UNIVERSITY OF CALIFORNIA, RIVERSIDE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

COLLEGE OF ENGINEERING
CENTER FOR ENVIRONMENTAL RESEARCH AND TECHNOLOGY
1084 Columbia Ave. Riverside, CA 92507

(951) 781-5791 FAX (951) 781-5790
http://www.cert.ucr.edu

January 25, 2012

RE: Peer review comments on in-use sulfur effects study

Dear Mr. Menard:

Please find attached my comments on the document "The Effects of Fuel Sulfur Level on Emissions from the In-Use Tier 2 Vehicles". I reviewed the document itself, as well as the underlying statistical concepts used in the analysis. I work as a research engineer for the University of California at Riverside, CE-CERT and have no real or perceived conflict of interest related to this evaluation. I have considerable expertise in emissions testing and have conducted several major emissions test programs related to gasoline fuel properties, and the analysis of such data. Please let me know if you would like further information or have further questions relating to my comments, or would like to discuss the comments via a conference call.

Regards,
Tom Durbin, Ph.D
Research Engineer
University of California
CE-CERT
Riverside, CA 92521

Comments on

US EPA Report "Effects of Fuel Sulfur Level on Emissions from the In-Use Tier 2 Vehicles"

This report presents an interesting approach to estimating the real effects of fuel sulfur in the Tier 2 in-use fleet. The study was very extensive in terms of the number of vehicles and level of testing that was done, and the results of the study some potentially important impacts of differing in-use sulfur levels. It would be interesting to evaluate these results further to determine the modeled impact of in-use fuel sulfur levels on ambient quality. The analysis methodologies in the report appear to be statistically sound, and appropriate for repeated measures types of testing. The presentation of the data is relatively extensive in terms of the descriptions of the statistical analysis procedures used. As such, the report would relatively straightforward to read for the some with a moderately strong background in statistics, but might a little too technical for some readers without a strong statistical background. Detailed comments related to this report are provided below in three different areas 1) general technical content 2) editorial comments, and 3) answers to assigned peer review questions.

1. Executive Summary. No significant comments outside of editorial comments provided below.

2. Introduction – See editorial comments below. Also, it might be useful to discuss the Umicore study cited in the Executive summary, since usually the executive summary is more condensed than the Background.

3. Study Design

   Was the fuel commercially available in the Ann Arbor area at the time of the study oxygenated or not. This is of relevance since the test fuel is non-oxygenated.

   30 ppm is given as the in-use sulfur average. Is this based on actual data of in-use sulfur samples, or is this based on regulatory average requirements.

   In discussing why the US06 x2 clean-out cycle was used, it might be worth discussing the clean-out procedures that were used in the MSAT and Umicore studies..

4. Test Vehicle Selection, Recruitment, and Delivery

   What is the average sulfur level for the as received vehicles.

   Section 4.1 should specifically mention that 19 makes/models were utilized in the study.

   Was any attempt made to characterize the vehicle categories by other factors such as level of long term fuel trim, or whether the vehicle used a closed coupled or double catalyst or only an underbody, or catalyst metals.

   What incentive was provided to owners to participate in the program? – add to section 4.2.

   More information should be given about the number of vehicles targeted for testing, and the number of vehicles in each vehicle make/model category. The first mention should be here,

as opposed to in section 6.2. In that section, it talks about 5 vehicles for each make/model class. It would be useful to say the total number of vehicles tested at this point.

5. Test Fuel Specs and Procurement

Some more details should be included about the fuel selection would be useful. How does this fuel compare to typical in-use fuel, in terms of say aromatics and RVP. Were oxygenates in use in the recruitment area? Or compared to averages of in-use used throughout the US?

6. Test Procedures

It would be useful to add in fuel change points into the test procedure flow charts.

There is some lack of clarity in reading through this section that can be shown by looking at the last paragraph. The last paragraph indicates that only 4 of 19 family used the modified short procedure, whereas the first sentence of the paragraph indicates that the change was made approximately halfway through the program. Then looking at Table 7.7, it indicates that between 2-4 vehicles in each make/model class were given the modified short procedure, whereas it seems to imply in the paragraph that once the change was made, it was applied thereafter. Was the change just made for a subset of the vehicles tested in each vehicle category. Also, in the last sentence it indicates that the change in the number of vehicles providing "sulfur level data" is seen in Table 7-7. Does this mean that the modified short vehicles were added to the "sulfur level data" as well as the Long or modified Long procedure vehicles.

Related to the above comment, why not simply just say at the end of each procedure description how many vehicles were tested on each procedure.

Figure 6-3 indicates that only the 28 ppm sulfur level data is used for the "clean out effect", but the very first sentence under section 7 indicates that there is a set of "clean-out" data at 5 ppm.

