Statistical Assessment of Vehicular Carbon Monoxide Emission Prediction Algorithms

SIMON WASHINGTON AND RANDALL GUENSLER

Increased concern about the ability to accurately model and predict emissions from motor vehicles prompted this research. The ability of the mathematical algorithms contained in version 4 of the CALINE line source dispersion model (CALINE4) developed by Caltrans to accurately predict carbon monoxide (CO) emissions from a fleet of motor vehicles is assessed. The CALINE4 model contains algorithms that predict CO emissions from discrete modal events of idle, cruise, acceleration, and deceleration. A BASIC computer program is used to assess and compare the performance of the CALINE4 algorithms with those incorporated in version 7F of the EMFAC model (EMFAC7F), which is used and developed by the California Air Resources Board. The statistical assessment includes comparisons of mean prediction bias, Theil's U-Statistic, and the linear correlation coefficient. The analyses demonstrate that the currently used CALINE4 algorithms perform similarly to those contained in EMFAC7F, but when modified to use individual emission rates (instead of fleet average emission rates), the CALINE4 algorithms generally outperform the EMFAC7F algorithms. For short- to medium-term microscale model improvements, it is recommended that the CALINE4 model be revised to (a) incorporate individual emission rates into its emission estimation algorithms, (b) update its statistically derived model coefficients, and (c) update the modal activity algorithms to cover all modeling scenarios. For long-term modeling improvements, it is recommended that a more robust modal model be estimated based on second-by-second data and additional causal variables, and true vehicle simulation models be used to estimate vehicle activity.

The Clean Air Act (CAA) requires metropolitan regions in nonattainment with National Ambient Air Quality Standards (NAAQS) for carbon monoxide (CO) to demonstrate timely reductions in regional CO emission inventories and zero increases in CO hot spots for project level air quality impact analyses (*1*). When regional and local planners are faced with making transportation growth and investment decisions, they are constrained to select only those projects that will decrease CO emission inventories and the severity and number of CO hot spots. Because these decisions often involve millions and sometimes billions of local, state, and federal dollars, there is a need for planners to conduct accurate, precise, and meaningful analyses. Accurate and precise estimates of CO inventories and CO hot spot impacts require statistically and theoretically robust CO estimation algorithms. The model California is using to estimate CO emission inventories is EMFAC7F-BURDEN, developed by the California Air Resources Board (CARB). A project-level CO impact analysis model commonly used in California is CALINE4, developed by the California Department of Transportation (Caltrans).

There are important differences between EMFAC and CALINE. First, CALINE is primarily a pollutant dispersion model used to estimate the CO impacts of transportation projects—it is intended for microscale applications. EMFAC, on the other hand, is an emission inventory model (when coupled with BURDEN and regional motor vehicle activity data) and estimates inventories of CO, nitrogen oxides (NO_x), and hydrocarbons (HC). In practice, EMFAC and CALINE emissions predictions are never directly compared because they operate under entirely different frameworks. The common thread between these two models, and the central focus of this paper, is that the EMFAC and CALINE models contain mathematical algorithms that predict CO emissions given vehicle activity estimates.

This paper examines how well these algorithms predict CO emissions. There are two motivations for this research. First, concern over the ability of CARB's EMFAC7F emissions model to estimate accurately modal emission inventories from motor vehicles (*2,3,4,5,6*) has prompted the need to statistically quantify the performance of the mathematical algorithms. In addition, the CO emission prediction algorithms contained in the CALINE4 model use modal correction factors, which correct the baseline emission rate employed in EMFAC (or MOBILE) based on estimates of acceleration, deceleration, cruise, and idle activity. This is in contrast to the EMFAC model algorithms, which use speed correction factors to correct baseline emission rates based on average speed estimates.

Second, the CO-emission prediction algorithms that are embedded in the CALINE4 model and that are used when the intersection option portion of the model is selected have never been statistically verified using real emissions data. Because Caltrans uses these models to perform project-level CO emission analyses, Caltrans staff wanted to verify the CO emission prediction algorithms by comparing their predictions to those predicted by the EMFAC model—expecting that the CO-emission predictions between models would be consistent.

This paper presents the results of a technical and statistical assessment of the ability of the CALINE4 and EMFAC7F mathematical algorithms to adequately predict measured CO emission rates from motor vehicles tested on numerous laboratory testing cycles. The algorithms are dissected to determine where prediction errors are likely to originate and how the algorithms could be improved. The statistical assessment uses measures of performance such as mean prediction bias, Theil's U-Statistic, and the linear correlation coefficient to compare predicted CO emissions with measured CO emissions. These performance tests use the CO emission test results from 14 standardized testing cycles (*2*). To aid in the analyses, a BASIC computer program was developed to reproduce the internal algorithms for both the CALINE4 and EMFAC7F models (*7*).

