UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
NATIONAL VEHICLE AND FUEL EMISSIONS LABORATORY
2565 PLYMOUTH ROAD
ANN ARBOR, MI 48105-2498

OFFICE OF
AIR AND RADIATION

December 20, 2013

MEMORANDUM

SUBJECT: EPA Response to Comments on the 2nd peer review of *Estimated Summer Hot-Soak Distributions for Denver's Ken Caryl I/M Station Fleet*

FROM: Constance Hart, Assessment and Standards Division
Office of Transportation and Air Quality, U.S. Environmental Protection Agency

Constance Hart

THRU: Angela Cullen, Director, Data and Testing Center, Assessment and Standards Division
Office of Transportation and Air Quality, U.S. Environmental Protection Agency

Angela Cullen

TO: William Charmley, Director, Assessment and Standards Division
Office of Transportation and Air Quality, U.S. Environmental Protection Agency

EPA contracted with ICF International to conduct and coordinate the peer review for the subject report. The three peer reviewers selected by ICF were Dr. Leela Rao (California Air Resources Board), Dr. Stephen Stewart (Operations Manager of British Columbia AirCare Program), and Dr. Michael Tschantz (MeadWestvaco Corporation). EPA would like to extend its appreciation to all three reviewers for their efforts in evaluating this report. The three reviewers brought useful and distinctive views in response to the charge questions. Following completion of the review, ICF submitted a report, 'Peer Revew of Draft Report "Estimated Summer hot Soak Distributions for Denver's Ken Caryl IM Station Fleet"', which summarizes the process and includes all comments.

In cooperation with the contractor who performed the study, Eastern Research Group (ERG), the report has been updated in response to the comments of the three peer reviewers. Responses are given in bold below for individual comments to point to where this was addressed in the revised report.

Leela Rao Review Comments:

The draft report "Estimated Summer Hot-Soak Distributions for Denver's Ken Caryl I/M Station," prepared by ERG under contract for the U.S. Environmental Protection Agency, details the results of testing that was conducted to evaluate evaporative emissions on a selection of the vehicle fleet. Vehicles were prescreened using remote sensing in order to more heavily sample vehicles with elevated evaporative emissions. Once accepted into the test program, vehicle hot soak emissions were measured using a portable SHED (PSHED). Based on the sample and response weights, the measurements from the PSHED were then used to estimate the distribution of hot-soak emissions for the entire sampled fleet. The analysis was also broken down by model-year group, with model year (MY) used as a surrogate for fuel-system and emissions-control technology.

The results showed that all 1981 MY and older vehicles are expected to have emissions corresponding to cumulative leak of 0.020 inches. In contrast, 26% of 1981-1995 MY vehicles and 3.3% of 1996-2003 MY vehicles are expected to have a leak of that size. In addition to characterizing the hot soak evaporative emissions, a physical inspection of the major fuel system components was conducted and the sources of vapor emissions identified where possible. In 66% of the cases, vapor sources were identified as the fuel tank, fill pipe, and canister.

While the results from this study demonstrate the ability to characterize hot soak emissions from the population of vehicles entering the Ken Caryl I/M station, the ability to generalize these results to larger fleet is limited due to the fact that controlling variables such as ambient temperature, fuel volatility, and barometric pressure were not standardized. Despite this, the hot soak emission characterization combined with the inspection results can provide some insight into the level and sources of evaporative emissions from the local area fleet.

In general, the report is well written and the analyses clearly described. There do not appear to be any major flaws in the methodology given the relatively narrowly defined project goals, and potential biases and sources of uncertainty have been identified. Throughout the analysis description examples are provided to illustrate the methodology, and figures and tables used well to assist the reader in understanding the approach taken. The conclusions drawn are conservative in that they are limited to the narrowly defined project goals.

RESPONSE: The scope of this report was limited by design to the primary analyses of the study results. We limited the interpretation and conclusions in the report to the estimation of elevated hot-soak frequencies for the sampled fleet. Application requires additional steps that are better discussed in the specific context. Some assumptions and adjustments are necessary, such as for temperatures and fuel volatility, as well as altitude.

The project goals were identified as estimation of the cumulative distributions of evaporative emissions in the light-duty fleet and utilization of a cost-effective and efficient method of identifying vehicles with elevated evaporative emissions with subsequent measurement of those emissions. While it is tempting to extrapolate the results from the fairly limited testing conducted to the fleet at large, the researchers recognized the limitations of the test program (e.g., only one population sampled, ambient testing conditions not controlled, test fuel conditions not controlled) and avoided this trap. From the descriptions of the methods and results it appears that the project and methodological goals were achieved.

Sampling was conducted using a "probability proportional to evaporative index" (ppEI). The index is well described and seems appropriately applied as a method to strategically sample vehicles that are likely to have elevated evaporative emissions. Given the sampling fractions and the response rate, it was then possible to estimate distribution of evaporative hot soak emissions from vehicles visiting Ken Caryl I/M station. Because sampling was limited to one location with a possibly unrepresentative population of vehicles, it is difficult to extrapolate these results to the fleet at large.

RESPONSE: The biases in the fleet are discussed in Section 4.3.1. We feel the issues are mitigated by using the prevalence rates by model year. The section concludes that if

anything the rates are conservative since Ken Caryl is an affluent area where vehicles tend to be well maintained.

Hot soak emissions were used as a surrogate measure of evaporative emissions from each vehicle. As discussed in the report, hot soak emissions are just one part of a vehicle's evaporative emissions, but nonetheless should identify those vehicles with fuel system leaks and possibly even purge issues. Measurement of hot soak emissions will also correctly identify those vehicles with low overall evaporative emissions. However, vehicles with moderate fuel system defects, such as bleed emissions that only are observable on the second day of a multi-day diurnal test, will not be identified by a single hot soak test. Despite this, the use of a hot soak as a quick screening tool for identifying vehicles with fuel system leaks seems appropriate and cost effective.

