MEMORANDUM

Date: August 23, 2001

Subject: Investigation of the relative impacts of mono and polyaromatics on emissions and potential methods for their inclusion in a diesel emissions model

From: David J. Korotney
Assessment and Standards Division
U.S. Environmental Protection Agency

To: Richard Rykowski
Assessment and Standards Division
U.S. Environmental Protection Agency

In our recent Staff Discussion Document entitled, "Strategies and Issues in Correlating Diesel Fuel Properties with Emissions," we discussed the issue of the relative impact of mono and polyaromatic hydrocarbons on emissions of regulated pollutants (Section III.C.1). In that document we pointed out that the inclusion of monoaromatics and polyaromatics as terms in our models would have cut the available data by approximately 50%, since the curve-fitting techniques make use of only those observations that include all of the fuel properties being considered. In order to have the most robust model possible, we wanted to make use of as much data as possible, and so included only total aromatics. Still, a number of studies offered evidence that mono and polyaromatics have different impacts on emissions, particularly for particulate matter (PM). I have therefore investigated ways to include monoaromatics and polyaromatics terms in our models. However, I have not investigated every possible approach, and there remain several open issues that we should address in cooperation with our stakeholders. Please consider this memorandum a starting point for discussion.

The investigation into whether and how to include mono and polyaromatics in our models can proceed most effectively if we answer the following questions in sequence:

1. Do mono and polyaromatics exhibit different effects on emissions?
2. If so, can we include monoaromatics and polyaromatics terms in the model without reducing its predictive power?

For this investigation, I focused on the PM model. Potential approaches to including mono and polyaromatics in the PM model can be extended conceptually to the NOx and HC models. I focused on fixed (least-squares) models despite the fact that our final "Unified" models were based on mixed (maximum likelihood) models. Fixed models are easily automated and widely understood. I believe that the trends exhibited by the fixed models are indicative of what we would expect to see if we repeated the process with mixed models. To remain consistent with our Unified Model approach, I used the version of our database in which all repeat measurements

were averaged. I standardized all fuel terms and included engine dummy variables in the regressions. Also, I included in the models only those non-aromatic terms that appeared in our final Unified Model for PM for the purposes of this analysis.

The first step was to determine if the inclusion of mono and polyaromatic terms in the PM model would result in a model in which these two terms exhibit different effects from one another. To that end, I used a fixed model approach to produce a PM model containing mono and polyaromatics terms in lieu of total aromatics. This regression utilized only 162 out of the 294 available observations, since only these 162 observations included measurements for mono and polyaromatics. For comparison, I also produced a model with total aromatics, which was based on the full set of 294 observations. Both models are shown in Table 1 with unstandardized terms. I have transformed the equations into ones providing % change in PM emissions to be consistent with the intended application of our model, as described in Section III.D of the Staff Discussion Document. The baseline monoaromatics level is assumed to be 25.8 vol%, while the baseline polyaromatics level is assumed to be 8.6 vol%.

Table 1
Regression coefficients for PM model with mono and polyaromatic terms
% change in PM = C × exp(a × natural cetane + b × cetane difference + ⋯) - 100

	PM model with mono and poly aromatic terms	PM model with total aromatics
Data points used	162	294
Transformation constant C	12.392	23.788
Natural cetane	-0.00054(ns)	-0.00285(ns)
Cetane difference	-0.04163(ns)	-0.03383(ns)
Nat cetane × cetane difference	0.00090	0.00069
Sulfur, ppm	0.00007	0.00007
Specific gravity	2.30760	1.61515
Oxygen, wt%	-0.07567	-0.08284
Total aromatics, vol%		0.00489
Monoaromatics, wt%	0.00207(ns)	
Polyaromatics, wt%	0.00876	
Effective natural cetane coefficient when cetane difference = 0.8	0.00018	- 0.00230

ns = not significant at p = 0.05 level

The linear cetane terms were retained in order to maintain hierarchy with the interactive term. I did not re-regress the mono/poly model to exclude the insignificant monoaromatics term.

As can be seen from Table 1, mono and polyaromatics do in fact exhibit different impacts from one another on PM emissions, based on the regression on the smaller subset of 162 observations. The standardized coefficients for these two terms also differ significantly. The

monoaromatics term is nearly significant, though its p-value does exceed 0.05. For both standardized and unstandardized terms, the ratio of the polyaromatics term coefficient to the monoaromatics term coefficient is approximately 4:1.

