'MODAL' ACTIVITY MODELS FOR PREDICTING CARBON MONOXIDE EMISSIONS FROM MOTOR VEHICLES

Simon P Washington, PhD
and
Troy M Young

Institute of Transportation Studies
University of California at Davis
Davis CA 95616

ABSTRACT

Motor vehicle emission prediction models have been shown to underestimate emissions by factors of 2 to 3 in some cases. There are many reasons why current models do not predict automobile emissions under 'real-world' driving conditions accurately. Among them are: non-representative driving activity used to derive emission prediction algorithms; lack of driving activity input variables; statistical shortcomings of models; and non-representativeness of tested vehicles compared to on-road vehicles.

Given the inaccuracy of current emission prediction models and the need to accurately assess transportation control measures, incremental transportation supply changes, and intelligent transportation technologies, new activity-based emission prediction algorithms are required. Improved algorithms need to be sensitive enough to capture the effects of microscopic flow adjustments, or flow smoothing, that are now commonly considered among transportation and air quality planners.

This paper first presents some results of second-by-second emissions data collected for a 1991 Toyota Camry in Australia. This data is used to demonstrate the importance of modal activity, and the wisdom of incorporating modal variables into an emissions model. The data was used to develop an 'elemental' model for prediction of emissions of CO, HC and NOx; however, only the models for acceleration and deceleration CO emissions are presented here.

As a way to encapsulate the modal modeling approach into the existing modeling framework, a modal model estimated using traditional 'bag data' is introduced. The new model, comprised two linear regression modal models, collectively dubbed DITSEM (Davis Institute of Transportation Studies Emission Model), employs modal explanatory variables such as acceleration, positive kinetic energy and proportion of cycle at idle to predict CO emissions from both 'high' and 'normal' emitting vehicles.

To measure the model performance of DITSEM, CO emission prediction algorithms embedded in 'competing' models (CALINE4 and EMFAC7F) are used to predict emission test results over a wide range of driving cycles. Measures of model performance compared are mean squared error, mean absolute error, Theil's U-Statistic and the linear correlation coefficient. Statistical comparisons show that DITSEM CO prediction algorithms are superior and capture the effect of microscopic flow changes well.

The authors suggest that significant interim improvements can be made using the existing 'bag' collected data—making them sensitive to microscopic changes in travel behavior. A more sophisticated modeling approach, one based on second by second data and similar to the one presented here, could be used for a more long term model improvement program. This large scale second-by-second effort should only be undertaken with specific measurement and modeling objectives in mind, since a great deal of improvement can already be achieved with the current data.

MODAL EMISSIONS DATA FOR MODERN PASSENGER VEHICLE

Changes in driving conditions and speed profile are reflected as changes in the distribution and characteristics of modal activity. Many authors (for example, LeBlanc et al., 1994; Kelly and Grobliki, 1993; Darlington et al., 1992; Austin et al., 1992) have shown that a small proportion of modal activity may be responsible for a large proportion of the pollution generated under a set of driving conditions.

Energy use and exhaust emission data were collected for an Australian-made, 1991 model, Toyota Camry sedan. This vehicle has a 2.0 L electronic fuel injected engine and a four-speed automatic transmission and at the time of testing was 19 months old and had traveled approximately 20,000 km. The vehicle was tested on the chassis dynamometer operated by the Environment Protection Authority of Victoria (EPA-Vic), in Melbourne.

The emission testing involved operating the vehicle on the urban driving cycle (the FTP cycle is currently the standard test cycle used in Australia) through the cold start, stabilized and hot start phases and then the 'cold' start and stabilized phases were repeated with the engine hot. The main purposes of this stage of the testing was twofold: (a) to check that the emissions from this vehicle were comparable with similar vehicles previously tested; and (b) to check that the results were repeatable, by comparing the first hot start with the second 'cold' start and comparing the two stabilized phases. The result from the second stabilized phase was within 4.2% of the first result for HC, 1.0% for CO, 2.9% for NOx, 2.7% for CO_2 and 7.0% for energy use. The standard driving cycle testing produced Bag1, Bag2, Bag3 and the weighted emission rates. The data processor also analyzed the emission mass for each of eighty modes throughout the cycle.

After completion of the standard tests, a series of controlled modal tests were performed. These modal tests included acceleration phases from rest to 100 km/h, decelerations from 100 km/h to rest and cruise (steady-speed) tests at speeds from 10 km/h to 100 km/h in 10 km/h increments. The acceleration and deceleration phases were performed at a constant rate of acceleration and covered a range of acceleration/deceleration rates from 2 km/h/s to 6 km/h/s. Due to the vehicle performance and limitations of the dynamometer (not having large diameter rollers), it was not possible to hold an acceleration rate greater than 6 km/h/s over the full range of speeds. The idle emission rate was determined from the idle modes within the urban driving cycle. The average idle CO emission rate was calculated to be 0.007 g/s.

During the controlled modal tests, emission concentration and mass, exhaust volume, air-fuel ratio, fuel mass, oxygen content and dynamometer speed were recorded on a second-by-second basis. The emissions and fuel consumption data were adjusted appropriately to account for the respective lags in the exhaust system and the analyzers. These lag times had been previously determined for each pollutant by EPA-Vic staff.

ACCELERATION CO EMISSIONS MODEL DEVELOPMENT

Each data point recorded during the acceleration phases was plotted as the total CO emissions produced for an acceleration from rest as a function of the final speed of the acceleration. Figure 1 shows the data set collected for the Toyota Camry.

An analysis of this data showed that much of the variation in the acceleration CO emissions for a given final speed could be explained by differing acceleration rates. The cumulative acceleration rate was calculated at each data point; that is, the total change in speed since the start of the acceleration over the time of the phase. The use of such an acceleration rate as a parameter to describe emissions assumes that the change in speed occurs at a constant rate. This is not the case in real-world traffic conditions; the speed profile for most acceleration phases is an 's-shaped' curve with the lowest instantaneous acceleration rates at the start and end of the phase. However, most available traffic models do not even give details of modal activity, much less average acceleration rates or acceleration profiles. Hence, at least until traffic models are modified, typical acceleration rates need to be assumed.

