Review of EPA Draft Report

The Effects of Fuel Sulfur Level on Emissions from the In-Use Tier 2 Vehicles

Report Dated: November, 2011

This draft report describes a vehicle test program conducted to assess the emissions impacts of reducing gasoline sulfur levels in Tier 2 passenger cars and light-duty trucks. The experimental test design was suitable for the intended purposes, although clearly, other testing procedures could have been used as well. For example, the standard FTP test procedure was used, at 75°F, to determine the emissions results from each vehicle test. Some comments should be offered regarding the suitability of the FTP driving cycle and a 75°F test temperature in providing representative/realistic operating conditions. For the in-use vehicle fleet, a set of standardized test conditions is clearly necessary to obtain repeatable and statistically significant results. However, the FTP test is now over 35 years old, and 75°F is just one of many temperatures that could be used. The overall quantitative effect of fuel sulfur on emissions as determined in this study would be more convincing if it were shown that similar results apply to other realistic operating conditions.

The use of two, back-to-back US06 dynamometer cycles was an appropriate choice for catalyst regeneration. Although a somewhat more aggressive cycle might provide more complete desulfurization, potential damage to the vehicle was a significant concern, as explained in this report. Also, the approach of catalyst clean-out, followed by repeated testing with a single fuel is a valid way to determine a new “equilibrium emissions level” at a specific fuel sulfur level. However, in real-world driving, such an equilibrium level is never attained, because operating conditions are constantly changing, and fuel sulfur varies from tank-to-tank.

It is not clear why this report places so much emphasis on Bag 2 results – especially Bag 2 NOx. Typically, Bag 2 emissions concentrations are extremely low from properly functioning Tier 2 vehicles. In fact, as shown in this study, it is often difficult to distinguish Bag 2 levels from background levels. Large percentage differences between two very small numbers may not be very meaningful. To help put these results in perspective, it would be useful to show the relevant certification emissions levels for these Tier 2 vehicles. Also, something should be said about the impact of fuel sulfur reduction on fleet-wide emissions, and how large a reduction this represents in comparison to the entire mobile-source emissions inventory. In reality, this test program only captured a small portion of the entire in-use fleet (only Tier 2 vehicles of model years 2007 - 2009).

Some historical perspective about the “FTP bag method” would be helpful here. This sampling and analysis method was developed about 40 years ago, to measure emissions from uncontrolled (or slightly controlled) vehicles. Prior to introduction of emissions control systems, Bag 2 emission levels were routinely 2 orders of magnitude higher than today. Under such conditions, the FTP bag method was perfectly suitable for the intended purpose. Now, with Bag 2 emission levels being nearly indistinguishable from background, a different method should be employed for quantifying vehicle emissions.

In retrospect, it seems unfortunate that only non-ethanol gasolines were used, as such fuels are increasingly uncommon. It is stated that similar results would be expected from ethanol-containing gasolines, but no data are provided to support this. If any subsequent testing has been done with ethanol-containing gasolines, this should be mentioned.

Overall, the statistical modeling approach seems appropriate. As described in Section 7.2 (pages 26-28) the structure and limitations of the emissions dataset make the linear mixed model a good choice for analysis, as this better accommodates missing data, irregularly spaced measurements, and within-vehicle effects.

Also, the method used to impute emissions values in cases where measured levels were lower than background is probably the best that can be done. The approach of developing an imputation value based upon each specific vehicle family seems better than using a single, fleet-wide imputation value. The fact that imputed values were applied to only about 20% of the vehicles provides some assurance that this imputation method did not excessively distort the results. This was further confirmed by performing the statistical analyses with and without inclusion of the imputed values. One additional test that could be useful is to conduct the statistical analyses using values of zero instead of the imputed emissions values.

A significant question about the statistical modeling approach is the failure to include any information about the vehicles' catalyst formulations, or other catalyst properties. It is not surprising to observe that vehicle type (car vs. truck) was not a significant model term, but it is surprising that there is no term related to the emissions control systems used across the range of vehicles. Surely not all catalysts were the same, and different catalysts would be expected to respond to changes in sulfur levels in different ways. This is an area that should be addressed. Inclusion of a catalyst term could result in quite different statistical models.

