California Inspection and Maintenance Review Committee

Improving Evaluation of Mobile Source Policies

• • • • • • • • • • • •

Comments to the National Research Council on Its Review of EPA's Mobile Source Emissions Factor Model

Prepared by

Joel Schwartz
Executive Officer

Joel Schwartz, Executive Officer

915 L Street, Suite 1435
Sacramento, CA 95814

Phone: (916) 322-8181
Fax: (916) 445-7124

March 4, 1999

Introduction

The California I/M Review Committee (IMRC) is pleased to provide comments on the NRC Committee's review of EPA's MOBILE model.[1] The IMRC was created to evaluate the effectiveness of California's Smog Check I/M program, and recommend program improvements to the Legislature and Governor.

Our comments will focus on two propositions: First, human behavior can have a substantial effect on mobile source emissions, but MOBILE is currently ill equipped to account for these dynamic effects. Second, the goal of the NRC's evaluation of MOBILE should focus more broadly on how to achieve better mobile source policy outcomes.

The path to this goal might be improvements to MOBILE, supplemental techniques in addition to MOBILE, or a different approach all together. As a result, the NRC Committee should modify its Project Scope toward determining (1) whether and under what circumstances MOBILE is or could be an appropriate tool for evaluating mobile source policy and (2) more generally, how can mobile source policy development and evaluation be improved.

MOBILE Modeling Issues

There are at least four major reasons why we should be concerned about the validity of MOBILE's output:[2]

Unrepresentative Input Data

The samples of vehicles used to collect vehicle emissions data may not be representative of some or all of the vehicle populations that the model purports to describe. This is due the recruitment methods, the recruitment locations, and/or the sample sizes. For example, much of EPA's emissions data are collected from vehicles arriving at an I/M inspection lane for a scheduled inspection. However, a number of studies have demonstrated that some motorists attempt to alter their vehicles in ways that allow their cars to pass the test, but without making repairs that would lower emissions on the road. In addition, some motorists never pass their test, or do not show up for testing at all. Examples of this include the following:

- In a survey of motorists and mechanics, the Arizona Auditor General found the following (Arizona Auditor General 1988): (1) 26 percent of motorists who failed their inspection said they had their car readjusted back to its original state after passing; (2) over 80 percent of mechanics said that they were frequently asked to either adjust cars to pass the test without conducting needed maintenance or to re-adjust vehicles after they passed the test.

---

[1] Project BEST-U-97-01-A, "Review of the EPA's Mobile Source Emissions Factor Model." Information can be found on the World Wide Web at http://www4.nas.edu/cp.nsf/57b01c7b1b6493c485256555005853cf/c8b09b6bce55181b852566a3000b07e6?OpenDocument

[2] For convenience, we use "MOBILE" to refer to the entire model, including the TECH module.

- On-road ambient CO measurements in Minnesota's I/M program indicated the program was reducing on-road CO emissions by about one percent (Scherrer and Kittelson 1994). However, data from the I/M test lanes indicated that CO emissions had dropped by about 50 percent between the first and second years of the program's existence.[3]
- Recent data from the Colorado, Arizona, and British Columbia I/M programs indicate that between 15 percent and 30 percent of failing vehicles never pass their inspection. However, remote sensing data show that many of these vehicles are still on the road (Stewart and Gourley 1996; Stedman et al. 1997; Stedman et al. 1998; Wenzel 1998).
- Data supplied to the IMRC by the California Department of Motor Vehicles indicate that when, in 1997, California required motorists to supply proof of insurance in order to register their cars, the number of non-renewals increased by about 570,000 vehicles over the previous year. This might indicate that some motorists elected to drive an unregistered vehicle, rather than pay for insurance.