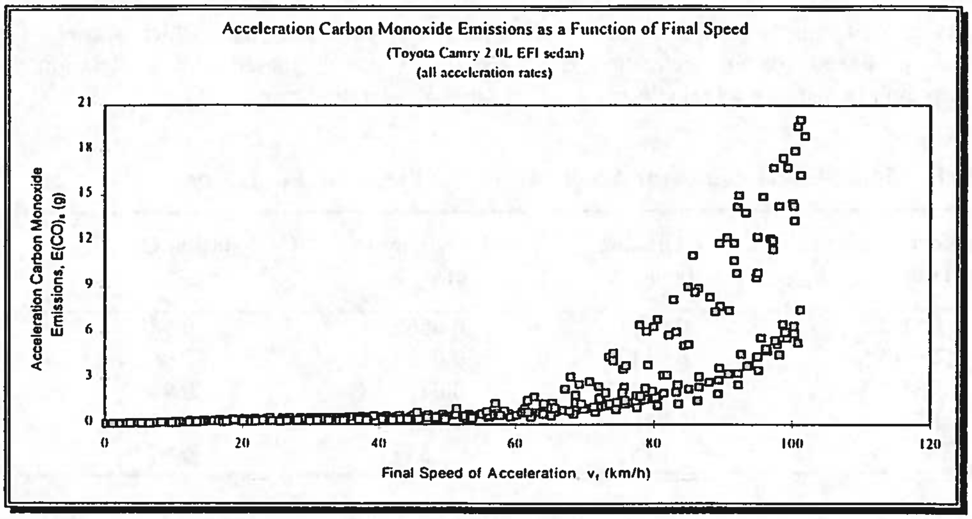

Figure 1 - Data from Modal Emission Testing (Acceleration CO Emissions)

The SPSS for Windows statistics package was used to develop a model describing the complete data set in terms of the final speed of each acceleration, regardless of acceleration rate. The data set was also filtered into 4 subsets based on the average acceleration rate. The subsets were defined by acceleration rates from 1.51 to 2.5 km/h/s, 2.51 to 3.5 km/h/s, 3.51 to 4.5 km/h/s and 4.51 to 5.5 km/h/s. The non-linear regression procedure was used to develop a model for each subset and the shape of these models and the model for all acceleration rates is shown in Figure 2 below. The model form that best described the acceleration CO emissions data was found to be an exponential of the form:

$$E(CO)_a = Ae^{Bv_f}$$

where: $E(CO)_a$ is the acceleration CO emissions in grams; v_f is the acceleration final speed in km/h; and A and B are regression coefficients, shown in Table 1.

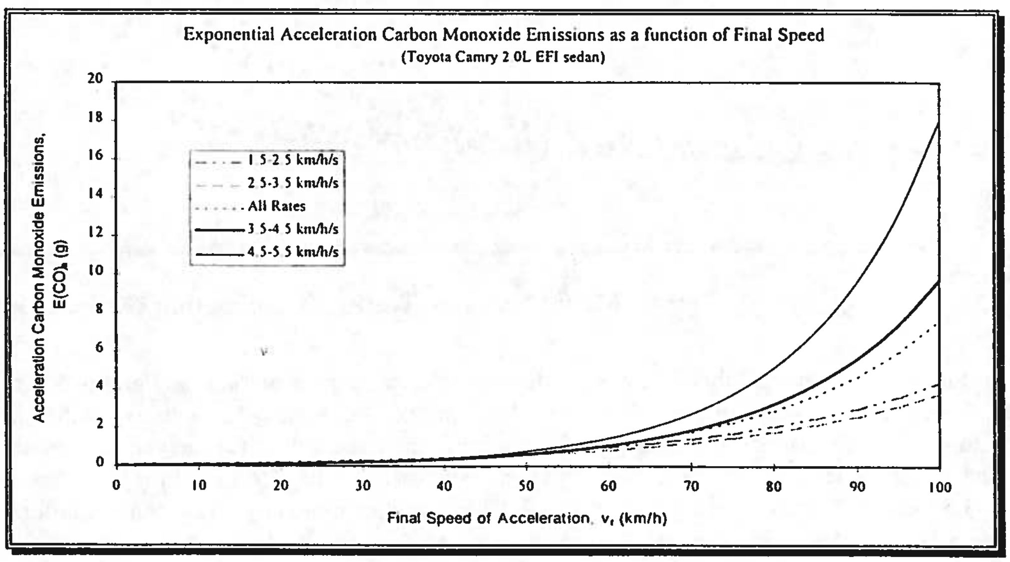

Figure 2 - Acceleration CO Emission Functions

This figure highlights the effect of acceleration rates on the emissions produced during acceleration modes within a driving cycle. It appears that CO emissions increase with average acceleration rate, although this data showed an exception for the two subsets of low acceleration rates (the bottom two curves), for some speeds. It should be noted that for higher acceleration rates the distance traveled to reach a given final speed is less. Hence, the emission rate per distance (e.g. g/km) may in some cases be less for higher acceleration rates. The reverse was seen from an analysis of fuel consumption and carbon dioxide (CO_2) emissions data which showed the fuel used, or CO_2 emissions produced, for an acceleration to a given final speed decreased with acceleration rate. However, the fuel consumption per distance traveled is greater for higher acceleration rates.

Table 1 - Model Coefficients for Acceleration CO Emission Functions

Acceleration Rate (km/h/s)	Constant term, A	Coefficient of v_f, B	Correlation Coefficient, r
1.51-2.5	0.0211	0.0562	0.988
2.51-3.5	0.0112	0.0628	0.962
3.51-4.5	0.0126	0.0712	0.985
4.51-5.5	0.0754	0.0552	0.982
all rates	0.0181	0.0651	0.857

DECELERATION CO EMISSIONS MODEL DEVELOPMENT

The deceleration phase data was plotted on axes of total CO emissions produced during a deceleration to rest and the initial speed of the deceleration. The complete data set is shown in Figure 3.