RESPONSE: Identifying purge issues is out of scope for this project. Evaluations for purge issues through intensive techniques such as multi-day diurnals would have been far more laborious and expensive, as well as burdensome to respondents, thus infeasible, therefore we agree that what we did was cost-effective

The primary concern I have is with the use of model year as a surrogate for emission control technology. While it is true that on a gross level vehicles manufactured prior to 1981 had no controls, those manufactured between 1981 and 1995 had basic evaporative emission controls, and 1996 and subsequent vehicles had enhanced evaporative emission controls in addition to OBD monitoring, the reality is not as straightforward. Each regulatory action contains a phase-in period during which only part of the vehicle fleet must comply with the new standards. As such, it may be 2-4 years after the start of a new rule until all vehicles contain the newest emission control technology. A second issue is that some 2001-2010 MY vehicles may have been partial zero emission vehicles (PZEVs), with additional emission control technology above and beyond that found on vehicles with enhanced evaporative emission controls. Thus, rather than using model year as a surrogate for emission control technology, a better approach would have been to determine the certification classification of each tested vehicle. While this may not have appreciably affected the results, it is more accurate than the approach taken, which is more of an approximation of emission control technology based on model year. If it is overly burdensome to reanalyze the results based on actual emission control technology, a brief discussion of why model year is likely a simplification, along with any possible ramifications of this simplification, should be added to the report.

RESPONSE: Characterizing vehicles by standard could have been relatively easy for vehicles manufactured since 2000. It would have been difficult and very burdensome for vehicles in model years prior to 2000, and probably impossible for vehicles manufactured prior to 1990.

We agree that the use of model-year groups as a surrogate for technology and standard levels is approximate, and is imprecise during phase-in periods when the composition of vehicles entering the market is in flux.

Nonetheless, we see it as an acceptable practical compromise, because:

Characterizing vehicles by standard would have been ideal, and probably superior, but would also have been extremely burdensome for vehicles manufactured between 1994 and 2000, and probably impossible for vehicles manufactured prior to 1994.

Nonetheless, we have added text conveying these points in Section 4.2.1.

In summary, the distributions of hot-soak emissions estimated from vehicles sampled at Ken Caryl station were combined with mechanic's inspection information to characterize evaporative emissions and determine the primary sources of those emissions. The vehicle testing was conducted with a portable SHED, with test vehicles pre-screened using remote sensing. This methodology was a cost-effective way of characterizing the evaporative emissions from this fleet.

Michael Tschantz Comments:

Does the report meet its primary goal (to estimate distributions of hot-soak emission levels for gasoline fueled light duty vehicles and light duty trucks, using a quick and inexpensive procedure to conduct a survey of an in-use fleet)?

Yes, the report met its primary goal. The report presents a sound and reasonable methodology to effectively and efficiently obtain necessary data and construct a distribution of hot soak emissions. The authors correctly point out the limitations of extrapolating the data across the entire annual vehicle fleet. However, I think the report and its utility -- towards a presumed end-goal of extrapolation to an annual leak-emissions inventory – could be improved by adding a paragraph or section that summarizes the specific work that would be necessary or recommended to construct an inventory.

RESPONSE: We agree that additional work is needed to apply the results of this project in development of inventories. Specifically, these steps involve correcting the prevalence rates for altitude using the Wade-Reddy equation, then normalizing the emissions to the MOVES default RVP of 9.0 psi at sea level. As these steps are described in detail in the "Development of Evaporative Emissions Calculations for MOVES2014" report, we did not see a need to add this material to the project report under review.

Was the sampling methodology using the probability proportional to Index (ppEI) appropriately applied for the situation, allowing for appropriate distribution of the fleet in the end product?

Yes, the methodology was appropriately applied to redistribute the sample to be representative of the fleet.

Is the description of analytic methods and procedures clear and detailed enough to allow the reader to develop an adequate understanding of the steps taken and assumptions made to develop the fractions in Table 4-11? Are examples selected for tables and figures well-chosen and designed to assist the reader in understanding the approach and methods?

Appendix A and section 2.3.1 provide the critical information to understand the methodology and calculations used to generate EI23, bin assignments, etc. While I could eventually reconstruct the methodology, Appendix A seemed disjointed and difficult to follow. Some attention should be made to editing this section to make it easier to follow. Some specific issues identified include:

- A significant amount of effort was spent describing the differences between Method A and Method B using an RSD3000 and RSD4000. Section 2.3.1 mentions that the overall result was the quantity "A minus B." But after review of Appendix A, it does not appear there is any subtraction of Method B estimates from A. In fact, it appears that Method A is not used in the analysis at all. If Method A is used, then there should be clarification on how. If Method A is not used, then the report might be improved by reducing or eliminating the discussion surrounding Method A.

RESPONSE: The reviewer is correct in observing that Method A calculations were not used in screening or sampling in the project, and that discussion of these methods did not clarify the methods or results of the study, either in the body of the report or in Appendix A. Accordingly, we have removed all references to methods A and B in the final revision of the report.

In addition, we have revised Appendix A to improve clarity and flow.

- On page A-11, the authors describe how the bins were established, but it was unclear where the equations came from until reading through following sections. Appendix A might be improved by rearranging the sections (e.g. relocating the section starting on page A-14 to a location prior to the equations on page A-11).

  RESPONSE: We agree with the reviewer's observation. In the revision of the Appendix we have placed the section describing the derivation of the equations before the material that describes its application in calculation of the final index.

- On page A-16, the second to last paragraph states "Since exhaust Method B HC concentration [is] known ...the equation cannot be rearranged to estimate the propane release rate from the measured EI23 ..." It looks like the equation being referred to is the one at the top of the page (EI23 = 78.536 + ...). Why can't this be rearranged and solved for Propane_scfC3?

