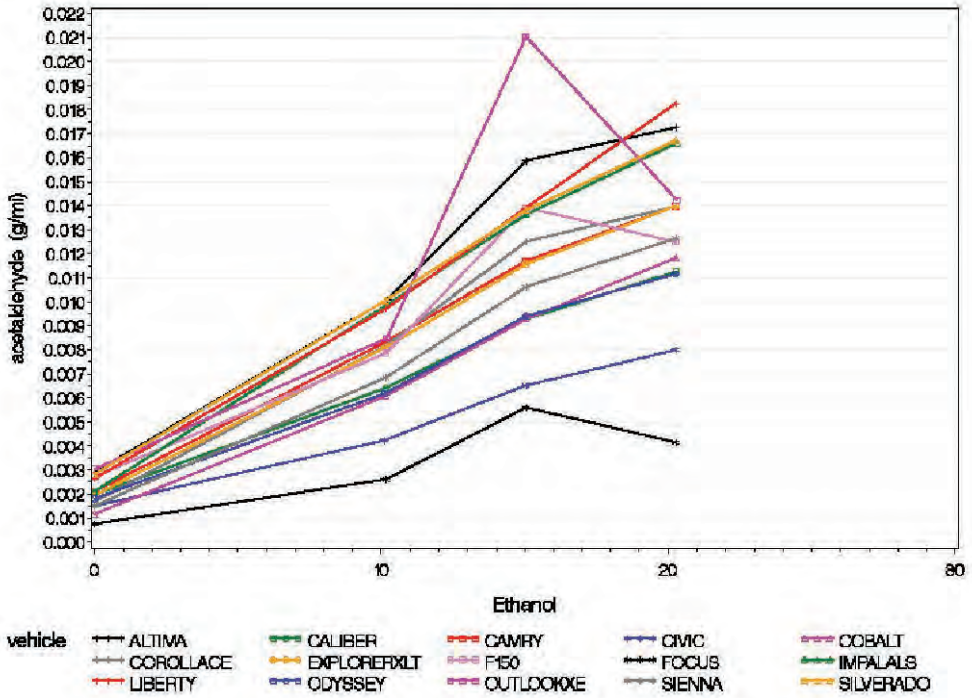


Figure 92. Linear-effects plots for Acetaldehyde (g/mi) and ln(Acetaldehyde) vs. ethanol content (%) (Bag 1, full design); data are averaged by four ethanol levels and by vehicle.



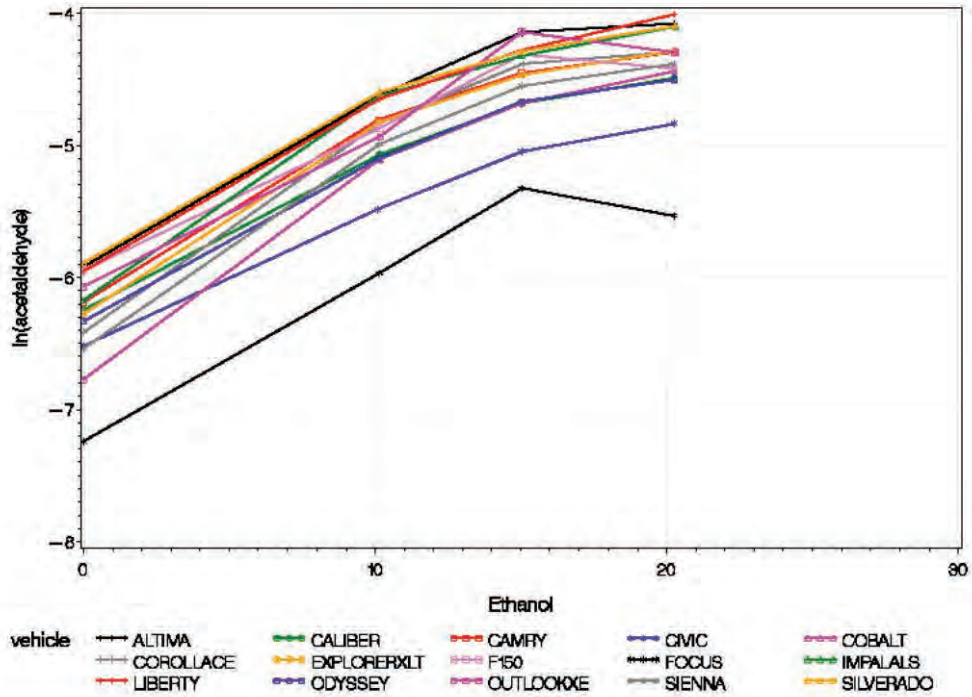
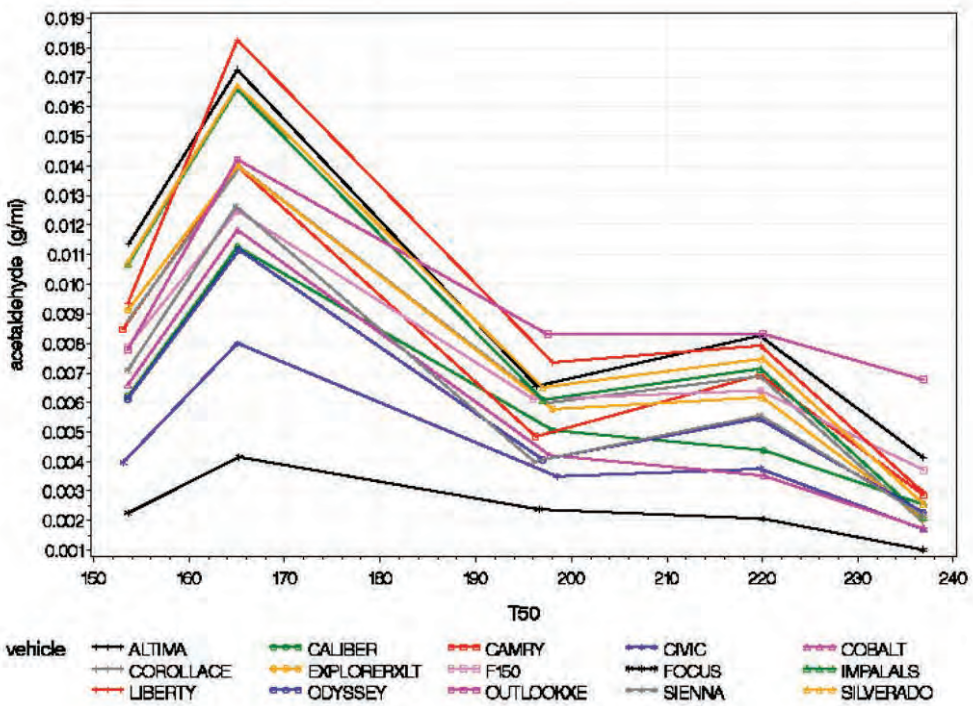


Figure 93. Linear-effects plot for Acetaldehyde and ln(Acetaldehyde) vs. T50 (°F) (Bag 1, full design); data are averaged by five T50 levels and by vehicle.



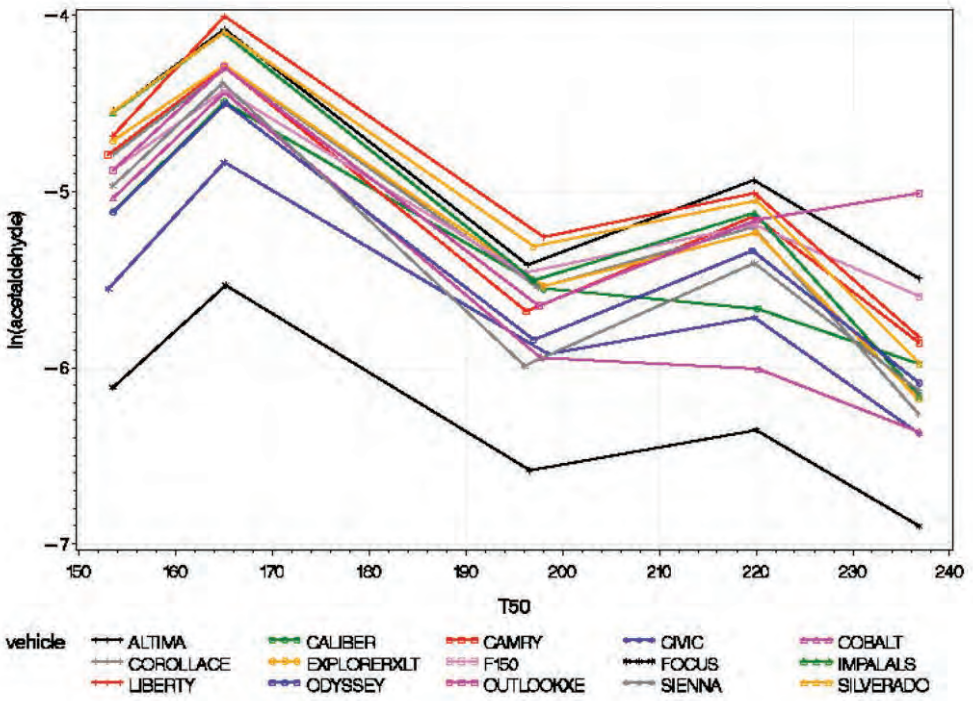
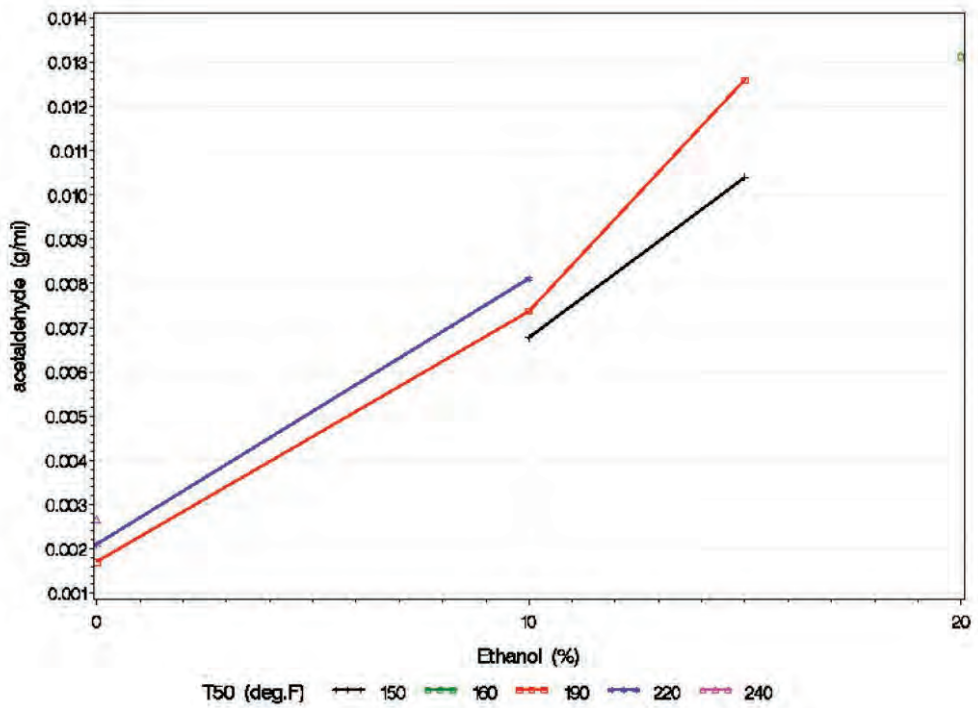


Figure 94. Interaction plots of Acetaldehyde and ln(Acetaldehyde) vs. ethanol, by T50 level; data averaged by four ethanol and five T50 levels.



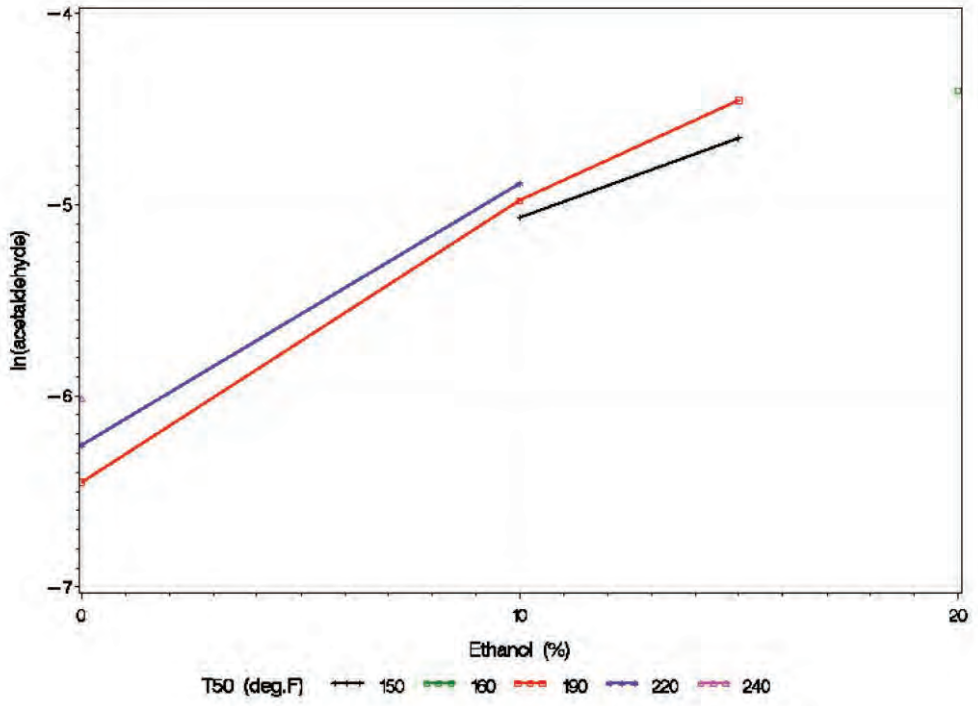
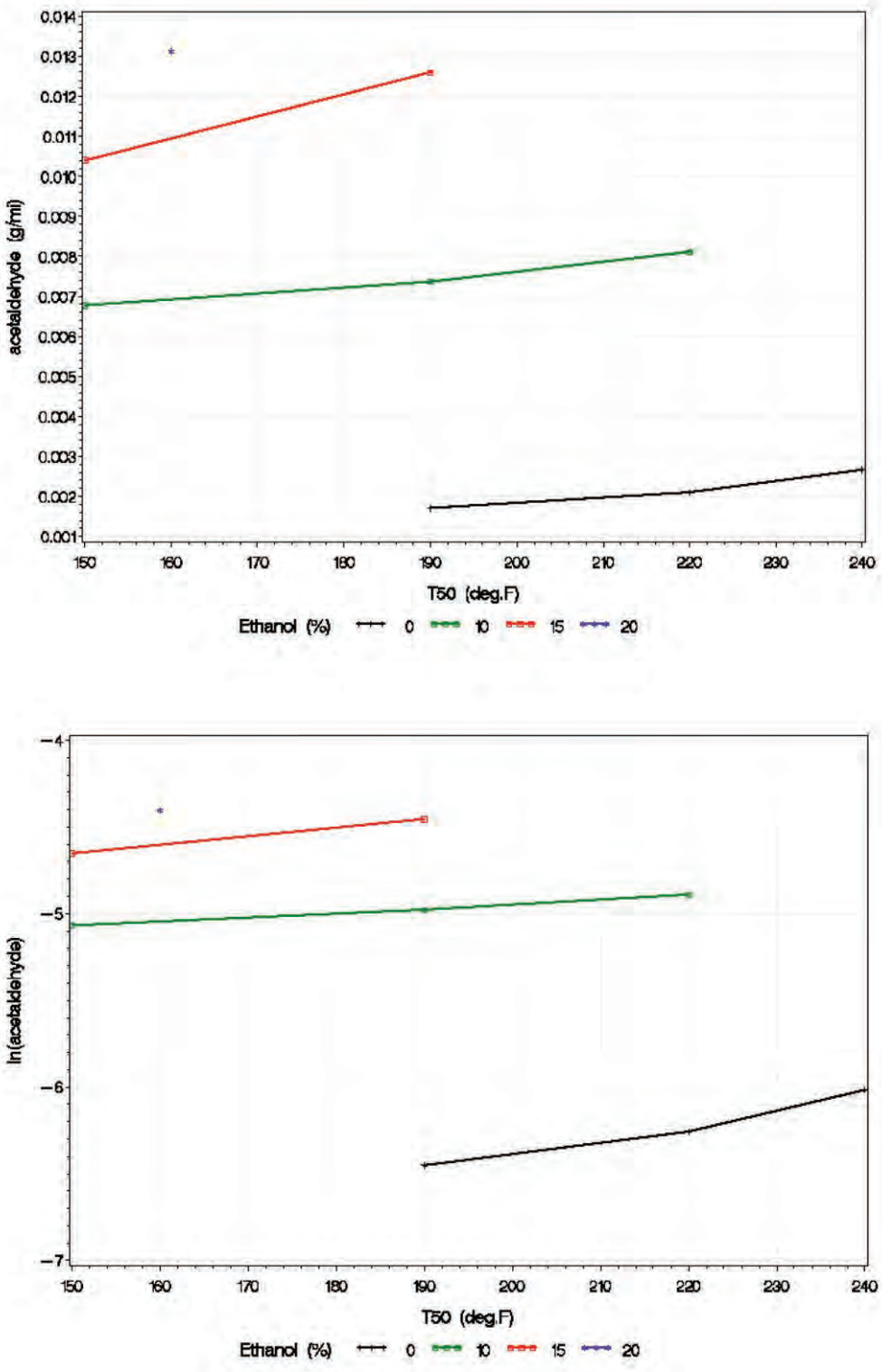


Figure 95. Interaction plots of Acetaldehyde and ln(Acetaldehyde) vs. T50, by Ethanol level; data averaged by five T50 and four Ethanol levels.



8.7 Model Fitting

In all cases the response variable was the natural logarithm of estimated emissions measurements, ($\ln \hat{Y}_i$), and the predictors were the standardized or doubly standardized fuel properties, as described above (see Equation 4 and Equation 7, page 29). The vehicles were included as class variables, as described below. The predictors considered for inclusion in models are listed in Table 80, along with the notation used to identify them.

Table 80. Description and notation for parameters included in model fitting.