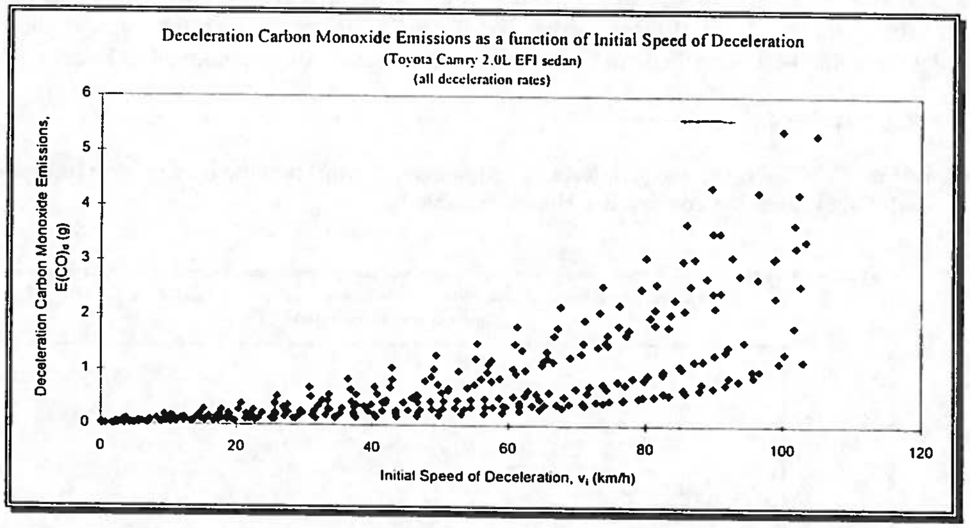

Figure 3 - Data from Modal Emission Testing (Deceleration CO Emissions)

As for the acceleration data, this data was subdivided by average deceleration rate; giving 5 subsets for deceleration rates from -1.5 to -2.5 km/h/s, -2.51 to -3.5 km/h/s, -3.51 to -4.5 km/h/s, -4.51 to -5.5 km/h/s and -5.51 to -6.5 km/h/s. Models of a quadratic form were found to best describe the variation in deceleration emissions for each subset. The linear coefficient of v_i was only significant at the 95% level for the subset of deceleration rates from -3.51 to -4.5 km/h/s. Figure 4 shows the model curves determined by linear regression for each data subset and the regression coefficients and model correlation coefficients are shown in Table 2.

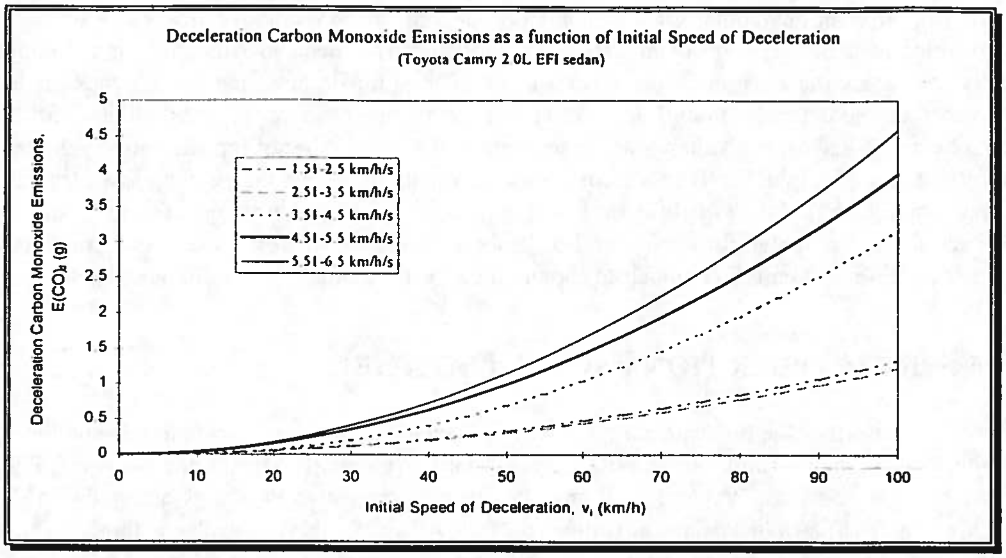

Figure 4 - Deceleration CO Emission Functions

Table 2 - Model Coefficients for Deceleration CO Emission Functions

Deceleration Rate (km/h/s)	Coefficient of v_f, A	Coefficient of v_f^2, B	Correlation Coefficient, r
1.51-2.5	ns	0.00013	0.935
2.51-3.5	ns	0.00012	0.961
3.51-4.5	-0.00421	0.00036	0.992
4.51-5.5	ns	0.00040	0.982
5.51-6.5	ns	0.00046	0.994

ns = not significant

Elemental Model for Emissions Estimation

CO emissions can be estimated as a function of driving modes by predicting the emissions for each discrete element of modal activity and summing them over the complete cycle. That is, separate acceleration, deceleration, cruise and idle functions are used to estimate the emissions generated for each acceleration, deceleration, cruise and idle mode and the results are summed over the cycle. This method was used successfully by Bowyer, Akçelik and Biggs (1985) to model fuel consumption under varying driving conditions.

The models developed for acceleration and deceleration CO emissions, shown in Figures 2 and 4 respectively, clearly show the importance of acceleration/deceleration rates in terms of enrichment of CO emissions during modal activity. Application of such an elemental model requires a driving cycle to be broken up into its component modes, and hence criteria for each mode need to be defined. The discretization of a cycle into its component modes can be performed with the "CYCLE" program mentioned earlier in this paper. In addition, since steady-speed cruise conditions are rare there is a need to model CO emission rates for cruise conditions including speed fluctuations. There is also a need to collect and analyze acceleration and deceleration emissions data for a wider range of acceleration/deceleration rates than covered in this work.