S. Washington, Institute of Transportation Studies, University of California at Davis, Davis, Calif. 95617. R. Guensler, Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Ga. 30332–0355

DESCRIPTION OF STANDARDIZED TESTING CYCLES USED TO ASSESS CALINE4 AND EMFAC7F

Summary statistics on the standardized testing cycles contained in the speed correction factor (SCF) data set are shown in Table 1. The table shows some of the pertinent characteristics unique to each the cycle, such as the distance of the test cycle and the duration of the cycle. The table also depicts modal attributes of each cycle. For example, almost half of Low Speed Test Cycle #3 is spent with vehicles in the idle mode of operation. The age of the vehicles tested on these cycles ranged from 1977 model years to 1990 model years. About half of the cycles have emission results for 464 vehicle tests, while a couple of the tests (High Speed Test Cycles #1 and #2) have only 25 vehicle test results. These emissions test results represent the current data set used by CARB and the Environmental Protection Agency (EPA) to develop their emission factor models, EMFAC7F and MOBILE5A respectively. The data are currently the most comprehensive and quality-controlled U.S. emissions data available over a variety of testing cycles.

The modal activity data shown in the table represent the percentage of time the total cycle spent in a particular mode of operation. For example, 21 percent of the 505 sec of the Federal Test Procedure (FTP) Bag 1 is spent with the vehicles accelerating. Acceleration (A) conditions are defined as increases in velocity (V) over two consecutive seconds of operation and can last any number of seconds. Derivation of deceleration ($-A$), idle ($A = 0$, $V = 0$), and cruise operations ($A = 0$, $V \neq 0$) are derived in a similar straightforward manner.

The 14 test cycles shown in the table represent unique profiles of modal activity. Each cycle was developed to approximate driving behavior under different conditions. For example, the New York City cycle approximates driving in New York City, which is characterized by a low mean speed (11.4 km/hr) and a lot of modal activity. The FTP Bags 1 through 3 are important fundamental components of the MOBILE and EMFAC emissions models. In the EMFAC7F model, for example, FTP Bag 2 test results are used as the base emission rate, which are then "speed corrected" to derive emissions at average speeds other than 16 mph (the FTP Bag 2 cycle average speed). FTP Bags 1 and 3 contain emission contributions from cold and hot starts, respectively, and FTP Bag 2 contains hot stabilized emissions only. The importance of these test cycles is evident later in the paper.

THEORETICAL BASIS OF EMISSION PREDICTION ALGORITHMS

CALINE4 Line Source Dispersion Model

The CALINE4 line source dispersion model has been developed over many years by Caltrans. The CALINE4 line model estimates CO, NO_x, and suspended particulate concentrations. It uses the Gaussian diffusion equation to distribute air pollution over and along modeled roadways (*8*). EPA has approved the model as a tool to assess impacts from CO hot spots. The model is used primarily for local project analyses in areas where its use was established before July 1993. The model contains algorithms that estimate CO emission contributions from modal events of idle, acceleration, deceleration, and cruise, and therefore contains a modal emissions model component. This modal emission model component is used only when the intersection link option is used when running the CALINE4 model; otherwise for main line sections, MOBILE- or EMFAC-derived average emission rates are used.

CALINE4 has undergone three revisions since the original version in 1972, and it uses the Gaussian dispersion equation to distribute estimated emissions along a roadway. When the intersection link option is used, CO emissions are estimated on a modal basis; that is, equations or algorithms are used to predict CO emissions from each modal event (idle, cruise, acceleration, and deceleration). Information required to derive the modal events are intersection specific and require information about acceleration and deceleration times (from link endpoint to intersection stopline), minimum and maximum idle times, traffic volumes, and the number of vehicles delayed. The program uses these inputs to generate the modal activity occurring at an intersection (*8*). When the intersection option is not chosen, CO emission predictions are based on the speed-corrected baseline emission rates provided by EMFAC or MOBILE (*8*).

TABLE 1 Summary Statistics on Standardized Testing Cycles Used in Analyses

CYCLE NAME	TIME (sec)	DIST. (km)	MEAN SPEED (kph)	% IDLE	% ACCEL.	% DECEL.	% CRUISE
Federal Test Procedure - Bag 1	505	6.65	41.2	19.6	21.0	20.4	39.0
Federal Test Procedure - Bag 2	866	7.15	25.8	18.6	25.3	19.3	36.8
Federal Test Procedure - Bag 3	505	6.65	41.2	19.6	21.0	20.4	39.0
Highway Fuel Economy Test	765	19.00	77.7	0.7	14.1	11.8	73.4
High Speed Test Cycle # 1	474	10.98	72.5	1.1	13.3	9.9	75.7
High Speed Test Cycle # 2	480	12.59	82.1	1.0	13.8	10.4	74.8
High Speed Test Cycle # 3	486	14.44	93.0	1.0	14.2	10.9	73.9
High Speed Test Cycle # 4	492	16.32	103.7	1.0	15.3	11.4	72.3
Low Speed Test Cycle #1	624	1.30	6.5	36.5	24.2	25.6	13.7
Low Speed Test Cycle #2	637	1.18	5.9	38.8	23.4	24.3	13.5
Low Speed Test Cycle #3	616	0.96	3.9	47.7	16.2	17.9	18.2
New York City Cycle	598	2.18	11.4	34.9	23.9	24.2	17.0
Speed Cycle 12	349	2.17	19.4	27.2	26.1	24.1	22.6
Speed Cycle 36	996	18.37	57.7	6.5	19.0	16.0	58.5

The latest version of the algorithms used in the CALINE4 model is similar to those in the Colorado Department of Highways (CDOH) model released in 1980. The data used to estimate the CDOH models were derived from 37 discrete modes driven by 1,020 light-duty vehicles ranging from 1957 model year to 1971 model year (*9*). A subset of 62 vehicles representing California cars for model years 1975 and 1976 was used to estimate the coefficients employed in the CALINE4 algorithms (*8*). In the Caltrans and CDOH model development efforts, the modal acceleration speed (AS) product demonstrated good explanatory power for CO emissions estimation. Consequently, the AS product, defined as the product of the average acceleration and average speed for the acceleration event, is one of the explanatory variables used in the CALINE4 model (*8*). For a more detailed description of the CALINE4 model, see the work by Benson (*8*).