  RESPONSE: At first glance, it appears as though the equation could be algebraically rearranged. However, when the equation represents a regression of y on x, it cannot simply be rearranged to predict x from y without producing biased predictions for x. One reason for this outcome is that the regression model assumes that x is "measured without error" and that the regression accounts for the mean response of y to x, while accounting for variability due to random error in y. Stated differently, the regression assumes that all error is in y. A regression of x on y would produce a different set of coefficients and mean squares, under the assumption that y was measured without error, and that all variation is in x. The author did attempt to regress propane release rate on EI23 and exhaust [HC]. However, the predictions were affected by large biases and errors (due to high variability in EI23), and were not considered useful. The final approach settled upon was the method described in Appendix A.

- On page A-17, the values 1.291 and 2.157 are what established the terms "+/- 2*0.091" (e.g 1.47-2*0.091 = 1.291). A few sentences could be added to help the reader see the connection.

  RESPONSE: The revised text for the Appendix clarifies the origins of both values. Specifically, the values 1.291 and 2.157 represent low and high bounds for the quantity across all test conditions. The value 0.091, on the other hand, represents the standard deviation of ln(ln(EI23)) across test conditions.

- The terminology for Method B HC was inconsistent. Attention should be made to making terminology for this and all terms consistent.

RESPONSE: As mentioned, references to "Method B" and "Method A" have been removed from the final revision.

- Section 3.6 discusses how 21 vehicles were chosen for repair to compare with as-received hot-soak emissions. Since this work is outside the scope of this report, maybe section 3.6 can be omitted.

  ***RESPONSE:** The reviewer's point is well taken. As section 3.6 made no contribution to the stated goals for the report, we have removed it from the final revision.*

- The 95% confidence intervals for the model year groupings of hot soak emissions are provided. I think it would be helpful to show how the limits could affect the overall distribution shown in Figure 4-7. I would recommend adding the 95% confidence intervals to Figure 4-7.

  ***RESPONSE:** One of the important results of the study is that the relationship between the screening index and the pSHED measurements differs by model-year group (as a surrogate for technology and emission standards). An implication of this result is that the fractions by model-year group are of much greater interest than those for the overall fleet. Thus, adding confidence intervals to this figure would contribute very little utility to the report.*

Are the methods and procedures employed technically appropriate and reasonable, with respect to the relevant disciplines, including physics, chemistry, engineering, mathematics, and statistics? Are you able to suggest or recommend alternate approaches? In making recommendations please distinguish between cases involving reasonable disagreement in adoption of methods as opposed to cases where you conclude that current methods involve specific technical errors.

In my view, the methods and procedures are reasonable and appropriate.

Is the use of hot-soak as a surrogate for elevated evaporative emissions a reasonable premise? Is it reasonable to conclude that there is potentially a link between fuel/evaporative control system leaks and high hot soak emissions?

I feel the relationship between high EI23 and high PSHED emissions is made obvious by the authors. One familiar with fuel systems and evaporative systems should also assume there would be a link. One thing to consider when comparing running loss emissions with engine-off hot soak emissions, however, is that the pressure of the fuel system can fluctuate between positive and negative pressures during various driving and parking situations. When a vehicle is driven and purge is being pulled through the evaporative system by the engine, it is possible that the fuel tank and evaporative system could operate under slight vacuum (it is also possible that the fuel tank is under slight positive pressure when there is no purge and sometimes when purge is being pulled). When the fuel tank or canister is slightly evacuated, air should leak <u>into</u> the system through small openings (and not leak out) and canister breakthrough would not occur; there would be no vapor emissions from the fuel tank or canister during these conditions. When the vehicle is placed hot in the PSHED with the engine off, the vehicle fuel tank likely warms and develops a very slight positive pressure (although it might be possible that the fuel tank could cool and develop a very slight vacuum). The PSHED temperature data show an average of about +10°F temperature rise over the 15 minutes of hot soak, suggesting the fuel tank temperature would increase slightly. In this case of positive tank pressure, vapor would leak out of small openings. One might argue that this tank pressure effect must be accounted for when extrapolating hot soak emissions to driving/running-loss emissions to avoid over-estimating emissions. Perhaps the consideration could take the form of reducing the estimate based on some average proportion of time that the fuel tank is operated under vacuum versus when operated under pressure. This proportion would vary by technology grouping. Purge rates (and range of engine operating conditions that purge air is being pulled) have generally increased since the addition of the 48-hour diurnal test in 1996. In addition to technology improvements, the higher average purge rates could reduce running loss emissions even if leaks exist.

RESPONSE: The descriptions of the behavior of the system during driving and while parked are helpful in understanding the observed variability in the relationship between the index and the PSHED measurements, and why the predictive power of the index declines as control technology improves. In concept, we agree that it could improve inventory modeling if it were feasible to account for changes in system behavior and purge in the study results. However, at present we lack the necessary data and sophistication to incorporate such detailed refinements in the application of the results.

The description of the RSD test states that the test vehicle comes to a stop at a location 40 ft before the RSD test beam and is then accelerated past the beam. Depending on the calibration of the vehicle, the evaporative system may not be purging during this acceleration and could be under slight positive pressure; or, the fuel tank might be under vacuum since the vehicle was first idled at a stop and manifold pressure is at maximum vacuum. This effect might explain why some vehicles with low EI23 had high PSHED emissions and vice versa. Nonetheless, the data certainly show that a high EI23 is most often predictive of high hot-soak emissions.