The acceleration and deceleration functions were developed to predict the CO emissions for accelerations from, and decelerations to, rest. However, it has been shown that the acceleration CO emissions for an acceleration from an initial speed, v_i to a final speed v_f can be calculated by subtracting the predicted emissions for an acceleration from rest to v_i from the emissions for an acceleration from rest to v_f. A similar calculation can be used for decelerations to positive speeds.

Clearly the effort in developing such elemental models for a representative fleet of vehicles would be substantial. The development of this form of modeling was initiated by a need in Australia for a computer-based system to predict and assess the environmental impacts of alternative traffic management schemes for local areas. Emphasis was placed on the determination of the *relative* impacts of two or more schemes, rather than the absolute impact of each scheme. Hence, traffic volumes are described by a small number of representative vehicles in their appropriate proportions; for example, the Toyota Camry was representative of medium-sized passenger vehicles. This emission modeling application is notably different to the application of emission models for air quality conformity analyses. However, there is potential for the use of the elemental modeling framework in conformity analyses to substantially improve the ability of emission models to capture the effects of small changes in modal activity.

THE SPEED CORRECTION FACTOR DATA SET

This section describes the fundamental set of data collected over the past decade to develop the internal speed related algorithms in both the California Air Resources Board's (CARB) EMFAC and the United States Environmental Protection Agency's (USEPA) MOBILE models. The data consist of light duty automobile CO emissions collected over the course of various testing or 'driving cycles'. All of the driving cycles comprise a unique combination of starts, stops, cruise, acceleration and deceleration events. Carbon monoxide (CO), hydrocarbons (HC) and oxides of nitrogen (NOx) emissions are generally collected while vehicles are driven on the cycles; however, CO emissions are the focus in this research effort.

The data set used in this model development effort consists of CO emissions collected from 13 cycles. The 13 cycles are shown in Table 3. The table shows the number of vehicles tested on each cycle, the sponsoring agency, the length of each cycle in seconds and the average speed of the cycle in miles per hour. The age of the vehicles tested on these cycles ranges from 1977 model years to 1990 model years. It should be noted that the vehicles tested here represent the majority of the current data set used by both the CARB and the USEPA for development of the activity portion of the emission inventory models, EMFAC7F and MOBILE5a respectively.

Table 3 - Summary Information on Testing Cycles in Speed Correction Factor Database

Cycle Number and Name	Number of Vehicles Tested	Source of Test Cycle	Length of Cycle in Seconds	Average Speed of Cycle in MPH
0 Unified Cycle	56	CARB	1135	27.47[1]
1 Federal Test Procedure - Bag 1	464	USEPA	505	25.60
2 Federal Test Procedure - Bag 2	464	USEPA	866	16.04
3 Federal Test Procedure - Bag 3	464	USEPA	505	25.60
4 Highway Fuel Economy Test	464	USEPA	765	48.27
5 High Speed Test Cycle # 1	25	CARB	474	45.07
6 High Speed Test Cycle # 2	25	CARB	480	51.03
7 High Speed Test Cycle # 3	69	CARB	486	57.77
8 High Speed Test Cycle # 4	69	CARB	492	64.44
9 Low Speed Test Cycle #1	236	USEPA	624	4.02
10 Low Speed Test Cycle #2	236	USEPA	637	3.64
11 Low Speed Test Cycle #3	236	USEPA	616	2.45
12 New York City Cycle	464	USEPA	598	7.10
13 Speed Correction Factor Cycle 12	464	USEPA	349	12.07
14 Speed Correction Factor Cycle 36	464	USEPA	996	35.85

[1] The Unified Cycles actually consists of 3 bags like the long FTP, but only the hot stabilized (Bag 2) portion of the cycles is used.

All of the CO emissions data reflected in Table 3 were collected in laboratories utilizing dynamometers and constant volume sampling emissions equipment. The data were subjected to state and federal agency quality control criteria and are likely the most robust and reliable set of 'bag collected' tailpipe emissions data available for US automobiles.

It is important to note that the FTP Bag 2 test was performed on all vehicles contained in the Speed Correction Factor (SCF) data set, while some of the other tests were performed on a subset of vehicles. For example, 69 vehicle test results are available from the High Speed test cycle #3. For this reason, and others related to the way in which the FTP Bag 2 test cycle was derived, the FTP Bag 2 test result is a pivotal and fundamental component to both USEPA's and CARB's emission inventory models [Guensler, 1993].

To provide a feel for various cycle attributes, descriptive and 'modal' characteristics are shown in Table 4 and Table 5 respectively. The tables show how test cycle characteristics differ across tests. For example, the maximum speed on Low Speed test cycle #3 is 10 mph, while the maximum speed on High Speed test cycle #3 is roughly 67 mph. Almost 50% of the Low Speed test cycle #3 is performed with the vehicle in idle, while the High Speed test cycle #3 only spends about 1% of its testing time in idle.

Table 4 - Descriptive Statistics of Speed Correction Factor Testing Cycles

Cycle Name	Mean Speed	Maximum Speed	Standard Deviation of Speed	Coefficient of Variation of Speed (SD/Mean)
0 Unified Cycle	27.47	68.00	28.26	1.03
1 Federal Test Procedure - Bag 1[1]	25.58	56.67	18.23	0.71
2 Federal Test Procedure - Bag 2	16.04	34.30	10.72	0.67
3 Federal Test Procedure - Bag 3[1]	25.58	56.67	18.23	0.71
4 Highway Fuel Economy Test	48.27	59.90	10.09	0.21
5 High Speed Test Cycle # 1	45.07	53.30	9.67	0.21
6 High Speed Test Cycle # 2	51.03	59.90	11.22	0.22
7 High Speed Test Cycle # 3	57.77	67.40	13.03	0.23
8 High Speed Test Cycle # 4	64.44	74.90	14.95	0.23
9 Low Speed Test Cycle #1	4.02	16.00	4.38	1.09
10 Low Speed Test Cycle #2	3.64	14.00	4.15	1.14
11 Low Speed Test Cycle #3	2.45	10.00	3.07	1.25
12 New York City Cycle	7.10	27.70	8.00	1.13
13 Speed Correction Factor Cycle 12	12.07	29.10	10.23	0.85
14 Speed Correction Factor Cycle 36	35.85	57.00	18.88	0.53

[1]The data collected under the Bag 1 and Bag 3 cycles of the Federal Test Procedure contain cold and hot engine start emission contributions.