Thus, recruitment at I/M lanes results in vehicle samples that differ from vehicles on the road, both in fleet composition, and in emissions on the day of the test. Many of MOBILE5's assumptions regarding vehicle emissions are also based on very small samples. For example, 274 vehicles were used to generate the model's IM240 identification rates. Similarly, data on repair effectiveness were based on repair results for 266 vehicles. In both cases, the data were further stratified into smaller groups based on technology and emitter category. Many of these sub-groups, particularly the high emitters, which contribute most of the emissions and I/M benefits, contained very few vehicles (Sierra Research 1994).

Studies have also shown biases in vehicle recruitment by mailed solicitation letters. For example, in a 1994 study in Sacramento, vehicles that did not show up for an IM240 test turned out to have higher emissions in on-road RSD measurements than vehicles that did show up (Klausmeier et al. 1995). Other types of biases occur in EPA data collection efforts. For example, the repair estimates in MOBILE5a are based on repairs made by highly trained mechanics working in EPA-contracted labs. Thus, both the skill level of real-world mechanics, and the economic incentives of the real-world vehicle repair market were taken out of the equation (Harrington et al. 1998).

Unsound Data Analysis Techniques

Some of the techniques EPA uses to analyze data and incorporate results into its models are unsound. For example, in its repair studies, 46 percent of the vehicles in the data set were not repaired to below the emission standards required by EPA. In developing its repair effectiveness estimates for MOBILE5, EPA adjusted the final emissions values downward for these vehicles (Sierra Research 1994). For its estimates of super-emitter emissions in MOBILE4, EPA collected data on emissions versus mileage for a sample of 17 gross emitters. But rather than draw a regression line through the data, EPA drew a line through the origin and the centroid of the data points (Glover and Brzezinski 1989). A regression of these data gives an r^2 of zero.

---

[3] Personal communication with Huel Scherrer, University of Minnesota, March 5, 1995

EPA also has not taken account of regression to the mean is developing I/M credits. When measurements are imperfectly correlated, extreme values tend, on average, to regress towards the mean on repeated measurements (Kachigan 1986; Kruger et al. 1999). Vehicle emissions vary from test to test in repeated testing, particularly for high emitting vehicles (Knepper et al 1993; Stedman et al. 1994).[4] This is an issue in I/M programs because both the data underlying MOBILE and many evaluations of I/M effectiveness involve before and after test results on only the failing vehicles—the ones on the tail of the emissions distribution. Due to regression to the mean, failing vehicles that are measured a second time under the same conditions will generally appear to have achieved emissions reductions even with no intervention of any kind.

These questionable methods are not isolated incidents, but have been applied by EPA in a wide range of mobile source analyses. For example, in the arena of I/M evaluation, EPA compared emissions of vehicles in Phoenix at a centralized test lane with emissions of vehicles in southern California solicited by a promise of free repairs, as a means of comparing the effectiveness of California's and Arizona's I/M programs (EPA 1995). EPA's analysis did not discuss any issues regarding the differences in sample recruitment, or the possibility that there could be many differences between Phoenix and southern California besides I/M effectiveness that could affect vehicle emissions.

Combining the sampling biases and the unsound data analysis techniques, we can conclude that MOBILE (1) is based on data that are not representative of vehicles on the road, and (2) includes some input numbers that are not tied in any legitimate way to measurements of the physical world.

Model Structure

There are several ways in which the structure of MOBILE makes it inherently suspect as a tool for mobile source emissions analysis. Perhaps the most fundamental of these is that MOBILE is not really a "model" in the sense that builders of scientific models typically use that term. It is not an abstract representation of the key phenomena that affect real-world vehicle emissions. Rather, many of the effects and relationships in MOBILE are based on *ad hoc* assumptions built into the model by EPA. The output of the model depends largely on these *input* assumptions, rather than on the *output* of a set of equations that represent dynamic interactions among vehicle technology, the goals and interests of motorists and mechanics, and the behavioral incentives created by the social, legal, and economic environment.

MOBILE does not explicitly "model" the effect of human behavior on vehicle emissions.[5] To the extent that MOBILE includes human behavioral factors, they are there only implicitly as "fudge factors" that are tacked on after the basic emission rates have been calculated. As shown in Equation 1, these two factors are the compliance rate and the discount (from Harrington et al. 1998).