Fuel Parameter	Model term	In optimized design	Notation	Standardization[1]
Ethanol content (%)	etOH	YES	Z_e	One-stage
Aromatics content (%)	Arom	YES	Z_a	One-stage
RVP (psi)	RVP	YES	Z_r	One-stage
T50(°F)	T50	YES	Z_5	One-stage
T90 (°F)	T90	YES	Z_9	One-stage
	etOH × etOH	YES	ZZ_{ee}	Two-stage
	T50 × T50	YES	ZZ_{55}	Two-stage
	etOH × Arom	YES	ZZ_{ea}	Two-stage
	etOH × RVP	YES	ZZ_{er}	Two-stage
	etOH × T50	YES	ZZ_{e5}	Two-stage
	etOH × T90	YES	ZZ_{e9}	Two-stage
	Arom × RVP	NO	ZZ_{ar}	Two-stage
	Arom × T50	NO	ZZ_{a5}	Two-stage
	Arom × T90	NO	ZZ_{a9}	Two-stage
	T90 × T90	NO	ZZ_{99}	Two-stage
	T50 × T90	NO	ZZ_{59}	Two-stage
	RVP × T90	NO	ZZ_{r9}	Two-stage

[1] For one-stage standardization, see Equation 5, for two-stage standardization, see Equation 6 and Equation 7.

For each compound and bag, the procedures for model-fitting varied, depending on the degree of censoring, and whether media contamination applied.

With respect to censoring, the following rule was applied. If the number of censored measurements was ≤ 5, we substituted the smallest measured positive value for the missing values, and proceeded with model fitting, using a mixed-model approach. However, if the number of censored measurements was > 5, we fit a model using Tobit regression (i.e., "censored normal regression"), an established technique for analysis of left-censored datasets. For compounds affected by media contamination, we integrated the approach to censoring with an "Estimated Dependent Variable Model," an approach to modeling datasets with measurement uncertainty in the response variable. Table 81 summarizes the modeling approaches used, by compound and bag. Note that if the level of censoring was considered too severe to allow for model fitting, "modeling approach" is assigned as "none."

Table 81. Summary of Modeling approaches, by Compound and Bag.

Compound	Media contamination	Bag 1			Bag 2		
		No. censored values	Modeling approach	EDV	No. censored values	Modeling approach	EDV
Acetaldehyde	YES	≤5	mixed	YES	≤5	mixed	YES
Formaldehyde	YES	≤5	mixed	YES	≤5	mixed	YES
Acrolein	YES	>5	Tobit	YES	>5	none	
Ethanol	YES	>5	Tobit	YES	>5	Tobit	YES
Benzene	NO	≤5	mixed	NO	>5	none	
1,3-butadiene	NO	≤5	mixed	NO	>5	none	
Ethane	NO	≤5	mixed	NO	≤5	mixed	NO

8.7.1 The Estimated Dependent Variable Model (EDV)

This approach has been developed for situations in which measurement uncertainty plays a substantial role in generating the set of values for the response variable[18]. After estimating measurements and variances as described above, development of the EDV involves two steps. The first step is to fit a preliminary model of the compound in terms of the fuel properties. For this purpose, we fit "full" models, containing all candidate fuel properties. The full models were the "17-term" full for the full design, and the "4-term" full model for the reduced design. In addition, the model included a dummy variable for each vehicle.

We solved the model by least squares (using Proc Reg in SAS 9.2) and obtained simple residuals r_i. Using the residuals, we re-estimated the random error ($\hat{\sigma}_\varepsilon^2$) using the following expression (Equation 24),

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_i r_i^2 - \sum_i \hat{\sigma}_k^2 + tr\left(\left(\mathbf{X}`\mathbf{X}\right)^{-}\mathbf{X}`diag\left(\boldsymbol{\sigma}_k^2\right)\mathbf{X}\right)}{n-p-1} \qquad \textbf{Equation 24}$$

where n = the total number of measurements (on all fuels and vehicles) and p = the total number of parameters in the model, including the intercept. After re-estimation of the random error, we recombined it with the variance of the media contamination to calculate a set of "variance-based" weights, as shown in Equation 25.

$$w_i = \frac{1}{\sqrt{4\hat{\sigma}_k^2 + \sigma_\varepsilon^2}} \qquad \textbf{Equation 25}$$

In calculation of the weights, the variance of the media contamination is multiplied by 4.0, to account for the four times that the contamination can affect the total measurement, for the two

sets of two media used to measure sample and background, respectively (Equation 21). When applicable, the weights w_i were applied to all subsequent models during the model-fitting process, for compounds affected by media contamination.

8.7.2 Fitting by Backwards Elimination

The process started with the appropriate full model (17-term or 4-term), and proceeded by backwards elimination. In each step, one or more parameters were removed, and the model was refit. Models with one or more terms removed will be referred to as "reduced models." Terms were selected for removal based on the *p*-value for their respective *t*-test of significance ($p > 0.10$), starting with parameters with the highest *p*-values.

At each step, we tested the goodness-of-fit of each reduced model against that of the full model using a likelihood ratio test. In addition, each successive reduced model was tested for goodness-of-fit against the preceding reduced model. At each step, if the current reduced model was not a significantly poorer fit than its predecessor, it was accepted as the current "best fit." To interpret the goodness-of-fit test, the current reduced model was considered a poorer fit than its predecessor if the *p*-value for the likelihood ratio test was < 0.10. The process was repeated until the current reduced model was a significantly poorer fit than its predecessor.

In performing the likelihood ratio tests, it was necessary that the two models included in the test be "nested," i.e., that both models have all terms in common except the subset of terms whose inclusion is the subject of the test. This condition always applied, in that all reduced models were nested within full models, and each reduced model was nested within the preceding reduced model. For a specific test, the model with more parameters is designated as the "reference" model, and the model with fewer parameters as the "nested" model. The test was fit in standard fashion, using the log-likelihood statistics output as the primary fit statistics for the models (all models were fit by maximum likelihood estimation). The test statistic is calculated as the difference in the -2log-likelihood (-2ln*L*) between the nested and reference models, and which is assumed to be distributed as a χ^2 statistic with *d* degrees of freedom, where *d* is the difference in the numbers of parameters between the two models ($p_{\text{ref}} - p_{\text{nested}}$).

$$\chi^2_{test} = -2\ln\left(\frac{L_{\text{nested}}}{L_{\text{reference}}}\right) = -2\ln L_{\text{nested}} - \left(-2\ln L_{\text{reference}}\right) \sim \chi^2_d \qquad \textbf{Equation 26}$$

The test was considered significant if the *p*-value was less than 0.10.

8.7.3 Mixed models

For the compounds and bags indicated in Table 81, we fit mixed models. Fuel properties were treated as continuous numeric variables and assigned as fixed factors; each vehicle was treated as a class variable and assigned as a random factor. We fit the model as a random coefficients model, in which the random effect is a random intercept fit for each vehicle. However, we did

not attempt to fit random slope coefficients for individual vehicles. For the speciated compounds, the mixed model structure was applied as described in 5.3.1 (page 97).

Table 82 shows the set of nested models fit for Bag 1 Acetaldehyde. The model including all potential candidate terms, both those included in the optimized design, plus additional terms, is denoted as the "Full model." The sequence of models fit by sequentially removing terms from the full is listed in order. The reduced models are identified by the number of terms removed from the full. For example the first model fit by removing the aromatics×T50 interaction from the full is denoted as "Full minus 1," abbreviated as "FM1." In this case, eight reduced models were fit, from "FM1" to "FM8," with each successive model nested in its predecessor. In this case, it is interesting to note that only one parameter not in the optimized design is included in final four models (FM5-FM8).

For each model listed in Table 82, Table 83 portrays the fit statistics and tests of fit. The table shows the "fitting history" of the model, beginning with the full model and ending with the selection of the reduced model giving the "best fit." In this context, the best fit model represents the model giving the best statistical fit to the EPAct Phase-3 dataset, under the assumptions and procedures adopted and implemented during the model-fitting process, as described. In the block containing "fit parameters" the table includes the number of terms p (including the intercept), the -2 log likelihood ($-2\ln L$), which represents the basic fit criterion from the maximum-likelihood fitting procedure, and the Bayesian Information Criterion (BIC), as reported by the MIXED procedure. Based on these parameters, likelihood-ratio tests of fit were conducted at each fitting step (Equation 26). The first block shows results of tests conducted using the full model as the reference model, and each successive reduced model as the nested model. The first column, the "deviation," represents the difference in the $-2\ln L$ statistics between the nested and reference models. The second column, "d" represents the difference in the number of terms between the reference and nested models; this value increases by 1 for each successive model. The third column represents the p-value for the test statistic, given as a chi-square statistic with d degrees of freedom (χ_d^2). The third block of the table also contains results of likelihood-ratio tests, with each model tested not against the full but against its immediate predecessor in the series. Within this block, the deviation, d and p-value are calculated in the same way, except that the value of d is 1 for all tests.

In this procedure, the BIC and "tests-against-full" are included for completeness and as corroborating information. However, the "test-against-previous" is governing for selection of the best-fit model. Note that the test-against-full qualifies as the test-against-previous for the first reduced model (FM1). At each step, the models were fit, and the tests conducted. If the p-value for the test-against-previous is greater than the critical value, the null hypothesis of no significant difference in fit between the reference and nested models is retained, and the nested model is retained as the current best fit. This process was repeated to the point of "one step past best" i.e., to the point at which the test-of-fit was clearly significant. When this result is obtained, the null hypothesis of no difference in fit is rejected in favor of the alternative

hypothesis of a significant difference in fit, and interpreted to mean that dropping the last model term led to a decrease in goodness-of-fit. The last model with an insignificant result is selected as the "best-fit."

In this example, the BIC and tests-of-fit generally corroborate closely, in that the BIC drops steadily from the full to FM7, and then increasing at the end of the series. However, Table 83 shows an apparent anomaly at step "FM6." For FM6, the BIC increases slightly relative to FM5, suggesting a decrease in fit, and the test-against-previous is marginally significant, suggesting that FM6 might qualify as the best fit. However, when pushing "one-step-past," and fitting FM7, the BIC drops again to a value lower than that for FM5, and the test-against-previous again exceeds the critical value. Seeing this result, we proceeded to an additional step, fitting FM8. For FM8, the BIC increases sharply, and the test-against-previous is well below the critical value, indicating a significant decline in fit. Based on these results, the reduced model FM7 was selected as the best fit.

For the full and best-fit models, Table 84 shows the coefficients and type-III tests of effect. For the most part, coefficients for terms in both models are similar in value. Exceptions include the T50 (Z_5), etOH×aromatics (ZZ_{ea}) and the T50×T90 (ZZ_{59}) terms, which change by margins of -13%, -39% and 45%, respectively. These changes might be explainable in terms of the dropping of interactions involving etOH, aromatics, T50 and T90, including etOH×T50, aromatics×T50, etOH×T90 and aromatics×T90. In general, though, the relative stability of the other coefficients suggests that the process of standardization induces the model terms to act as though effectively independent, as required by the multiple regression model. For all parameters retained in the best-fit, all standard errors are lower than corresponding values in the full, suggesting that model-fitting has resulted in improved precision of estimation for these parameters. In contrast, note that all terms dropped during model fitting had high type III *p*-values in the full, with the exception of RVP×T90, which has a low *p*-value in the full, but a higher value at step FM5, when it was dropped (0.087). Despite the fact that this type-III *p*-value is less than 0.10, the decision to drop this term is based on the corresponding tests-of-fit between FM6 and FM7, as previously described, not the type-III test at FM5. Another result of interest is that for this model, the principal of hierarchy is maintained with full consistency, in that all linear terms are highly significant, as are all included interactions. In reviewing these models, an advantage of the standardization is that the coefficients for different terms can be compared in terms of magnitude. When reviewing coefficients, then, a striking result is that the effects for etOH (Z_e) and its squared term (ZZ_{ee}) are by far the strongest effects in the model; this result is not surprising, however, after having reviewed Figure 92 through Figure 95.

Table 82. Models fit for Bag-1 Acetaldehyde

Model term	Notation	Model								
		Full	FM1[1]	FM2	FM3	FM4	FM5	FM6	FM7	FM8
etOH	Z_e	●	●	●	●	●	●	●	●	●
Arom	Z_a	●	●	●	●	●	●	●	●	●
RVP	Z_r	●	●	●	●	●	●	●	●	●
T50	Z_5	●	●	●	●	●	●	●	●	●
T90	Z_9	●	●	●	●	●	●	●	●	●
etOH × etOH	ZZ_{ee}	●	●	●	●	●	●	●	●	●
T50 × T50	ZZ_{55}	●	●	●	●	●	●	●	●	●
etOH × Arom	ZZ_{ea}	●	●	●	●	●	●	●	●	×
etOH × T50	ZZ_{e5}	●	●	●	×					
etOH × T90	ZZ_{e9}	●	●	●	●	●	×			
etOH × RVP	ZZ_{er}	●	●	●	●	●	●	●	●	●
Arom × RVP	ZZ_{ar}	●	●	●	●	×				
Arom × T50	ZZ_{a5}	●	×							
Arom × T90	ZZ_{a9}	●	●	×						
T90 × T90	ZZ_{99}	●	●	●	●	●	●	●	×	
T50 × T90	ZZ_{59}	●	●	●	●	●	●	●	●	●
RVP × T90	ZZ_{r9}	●	●	●	●	●	●	×		
1 Represents "Full minus 1," etc.										

Table 83. Fitting history for Bag-1 Acetaldehyde – with "FM7" selected as best fit model.

Fit Parameters				*Test with respect to Full*			*Test with respect to Previous Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.[1]	*d*	Pr>χ^2	Dev.	*d*	Pr>χ^2
Benchmark	18	-31.467	22.69						
FM1	17	-31.389	20.06	0.078	1	0.78			
FM2	16	-30.460	18.29	1.007	2	0.60	0.929	1	0.34
FM3	15	-29.266	16.77	2.201	3	0.53	1.194	1	0.27
FM4	14	-27.873	15.46	3.594	4	0.46	1.393	1	0.24
FM5	13	-25.898	14.72	5.569	5	0.35	1.975	1	0.16
FM6	12	-22.970	14.94	8.497	6	0.20	2.928	1	0.09
FM7	**11**	**-20.561**	**14.64**	**10.906**	**7**	**0.14**	**2.409**	**1**	**0.12**
FM8	10	-14.980	17.52	16.487	8	0.03	5.581	1	0.02

[1] A lower value indicates a better fit.

[1] The deviation is the difference in the -2loglik statistics for the nested and reference models, respectively, per Equation 13.

Table 84. Acetaldehyde (Bag 1): Coefficients and Type-III Tests of Effect for the Full and Best-Fit Models.

Effect	*Full Model*					*Best-Fit Model (FM7)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr>*t*	Estimate	Std.Err.	d.f.	*t*-value	Pr>*t*
Intercept	-5.2324	0.08802	15	-59.4	0.000000	-5.2323	0.08785	15	-59.6	0.000000
Z_e	0.8250	0.01297	898	63.6	0.000000	0.8145	0.01020	898	79.9	0.000000
Z_a	0.03999	0.009279	898	4.31	0.000018	0.03484	0.008249	898	4.22	0.000027
Z_r	-0.03667	0.01297	898	-2.83	0.0048	-0.04170	0.008833	898	-4.72	0.000003
Z_5	0.09927	0.01826	898	5.44	0.000000	0.08670	0.01063	898	8.16	0.000000
Z_9	0.04235	0.01115	898	3.80	0.00016	0.03801	0.007764	898	4.90	0.000001
ZZ_{ee}	-0.1716	0.01548	898	-11.09	0.000000	-0.1669	0.007849	898	-21.3	0.000000
ZZ_{55}	0.07115	0.01314	898	5.42	0.000000	0.06665	0.007993	898	8.34	0.000000
ZZ_{ea}	0.03016	0.01304	898	2.31	0.021	0.01840	0.007777	898	2.37	0.018
ZZ_{er}	0.02020	0.008769	898	2.30	0.021	0.02194	0.007845	898	2.80	0.0053
ZZ_{e5}	-0.01614	0.01673	898	-0.965	0.33					
ZZ_{e9}	-0.01486	0.01072	898	-1.39	0.17					
ZZ_{ar}	0.01738	0.01618	898	1.07	0.28					
ZZ_{a5}	0.004828	0.01729	898	0.28	0.78					
ZZ_{a9}	0.008759	0.008852	898	0.99	0.32					
ZZ_{99}	0.01270	0.01503	898	0.84	0.40					
ZZ_{59}	0.02718	0.01132	898	2.49	0.013	0.03959	0.008256	898	4.80	0.000002
ZZ_{r9}	-0.0206	0.009971	898	-2.07	0.039					
σ^2_{veh}	0.1154					0.1149				
σ^2_{ε}	0.08743					0.08850				

8.7.4 Tobit regression

For compounds and bags with high levels of censoring, we fit "censored normal regression," or "Tobit" models, a technique commonly used for left-censored data[19,20]. We fit the models using the LIFEREG procedure in SAS 9.2, as applied for left censoring.

As with the mixed models, the procedure solves for the model parameters using maximum likelihood estimation. However, the Tobit approach does not attempt to estimate the missing values. Rather, the formulation of the maximum likelihood function (L) is modified so as to compensate for the absence of the censored values and to estimate values for the model coefficients accordingly. In the Tobit model, each measurement is represented by its probability density (standard normal), given an assumed set of parameters, and each censored value is represented by the cumulative probability that the value would be less than the effective censoring level.

As the LIFEREG procedure is not able to handle random factors, it was necessary to enter each vehicle as a fixed factor, represented as a dummy variable. Thus, the model outputs an intercept for each vehicle, and an estimate of random error variance. It does not estimate a component of variance representing the between vehicle variability (the variance of the random intercepts). This step must be performed manually, as described below.

In cases when the EDV applied, the model incorporated weights estimated from the preliminary model, as described above, except that the preliminary model was run as a Tobit model, rather than as a least-squares model. In all other respects, estimation of the weights was identical.