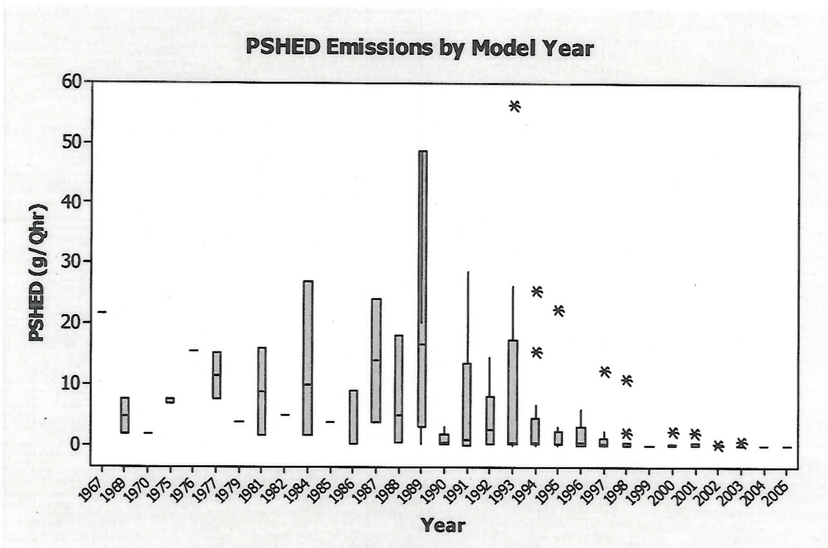
RESPONSE: See the response to the previous comment.

Is stratification of the results by model year group a reasonable approach to distinguish fuel system and emission control technology changes?

I think that grouping the vehicles is a reasonable approach, because technology advancements necessary to meet more stringent evaporative requirements and powertrain efficiency should have a positive impact on hot-soak and running loss emissions. I feel the effect of vehicle aging needs to be addressed to strengthen the argument that gross technology advancements are the primary reason for lower hot soak emissions. Since this study was a single snapshot in time, it is not possible to directly track how emissions for a single model year of vehicles (or a specific set of vehicles) will or will not increase with time. I plotted the PSHED emissions for the 175 vehicles by model year (below). It certainly appears that emissions from vehicles manufactured since 1998-1999 are much lower in emissions than those manufactured before 1994. Without using any statistical analysis, emissions for vehicles manufactured from 1994 to 1997 appear to be transition years. The transition could be due to early program implementation and phase-in of enhanced evap and ORVR enabling technologies. Since these vehicles were 12-15 years old at the time of the test, one might argue that the vehicle is approaching its design life and components are simply wearing out. The report could be improved by strengthening the argument that vehicle technology plays a larger role in reducing hot-soak emissions than vehicle age.

RESPONSE: Again, without questioning the conclusion regarding emissions of post-1998 vehicles relative to pre-1995 vehicles, we would urge that it is necessary to apply sample weighting when calculating averages of the PSHEDs. Additionally, given the very high degree of variability in these data, we would suggest that averaging by individual model years may be a higher level of disaggregation than this dataset can support.

Below, we present an analysis of age trends for two model-year groups, which may give some insight into the relative influences of technology and age.



The authors provide some insight on how one might interpret what a bin represents. In summary, bins are broadly considered groupings of vehicles with common expected severities of leaks. One might first assume that a brand new vehicle would be cast into a certain bin based upon the technology applied on that vehicle and the quality of workmanship. One might then envision two paths by which that vehicle may migrate to or towards higher bin levels as the vehicle ages: (1) step change increases from individual failure events (e.g. a connection fails, purge valve fails), or (2) a relatively slow degradation of components that causes the vehicle to slowly pass from one bin to another across a continuum (e.g. slow fatigue of a hose or joint, hose connection slowly opening up).

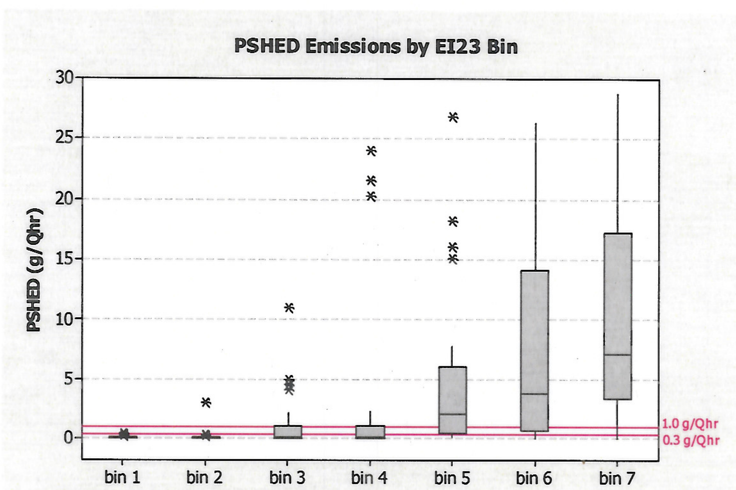
RESPONSE: We had not envisioned applying a physical interpretation to the bins, rather, we conceived them as a way to discount the high variability in values of the index for individual vehicles (inherent in remote sensing) and in the resulting uncertainty in predicting vehicles' "actual" emissions on the basis of index values.

The interpretation proposed by the reviewer could be helpful, but we would suggest applying it within model-year or tech groups, rather than between them. The reason for this conclusion is that the relation between the index and the "truth" measure changes with model year (and age). Thus, an assumption that assignments to the same bin have the same significance across model-year groups is fallacious.

If we assume that both sets of failure processes are occurring simultaneously, then it might help explain why the comparison of model year emissions is challenging. A step-change failure would likely mask any slow fatigue that could simultaneously be occurring and would not occur on every vehicle. If one could track the hot soak emissions from a large model year sample, one might expect to see the distribution in emissions widen and shift towards higher values with age. Step change failures would cause the distribution to widen on the high end of emissions; aging would cause the low end to shift towards higher emissions. In aggregate, the mean should shift higher with age. The variation in failures and severity would likely blur the distribution. We see this trend in the figure above. With small samples, one "high" emissions data point can have a significant impact on the distribution and interpretation of results. Also, variation in vehicle design and in-use histories will result in variability of in-use hot soak emissions.