A number of other specific comments, questions, and suggestions regarding this draft report are offered below. The ordering of these items is chronological, as they appear in the report.

- On page 2 (1st paragraph) the quantity of sulfur present on the vehicle's catalyst is said to be a function of temperature and fuel sulfur level. Isn't the catalyst formulation/metallurgy another important factor? (Catalyst formulation is mentioned on page 7.) What is known about the catalyst formulations used in each of the test vehicles of this study?

- It is interesting to note that when reducing fuel sulfur from 28 ppm to 5 ppm, significant emissions reductions were observed for all pollutants except PM, (See Tables ES-2 and ES-3.) Is there a clear mechanistic reason for this lack of effect for PM? Are the catalyst systems used in Tier 2 vehicles ineffective in reducing PM (and PM precursors)? This lack of a PM effect should be mentioned along with the major findings that are shown in bullet form on page 6.

- On page 7 (final paragraph) it is stated that the impact of fuel sulfur on emissions was considered negligible under the Tier 0, Tier 1, and NLEV programs. I don't think this is true. The emissions impacts of fuel sulfur have been well known for a long time, but more severe sulfur reductions were not thought to be a cost-effective emissions control approach until recently. To provide greater context, it would be useful to include a brief summary of gasoline regulations (including sulfur levels), and how they have evolved over the past 30-years. Along with this, a history of LD vehicle emissions standards should also be presented.

- On page 8, various detrimental effects of hot/rich catalyst operation are described – including catalyst degradation and increased emissions of PM, NMOG, and CO. However, nothing is said about fuel economy effects – either here or elsewhere in the report. In general, fuel economy is a concern with any type of fuel or vehicle modification. This topic should at least be mentioned. In addition, the dataset generated in this study provides an excellent opportunity to evaluate the impacts of sulfur reduction on fuel economy, by analyzing CO_2 as another pollutant, along with CO, THC, NOx, and PM.

- The 2005 MSAT study, which used nine Tier 2 vehicles, is described on pages 8-9. The emissions reductions upon lowering fuel sulfur from 32 to 6 ppm are said to be 33% for NOx, 11% for THC, 17% for CO, and 32% for CH_4. Are these values simple averages of the nine vehicles, or were they computed by a more sophisticated method? Were similar statistical methodologies used to compute the fuel sulfur effects in both the MSAT study and the present study? Were PM emissions also measured in the MSAT study?

- When first describing the two test fuels on page 10, it would be helpful to refer the reader to Table 5-1 (page 15), which provides a more complete listing of fuel properties. Also, did these fuels contain anything unusual in the way of additives; e.g. antioxidants, detergents, dyes, etc.?

- In selecting the vehicles, was the ratio of LD trucks/LD passenger cars representative of the in-use fleet? Table 4-1 (page 13) which lists the 19 makes/models that were recruited, shows that the No. 1 U.S. sales rank vehicle is missing. Was this an oversight, or are the rankings shown in this table incorrect? Notice also that the Toyota Camry is shown as No. 23 in U.S. sales rank. This seems unbelievably low for what has traditionally been a "top seller." Also, it would be useful to add a column to Table 4-1 to indicate the number of vehicles in each category that were tested in this study, comprising the total test fleet of 81. It might also be useful to indicate here the number of vehicles in each class that underwent the various test procedures: i.e. short procedure, long procedure, modified short procedure, and modified long procedure.

- The modified long testing procedure shown in Figure 6-2 (page 19) is confusing. The blue box indicating the short procedure shows only two post cleanout FTPs, although the wording still indicates "triplicate FTPs at 28 ppmS." Also, in describing the modified short procedure on page 20, it is said that "… the change in the number of vehicles providing sulfur level data can be seen in Table 7-7 starting with Family ID N513." It is not clear what this means. Does this table list the vehicle families in chronological testing order, so that the first 13 families listed were tested using the standard short procedure, and the last 4 families listed were tested using the extended short test? It is also confusing that Tables 7-5 and 7-7 provide identical information.