Table 5 - Modal Characteristics of Speed Correction Factor Testing Cycles

Cycle Name	Time (Sec)	Dist. (Miles)	% Cycle Idle	% Cycle Accel.	% Cycle Decel.	% Cycle Cruise
0 Unified Cycle	1135	8.66	12.6	30.2	29.0	28.2
1 Federal Test Procedure - Bag 1	505	3.59	19.6	21.0	20.4	39.0
2 Federal Test Procedure - Bag 2	866	3.86	18.6	25.3	19.3	36.8
3 Federal Test Procedure - Bag 3	505	3.59	19.6	21.0	20.4	39.0
4 Highway Fuel Economy Test	765	10.26	0.7	14.1	11.8	73.4
5 High Speed Test Cycle # 1	474	5.93	1.1	13.3	9.9	75.7
6 High Speed Test Cycle # 2	480	6.80	1.0	13.8	10.4	74.8
7 High Speed Test Cycle # 3	486	7.80	1.0	14.2	10.9	73.9
8 High Speed Test Cycle # 4	492	8.81	1.0	15.3	11.4	72.3
9 Low Speed Test Cycle #1	624	0.70	36.5	24.2	25.6	13.7
10 Low Speed Test Cycle #2	637	0.64	38.8	23.4	24.3	13.5
11 Low Speed Test Cycle #3	616	0.42	47.7	16.2	17.9	18.2
12 New York City Cycle	598	1.18	34.9	23.9	24.2	17.0
13 Speed Correction Factor Cycle 12	349	1.17	27.2	26.1	24.1	22.6
14 Speed Correction Factor Cycle 36	996	9.92	6.5	19.0	16.0	58.5

DEVELOPING THE DITSEM MODAL EMISSION MODEL

Unique to the treatment of the SCF data is the notion of adding modal variables for use in model estimation. To break down the 13 cycles into modal components, a FORTRAN computer program entitled "CYCLE" was written, debugged and compiled. The program discretizes any driving cycle into idle, cruise, acceleration, positive kinetic energy, power and deceleration events. The program is very flexible and allows the user to select any number of 'cutpoints' with which to discretize cycles. For example, the user can request the program to discretize acceleration events for the Unified cycle at 1.0 mph/sec and at 2.0 mph/sec. The program will then compute the number of seconds (which can be converted to percent of cycle) spent accelerating between 0 and 1.0 mph/sec, between 1.0 and 2.0 mph/sec and greater than 2.0 mph/sec.

Summary statistics were calculated using BMDP statistical software for all variables used in the model development. Summary statistics for all modal variables, as well as definitions for each of the modal variables, are shown in Table 6. The table shows all of the variables available for use in the statistical model building process and provides a definition of each variable in the table footnotes. The explanatory variables consist of testing cycle variables, vehicle variables and emission result variables. For example, 4126 vehicles in the data set contain idle in neutral emission test results and 2365 vehicles contain idle in drive emission test results. The predictor variable CO, on the other hand, consists of the emission rate of carbon monoxide in grams of pollutant per mile driven. These variables are shown in different forms, ranging from straight grams per second emission rates to micro-grams per second per cubic inch of engine displacement.

Table 6 - Summary Statistics for Variables Developed for Dissertation Research

Variable Name[1]	Total Number of Vehicles	Mean	Standard Deviation	Standard Error of Mean	Smallest Value	Largest Value	Range
CYCLE	4695	7.7465	4.7652	.06954	0.0000	14.000	14.000
SECONDS	4695	654.40	200.15	2.9210	349.00	1135.0	786.00
AVGSPEED	4695	22.575	15.764	.23006	2.4500	64.440	61.990
DIST	4695	4.3532	3.5784	.05222	.41922	10.257	9.8382
PKE>105	4695	.02190	.17398	.00254	0.0000	1.5900	1.5900
PKE>90	4695	.26366	.38530	.00562	0.0000	2.5600	2.5600
PKE>75	4695	.64602	.80551	.01176	0.0000	4.6700	4.6700
PKE>60	4695	1.4612	1.5550	.02269	0.0000	8.9900	8.9900
PKE>45	4695	3.4774	2.7529	.04018	0.0000	14.270	14.270
POW>16	4695	3.0479	3.5297	.05151	0.0000	16.260	16.260
POW>14	4695	4.1296	4.3795	.06392	0.0000	19.920	19.920
POW>12	4695	5.3140	5.3706	.07838	0.0000	22.560	22.560
POW>10	4695	6.9115	6.5196	.09515	0.0000	26.220	26.220
POW>8	4695	9.5317	8.0597	.11763	0.0000	27.440	27.440
ACC>3	4695	2.6829	2.0213	.02950	0.0000	5.1500	5.1500
ACC>2	4695	5.3341	3.2552	.04751	.49000	9.1100	8.6200
ACC>1	4695	11.720	4.6163	.06737	2.4800	15.720	13.240
DEC>3	4695	3.9247	2.6604	.03883	0.0000	8.7200	8.7200
DEC>2	4695	6.9079	4.0990	.05982	.16000	15.680	15.520
DEC>1	4695	13.208	4.9335	.07200	4.8500	28.990	24.140
CRZ>60	4695	.57195	2.9591	.04319	0.0000	18.290	18.290
CRZ>50	4695	3.0861	4.5945	.06705	0.0000	18.520	18.520
CRZ>40	4695	4.3095	5.8373	.08519	0.0000	18.750	18.750
CRZ>30	4695	5.3697	6.2638	.09142	0.0000	18.990	18.990
%IDLE	4695	19.421	12.480	.18214	.52000	45.290	44.770
CO	4695	16.382	52.020	.75920	.01250	971.00	970.99
FTPBAG2	4592	9.7107	30.511	.45026	.01250	321.50	321.49
IDLENEUT	4126	138.03	375.20	5.8411	0.0000	3362.1	3362.1
IDLEDRV	2365	147.47	308.16	6.3368	0.0000	2414.9	2414.9
MODYR	4695	83.753	3.2427	.04732	77.000	92.000	15.000
INERTIA	4687	3280.4	606.73	8.8624	0.0000	5500.0	5500.0
DYNOHP	4687	8.6083	2.0781	.03035	0.0000	16.400	16.400
CID	4431	183.18	74.139	1.1138	61.000	460.00	399.00
TRAN	4011	1.5613	1.3534	.02109	1.0000	5.0000	4.0000
CYLN	4011	5.3949	1.5571	.02459	4.0000	8.0000	4.0000
SAR	3970	2.7234	.68756	.01091	2.0000	4.0000	2.0000
FINJ	4431	1.9790	.75826	.01139	1.0000	3.0000	2.0000
CATCONV	3955	3.2599	.70671	.01124	1.0000	4.0000	3.0000
COLC	4375	.92594	.42283	.00639	0.0000	2.0000	2.0000
COPERSEC	4695	.06076	.15936	.00233	851E-8	2.6709	2.6709
COPERCID	4431	366.87	1095.6	16.458	.02789	19785.	19785.
B2PERSEC	4592	.04316	.13561	.00200	556E-7	1.4289	1.4288
B2PERCID	4431	259.29	907.58	13.634	.18215	10584.	10584.
IDPERCID	3955	178.32	466.46	7.4171	0.0000	5173.5	5173.5

