Validation of the MEASURE Automobile Emissions Model: A Statistical Analysis

IGNATIUS FOMUNUNG

Clark Atlanta University

SIMON WASHINGTON

RANDALL GUENSLER

WILLIAM BACHMAN

Georgia Institute of Technology

ABSTRACT

This paper details the results of an external validation effort for the hot stabilized option currently included in the Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE). The MEASURE model is one of several new modal emissions models designed to improve predictions of CO, HC, and NO_x for the on-road vehicle fleet. Mathematical algorithms within MEASURE predict hot stabilized emission rates for various motor vehicle technology groups as a function of the conditions under which the vehicles are operating, specifically various aggregate measures of their speed and acceleration profiles. Validation of these algorithms is performed on an independent data set using three statistical criteria. Statistical comparisons of the predictive performance of the MEASURE and MOBILE5a models indicate that the MEASURE algorithms provide significant improvements in both average emission estimates and explanatory power over MOBILE5a for all three pollutants across almost every operating cycle tested. In addition, the MEASURE model appears to be less biased, the most critical model performance measure for point-estimate forecasts, than MOBILE5a.

Ignatius W. Fomunung, Department of Engineering, Clark Atlanta University, Atlanta, GA 30314. E-mail: ifomunun@cau.edu.

INTRODUCTION

Emission rate model uncertainties in currently employed regional emissions models arise in part because emission rates rely primarily on average speed as the dominant, continuous, independent variable in the regression analysis. However, many factors, both continuous and discrete, in addition to average speed, affect the net load demanded of an engine, which in turn affects a vehicle's resultant emissions. These factors include roadway grade, rolling resistance, aerodynamic drag, engine speed, engine friction losses, transmission losses, vehicle mass, power consumption of accessories, and so forth. Numerous references identify these factors as influential in the formulation of various pollutants; however, they are largely omitted in currently employed emission prediction algorithms (Guensler 1993).

Cicero-Fernandez et al. (1997a; 1997b) demonstrated that emissions from an individual vehicle may increase by a factor of two when driven on an uphill grade, yet current inventory models do not account for grade. In addition, real-world driving conditions, in terms of speed/acceleration distributions and/or traces, are not well represented in the current models. The Federal Test Procedure (FTP), appropriately used to develop baseline emissions factors, does not capture the extremes of emission-producing activities associated with aggressive driving. Jimenez-Palacio (1999), using a new definition of specific power, calculated the maximum specific power of the FTP to be approximately 22 kilowatts per metric ton. More telling, the research indicates that the onset of commanded enrichment for many vehicles occurs at this maximum. Commanded enrichment is responsible for elevated or "super" emissions, which can be one to several orders of magnitude higher than emissions obtained under stoichiometric engine operation. As a result, a large proportion of commanded enrichment is not likely to appear under the FTP.

Driver behavior may also be an important source of uncertainty and variability in motor vehicle emissions (Bishop et al. 1996). A study of repeated measurements on the IM240 driving cycle indicates that driver behavior may be responsible for potentially order-of-magnitude differences in emissions for clean low-emitting vehicles (Webster and Shih 1996). Despite this recognition, few advances have been made in quantifying the effect of driving behavior on emissions, except for Shih et al. (1997), who used throttle position distributions to represent driver behavior, albeit with mixed results. Their research provides evidence that throttle position distributions might be used to reflect differences in driving behavior, but such models still need refinement. The forecasting of throttle position distributions, which interact with specific driver types, facility types, and trip purposes, may prove too difficult.

Emerging Models

Efforts at improving motor vehicle emissions have occupied researchers for quite some time. Cadle et al. (1997) recently summarized advances in real-world motor vehicle emissions modeling. The U.S. Environmental Protection Agency (EPA) is currently revising the MOBILE5a emissions rate model. MOBILE6 promises to provide significant improvements in terms of representing modal impacts on emissions rates because supplemental driving cycles that mimic on-road conditions under various levels of congestion are being used to develop cycle-based speed correction factors. New certification testing cycles also promise to reduce the frequency of on-road enrichment. The USO6 cycle represents emissions in aggressive driving, and the SCO3 cycle reflects the effects of accessory loads like air conditioning usage, power steering, and so forth.

Modal modeling approaches are also currently under development. A modal emissions model being developed at the University of California, Riverside by An et al. (1997) is based on 300 vehicles tested under a variety of laboratory driving cycles. Two modal approaches developed at the Georgia Institute of Technology are included in the GIS-based modal emission model: an aggregate modal model based on statistical analysis of historic laboratory data (Guensler et al. 1998) and a load-based prediction module based on analysis of instrumented vehicle data (Rodgers 1995).

For the past six years, the Georgia Tech Research Partnership has been developing a research-grade motor vehicle emissions model within a geographic information system (GIS) framework. Once validation and peer review efforts are complete, MEASURE may serve as an alternative or supplement to the current MOBILE5a model. The aggregate modal model within MEASURE predicts emissions from light-duty vehicles. The Georgia Tech aggregate modal model predicts emissions as a function of vehicle operating mode, representing a spectrum of vehicle operating conditions including cruise, acceleration, deceleration, idle, and the power demand conditions that lead to enrichment, that is, high fuel to air ratios. The model accounts for interactions between specific vehicle fleet characteristics and vehicle operating modes. For each technology group within a light-duty motor vehicle fleet, the relationships between modal activity and emissions can differ significantly. The framework allows for facility-level aggregations of microscopic traffic simulation or disaggregation of traditional macroscopic four-step travel demand forecasting models to develop emission-specific vehicle activity data.

The aggregate modal model within MEASURE employs emission rates based on theoretical engine-emissions relationships that have been modeled using various statistical techniques (Fomunung et al. 1999). The emissions rate models have been estimated through a process that utilizes the best aspects of hierarchical tree-based regression (HTBR) (Breiman et al. 1984) and ordinary least squares (OLS) regression. The relationships are dependent on both modal and vehicle technology variables, and they are "aggregate" in the sense that they rely on bag data to derive their modal activities (Washington 1994). Thus, they are suitable for existing aggregate approaches contained within the travel demand modeling (TDM) framework.

Although much effort has been conducted and reported in the literature on the emission algorithms within MEASURE, little has been done toward the external validation of the MEASURE emissions predictions components or to compare the performance of MEASURE with that of MOBILE5a. Model validation, the use of a sample of external data to assess model predictive abilities, is perhaps the single most important measure of a model's ability to capture relationships across space and time. In addition, it is the only way to compare two competing models fairly. This paper details the results of an external validation effort for the hot stabilized exhaust option currently included in MEASURE. The performances of MEASURE and MOBILE5a are compared using mean absolute prediction errors, linear correlation coefficients between observed and predicted emissions, and mean prediction errors. Results are provided for each driving cycle and for vehicle technology classes.

MEASURE AGGREGATE MODAL MODELS

In the context of this paper, the term "model" refers to a mathematical algorithm or expression that relates emissions measurements to various explanatory variables. The model estimation data consisted of more than 13,000 laboratory tests conducted by EPA and the California Air Resources Board (CARB) using standardized test cycle conditions, as well as alternative driving cycles (Fomunung et al. 1999). The aggregate modal model algorithms presented below were estimated using the logarithm of the emission rate ratio for each pollutant as a response variable (Fomunung et al. 1999). The ratio is the emission rate (in grams per second) (g/sec) for a vehicle driven on a given cycle (or equivalently across a specified speed/acceleration matrix), divided by that vehicle's emissions rate (g/sec) obtained from the FTP bag 2 testing cycle. MEASURE's Aggregate Modal Model predicts the ratio of g/sec emission rates for several vehicle technology groups. The following sequence of equations shows the method for calculating the predicted emissions rates for each pollutant in units of either g/sec (ψ) or g/mile ($\tilde{\psi}$):

$$\psi_i = \tilde{\psi}_i \times \mathrm{DIST} / \mathrm{DUR} \tag{1},$$

$$\psi_{i\mathrm{bag2}} = \tilde{\psi}_{i\mathrm{bag2}} \times 3.91/866 \tag{2},$$

and

$$R_i = P_i \; / \; \psi_{i\mathrm{bag2}} \tag{3}.$$

In these equations, ψ_i and $\tilde{\psi}_i$ are the observed or measured pollutant (i is the index for CO, HC, or NO_x); P_i is the predicted value of pollutant i; $\psi_{i\mathrm{bag2}}$ and $\tilde{\psi}_{i\mathrm{bag2}}$ are the observed FTP bag 2 rates for pollutant i in a given vehicle; DIST is the driving cycle distance in miles; DUR is the cycle duration in seconds; 3.91 is the hot stabilized FTP bag 2 subcycle distance in miles; and 866 is the FTP bag 2 subcycle duration in seconds.

The emissions models in MEASURE are presented in two formats: in the form in which they were estimated (suitable for making statistical inferences) and in original variable units, often more intuitive for use in emissions rate prediction and for interpretation of results. The statistical details of the models are provided in tables 1, 2, and 3 for CO, HC, and NO_x, respectively. Details of the model development process including goodness-of-fit, analysis of residuals, and interpretation of coefficients are published elsewhere (Fomunung et al. 1999; Fomunung 2000).