- How is NMOG different from NMHC? What analytical procedures were used to measure these two pollutants?

- Displaying the background and sample measurements by FTP Bag (as done in Figure 7-1 and Appendix B) is very instructive. However, only NOx and THC results are shown in this way; it would be useful to also include CO and PM. It would also be helpful to draw ovals in Figure 7-1 to capture all background measurements for each vehicle family shown. Although

it is not possible to tell which sample points correspond to which background points, this graphical approach clearly illustrates the problem of very low emissions measurements relative to background in Bag 2. It also begs the question of why background levels of NOx are so variable. Figure 7-1 shows that these background concentrations varied by over an order of magnitude for many of the vehicle families tested.

- In the determination of outlier data points (described on page 25) what is the rationale for choosing an outlier screening criterion value of ± 3.5 for the studentized residuals? Also, although Table 7-2 identifies the number of outlier points for each pollutant/bag, it doesn't indicate how many of these outlier points were actually excluded from the statistical analyses.

- As explained on page 26, log-transformation of emissions measurements has commonly been used to analyze vehicle emissions data. However, many previous emissions modeling studies have utilized test fleets that included a variety of technology types having a wide range of emission levels. In the present study, only Tier 2 vehicles were used, and the emission levels did not vary drastically across the test fleet. Given this situation, is log-transformation still necessary (and helpful)?

- On page 29 it is stated that the average starting odometer reading of the vehicles used to assess the clean-out effect at 28 ppmS was 31,470 miles. Is this value the average of the 17 vehicle families shown in Table 7-3, or the average of all 81 vehicles? Similarly, is the ± value of 1,578 miles the standard deviation of the 17 vehicle families or the 81 individual vehicles? (The same questions pertain to the vehicles used to assess the clean-out effect at 5 ppmS, described on page 35; and the vehicles used to assess the sulfur effects, described on page 39.)

- The box plots shown in Figure 7-2 (page 31) are quite informative in displaying the relative variances between the pre- and post-cleanout vehicle tests. It would be useful to show similar plots for other pollutants and Bags. Also, a legend should be included to explain the different symbols used in these plots.

- When describing the dependent variable (Y_i) and effects (X_i and Z_i) on pages 31-32, the reader should be reminded of the mixed model being used, shown as Equation 7-1 on page 28. Also, it is stated on page 32 that "The significance of between-family variation was observed graphically in Figure 7-2 …" But this figure only show Bag 2 NOx results. Was similar between-family variation observed with other Bags and pollutants?

- Figure 7-3 (pages 37-38) is not easy for the reader to process. I realize that the data are lumped by emission level, but displaying the data in three separate charts, over two pages, is confusing and difficult to understand. It would be preferable to show these data on a single page (similar to Figure 7-2, page 31) or even in a single chart (similar to Figure 7-1, using a logarithmic scale). It would also be better to order the vehicles along the x-axis in the same way they are listed in Table 7-5. Finally, it would be helpful to see more Bags/pollutants presented in this way – not just Bag 2 NOx.

- On page 38, the enhanced emissions reduction benefits of the 5 ppmS clean-out are described as compared to the 28 ppmS clean-out (comparing Tables 7-4 and 7-6). It is

interesting to note that this does not apply to PM emissions. Some explanation should be offered as to why PM emissions show such a different behavior. Also, the last sentence on page 38 mentions confirmation of results that have not even been presented yet. The placement of this sentence seems odd.

- The same comments given above regarding data display in Figure 7-3 apply to Figure 7-4 (pages 41-42). The current grouping of vehicles shown in Figure 7-4 makes it very difficult to see the important point being made on page 40 that some vehicles within the same family had markedly different emission profiles. This would be much easier to see if a single chart were used to display all vehicles – similar in structure to Figure 7-1. Also, the box plots in Figure 7-5 (pages 43-44) should all appear on a single page, and should have a legend to explain the symbols being used. Further, it would be helpful to identify each vehicle ID number shown in Figure 7-5 by make/model; e.g. Vehicle ID 0003 is a Toyota Corolla. Again, additional figures should be shown for other Bags and pollutants.