[1] The variables are defined below:
CYCLE is the cycle number corresponding to the numbering identified in Table 3,
SECONDS is the duration of the test cycle,
AVGSPEED is the average speed of the test cycle in miles per hour,
DIST is the distance of the test cycle in miles,
PKE>X is the percent of the cycle that is spent with positive kinetic energy (acc. x velocity) greater than X in mph^2/sec,
POW>X is the percent of the cycle that is spent with power (acc. x $velocity^2$) greater than X in mph^3/sec,
ACC>X is the percent of the cycle that is spent with acceleration rates greater than X,
DEC>X is the percent of the cycle that is spent with deceleration rates greater than X,
CRZ>X is the percent of the cycle that is spent with cruising speeds greater than X,
%IDLE is the percent of the cycle spent at idle,
CO is the grams per mile emission rate for a vehicle on a testing cycle,
FTPBAG2 is the grams per mile emission rate for a vehicle on the Federal Test Procedure, Bag 2 test cycle,
IDLENEUT is the grams per hour emission rate for an idling vehicle in neutral on a testing cycle,
IDLEDRV is the grams per hour emission rate for an idling vehicle in drive on a testing cycle,
INERTIA is the curb weight of the vehicle loaded with fuel and oil in pounds,
MODYR is the last two digits of the model year for a vehicle,
DYNOHP is the dynamometer measured horse power, including factors such as drag, coast down ,friction, and frontal area,
CID is the cubic inch displacement of the tested vehicles engine,

TRAN is transmission type of the vehicle (1= automatic, 2= semi-automatic, 3=3-speed manual, 4=4-speed manual, 5=5-speed manual),
CYLN is the number of cylinders in the vehicle, 4, 6, or 8,
SAR is supplemental air recirculation (1=pre circa 1980, 2=no, 3=yes, air pump, 4=yes, pulse),
FINJ is the fuel injection type (1= port, 2= carburetor, 3=throttle body),
CATCONV is the catalytic converter type (1=none, 2=oxidation only, 3=3-way catalyst, 4=oxidations and 3-way catalyst),
COLC is closed or open look catalyst type (0=pre circa 1980, closed, 1=closed, 2=open),
COPERSEC is the grams per second emission rate for a vehicle on a cycle,
COPERCID is the micro grams per second per cubic inch of displacement emission rate for a vehicle on a cycle,
B2PERSEC is the grams per second emission rate for a vehicle on the FTP Bag 2 cycle,
BS2PERCID is the micro grams per second per cubic inch of displacement emission rate for a vehicle on the FTP Bag 2 cycle,
IDPERCID is the micro grams per second per cubic inch of displacement emission rate for a vehicle on the idle in neutral cycle.

Statistical Approach and Methodology

The first question that is addressed here is why use statistics? Why not develop a *theoretical* instead of *statistical* model that can estimate CO emissions from the internal combustion engines of motor vehicles?

First and foremost, a theoretical model requires that cause-effect variables be known about vehicles in the vehicle fleet. From combustion theory we know that the most important variables affecting carbon monoxide emissions are the fuel to air equivalence ratio ϕ, engine revolutions per minute, volume and number of the engine cylinders, friction and efficiency losses, engine operating temperature, engine load, catalytic converter equipment, and various measures of maintenance state of the engine and vehicle. In addition, fuel chemical composition plays a role in determining how much pollutants are formed in the combustion process [Heywood, 1988].

The most obvious immediate problem is that most of this information is not available for vehicles in the SCF data base and obtaining this information requires a level of data collection not possible for a large vehicle fleet. It is impractical to suggest that the gamut of variables necessary to actually predict emissions could be collected for a large fleet of vehicles. Instead, we collect 'surrogate' measures of the above variables and hope that these surrogates will do a reasonable job of substituting for the 'real' causal variables, enabling prediction of CO emissions. Variables such as engine size, inertial weight, catalytic converter type and number of cylinders are used to 'capture' the same effects as the real causal variables. Since these variables are observational and encompass a random sampling scheme, statistical methods are most appropriate. In addition, one of our goals is CO emission prediction from vehicles in a hypothetical fleet on a hypothetical driving route, so again statistical methods are most appropriate.