Model Estimation Forms

Equation (4) shows the estimation form for CO. Equation (5) shows the estimation form for HC, and equation (6) shows the estimation form of NO_x.

For CO,

$$\text{Log}R_{CO} = 0.0809 + 0.002 \times \text{AVGSPD} + 0.0461 \times \text{ACC.3} + 0.0165 \times \text{IPS.60} - 0.0283 \times \text{ips45sar2} + 0.3778 \times \text{ips90tran1} - 0.0055 \times \text{tran3idle} + 0.1345 \times \text{tran5km1} + 0.3966 \times \text{finj3sar3} - 0.0887 \times \text{cat3tran1} - 0.2636 \times \text{sar3tran4} - 0.481 \times \text{flagco} \quad (4)$$

where

AVGSPD is the average speed of the driving cycle in mph,

ACC.3 is the proportion of the driving cycle on acceleration greater than three mph per second,

IPS.X is the proportion of the driving cycle on inertial power surrogate (IPS) (speed times acceleration) greater than X mph^2/sec (Washington 1994) (thus, IPS.60 implies IPS greater than 60 mph^2/sec),

ips45sar2 is an interaction between IPS.45 and a vehicle with no air injection,

ips90tran1 is an interaction variable for a vehicle with automatic transmission on IPS.90,

cat3idle is an interaction variable for a three-speed manual transmission at idle,

tran5km1 is an interaction variable for a five-speed manual transmission vehicle with mileage ≤ 25,000 miles,

finj3sar3 is an interaction variable for a vehicle that has throttle body fuel injection and pump air injection,

cat3tran1 is an interaction variable for a vehicle with automatic transmission and three-way catalyst (TWC),

TABLE 1 CO Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
Intercept	0.0809	0.0154	5.2382	<0.001
AVGSPD	0.0020	0.0004	5.0514	<0.001
ACC.3	0.0461	0.0026	17.8998	<0.001
IPS.60	0.0165	0.0066	2.4909	<0.013
ips45sar2	–0.0283	0.0067	–4.2136	<0.001
ips90tran1	0.3778	0.0265	14.2899	<0.001
cat3idle	–0.0055	0.0004	–13.8299	<0.001
tran5km1	0.1345	0.0134	10.0067	<0.001
finj3sar3	0.3966	0.0314	12.6305	<0.001
cat3tran1	–0.0887	0.0145	–6.1218	<0.001
sar3tran4	–0.2636	0.1177	–2.2401	<0.025
flagco	–0.4810	0.0290	–16.5777	<0.001

Residual standard error: 0.9177 on 12,965 degrees of freedom
R^2 (adjusted): 0.1726*
F-statistic: 245.9 on 11 and 12,965 degrees of freedom, the p-value is 0
*The low R^2 for the CO model is an indication that the model doesn't fit very well. It is low relative to the values for the HC and NO_x models because the production mechanism for CO emissions in the engine and exhaust manifold is more complex than for HC and NO_x emissions. The EPA testing protocol that generated the current database did not include important variables such as catalyst efficiency under varying load conditions and various transient oxygen effects, which research has shown account for much of the variability in CO emissions. It is expected that a CO model estimated using a data set with these additional variables would result in a much improved R^2.

TABLE 2 HC Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
intercept	0.1685	0.0098	17.1164	<0.001
my79	0.3601	0.0098	36.5975	<0.001
finj2tran4	–0.0732	0.0196	–3.7260	<0.001
cat2sar1	0.3324	0.0206	16.1707	<0.001
cat3sar1	–0.4201	0.0247	–17.004	<0.001
cat3sar2	–0.1188	0.0123	–9.6257	<0.001
sar3tran1	–0.3602	0.0194	–18.5248	<0.001
cyl8	–0.2349	0.0115	–20.4826	<0.001
sar3km1	–0.2175	0.0152	–14.3368	<0.001
finj2km3	–0.0290	0.0034	–8.4548	<0.001
acc1finj2	–0.0550	0.0030	–18.3900	<0.001
acc3cat2	–1.3528	0.0234	–57.7883	<0.001
ips90sar3	–0.9201	0.0566	–16.2530	<0.001
dps8finj2	0.0391	0.0007	54.0156	<0.001

Residual standard error: 9.414 on 12,350 degrees of freedom
R^2 (adjusted): 0.6094
F-statistic: 1,482 on 13 and 12,350 degrees of freedom, the *p*-value is 0

TABLE 3 NO_x Model Details

Variable	Estimated coefficient	Standard error	*t*-value	Pr(>|*t*|)
(intercept)	–0.5864	0.0068	–85.9273	<0.001
AVGSPD	0.0225	0.0002	131.6271	<0.001
IPS.120	0.3424	0.0452	7.5684	<0.001
ACC.6	0.6329	0.1683	3.7595	<0.002
DEC.2	0.0247	0.0007	34.8026	<0.001
finj2km1	0.0083	0.0008	10.4205	<0.001
finj2km2	0.0028	0.0004	6.8670	<0.001
cat2km3	–0.0021	0.0004	–5.9243	<0.001
cat3km2	0.0026	0.0002	13.5707	<0.001
cat3km3	0.0003	0.0001	2.9355	<0.001
finj1km3flagnox	–0.0085	0.0015	–5.7854	<0.001
finj3km3flagnox	–0.0068	0.0009	–7.4491	<0.001

Residual standard error: 0.3479 on 12,962 degrees of freedom
R^2 (adjusted): 0.623
F-statistic: 1,947 on 13 and 12,962 degrees of freedom, the *p*-value is 0 under varying load conditions and various transient oxygen effects, which research has shown account for much of the variability in CO emissions. It is expected that a CO model estimated using a data set with these additional variables would result in a much improved R^2.

sar3tran4 is an interaction variable for a vehicle with four-speed manual transmission and pump air injection, and

flagco is a flag used to tag a vehicle emitting high CO emissions (Wolf et al 1998).

For HC,

$$\text{Log}R_{HC} = 0.1685 + 0.3601(\text{my79}) - 0.0732\ (\text{finj2tran4}) + 0.3324\ (\text{cat2sar1}) - 0.4201\ (\text{cat3sar1}) - 0.1188\ (\text{cat3sar2}) - 0.3602\ (\text{sar3tran1}) - 0.2349\ (\text{cyl8}) - 0.2175\ (\text{sar3km1}) - 0.0290\ (\text{finj2km3}) - 0.055\ (\text{ACC.1finj2}) - 1.3528\ (\text{ACC.3cat2}) - 0.9201\ (\text{IPS.90sar3}) + 0.0391\ (\text{DPS.800finj2}) \quad (5)$$

where

my79	is model year < 79
finj2tran4	is an interaction variable for a four-speed manual transmission vehicle with a carburetor,
cat2sar1	is a variable for a pre-1981 model year vehicle with an "oxidation only" catalyst and of unknown air injection type,
cat3sar1	is a variable for a pre-1981 model year vehicle with a TWC and of unknown air injection type,
cat3sar2	is a variable for a vehicle with TWC and no air injection,
sar3tran1	is an automatic transmission vehicle with pump air injection,
cyl8	is a vehicle with an eight-cylinder engine,
sar3km1	is a vehicle with pump air injection and mileage ≤ 25,000 miles,
finj2km3	is a vehicle with pump air injection and 50,000 < mileage ≤ 100,000 miles,
acc1finj2	is a carburetor-equipped vehicle operating with acceleration greater than one mph per second,
acc3cat2	is an "oxidation only" catalyst vehicle on ACC.3,
ips90sar3	is a vehicle with air pump on IPS.90, and
dps800finj2	is the proportion of drag power surrogate (DPS) (speed times speed times acceleration) greater than 800 mph^3/sec.

For NO_x,

$$\mathrm{Log}R_{NO_x} = -0.5864 + 0.0225 \times \mathrm{AVGSPD} + 0.3424 \times \mathrm{IPS.120} + 0.6329 \times \mathrm{ACC.6} + 0.0247 \times \mathrm{DEC.2} + 0.0083 \times \mathrm{finj2km1} + 0.0028 \times \mathrm{finj2km2} - 0.0021 \times \mathrm{cat2km3} + 0.0026 \times \mathrm{cat3km2} + 0.0003 \times \mathrm{cat3km3} - 0.0085 \times \mathrm{finj1km3flagnox} - 0.0068 \times \mathrm{finj3km3flagnox} \quad (6)$$

where

IPS.120	is IPS greater than 120 mph^2/sec,
ACC.6	is the proportion of acceleration greater than six mph per second,
DEC.2	is the proportion of deceleration greater than two mph per second,
finj2km1	is a carburetor-equipped vehicle with mileage less than 25,000 miles,
finj2km2	is a carburetor-equipped vehicle with 25,000 < mileage ≤ 50,000 miles,
cat2km3	is an "oxidation only" catalyst vehicle with 50,000 < mileage ≤ 100,000miles,
cat3km2	is a TWC vehicle with 25,000 < mileage ≤ 50,000 miles,
cat3km3	is a TWC vehicle with 50,000 < mileage ≤ 100,000 miles,
finj1km3flagnox	is a second-order interaction variable for a high-emitting vehicle with port fuel injection and 50,000 < mileage ≤ 100,000 miles, and
finj3km3flagnox	is a second-order interaction variable for a high-emitting vehicle with throttle body fuel injection and 50,000 < mileage ≤ 100,000 miles.