The procedure outputs the log likelihood (L) as a goodness-of-fit parameter. It does not output an estimate of the BIC, but the BIC is readily calculated from L, the number of model terms p and the total number of (non-missing) observations n, as

$$\mathrm{BIC} = -2\ln L + p\ln n \qquad \textbf{Equation 27}$$

Using these parameters, the backwards elimination model-fitting was performed as with the mixed models, as previously described, with the exception that we did not use the Type-III tests of effect output by the procedure, because vehicles were not appropriately treated as random factors. To compensate, we estimated standard errors numerically, by applying jackknife repeated replication (JRR), using vehicle as the sampling unit. The procedure is simple. With n_{veh} vehicles included in the sample, we ran the model n_{veh} times, excluding one vehicle in each run, and saving the coefficients from each run. In each replicate, the remaining vehicles as assigned as weight w_{boot}, calculated as

$$w_{\mathrm{boot}} = \frac{n_{\mathrm{veh}}}{n_{\mathrm{veh}} - 1} \qquad \textbf{Equation 28}$$

Note that in cases when the EDV was incorporated, the entire process was repeated for each replicate model, including the fitting of the preliminary model, and estimation of variance-based weights. After fitting the preliminary model, a final "jackknife weight" w_{jk} was calculated as the product of the variance-based and sampling weights w_i and w_{boot} (Equation 29).

$$w_{\mathrm{jk}} = w_i w_{\mathrm{boot}} \qquad \textbf{Equation 29}$$

After fitting all replicates, the variance of estimation for each of the parameter estimates was estimated by calculating sample variances on the set of replicate coefficients,

$$s^2_{\hat{\beta}} = \frac{n_{\mathrm{veh}} - 1}{n_{\mathrm{veh}}} \sum_{j=1}^{n_{\mathrm{veh}}} \left(\hat{\beta}_j - \hat{\beta}_0\right)^2 \qquad \textbf{Equation 30}$$

where $\hat{\beta}_j$ is the parameter estimate for replicate j and $\hat{\beta}_0$ is the parameter estimate from a model fit using all vehicles. The square roots of these variances gave standard errors that served to develop tests of effect used to guide the backwards elimination process. Test statistics were calculated in the typical fashion as

$$t_{\text{test}} = \frac{\hat{\beta}}{s_{\hat{\beta}}} \qquad \textbf{Equation 31}$$

with corresponding two-tailed *p*-values estimated by

$$p = 2\left(1 - \Pr\{t > |\, t_{\text{actual}}\,|\}\right) \qquad \textbf{Equation 32}$$

where the *t*-statistics are taken from a *t* distribution with n_{veh} degrees of freedom. After calculating the tests of effect, the model fitting procedure proceeded as with the mixed models.

Below, we present the results of a Tobit model fitting process, using Bag 1 acrolein as an example. Table 85 shows the models fit for Bag 1 acrolein. The best-fit model, FM8, has nine terms, five linear effects plus one quadratic term and three interactions. As with acetaldehyde, only one term outside the 11 design parameters (T50×T90) remains in the best-fit model. For this compound, we fit a model containing only the five linear effects, to examine the adequacy of a model without any 2nd-order terms.

The model fitting process is illustrated in Table 86. The format is identical to that for acetaldehyde, using parameters calculated for the Tobit model as described above. Again, the sequence shows the improvement in fit from the full model to the best-fit (FM8), through a steady decline in the BIC. What is striking is that the linear-effects model gives a dramatic loss of fit in relation to any of the other models, including the full model. Table 87 shows the coefficients and tests of effect for the full and best-fit models, with the standard errors and *t*-statistics based on the jackknife replication process described previously.

As mentioned, because the vehicles were entered into the Tobit model as fixed effects, the procedure could not calculate a grand intercept for all vehicles, as does the mixed model procedure. It was thus necessary and straightforward to calculate the grand intercept as the mean of intercepts for all vehicles. As is typical with dummy variables, the intercept reported by the model represented the intercept for the last vehicle entered, and the intercepts for the remaining vehicles represent differences between their intercepts and those for the reference vehicle, Accordingly, the grand intercept was calculated by adding the intercept differences for all vehicles ($\Delta\beta_{0,j}$) to the reference vehicle intercept ($\beta_{0,\text{ref}}$), and averaging (Equation 33).

$$\hat{\beta}_0 = \frac{\sum_{j=1}^{n_{\text{veh}}} \left(\Delta\hat{\beta}_{0,\,j} + \hat{\beta}_{0,\text{ref}}\right)}{n_{\text{veh}}} \qquad \textbf{Equation 33}$$

Similarly, to calculate a variance to represent the random covariance parameter fit by the mixed model for vehicle, we substituted a simple variance of the vehicle intercepts, calculated as

$$V(\hat{\beta}_0) = \frac{\sum_{j=1}^{n_{\text{veh}}} \left(\left(\Delta\hat{\beta}_{0,j} + \hat{\beta}_{0,\text{ref}} \right) - \hat{\beta}_0 \right)^2}{n_{\text{veh}} - 1}$$

Equation 34

This result is recorded as σ^2_{veh} in the table, along with the random error variance σ^2_{ε} estimated by the LIFEREG procedure.

Finally, to illustrate the results of the jackknife replication procedure, Figure 96 shows cumulative distributions of coefficients for each jackknife replicate for the five linear effects. The distributions do not appear to be entirely symmetrical, showing somewhat shortened upper tails. The distribution for ethanol (Z_e) shows a shortening, or clumping in the upper tail, as though three vehicles were influential in raising the coefficient. The distributions for aromatics (Z_a),T50 (Z_5), and T90 (Z_9) are similar in that they show noticeable lengthening in the lower tail, suggesting that 2-3 vehicles may be influential in decreasing the values of the coefficients. The distribution for RVP (Z_r) is the most symmetric, although the lower tail is slightly longer than the upper.

Table 85. Models fit for Acrolein, Bag 1 (all models include an intercept term)

Model term	Notation	Model				
		Full	FM4	FM7	FM8	Linear Effects
etOH	Z_e	●	●	●	●	●
Arom	Z_a	●	●	●	●	●
RVP	Z_r	●	●	●	●	●
T50	Z_5	●	●	●	●	●
T90	Z_9	●	●	●	●	●
etOH × etOH	ZZ_{ee}	●	●	●	●	×
T50 × T50	ZZ_{55}	●	●	●	×	
etOH × Arom	ZZ_{ea}	●	●	×		
etOH × RVP	ZZ_{er}	●	●	×		
etOH × T50	ZZ_{e5}	●	●	●	●	×
etOH × T90	ZZ_{e9}	●	●	●	●	×
Arom × RVP	ZZ_{ar}	●	●	×		
Arom × T50	ZZ_{a5}	●	×			
Arom × T90	ZZ_{a9}	●	×			
T90 × T90	ZZ_{99}	●	×			
T50 × T90	ZZ_{59}	●	●	●	●	×
RVP × T90	ZZ_{r9}	●	×			

[1] Indicates "Full minus 4," etc.

Table 86. Acrolein (Bag 1): Model Fitting History (FM8 selected as best-fit model).

Fit Parameters				*Test with respect to Full Model*			*Test with respect to Previous Model*		
Model	*p*	-2ln*L*	BIC[1]	Dev.[1]	*d*	Pr>χ^2	Dev.	*d*	Pr>χ^2
Full	18	1105.40	1227.52						
FM4	14	1106.84	1201.82	1.4446	4	0.8364			
FM7	11	1111.16	1185.79	5.7634	7	0.5676	4.3188	3	0.2290
FM8	10	1114.31	1182.15	8.9122	8	0.3498	3.1488	1	0.0760
Linear Effects	6	1264.37	1306.08	159.9748	12	0.0000	151.0627	4	0.0000

[1] A lower value indicates a better fit.

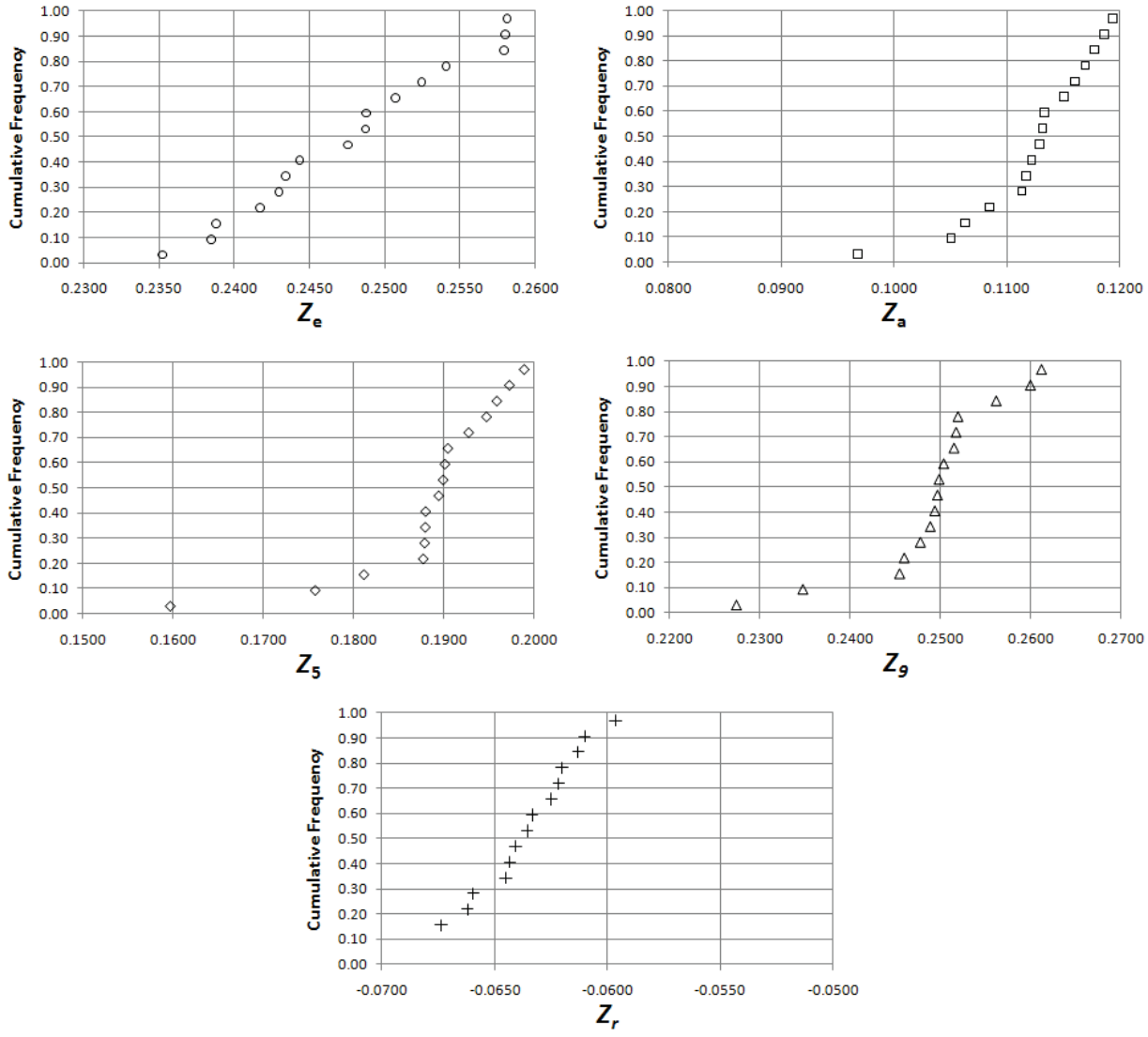
[1] The deviation is the difference in the -2ln*L* statistics for the nested and reference models, respectively, per Equation 13.

Table 87. Acrolein (Bag 1): Coefficients and Tests of Effect for the Full and Best-Fit Models.

Effect	*Full Model*					*Best-Fit Model (FM8)*				
	Estimate	Std.Err.	d.f.	*t*-value	Pr>*t*	Estimate	Std.Err.	d.f.	*t*-value	Pr>*t*
Intercept[1]	-7.9337					-7.9338				
Z_e	0.2571	0.02638	15	9.74	0.000000	0.2476	0.02738	15	9.04	0.000000
Z_a	0.1149	0.02128	15	5.40	0.000074	0.1122	0.02184	15	5.14	0.00012
Z_r	-0.05815	0.01799	15	-3.23	0.0056	-0.0645	0.01364	15	-4.73	0.00027
Z_5	0.1979	0.03123	15	6.34	0.000013	0.1881	0.03554	15	5.29	0.000091
Z_9	0.2465	0.02979	15	8.28	0.000000	0.2488	0.03125	15	7.96	0.000000
ZZ_{ee}	-0.06009	0.01880	15	-3.20	0.0060	-0.08306	0.01392	15	-5.97	0.000026
ZZ_{55}	0.02735	0.01709	15	1.60	0.13					
ZZ_{ea}	0.01716	0.01838	15	0.93	0.37					
ZZ_{er}	0.01253	0.01404	15	0.89	0.39					
ZZ_{e5}	-0.09661	0.02096	15	-4.61	0.00034	-0.1185	0.02415	15	-4.91	0.00019
ZZ_{e9}	0.04178	0.01618	15	2.58	0.021	0.04618	0.01120	15	4.12	0.00091
ZZ_{ar}	0.02002	0.01562	15	1.28	0.22					
ZZ_{a5}	0.01127	0.01822	15	0.62	0.55					
ZZ_{a9}	-0.007484	0.01726	15	-0.43	0.67					
ZZ_{99}	0.0004162	0.01481	15	0.028	0.98					
ZZ_{59}	0.06274	0.01552	15	4.04	0.0011	0.05985	0.01271	15	4.71	0.00028
ZZ_{r9}	0.0002551	0.01709	15	0.015	0.99					
σ^2_{veh} [1]	0.3633					0.3629				
σ^2_{ε}	0.03206					0.3213				

[1] Not fit by the model, but manually recalculated from intercepts for individual vehicles.

Figure 96. Acrolein (Bag 1): Cumulative Distributions of Coefficients for the five Linear Effects (Each data point represents a single jack-knife replicate).



9 Summary and Conclusions

9.1 Regulated Emissions, Total Hydrocarbons and Methane

9.1.1 Modeling Results

To aid in gaining a broad view of the results of this project, the sets of coefficients presented above in tabular form are also presented in this section in the form of bar charts, in which the bars represent both the magnitude and sign of the coefficients. It is important to note that the coefficients presented are those for one- or two-stage standardized terms. As a reminder of this critical step, we have retained the "Z_x" notation for linear terms and the "ZZ_{xy}" notation for quadratic and linear terms in all tables and figures throughout. Thus, all fuel properties are centered on means of 0.0 and expressed in terms of their own standard deviations (which take the value of 1.0). While more abstract, this approach has the advantage of keeping coefficients for different effects comparable in terms of magnitude, allowing direct comparison of effects in terms of importance. Such direct comparisons would not be possible were the terms scaled in terms of each of the properties' units (vol.%, psi, °F).

Reduced Models: As in the tables, results for reduced models fit starting with the 11-term full model and the 16-term design model are presented. For convenience, these reduced models will be referred to throughout as the $\underline{\text{reduced}}_{11}$ and $\underline{\text{reduced}}_{16}$ models, respectively.

9.1.1.1 Hydrocarbons (THC, NMOG, NMHC and CH_4)

Coefficients for reduced models for the hydrocarbon species are presented graphically in Figure 97 to Figure 100 for THC, NMOG, NMHC and CH_4, respectively. The plots include results for Bags 1-3 and constitute a graphic presentation of the tabular results previously shown in 7.2, (page 137 ff.). In addition, Figure 101 presents coefficients for the three species juxtaposed on single graphs for each bag, to facilitate comparisons among species, as described below.

Cold-Start Emissions. For Bag 1 results, the sets of coefficients are broadly similar for THC, NMOG and NMHC . For the linear effects, all coefficients are positive, except for RVP. In terms of absolute magnitude, the most important linear effect is T50, followed by aromatics. Ethanol is ranked third, and RVP fourth, except for NMHC, for which RVP is third and ethanol fourth. T90 is consistently the smallest term. NMOG has the largest etOH effect, and NMHC the smallest, whereas NMHC has the largest aromatic and T50 effects, and THC the smallest. Both quadratic terms are included in all three models and have similar magnitudes. As the signs of the etOH×etOH and T50×T50 quadratic terms are both positive, they impart upwards curvature to their respective linear trends. In addition, all three models include the same three interaction terms retained from the design model (etOH×arom, etOH×T50 and etOH×T90); the magnitudes of the interactions are also similar, but with NMOG and NMHC coefficients slightly larger than those for THC, although the differences do not appear statistically significant. As the signs of these interactions are all positive, as are all three linear terms, these interactions qualify as

reinforcement effects. All three models also retain the same three interaction terms from among the five additional terms in the extended model; all three effects are positive and similar in size in all three models.

In Bag 1, a different pattern is evident for CH_4. The coefficient for aromatics differs in size from those for the other three species, as well as in sign, with a strong negative aromatics coefficient as the most striking feature of the CH_4 model. The ethanol effect is similar to that for NMOG, whereas the T50 and RVP effects are just over half the size of their counterparts. The CH_4 model does not retain a linear effect for T90. The CH_4 model retains both quadratics terms, with magnitudes equal in sign but about half the size of those for the other species. The pattern for the interactions in the $reduced_{11}$ model is similar, except that CH_4 retains the etOH ×RVP interaction rather than the etOH×T90 interaction. The $reduced_{16}$ model retains four of the five additional interactions; all effects are positive and similar in magnitude.

Hot Running Emissions. For Bag 2 results, the pattern is not as consistent among THC, NMOG and NMHC, although similarities remain. The effects describing bulk properties (RVP, T50, T90) have larger effects than those describing fuel composition (etOH, Arom), although differences are not always large. T90 is the largest effect for all three species, with T50 and RVP ranked second and third for THC and NMOG. NMHC differs slightly, in that this ranking is reversed. Ethanol and aromatics effects are ranked fourth or fifth. Again, NMHC differs in that an ethanol effect is not retained in the $reduced_{11}$ model. In the $reduced_{16}$ model, the ethanol linear effect is not significant but is retained on the strength of two interactions involving ethanol. With respect to aromatics, the THC model differs in that the aromatics coefficient is negative, rather than positive, as for the other two species. The T50×T50 quadratic term is retained in all three models, but differs in size, with the THC coefficient smaller than those for NMOG and NMHC. The model for NMOG retains two interactions from among the terms in the design model, that are not included in the THC or NMHC models. When considering all interactions, including those from the extended model, the set of interactions retained is similar. All three models retain the etOH×Arom, etOH×RVP and arom×T50 interactions, which are positive, negative and positive in sign, respectively. Additionally, the NMOG and NMHC models retain the etOH×T50 interaction. However, this term is marginally significant.