[4] Also based on analysis of IM240 data collected by EPA by soliciting vehicles at an Indiana I/M lane in 1990 and 1991.

[5] MOBILE *does* include correction factors based on *driving* behavior. We refer here to the way motorists and mechanics respond to incentives within and external to I/M programs.

(1) $BER_{I/M} = B_o(1 - (CRED\ (1\text{-}w)\ f(c)\ d))$

Where,

$BER_{I/M}$ is the fleet-average base emission rate with I/M

B_o is the fleet-average corrected base emission rate without I/M

CRED is the I/M credit level determined by the model

w is the waiver rate

f(c) is the adjustment factor based on the user-supplied compliance rate

d is the discount for decentralized I/M programs

Equation 1 illustrates two structural problems with the model: First, the discount and the compliance rate are externally imposed, *ad hoc* factors. The discount is not determined by any explicit analytical procedure, but is assumed by EPA to be 50 percent for decentralized programs (Schwartz 1995a; Schwartz 1995b).[6] The compliance adjustment is based on an assumed effectiveness reduction for a given non-compliance rate. Harrington et al. (1998), have shown based on real-world data that the assumed effect in the model of non-compliance is too small. As we showed above, there are many ways in which motorists avoid I/M, both in centralized and decentralized programs. A single, static factor applied across the board can not encompass these complex behavioral responses.

Second, even if we find that the data and methods for generating the emission factors are valid, these "fudge factors" subvert any presumed accuracy of the model. Whether accurate or not, the emission factors can be arbitrarily altered by substantial amounts. The output of MOBILE thus depends largely on the assumptions imposed by the modelers.

There are many factors not included in the model that could affect emissions. Mechanics may improve their abilities with experience in a new program, and fraud might go down with increased enforcement. Some motorists might avoid registration if fees are increased, or search harder for a way to avoid repair if cost limits are raised. Motorists might keep their cars in better repair in a boom economy than during a recession. Greater non-registration enforcement or on-road testing might reduce non-compliance. Although these factors might have a substantial effect on vehicle emissions, MOBILE can not be used to assess their impact (Harrington et al. 1998).

MOBILE also includes at least two built-in assumptions about vehicle identification and repair that are not true. First, MOBILE assumes that if a vehicle is not identified the first time it "should" fail, then it is never identified in future tests (Harrington and McConnell 1994). This assumption ignores both the intrinsic emissions variability of vehicles, and the opportunity to fail high emitters in future inspections. EPA data of multiple IM240 tests on the same vehicles show that without careful control of test conditions, r^2 values for repeated tests on the same vehicles can be as low as 0.66.[7]

---

[6] The National Highway Systems Designation Act now prevents EPA from unilaterally imposing the discount.

[7] See footnote 4.

Emissions variability is higher for cars with higher average emissions. For example, a study of multiple FTPs of several high emitters found that emissions varied by up to a factor of eight from test to test on the same vehicle (Knepper et al. 1993). As a result, MOBILE overemphasizes a test's ability to identify high emitters the first time around. This effect would result in an underestimation of I/M effectiveness.

Emissions variability can affect I/M performance in other ways. If emissions variability is significant, a car with high average emissions might pass by taking the test multiple times. There is evidence to suggest that at least some motorists are using this strategy in I/M programs (Bishop 1998; Wenzel 1998).[8] MOBILE's use of a fixed "identification rate" for high emitters on a given test would overestimate emission reductions under these conditions. On the other hand, a test with a relatively low high-emitter identification rate might not hinder long-term program performance because the program will have multiple opportunities to identify such a vehicle (Harrington and McConnell 1994). The MOBILE model does not take account of these factors.

Second, MOBILE assumes that the emissions deterioration rate of a fleet that has gone through a cycle of I/M is the same as that of a fleet that has not. In other words, MOBILE assumes that emission reductions achieved in each cycle of I/M persist forever (Glover and Brzezinski 1997). Real-world data suggest that this is not the case. For example, 40 percent of cars that failed in Arizona's I/M program in 1995 failed again in 1997 (Wenzel 1998).