This implies that on a vehicle-by-vehicle basis after calculating R_i from the response variable, the predicted rate P_i in g/sec is

$$P_i = R_i \times \psi_{i\mathrm{bag2}} \quad (7).$$

Note that equation (7) is similar in form to the embedded algorithm in MOBILE5a, which gives emission rates as Correction Factors times Base Emission Rate (BER). BER is similar to $\psi_{i\mathrm{bag2}}$; R_i represents all the variables which figure into the models for each pollutant and can be thought of as speed, load, and technology correction factors. The conversion to g/mile is straightforward and given by

$$\tilde{P}_i = R_i \times \psi_{i\mathrm{bag2}} \times 1/\mathrm{AVGSPD} \quad (8).$$

Model Prediction Forms

The prediction forms for CO, HC, and NO_x are shown in equations (9), (10), and (11), respectively, and the variables are as previously described.

The prediction equations are no more than the antilogs of the estimation equations.

For CO, in g/sec,

$$P_{CO} = 1.205 \times \text{FTP bag2} \times \text{antilog}\,[0.002 \times \text{AVGSPD} + 0.0461 \times \text{ACC.3} + 0.0165 \times \text{IPS.60} - 0.0283 \times \text{ips45sar2} + 0.3778 \times \text{ips90tr1} - 0.0055 \times \text{tran3idle} + 0.1345 \times \text{tran51} + 0.3966 \times \text{finj3sar3} - 0.0887 \times \text{cat3tran1} - 0.2636 \times \text{sar3tran4} - 0.481 \times \text{flagco}] \quad (9).$$

For HC,

$$P_{HC} = 1.474 \times \text{FTP bag2hc} \times \text{antilog}[0.3601(\text{myhc81}) - 0.0732\,(\text{finj2tran4}) + 0.3324\,(\text{cat2sar1}) - 0.4201\,(\text{cat3sar1}) - 0.1188\,(\text{cat3sar2}) - 0.3602\,(\text{sar3tran1}) - 0.2349\,(\text{cyl8}) - 0.2175\,(\text{sar3km1}) - 0.0290\,(\text{finj2km3}) - 0.055\,(\text{ACC.1finj2}) - 1.3528\,(\text{ACC.3cat2}) - 0.9201\,(\text{IPS.90sar3}) + 0.0391\,(\text{DPS.800finj2})] \quad (10).$$

For NO_x,

$$P_{NOx} = 0.259 \times \text{FTP bag2} \times \text{antilog}\,[0.0225(\text{AVGSPD}) + 0.3424(\text{IPS.120}) + 0.6329(\text{ACC.6}) + 0.0247(\text{DEC.2}) + 0.0083(\text{finj2km1}) + 0.0028(\text{finj2km2}) - 0.002(\text{cat2km3}) + 0.0026(\text{cat3km2}) + 0.0003(\text{cat3km3}) - 0.0085(\text{finj1km3flagnox}) - 0.0068(\text{finj3km3flagnox})] \quad (11).$$

The algorithms shown in equations (4) to (6) indicate that, apart from AVGSPD, which appears in both the CO and NO_x models, a different collection of variables is needed to model each pollutant. This finding is in agreement with theoretical expectations. The production and distribution of all three pollutants are functions of the physico-chemical processes occurring in the engine. While CO and NO_x are principally produced as a result of chemical and kinetic mechanisms within the engine, the production of HC is heavily dependent on the physical processes within the engine. The phrase "physical processes" is used in an inclusive sense to embody both the physical structure of the engine combustion chamber and the physics of the combustion process within the combustion chamber. It has long been recognized that the crevices within the combustion chamber are a significant source of exhaust hydrocarbons (Heywood 1988). Therefore, it is not surprising that different variables are needed to model each pollutant. For example, the variable cyl8, which is positively correlated with the number of crevices in the engine, is a significant predictor variable in the HC model but is insignificant in both the CO and NO_x models.

VALIDATION DATA SET DESCRIPTION

Model validation consists of two types, internal and external. Internal validation consists of model-checking for plausibility of signs and magnitudes of estimated coefficients, agreement with past models and theory, and model diagnostic checks such as distribution of error terms, normality of error terms, and so forth. Internal validation was performed as part of the model estimation procedure and is documented in Fomunung et al. (1999) and Fomunung (2000). External validation is the process whereby a model is compared to data collected "outside" the modeling framework (i.e., data from another location or time). External validation is the only way to check if a model has been "overfit" to data, thus capturing spurious rather than real relationships or underlying structure in the data. It is also the only way to determine whether the relationships captured in the estimated model reflect the same relationships elsewhere or over time. Finally, external validation is the only objective way to compare two models estimated using different data. These objectives have motivated the validation of the MEASURE emission prediction algorithms: to assess its transferability and to compare its performance to the current in-practice emission predictions model, MOBILE5a.

The data used for MEASURE and MOBILE5a validation consist of 50 vehicles tested across 16 different hot stabilized driving cycles. Of the 50 vehicles, 4 are Chrysler-manufactured cars, 13 are Ford cars, 21 are GM cars, and the rest are imports. One of the four Chrysler cars is a 1983-model year car with 94,399 miles. Another is a 1989-model year car with 118,586 miles, and two are 1995-model year cars with 20,855 miles and 28,525 miles, respectively. The Ford cars are from model years 1985 to 1992 with between 53,000 and 123,000 miles. The GM cars are from model years 1985 to 1996 with 16,000 to 180,000 miles.

TABLE 4 Number of Tests, Average Speeds, Maximum Speeds, and Maximum Instantaneous Acceleration of Each Test Cycle for the Validation Data Set

Test cycle description	Name	Number of tests	Average speed (mph)	Maximum speed (mph)	Maximum acceleration (mph/sec)
Arterial LOS A-B cycle	ARTA	50	24.8	58.9	5.0
Arterial LOS C-D cycle	ARTC	50	19.2	49.5	5.7
Arterial LOS E-F cycle	ARTE	50	11.6	39.9	5.8
Hot running 505	F505	50	25.6	56.3	3.4
New York City cycle	FNYC	50	7.1	27.7	6.0
Freeway LOS A-C cycle	FWAC	50	59.7	73.1	3.4
High-speed freeway cycle	FWHS	50	63.2	74.7	2.7
Freeway LOS D cycle	FWYD	50	52.9	70.6	2.3
Freeway LOS E cycle	FWYE	50	30.5	63.0	5.3
Freeway LOS F cycle	FWYF	50	18.6	49.9	6.9
Freeway LOS G cycle	FWYG	50	13.1	35.7	3.8
CARB "unified" LA92 cycle	LA92	49	24.7	67.2	6.9
Local roadways cycle	LOCL	50	12.8	38.3	3.7
Freeway ramp cycle	RAMP	50	34.6	60.2	5.7
Start cycle	ST01	49	20.1	41.0	5.1
Areawide non-freeway cycle	WIDE	49	19.4	52.3	6.4

The model years of the imports are from 1987 to 1993 with 30,000 to 197,000 miles.

Neither the MEASURE nor MOBILE5a models were originally estimated using data from these vehicles. Table 4 lists the different cycles used and shows their average speed s, maximum speeds, and acceleration characteristics. EPA tested each vehicle on every cycle (three cycles only included 49 of the 50 vehicles), and the near-balanced sampling design results in the ability to segregate vehicle-to-vehicle, within vehicle, and cycle-to-cycle variability. A similar list of cycles used in the MEASURE model development is shown in table 5. There are minor differences between the two data sets. First, only two driving cycles, NYCC and the CARB "unified" cycle, were used in both instances. Second, the data ranges for the parameters of interest are slightly different: average speeds range from 2.45 to 59.9 mph in the model data and from 7.1 to 63.2 mph in the validation data; maximum speeds range from 10 to 71.3 mph in the model data and from 27.7 to 74.7 mph in the validation data; and maximum acceleration ranges from 1.5 to 6.9 mph per second in the model data and from 2.3 to 6.9 mph per second in the validation data. These differences notwithstanding, the independence of the validation data set lends itself well to purposes of model evaluation.

The aggregation of existing in-use EPA data obtained from past testing efforts by both the EPA and CARB and used to develop the aggregate modal emission models in MEASURE is different from that of the validation data set in several respects. First, not all vehicles were tested on all cycles. Second, the vehicles recruited, in aggregate, are not representative of the national on-road vehicle fleet. Finally, there is very little replication testing, so within-driver variability is not known. However, the relatively large size of the aggregate database provides an opportunity to 1) obtain precise estimates of a multitude of vehicle-specific technology effects, 2) predict emission rates over a wide range of makes and model years, and 3) assess the effect of mileage accrual.