The CALINE4 modal emission algorithms can be written as

$$TE_{ik} = EI_{ik} + EA_{ik} + EC_{ik} + ED_{ik} \quad (1)$$

where

- TE_{ik} = total CO emission estimate for vehicle i on cycle k in grams,
- EI_{ik} = CO emissions from idle events for vehicle i on cycle k in grams,
- EA_{ik} = CO emissions from acceleration events for vehicle i on cycle k in grams,
- EC_{ik} = CO emissions from cruise events for vehicle i on cycle k in grams, and
- ED_{ik} = CO emissions from deceleration events for vehicle i on cycle k in grams.

The emission contributions from the discrete modal events can be defined as

$$EI_{ik} = (IR_{[\mathrm{g/sec}]}) * (t_{i[\mathrm{sec}]}) \quad (2)$$

where IR is measured idle emission rate and t_i is time spent in the idle operating mode.

$$EA_{ik} = [(FTPB2_{[\mathrm{g/min}]}) * (C1) * EXP\,(C2 * AS)] * t_{a\,[\mathrm{sec}]} * 1_{[\mathrm{min}]}/60_{[\mathrm{sec}]} \quad (3)$$

where

- $FTPB2$ = measured emission rate on FTP Bag 2,
- Coefficients $C1$ = 0.75 and $C2$ = 0.0454 for acceleration condition 1,
- Coefficients $C1$ = 0.027 and $C2$ = 0.098 for acceleration condition 2,
- Acceleration condition 1 is for vehicles starting at rest and accelerating up to 45 mph (72.42 km/hr),
- Acceleration condition 2 is for vehicles starting at 15 mph (24.14 km/hr) or greater and accelerating up to 60 mph (96.56 km/hr),
- AS = acceleration speed product based on average speed and average acceleration rate of the accel mode in mi²/hr²/sec, and
- t_a = time spent in the acceleration mode.

$$EC_{ik} = (FTPB2_{[\mathrm{g/min}]}) * [(0.494 + 0.000227 * S_{[\mathrm{km/hr}]})^2] * (t_{c\,[\mathrm{sec}]} * 1_{[\mathrm{min}]}/60_{[\mathrm{sec}]}) \quad (4)$$

where

- $FTPB2$ = measured emission rate on FTP Bag 2,
- t_c = time spent in the cruise event, and
- S = average speed of the vehicle in the modal event in mph.

$$ED_{ik} = (IR_{[\mathrm{g/sec}]}) * (t_{d[\mathrm{sec}]}) * 1.5 \quad (5)$$

where IR is measured idle emission rate and t_d is time spent in the deceleration operating mode.

It is critical to note that the FTP Bag 2 emission rate and the IDLE emission rate used in the CALINE4 model program are estimated average values for the on-road fleet of motor vehicles. The CALINE4 algorithms do not contain emission factors that differentiate between technology groups or model year. The result is that the CALINE4 modal emission prediction algorithms predict equivalent modal contributions of emissions for all modeled vehicles, the average emission rate over a given driving segment. In other words, a 1980 Cadillac Seville is predicted to emit the same as a 1993 Geo Metro, the Seville's emissions being underpredicted and the Metro's emissions being overpredicted using the fleet mean value.

EMFAC7F—California Regional Emissions Model

The EMFAC7F emissions model developed by CARB operates differently from CALINE4. Instead of taking a modal approach, EMFAC7F uses average operating speed and fuel delivery technology and model year as explanatory variables in the model. For each of four technology group classifications based on fuel delivery technology and model year, EMFAC7F predicts a modal emission ratio, based on the ratio of emissions on the FTP to emissions at other cycle average speeds. The resultant ratios are called SCFs and are used to estimate emissions at speeds other than 16 mph (at 16 mph measured emissions are predicted). For a complete description of the operating characteristics and analyses of the recent EMFAC7F model, see the work by Guensler (*2*).

The regression form of the EMFAC7F model for predicting CO emissions can be written as

$$TE_{mn} = \{BAG2_n * [EXP\,(\mathbf{B1}_n * SADJ1) + (\mathbf{B2}_n * SADJ2) + (\mathbf{B3}_n * SADJ3) + (\mathbf{B4}_n * SADJ4)]\} + \mathrm{error} \quad (6)$$

where

- TE_{mn} = total CO emissions for vehicle m from technology group n,
- $BAG2_n$ = average measured Bag 2 result for technology group n vehicles,
- $SADJ1$ = (16 − average prediction speed),
- $SADJ2$ = (16 − average prediction speed)²,
- $SADJ3$ = (16 − average prediction speed)³,
- $SADJ4$ = (16 − average prediction speed)⁴,
- $\mathbf{B1}_n, \ldots, \mathbf{B4}_n$ = least squares estimated coefficients, and
- error = the disturbance term.