In addition to the replacement of total aromatics with mono and polyaromatics, the coefficients for several other terms are significantly different between the two models. For instance, the specific gravity term rose by 40% when mono and polyaromatic terms were introduced, and the slope of the natural cetane effect in the region of the baseline fuel actually changes sign when total aromatics was replaced by mono and polyaromatics. Whether this result is due to the fact that only a subset of the database was used to develop the model with mono and polyaromatic terms, or whether it is due to the fact that the non-aromatics term coefficients have been estimated in the presence of mono and polyaromatics instead of total aromatics, is unclear. Regardless, one concern with simply using the subset of the database (the 162 observations) to develop the PM model is the dramatic effect that this more limited set of data has on the model coefficients in comparison to a model based on the larger set of data.

In order to more precisely determine the cause of the change in non-aromatic term coefficients, I conducted another regression on the subset of the database (162 observations), but included only total aromatics. The results are shown in Table 2, along with a repeat of the two models from Table 1.

Table 2
Regression coefficients for PM model with total aromatics on subset of data
% change in PM = C × exp(a × natural cetane + b × cetane difference + ⋯) - 100

Data points used	162	294	162
Transformation constant C	10.512	23.788	12.392
Natural cetane	-0.00067(ns)	-0.00285(ns)	-0.00054(ns)
Cetane difference	-0.04549(ns)	-0.03383(ns)	-0.04163(ns)
Nat cetane × cetane difference	0.00098	0.00069	0.00090
Sulfur, ppm	0.00008	0.00007	0.00007
Specific gravity	2.45821	1.61515	2.30760
Oxygen, wt%	-0.07505	-0.08284	-0.07567
Total aromatics, vol%	0.00488	0.00489	
Monoaromatics, wt%			0.00207(ns)
Polyaromatics, wt%			0.00876
Effective natural cetane coefficient when cetane difference = 0.8	0.00011	- 0.00230	0.00018

ns = not significant at p = 0.05 level

As Table 2 shows, the two models with identical terms nevertheless have dramatically different coefficients for some non-aromatics terms. As before, specific gravity becomes more important when only the subset of the database is used, and the slope of natural cetane in the region of the baseline fuel once again changes sign when the subset of the database is used. These results

suggest that the primary factor affecting the specific gravity and natural cetane coefficients is not whether total aromatics or mono/poly aromatics are modeled, but rather the fact that less data is being used to develop the model. In other words, the use of the subset of the database (the 162 observations) may not produce as robust an estimate for non-aromatics terms as the larger set of data (the 294 observations).

Despite this difficulty, it would still be valuable to have a PM model that contains mono and polyaromatics terms, both for the increased precision that it offers for emission impact estimates and for the cost-effectiveness of aromatics control. Therefore, I have investigated potential approaches for including mono and polyaromatics terms in the PM model that avoid the problems introduced by using only that portion of the database which includes measurements of monoaromatics and polyaromatics.

Ideally, the coefficients for the non-aromatic terms will be based on the widest possible volume of data, i.e. the 294 observations, while the coefficients for the mono and polyaromatics terms will be based on the subset of the database (the 162 observations). I have identified two potential approaches that would accomplish this. These are described below.

Approach #1

I combined the coefficients for non-aromatics terms which are based on the full dataset (294 observations) with the coefficients for mono and polyaromatics from the subset (162 observations). Doing so produced the following model:

$$\begin{aligned}\%\text{ change in PM emissions} = 24.746 \times \exp(\ & -0.00285 \times \text{natural cetane} \\ & -0.03383 \times \text{cetane difference} \\ & +0.00069 \times \text{natural cetane} \times \text{cetane difference} \\ & +0.00007 \times \text{sulfur, ppm} \\ & +1.61515 \times \text{specific gravity} \\ & -0.08284 \times \text{oxygen, wt\%} \\ & +0.00207 \times \text{monoaromatics, vol\%} \\ & +0.00876 \times \text{polyaromatics, vol\%} \ ) - 100\end{aligned}$$

This approach has the obvious advantage of simplicity. However, it has the drawback that coefficients for the mono and polyaromatics terms actually assume the presence of a different set of non-aromatic term coefficients than those in the above equation. This model cannot said to provide the best fit for either the data in the subset (162 observations) or the full dataset (294 observations). It is unclear how this might impact the predictive capabilities of the model, though a correlation coefficient could be calculated to provide some indication of comparative fit.

Approach #2

This approach is similar to the first in that it provides a means for combining the coefficients for non-aromatics terms, which are based on the 294 observations, with the coefficients for mono

and polyaromatics, which are based on the 162 observations. However, in this approach, the mono and polyaromatics coefficients are estimated in such a way that they are more consistent with the presence of the non-aromatics term coefficients.