PREPARING THE MEASURE AND MOBILE5A MODELS FOR VALIDATION AND COMPARISON

Before being able to assess the predictive abilities of both MEASURE and MOBILE5a, it was necessary to set some ground rules for model validation and comparison. First, it was necessary to determine which "classes" of vehicles would be compared. In other words, it seemed that for at least some comparisons it would be useful to see how the two models predict emission rates for classes of

TABLE 5 Number of Tests, Average Speeds, Maximum Speeds, and Maximum Instantaneous Acceleration of Each Test Cycle for the Model Development Data Set

Cycle name*	Number of tests	Average speed (mph)	Maximum speed (mph)	Maximum acceleration (mph/sec)
Arterial 1	23	14.30	44.9	6.9
Arterial 2	21	24.06	46.3	5.8
Arterial 3	23	34.39	54.9	6.9
CCDH (bag 2)	58	13.40	29.8	3.0
Cycle 1	21	59.90	71.3	1.5
Cycle 2	23	53.31	68.0	2.0
Cycle 3	22	39.28	68.9	4.6
Cycle 4	20	31.54	61.9	3.3
Cycle 5	23	23.60	56.5	3.9
Cycle 6	21	15.94	40.9	5.0
Cycle 7	23	9.17	39.7	3.1
HFET	6586	48.20	59.9	3.2
LSP 1	813	2.45	10.0	2.4
LSP 2	814	3.62	14.0	2.5
LSP 3	815	4.04	16.0	3.4
NYCC	1218	7.07	27.7	6.0
SC12	1199	11.70	29.1	3.3
SC36	1201	36.50	57.0	6.0
Unified cycle (bag 2)	88	27.40	67.2	6.9
FTP (bag 2)	All	16.20	34.3	3.3

*Arterial 1, 2, and 3 denote cycles developed in California for roadway specific testing.
CCDH denotes a cycle developed for use by the Colorado Department of Health for high altitude testing.
Cycles 1, 2, 3, 4, 5, 6, and 7 represent high-speed cycles developed in California for roadway facility testing.
HFET stands for Highway Fuel Economy Test.
LSP 1, 2, and 3 refer to EPA's low-speed testing cycles.
SC12 and SC36 refer to EPA speed correction factor cycles.
Unified Cycle (LA92) refers to a new California laboratory testing cycle providing greater engine loads.

vehicles that are fundamentally different since emissions are characteristically different across such classes. Second, emission factors need to be converted to comparable and meaningful units, i.e., emissions in grams per second. Finally, appropriate criteria for comparison needed to be established.

Technology Class Definition

Four different emissions-control technology types were investigated during model development: fuel injection, catalytic conversion, transmission, and supplemental air injection. Each technology can be represented by several different types, as indicated below (with coding shown):

- Fuel Injection (FINJ)
    1. Port fuel injection (PFI), coded as finj1
    2. Carburetor and all pre-1980 domestic cars, coded as finj2
    3. Throttle body fuel injection (TBI), coded as finj3
    4. Unknown type pre-1980 import and both 1980 domestic and import, coded as finj4
- Catalytic Converter (CAT)
    1. None, coded as cat1
    2. Oxidation only, coded as cat2
    3. Three-way catalyst, coded as cat3
    4. Oxidation and three-way catalyst, coded as cat4
- Supplemental Air Recirculation (SAR)
    1. Pre-1980 of unknown type, coded as sar1
    2. None, coded as sar2
    3. Air pump, coded as sar3
    4. Pulse, coded as sar4
- Transmission Speed (TRAN)
    1. Automatic, coded as tran1
    2. Semi automatic, coded as tran2

3. Three-speed manual, coded as tran3
4. Four-speed manual, coded as tran4
5. Five-speed manual, coded as tran5

To capture the effects of deterioration, accrued test vehicle mileage was used as a surrogate for deterioration. Fomunung et al. (1999) have previously determined that deterioration appears to occur more like a step function rather than a constant deterioration over time, so four deterioration mileage groups (or bins) are employed in the models. These groups are "25,000 miles or less," "25,000 to 50,000 miles," "50,000 to 100,000 miles," and "100,000 miles or more." They are labeled km1, km2, km3, and km4, respectively.

It was a fairly complex task to implement the regression equations inside the MEASURE model. First, it was necessary to define mutually exclusive technology groups that would interact uniquely with vehicle operating modes. In essence, it was necessary to employ classification rules that resulted in mutually exclusive and collectively exhaustive technology groups. To define preliminary classification rules, a matrix of all possible combinations of the four technology variables plus the mileage bins and high-emitter status that appear in the regression model (a total of $4 \times 4 \times 4 \times 5 \times 4 \times 2$ or 2,560 technology rules) was created for each pollutant. Then, using equation (3) and each of equations (9), (10), and (11), which include technology and modal interactions, for CO, HC, and NO_x, respectively, the predicted emission rate ratio for each pollutant was computed for each of the 2,560 initial classification rules using the modal variables from the highway fuel economy test.

Classification rules that yielded the same predicted emission rate ratio for any given cycle were then clustered together; that is, they were collapsed into the mutually exclusive technology groups that are represented in the regression equation. A cross-check with modal variables from other driving cycles (LA4, Low Speed 1, and High Speed Cycle 1) produced the same technology groups. Each technology group cluster was then assigned an aggregate definition to represent a "technology group," as distinct from the former "classification rule." Consequently, 44 technology groups were defined for CO, 120 for HC, and 13 for NO_x, and all were assigned consecutive numerical labels beginning from 1. Thus, CO technology groups were labeled from 1 to 44; HC, from 1 to 120; and NO_x from 1 to 13. The definition of each technology group can be found in Fomunung (2000). The vehicle activity of each of these technology groups is then tracked separately within the MEASURE model because the technology and modal activity interaction variables appearing in the regression equations are different for each group.

Emissions Rates

The next step was to predict emissions for each pollutant for any given cycle and technology group. To predict emissions for each technology class one at a time, equation (7) is modified to

$$P_i = R_{ij} \times \psi_{ij\text{bag2}}, \qquad (12)$$

where P_i is measured in g/sec and $\psi_{ij\text{bag2}}$ is now defined as the average of the base emissions rate (FTP bag 2), in g/sec, of pollutant i for technology class j. Note again that in this form, the term R_{ij} in equation (12) represents a cycle-specific correction factor for each technology class. The R_{ij} is the predicted rate ratio of pollutant i for technology class j. The values for $\psi_{ij\text{bag2}}$ are obtained from the FTP bag2 measurements in the original data set, while values for R_{ij} depend on the modal variables put into the model.

Criteria for Model Validation and Comparison

There are a number of model goodness-of-fit criteria that can be used to assess the difference between the emissions predicted by MEASURE and MOBILE5a and the emissions observed in the validation data. The focus in this paper is on point estimates of emissions. That is, an independent validation sample is used to compare the performance of MEASURE and MOBILE5a in predicting emissions of CO, HC, and NO_x. Overall model bias, the mean difference between predicted and observed emissions for a sufficiently large validation sample, reflects perhaps the most important criterion for comparing whether a model is working well in practice.

This study assesses the relative performance of the two models, MEASURE and MOBILE5a, using three statistical measures of effectiveness: the linear correlation coefficient, the root mean squared error (RMSE) (Neter et al. 1996), and the mean prediction error. The linear correlation coefficient reflects the degree to which a linear relationship exists between observed and predicted emissions. A high linear correlation coefficient would imply a close correspondence between paired data (predicted and observed emissions for vehicle *i*), whereas a low coefficient would imply the reverse. The RMSE is a measure of the prediction error. When comparing two models, the model with a smaller RMSE is a better predictor of the observed phenomenon. In addition, low values of RMSE accompanied by a high linear correlation coefficient is a good indicator that a model predicts well. The third measure of comparison is mean prediction error, ideally close to zero.

The MOBILE5a hot stabilized emission rates for each vehicle in the data set were predicted from the FTP bag 2 hot stabilized emission rate for each vehicle. The MOBILE5a input file provided by the EPA Region 4 office for Atlanta was modified to reflect 100% hot stabilized operations by setting the fractions of cold and hot start vehicle miles traveled (VMT) to zero. The model was set in a model year mode to predict emission rates for each model year. The model was then run in five-mph average speed increments to develop an emission rate matrix by model year and average speed for calendar year 1997. A matrix of emission rate ratios was developed from the emission rate matrix, with the 20-mph emission rate serving as the baseline emission rate (to conform with MOBILE5a internal assumptions related to the 19.6-mph average speed of the composite FTP). The emission rate ratio is equivalent to the speed correction factor implemented by MOBILE5a for each model year. The emission rate ratio for the average speed of the test cycle (found in the matrix using cubic spline interpolation) was then multiplied by the hot stabilized FTP bag 2 emission rate for that vehicle to estimate emissions on the alternative test cycle.

It is worth mentioning briefly that models were not compared based on the confidence in mean emission predictions, despite the fact that these comparisons may be useful. These comparisons are omitted for two important reasons. First, the data set used to estimate the emissions models within MEASURE is much larger than that used for estimating MOBILE5a, and thus statistical estimates are likely to be inherently more precise for MEASURE. Second, the regulatory arena in which models are employed has yet to embrace the use of confidence intervals on model outputs; therefore, comparisons of model efficiency would not likely lead to a strong argument for one model over another since precision is not applied in practice. It is not without hesitation that these comparisons have been omitted; the authors strongly believe that these types of comparisons are valid criteria for mounting evidence in favor of one model over another and could be useful in policy arenas.