Modeling Objectives

The objectives of this model estimation process are several. First, the model algorithms should be able to forecast emissions from vehicles driven on cycles with significantly different profiles of modal activities (idle, cruise, acceleration, deceleration). In essence, the models should capture the modal component of the driving cycle. This is an important departure from the existing set of models, which can only differentiate emissions from cycles with different average speeds [Guensler, 1993]. This means that the estimated models must contain variables that reflect, at least to some extent, the modal activities contained in a given driving cycle. The second objective is to estimate models that have all the properties of a statistically robust model; that is, parameter estimates that are un-biased, consistent and efficient. The third objective is to utilize model variables that can readily be obtained in future data collection efforts and that could easily be included in an interim model improvement effort. These variables should also facilitate updating of a vehicle fleet.

Modeling Hypotheses

Hypothesis formulation allows a researcher to test previous research findings, or educated 'hunches' about relationships reflected in the data. Perhaps the most important hypothesis to be tested here is the notion that the sample of vehicles contained in the SCF data set can be statistically separated into two defendable and natural subgroups, 'normal' emitting vehicles and 'high' emitting vehicles. This hypothesis is supported by previous research [Washington, Guensler and Sperling, 1994] suggesting that high emitting vehicles are dominating currently employed model algorithms. In addition, the USEPA in their MOBILE model treat vehicles differently with respect to their emissions behavior. To this end, the sample is explored to see if it can be divided into two statistically different groups.

Outliers in these analyses are used only to identify alternative variable functional forms and alternative model specifications. This is to say that 'outliers' will not be discarded from the data set. There are two reasons for this. First, the data have previously been rigorously checked for quality control by USEPA and CARB staff, and by Dr. Randall Guensler during his dissertation effort. In addition, data screening techniques were employed to locate and rectify typographical errors. The data set should by all standards be complete and accurate, and devoid of any 'incorrect' entries. The second reason for keeping outliers in the model is that they are very likely to represent 'real' vehicles in the fleet and removing them would also remove the ability of an estimated model to 'capture' their effect. This philosophy is the primary impetus for hoping to statistically separate high emitting vehicles from normal emitting vehicles.

It is desired at the outset, to construct a parsimonious model to predict emissions. This means that a model with many explanatory variables and robust explanatory power is good, but a model with fewer explanatory variables and a slight reduction in predictive power is better. To estimate a model with few, yet powerful, explanatory variables, a methodology is employed to determine whether explanatory variables are contributing 'enough' to the model to justify their inclusion. Although very subjective, it is decided to only include explanatory variables that contribute more than around 0.5% to the overall adjusted r-square value of the model. This means that models selected throughout do not have the highest adjusted r-square values, but include the variables determined to meet the parsimony criteria established above.

In addition, 2nd order interactions between independent variables are considered throughout model estimation, but are only included in the models and the accompanying discussion if the interaction terms meet the criterion of parsimony established above and the interaction term is independently significant in the model using the t-test. Interactions are not expected to be important in the modeling process, but will be considered and assessed for completeness.

Research Design for a Carbon Monoxide Modal Emission Model

There are several important reasons for choosing CO as the 'best' pollutant to model. First, CO has traditionally proven to be the hardest pollutant to model. Both the USEPA and CARB model algorithms, upon inspection, reflect the fact that estimation of both HC and NOx pollutants result in more robust and more easily estimable models. In essence, there is not as much random error with respect to NOx and HC as there is with CO. Secondly, CO is currently modeled on both the local and regional level and is therefore of critical concern for micro-scale as well as macro-scale analyses. The Clean Air Act mandates that regions determine emission inventories of CO, as well as perform local impact analyses of CO 'hotspots' [USDOT, 1993]. To this end, the model estimated for CO is compared to a local impact model, CALINE4 [Benson, 1989] and a regional model, EMFAC7F [CARB, 1992] to evaluate different performance characteristics of the models (see The Performance of DITSEM). In addition, CO is almost entirely emitted from the vehicle tailpipe, whereas hydrocarbon emissions are only 65% emitted from vehicular tailpipes. Finally, the methodology described here is the most important contribution to theory, its application being the use of CO emission data. The same exact exploratory model development procedure can also be applied to develop models for HC and NOx, and should be done if a model improvement effort is to be initiated by a regional or local agency.

THE DITSEM MODEL AND ITS INTERPRETATION

After weeks of data manipulation, transformation and error plots, and after meeting all of the modeling objectives and statistical properties, two linear regression models were developed to predict CO emissions from a fleet of vehicles; one for high emitting vehicles (Model I), and one for normal emitting vehicles (Model II). The two models are shown below.

Model I - High Emitter Model (COPERCID* > 2.5)

$$LOG_{10}[(CO/CID) + 1] = 1.5720 - 0.5503(BAG2) + 0.1775(BAG2^2) + 0.0128(MODYR) + 0.0112(\%IDLE) + 0.0104(AVGSPD),$$

Model II - Normal Emitter Model (COPERCID* <= 2.5)

$$LOG_{10}[(CO/CID) + 1] = 2.2360 + 0.5132(BAG2) + 0.0835(PKE>60) - 0.0170(MODYR) - 0.0067(\%IDLE) + 0.04093(ACC>3),$$

where:

ACC>3 = percent of cycle spent with acceleration rate greater than 3 mph/sec;
AVGSPD = average speed of cycle in miles per hour;
BAG2 = LOG_{10}(B2PERCID + 1);
B2PERCID = carbon monoxide emissions in micro-grams per cubic inch displacement per second on the Federal Test Procedure, Bag 2;
CO = micro-grams per second of carbon monoxide emissions;
CID = cubic inches of engine displacement;
MODYR = last 2 digits of model year of vehicle;
%IDLE = percent of test cycle time spent at idle;
LOG_{10}[(CO/CID) + 1] = COPERCID*
PKE>60 = percent of cycle spent with positive kinetic energy (velocity x acceleration) greater than 60 mph^2/sec.