7. Data Analysis and Results

The comments for this section are provided in a combination of some comments here, which are more related to the presentation, flow, or missing information in the section. At the end of this document, addition information is provided to address the questions specifically provided under the reviewer,

p. 21 – 3rd paragraph – A statement should be added to answer the following question – Why do the results of the Bag 2 NOx emissions have more substantial implications than those of other pollutants/bags?

While it is useful to discuss the results at low concentrations, further information on the experimental methods might merit a section. For example, what analyzer are used? Can the detection limits of the analyzers be included in an Appendix? What methods/microbalance was used for the PM measurements, etc.
p. 24 – top of page. There should be a brief paragraph discussing Table 7-1 and the number of zero value measurements before discussing the imput value methodologies.

P. 24 – last sentence – It says that the statistical analyses were run both with and without imputed values. In terms of the "without" input value case, does that mean that the values were eliminated, were zeros used as the imput values, or were the original values used, whether they are positive or negative.

p. 25 – second paragraph – A new terms vehicle "families" is introduced. Presumably this term is the same as the vehicle make/models used earlier. Presumably these also have the same engine family and this should be included along with engine size, vehicle configuration, and weight.

p. 30 – There is no discussion on the symbols in the plots presumably the plots show the average and median inside the bar. Then the bars represent the 95% confidence levels and the error bars the full data range. Then there are some other dots on the plot?

p. 30 – It is not clear what is meant by "Some vehicle families show the presence of within-family variability" Presumably this could also be due to differences in in-use sulfur levels as well, since a handful of vehicles came in with in tank fuels with sulfur levels much higher than those for others. For the post-cleanout variability, it seems like this should be independent of the in-use sulfur loading. Could some of the variability be related to the condition of the vehicle? Vehicle family M504 seems to show a very variability post-cleanout.

P. 32 – Related to BIC criterion. It indicates that the BIC performs relatively better for small sample size settings. How would one define 5 here? Is the current experiment a small sample size?

Figure 7-4 does not appear to add significant value. Much of the same information is available in the box plots in Figure 7-5, and that is a little easier to read. It talks about the vehicles with similar emissions profiles being grouped. However, it is not immediately obvious what grouping were used for the Figures 7-4 and 7-5.

8. Summary and Conclusions

Under the first bullet point. Performing a clean-out cycle **with a 28 ppm fuel**

The second bullet point could be clearer. For example: for a subset up vehicles tested on both 5 ppm and 28 ppm fuels, it was found that additional FTP composite reductions of 18% for NO_x, 9% for NMHC, and 8% for CO were for the 5 ppm fuel in comparison with the 28 ppm fuel.

9. Grammar and other editorial things.

--- Executive summary – the number of spaces between sentences differs from to in some cases 3 or even 4. This should be made consistent.
--- p. 2 The end of the first sentence and start of the second redundantly talk about catalyst efficiency.
--- p. 2 The final sentence of the first paragraph should be split into two sentences.

--- p. 2 Second to last sentence -- rewrite… for a PZEV operated on a 3 ppm fuel compared to a 33 ppm fuel.

--- p. 3 third sentence – ~~In response~~….**To address this question, the** EPA (spell out EPA first use)

--- p. 3 last sentence – split it – performance**. F**or Example,

--- p. 4 2nd line – fuel**s** (add s); Also split the second to last sentence.

--- p. 5 need space for 5 ppm and 28 ppm, and period after "overall". This sentence should probably also be split.

--- p. 5 3rd sentence – rewrite – a significant miles-by-sulfur interaction was not found from the model fitting….

--- p. 5 last sentence – rewrite – In this case, **the relatively differences with sulfur level varied as a function of mileage, so** determination…..

--- p. 7 – 1st sentence – "has long been shown" is ackward. Also in executive summary.

--- p. 8, 9, and 10 – comma after e.g.**,**

--- p. 8 comma after idles)**,** comma after in 2005**,**

--- p. 9 – 1st line – spell out MSAT acronym; line 7 – word "specimens" is a bit ackward; 6th line form bottom "benefits of further sulfur control"

--- p. 11 – about line 10 comma after high temperatures), line 15 "conditions **that are** favorable" line 16 lines further down "A vehicle with **a** relatively…, reference 10 on the E-60 report is not referenced properly and should contain more author names.

--- p. 12 – 1st line operation**,** (add comma); 1st line of section 4.1 "~~with the intent of being~~ **to be** representative of **the** latest…";

--- p. 12 – last sentence table 4-1 should not be underlined. This same changed should be incorporated throughout the report, as it appears there is something in the formatting underlining table and figure references in the text.

--- p. 13 - First line – "Vehicles recruited for testing were targeted to have a mileage between 12,000 and 40,000 miles and an age of less than three years old."