In Bag 2, the pattern again differs for the CH_4 model. The fuel composition terms are largest, followed by those for the bulk properties. Aromatics has the largest coefficient, although negative in sign, followed by ethanol, which is positive. The distillation parameters follow in third and fourth places, with RVP in fifth. The strength of the aromatics term may help explain the negative coefficient for THC which stands out in contrast to the positive coefficients for NMOG and NMHC. The CH_4 model contains one quadratic term ($T50^2$), although its size is small. The CH_4 model includes one interaction from the design model, and two from the extended model, although all three are small in size.

Hot-Start Emissions. For the Bag 3 results, the overall pattern is more similar to Bag 2, rather than to Bag 1, although less consistent. Only CH_4 has a significant linear effect for ethanol. The effect for THC is quite small but insignificant, whereas those for NMOG and NMHC are larger but also insignificant. For these species, the ethanol effects are retained only to maintain hierarchy with respect to two interactions involving ethanol. For aromatics, THC and CH_4 have strong negative aromatics effects; the effect for THC may be explained by the presence of its CH_4 component. In contrast, neither NMOG nor NMHC retain aromatics effects in the Bag 3 model, which contrasts with both models for Bags 1 and 2. None of the reduced models include significant RVP terms; again the RVP effect is retained on the strength of the etOH×RVP interaction, which is large and significant for NMOG and NMHC, small and marginally significant for THC, and absent for CH_4. For the distillation parameters, the relatively small positive effects for THC may be explainable by the contrast between the behavior in its non-methane and methane components, which show large and small coefficients for these terms, respectively. None of these four models include either quadratics term.

Several additional observations can be made. One is that the magnitudes of corresponding coefficients are generally larger for Bag 1 than for Bag 2 emissions, suggesting that the effects of fuel properties are more pronounced for "cold start" than for "hot running" emissions. This point also generally applies to Bag 3 relative to Bag 1, with some exceptions. One exception is that the coefficients for distillation parameters in the NMHC models are as large as those in the Bag 1 model. Another interesting example is that the negative coefficients for aromatics in the CH_4 models are very close in size in all three models.

Secondly, it is apparent that relative patterns among coefficients are more similar between Bags 1 and 2 than between Bag 3 and the other two Bags. The reasons for this difference may be related to the relative importance of measurement error in Bag 3 relative to the other two Bags. In the charts, this result may be illustrated by the fact that the confidence intervals for coefficients in Bag 3 are generally larger than their counterparts in the Bag 1 or Bag 2 models. Again CH_4 is an exception, having been measured with roughly similar precision in all three bags.

Finally, the similarities between the NMOG and NMHC models are expected, given the close correlation of these two species to $NMHC_{FID}$, as described above in 3.2 (page35 ff.), and the fact that a majority of values for Bags 2 and 3 were imputed from $NMHC_{FID.}$ However, the similarities are also highly pronounced for Bag 1, for which only a small fraction of measurements were imputed.

Figure 97. THC: Model Coefficients for Reduced Models, based on 11-term and 16-term full models. Tabular results shown in Table 56 and Table 57 (error bars represent 90% confidence intervals).

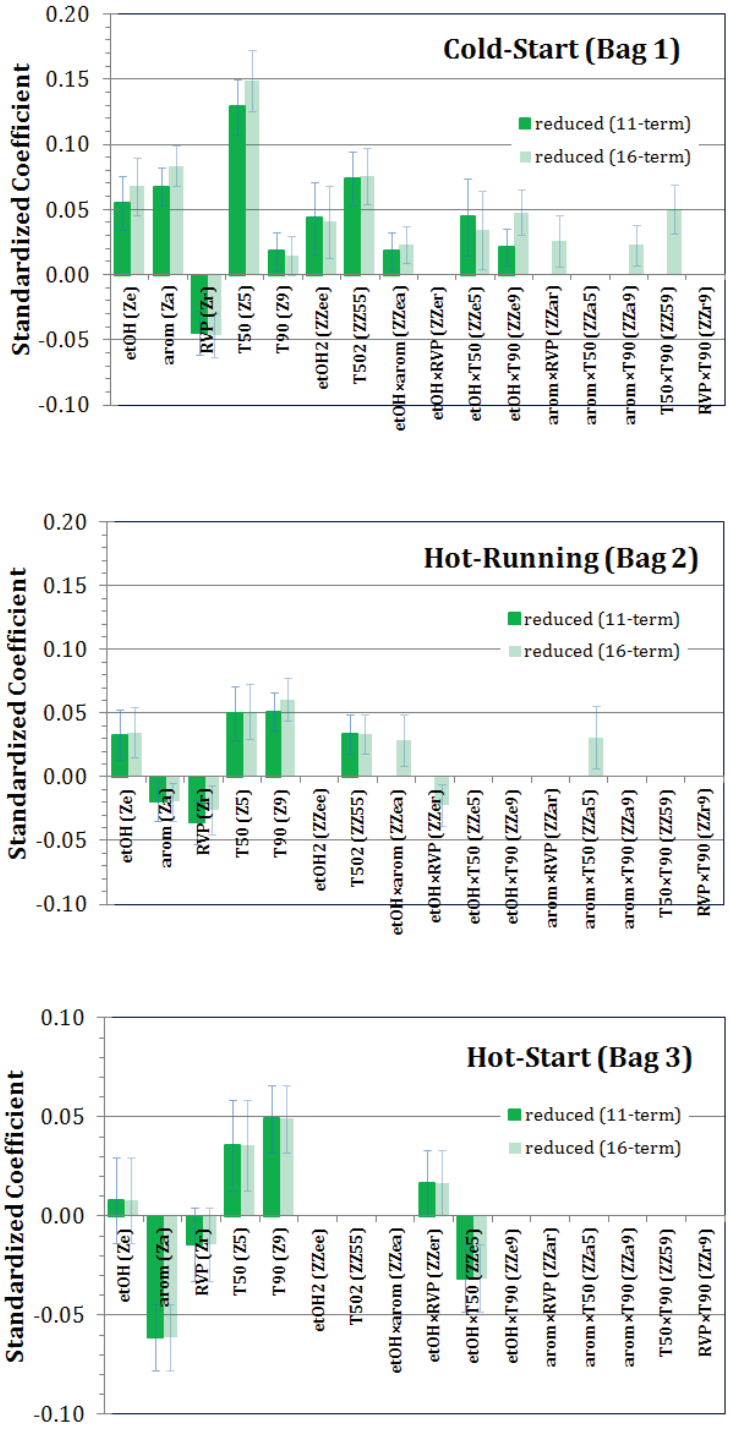


Figure 98. NMOG: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 58 and Table 59 (error bars represent 90% confidence intervals).

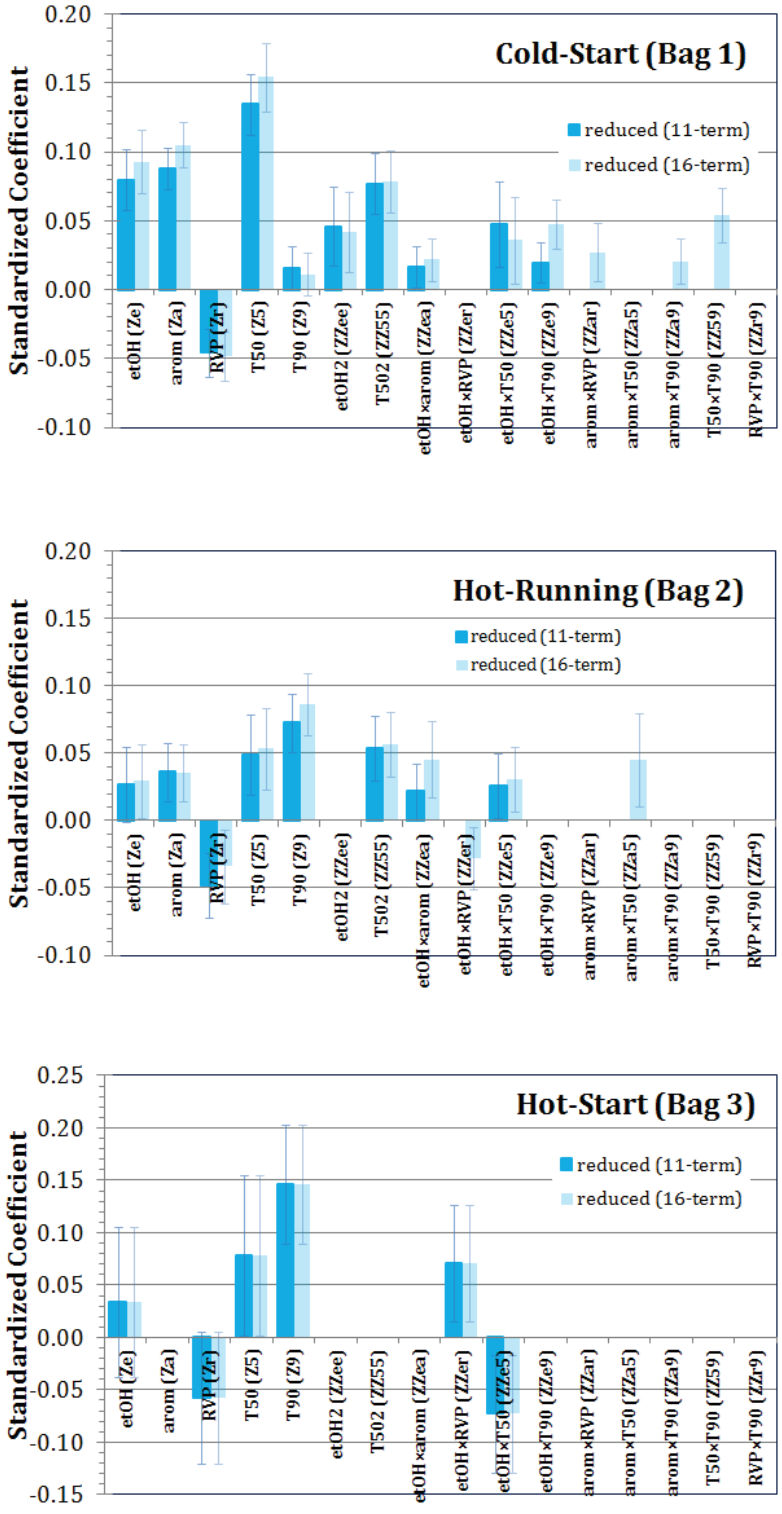


Figure 99. NMHC: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 60 and Table 61 (error bars represent 90% confidence intervals).

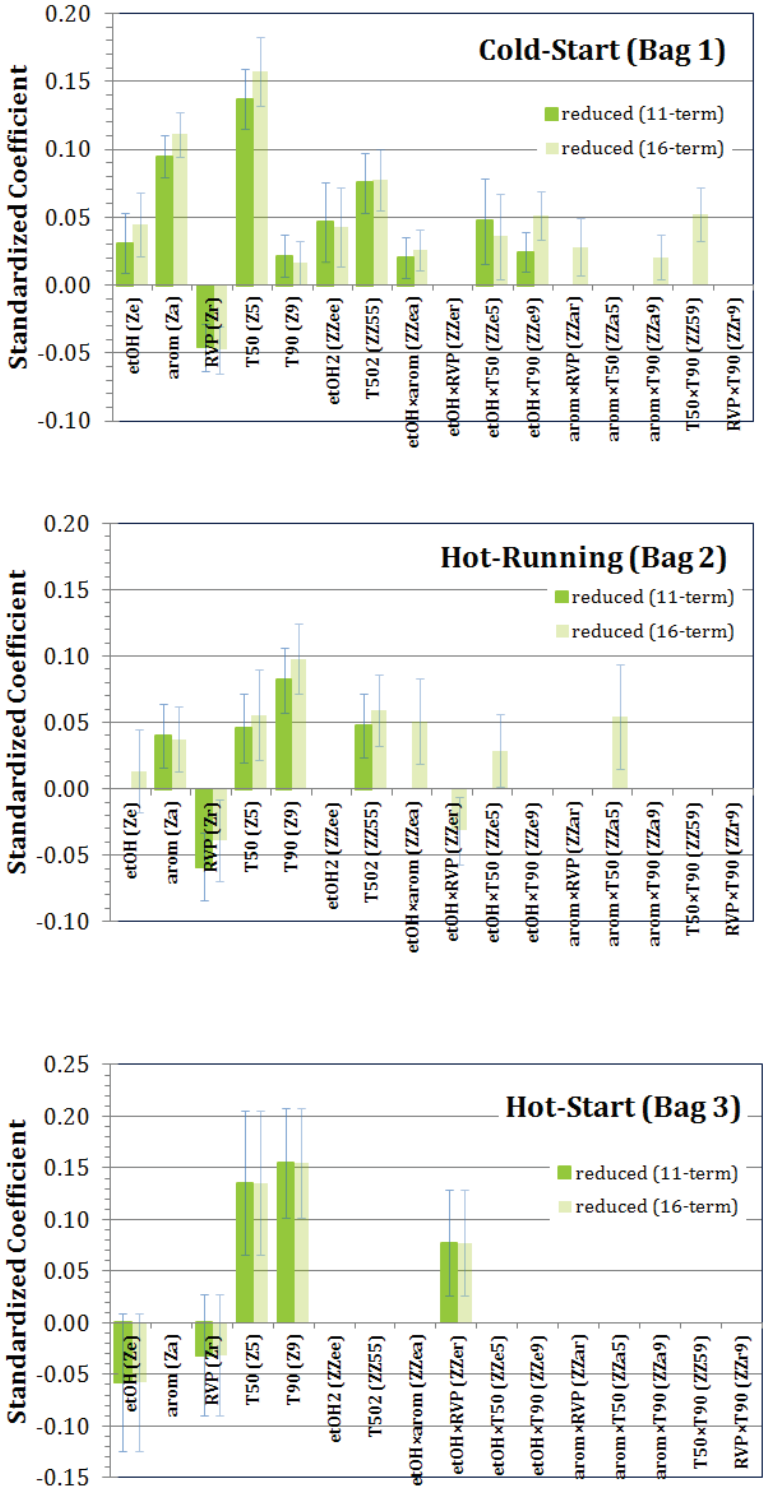


Figure 100. CH_4: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 62 and Table 63 (error bars represent 90% confidence intervals).

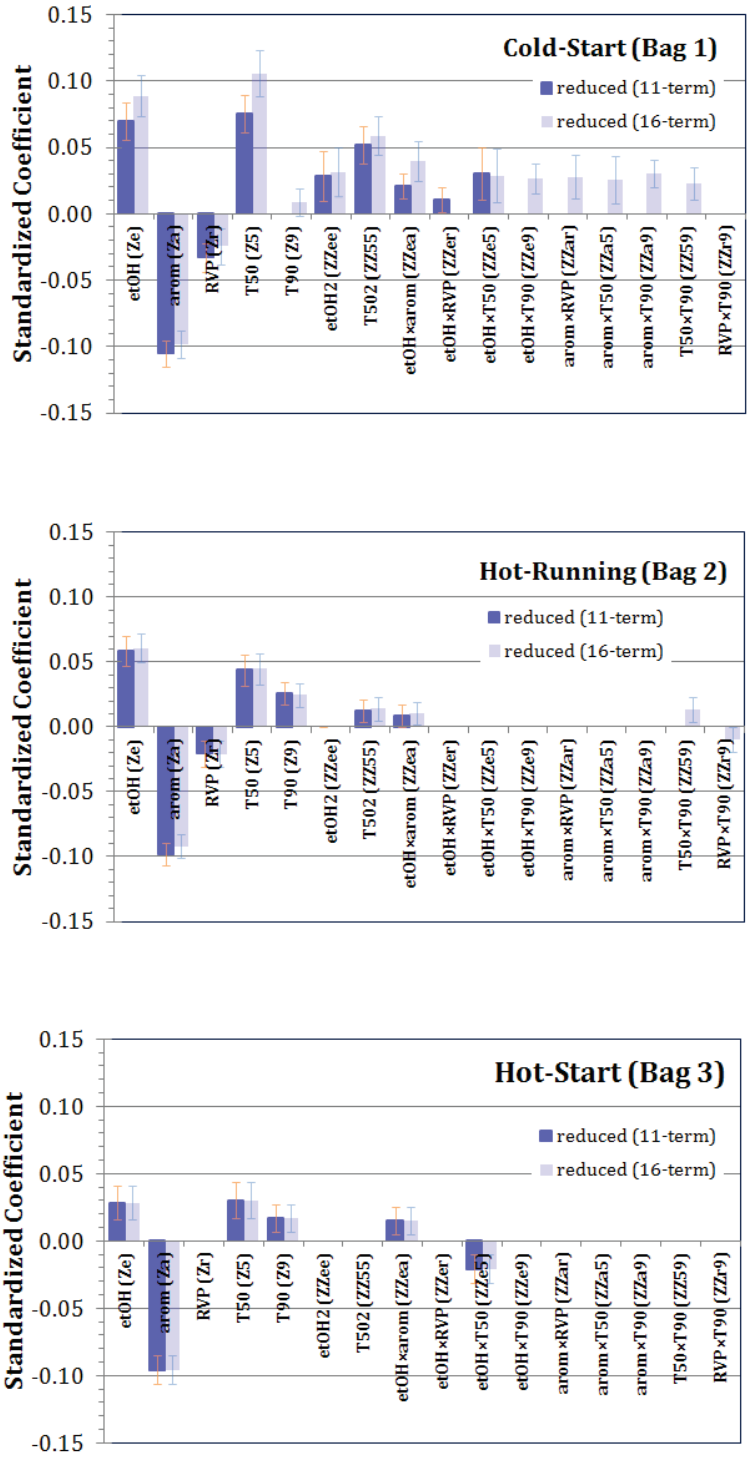
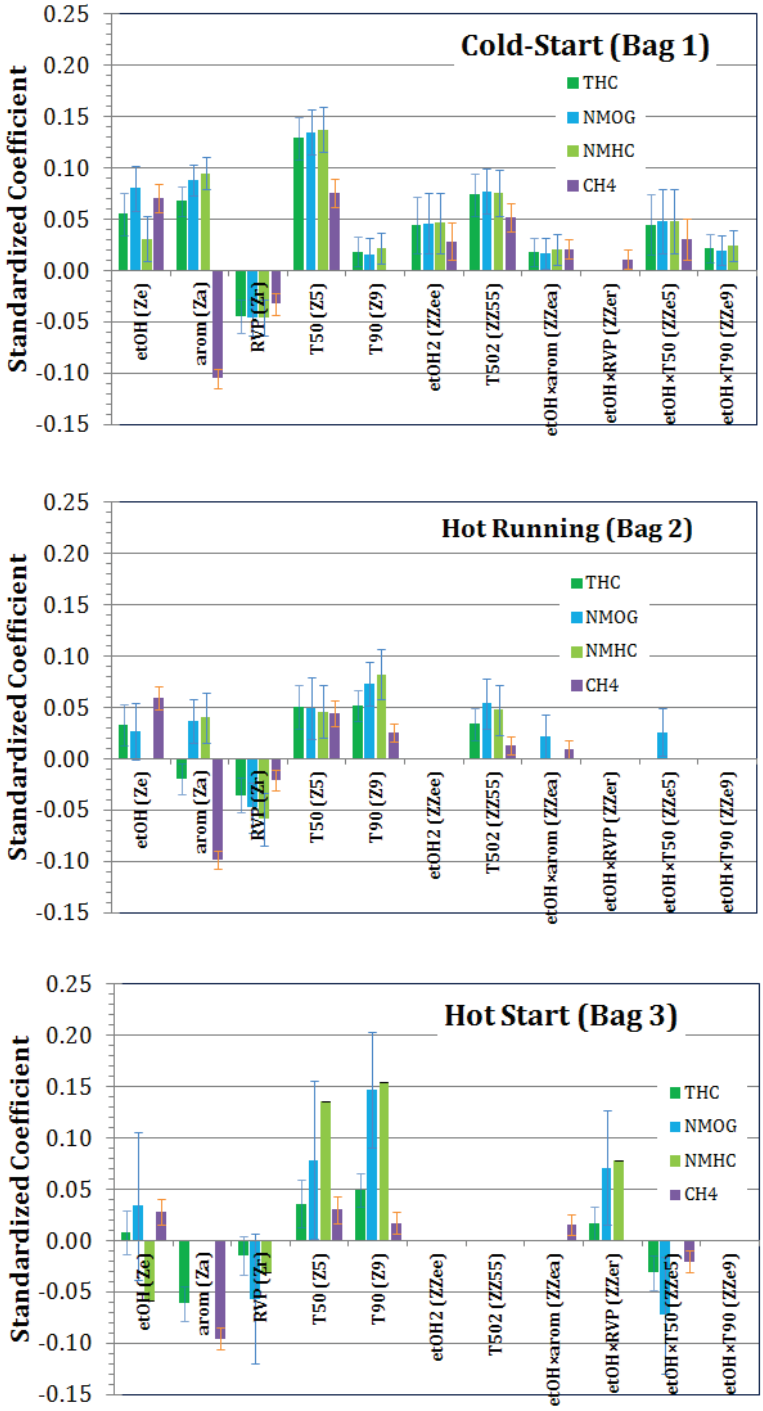