RESULTS OF THE MODEL VALIDATION EXERCISE

This section describes the results of the validation of the MEASURE and MOBILE5a emission factor modules by comparing their prediction abilities across the set of validation data. Using validation vehicle characteristics and emissions results for each of the three pollutants CO, HC, and NO_x, the MEASURE and MOBILE5a emissions algorithms, shown in equations (9), (10), and (11), respectively, for the MEASURE model algorithms, were used to predict the observed data.

Because a vehicle fleet is usually tracked, in practice, by characterizing the number of vehicles in each technology class and by model year, model validation results were computed both for aggregate data (all vehicles) by driving cycle and by technology class. The results provided on a driving cycle basis yield information on how well the models explain variability in emissions due to differences in modal activity or driving profiles, while technology-class based results yield information on how well the models explain emission differences caused by disparate vehicle technologies.

The results of the performance evaluation are shown in tables 6 through 11. The linear correlation results on a cycle basis are shown in table 6, while table 7 (a, b, c) lists the results on a technology class basis for CO, HC, and NO_x, respectively. The number of vehicles tested on each cycle is shown in table 4, whereas table 7 (a, b, c), shows that the 797 vehicle tests in the validation data set are distributed into the following: 16 CO technol-

TABLE 6 Correlation Coefficients:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Cycle

Cycle	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)		
	CO	HC	NOx	CO	HC	NOx
ARTA	**0.559**	**0.702**	**0.391**	0.268	0.243	0.339
ARTC	**0.463**	**0.577**	**0.411**	0.368	0.199	0.269
ARTE	**0.432**	**0.606**	**0.398**	0.314	0.252	0.280
F505	**0.602**	**0.688**	**0.372**	0.266	0.302	0.313
FNYC	**0.399**	**0.601**	**0.446**	0.314	0.263	–0.001
FWAC	**0.642**	**0.647**	**0.496**	0.221	0.255	0.282
FWHS	**0.634**	**0.686**	0.428	0.229	0.232	**0.519**
FWYD	**0.581**	**0.522**	0.400	0.395	0.373	**0.513**
FWYE	**0.545**	**0.672**	0.428	0.370	0.323	**0.451**
FWYF	**0.464**	**0.598**	**0.465**	0.368	0.225	0.299
FWYG	**0.458**	**0.581**	**0.389**	0.336	0.264	0.265
LA92	**0.579**	**0.630**	**0.424**	0.355	0.230	0.321
LOCL	**0.465**	**0.616**	**0.434**	0.314	0.235	0.260
RAMP	**0.610**	**0.630**	0.361	0.306	0.357	**0.379**
ST01	**0.665**	**0.689**	**0.509**	0.192	0.151	0.205
WIDE	**0.561**	**0.682**	**0.424**	0.323	0.269	0.273

*Highest value for each comparison is **bolded**

TABLE 7a Linear Correlation Coefficients* for CO: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
3	16	0.512	**0.514**
6	16	**0.781**	0.548
11	32	0.164	**0.533**
14	190	**0.293**	0.120
19	16	**0.626**	0.467
20	32	0.599	**0.635**
21	64	0.578	**0.877**
22	112	**0.501**	0.456
23	176	0.433	**0.476**
27	16	**0.849**	0.765
33	15	**0.975**	0.908
36	32	0.500	**0.535**
39	16	**0.880**	0.809
40	16	**0.952**	0.908
41	32	**0.735**	0.534
42	16	**0.624**	0.439

*Highest value for each comparison is **bolded**

ogy classes out of a possible 44, 13 HC technology classes out of 120, and 5 NO_x technology classes out of 13. In addition, the number of vehicle tests in each technology class is shown. The results in table 6 show that for CO and HC, the MEASURE model outperforms MOBILE5a across all test cycles (the highest linear correlation coefficient in each comparison is bolded), while for NO_x both models perform equally well across almost all cycles, with MEASURE doing better in the rest of

TABLE 7b Linear Correlation Coefficients* for HC: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
32	16	**0.597**	0.555
34	16	**–0.099**	–0.065
38	16	0.095	**0.108**
51	16	**0.126**	0.115
54	16	–0.452	**–0.459**
77	304	**0.370**	0.092
80	191	0.145	**0.213**
84	79	**–0.110**	–0.042
95	64	**0.296**	0.111
96	16	**0.539**	0.460
97	15	**0.946**	0.915
108	16	0.075	**0.126**
112	32	0.085	**0.306**

*Highest value for each comparison is **bolded**

Table 7c Linear Correlation Coefficients* for NO_x: Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

Tech class I.D.	Number of tests	Observed vs. predicted MEASURE (g/sec)	Observed vs. predicted MOBILE5a (g/sec)
4	16	–0.288	**–0.296**
5	161	0.368	**0.449**
6	556	0.452	**0.497**
7	48	0.746	**0.939**
8	16	**0.926**	0.952

*Highest value for each comparison is **bolded**

the cycles. For the CO and HC results in table 7, no general trend is discernible, but it can be noted that for a majority of the results MEASURE performs equally well or better than MOBILE5a. For NO_x, however, MOBILE5a performs slightly better than MEASURE in four technology classes and significantly better in technology class seven.

Tables 8 and 9 contain the results from the root mean square error analysis (the smallest RMSE is bolded in each comparison). Table 8 shows the results on a cycle basis and table 9, on a technology class basis. As with the case of the linear correlation coefficient, the results on a cycle basis indicate that for CO and HC, MEASURE performs better than MOBILE5a, but for NO_x, MEASURE performs equally well or slightly better than MOBILE5a. On a technology class basis, MEASURE is only marginally better than MOBILE5a for CO and HC, and results are mixed for NO_x.

Table10 shows the result of the mean prediction error on a cycle basis, and table 11 shows the results on a technology class basis (smallest mean prediction error is bolded in each comparison). Also shown in both tables in underlined italics are the overall weighted average mean prediction errors per pollutant. To provide the reader with a quick assessment of the relative improvement of one model over the other, a column with the ratio of mean prediction error using MOBILE5a to that of MEASURE is highlighted in table 10. The same comparison on a technology class basis is shown in table 12. When comparing mean prediction error, it can be seen that MEASURE consistently overpredicts, while MOBILE5a consistently underpredicts, both on cycle and technology class bases. However, the same results indicate that across all cycles and

TABLE 8 Root Mean Square Prediction Error:* Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Cycle

Cycle	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)		
	CO	HC	NO_x	CO	HC	NO_x
ARTA	**0.1038**	**0.0069**	**0.0071**	0.1362	0.0084	0.0072
ARTC	**0.0900**	**0.0078**	**0.0064**	0.1149	0.0086	0.0065
ARTE	**0.0763**	**0.0063**	**0.0048**	0.0980	0.0071	0.0050
F505	**0.0990**	**0.0056**	**0.0072**	0.1268	0.0068	0.0075
FNYC	**0.0679**	**0.0056**	**0.0024**	0.0830	0.0064	0.0029
FWAC	**0.2763**	**0.0095**	0.0166	0.3232	0.0110	**0.0162**
FWHS	**0.3271**	**0.0107**	0.0194	0.3615	0.0121	**0.0180**
FWYD	**0.1947**	**0.0084**	**0.0129**	0.2429	0.0099	0.0138
FWYE	**0.1189**	**0.0073**	**0.0074**	0.1529	0.0089	0.0080
FWYF	**0.0787**	**0.0069**	0.0059	0.1029	0.0078	0.0059
FWYG	**0.0658**	**0.0057**	0.0029	0.0784	0.0061	0.0029
LA92	**0.1099**	**0.0073**	**0.0080**	0.1505	0.0089	0.0085
LOCL	**0.0684**	**0.0058**	0.0042	0.0832	0.0064	**0.0041**
RAMP	**0.2153**	**0.0107**	**0.0121**	0.2666	0.0130	0.0128
ST01	**0.2031**	**0.0202**	**0.0092**	0.2564	0.0238	0.0093
WIDE	**0.0923**	**0.0064**	**0.0057**	0.1266	0.0080	0.0058

*Smallest value for each comparison is **bolded.**

TABLE 9 Root Mean Square Error (RMSE):* Observed vs. Predicted Using MEASURE, and Observed vs. Predicted Using MOBILE5a, by Technology Class

MEASURE									MOBILE5a								
CO			HC			NO_x			CO			HC			NO_x		
Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE	Tech Class	n**	RMSE
3	16	0.109	32	16	**1.09E-02**	4	16	**2.5E-03**	3	16	**0.020**	32	16	1.2E-02	4	16	3.4E-03
6	16	**0.071**	34	16	**7.41E-03**	5	161	8.4E-03	6	16	0.104	34	16	7.7E-03	5	161	**7.5E-03**
11	32	0.153	38	16	8.33E-03	6	556	7.9E-03	11	32	**0.145**	38	16	**1.1E-03**	6	556	**7.6E-03**
14	190	**0.076**	51	16	**1.10E-02**	7	48	**9.9E-03**	14	190	0.079	51	16	1.2E-02	7	48	1.3E-02
19	16	**0.121**	54	16	1.20E-02	8	16	**1.9E-02**	19	16	0.129	54	16	1.2E-02	8	16	3.5E-02
20	32	**0.225**	77	304	**6.28E-03**				20	32	0.230	77	304	6.4E-03			
21	64	0.083	80	191	3.80E-03				21	64	**0.072**	80	191	**3.7E-03**			
22	112	**0.055**	84	79	5.11E-03				22	112	0.064	84	79	**5.8E-03**			
23	176	**0.038**	95	64	**4.68E-03**				23	176	0.040	95	64	4.8E-03			
27	16	**0.109**	96	16	**2.14E-02**				27	16	0.123	96	16	3.8E-02			
33	15	**0.575**	97	15	**3.82E-02**				33	15	0.930	97	15	4.2E-02			
36	32	**0.420**	108	16	**5.49E-03**				36	32	0.557	108	16	2.5E-02			
39	16	**0.382**	112	32	**4.33E-02**				39	16	0.588	112	32	5.8E-02			
40	16	**1.255**							40	16	1.303						
41	32	**0.251**							41	32	0.480						
42	16	**0.088**							42	16	0.338						

*Smallest value for each comparison is **bolded.**
**n is the number of tests in each technology class of the validation data set.

