Stephen L. Gerritson
Environmental Consulting
11016 NE 164th Place
Bothell, WA 98011
Phone 425-486-9784
Fax 425-483-4525
email sgerritson@earthlink.net

December 29, 1998

Ms. Penny Carey
U.S. EPA
Assessment and Modeling Division/SAG
2000 Traverwood Drive
Ann Arbor, MI 48105

Dear Ms. Carey,

In a letter from Phil Lorang, dated November 3, 1998, I was asked to review and offer comments on several documents pertaining to the revision of the U.S. EPA's MOBILE model. Specifically, I was asked to review the following four reports:

+ M6.STE.002, "Determination of Hot Running Emissions from FTP Bag Emissions"
+ M6.STE.003, "Determination of Start Emissions as a Function of Mileage and Soak Time for 1981-1993 Model Year Light-Duty Vehicles"
+ M6.EXH.001, "Determination of Running Emissions as a Function of Mileage for 1981-1993 Model Year Light-Duty Cars and Trucks"
+ M6.EXH.002, "Analysis of Emissions Deterioration Using Ohio and Wisconsin IM240 Data"

My experience and expertise in model development and testing include more than eight years as Executive Director of the Lake Michigan Air Directors Consortium, during which time I oversaw the development and validation of the Urban Airshed Model, Version 5, and the emissions modeling system known as EMS-95. I also served as a member of the OMS Technical Advisory Subcommittee of the Clean Air Act Advisory Committee.

I hereby certify that I have no conflicts of interest, real or perceived, in this effort or in the revision of the MOBILE model. I am an independent consultant, doing business as an individual with no other corporate name or identity, from the address above. My comments are provided in the enclosed documents, and have been filed on the enclosed diskette in Word Pefect 6.0. Please call me if you have any questions. Thank you for the opportunity to perform this review.

Sincerely,

Stephen Gerritson

M6.STE.002
"The Determination of Hot Running Emissions from FTP Bag Emissions"

I. Introduction

This paper is an attempt to develop a simple model for estimating hot running emissions (HR505) from FTP data. Because the HR505 portion is identical in all respects to the first and third portions of the FTP, but does not contain an engine start, the results from the HR505 portion could be used to extrapolate the emissions associated with cold and hot starts.

II. Report Clarity

Although the objective of the paper is fairly straightforward, the argument would have benefitted from a discussion of the options available and the rationale for selecting the sample and methodology that was used. The impression given, particularly with the statistical analysis, is one of trial and error until something worked. The writing is clear and easily understood, with no awkward constructions. Finally, the paper should contain a clear statement of the conclusions.

III. Overall Methodology

As noted above, the goal of this paper is to develop a means of separating analytically the running emissions from the start emissions by comparing datasets containing running emissions only to FTP datasets containing both running and start emissions. The methodology used was to compare test data from 77 vehicles, all of which have undergone the three segments of the FTP, plus an additional segment (HR505). The data were then subjected to several regression analyses to determine the level of correlation.

In my opinion, the overall study design and methodology used is appropriate. While I have some comments on specific aspects of the work, this approach is straightforward, takes advantage of existing tests, and for the most part excludes what could be confounding variables.

IV. Appropriateness of the Datasets Selected

The data used came from testing performed at two facilities: the Automotive Testing Labs, Inc, in Ohio, and the U.S. EPA lab in Ann Arbor. A total of 77 vehicles participated, although data from one of these were subsequently excluded. I have several comments.

Sample Size. A question concerning the adequacy of the sample size might be asked, and some explanation of why 77 (as opposed to 78 or 80) would be helpful. In hindsight, since model year had no significant effect on the results, the sample size may be adequate.

Sample Selection. There are several issues with respect to sample selection. First, the sample does not appear to be random, but self-selecting (e.g. volunteers) from a larger pool of potential recruits. This may introduce sample bias. Second, it is unclear whether this sample is

representative of the fleet at large, or even whether the two subsamples (Ohio and Ann Arbor) are comparable. The Ohio sample was drawn from volunteers who had gone through the I/M lanes in Ohio. The Ann Arbor vehicles, presumably, were not subject to I/M. Some discussion of this would be helpful. Third, it was acknowledged that many of the Ohio vehicles were I/M failures, and therefore not a random sample.