Figure 101. Reduced Models for Hydrocarbon species, fit based on the 11-term design model (reduced$_{11}$) (error bars represent 90% confidence intervals for standardized coefficients)



9.1.1.2 Oxides of Nitrogen (NO_x)

Coefficients for reduced models for NO_x are presented graphically in Figure 102. The plots include results for Bags 1-3 and constitute a graphic presentation of the tabular results previously shown in 7.2, (page 137).

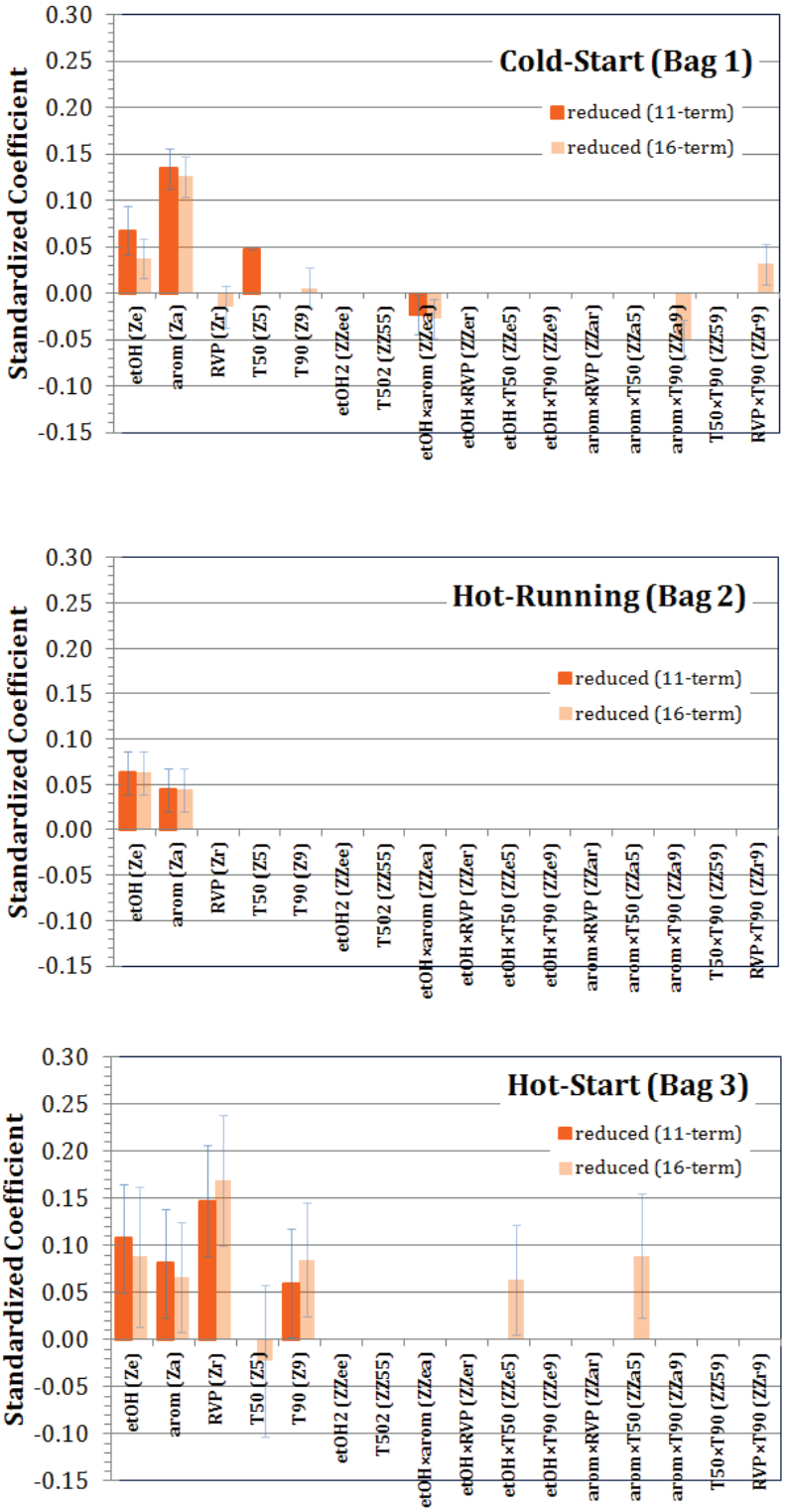
Cold-Start Emissions. For Bag 1 results, aromatics is the single most important effect, by a wide margin, followed by ethanol. As described, the reduced model differs depending on whether the 16-term or 11-term models are taken as the starting point. When starting with the 16-term model, the $reduced_{16}$ model retains RVP and T90 linear terms (both insignificant), plus two interactions not included in the $reduced_{11}$ model: arom×T90 and RVP×T90. In contrast, when starting with the design model, neither RVP nor T90, nor any of their interactions are retained, but the model fits a fairly strong (and significant) T50 term. Both models fit a negative etOH×arom interaction. The reasons for these differences are not apparent, but it is clear that the relations among NO_x, etOH and T50 are complex (See Figure 84-Figure 86 above).

Hot-Running Emissions. By contrast, the reduced model for Bag 2 is very simple. It retains only two linear terms, etOH and aromatics, which are also the two most important terms for Bag 1, although the relative importance of these two properties is reversed. In addition, this model is unique in that starting with the 16-term or 11-term full models does not affect the outcome. This result obtains because in this case, a five-term linear effects model gives a better fit than models containing any interactions, meaning that none of the interaction or quadratic terms contributes to improvement in fit. For example, the BIC for the 16-term model, the linear-effects model and the reduced model are 1134, 1112 and 1105, respectively.

Hot-Start Emissions. The models for Bag 3 show a markedly different pattern, however. While the etOH and aromatics coefficients show a pattern similar to that in Bag 2, although somewhat larger, this model includes positive RVP and T90 terms not present in either the Bag 1 or Bag 2 model. In fact, the RVP term is the most important term in the model, with T90 on par with aromatics and ethanol. The Bag 3 model also fits two interactions not present in either the Bag 1 or Bag 2 models. It may be appropriate to consider whether the Bag 3 results may be more vulnerable to measurement error attributable to low sample measurements relative to background, given the issues with measurement discussed in 6.1.1 (page 114).

Overall, it appears that the Bag 1 model is dominated by the effect of aromatics, and the Bag 2 model by ethanol, followed closely by aromatics. The magnitude of the aromatics effect in Bag 1 is similar between the $reduced_{16}$ and $reduced_{11}$ models. However, the ethanol coefficient differs between them, with the inclusion of the T50 term in the $reduced_{11}$ model associated with an increase in the ethanol coefficient.

Figure 102. NO_x: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 66 and Table 67.



9.1.1.3 Particulate Matter (PM)

Coefficients for reduced models for PM are presented graphically in Figure 103. The plots include results for Bags 1-3 and constitute a graphic presentation of the tabular results previously shown in 7.2, (page 137).

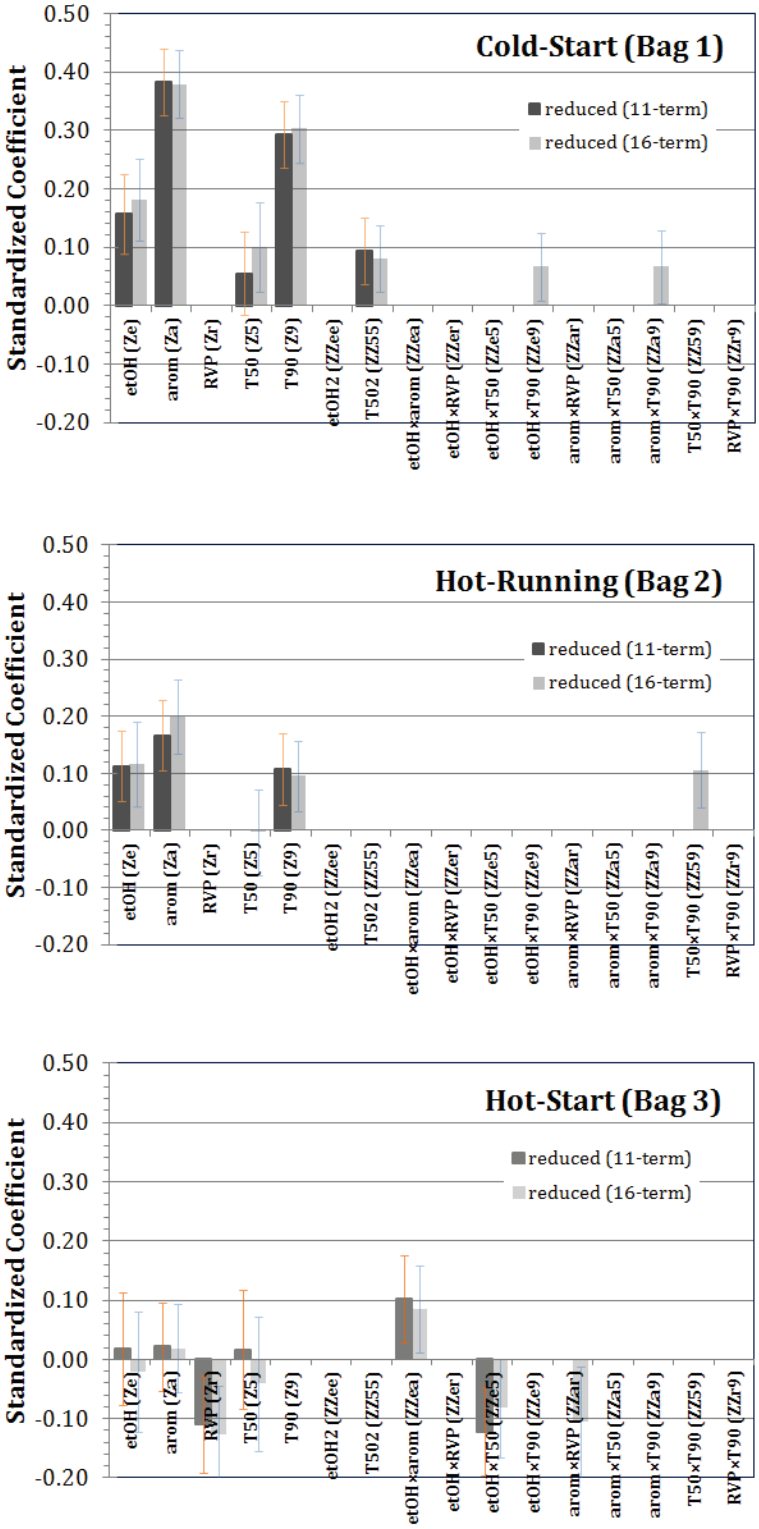
Cold-Start Emissions. For Bag 1 results, as with NO_x, aromatics is the single most important effect, by a wide margin, followed by T90 and ethanol. A relatively small effect is also fit for T50, which differs by 50% between the $reduced_{16}$ and $reduced_{11}$ models, whereas the other three linear effects are relatively consistent in this respect. Neither model includes an RVP term, whereas both include a quadratic term for T50. The quadratic term is positive, as is the linear effect, denoting an upward curvature. The $reduced_{16}$ model includes two interactions not included in the $reduced_{11}$ model. One term, etOH×T50, is a design-model term, and the second, arom×T90 is not. The presence of the arom×T90 interaction may indicate that the strong aromatics and T90 effects reinforce each other, although the magnitude of the interaction is not large relative to the linear effects.

Hot-Running Emissions. The model for Bag 2 is relatively simple. It retains three linear terms, etOH, aromatics and T90, which are also the two most important terms for Bag 1, although the relative importance of ethanol and aromatics are reversed. The $reduced_{16}$ model also includes a fairly strong T50×T90 interaction not retained in the Bag 1 model. Aside from relatively minor interactions, the Bag 1 and Bag 2 models appear generally similar.

As with NO_x, the models for Bag 3 show a markedly different pattern, with even more pronounced departures. The Bag 3 model lacks a term prominent in Bags 1 and 2 (T90) and contains a term lacking in both of the other bags (RVP). In fact, RVP is the strongest and the only significant linear term in the model. The terms for ethanol, aromatics and T50 are very small, and insignificant, with the signs of the effects varying between the $reduced_{11}$ and $reduced_{16}$ models for ethanol and T50. In fact, aside from RVP, the linear terms are retained only to maintain hierarchy with three interaction terms (etOH×arom, etOH×T50 and arom×RVP). These results suggest that the Bag 3 results are probably unduly affected by measurement error due to very low filter masses. It appears appropriate to discount the Bag 3 results for purposes of assessing the effects of fuel properties on particulate emissions.

Overall, it appears that the both the Bag-1 and Bag-2 models are dominated by the effect of aromatics. The T90 effect is also important, but more so for Bag-1 than Bag-2 emissions. Ethanol is important in both models, although somewhat more so in Bag 2. As with HC and NO_x emissions, the coefficients are larger in Bag 1 than in Bag 2, although the general pattern among effects is similar.

Figure 103. PM: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 68 and Table 69.



9.1.1.4 Carbon Monoxide (CO)

Coefficients for reduced models for CO are presented graphically in Figure 104. The plots include results for Bags 1-3 and constitute a graphic presentation of the tabular results previously shown in 7.2, (page 137). The sets of coefficients for CO appear unique in that they show markedly different patterns for the different Bags.

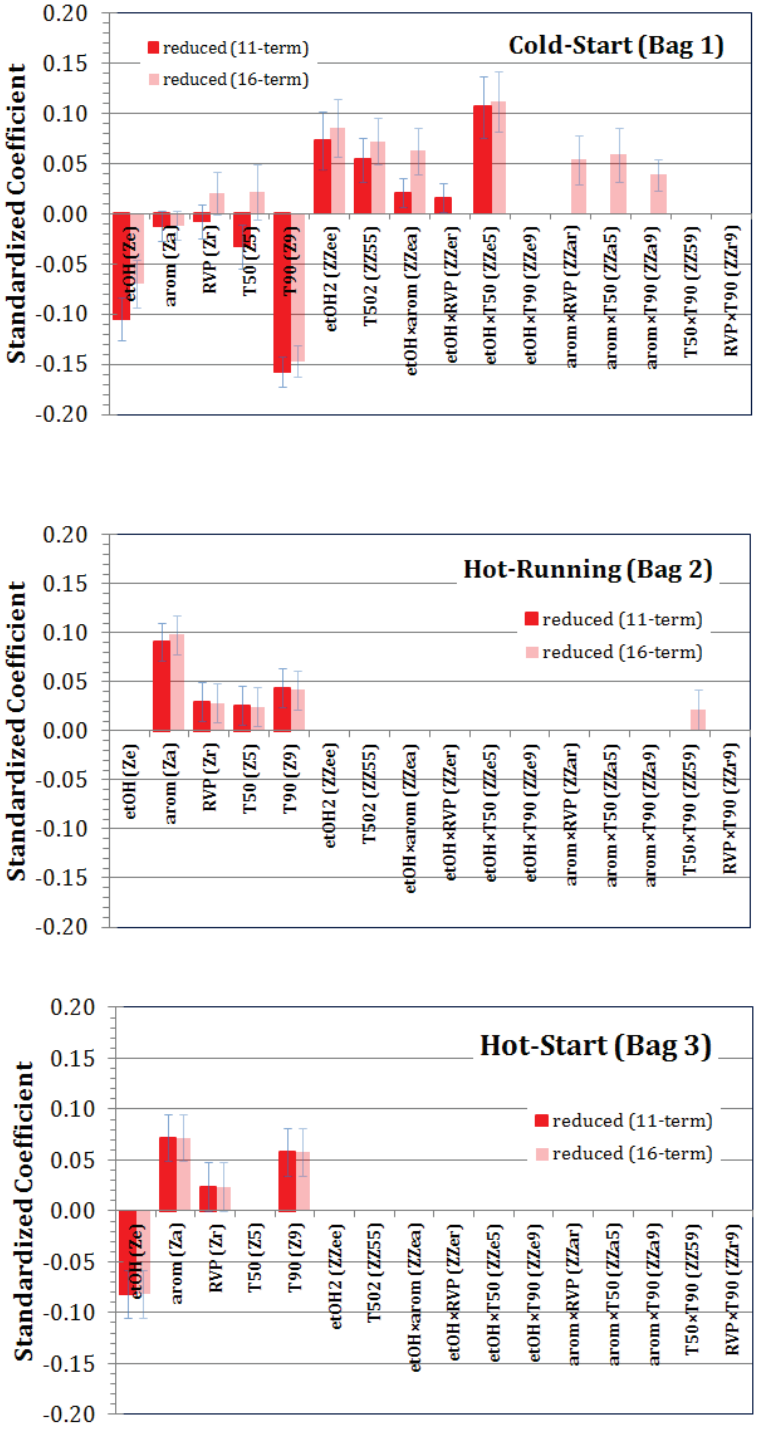
Cold- Start Emissions. For Bag 1 results, a striking feature is that the linear effects for ethanol and T90 are strongly negative, a pattern that occurs for no other emission or bag. The effect for aromatics is also small but negative. Linear effects for RVP and T50 are relatively small, but are unusual in that they change sign from positive to negative between the $reduced_{16}$ and $reduced_{11}$ models. Both models include quadratic terms for ethanol and T50. Both models include interaction terms for etOH×arom and etOH×T50 effects. Interestingly, the etOH×Arom coefficient in the $reduced_{11}$ model is less than a third of than in the $reduced_{16}$ model, whereas the coefficients for the etOH×T50 terms are very close in size. The $reduced_{11}$ model also contains an etOH×RVP interaction lacking in the $reduced_{16}$ model. The $reduced_{16}$ model retains three interactions from among the additional interactions in the extended model.