Overall, MOBILE appears to be an *ad hoc* combination of out-of-context emissions data, hardwired input assumptions, and arithmetic operations. Many real-world factors that affect emissions could change without changing the output of the model. Some factors input to the model rest on a weak analytic foundation. Given that MOBILE does not explicitly model or even implicitly include many of the key factors that affect vehicle emissions, it is unlikely that MOBILE can be useful for teasing out the effects of various mobile source policies.

Model Validation

Results of validation studies of MOBILE have been mixed, but MOBILE has often been found not to accurately predict the results of on-road measurements (see, for example, Pierson et al. 1990, and Pierson et al. 1995). But even if MOBILE sometimes accurately predicts measured emissions rates, there is reason to believe that correct predictions occur for the wrong reasons. There are at least three reasons for this.

First, MOBILE has dozens of input parameters, some of which can be changed by the user and some of which are hardwired into the model's code. These include (1) base emission rates for each combination of model year, age, vehicle class, and engine technology; (2) weighting factors for the fraction of miles driven by each combination of vehicles in each year; (3) correction factors for speed, temperature, fuel, and driving mode; (4) deterioration rates as a function of age by vehicle class, pollutant, technology and emitter group; (5) assumptions regarding repair effectiveness that determine I/M credit; and (6) waiver, compliance, and discount rates (Harrington et al. 1998). The data

[8] This is also another example of a human behavior effect on vehicle emissions that is not encompassed by the model.

collection and processing problems discussed above likely ensure the inaccuracy of at least some of these numbers. Accurate real-world values for some parameters might be difficult to determine under any circumstances.

Second, even if the data are representative of the on-road fleet when and where they are collected, it is highly unlikely that they would be representative of all sub-fleets at all times, due to the dynamic factors detailed above. Third, we discussed several structural shortcomings of the model that affect its potential validity. Human behavior is not explicitly modeled and many potential behavioral effects are not accounted for at all. The output of the model largely depends on the assumptions of the modelers. Some key assumptions in the model are untrue. The magnitudes of these effects are uncertain. However, some of them would tend to cause overestimates of emissions, while some would tend to cause underestimates. As a result, some of the errors might offset each other. Overall, MOBILE is unlikely to be capable of predicting emissions for the right reasons or predicting the effect of policy or economic changes on future emissions.

MOBILE Drives Mobile Source Policy Decisions

There are two ways in which an invalid model can result in ill-advised policy decisions. First, policies that might be effective, but that receive no credit in the model, are unlikely to be considered by local air pollution authorities. For example, repair effectiveness is built into MOBILE, with no consideration of the effects of mechanic training, enforcement, or repair costs (Harrington et al. 1998). Second, I/M can crowd out other effective measures if the emissions reductions from the alternative policy overlap with those credited to I/M. This problem is exacerbated to the extent that MOBILE assigns more credit to I/M than is warranted by real-world emission reductions.

MOBILE is used for making high-stakes regulatory judgements regarding inspection and maintenance, conformity and overall emissions inventories. Even if the data problems discussed above are remedied, it appears that the fundamental structure of MOBILE will prevent it from generating valid predictions about emissions or the effect of alternative mobile source policies. MOBILE is therefore probably not effective for the purposes to which it is employed.

EPA is currently developing MOBILE6, the latest version in the MOBILE series. If the deficiencies of MOBILE5 described above were being remedied, there would be less reason for concern over the future use of MOBILE. A recent EPA overview of changes being incorporated into MOBILE6 suggests that mainly input data and data analysis issues are being addressed. EPA is also including some structural changes to the model, but these are technical changes, such as separating out cold start and running emissions, and expansion of the number of vehicle classes (Brzezinski and Newell 1998). The static structure of the model, inadequate treatment of human behavior effects, and the validation problem do not appear to be adequately addressed.