- On page 46, the term “BIC” is used without definition. The reader must refer to page 32 to see that BIC refers to the Schwarz Bayesian Criterion. It would be helpful to include a table of acronyms and abbreviations, as many rather obscure terms are used throughout this report.

- Figure 7-8 (page 50) is very helpful in convincingly demonstrating the reasonableness of the derived statistical model in predicting Bag 2 NOx emissions with both high and low sulfur levels. It would be useful to include similar plots for other Bags/pollutants.

- At the bottom of page 50, where Figure 7-9 is being discussed, it is stated that “the rate of sulfur loading is the same for both high and low sulfur levels.” What is the basis for this statement? Sulfur loading is not measured directly. Is NOx emission rate taken as a surrogate for sulfur loading? (This statement about sulfur loading is also given in the Summary and Conclusions Section on page 59.) Also, Figure 7-9 shows quite clearly that for some vehicles, not only are the emission rates higher with the high sulfur fuel, but also the increase in emissions with mileage accumulation is higher. For example (see vehicles 0075, 0123, 0264, 0178, and 0179). This seems inconsistent with the statement that rate of sulfur loading is the same for both high and low sulfur levels.

- In Table 7-10 (page 53) the fixed effects in the NOx statistical model include only sulfur level and miles for all three FTP Bags. Yet the FTP Composite result includes a sulfur level by miles interaction term. This seems strange, since the FTP Composite emissions are simply calculated as a combination of the individual Bag emissions. Some explanation should be offered.

- Pages 53-54 discuss the problem of comparing emissions from high and low sulfur levels when there is a sulfur level by mileage interaction. The approach taken to address this (as illustrated in Figure 7-10) seems rather arbitrary, although it may be as good as any other approach. However, to explore the impact of this selection, it might be good to conduct some sensitivity analyses using other methods.

- On page 55, it is stated that "For all models except CO Bag 1, CO Bag 1 – Bag 3, and CH_4 Bag 1 – Bag 3, the reduction estimates are statistically significant …" However, this is not true for the pollutants NOx + NMOG or PM.

- Were the sensitivity analyses described in Section 7.3.4 (pages 56-59) only conducted for Bag 2 NOx? If so, this seems like a major limitation. Also, at the top of page 59 it is mentioned that even when removing the influential vehicles from the analysis, sulfur effect "is still highly significant." What are the significance levels before and after removing these influential vehicles?

- In the Summary and Conclusions Section (page 59) the first bulleted finding should be clarified to indicate that the stated NOx, NMHC, and PM reduction values were obtained using a clean-out procedure with 28 ppmS. The second bulleted finding should be clarified to indicate that although clean-out with 5 ppmS fuel rather than 28 ppmS fuel further reduced NOx, NMHC, and CO, no effect was observed for PM.

- In discussing the sensitivity analyses within the Summary and Conclusions Section (page 60) it should be stated that these analyses were only performed for Bag 2 NOx emissions (if this is true).

A number of typographical and other errors were also noted:

- Page 3, footnote: "The program has collected additional data that is being incorporated …"
- Page 5, 5th line: Insert period after "overall."
- Page 6, footnote to Table ES-3: It is stated that reduction estimates were computed differently for Bag 1 THC and CH_4 because these clean-out effects were not statistically significant. Table ES-1 shows that the clean-out effects were also not statistically significant for Bag 1 NOx and NMHC.
- Page 6, next to last line: "… are consistent with those formed in the MSAT and Unicor studies …"
- Page 17, Figure 6-1: The sulfur clean-out cycle in the "short procedure" (blue-colored box) should indicate 28 ppmS, not 30 ppmS.
- Page 29, 2nd line: "In addition, the results from the clean-out data were used to supplement …"
- Page 30, 4th line: "… representing the as received sulfur level …"
- Page 32, 2nd line: "… vehicles from the same vehicle family have ~~the~~ similar emission profiles."
- Page 40, 4th line: "However, the sulfur loading effect certainly varies by vehicle ..."
- Page 58, 1st ¶, 5th line: "… which is calculated after an iterative process …"