RESPONSE: We agree with this description of how emission levels increase as a fleet of vehicles (of comparable technology) ages and experiences both normal wear as well as individual "catastrophic" failure events. We also find it plausible that normal wear and minor failures would contribute to shifts in the lower regions of the fleet distribution, and that major failures would contribute to marked spread on the high end; both processes, manifesting in various ways, contribute to the overall spread of the distribution and the increase in the (arithmetic) mean over time. We also agree that emissions are highly variable, even within technology groups, and that given this inherent variability, it is very difficult to discern and interpret trends based on small samples.

The seven bins assigned by EI23 are clearly shown in the report as being suitable to distinguish between high and low hot-soak polluters. Distributions for each bin are made up from an average of 175/7=25 vehicles spanning many model years and the variation can be very large, as shown below.



RESPONSE: A presentation such as this figure is useful in showing clearly the wide range in the PSHED measurements relative to the index bins but must be viewed with caution because it does not account for differences among model-year or technology groups. In viewing this chart, it is important to be aware that differing model-year groups are not distributed evenly across the seven bins. Nor does a simple view account for the differences in sampling weights among individual vehicles, which is necessary to derive representative results from these data.

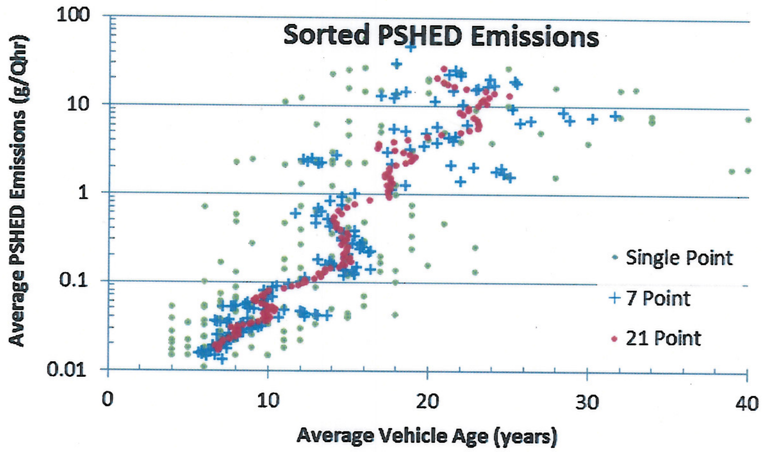
A way to possibly get around these issues is to expand the bin concept and analyze these bin levels of PSHED emissions as a function of the average vehicle age within that bin. That is, for a given level of PSHED emissions, the average age of vehicle that would produce that level of emissions can be calculated. This would, in practice, be a way to sort those vehicles that experience very high emissions (presumably from a step-change causing failure) from the lower-emissions, normally aged vehicles.

***RESPONSE:** It is not clear to us that estimating average ages within bins would effect this separation. In fact, we would also suggest that this step would not be meaningful if average ages were calculated across technology groups. While making such a separation could provide insights into fleet behavior, we do not see it as necessary, as both catastrophic events and normal wear-and-tear factor into the overall fleet mean. The goal of screening and representative sampling is to incorporate and account for both processes.*

The issue with variation can be resolved by removing the large variation by making the bins very small (e.g. one, seven, or 21 vehicles) and permitting the bins to overlap one another to compare the incremental change of the average response.

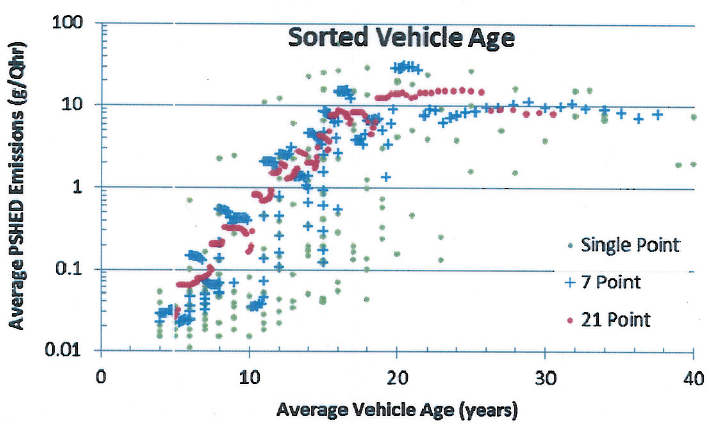
***RESPONSE:** As the high variability is an inherent characteristic of the emissions data, it cannot be "removed" by analytic techniques. Nonetheless, it may be possible to devise ways to minimize the uncertainty in averaging of variable data and so to facilitate the assessment of trends. Calculating moving averages based on overlapping bins as proposed by the reviewer may be one means of achieving this goal, although the reviewer did not assess the uncertainties in the calculated averages. This treatment of the data may provide useful insights when viewing the dataset as an aggregate during analysis, after sampling and measurement is completed. However, we would suggest that caution be taken to avoid averaging across, rather than within model-year groups. In addition, to obtain representative results, it is necessary to apply the unequal sampling weights to the individual measurements.*

I sorted the PSHED data from lowest to highest and maintained tags to model year. I then took single point, seven-point, and 21-point running averages for both PSHED emissions and vehicle age and plotted them (shown in the top figure below) against one another; this effectively made 175, 169, and 155 overlapping bins, respectively. The results suggest that a step-change in emissions occurred for vehicles with an average age of about ten years (increasing from 0.04 g/Qhr to 0.08 g/Qhr), and a second step change occurred for vehicles with an average age of about 15 years (increasing from 0.2 g/Qhr to 0.8 g/Qhr). Between these time frames, the emissions increased rather smoothly with increasing average vehicle age. These step-change increases might be corresponding to enhanced evap, OBD, and ORVR. The moderate changes might be caused by aging. A cursory glance suggests that hot soak emissions could be increasing at a nominal rate of about 1-fold per two years. It is just that new vehicles with state-of-the-art technology may start in lower bins than their predecessors.