--- p. 14 - line 5 up front bottom – exhaust leaks**,** (add comma)

--- p. 16 - First full paragraph – Following **the** fuel change**,**

--- p. 16 - last sentence – rewrite to "The Long and Short procedures are shown in Figure 6-1 and are discussed in greater detail below. The Long and Short procedures are identical in structure for the first six emissions tests.

--- p. 17 - third line – start a new paragraph about the Long procedure.

--- p. 20 – last sentence – Split into two sentences.

--- p. 21 – 3rd paragraph – detail**.** ~~but t~~The

--- p. 27 – 2nd paragraph line 4 "vehicles which **was** ~~were~~ "…., later in same paragraph Besides this**,** (add comma), 3rd paragraph line 2 – effects**,** (add comma)

--- p. 28 – line right after equation - respectively**,** (add comma)

3rd paragraph – line 2 – model**,** (add comma) – line 4 structure**,** (add comma)

--- p. 29 1st paragraph line 3 exchange "were" for ~~was~~ - line 4 – levels**,** (add comma) -

--- p. 32 – last paragraph line 3 – covariance**,** (add comma)

--- p. 33 line 6 – selected**,** (add comma)
--- p. 35 last paragraph 1st line – "dataset include**d** 17" ……
--- p. 38 – 2nd line from bottom – Section 7.3.3**,** (add comma)
--- p. 40 – 3rd line from bottom – i.e.**,** (add comma)
--- p. 40 – 5th line from bottom – rewrite - considering each vehicle as **a** random effect**s** might be useful
--- p. 42 – The sentence that includes Figure 7-6 should be broken up into two sentences. Example: This is shown in figure 7-6, which shows the log-transformed emissions from individual vehicles by sulfur level.

5th sentence from bottom – add comma … some vehicles**,** and suggests….

--- p. 44 – 2nd line from bottom – add comma … vehicle to vehicle**,** and simple descriptive statistics…
--- p. 45 – 1st line – rewrite .. a similar top-down model fitting statistical approach was applied to the clean-out data, as described in Section 7.3.1.
--- p. 45 – 2nd sentence – rewrite .. Furthermore, additional analysis was done that used a subset of the sulfur level dataset that isolates the emission measurements immediately following the clean-out to address the effectiveness of the clean-out cycle in reducing emissions.

--- p. 46 – 2nd paragraph 4th sentence – add comma … irregularly**,** where
--- p. 46 – 3rd paragraph 2nd sentence – ….exponentially with time**, i.e.,** the variability….
---- p. 50 – 2nd to last line – operation**,** (add comma) causing **an** increase in emissions
--- p. 58 – 1st paragraph – line 3 through **the** presence, line 4 (RLD), (add comma), line 5 after **an** iterative…

--- p. 58 – 1st paragraph – line 3 through **the** presence, line 4 (RLD), (add comma), line 5 after **an** iterative…
--- p. 59 – 2nd paragraph under "summary and conclusions" - …US06 cycles**. This data was used** to examine the existence…

Peer Review Charge Questions

1. **Was the imputation method used for replacement of measurements with low concentration reasonable? What other alternatives may have been better in this case? (Section 7.1)**

The text talks about tailpipe emissions being greater than zero, while the actual question of relevance may be – Are the tailpipe emissions greater than the ambient background concentrations? (presuming a relatively clean background) If the vehicle's emissions is below background levels, than it would not be making a additive contribution to pollution levels. In this case, you could get a distribution of both positive and negative values fluctuating around zero.

My concern with the imput value procedure is that it might introduce a slight positive bias. My preference would probably be to use zeros as the imput values, since this would represent a distribution of positive and negative values that would average to zero over a larger dataset. For the cases where the dilute exhaust concentrations are lower than the measured background. By eliminating, or essentially not allowing, negative or zero values, doesn't this add a positive bias to the results. Or how do you ensure there is a not a positive bias.

p. 24 – first paragraph – its states that a data point can either be deleted or replaced with an imput value. Couldn't the value just simply be left as is? Also, if an input value is deleted for a particular bag, presumably no composite emissions would be available. Would the corresponding test be eliminated in its entirety?

p. 25 last line – its unclear how an outlier could be an input value. Is this a case where the input value is a very low value? A description of what these values might be should be added to the text. Again, if the imput outlier value is eliminated, how does that impact the result of the test?

Can Table 7-2 be modified to include not only the number of outliers identified, but also the number that were eliminated? This should also be in the text in the paragraph for section 7.1.2.