Hot-Running Emissions. The model for Bag 2 is relatively simple but presents a very different profile than the Bag 1 model. A prominent difference is that neither reduced models retain an ethanol effect, either positive or negative. Also in contrast to the Bag 1 model, the aromatics and T90 terms are positive and strong, with aromatics as the most important effect. The RVP and T50 terms are both positive but small, although significant. The $reduced_{16}$ model retains a small T50×T90 interaction from the extended model, aside from which the two reduced models are very similar, much more so than the Bag 1 model.

Hot-Start Emissions. The models for Bag 3 have similarities to both the Bag-1 and Bag-2 models. They are similar to Bag 1 in that the ethanol effect is strong and negative, but also similar to Bag 2 in that the aromatics and T90 effects are strong and positive. The RVP effect is small but positive, but the T50 effect is not present. Both reduced models have no quadratics or interaction terms.

If we accept the results of these models, especially those for Bags 1 and 2, it could suggest that the effects of changing fuel properties differs markedly between cold start and running emissions, more so than for any other emission or bag.

Figure 104. CO: Model Coefficients for Reduced Models, based on 16-term and 11-term full models. Tabular results shown in Table 64 and Table 65.



9.1.1.5 Summary

The study was conducted to assess the effects of fuel properties on the emissions of vehicles certified to Tier-2 standards, primarily the Bin-5 standards. Reviewing the results, it is clear that such effects exist and are measurable. In this section, we will attempt to review and summarize the results.

As mentioned above, the application of the study design supplemented by the standardization of parameters allows assessment of fuel effects as though they were independent. These devices provide powerful tools to aid in interpretation of the effects of changing fuel properties on emissions. As such, the models will aid in the interpretation of the results of other studies, as well as the results of this project.

In reviewing Table 88 and Table 89, as well as Figure 97 through Figure 104 above, we can make some generalizations with respect to the individual fuel properties:

Ethanol: In most models, the linear-effect coefficients for ethanol are positive for both running and start emissions, implying that increases in ethanol content would be associated with increases in emissions, if the remaining fuel properties could be kept constant while increasing the ethanol level. A conspicuous exception to the pattern is CO, which has a negative coefficient for start emissions and no ethanol term for running emissions. Another exception is NMHC, which no ethanol term for running emissions. For start emissions, the terms are largest for PM, NO_x and NMOG, whereas for running emissions, the terms are largest for PM, NO_x and CH_4, although presumably, the underlying physical processes could vary among pollutants and processes. The linear effects for NO_x are unique in that the coefficients are nearly equal in size for both start and running emissions, implying that the relative effect of ethanol is similar in both processes, even though start emissions are considerably higher. For start emissions, the etOH×etOH quadratic term is present for all HC species and CO, imparting some curvature to ethanol trends for these species. It is consistently positive. Its size is similar for THC, NMOG and NMHC, smaller for CH_4 and larger for CO. For running emissions, no models include the quadratic term.

Aromatics: The patterns for aromatics are less consistent. Coefficients are positive for most models, with several exceptions for both start and running emissions. One exception is CO, which has a small negative coefficient for start emissions and a larger positive coefficient for running emissions. A second exception is THC, for which the start coefficient is positive and the running coefficient negative. Thirdly, coefficients for CH_4 are large, negative and similar in size for both start and running emissions, which is unique in implying that changes in ethanol have similar relative effects on both start and running emissions. In terms of magnitude, the pattern is similar to ethanol, with PM, NO_x and NMOG showing the strongest effects. For start emissions, the interaction between aromatics and ethanol appears in all models except PM. The start interaction terms are consistent in size and positive in sign for all emissions except NO_x. Given

the signs of the linear terms, the interaction qualifies as a small reinforcement for all models except CH_4 and NO_x, for which it qualifies as an interference. For running emissions, this interaction appears only for NMOG and CH_4, for which it acts as a reinforcement and as an interference, respectively.

RVP: A linear term is included for all pollutants except NO_x and PM. The sign of the term is consistently negative with a single exception for running CO, which has a positive term. For the hydrocarbons, the size of the term is relatively consistent, although the coefficients for running models tend to be somewhat smaller than those for start emissions. The interaction with ethanol appears in two models for start emissions, but in no models for running emissions. In both start models, the terms are small and positive.

T50: Coefficients for this property consistently positive, with the single exception of start CO, and appears in all models except running NO_x and PM. For start emissions, the effects are largest for THC, NMOG and NMHC, and smaller for CH_4, NO_x and PM; for CO, the term is negative and relatively small. For running emissions, coefficients are positive but smaller than for start emissions; for the hydrocarbons except CH_4, T50 is the largest single term. For the hydrocarbon species, T50 shows a consistent reinforcement interaction with ethanol for start emissions, for running emissions, the interaction applies only to NMOG. For start CO, the interaction is present but acts as an interference, in that both linear terms are negative and the interaction is positive.

T90: This term is unique in that it appears more frequently in models for running than for start emissions, and in that it is sometimes larger in running models than start models. In the start models, the term is large and positive for PM, small and positive for THC and NMHC, large and negative for CO, and absent for the remaining models. In the running models, the term is large and positive for the hydrocarbons except methane, small and positive for PM and CO, and absent for NO_x. The interaction between T90 and ethanol is retained in only two models for start emissions, THC and NMHC, in which it is positive and similar in size. In both models, the linear and interaction terms are all positive, qualifying this effect as a reinforcement interaction. The T90 coefficient is largest for start PM, where it has a reinforcement interaction with the even stronger aromatics effect.

In addition, it is possible to make some general points about the responses of exhaust emissions to changing fuel properties that apply across the measured compounds and species, and for both start and running emissions.

- Other factors being equal, increasing ethanol is associated with an increase in emissions, as indicated by the positive ethanol coefficients in most models, both for running and start emissions.
- Other factors being equal, increasing volatility is associated with reductions in (exhaust) emissions, as indicated by generally negative coefficients for RVP (and generally positive coefficients for T50).

- In relative terms, fuel effects are generally more pronounced for start than for running emissions, as indicated by the fact that in most cases, the coefficients for Bag 1 models are larger than their counterparts for Bag 2 models. If we assume that we can validly make direct comparisons between coefficients between Bag-1 and Bag-2 models, this result may suggest that the effects of fuel properties are more pronounced during engine starts than during running operation. One interpretation might be that fuel effects could be damped by efficient operation of the catalyst after the engine comes up to temperature.

9.1.2 Models Selected for Application

Based on our review and analyses of study results to date, our intent is to make use of reduced models fit with respect to the 11-term design model ($reduced_{11}$) for purposes of description and prediction. For application, we are retaining reduced models for Bags 1 and 2, based on concerns that Bag 3 models may not reliably represent fuel-parameter effects on emissions.

Given the low design efficiency for the extended model, we presume that it is more probable that terms not included in the design model may be more liable to bias or to represent artifacts of design or measurement. Until such concerns have been ruled out, it is reasonable to retain models fit within the space defined by the study design.

Coefficients for the reduced models, previously presented in 7.2.3 above, are summarized below. Coefficients for Bag 1 and Bag 2 models are presented in Table 88 and Table 89, respectively.

As noted, the results of this project directly apply to vehicles certified to Tier-2 standards. In considering the applicability and representativeness of the results, several questions arise. Firstly, as the vehicles employed in the study were relatively new, with less than 10,000 miles accumulated, we can ask whether these modeling results would apply to similar vehicles having aged, deteriorated or having accumulated high mileage. Although not as important, we can also ask whether the results might also be applicable to vehicles certified to earlier standards, such as NLEV or LEV-I standards.

An important characteristic of emissions, including the results of this project, is that they typically follow multiplicative or logarithmic scaling. Given that the coefficients represent differences in logarithms, which represent relative multiplicative differences, i.e., ratios, it is plausible that the model coefficients, expressed as relative differences, should prove transportable to either aged Tier-2 vehicles or even pre Tier-2 vehicles.

It is also necessary to consider the composition of the vehicle sample. As described above, it comprises a judgment sample of high-sales models from major manufacturers in model year 2008. In terms of standards, the vehicles represent the emissions standards that are most

prevalent for light-duty vehicles, including Bins 3 and 5 (or equivalent LEV and ULEV standards under LEV-II), as well as a single Bin 8. The selection of makes and models does not qualify as a random sample in a strict sense. One practical limitation is that the effort and expense involved in measurement precluded drawing a reasonably sized random sample of makes and models. Nonetheless, given the size of the sample, it is plausible that a well-designed judgment sample may perform as well as a small random sample, as noted by one of the pioneers of survey design: "*... No clear rule exists for deciding exactly when probability sampling is necessary, and what price should be paid for it. The decision involves scientific philosophy and research strategy ... ... If a research project must be confined to a single city in the United States, I would rather use my judgment to choose a "typical" city than select one at random. Even for a sample of 10 cities, I would rather trust my knowledge of U.S. cities than a random selection. ...*"[21] Additional clarification of the utility of the models may be provided by subsequent validation using independent data, including the aspects of applicability and transportability.

Table 88. Bag 1: Reduced Models, based on the 11-term Design Model.

Model term	Notation	THC	CH_4	NMOG	NMHC	CO	NO_x[1]	PM
Intercept	Intercept	-0.8664	-3.0074	-0.95209	-1.0315	1.3466	-2.8594	0.6559
etOH	Z_e	0.0548	0.06994	0.08019	0.03094	-0.1049	0.06750	0.1582
Arom	Z_a	0.0676	-0.1053	0.08782	0.09461	-0.01242	0.1339	0.3833
RVP	Z_r	-0.0445	-0.03275	-0.04224	-0.04568	-0.00762		
T50	Z_5	0.1288	0.07554	0.1345	0.13689	-0.03273	0.04783	0.0550
T90	Z_9	0.0183			0.02160	-0.1571		0.2923
etOH × etOH	ZZ_{ee}	0.0436	0.02844	0.04432	0.04612	0.07304		
T50 × T50	ZZ_{55}	0.0736	0.05170	0.07579	0.07534	0.05358		0.0935
etOH × Arom	ZZ_{ea}	0.0179	0.02088	0.01693	0.02045	0.02086	-0.02369	
etOH × RVP	ZZ_{er}		0.01082			0.01596		
etOH × T50	ZZ_{e5}	0.0445	0.03048	0.04653	0.04729	0.1064		
etOH × T90	ZZ_{e9}	0.0214			0.02441			
Vehicle variance	σ^2_{veh}	0.1325	0.2855	0.1224	0.1266	0.3920	0.5925	0.4251
Residual error	σ^2_{ε}	0.06872	0.03014	0.07538	0.07624	0.07214	0.1458	1.0359

[1] Fit excluding the Ford Focus.

Table 89. Bag 2: Reduced Models, based on the 11-term Design Model.

Model term	Notation	THC[1]	CH_4	NMOG[1]	NMHC[1]	CO	NO_x[2]	PM
Intercept	Intercept	-4.6533	-5.7075	-5.2360	-5.3253	-1.3893	-4.5692	-1.3107
etOH	Z_e	0.0327	0.05860	0.02673			0.06299	0.1126
Arom	Z_a	-0.0195	-0.09836	0.03634	0.03987	0.0913	0.04407	0.1662
RVP	Z_r	-0.0355	-0.02049	-0.04786	-0.05881	0.0299		
T50	Z_5	0.0501	0.04394	0.04915	0.04548	0.0261		
T90	Z_9	0.0514	0.02575	0.07252	0.08202	0.0440		0.1072
etOH × etOH	ZZ_{ee}							
T50 × T50	ZZ_{55}	0.0337	0.01227	0.05349	0.04774			
etOH × Arom	ZZ_{ea}		0.008769	0.02171				
etOH × RVP	ZZ_{er}							
etOH × T50	ZZ_{e5}			0.02586				
etOH × T90	ZZ_{e9}							
Vehicle variance	σ^2_{veh}	0.8384	1.1108	0.8502	0.9691	1.9187	0.4720	0.7827
Residual error	σ^2_{ε}	0.06717	0.02518	0.1310	0.1708	0.1256	0.1836	1.1337

[1] Fit excluding the Honda Odyssey and Toyota Sienna.
2 Fit excluding the Chevrolet Cobalt.

9.1.3 Comparison to Previous Results

It is valuable to place the results of a new study in the context of the existing body of knowledge. Results of this program generally show that Tier 2 vehicles continue to exhibit sensitivity to fuel parameters in many of the same ways as older vehicles certified to previous standards, with some exceptions that will be examined here. This discussion will focus on the direction of specific linear-effects coefficients in the models, i.e., effects of each property one at a time, as though the other properties were held constant. Attempts to characterize interactive effects are more complicated and are thus more difficult to interpret.

9.1.3.1 Ethanol

Over the past two decades a relatively large number of programs have studied the effects of ethanol. At the same time, since blending ethanol into gasoline also affects many other fuel properties, and given that ethanol is blended in into gasolines in different ways that affect the collateral property changes differently, it is difficult to interpret trends across the body of literature without more information on multiple fuel property changes. A recent summary of the literature cites several studies from the late 1980s through the early 2000s (covering Tier 1 and earlier vehicles) that consistently show ethanol blends as having increased NO_x emissions.[22] More recently, in the package for the 2009 Renewable Fuels Standard, EPA also summarized a

number of data sources and found the same trend[23]. The results of the present study are consistent with the published results, in that the coefficient for ethanol is positive for both start and running emissions. However, it is important to note that the models also suggest that reductions in NO_x could occur with corresponding reductions in aromatics, particularly for start emissions, for which the aromatics coefficient is larger than that for ethanol. Thus, despite much lower overall emission levels that have been achieved in Tier 2 vehicles through improved fuel control and catalyst efficiency, the effect of ethanol on tailpipe emissions appears to persist in both cold-start as well as hot-running operation.

Many of the past studies have simply measured THC or NMHC by flame ionization detector and therefore have not fully captured the impacts of ethanol on VOC or NMOG emissions.[r] In the current study, we fit models to HC as THC, CH_4, NMHC and NMOG, and found positive coefficients for ethanol for both start and running emissions, although the linear-effects coefficients for running emissions were significant for THC and CH_4, marginally significant for NMOG and insignificant for NMHC. However, if typical collateral fuel changes (lower T50 and aromatics) are accounted for, we might project that blending ethanol would tend to reduce THC, NMHC and NMOG emissions (highlighting the important sensitivities to these other fuel parameters). Potential changes for CH_4 are more difficult to project, as the aromatics and T50 coefficients are opposite in sign.

9.1.3.2 Aromatics

Aromatic content of gasoline has long been understood to affect regulated and unregulated emissions. The literature review shows consistent findings of reduction in hydrocarbon emissions with reduced aromatics level, however the trend is more variable for NO_x. The literature summary suggests that higher aromatics content is likely to increase engine out NO_x due to the higher flame-front temperatures produced by aromatic compounds, but they also appear to increase the efficiency of three-way catalysts that reduce NO_x to nitrogen.

For NO_x, this study finds positive coefficients for both cold-start and hot-running operation, which is consistent with effects reported in some of the more recent studies described in the review. This pattern suggests that the effect might be due primarily to the engine-out effect during cold-start and transients during hot-running operation. Recent studies performed by Honda show that PM emission increases are amplified with higher heavy aromatics (C9+) content, an effect that may be reflected in the positive coefficients for aromatics (and T90) reported in the present study, pointing to aromatics composition as an area for further investigation.[24] It is interesting to note that aromatics is the only fuel property studied for which

[r] The typical hydrocarbon analyzer used for emission testing contains a flame ionization detector (FID), which is calibrated (typically using propane) to accurately count carbon atoms within H-C bonds. Carbons bonded to oxygen, which occur in carbonyl and alcohol emissions from burning ethanol fuels, produce a much smaller response in the FID, and thus emissions from ethanol fuels require additional characterization methods to properly quantify as NMOG or VOC.

an increase is associated with increases in all regulated emissions, with the exception of cold-start CO.