TABLE 10 Mean Prediction Error:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Cycle

	Observed vs. predicted MEASURE (g/sec)			Observed vs. predicted MOBILE5a (g/sec)			MOBILE5a error/MEASURE error (absolute ratio values)		
Cycle	CO	HC	NO_x	CO	HC	NO_x	CO	HC	NO_x
ARTA	**0.074**	**0.004**	0.006	–0.134	–0.008	–0.006	1.8	2.0	1.0
ARTC	**0.054**	**0.005**	0.006	–0.113	–0.008	–0.006	2.1	1.6	1.0
ARTE	**0.037**	**0.003**	**0.004**	–0.096	–0.007	–0.005	2.6	2.3	1.3
F505	**0.064**	**0.003**	0.006	–0.125	–0.006	–0.006	2.0	2.0	1.0
FNYC	**0.022**	**0.002**	**0.002**	–0.082	–0.006	–0.003	3.7	3.0	1.5
FWAC	**0.265**	**0.007**	**0.002**	–0.323	–0.010	–0.011	1.2	1.4	5.5
FWHS	**0.314**	**0.008**	**0.001**	–0.360	–0.011	–0.011	1.2	1.4	11.0
FWYD	**0.174**	**0.005**	**0.005**	–0.242	–0.009	–0.010	1.4	1.8	2.0
FWYE	**0.089**	**0.005**	**0.006**	–0.151	–0.008	–0.007	1.7	1.6	1.2
FWYF	**0.042**	**0.004**	0.005	–0.100	–0.007	–0.005	2.4	1.8	1.0
FWYG	**0.017**	**0.002**	0.002	–0.075	–0.006	–0.002	4.4	3.0	1.0
LA92	**0.087**	**0.005**	**0.007**	–0.150	–0.009	–0.008	1.7	1.8	1.1
LOCL	**0.021**	**0.002**	0.004	–0.079	–0.006	–0.004	3.8	3.0	1.0
RAMP	**0.201**	**0.009**	**0.011**	–0.266	–0.013	–0.012	1.3	1.4	1.1
ST01	**0.203**	**0.020**	0.009	–0.256	–0.024	–0.009	1.3	1.2	1.0
WIDE	**0.004**	**0.004**	0.005	–0.124	–0.008	–0.005	31.0	2.0	1.0
Weighted average	*0.104*	*0.005*	*0.005*	*–0.167*	*–0.0091*	*–0.007*	*1.6*	*1.8*	*1.4*

*Smallest value for each comparison is **bolded.**

TABLE 11 Mean Prediction Error:* Observed vs. Predicted Using MEASURE and Observed vs. Predicted Using MOBILE5a, by Technology Class

MEASURE									MOBILE5a								
CO			HC			NO_x			CO			HC			NO_x		
Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error	Tech Class	n**	error
3	16	–0.109	32	16	**0.011**	4	16	**–0.001**	3	16	**–0.019**	32	16	–0.012	4	16	0.003
6	16	**0.045**	34	16	**0.007**	5	161	0.007	6	16	–0.102	34	16	–0.008	5	161	**–0.006**
11	32	**0.120**	38	16	–0.008	6	556	**0.004**	11	32	–0.141	38	16	**–0.001**	6	556	–0.006
14	190	**0.051**	51	16	**0.011**	7	48	**0.006**	14	190	–0.077	51	16	–0.012	7	48	–0.013
19	16	**0.118**	54	16	0.012	8	16	**0.009**	19	16	–0.129	54	16	–0.012	8	16	–0.035
20	32	**0.221**	77	304	**0.005**				20	32	–0.230	77	304	–0.006			
21	64	**–0.029**	80	191	**0.001**				21	64	–0.072	80	191	–0.003			
22	112	**0.050**	84	79	**–0.000**				22	112	–0.064	84	79	–0.005			
23	176	**0.028**	95	64	**0.003**				23	176	–0.036	95	64	–0.005			
27	16	**0.109**	96	16	**0.021**				27	16	–0.123	96	16	–0.038			
33	15	**0.575**	97	15	**0.038**				33	15	–0.930	97	15	–0.042			
36	32	**0.318**	108	16	**0.001**				36	32	–0.557	108	16	–0.025			
39	16	**0.382**	112	32	**0.036**				39	16	–0.588	112	32	–0.058			
40	16	**1.255**							40	16	–1.303						
41	32	**0.244**							41	32	–0.480						
42	16	**0.081**							42	16	–0.338						
		0.108			*0.005*			*–0.005*			*–0.167*			*–0.009*			*–0.007*

*Smallest value for each comparison is **bolded**.
** n is the number of cases in each technology class of the validation data set.

TABLE 12 Absolute Ratios of Prediction Error of MOBILE5a Results to MEASURE Results

CO MOBILE5a/MEASURE error		HC MOBILE5a/MEASURE error		NO_x MOBILE5a/MEASURE error	
Tech Class I.D.	Absolute ratio values	Tech Class I.D.	Absolute ratio values	Tech Class I.D.	Absolute ratio values
3	0.2	32	1.1	4	3.0
6	2.3	34	1.1	5	0.9
11	1.2	38	0.13	6	1.5
14	1.5	51	1.1	7	2.2
19	1.1	54	1.0	8	3.9
20	1.1	77	1.2		
21	2.5	80	3.0		
22	1.3	84	*		
23	1.3	95	1.7		
27	1.1	96	1.8		
33	1.6	97	1.1		
36	1.8	108	25.0		
39	1.5	112	1.6		
40	1.04				
41	2.0				
42	4.2				
Weighted average	*1.6*		*1.8*		*1.4*

* The ratio in this cell is a number divided by zero, which is undefined.

FIGURE 1 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Carbon Monoxide Emission Rates)

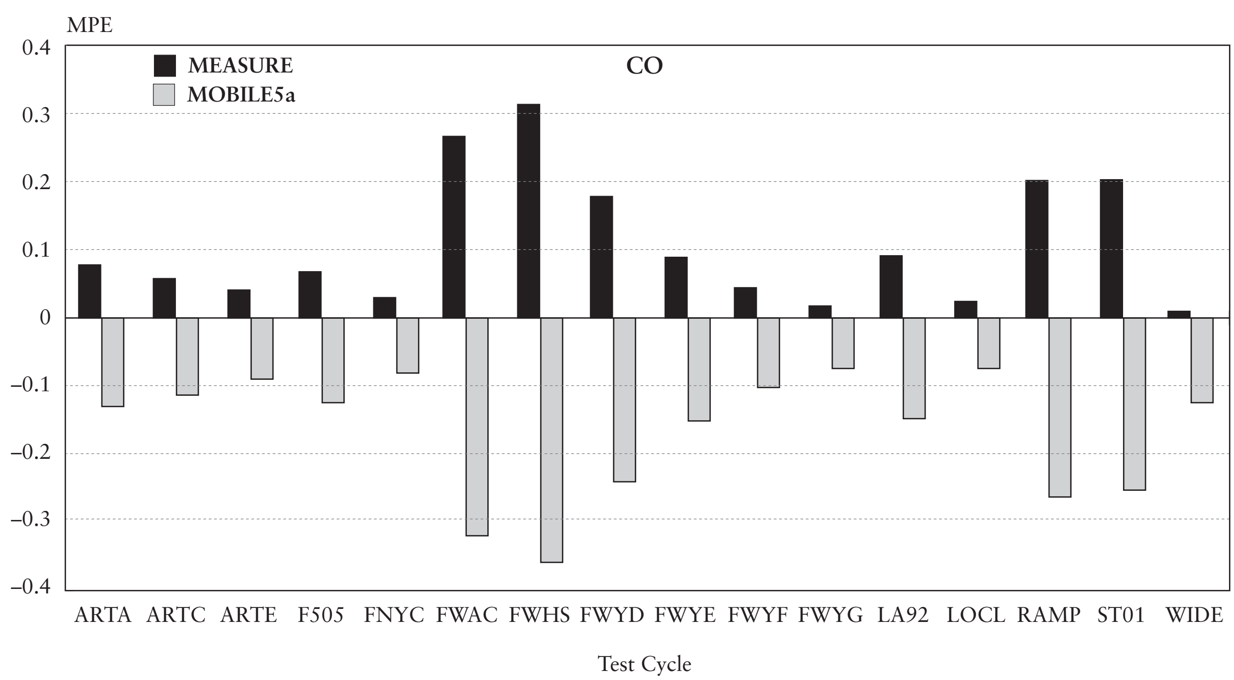


FIGURE 2 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Unburned Hydrocarbons Emission Rates)