Datasets Generated. The datasets generated for this analysis consisted of measurements of three pollutants for each portion of the FTP and the same three pollutants for the HR505. Presumably, variables such as fuel, operator differences, and so on were controlled or minimized. Other data might have included odometer readings. In general, I have no comments on this aspect of the study. However, I believe that the dataset in Table 1 is incomplete.

V. Data Analyses Conducted

The use of regression analysis to assess the correlation between the FTP and HR505 results is appropriate. Some explanation of the specific aspects of the T statistics would have been helpful (they are merely mentioned in the text and given as final results in the tables). It is difficult to determine exactly what is being tested for significance.

More discussion of the standard deviations would have been helpful. In looking at the summary statistics, it appears that HC and NOx are much more closely grouped than CO. Were these results expected? Can the anomalies in the CO measurements be explained? Do they affect the validity of the formula?

Finally, one might question the decision to drop the data from Vehicle #16. In the summary tables, it is difficult to see that this decision had any significant impact on the results, and, since there were other vehicles in the study with unrepaired problems, it might have been better to leave it in.

VI. Appropriateness of the Conclusions

The major conclusion of the paper is that the HR505 emissions can be calculated using emissions from the three portions of the FTP and the derived formula. Assuming that the questions raised above concerning sample selection and methodology are answered satisfactorily, this conclusion is appropriate. Intuitively, one would expect the relationship that is in fact seen, and the correlations (r-squared) are quite high.

VII. Recommendations

To lay any questions concerning sample size and methodology to rest, the U.S. EPA might consider testing another sample and comparing predicted values using the formula to actual values obtained from measurements. In the alternative, you might wish to recalculate the formula using only the Ohio data, then test it on the Ann Arbor vehicles. In this manner, no other testing would be required, although the sample size used to derive the formula would be even smaller.

M6.STE.003
"Determination of Start Emissions as a Function of Mileage and Soak Time for 1981-1993 Light Duty Vehicles"

I. Introduction

This paper is an attempt to develop a means of predicting or estimating start emissions as a function of mileage and soak time. It uses the results of the paper entitled "The Determination of Hot Running Emissions from FTP Bag Emissions." Based upon this work, on three FTP data sources, and on several other factors, formulas for calculating start emissions as a function of deterioration and soak time are derived.

II. Report Clarity

The paper is well organized, thorough in its discussions and complete. It is well written and contains no awkward constructions or unclear statements. (The Introduction does contain an error in the description of subsequent sections, however, but this is a minor point.) The same cannot be said for Appendix A, which is awkward and difficult to follow, yet which is important to the derivation in the main body of the report.

Although I recognize that it was impossible to include all tables and information used in preparing this report, I was somewhat troubled by the sudden introduction of the CARB report and the use of its conclusions without a critical review. Finally, I appreciated the example as a means of emphasizing the results, although the paper could benefit from a clear listing of the conclusions.

III. Overall Methodology

The overall methodology was to apply the findings from the HR505 tests (paper cited above) to data from three sets of FTP tests (EPA lab in Ann Arbor, AAMA data from Michigan and Arizona, and API data from Arizona), adjusted for recruitment bias with data from Dayton, Ohio, I/M testing, to calculate start emissions after a twelve hour soak and a ten minute soak. These results, plus the CARB report, were then used to calculate start deterioration with mileage and start emissions as a function of soak time.

IV. Appropriateness of the Datasets Selected

Since no description of the datasets beyond their origin and responsible agency is provided, it is difficult to comment on their appropriateness. While the overall total (4416 cars and 1205 trucks) seems sufficient, I would question the use of the small API sample from Arizona solely on the basis of odometer readings. The distribution of vehicles between Michigan and Arizona could be significant. Finally, the use of data from Ohio to correct for recruitment bias may not be appropriate.

Sample Size. Although no numerical breakdown is offered, the statement is made in the

text and the Appendix that the API and EPA samples are relatively small in size.