9.1.3.3 Distillation parameters

Relative to ethanol and aromatics, the effects of distillation properties require more complex study designs and fuel-blending procedures in order to properly evaluate their effects. In addition, as bulk properties of the fuel blend, their effects on combustion are more difficult to interpret and understand mechanistically. The literature review summarizes a number of older studies that suggest reducing T50 and T90 (i.e., lighter, more volatile fuel) increases NO_x emissions, while the present study suggests the opposite. The literature review doesn't distinguish cold-start vs. running emissions, but the current study suggests that reduced T50 is associated with NO_x reduction during the cold start, and has no effect after warm-up. With the understanding that emission controls on Tier 2 vehicles are highly efficient once active, we could surmise that a more volatile fuel combusts more readily during the first several seconds after a cold start, and thus causes the catalyst to become active earlier than with less volatile fuels. Once the catalyst is active, subsequent changes in combustion behavior and engine-out emissions due to distillation properties may be much less important than in older vehicles where after-treatment became active later in the cycle. The present study shows little or no effect of T90 on NO_x, which is reasonable considering that NO_x production and control are largely influenced by release of energy as heat. As T90 represents the less volatile fractions composing the "higher end" of the distillation curve, it represents to a lesser degree the volatile fractions primarily combusted under cold-start conditions, and thus less energy during starts, than T50, which represents the center of the distribution curve.

Consistent with the older studies described, the current study suggests that reducing T50 will reduce both start and running HC, for all four species modeled, which can be plausibly attributed to more rapid and complete evaporation and combustion of the bulk of the fuel mixture during all phases of combustion. For T90, the present study is consistent with older data during hot-running operation, where reducing T90 reduces HC emissions for all four species, although in varying degrees. In addition, for cold-start emissions, the models for THC, NMHC and NMOG suggest that the distillation parameters interact with ethanol so as to reinforce the effects of both (both linear-effects and interaction coefficients are positive in all cases). This situation highlights the fact that combustion is quite complicated on a microscopic level, involving complex interactions of multiple fuel parameters and other dynamic phenomena occurring in rapidly changing conditions during the first several seconds of engine operation.

9.1.3.4 Vapor Pressure

Fuel vapor pressure (as RVP or DVPE) primarily affects evaporative emissions (both direct and permeation), but it has been studied in some exhaust programs as well. The literature reports that a program conducted in the late 1980s found lower RVP associated with reduced exhaust

hydrocarbon emissions, but that other studies since conducted since have found variable or non-significant effects. For NO_x, the literature shows little or no effect of RVP.

The present study confirmed the previous findings of no (linear) effect of RVP on NO_x emissions; however, the coefficients for the HC species are consistently negative for both cold-start and hot-running emissions, suggesting increased HC emissions with reduced RVP (other properties constant), a trend apparently opposite to that described for hydrocarbons in the older studies, although concurrent fuel property changes in the cited studies have potential to confound this comparison. In the current study, the direction of the RVP coefficients seems directionally consistent with the positive T50 coefficients, suggesting that less volatile fuel (higher T50, lower RVP) would tend to show higher hydrocarbon emissions.

9.2 Speciated Hydrocarbons and Air Toxics

Summary results for several compounds in Bags 1 and 2 are presented below. We refer the reader to several tables, previously presented, for information on samples and model fitting approaches. Table 79 shows sample sizes, including numbers of total and censored measurements. Table 80 presents the notation used to identify each model term, and its level of standardization (one- or two-stage). Table 81 summarizes model-fitting approaches, for all models fit.

Sets of model coefficients for best-fit models are presented in Table 90 and Table 91. Detailed results, including models fit, fitting histories, coefficients and tests of effect are presented in Appendices K to Q.

In reviewing the coefficients for cold-start emissions in Bag 1(Table 90), it is interesting to note that the additional terms not included in the optimized design are not retained in best-fit models, with the exception of T50×T90 for acetaldehyde, formaldehyde and acrolein. In reviewing tests of effect for the full models (Table 84, Table 87), the additional parameters are often estimated with low precision. This result is not unexpected, given that the matrix was optimized to maximize the precision of estimation for those effects included in the design.

The model results reflect the study design applied to each compound as well as the underlying physico-chemical processes. The reduced model structures are more complex for those compounds fit with the full design, specifically the aldehydes, acrolein and ethanol.

Ethanol. The ethanol coefficients are positive and large for the aldehydes, acrolein and ethanol. For acetaldehyde and ethanol, the ethanol effects are clearly dominant. These results are not surprising, given the structural affinity between acetaldehyde and ethanol, and that the strongest indicator of ethanol in the exhaust is ethanol in the fuel. For formaldehyde and acrolein, the ethanol coefficients are important but not as dominant. Neither benzene nor 1,3-butadiene retain ethanol coefficients in their reduced models. All compounds except formaldehyde retain large and negative etOH×etOH quadratic terms, which are clearly required to fit the downward

curvature in the logarithmic trends. See Figure 92 for acetaldehyde, Appendix M.1 for acrolein and Appendix N.1 for ethanol.

Aromatics. In contrast to ethanol, the aromatics coefficients are small for the aldehydes, although several times stronger for acrolein. Ethanol does not retain an aromatics term in its reduced model. Not surprisingly, the aromatics coefficient for benzene is large, and the only effect retained in the reduced model (Note that fuel benzene is also a strong predictor of exhaust benzene, but was not a target study parameter). The two aldehydes retain small but significant reinforcement interactions between aromatics and ethanol.

RVP. The sign and strength of RVP coefficients are similar for all four compounds fit under the full design (but absent for those fit under the reduced design). As with the RVP terms in the models for aggregated hydrocarbons (THC, NMHC and NMOG, Table 88, page 230), the sign of the RVP linear effects are negative. The magnitudes for the individual species are also similar in size to those for the aggregate HC (-0.04 to -0.06). The interaction between ethanol and RVP is retained only in the acetaldehyde model, in which it is positive and small.

T50. For the four compound fit under the full design, linear-effect coefficients for T50 are positive. However, the pattern in the size of the coefficients mirrors that for ethanol, in that the two compounds with largest ethanol coefficients (acetaldehyde and ethanol) have smaller T50 coefficients than formaldehyde and acrolein, which have T50 coefficients about twice as large. These results may reflect similarities in structure between the two pairs of compounds, or similarities in formation processes during combustion. In addition to large linear coefficients, formaldehyde and acrolein have small interference interactions between T50 and ethanol.

T90. More so than for the other properties, linear coefficients for T90 differ among the compounds fit under the full design. The coefficients for acetaldehyde, formaldehyde and acrolein are positive, but increasing, respectively, with the values for formaldehyde and acrolein approximately 3 and 8 times larger than that for acetaldehyde. In contrast, the coefficient for ethanol is negative, suggesting reduced ethanol emissions for less volatile fuels. In addition to large linear effects for ethanol and T90, formaldehyde and acrolein have small reinforcement interactions between these properties.

The structures for reduced models are much simpler for benzene, 1,3-butadiene and ethane, reflecting the limits imposed by the reduced design. It is clear that in model fitting for these compounds, that only strong effects appear significant and are hence retained in the reduced models.

Covariance Parameters. The table also includes the two covariance parameters fit by the mixed model, or recalculated manually for the Tobit models. The “vehicle” component reflects the variance among vehicles, or the “between-vehicles” variance. It represents the variance of a normal distribution with mean 0, i.e., the random scatter of individual vehicles around the mean for all vehicles. The “residual” component represents random error, unexplained by the model

after accounting for fuel-parameter and vehicle effects. The relations between these two components are compound specific. For acetaldehyde, both components are relatively low and almost equal in size. Formaldehyde has a larger vehicle variance, whereas acrolein and ethanol show less variance between vehicles and larger residual error variances.

Corresponding sets of coefficients for the Bag 2 models are shown in Table 91. As with two of the Bag 1 models, these models reflect the simplicity of the reduced design. As mentioned, it was not possible to fit RVP effects for any compound, nor did we attempt to fit quadratic or interaction terms. For acrolein, benzene and 1,3-butadiene, no model fitting was attempted, given the high rates of censoring for these compounds (Table 79). Intercepts are much lower than in Bag 1, showing the much lower emission levels during hot-running operation. Models for acetaldehyde and ethanol are reduced to ethanol effects, whereas that for formaldehyde includes a negative term for T90. The relative sizes of the ethanol effects for the aldehydes and ethanol in Bag 2 are similar to those for Bag 1, i.e., ethanol has the largest effect, followed by acetaldehyde and formaldehyde. Residual error variances in Bag 2 are consistently higher than in Bag 1, showing the expected result that relative variability is considerably higher for running emissions in Bag 2 than in Bag 1 which is dominated by the start increment.

Table 90. Coefficients for Reduced Models, by Compound, for Bag 1 (Cold-start Emissions).

Model term	Notation	Compound						
		Acetaldehyde	Formaldehyde	Acrolein	Ethanol	Benzene	1,3-Butadiene	Ethane
Intercept	*Intercept*	-5.2323	-5.9771	-7.9338	-4.9080	-4.1074	-5.8365	-4.3412
etOH	Z_e	0.81449	0.2299	0.2476	1.4627			0.05222
Arom	Z_a	0.03483	0.02822	0.1122		0.4032		-0.1925
RVP	Z_r	-0.04170	-0.04718	-0.06450	-0.06054			
T50	Z_5	0.08670	0.1672	0.1880	0.07029		0.1334	0.1830
T90	Z_9	0.03801	0.1302	0.2489	-0.09923		0.09828	
etOH × etOH	ZZ_{ee}	-0.1669		-0.08310	-0.4970			
T50 × T50	ZZ_{55}	0.06665	0.05262		0.1108			
etOH × Arom	ZZ_{ea}	0.01840	0.01651					
etOH × RVP	ZZ_{er}	0.02194						
etOH × T50	ZZ_{e5}		-0.01627	-0.1186				
etOH × T90	ZZ_{e9}		0.02004	0.04617				
Arom × RVP	ZZ_{ar}							
Arom × T50	ZZ_{a5}							
Arom × T90	ZZ_{a9}							
T90 × T90	ZZ_{99}							
T50 × T90	ZZ_{59}	0.03959	0.03489	0.05986				
RVP × T90	ZZ_{r9}							
Vehicle	σ^2_{veh}	0.1149	0.3358	0.1032	0.1283	0.2739	0.0616	0.2454
residual	σ^2_{ε}	0.0885	0.1407	0.3629	0.5730	0.1896	0.0690	0.0296

Table 91. Coefficients for Reduced Models, by Compound, for Bag 2.

Model term	Notation	Compound						
		Acetaldehyde	Formaldehyde	Acrolein	Ethanol	Benzene	1,3-Butadiene	Ethane
Intercept	*Intercept*	-9.4192	-8.6574		-9.5634			-7.7241
etOH	Z_e	0.1910	0.07804	NO	0.8163	NO	NO	
Arom	Z_a			MODEL		MODEL	MODEL	-0.1092
T50	Z_5							0.1452
T90	Z_9		-0.08322					0.1270
Vehicle	σ^2_{veh}	0.05372	0.08239		0.4634			2.6669
residual	σ^2_{ε}	0.4153	0.3776		1.1682			0.1517

9.2.1 Model Fitting under the Reduced Design

In section 8.2 above, we described the evaluation of the *G*-efficiency for the reduced design. The conclusion was that the efficiency of the reduced design was adequate to allow modeling based on a 4-term full model (Model 1.c, Table 78, page 189). During analysis, it is possible to follow the preliminary evaluation of design efficiency with a more direct assessment of the adequacy of the reduced design in model fitting. Specifically, for selected compounds measured under the full design, it is possible to compare model coefficients for models fit under both full and reduced designs.

The secondary evaluation is useful because in Bag 1, three compounds were measured a variant of the reduced design. Benzene, 1,3-butadiene and ethane were measured on the reduced fuel set (11 fuels) , but on the entire set of fifteen vehicles.

We performed these analyses for NMOG (Bag 1) and (Bag 2), as well as Ethane (Bag 1), Acetaldehyde (Bag 1) and Formaldehyde (Bag 1).

9.2.1.1 NMOG (Bag 1)

To more directly evaluate the utility of the reduced design in estimating fuel property effects, we fit models for NMOG (Bag 1) using the full design and the reduced design as applied to benzene, 1,3-butadiene and ethane.

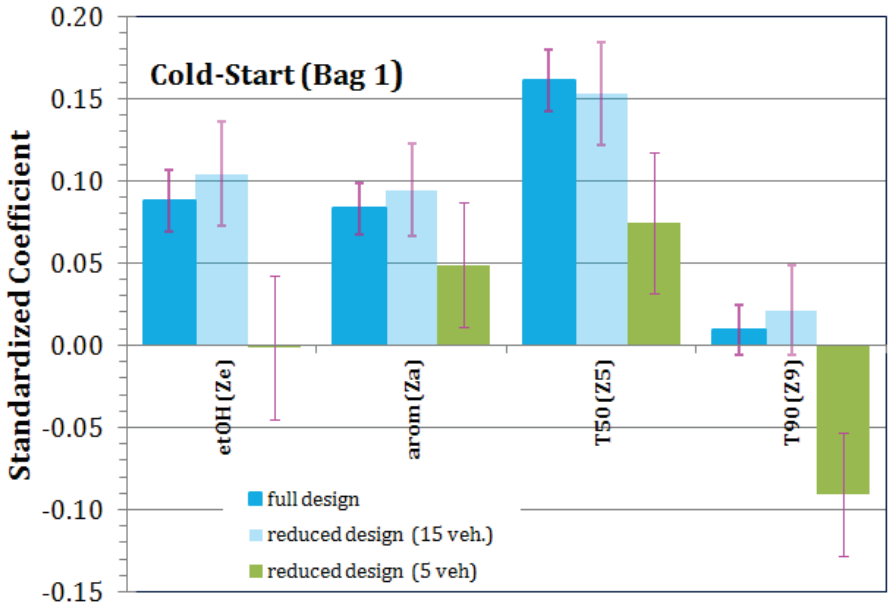
The results, shown in Table 92 and Figure 105, suggest that the fuel effects estimated from the reduced design (with 15 vehicles) are in agreement with those estimated from the full design. For the four fuel-property terms, the mean coefficients for the reduced design fall within the 90% confidence intervals for the full-design values. Results for both levels of the design also suggest that were model fitting performed, the T90 term would be dropped from the reduced model, while the remaining three linear effects would be retained. Not unexpectedly, the figure shows that the uncertainty of coefficients is larger for the reduced design than for the full design, reflecting the higher *G*-efficiency of the full design.

Table 92. NMOG (Bag 1): Coefficients for Four-term models fit using Full and Reduced designs.

Model Term	Notation	Full design[1]	Reduced Design (15 veh.)[2]		Reduced Design (5 veh.)[3]
		Full[3]	Full[4]	Best fit	Full[5]
Intercept	*Intercept*	-0.9548	-0.8943	-0.8939	-0.9987
etOH	Z_e	0.08786	0.1040	0.1020	-0.00148
Arom	Z_a	0.08325	0.09435	0.09193	0.04899
T50	Z_5	0.1611	0.1527	0.1518	0.07441
T90	Z_9	0.009286	0.02127		-0.09081
Vehicle	σ^2_{veh}	0.1222	0.1091	0.1090	0.07728
residual	σ^2_{ε}	0.07989	0.08907	0.08948	0.05717

[1] Includes 15 vehicles and 27 fuels, with replication.
[2] Includes fifteen vehicles measured on 11 fuels.
[3] Includes five vehicles measured over 11 fuels.
[4] All terms highly significant, except for T90.
[5] All terms significant, except for ethanol.

Figure 105. NMOG (Bag 1): Coefficients for fuel-property effects for model fits using full and reduced designs (as defined in Table 92) (Error bars represent 90% confidence intervals).



9.2.1.2 NMOG (Bag 2)

We repeated these steps for NMOG (Bag 2). However, in this case we applied the variant of the reduced design applied in the Bag 2 measurements. Accordingly, these models were fit using measurements for 11 fuels on five vehicles. For the "full design" the results represent the measurements on 13 vehicles on 27 fuels, with replication. The total included 13, rather than 15

vehicles, because we did not include the Odyssey or Sienna, as previously described (see 6.1.3, page 122).

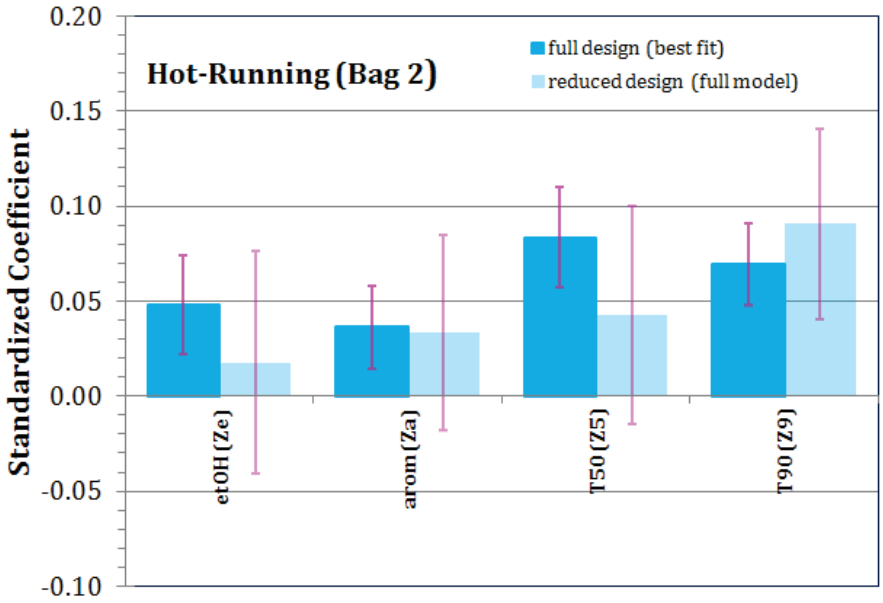
For the full dataset, coefficients for all four terms are highly significant. Thus in this case, the "full model" coincides with the "best-fit" or reduced model. However, model fitting with the reduced design gives a different picture. The lower efficiency of the reduced design is apparent in margins of error roughly twice as large as for the full design (Figure 106). If model fitting is performed for the reduced design, using the same criteria applied for the full design, a single term, T90 is retained in the best-fit model. It is clear that a reduced model under the reduced design does not give estimates of fuel effects similar to the corresponding reduced model under the full design.

Table 93. NMOG (Bag 2): Coefficients for Four-term models fit using Full and Reduced designs.