2. **Please comment on the use of mixed model in analyzing the data. Was the model fitting strategy in selecting the final model statistically sound? (Section 7.2)**

The use of a mixed model is a relatively standard method to treat data analysis of emissions, and appears to be appropriate for the study design given here. The fitting strategy of using a saturated model, then developing the most optimal covariance structure, and then fitting the final model appears to be statistically sound. The step-wise backward elimination approach for the "Sulfur Effects data" also appears to be statistically appropriate.

One thought relating to the discussion of the univariate ANOVA and the multivariate MANOVA in section 7.2. Overall, it does not seem to add significant value to the report. Looking at the audience for this report, it would likely be composed of a mix of readers that may or may not have a statistical background. For those with a statistical background,

the additional information on the univariate ANOVA and the multivariate MANOVA would not provide significant value. For those that are not as familiar with statistical methods, the added information would likely be more confusing than clarifying. It is suggested that this information, although interesting, might be better placed in an Appendix.

3. Were the model assumptions for the covariance structure reasonable given the data? (Section 7.3)

For the "Clean-out data", an unstructured covariance structure was initially used (which did not converge), and then a compound symmetry structure was selected. The compound symmetry structure is appropriate since the measurements from the same vehicle should have a homogeneous variance, as included in the text. For the "Effect of Sulfur data, where multiple measurements are made at different mileage accumulations, the autoregression covariance structure is appropriate since the since the correlations between measurements is expected to decline as the measurements are further apart in terms of mileage accumulation, as mentioned in the text of the report.

4. Please comment on the methodology used in determining the in-use sulfur effect for models with and without the sulfur and mileage interaction term. (7.3.3)

Overall, the methodology for determining the in-use sulfur effect for models with and without the sulfur and mileage interaction term appears to be sound. For cases where the sulfur level and mileage interaction term was not significant, the sulfur effect, sulfur loading and associated percentage differences between the high and low sulfur fuels should be constant as a function of increasing mileage. Thus, using the differences in the least squares means from the final model and reverse transforming them could be used to quantify the percentage differences between high and low sulfur for the case where the sulfur level and mileage interaction term was not significant.

For cases where the sulfur level and mileage interaction term is significant, the rate of sulfur loading between the low and high sulfur fuels would differ as a function of mileage, so a different approach would be needed. The methodology of using the in-use emission level upon arrival (pre-cleanout) from a larger clean-out dataset and projecting it out to the mileage where the two lines intersect (in-use equivalent loading) seems reasonable for estimating the in-use sulfur level effect. Some clarification should be given on what is meant by the "larger clean-out dataset" since it is not immediately obvious what data these actually are. Additionally, how do the incoming sulfur levels for the "larger clean-out dataset" compare with average in-use values for the US fleet, or the fleet in different regions of the country.

5. Is the interpretation of the mileage-by-sulfur interaction term presented correctly? (Section 7.3.3)

As discussed under question #4, the interpretation of the mileage-by-sulfur interaction term does appear to be presented correctly. On page 48, when the interpretation of the interaction term is first being discussed, I would also add in something about the percentage differences in emissions being constant as a function of mileage, just to make things clearer. For example, for the last line of the paragraph just after Table 7-9:

In other words, the effect of high fuel sulfur on Bag 2 NO_x exists immediately after clean-out and remains constant **on a percentage basis** during subsequent driving of a vehicle.

6. Are the sensitivity analyses on effects of low-level concentrations, imputed values, and influential vehicles sufficient as performed? Do the results from the sensitivity analyses provide additional support for the robustness of the conclusion? (Section 7.3.4)

Overall, the sensitivity analyses see to add additional support to the robustness of the conclusions. Here are a few thoughts on the sensitivity analyses.

If the sensitivity analyses were only done for the NO_x bag 2 emissions, it would be worth mentioning that the NO_x bag 2 emissions showed a higher percentage of measurements with zero values than most other pollutant/bag combinations, as illustrated in Table 7-1.

It would be interesting to get some feel for the vehicles/data eliminated in the sensitivity analysis for the low concentration measurements. Since these would be low emitting vehicles, it seems like they might have more robust and sulfur tolerant catalyst systems. This is consistent with the results showing an increase in the emissions reductions as these data are pulled out.

Relating to the imput values, it would be interesting to see what the effect would be of making the imput values simply zero. Given the small differences between the model with and without the imput model, there would likely be only a minor impact, but this would add to the robustness of the results, although probably not change any conclusions.

For the influential vehicles, it is interesting to look at the vehicles selected. Vehicles 0007 and 0178 show a relatively large sulfur effect, whereas vehicle 0046 shows a slight reverse sulfur effect with tight variability. It is interesting that other vehicles with relatively strong sulfur effects, such as 0165, 0179, and 0011, did not have a strong influence. The result of this sensitivity analysis is reasonably intuitive, with the NOx differences between sulfur levels shrinking, which would be consistent with removing some vehicles with a stronger sulfur effect.