As noted previously, four models are estimated based on CARB-defined technology groups. The technology groups are shown in Table 2.

Also, somewhat similar to CALINE4, EMFAC7F derives an average emission factor, in grams per mile, for an entire fleet of on-road vehicles. This average emission factor will result in over-

TABLE 2 Technology Groups Used in EMFAC7F SCF Model

CARB Technology Group	Model Year	Fuel Delivery Technology
1	1985 or earlier	Carbureted and Throttle Body Injection
2	1985 or earlier	Port Fuel Injection
3	1986 or later	Carbureted and Throttle Body Injection
4	1986 or later	Port Fuel Injection

prediction of emissions for "clean" vehicles and under-prediction of emissions for "dirty" vehicles.

CARB's model has been criticized for statistical and theoretical reasons. Among the statistical criticisms are non-normal error distributions, high multicollinearity among the explanatory variables, biased parameter estimates, and wide confidence intervals around the SCF curves (*2*). The theoretical criticisms are primarily concerned with the exclusion of causal explanatory variables, a nonrepresentative sample fleet of vehicles, and nonrepresentativeness of driving cycles as compared with real driving behavior.

MODEL PERFORMANCE EVALUATION

This research compares the ability of CALINE4 and EMFAC7F co-emission prediction algorithms to adequately predict measured emissions from a standardized and large data set. Using the SCF data base as the validation data set, the ability of both models to accurately predict measured emissions obtained from vehicles under numerous testing cycles is assessed. EMFAC7F has a slight advantage over CALINE4 because its emission algorithms were estimated using the SCF data set, while the CALINE4 model algorithms were estimated using a subset of the CDOH data set discussed earlier.

In practice, the CALINE4 and EMFAC7F model algorithms operate using fleet average FTP Bag 2 and idle rates (the CALINE4 user inputs values derived from EMFAC7F or MOBILE). In other words, the emission input data are aggregate data for a fleet of vehicles. For the analyses presented here, aggregate Bag 2 and idle rates were obtained by computing the average values for these variables for all vehicles contained in the SCF data set.

Using individual vehicle Bag 2 and idle rates, on the other hand, represents a significant modification to the way in which the model algorithms are used. Using individual vehicle test results, or disaggregate data, the model algorithms are allowed to capture the effect of algorithm prediction differences between vehicles, a degree of prediction flexibility not possible when aggregate data are used. Disaggregate analyses are performed here to investigate algorithm improvement possibilities.

Comparison of Mean Predicted Emissions

A desirable emission prediction algorithm will not be biased in its prediction of CO emissions. One indicator of bias in a model is the difference between true average emissions and estimated average emissions. Ideally, the mean value of the predicted emissions should be the same as the mean value of actual emissions. This is especially true when considering the current application of CALINE4 and EMFAC7F, which operate using average fleet emission rates. A great discrepancy in means over a large sample suggests that the model is consistently over- or underpredicting the actual emissions and that model predictions are biased.

To quantify biases for the CALINE4 and EMFAC model algorithms, estimated emissions were summed over a test cycle and then averaged according to the number of vehicles in the test cycle. As an example, the predicted emission estimates for vehicles tested on High Speed Cycle #2 are summed and then divided by 25 vehicles to compute the average emission estimate. The average emission estimate is then compared with the average observed emission result for the vehicles tested on that cycle to determine the mean bias. The formula for mean bias is given by

$$MPB_j = \left(\textstyle\sum_i Y_{ij} - \sum_i \Psi_{ij}\right)/n_j \qquad (7)$$

where

MPB_j = mean prediction bias on all vehicles on cycle j,
$\sum_i$ = summation over i vehicles,
Y_{ij} = predicted emissions for vehicle i on cycle j in grams,
Ψ_{ij} = observed emissions for vehicle i on cycle j in grams, and
n_j = number of vehicles tested on cycle j.

The mean prediction bias for the CALINE4 and EMFAC emission prediction algorithms is shown in Table 3. The table shows both the aggregate and the disaggregate model assessments. Disaggregate refers to the use of individual vehicle Bag 2 emission rates in model emission prediction algorithms, and aggregate refers to the use of average fleet Bag 2 emission rates, which is consistent with the manner in which the model algorithms are used in practice.

Note that EMFAC7F CO-emission prediction algorithms generally outperform the CALINE4 algorithms when mean prediction bias comparisons are studied. This suggests that when average fleet emission rates (aggregate) are used, EMFAC7F algorithms in general perform better than CALINE4 algorithms. (Recall, however, that model results are never compared in practice because of their distinctly different purposes.) The difference is not drastic, however, and the CALINE4 model algorithms have less prediction bias on several cycles. Furthermore, EMFAC7F CO-emission prediction algorithms have smaller biases for disaggregate model analyses. These findings are not surprising, because the coefficients embedded in the CALINE4 model were estimated using an older fleet of vehicles, whereas EMFAC7F coefficients were estimated using the SCF data base vehicle test results. Because the comparison provides an unfair advantage to the EMFAC7F model, whether the modal components embedded in the CALINE4 model are performing well cannot be adequately assessed. It can be speculated, however, that given the comparisons depicted in the table, if the coefficients in the CALINE4 emission model were updated using the SCF data base vehicles, CALINE4 would likely outperform EMFAC7F.