RESPONSE: As the reviewer suggests, the locations of the "steps" could be suggestive of major changes in technology. However, as the calculation does not incorporate sample weighting, we are not confident that either the sizes of the "steps" or the trends between them can be interpreted in terms of size, so as to assess the relative effects of technology and age.

For comparison, the plot below was produced by sorting the PSHED data by vehicle age, then plotting single-point, seven-point, and 21-point running averages. It appears to be much more difficult to interpret. I recommend the authors spend some time trying to address and differentiate effects of aging and technology advancements based on the data available.

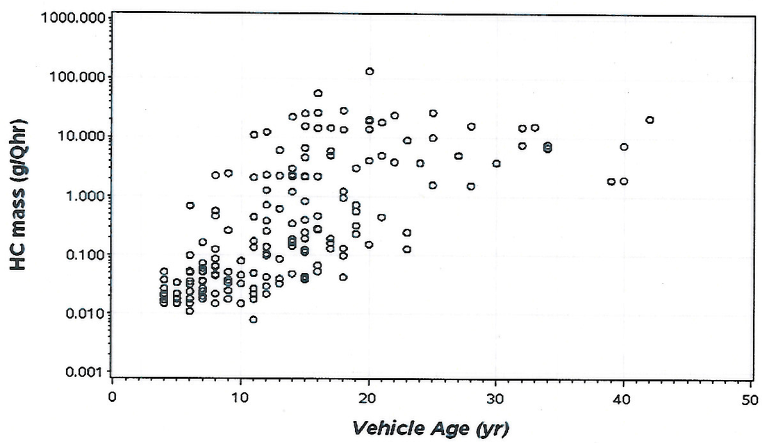


RESPONSE: We agree that the appropriate interpretation of this figure is not apparent.

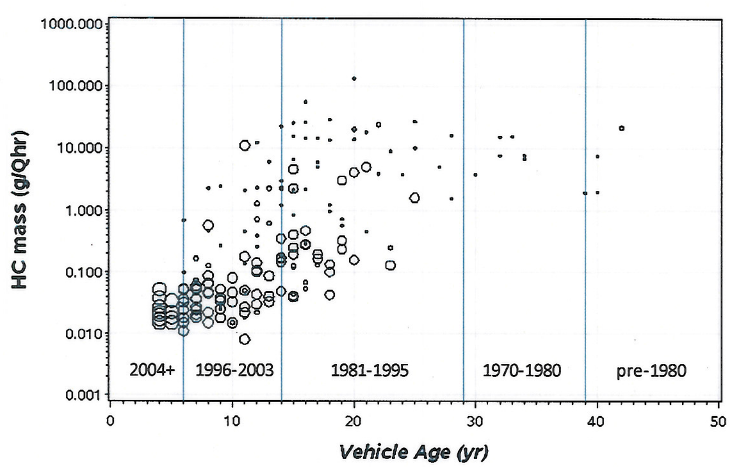
Given that two model-year groups have some temporal depth, it may be possible to examine the effect of age within the groups, relative to the effect of technology between the groups. An analysis is fairly straightforward if we accept the assumptions that technology is uniform with the groups and that age effects can be accounted for by a log-linear model. The first assumption is not entirely true can be accepted for a preliminary examination. In any case, it is necessary to apply sample

weighting in fitting the age models, for reasons that will be apparent in examining the data.

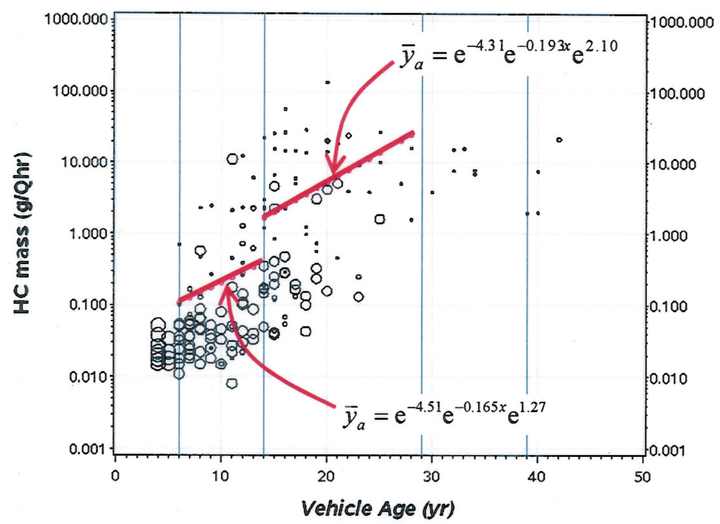
In the figure below, we present the PSHED data plotted vs. age. It appears as the single points do in the reviewer's figures above.



In the revised presentation below, however, we represent each measurement as a "bubble," in which the size of the bubble is proportional to the measurements sampling weight, i.e., the number of vehicles in the fleet represented by the measurement. This modification makes it clear that the majority of vehicles are "clean," especially in more recent model-year groups. In the figure, vertical lines denote the set of model-year groups.