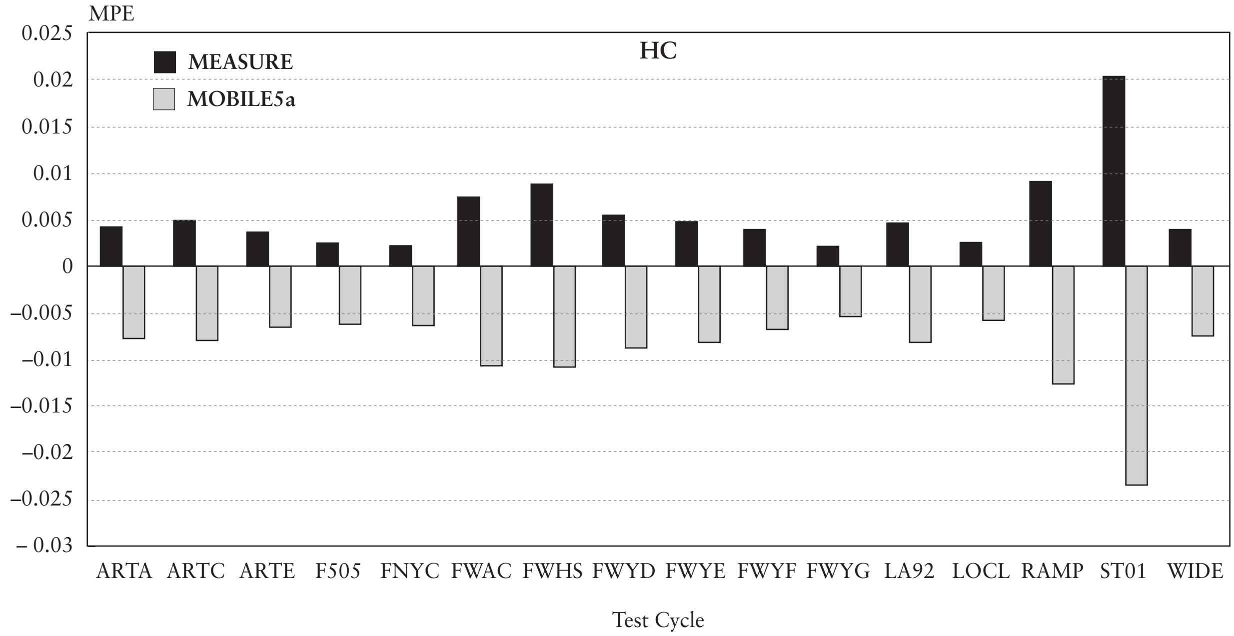


FIGURE 3 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Cycle Basis for Oxides of Nitrogen Emission Rates)

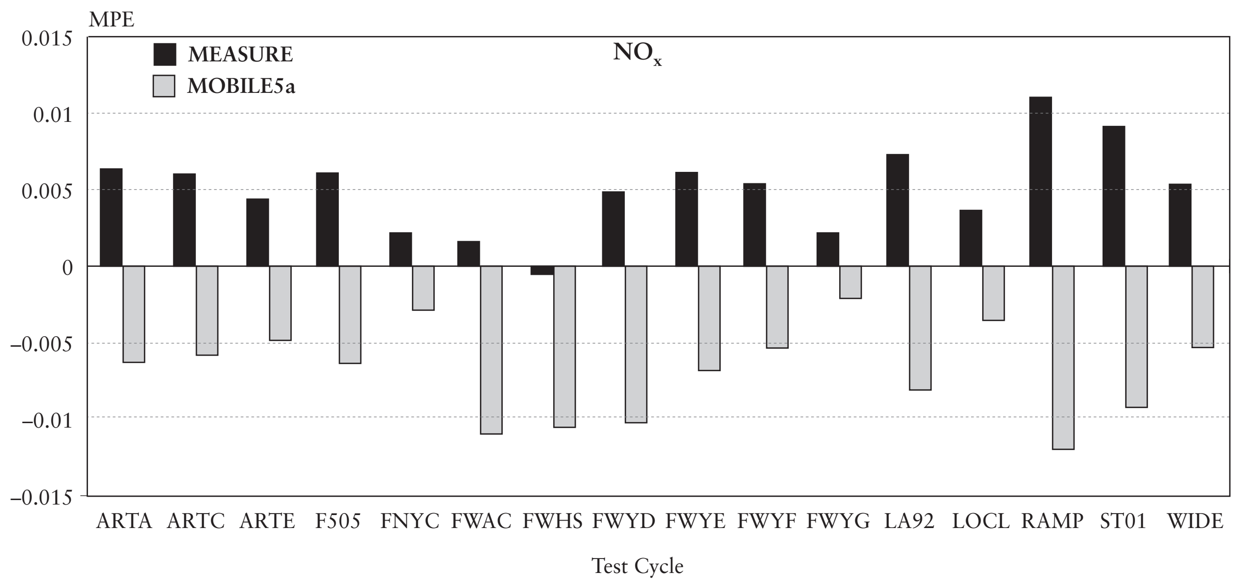
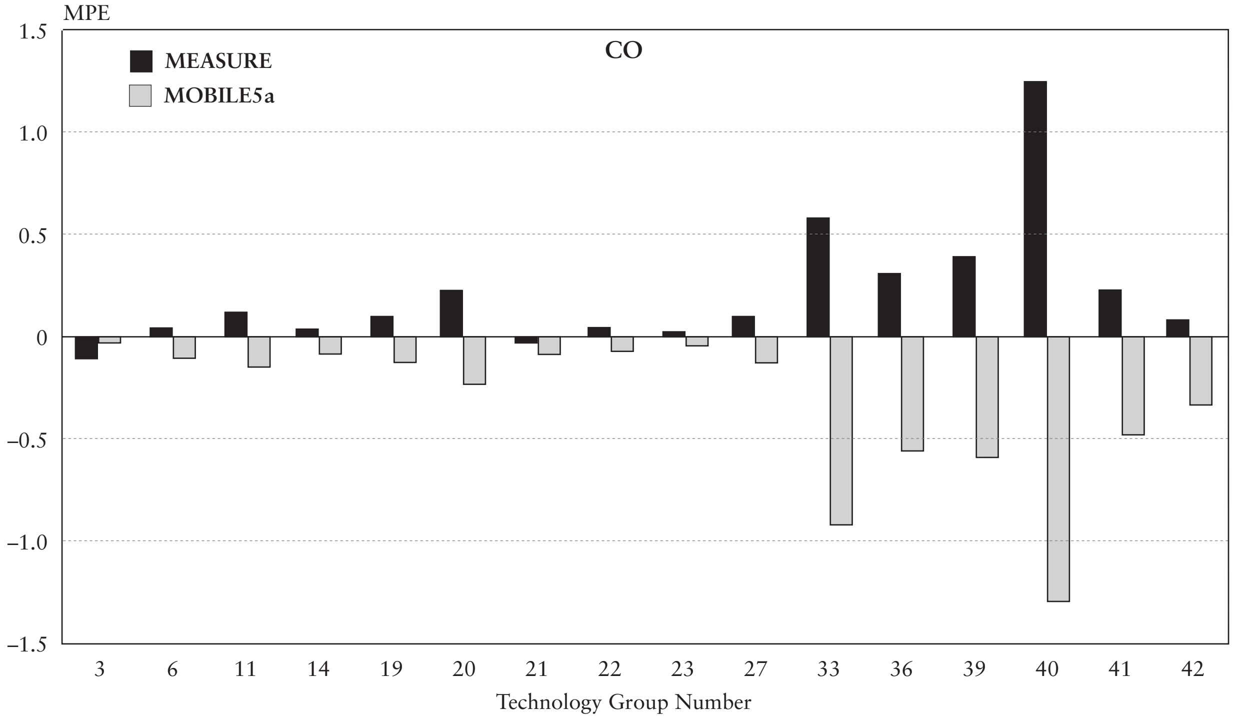


FIGURE 4 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for Carbon Monoxide Emission Rates)



technology classes the degree of overprediction (as measured by the magnitude of the errors) by MEASURE is lower than that of underprediction by MOBILE5a, demonstrating once again by this measure of assessment that MEASURE performs better than MOBILE5a. Pictorial representations of the mean prediction errors on a cycle basis are provided in figures 1 through 3 for CO, HC, and NO_x, respectively, and on a technology class basis in figures 4, 5, and 6, respectively.

CONCLUSIONS

The MEASURE model consistently showed larger correlation coefficients between observed and predicted emissions for the validation data set compared to MOBILE5a. The larger correlation coefficients suggest that the additional modal variables (beyond average speed) and their interactions employed in the MEASURE model provide additional explanatory power. The relatively smaller improvement in NO_x predictions stems from the fact that the average-speed approach to modeling NO_x emissions is not significantly inferior to using improved vehicle activity information; average speed seems to perform quite well for this pollutant.