Model Term	Notation	Full design[1]	Reduced Design[2]	
		Full[3]	Full	Best fit
Intercept	*Intercept*	-5.2388	-4.7775	-4.7768
etOH	Z_e	0.04792	0.01778	
Arom	Z_a	0.03600	0.03320	
T50	Z_5	0.08332	0.04258	
T90	Z_9	0.06919	0.09051	0.08393
Vehicle	σ^2_{veh}	0.8458	1.1405	1.1448
residual	σ^2_{ε}	0.1348	0.1026	0.1052

[1] Includes 13 vehicles and 11 fuels, with replication.
[2] Includes five vehicles and 11 fuels.
[3] All terms highly significant; full model also qualifies as "best fit."

Figure 106. NMOG (Bag 2): Coefficients for fuel-property effects for model fits using full and reduced designs (as defined in Table 92) (Error bars represent 90% confidence intervals).



9.2.1.3 Aldehydes

We performed similar analyses for the two aldehydes measured in the program. Because Bag 1 emissions were measured under the full design, they also provided an opportunity to compare modeling results obtained under the reduced design to corresponding results under the full design.

For acetaldehyde (Bag 1) we included a fifth term (etOH×etOH) in the full model (see Table 94). We made this exception because of the marked curvature in the ln(acetaldehyde) vs. ethanol trend, which requires inclusion of the quadratic term to get satisfactory model fits (see Figure 93). For the full design, the full five-term model also qualifies as a "reduced" model as all terms are significant. For the reduced design, the reduced model contains two fewer terms than the full model. The T50 linear effect is marginally significant at the 90% confidence level and the T90 linear effect is much smaller than its standard error (see Figure 107).

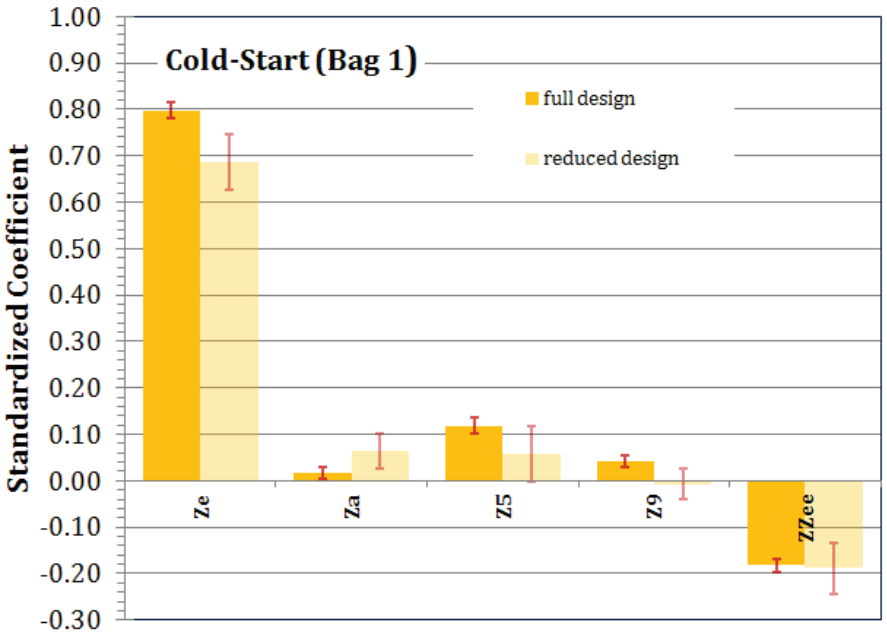
Relative to the full design, the reduced design estimates the dominant effects to within margins of 14% for the ethanol linear term, and -3% for the ethanol quadratic term. Differences are larger for the remaining effects, ranging from -50% for T50 to +280% for aromatics.

Table 94. Acetaldehyde (Bag 1): Coefficients for Five-term models fit using Full and Reduced Designs.

Model Term	Notation	Full design[1]	Reduced Design[2]	
		Full[3]	Full	Best fit
Intercept	*Intercept*	-5.1667	-4.9682	-4.9694
etOH	Z_e	0.7976	0.6862	0.6397
Arom	Z_a	0.01667	0.06360	0.08083
T50	Z_5	0.1181	0.05774	
T90	Z_9	0.04180	-0.00816	
etOH×etOH	ZZ_{ee}	-0.1823	-0.1882	-0.2289
Vehicle	σ^2_{veh}	0.1140	0.05896	0.05915
residual	σ^2_{ε}	0.1160	0.1027	0.1052

[1] Includes 15 vehicles and 27 fuels, with replication.
[2] Includes five vehicles and 11 fuels.
[3] All terms significant; full model also qualifies as "best fit."

Figure 107. Acetaldehyde (Bag 1): Coefficients for fuel-property effects for model fits using full and reduced designs (as defined in Table 92) (Error bars represent 90% confidence intervals).



Similar results for formaldehyde are shown in Table 95 and Figure 108. The full design has significant effects for all properties except aromatics, which is marginally significant at the 90% confidence level. For the reduced design, all four effects appear significant. All effects agree in sign between the two designs, but differ in size. The reduced-design effect is larger for aromatics, but smaller for the remaining three effects. Coefficients for ethanol, T50 and T90

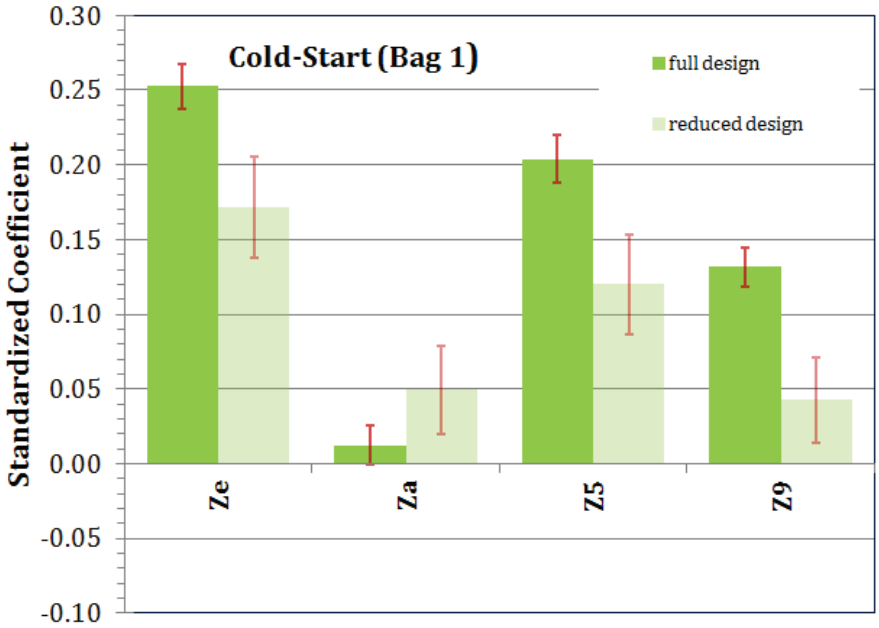
from the reduced design are 32%, 41% and 68% smaller than their counterparts from the full design, respectively.

Table 95. Formaldehyde (Bag 1): Coefficients for models fit using Full and Reduced Designs.

Model Term	Notation	Full design[1]		Reduced Design[2]	
		Full	Best fit	Full[3]	Best fit
Intercept	*Intercept*	-5.9671	-5.9671	-5.7087	
etOH	Z_e	0.2524	0.2507	0.1720	
Arom	Z_a	0.01231		0.04935	
T50	Z_5	0.2039	0.2018	0.1199	
T90	Z_9	0.1315	0.1313	0.04251	
Vehicle	σ^2_{veh}	0.3351	0.3353	0.2196	
residual	σ^2_{ε}	0.1690	0.1695	0.1216	

[1] Includes 15 vehicles and 27 fuels, with replication.
[2] Includes five vehicles and 11 fuels, without replication.
[3] All terms significant; full model also qualifies as "best fit."

Figure 108. Formaldehyde (Bag 1): Coefficients for fuel-property effects for model fits using full and reduced designs (as defined in Table 92) (Error bars represent 90% confidence intervals).



9.2.2 Models selected for Application

Results presented the sub-section 9.2.1 above raise questions about the adequacy of the reduced design to support model fitting, depending in some degree on the number of vehicles included in the subsamples of data.

In Bag 1, where data for all 15 vehicles is available for the reduced subset of fuels, the results for NMOG suggest that the reduced design with all vehicles can estimate fuel property effects that are comparable to those estimated under the full design (Figure 105).

However, results also suggest that the reduced design with only five vehicles is not adequate to generate estimates fuel property effects comparable to those from the full design. The main reason for this conclusion is that reduced, or "best fit" models lack terms that the full design indicates should be present. This point is illustrated by the results for NMOG (Bag 2) shown above in Table 93 and Figure 106. The best-fit model under the reduced design lacks terms for ethanol, aromatics and T50, retaining only a single term for T90. However, in the corresponding models fit under the full design, these terms are present and significant.

Results for aldehydes in Bag 1, in which models fit under the full and reduced designs can be compared, give a similar impression. The best fit model for acetaldehyde (Figure 107) lacks terms for T50 and T90 that are present in the full-design model. By contrast, the reduced model for formaldehyde (Figure 108) includes an aromatics term not present in the best-fit model under the full design. For the bag-1 aldehydes, the models under the reduced design may be better off than corresponding models in Bag 2, in that they contain replicate measurements lacking in Bag 2. In Bag 1, the presence of the replicates roughly doubles the number of measurements available, which may improve the models' precision and power.

Overall, it appears that model fitting under the reduced design is prone to errors in which terms that retained under the full design are rejected as insignificant. From these results we conclude that best-fit models under the reduced design are not adequate to represent the behavior of emissions in response to changing fuel properties.

An additional consideration is that the fuel effects estimated under the reduced design with five vehicles can differ from those estimated under the full design by margins ranging from 30% to 300%. This result suggests that coefficients estimated from the reduced design are subject to error in terms of magnitude. However, for compounds measured only under the reduced design, the available measurements comprise the best information available relating emissions to fuel properties.

Thus, for purposes of application, we have elected to use full rather than reduced models for the compounds included in this analysis and measured under the reduced design, including acetaldehyde, formaldehyde, acrolein, benzene, 1,3-butadiene and ethane. Coefficients for Bag-

1 models, representing cold-start emissions, and Bag-2 models, representing hot-running emissions, are presented below in Table 96 and Table 97, respectively.

For compounds in Bag 1 measured under the full design, reduced models are retained for application. Coefficients for these models, previously presented in Table 90 (page 236), are summarized in Table 98.

Table 96. Coefficients for Full Models, for Bag 1, for Compounds measured under the reduced Design[1].

Model term	Notation	Model		
		Benzene	1,3-Butadiene	Ethane
Intercept	*Intercept*	-4.1029	-5.8371	-4.3079
etOH	Z_e	-0.00468	-0.01729	0.1204
Arom	Z_a	0.4056	0.02673	-0.1728
T50	Z_5	0.04242	0.1247	0.2169
T90	Z_9	0.01133	0.1004	0.09531
Vehicle	σ^2_{veh}	0.2741	0.2192	0.1407
residual	σ^2_{ε}	0.1873	0.1089	0.04970

[1] For these models, "reduced design" signifies 15 vehicles measured on 11 fuels.

Table 97. Coefficients for Full Models, for Bag 2, for Compounds measured under the Reduced Design[1]

Model term	Notation	Compound						
		Acetaldehyde	Formaldehyde	Acrolein	Ethanol	Benzene	1,3-Butadiene	Ethane
Intercept	*Intercept*	-9.4189	-8.6574	NO MODEL	-9.3072	NO MODEL	NO MODEL	-7.7241
etOH	Z_e	0.1520	0.08456		0.9233			0.07345
Arom	Z_a	0.07991	0.01575		-0.3772			-0.1260
T50	Z_5	-0.02997	0.01863		-0.01910			0.1815
T90	Z_9	-0.07836	-0.08138		-0.3017			0.1322
Vehicle	σ^2_{veh}	0.05654	0.08205		0.3707			2.6785
residual	σ^2_{ε}	0.3814	0.3762		1.0889			0.1458

[1] For these models, "reduced design" signifies 5 vehicles measured on 11 fuels.

Table 98. Coefficients for Reduced Models, for Bag 1, for Compounds measured under the full Design[1].

Model term	Notation	Compound			
		Acetaldehyde	Formaldehyde	Acrolein	Ethanol
Intercept	*Intercept*	-5.2323	-5.9771	-7.9338	-4.9080
etOH	Z_e	0.81449	0.2299	0.2476	1.4627
Arom	Z_a	0.03483	0.02822	0.1122	
RVP	Z_r	-0.04170	-0.04718	-0.06450	-0.06054
T50	Z_5	0.08670	0.1672	0.1880	0.07029
T90	Z_9	0.03801	0.1302	0.2489	-0.09923
etOH × etOH	ZZ_{ee}	-0.1669		-0.08310	-0.4970
T50 × T50	ZZ_{55}	0.06665	0.05262		0.1108
etOH × Arom	ZZ_{ea}	0.01840	0.01651		
etOH × RVP	ZZ_{er}	0.02194			
etOH × T50	ZZ_{e5}		-0.01627	-0.1186	
etOH × T90	ZZ_{e9}		0.02004	0.04617	
Arom × RVP	ZZ_{ar}				
Arom × T50	ZZ_{a5}				
Arom × T90	ZZ_{a9}				
T90 × T90	ZZ_{99}				
T50 × T90	ZZ_{59}	0.03959	0.03489	0.05986	
RVP × T90	ZZ_{r9}				
Vehicle	σ^2_{veh}	0.1149	0.3358	0.1032	0.1283
residual	σ^2_{ε}	0.0885	0.1407	0.3629	0.5730

[1] For these models, "full design" signifies 15 vehicles measured on 27 fuels.

10 References

[1] USEPA Office of Mobile Sources. *Final Regulatory Impact Analysis for Reformulated Gasoline*. EPA-420-R-93-017. Assessment and Standards Division, Ann Arbor, MI. December, 1993.

[2] Technical Support Document: *Analysis of California's Request for Waiver of the Reformulated Gasoline Oxygen Content Requirement for California Covered Areas*. EPA420-R-01-016 (see also 40 CFR 80.45).

[3] Auto/Oil Air Quality Improvement Research Program. AQIRP Program Final Report. January 1997.

[4] USEPA Office of Transportation and Air Quality. *Assessing the Effect of Five Gasoline Properties on Exhaust Emissions from Light-Duty Vehicles certified to Tier-2 Standards: Pilot Studies (EPAct Phases 1 and 2, and FTP Pilot).* Memorandum to Docket EPA-HQ-OAR-2011-0135. Assessment and Standards Division, Ann Arbor, MI. March, 2013.

[5] USEPA Office of Transportation and Air Quality. *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. Assessment and Standards Division, Ann Arbor, MI; National Renewable Energy Laboratory, U.S. Department of Energy, Golden, CO; Coordinating Research Council, Alpharetta, GA. March, 2013.

[6] CRC report No. 648 (2006 CRC Hot-Fuel-Handling Program), Coordinating Research Council, Alpharetta, GA.

[7] Kuhfeld, Warren F. *Efficient Experimental Designs Using Computerized Searches*. Sawtooth Software Research Paper Series. Sawtooth Software, Inc. Sequim, WA. 1997.

[8] Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman, W. 1996. *Applied Linear Statistical Models*. Fourth Ed., Irwin, Chicago.

[9] Gunst, Richard. F. *Statistical Analysis of the Phase 3 Emissions Data Collected in the EPAct/V2/E89 Program: Final Report*. Subcontract LGC-0-40441-01. National Renewable Energy Laboratory, Department of Energy, Golden, CO. July 2011.

[10] Sluder, C. Scott; West, Brian H. 2011. *NMOG Emissions Characterizations and Estimation for Vehicles using Ethanol-blended Fuels*. ORNL/TM-2011/461. Fuels, Engines and Emissions Research Center, Oak Ridge National Laboratory, Oak Ridge, TN. October 15, 2011.

[11] Clarke, Joan U. 1998. Evaluation of censored data methods to allow statistical comparisons among very small samples with below detection limit observations. *Environmental Science and Technology*. 32: 117-183.

[12] Painter, L.J.; J. A. Rutherford. 1992. *Statistical Design and Analysis Methods for the Auto/Oil Air Quality Research Program*. SAE 920319. Society of Automotive Engineers, Warrendale, PA.

[13] Zhao, Yuchao; Frey, H. Christopher. 2004. Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. *Risk Analysis* 24(4): 1019-1034.

[14] Maddala; G.S. *Introduction to Econometrics*. Macmillan, New York. 1998. (p. 283 ff.).

[15] SAS Institute, Inc. SAS/STAT User's Guide, Version 6, Fourth Edition: Volume 2. Cary, NC. 1990. (PROC LIFEREG).

[16] Schabenberger, O. *Mixed Model Influence Diagnostics*. Paper 189-29 *In*: SAS Institute, Inc. Proceedings, Twenty-Ninth Annual SAS Users Group International Conference (SUGI 29), Montreal, Canada May 9-12, 2004.

[17] Mason, Robert L.; Buckingham, Janet P. *Re-design of Fuel Matrices for Epact Program*. Memorandum for Contract EP-C-07-028. Southwest Research Institute, Austin, TX. September, 2008. (Appendix A to Testing Report, citation 5).

[18] Lewis, Jeffrey B.; Linzer, Drew A. 2005. Estimating regression models in which the dependent variable is based on estimates. *Political Analysis* 13(4): 345-364.

[19] Maddala; G.S. *Introduction to Econometrics*. Macmillan, New York. 1998. (p. 283 ff.).

[20] SAS Institute, Inc. SAS/STAT User's Guide, Version 6, Fourth Edition: Volume 2. Cary, NC. 1990. (PROC LIFEREG).

[21] Kish, Leslie. *Survey Sampling*. John Wiley & Sons, New York. 1965. (page 29).

[22] Hochhauser, A.M. (2008). *Review of Prior Studies of Fuel Effects on Vehicle Emissions*. Coordinating Research Council, Alpharetta, Georgia. CRC Report No. E-84 (www.crcao.org).

[23] USEPA Office of Transportation and Air Quality. *Renewable Fuel Standard Program (RFS2): Regulatory Impact Analysis*. EPA-420-R-10-006. Assessment and Standards Division, Ann Arbor, MI. February, 2010. (Chapter 3).

[24] Aikawa, K.; Sakurai, T.; Jetter, J. 2010. Development of a predictive model for gasoline vehicle particulate matter emissions. *SAE Int J Fuels Lubr.* Vol. 3, Paper No. 2010-01-2115.