CALINE4 and EMFAC7F model algorithms generally underpredict CO on the low-speed test cycles. CALINE4 underpredicts on the two highest high-speed cycles and overpredicts on the two lowest high-speed cycles. EMFAC7F tends to overpredict on all high-

TABLE 3 Comparison of Mean Model Prediction Bias (in grams)

Cycle	Mean Aggregate CO[a]	MPB Dis-aggregate CALINE4	MPB Dis-aggregate EMFAC7F	MPB Aggregate CALINE4	MPB Aggregate EMFAC7F
Highway Fuel Economy Test	51.40	**-6.2**	-3.3	**-6.2**	**-2.9**
High Speed Test Cycle # 1	4.24	**2.4**	3.9	**2.4**	4.0
High Speed Test Cycle # 2	4.55	**3.9**	5.4	4.0	5.4
High Speed Test Cycle # 3	11.60	-2.4	**2.2**	-2.3	2.3
High Speed Test Cycle # 4	38.26	-25.3	**12.1**	-25.3	13.7
Low Speed Test Cycle #1	24.99	-9.5	-3.2	-9.9	**-3.0**
Low Speed Test Cycle #2	24.47	-9.8	-2.5	-10.3	**-2.4**
Low Speed Test Cycle #3	22.34	-8.6	-3.3	-8.8	**-3.2**
New York City Cycle	29.20	-0.9	**0.3**	-2.0	0.8
Speed Correction Factor 12	16.65	**0.0**	-0.3	**0.0**	-0.2
Speed Correction Factor 36	63.68	**2.3**	4.6	3.0	2.4

Bold = Smallest absolute mean bias in emission estimate in Dis-aggregate or Aggregate comparison
[a] Mean aggregate CO determined by computing the arithmetic mean of Bag 2 test results of vehicles in SCF data base

speed cycles. It is important to note that model algorithms are not performing consistently across testing cycles with varying characteristics. This suggests that there still may be cycle-related variables not included in model algorithms that may help to explain these emission variations.

Theil's U-Statistic Comparisons

A proposed measure of model performance that is not subject to the scaling problems of the previous measure is Theil's U-Statistic (*10,11*). Theil's U-Statistic is related to R-Square but is not bounded by 0 and 1. Large numbers of U reflect poor fit to the data, and small values of U indicate good fit. The U-Statistic formula is given by

$$U_j = \{[(1/n_j) \textstyle\sum_i(\Psi_{ij} - Y_{ij})2] / [(1/n_j) \textstyle\sum_i(\Psi_{ij}) 2]\}^{0.5} \tag{8}$$

where

U_j = Theil's U-Statistic for all vehicles on cycle j,
$\sum_i$ = summation over i vehicles on cycle j,
Y_{ij} = predicted emissions for vehicle i on cycle j in grams,
Ψ_{ij} = observed emissions for vehicle i on cycle j in grams, and
n_j = number of vehicles tested on cycle j.

Theil's U-Statistic results are shown in Table 4. The table shows that, for disaggregate comparisons, Theil's U is consistently smaller for CALINE4 than for EMFAC7F. Under aggregate model applications, however, the emission prediction algorithms in the EMFAC7F model are superior. These results suggest that the CALINE4 model works well when the values input for IDLE and FTP Bag 2 are allowed to vary simultaneously with vehicles but, when constrained to fleet average values, it is no better than EMFAC7F. In fact, because algorithm coefficients for CALINE4 were derived from a much older fleet, the EMFAC7F model performs better under contemporary model applications.

Linear Correlation Coefficient Comparisons

As a final useful statistical comparison of the two models, the linear correlation coefficient is used (*12,13*). The linear correlation coefficient reflects the degree of probability that a linear relationship exists between observed and predicted emissions. If a model can predict observed emissions well, then the linear correlation is expected to be high, whereas if a model predicts poorly, the linear correlation coefficient will be low. The formula for the correlation coefficient is given by

$$r_j = \textstyle\sum_i[\Psi_{ij} - Y_j(\text{ave})][Y_{ij} - Y_j(\text{ave})] / \{\textstyle\sum_i[\Psi_{ij} - Y_j(\text{ave})]2 \times \textstyle\sum_i[Y_{ij} - Y_j(\text{ave})]2\}^{0.5} \tag{9}$$

where

r_j = correlation coefficient between observed and predicted emissions for i vehicles on cycle j,
$\sum_i$ = summation over i vehicles on cycle j,
Ψ_{ij} = observed emissions for vehicle i on cycle j in grams,
$Y_j(\text{ave})$ = average observed emissions for all vehicles on cycle j in grams, and
Y_{ij} = predicted emissions for vehicle i on cycle j in grams.