Sample Selection. Although no explanation of how the samples were chosen is provided, there is indirect evidence that volunteers were recruited, and that this created a bias which had to be corrected for.

Datasets Generated. Datasets generated for this paper consist of calculations of start emissions as a function of mileage for normal and high emitters, and start emissions versus soak times for ten minute and twelve hour soaks. These results were then used to adjust the data in the CARB report.

Assuming the CARB results to be valid, I have reservations about using two data points to adjust an entire continuum.

V. Data Analyses Conducted

"In general, the regression correlation coefficients (r-squared) are not high (<0.10), and reflect the tremendous scatter in emissions." While the confidence intervals for normal emitters are narrow, the intervals for high emitters are acknowledged to be quite high due to large scatter and small sample sizes.

These statements are from the paper itself, and call into question the validity of the results, at least for high emitters. Correlation coefficients of less than 0.10, even if significant, indicate that there are other factors, which have not been tested for, which could account for the variances.

VI. Appropriateness of the Conclusions

The two major conclusions of the paper are the formula for calculating basic start emission rates for normal and high emitters, and the formula for calculating start emissions as a function of soak time. Although it goes beyond the scope of this review, I question the decision to divide the fleet into two categories (normal and high emitters). While normal emitters are narrowly defined, high emitters are defined as all vehicles the emissions from which are more than twice the new car standard. This open-ended category equates vehicles the emissions from which may be several orders of magnitude apart. This scatter, acknowledged in the report, has probably distorted the conclusions for high emitters.

Again, I question the use of two data points (ten minute and twelve hour soak results) to adjust a continuum of data from the CARB study. At the least, some explanation of why this adjustment is valid should be provided.

VII. Recommendations

Several recommendations can be made:

1) conduct further tests on high emitters, and develop a more robust statistic for this category;

2) conduct additional tests using soak times between ten minutes and twelve hours, to determine the validity of the adjustment factors; and

3) redo the calculations using only the AAMA data, then apply the results to the samples from the EPA and API.

M6.EXH.001
"Determination of Running Emissions as a Function of Mileage for 1981-1993 Model Year Light-Duty Cars and Trucks"

I. Introduction

This paper describes an effort to determine running emissions as a function of mileage or deterioration in light-duty vehicles. It makes use of the work described in another paper, entitled "The Determination of Hot Running Emissions from FTP Bag Emissions."

II. Report Clarity

The paper is well organized, thorough in its discussions and complete. It is well written and contains no awkward constructions or unclear statements. The examples at the end of the report were appreciated, as was the section entitled "Results." The charts and tables are clear and complete.

III. Overall Methodology

Basic running emissions were calculated using the FTP test results and the adjustment factors derived in the other paper cited above. Deterioration curves were then calculated using a least-squares regression for each model year and technology subgroup. A correction factor for high emitters, based on an assumed recruitment bias, was also developed.

IV. Appropriateness of the Datasets Selected

Three FTP datasets were used: from the EPA Lab in Ann Arbor, from the AAMA tests in Michigan and Arizona, and from API tests in Arizona. The API sample size was small, but was used because of the high mileage readings on the vehicles. These data were adjusted by applying high emitter factors derived from IM240 tests conducted in Ohio. The data were grouped according to model year and fuel delivery (carburator/fuel injection) systems.

Sample Size. Although the overall total of vehicles involved in the study was more than 5600, the EPA and API samples were relatively small.

Sample Selection. The samples were selected on the basis of volunteer recruitment. It was felt that this method created a bias which had to be corrected for.

Datasets Generated. The datasets generated are represented by the graphs and charts, and provide estimates of emissions at a range of mileage for each subgroup.

Overall, the datasets selected were appropriate, although I have some question about the use of the relatively small datasets when the AAMA data alone would have been sufficient. My questions about the validity of the correction factors based on the Ohio data have been raised in the discussion of the paper entitled "The Determination of Hot Running Emissions from FTP

Bag Emissions."