The NRC Committee's Analysis

According to the NRC's "Project Scope" document,[9] the NRC Committee will *in general* "consider the adequacy of the model's input data, assumptions, structure, and results used to characterize mobile source emissions. To the extent possible, the committee will consider ways to improve the reliability of the MOBILE model as a tool for assisting in the development of emission control strategies to meet air quality goals."

The Project Scope then goes on to list six *specific* issues that the Committee will address. These include, (1) types of mobile sources addressed, with a focus on under-represented sources, (2) strategies and methods for future data gathering, (3) alternative data sources and analytical techniques currently used for similar purposes, (4) incorporating the latest developments in modeling, (5) feasibility of incorporating modal modeling into MOBILE7, and (6) the overall accuracy of the model in predicting emissions.

We believe that improving the model *per se* should not be considered the fundamental goal of this process. MOBILE is not an academic model used for illuminating fundamental features of the natural world. It is a tool for selecting policies and making regulatory judgements. The goal of Congress in requesting this process is, presumably, to have the NRC present judgements on how to improve both the process of mobile source regulation and the policies that result from it. In this sense, the NRC Committee should see its charge as that of evaluating the model and assessing whether other (perhaps more reliable) means of evaluating mobile source emissions reduction policies exist.

With that in mind, we believe the specific issues that the NRC intends to address are not targeted in ways that could have the most beneficial impact on design of the MOBILE model and its role in mobile source policy and regulation. We believe the key questions to answer are the following:

1. Is MOBILE a valid representation of the processes that affect real-world vehicle emissions, including technological, sociological, and institutional factors?

2. Is the MOBILE model's output valid and accurate enough to justify its use as a measure of mobile source program effectiveness or for determining the effects of potential policy changes? If not, are there ways make the model useful for these purposes, and/or should EPA use other means to evaluate and credit mobile source emission reductions?

3. What is the appropriate way to collect emissions data so that they will be representative of vehicles on the road, and useful for determining the effect of mobile source emissions reduction programs?

[9] This document can be found at http://www4.nas.edu/webcr.nsf/ProjectScopeDisplay/BEST-U-97-01-A?OpenDocument

In terms of the model itself, the specifics of the NRC Committee's Project Scope do not explicitly address the effect of behavioral factors and the dynamic nature of the problem. To the extent that these factors are major determinants of vehicle emissions, the NRC's efforts might not improve the policy process.

If we take the NRC's charge more broadly, it is not clear that use of MOBILE is the best, or even a good, way to drive mobile source policy. MOBILE should be seen merely as one means to a goal. The NRC Committee's conclusion might indeed be that MOBILE would be adequate with a few improvements. But other conclusions, such as a totally different modeling approach, or more reliance on real-world measurements, are also possible as replacements or as supplements to the model.

There are several areas in which the NRC Committee should consider making recommendations:

- ***Draw Conclusions about the Model.*** To the extent that MOBILE6 will not resolve the structural problems of the model, the NRC should present its findings regarding whether MOBILE is a robust and valid tool for determining the effect of mobile source policies.
- ***Real-World Validation of Policies.*** The NRC should determine whether on-road measurements of unprepared vehicles should be required for final emission reduction credit to be issued. The high stakes involved in mobile source policy and the questionable accuracy of MOBILE suggest that EPA should require actual measurements of real-world emissions, rather than relying only on model results.
- ***Model Validation***. The NRC should recommend that MOBILE include more testable predictions. These tests should be designed to ensure that MOBILE generates its overall results for the right underlying reasons.

We realize that major changes to EPA's mobile source policies might not be possible. EPA is constrained by the nature of the SIP process and by past practices. However, this should not constrain the NRC Committee from recommending what it believes to be appropriate approaches. We believe consideration of the factors we suggest above are more likely to result in recommendations that enhance the validation and quality of mobile source policies.

References Cited

Arizona Auditor General (1988). Performance Audit, Vehicle Emissions Inspection and Maintenance Program, Report to Arizona Legislature by the Auditor General, Phoenix Arizona.