The correlation coefficients for the two emission prediction algorithms are compared in Table 5. The table shows that CALINE4 model algorithms generally outperform EMFAC model algorithms for disaggregate comparisons. Comparisons are not valid under aggregate conditions because the CALINE4 model predicts a constant value; thus the computation of the correlation coefficient yields 0. That the correlation coefficient varies over cycles with characteristically different modal activity suggests that a large proportion of modal activity is not explained by CALINE4's modal algorithms. This finding is magnified when it is considered that a great deal of observed modal activity is not represented in any of the test cycles contained in the SCF data set. For example, the greatest acceleration rate contained in the SCF data set is 3.3 mph/sec, whereas accelerations as high as 8.0 mph/sec have been observed in real world driving.

DISCUSSION OF RESULTS

This research effort has identified modeling deficiencies inherent in the algorithms contained in the CALINE4 and EMFAC7F emissions models. The CALINE4 model is used primarily for project-level analyses and is intended for microscale emission impact

TABLE 4 Comparison of Theil's U-Statistic (in grams)

Cycle Name	Dis-aggregate CALINE4 (Grams)	Dis-aggregate EMFAC7F (Grams)	Aggregate CALINE4 (Grams)	Aggregate EMFAC7F (Grams)
Highway Fuel Economy Test	0.605	**0.537**	0.967	0.966
High Speed Test Cycle # 1	1.107	1.524	**0.760**	0.968
High Speed Test Cycle # 2	1.524	1.930	**0.922**	1.125
High Speed Test Cycle # 3	0.799	0.935	0.752	**0.720**
High Speed Test Cycle # 4	**0.921**	1.054	0.940	0.936
Low Speed Test Cycle # 1	**0.689**	1.019	0.930	0.911
Low Speed Test Cycle # 2	**0.655**	0.964	0.943	0.923
Low Speed Test Cycle # 3	**0.702**	1.035	0.942	0.922
New York City Cycle	**0.389**	0.549	0.919	0.917
Speed Correction Factor 12	**0.413**	0.424	0.932	0.933
Speed Correction Factor 36	**0.533**	0.554	0.952	0.950

Bold = Smallest U-Statistic in emission estimate

assessment. It is often used to determine worst-case CO impact assessments of transportation projects. The CALINE4 model, furthermore, is not used for emission inventory purposes.

EMFAC7F, on the other hand, is primarily used for performing regional analyses. Used with transportation network models (UTPS type models), EMFAC7F estimates CO emission rates applied to activity in the air basin. Although both models are used to satisfy air quality modeling requirements stipulated in the Clean Air Act, their purposes are different. This distinction is important when considering recommendations for improving the models and technical improvements.

Several important deficiencies in the current modeling methodologies were illustrated, including the impact of errors in predicting mean emission rates on regional inventories, the use of fleet averages instead of individual vehicle emission rates, and the lack of causal variables in model formulations.

Statistical comparisons between the two models' algorithms included comparisons of mean prediction bias, Theil's U-Statistic, and the linear correlation coefficients between predicted and observed emissions. The assessment looked at both the aggregate model algorithms using average fleet emission rates and a disaggregate version of the algorithms using individual vehicle emission rates.

When making across-the-board comparisons between aggregate EMFAC7F and CALINE4 algorithms, it can be seen that EMFAC7F performs slightly better on almost all performance measures. This is not surprising, however, because the data set used to compare algorithms was also used to estimate the EMFAC7F algorithms, and the CALINE4 algorithms were estimated using a much older and smaller data set. Considering emissions algorithms using disaggregate data, however, CALINE4 algorithms predict emission rates better than do EMFAC7F algorithms. This difference is attributable to the inclusion of the idle variable in the CALINE4 model,

TABLE 5 Comparison of Correlation Coefficients (*r*)

Cycle Name	r Dis-aggregate CALINE4 (Grams)	r Dis-aggregate EMFAC7F (Grams)	r Aggregate CALINE4 (Grams) [a]	r Aggregate EMFAC7F (Grams) [b]
Highway Fuel Economy Test	0.792	**0.835**	0	0.006
High Speed Test Cycle # 1	**0.843**	0.836	0	0.126
High Speed Test Cycle # 2	**0.786**	0.774	0	0.028
High Speed Test Cycle # 3	0.310	**0.361**	0	0.328
High Speed Test Cycle # 4	0.201	**0.266**	0	0.149
Low Speed Test Cycle # 1	**0.702**	0.635	0	0.128
Low Speed Test Cycle # 2	**0.734**	0.634	0	0.141
Low Speed Test Cycle # 3	**0.684**	0.515	0	0.161
New York City Cycle	**0.911**	0.878	0	0.092
Speed Correction Factor 12	0.900	**0.913**	0	0.102
Speed Correction Factor 36	**0.832**	0.828	0	0.068

Bold = Greatest correlation coefficient between observed and predicted emissions

[a] The correlation coefficient for the CALINE4 model is 0 since the prediction uses the constant FTP Bag2 fleet average rate, the constant fleet average idle rate, and coefficients that are determined by cycle modal characteristics. The result is no variation in emissions predictions within a cycle (see Washington, Guensler, and Sperling, 1994).

[b] The correlation coefficient for EMFAC7F is non-zero since different within-cycle predictions result from the differences brought about by different vehicle technology groupings and their associated unique model coefficients.

which varies independently of the FTP Bag 2 emission rate and therefore captures more of the variation in emissions performance between vehicles.