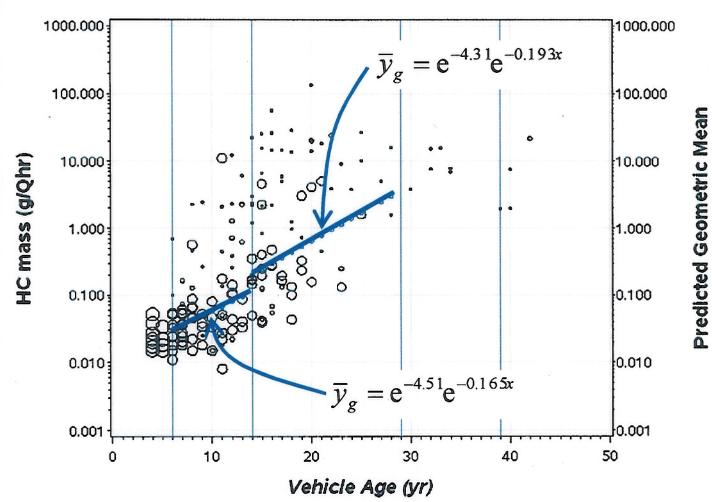


The next figure shows the same data with modeled age trends superimposed. For both model-year groups, the trends are log-linear with age (using natural log). The predicted values are arithmetic means, meaning that they reflect a value higher than that typical of vehicles in the group, due to the strongly skewed nature of the *underlying distributions. The age trends are highly significant, with p=0.002 and p=0.0067 for the 1981-95 and 1996-2003 groups, respectively. The very high degree of scatter around the trends is indicated by the respective R^2 values of 0.12 and 0.104.*



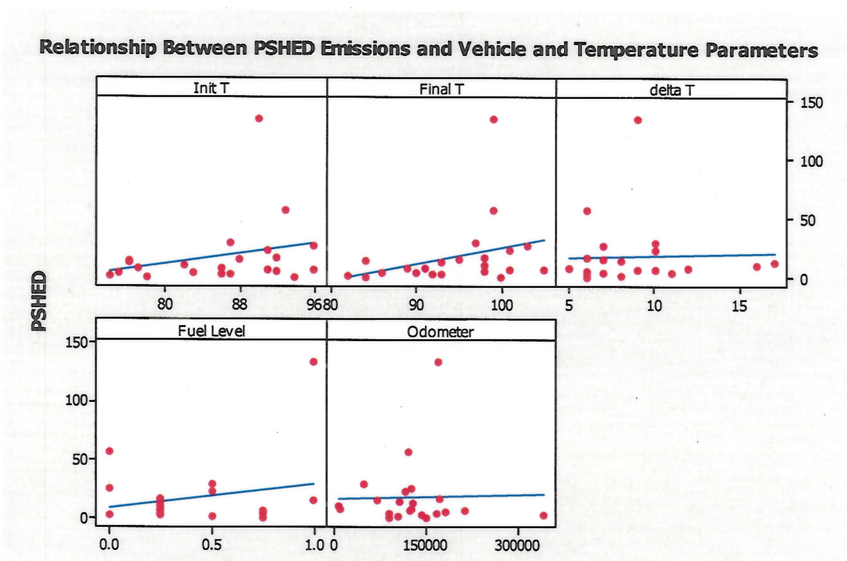
The final plot below shows predicted geometric means. As these values reflect the center of the distribution, these trends are more indicative of the expected emissions level for a typical vehicle in each model-year group. In addition, the differences between the arithmetic and geometric means reflect the degree of skewness in the distribution.

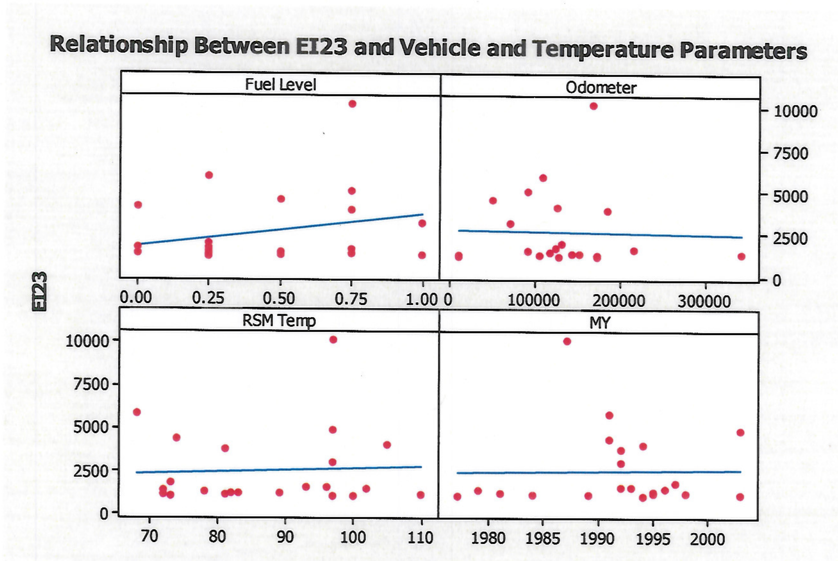
Taken overall, the results of this preliminary analysis suggest that the effects of aging, i.e., normal wear combined with "step failures" may be as large as or even larger than the effects of technology as such.



Does the methodology, data, and analyses support the report's conclusion?

I think the methods used to generate EI23 and assign to a bin are clearly appropriate and useful for an efficient and cost-effective evaluation of hot soak emissions. I think the use of the PSHED and the general tie between running loss and hot soak emissions is appropriate (although as mentioned earlier, data and analysis or an explanation addressing in-use driving tank pressure versus parked hot-soak pressure is needed). I think the authors suitably tied the sample population to a fleet population using appropriate proportioning. Therefore, I think the authors did a reasonable job developing a distribution of hot soak emissions, under the conditions evaluated in Denver during the summer of 2009. I also think that the authors clearly showed where a large number of leaks were arising, particularly from the pre-1996 vehicles.
The authors also suggest that temperature, pressure, fuel RVP, etc. can also impact the results. On page 4-15, the authors report that these factors were not controlled but recorded. If the authors can show correlations or lack of correlations, then the findings might usefully be applied outside of summer 2009 Denver. I quickly plotted PSHED emissions and EI23 versus various factors below. Again, without any formal analysis, there doesn't appear to be any strong correlation. I think the authors might consider identifying whether any correlations can be identified. The correlations (or lack thereof) would provide the reader with a better sense on how the findings can be extrapolated and what data should be collected during future evaluations.