Brzezinski, D. J., and T. P. Newell (1998). MOBILE6: A Revised Model for Estimation of Highway Vehicle Emissions, paper presented at the AWMA 8th Annual Conference on Emission Inventories, New Orleans, December 9, 1998.

EPA (1995). I/M Briefing Book: Everything You Ever Wanted to Know About Inspection and Maintenance. Washington, D.C., United States Environmental Protection Agency (EPA).

Glover, E. L. and D. J. Brzezinski (1989). Mobile4 - Exhaust Emission Factors and Inspection/Maintenance, Benefits for Passenger Cars, U.S. Environmental Protection Agency.

Glover, E. L. and D. J. Brzezinski (1997). Analysis of the Arizona IM240 Test Program and Comparison with the TECH5 Model, EPA report #420-R-97-001 (February).

Harrington, W. and V. McConnell (1994). Modeling In-Use Vehicle Emissions and the Effects of Inspection and Maintenance Programs. Air & Waste Management Association **44**(5): 791-799.

Harrington, W., V. D. McConnell, et al. (1998). A Behavioral Analysis of EPA's MOBILE Emission Factor Model - Discussion Paper 98-47, Resources for the Future, Washington, DC.

Kachigan, S. K. (1986). Statistical Analysis, Radius Press, New York.

Klausmeier, R., S. Kishan, et al. (1995). Draft Final Report - Evaluation of the California Pilot Inspection/Maintenance (I/M) program. Sacramento, California, California Bureau of Automotive Repair (BAR), Radian Corporation.

Knepper, J. C., J. D. Benson, et al. (1993). Fuel Effects in Auto/Oil High Emitting Vehicles, Auto/Oil Air Quality Improvement Research Program (AQIRP).

Kruger, J., et al. (1999). Superstition and the Regression Effect, Skeptical Inquirer, (March/April) .

Pierson, W.R., Gertler, A.W., et al.(1996) Real-world Automotive Emissions--Summary of Studies in the Fort McHenry and Tuscarora Mountain Tunnels, Atmospheric Environment, **30**(12).

Pierson, W. R., A. W. Gertler, and R. L. Bradow, Comparison of the SCAQS Tunnel Study with Other On-Road Vehicle Emission Data, Journal of the Air and Waste Management Association, **40**(11), 1495-1504.

Scherrer, H. C. and D. B. Kittelson (1994). I/M Effectiveness as Directly Measured by Ambient CO Data. International Congress & Exposition, Detroit, Michigan, SAE Technical Paper Series, University of Minnesota.

Schwartz, J. (1995). An Analysis of the USEPA's 50-Percent Discount for Decentralized I/M Programs, California Inspection and Maintenance Review Committee.

Schwartz, J. (1995). California Inspection and Maintenance Review Committee Staff Reply to "EPA Summary Response to the California Review Committee Report on I/M Effectiveness", California Inspection and Maintenance Review Committee.

Sierra Research (1994). Investigation of MOBILE5a Emission Factors: Evaluation of I/M and LEV Program Benefits, report # SR94-06-05, prepared for the American Petroleum Institute.

Stedman, D. H., G. A. Bishop, et al. (1997). "On-Road Evaluation of an Automobile Emission Test Program." Environmental Science & Technology **31**(3): 927-931.

Stedman, D. H., G. A. Bishop, et al. (1998). "Repair Avoidance and Evaluating Inspection and Maintenance Programs." Environmental Science & Technology **32**(10): 1544-1545.

Stedman, D. H., et al. (1994). On-Road Remote Sensing of CO and HC Emissions in California, report prepared for the California Air Resources Board, contract #A032-093.

Stewart, S. J. and D. I. Gourley (1996). Technical Review of the AirCare Program, Program Year Three, Ministry of Transportation and Highways

Wenzel, T. (1998). Evaluation of Arizona's Enhanced I/M Program, Lawrence Berkeley National Laboratory.