The use of individual vehicle emission test results in the model algorithms brings about drastic improvements in overall performance of both models' algorithms. This improvement can be seen in Tables 3, 4, and 5, where theoretical modifications using disaggregate data result in greatly improved statistical performance over models using aggregate data. This improvement is attributable to the algorithms' ability to predict the wide fluctuation in emissions between clean and dirty vehicles, largely reflected in their FTP Bag 2 emission test results.

CALINE4 emission prediction algorithm performance is perhaps more impressive when it is noted the EMFAC7F model algorithms were estimated using the SCF data base, but CALINE4's algorithms were estimated using a much older and smaller data set. Both statistical and practical factors were taken into account, and the improved CALINE4 algorithms represent a more powerful approach for estimating CO emissions for individual vehicles, provided that the algorithms are based on comprehensive testing of a representative sample fleet. Note, however, that the CALINE4 modal model still does not capture the effect of different modal activity reflected in the different testing cycles, as evidenced by the vast differences in correlation coefficients across cycles. This suggests the transportation air quality modeling community still needs an improved modal model.

CONCLUSIONS AND RECOMMENDATIONS

To put the research findings presented in this paper to effective and productive use, the transportation community must consider the current regulatory framework. The transportation community must also consider the current direction that complementary modeling efforts are taking and how simultaneous modifications will benefit future air quality analyses. Finally, the findings must be considered with respect to both short- and long-term solutions to current air quality analyses problems.

Research/Modeling Arena

In the short to medium term, the next CALINE4 model revision effort should include an upgrade to its modal emission algorithms. Among its improvements should be the addition of individual vehicle Bag 2 and idle emission rates and the recalculation of the modal model coefficients.

Incorporating individual vehicle Bag 2 and idle rates into model algorithms would require several steps. As an example of how this could be done, consider the following. A sample of tested vehicles (e.g., an expanded SCF data set) would need to be broken down into subsamples by emitter class. For example, four or five subsamples could separate vehicles by emission results on testing cycles, with classes of ultra-high emitters, high-emitters, normal emitters, low emitters, and ultra-low emitters. These subsamples of vehicles would constitute the sample bins from which local vehicle fleets could be approximated. Support files would be included with the CALINE4 software containing local or regional fleet characteristics necessary for subroutine calls from the main program. The subroutine would randomly sample vehicles from the five bins of emitters in the correct proportion to represent the local or regional fleet. These support files constituting the five bins would contain individual vehicle Bag 2 and idle test data (and additional variables needed in the model).

To obtain local or regional fleet characteristics, local or state DMV and BAR records could be used to determine critical determinants of the vehicle fleet composition. The end user could then select default fleet characteristics (a dirty vehicle fleet for worst-case analyses) or enter local or regional fleet characteristics for more accurate analyses. This formulation would require careful classification of emitter subsamples listed in the previous step. This improvement to CALINE would avoid, to the extent possible, miscomputation of average fleet FTP Bag 2 rates and subsequent emission impacts.

The coefficients contained in the CALINE4 model's algorithms were estimated using an older and smaller data set. These coefficients could be verified against a new data set (i.e., the SCF data set) to see whether they still characterize emissions behavior of these vehicles. Using mathematical search procedures, the coefficients could be simultaneously adjusted to minimize the mean squared prediction error and therefore optimize modal algorithms to the current vehicle fleet. Of course, there still remain questions of how representative the SCF data base is of the current vehicle fleet and whether the functional form of the CALINE4 model is the best available modal model formulation. There is reason to believe that improving the coefficients could improve the robustness of CALINE4's explanatory power, providing better estimates of CO emissions from modal events. With updated model coefficients, the modal model could be reassessed to determine whether it captures the emissions variations associated with the range of modal activities.

The CALINE4 modal model algorithms should be used during all assessments, not just those incorporated with intersections (assuming coefficients have been updated and prediction improvements follow). Because average emission outputs from current EMFAC7F and MOBILE models are questionable, their use will increase the uncertainty associated with cruise-related emissions on roadway segments. The cruise emission factor incorporated in the CALINE4 model is likely to yield more accurate results than the method now used, although this should be verified.

In the long term, CALINE's vehicle activity algorithm's should be upgraded to use traffic simulation algorithms for all vehicle activity estimation (not just intersections). In addition, a new, more robust modal model derived from second-by-second emissions data should be used. These upgrades, in addition to the dispersion component of CALINE4, would allow a more accurate assessment of project-level CO emission impacts under worst-case conditions.

Any new model development effort should explore the impact and role of high-emitters in the vehicle fleet. Research of this nature would involve random testing from vehicle fleets in various regions. Factors such as tampering rates, average condition of vehicles, average age of vehicles, accrued mileage, and types of vehicles will likely play a large role in the results. These influential factors are likely to help characterize a local or regional fleet of vehicles and help determine the discrepancies between a regional fleet and the fleet used to estimate models now in use. Research currently under way is assessing the differences between the true vehicle fleet and the fleet used to estimate regional emissions models (Smith et al., unpublished draft research report, Institute of Transportation Studies, University of California at Davis).

Finally, new model development efforts should include outputs that provide uncertainty bounds associated with predictions. Although providing more information to decision makers will make their task more difficult, it will aid in more effective policy decisions. It will also provide policy makers the information with which to devise more sensitive and reasonable policies, which explicitly account for technical uncertainty. Monte Carlo techniques could be used in this sort of model development effort to estimate confidence bounds around predicted values (*14*), or repeated random sampling and model runs to develop long-run average impacts with a measure of confidence could also be used. In either case, the technical uncertainty currently associated with emission impact assessment should be quantified and provided as part of standard assessment outputs.