Some of the driving cycles used in the validation study were designed by EPA contractors to represent on-road driving conditions under varying levels of congestion. Many of these cycles are significantly different from those that were used to develop the MOBILE5a and MEASURE models. The strong performance of the MEASURE model on these new cycles reveals the strength of applying the model to cycles outside those used to develop the model. These findings provide empirical support for the underlying principle that, although the models are cycle-based and aggregate, the discrete contributions of various modal contributions have been well modeled in MEASURE's modeling algorithms and can be used to model the emissions resulting from a variety of "off-cycle" vehicle activities.

In general, the results provided here are encouraging for MEASURE. The general superiority of MEASURE on mean prediction error suggests that if MEASURE and MOBILE5a were applied in practice for forecasting, MEASURE predictions would be more accurate, on average, by a factor of 1.6, based on the validation sample. On the basis of each pollutant, MEASURE would be more accurate

FIGURE 5 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for Hydrocarbon Emission Rates)

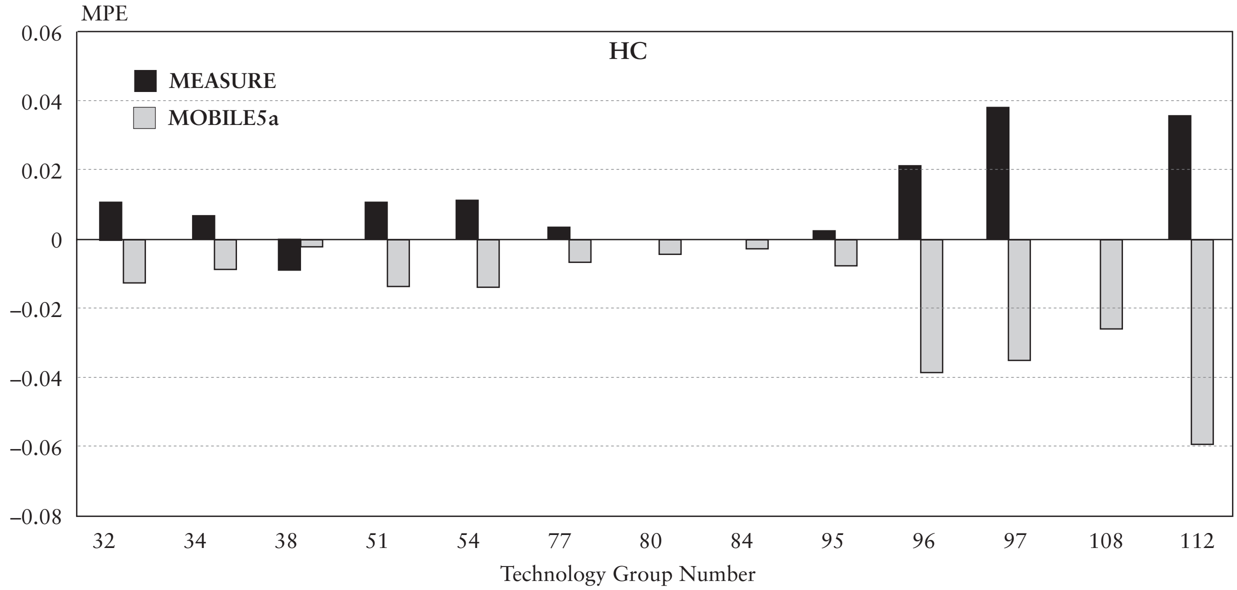
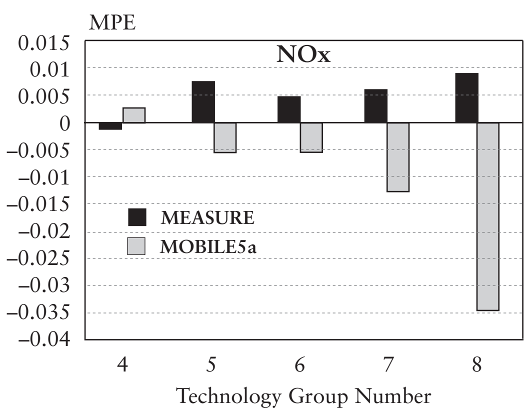


FIGURE 6 Differences in Mean Prediction Error (MPE) Between MEASURE and MOBILE5a (Predictions Performed on a Technology Group Basis for NOx Emission Rates)



by a factor of 1.6 for CO, 1.8 for HC, and 1.4 for NO_x. These factors are shown in underlined italics in the last column of table 10 and at the bottom of table 12. This is a compelling reason to favor MEASURE over MOBILE5a since systematic errors in emission rates will in practice be multiplied by the number of vehicles in an urban area and then again by the amount of mileage driven on a "typical day." MEASURE does slightly overpredict emissions for the validation sample, but this is not a significant concern since MEASURE would also slightly overpredict emissions reductions likely to be garnered from proposed control strategies. Thus, there is no expected major impact from using the model for control strategy modeling (i.e., as a comparative tool across control strategies and time).

Furthermore, the data used to develop MEASURE contained very few test results from 1994 and later model year vehicles. When new data from laboratory studies, such as the University of California, Riverside study by Barth et al. (1997), are included in the data set and the MEASURE algorithms are re-derived, the authors expect further improved performance in applications to the modern vehicle fleet.

REFERENCES

An, F., M. Barth, J. Norbeck, and M. Ross. 1997. The Development of a Comprehensive Modal Emissions Model: Operating Under Hot-Stabilized Conditions. *Transportation Research Record* 1587:52–62.

Barth, M., T. Younglove, T. Wenzel, G. Scora, F. An, M. Ross, and J. Norbeck. 1997. Analysis of Modal Emissions from a Diverse In-Use Vehicle Fleet. *Transportation Research Record* 1587:73–84.

Bishop, G.A., D.H. Stedman, and L. Ashbaugh. 1996. Motor Vehicle Emissions Variability. *Journal of the Air & Waste Management Association* 46:667–75.

Brieman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wardsworth International Group.

Cadle, S.H., R.A. Gorse, and T.C. Belian. Real-World Vehicle Emissions: A Summary. Paper presented at the Seventh CRC On-Road Vehicle Emissions Workshop, San Diego, CA, April 1997.

Cicero-Fernandez, P., J. Long, and A. Winer. 1997a. Effects of Grades and Other Loads on On-Road Emissions of Hydrocarbons and Carbon Monoxide. *Journal of the Air & Waste Management Association* 47.

Cicero-Fernandez, P., W. Wong, and J. Long. 1997b. Fixed Point Mobile Source Emissions Due to Terrain Related Effects: A Preliminary Assessment, in *Proceedings of the 90th Annual Meeting of the AWMA*, Pittsburgh, PA, June.

Fomunung, I., S. Washington, and R. Guensler. 1999. A Statistical Model for Estimating Oxides of Nitrogen Emissions from Light-Duty Motor Vehicles. *Transportation Research D: Transport and Environment* 4/5:333–52.

Fomunung, I. 2000. Predicting emissions rates for the Atlanta on-road light-duty vehicular fleet as a function of operating modes, control technologies, and engine characteristics. Ph.D. diss., Georgia Institute of Technology.

Guensler, R. 1993. Data Needs for Evolving Motor Vehicle Emission Modeling Approaches. *Transportation Planning and Air Quality II*. New York, NY: American Society of Civil Engineers.

Guensler, R., S. Washington, and W. Bachman. 1998. Overview of the MEASURE Modeling Framework. *Transportation Planning and Air Quality III, Emerging Strategies and Working Solutions*. New York, NY: American Society of Civil Engineers.

Heywood, J.B. 1988. *Internal Combustion Engine Fundamentals*. New York, NY: McGraw-Hill.

Jimenez-Palacio, J.L. 1999. Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing. Ph.D. diss., Massachusetts Institute of Technology.

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. *Applied Linear Statistical Models,* 4th edition. Boston, MA: Richard D. Irwin, Inc.

Rodgers, M.O. 1995. An Improved Power Demand Model for Evaluation of On-Road Vehicle Emissions. The Emission Inventory: Programs and Progress, EPA/AWMA. October.

Shih, R., S. Fable, and R.F. Sawyer. 1997. Effects of Driving Behavior on Automotive Emissions, in *Proceedings of the Seventh CRC On-Road Vehicle Emissions Workshop, Coordinating Research Council,* Atlanta, GA.

Washington, S. 1994. Estimation of a vehicular carbon monoxide modal emissions model and assessment of an intelligent transportation technology. Ph.D. diss., University of California at Davis.

Webster, W.J. and C. Shih. 1996. A Statistically-Derived Metric to Monitor Time-Speed Variability in Practical Emissions Testing, in *Proceedings of the Sixth CRC On-Road Vehicle Emissions Workshop,* San Diego, CA, 18–28 March, Atlanta, GA: Coordinating Research Council.

Wolf, J., R. Guensler, S. Washington, and W. Bachman. 1998. High-Emitting Vehicle Characteristics Using Regression Tree Analysis. *Transportation Research Record* no. 1641:58–65.