Beginning with the model which was based on the full dataset (the 294 observations), I calculated the pseudo-residuals as the difference between observed PM values in the subset of database (the 162 observations) and the predicted PM values for those same observations without including the total aromatics term in the predictions. This is similar to calculating traditional residuals, except that the total aromatics term is not included in the regression equation used to predict PM emissions. Mathematically, this can be described as follows:

$$PR_i = \ln(PM)_{o,i} - \ln(PM)_{p\ sans\ aro,i}$$

Where:

PR_i = The pseudo-residual for observation i in the database

$\ln(PM)_{o,i}$ = The observed value of $\ln(PM)$[1] for observation i in the database

$\ln(PM)_{p\ sans\ aro,i}$ = The predicted value of $\ln(PM)$[1] for observation i in the database, using the regression equation that was based on the 294 observations, but excluding the total aromatics term. This equation is:

$\ln(PM)$ = + engine-specific intercept
- 0.00285 × natural cetane
- 0.03383 × cetane difference
+ 0.00069 × natural cetane × cetane difference
+ 0.00007 × sulfur, ppm
+ 1.61515 × specific gravity
- 0.08284 × oxygen, wt%

The pseudo-residuals PR_i will be a function only of aromatics, since the total aromatics term was excluded from the calculation and all other fuel properties and engine offsets have been accounted for. As a result, the pseudo-residuals can be used as dependent variables in a new regression in which only mono and polyaromatics are included as terms in the model. The results are shown in Table 3.

Table 3
Regression coefficients for pseudo-residuals model
Pseudo-residual [delta ln(g/bhp-hr)] = exp(a × Monoaromatics + b × Polyaromatics) - 100

Monoaromatics, wt%	0.00250
Polyaromatics, wt%	0.00724

ns = not significant at p = 0.05 level

[1] Since it is the natural logarithm of emissions that is used in our regression analyses, it is likewise used in traditional residuals analyses and, by extension, in our pseudo-residuals analysis.

The Table 3 coefficients are both significant at the p = 0.05 level, and are based on the 162 observations which include monoaromatics and polyaromatics measurements. The ratio of poly to mono aromatics for the values in Table 3 is approximately 3:1, indicating that this approach mutes the relative poly:mono effects exhibited by the coefficients in Table 1. However, polyaromatics remain more important than monoaromatics.

The two terms from Table 3 can be recombined with the equation which was based on the 294 observations, minus the total aromatics term, to produce the following:

$$\begin{aligned}\%\ \text{change in PM emissions} = 24.795 \times \exp(\ & -0.00285 \times \text{natural cetane} \\ & -0.03383 \times \text{cetane difference} \\ & +0.00069 \times \text{natural cetane} \times \text{cetane difference} \\ & +0.00007 \times \text{sulfur, ppm} \\ & +1.61515 \times \text{specific gravity} \\ & -0.08284 \times \text{oxygen, wt\%} \\ & +0.00250 \times \text{monoaromatics, wt\%} \\ & +0.00724 \times \text{polyaromatics, wt\%}\ ) - 100\end{aligned}$$

In effect, the total aromatics term in the original equation has been replaced by mono and polyaromatics terms. The coefficients for the mono and polyaromatics terms were estimated under the assumption that the variation in PM levels due to changes in non-aromatic fuel properties have already been accounted for according to the specific non-aromatic term coefficients in the above equation. On this basis, then, coefficients in the above equation are internally consistent with one another.

Other issues

There are a number of issues that arise when considering how to include the effects of mono and polyaromatics in correlations between fuel properties and emissions. Some of these are listed below.

- If the only available fuel property information on aromatics is for total aromatics (such as from fuel surveys or a State's diesel fuel specifications), and yet the model includes separate terms for mono and polyaromatics, we would need to provide guidance on how to evaluate fuels in this case. There is a significant amount of scatter in mono/poly ratios for in-use fuels, with the average being approximately 3. It may not be appropriate to permit total aromatics to simply be split into mono and polyaromatic components based on this static ratio.

- The fact that the subset of the data produces significantly different coefficient estimates than the full dataset may be a function of the engine technologies that are in the subset as compared to those in the full database. I did not investigate this possibility.

- The database currently gives total aromatics in terms of volume percent as measured by

an FIA test method, while mono and polyaromatics are given in units of weight percent as measured by an SFC test method. It may be appropriate to apply some type of conversion to these aromatics measurements to place them all into the same units.