Policy/Regulatory Arena

For there to be an incentive to develop more robust project-level impact models, regulators must demonstrate that they are willing to commit resources to develop improved models, commit resources to run models, maintain in-house expertise, and approve model improvement efforts for future conformity analyses. Although there is motivation for new model development from a theoretical and academic standpoint, new models are of no use to practitioners if they are not allowed to use them. Regulatory agencies such as CARB and EPA should be urged to remain flexible, yet rigorous, when considering new models for the extremely timely and difficult air quality analyses now predominant in nonattainment regions throughout the United States.

In addition, many of the benefits and methodologies developed for improved project-level modeling are likely to benefit regional modeling improvements as well. Regional models are perhaps in more critical need of improved emission estimation procedures than are project-level models, and therefore a model development effort should keep both modeling arenas in mind.

As a final and critical note, the link between evolving transportation activity models (microsimulation and regional) and evolving air quality models (local impact and regional) must be considered. Currently, the outputs from transportation activity models are seriously deficient for inputs into air quality models and have contributed to emission estimation uncertainties (*2,15*). The link between these two models is absolutely critical to the accurate assessment of emission inventories. If an overall improvement to local and regional air quality models is not accompanied by parallel improvements in transportation activity and simulation models, then few accuracy and precision gains in air quality analyses will be realized. The evolution of these models is likely to take an interesting and exciting path.

ACKNOWLEDGMENTS

This research was funded by the California Partnerships for Advanced Technology and Highways (PATH) Program, Memorandum of Understanding #112. Some of the findings presented here can also be found in the corresponding PATH report. This research benefited from the continuous and generous consultation and feedback from John Nuyen, Caltrans, Division of New Technology, Materials, and Research. The authors also thank the reviewers, Paul Benson of Caltrans, Division of New Technology, Materials, and Research, and Stein Weissenberger of PATH, and PATH editorial review staff for providing detailed and insightful comments during the draft review stages. The authors acknowledge contributions from Cameron Yee and Matt Smith of the Institute of Transportation Studies, University of California at Davis. Finally, the authors gratefully acknowledge PATH, Caltrans, and the FHWA for sponsoring this research effort.

REFERENCES

1. *Air Quality: Transportation Plans, Programs, and Projects: Federal or State Implementation Plan Conformity.* U.S. Environmental Protection Agency, rule. 40 C.F.R. Federal Register, Nov. 24, 1993.
2. Guensler, R. *Vehicle Emission Rates and Average Vehicle Operating Speeds.* Ph.D. dissertation. University of California at Davis, Davis, 1993.
3. Gammariello, T., and J. Long. An Emissions Comparison Between the Unified Cycle and the Federal Test Procedure. Presented at Emission Inventory: Perception and Reality specialty conference, Pasadena, Calif., Oct. 18–20, 1993.
4. Leblanc, D., M. Meyer, F. Saunders, and J. Mulholland. Carbon Monoxide Emissions from Road Driving: Evidence of Emissions Due to Power Enrichment. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1994.
5. Gertler, A., and W. Pierson. Motor Vehicle Emissions Modeling Issues. *Proc., Air and Waste Management Association 84th Annual Meeting,* Pittsburgh, Pa., June 1991.
6. Washington, S. *A Cursory Analysis of EMFAC 7G, Reconciling Observed and Predicted Emissions.* Institute of Transportation Studies, University of California at Davis, Davis, 1994.
7. Washington, S., R. Guensler, and D. Sperling. *Modeling IVHS Impacts—Volume II: Assessment of the CALINE4 Line Source Dispersion Model.* PATH Report MOU #112. Submitted July 28, 1994.
8. Benson, P. *CALINE4, A Dispersion Model For Predicting Air Pollutant Concentrations Near Roadways.* Report FHWA/CA/TL-84/15. FHWA, U.S. Department of Transportation, 1989.
9. Rutherford, J.A. *Automobile Exhaust Emission Surveillance—Analysis of the FY 1975 Program.* Report EPA-460/3-77-022. Environmental Protection Agency, Dec. 1977.
10. Greene, W.H. *Econometric Analysis.* Macmillan Publishing Company, New York, 1990.
11. Fair, R. *Specification and Analysis of Macroeconomic Models.* Harvard University Press, Cambridge, Mass., 1984.
12. Bevington, P., and D. Robinson. *Data Reduction and Error Analysis for the Physical Sciences* (2nd ed.). McGraw-Hill, Inc., 1992.
13. Neter, Wasserman, and M. Kutner. *Applied Linear Models.* Richard D. Irvin, Inc., 1990.
14. Guensler, R. *A Monte Carlo Technique for the Assessment of Motor Vehicle Emission Inventory Uncertainty.* Institute of Transportation Studies, University of California at Davis. Davis, 1994.
15. Replogle, M. Improving Transportation Modeling for Air Quality and Long-Range Planning. Presented at the 72nd Annual Meeting of the Transportation Research Board, Washington, D.C., 1993.