V. Data Analyses Conducted

For each group, a regression line was estimated for emissions versus mileage. These lines took one of three forms; most situations were covered by the first (positive slope, intercept less than the low mileage mean emissions). High emitter correction factors were developed based on data collected in the IM240 program in Dayton, Ohio, and were themselves adjusted. The final adjusted values were then applied to the FTP results to obtain corrected values.

The correction factors based on the Ohio data continue to raise questions, and their use may not be appropriate. While there may be sample bias, there are better approaches to deal with it.

VI. Appropriateness of the Conclusions

The conclusions for the "normal" emitters seem to be appropriate. Those for the high emitters are doubtful, for the reasons discussed above.

VII. Recommendations

The problem of recruitment bias and the need to correct for high emitters should be dealt with in another study, designed to eliminate this problem. While I understand the expense and time involved, the additional effort is probably necessary to obtain a more accurate picture of the emissions deterioration of this subgroup.

M6.EXH.002
"Analysis of Emissions Deterioration Using Ohio and Wisconsin IM240 Data"

I. Introduction.

This paper attempts to document a means of using IM240 data, which is widely available, to estimate emissions deterioration, by developing an algorithm for adjustment based on a comparison to FTP data.

II. Report Clarity

Although the report itself is well written and fairly easy to follow, it contains neither a clear exposition of the aims or objectives of the work, nor a presentation of the conclusions (in spite of the section entitled "Results and Conclusions"). The discussion of the difficulties encountered in working with the data, and of the alternative approaches to data analysis, was welcome.

III. Overall Methodology

The overall approach was to use IM240 data, compared to FTP results, to develop an adjustment factor so that IM240 data can be used to estimate emissions deterioration. After acknowledging the many difficulties and uncertainties associated with IM240 data, several attempts were made to "correct" the IM240 values before comparison to the FTP data.

The overall objective - to find a way to use the wealth of information generated by IM240 tests in estimating emissions deterioration - appears to have become less viable as the study progressed. Much effort seems to have been spent on developing intermediate conversion factors to deal with problems uncovered during the course of the study. Each of these efforts led to another level of uncertainty in the final results.

IV. Appropriateness of the Datasets Selected

Data from the Colorado, Arizona, and Ohio IM240 programs were used. In Colorado, most vehicles are tested using a "fast-pass" procedure; therefore, only a random sample of those vehicles completing the full test were used. How the vehicles required to complete the full test were selected is not explained. This could be a source of bias. In Arizona, a random 2% of all vehicles are subjected to the full IM240 procedure; these data were used. Problems with odometer readings and emissions outliers were identified.

Data collected in the Dayton, Ohio, IM240 program were used, although no random full tests were conducted. As a result, data from Wisconsin's IM240 program, which records second-by-second results for full 240-second tests, were used to "convert" the Ohio varied-length test results to 240-second results.

The selection of these datasets and their inherent problems has introduced an unacceptable level of uncertainty into these analyses. Differences in operation and in climate, variations in the length of the tests, problems with odometer readings, and selection bias, all leading to the development of "corrective" algorithms, have added too many confounding variables.

V. Data Analyses Conducted

Analyses of the Colorado and Arizona datasets were conducted using a variety of regression models. "In general, the models do not fit the data especially well due to the lack of good baseline values for individual vehicles."

Conversion of fast-pass to full 240-second tests were also accomplished using regression analysis. This effort produced better results (correlations of about 0.80), although an unexplained variance of 20% remains. This is still a significant level of uncertainty.

Conversion of the adjusted IM240 results to FTP (LA4) results was accomplished in two steps. First, regression analyses were developed using the results of 78 tests described in another paper. Second, the results were applied to a sample of FTP and IM240 paired tests from Hammond, Indiana, and Phoenix.

VI. Appropriateness of the Conclusions

The conclusion that "currently available IM240 test data suffer from shortcomings that need to be addressed before the data can be used directly in modeling emissions" is appropriate. Given the uncertainties introduced in this study through efforts to "adjust" these data, I would not recommend that the results be used at this time.

VII. Recommendations

The availability of a great deal of IM240 data makes the use of these data for emissions analysis quite attractive. Knowing the difficulties and uncertainties inherent in their use should make the prospective design of a study easier.