RESPONSE: With respect to the PSHEDS, there may well be some correlations with some of the factors listed, although they are not necessarily linear correlations . Given the variability of the data, they may not be negligible.

As with the PSHED results themselves, however, it would probably be more useful to view these correlations within the context of model-year groups. In simple views such as those presented above, the tendency of older vehicles to have higher PSHEDS may make the apparent correlation appear substantially stronger than it actually would be if technology were accounted for. In addition, accounting for the unequal probability structure of the sample may affect the strength of apparent correlations.

With respect to the values of the index, a similar caution applies. More generally, however, it would not be surprising were the general variability of remote-sensing to obscure the effects of other influences such as ambient temperature.

Stephen Stewart Comments:

Primary and Secondary Goals

The primary goal was to estimate the cumulative distributions of evaporative emissions in the light-duty vehicle and light-duty truck fleets.

There were also two secondary goals. One was to apply a cost effective and efficient method of measuring evaporative emissions, and this was by measuring hot-soak emissions using a PSHED. The other was to use RSD screening and a probability index to improve efficiency of sampling vehicles with elevated evaporative emissions.

The report meets its primary goal in presenting the fleet fractions that are shown in Table 4-11 It also meets both its secondary goals, and this is demonstrated through the process and analysis necessary in order to reach to primary goal.

Possible Limitations

One limitation is that the project did not include any full-size light-duty trucks, because of the physical size of the PSHED, but it would be reasonable to assume that the results presented would be valid for LDT3 and LDT4 because their evaporative emission standards and control technology is the same as for LDV, LDT1 and LDT2.

The study did not include vehicles that had been exempted from the IM program by the clean-screening provisions. The report shows that high evaporative emissions would be unlikely to have affected the clean-screening results because of the vehicle speed during clean-screening. So clean-screening should not have biased the sampling process.

There were no participating vehicles of model year 2006 or newer, but evaporative problems with these vehicles should be even less probable than with the thirteen 2004-2005 vehicles that did participate, so even though there must be some vehicles in this age group in the fleet that have evaporative emission problems, this is unlikely to have caused any significant error in the overall estimate of the incidence of high evaporative emitters.

At the other end of the age scale, there were few participating vehicles older than model year 1981, and all had high PSHED results, leading to the conclusion that the entire fleet of 1980 and older vehicles have hot-soak emissions over 1g/Qhr. In reality there must be some vehicles in this age group that do not have evaporative emission problems. However, the order of magnitude comparison of age groups is more important here than a comparison to an absolute value of 1g/Qhr, so for these vehicles it is more important to look at their fractions above 2, 5, 10 or 20 g/Qhr.

Sampling and ppEI

This approach increases the chances of selection for vehicles that are likely to have higher evaporative emissions. It therefore reduced the total number of vehicles that needed to participate compared to completely random selection. The probability of selection is inverted during analysis to become part of the weighting used to apply test results to the overall fleet. This is an efficient approach. It is valid even though the screening index is much less useful for identifying vehicles with high evaporative emissions than for identifying those with low evaporative emissions. A second stage of sampling probability was created by the response rate of vehicle owners to selection. This too was inverted during analysis to provide a weighting.

Analytic Methods and Procedures

The testing process and methods are all logical and are well explained in the report. The technical bases for the methods are described step by step, and the appendices provide additional information.

The sampling approach and response rates are sound and the report explains how these carry through into the analysis of results as weightings to be used in applying the sample data to the overall fleet. The examples help in understanding what was done.

RSD for Evaporative Emissions

The way that RSD can be used to look for high evaporative emitters is explained in an appendix, and the main report considers how well high and low values of the EI23 index match up with actual high and low PSHED results. It does show that low values of the EI23 index are indeed likely to correspond to low PSHED results. However, high values of the EI23 index are almost equally likely to have high or low

PSHED results, but this is a much higher likelihood than would come from a completely random sample and therefore the EI23 value is an effective way to screen for sampling.

Modified California Method

MCM was used to look for the source of evaporative emissions. The results are interesting and will be relevant to later consideration of how to identify or inspect vehicles in order to find and repair high evaporative emitters. They are also the only data that links high hot-soak emissions to the overall condition of the vehicle's evaporative control system.

Hot-Soak

An appendix details a comparison of hot-soak emissions from PSHED and LSHED testing, and shows that PSHED results tend to be a little lower than LSHED results, and the analysis supports use of PSHED results as a good surrogate for LSHED. However, this is only for the hot-soak evaporative emissions. No measured data is presented that would allow examination of the validity of using hot-soak as a surrogate for elevated evaporative emissions in general, and as the report mentions, there are also running losses, permeation and diurnal emissions that need to be considered. However, these are convenient testing categorisations rather than being completely separate emission issues. The MCM results show some problems that are likely to cause both running losses and hot-soak emissions, and some that are likely to cause both diurnal and hot-soak emissions. I would speculate that in most cases permeation is a relatively minor part of high evaporative emissions, but it too contributes to hot-soak results. Overall, I believe that it is very reasonable to assume that there is a link between fuel and evaporative leaks in general and high hot-soak emissions, and therefore hot-soak is a reasonable surrogate for elevated evaporative emissions.

Stratification by Model Year

The certification standards for light-duty vehicles are almost completely related to vehicle model year, and the evaporative standards for light-duty trucks have been the same as light-duty vehicles. This has meant that the evaporative control technology and its ongoing performance as vehicles age also continues to be tied to vehicle model year. It is therefore completely appropriate to stratify the light-duty fleet by model year as a surrogate for evaporative emission control technologies.

Conclusion

The report describes a project that used a multi-stage testing process, some parts of which yielded data that was not used in this report's analysis. The stages of testing were logical and sequential. The analysis of results was also logical and sound.

The methodology, data and analyses support the report's conclusions.

RESPONSE: Dr. Stewart's comments are overall very positive and do not require updates. They are included here for completeness.