When model specifications between high and normal emitter models are compared and contrasted, several interesting differences can be seen. The first is that the sign for the variable MODYR is different for the two models. We would expect newer vehicles to emit less carbon monoxide per cubic inch of engine displacement than older vehicles due to more stringent emissions control equipment, more sophisticated fuel delivery systems and generally better condition of the engines. However, we see that for high emitters, emissions increase with an increase in MODYR. This is likely explained by the fact that newer vehicles are more likely to be controlled by computer processors, which when malfunctioning or inoperative, result in poor contol of the entire emissions control process, from combustion to spark retardation.

When looking at the normal-emitter model, we note that CO emissions increase as percent of cycle spent at idle decreases. This 'negative' correlation may be explained by a couple of factors. First, as percent idle decreases, activity in other modes, i.e. acceleration, deceleration, increases. If the other modal activities result in emission rates that are higher than those associated with idling, then an increase in idling will reduce overall emissions from other modes of activity. Conversely, if idling is an extreme model activity, then an increase in idling will outweigh the increases in emissions from other modal activities. So the sign of the percent of cycle spent at idle variable is largely influenced by the relative difference between emission rates at idle versus emissions rates from other modes of activity. In the models described above, the average idle rate for high emitting vehicles is slightly higher than the emission rate for the composite of other modes, while the average idle rate for normal emitting vehicles is lower than the composite emission rate for other modes of activity.

The collection of significant variables in the two models is different. The high emitter model has fewer cycle attribute variables than does the normal emitter model. This makes intuitive sense, since we expect that high emitting vehicles are less affected by cycle characteristics than are smoothly operating vehicles. This difference is evident in that the high emitter model contains two cycle attribute variables, %IDLE and AVGSPD, while the normal emitter model contains PKE>60, %IDLE and ACC>3. We also note that AVGSPD is a significant explanatory variable for high emitters, but more specific cycle attributes such as acceleration and positive kinetic energy percentages are more important for normal emitting vehicles.

The partial slope coefficients (beta's) in a properly estimated regression model can be interpreted to be the absolute change in the response Y as a function of unit changes in the variable of concern, provided that intercorrelation does not exist and that real world changes in one variable do not affect values of other variables. For example, in the normal emitter model the partial slope coefficient for the independent variable MODYR is -0.0170. This suggests that if all other variables in the model are held constant and we increase MODYR by one unit, from 78 to 79 for example, then the predicted COPERCID* will decrease by -0.0170 units. Care must be taken when interpreting partial slope coefficients in this manner, since in some cases it may be unlikely to vary just one of the variables without affecting other variables in the model.

This is especially true when intercorrelation is present (not a problem here). As an example, it would be hard to vary ACC>3 while holding all other variables constant, since PKE>60 is itself a function of ACC>3. Nonetheless, this interpretation of partial slope coefficients can be a very useful tool in which to derive relationships embedded in the model.

It is unknown whether the true population model has been sufficiently approximated with the DITSEM model, but there is indication to believe that it has. First, the error component has been reduced to randomness that fits a normal distribution of errors with estimated mean and variance parameters. These errors have been shown to be homoscedastic for the high emitter model and heteroscedastic for the normal emitter model [Washington, 1994].

Heteroscedasticity is not uncommon when dealing with observational data and makes sense with these emissions data. This natural tendency can be seen with the data in the original grams per mile format. Some vehicles emit up to 1000 times more emissions than other vehicles under certain conditions (hence the log transformation). This translates to large variability in CO emissions for high emitting vehicles, where the variability for 'cleaner' vehicles is much smaller. Heteroscedasticity results in inefficient parameter estimates; that is, high standard errors. However, the heteroscedastic normal regression model that resulted was rectified with weighted least squares, which weighted observations according to their individual error terms and reduced the standard errors of the estimated coefficients.

Although heteroscedastic for the normal emitter group of vehicles, the error terms have been shown to be approximately normally distributed for both normal and high emitting groups [Washington, 1994]. The slight departure from normality is not serious, as serious departures require significant skews of distributions, as indicated by large differences between the mean and median of the distribution.

Possessing Best Linear Unbiased Estimators (BLUE) parameters in the regression models, theoretical concerns of making statistical inferences and forecasts with DITSEM are satisfied. The most logical question that arises is, how could DITSEM be used to forecast emissions from a fleet of vehicles and furthermore, how could it replace existing emission model algorithms in light of its position in the stream of transportation-air quality models?

To answer this, we must consider how statistical inferences from the model can be made to a hypothetical vehicle fleet population. First of all, we know that we must collect or know information about the model year distribution of vehicles, but we must also know the FTP Bag 2 test result for vehicles. This illustrates a more complicated issue. Difficulties arise since theoretical constraints prohibit estimation of values outside the domain of values used to estimate the model. For example, if we know the model year distribution of vehicles, how do we know that combinations of model year and Bag2 results for vehicles in the model estimation data set are represented in the vehicle fleet population? We don't. All we can do is collect as many vehicles as possible to perform vehicle tests on and periodically 'update' the coefficients in the model if necessary to reflect the changing domain of variables and emissions characteristics of the vehicle fleet.

There is another constraint that confounds the domain problem described above. In order to make inferences to a population of vehicles, we must be confident that the vehicles used to estimate the model are representative of the 'universe' population that we are trying to forecast. In the case of the SCF data set, we know this is not the case. In fact, recent work [Smith, Guensler and Washington, 1994] indicates that the current data set is heavily skewed toward 1981 to 1989 model year vehicles in the 2000 to 4000 pound weight class. In the current national fleet of vehicles, for example, model years 1981 to 1989 make up about 50% of the total fleet, whereas they constitute about 85% of the SCF data set. We note that late model years, from 1991 and 1994, make up about 30% of the current California vehicle fleet, yet comprise 0% of the SCF data set. With these large discrepancies it is tenuous, at best, to count on reliable statistical inferences about existing fleet emission rates. However, it is currently the best data set available and now the best model available; therefore we should make the best predictions we can until more representative data to estimate a model become available. The risk, of course, is model prediction bias in an unknown direction; although if the data are skewed toward older vehicles, model predictions are likely to be conservative and err on the side of over-prediction of emissions.

To estimate emissions using DITSEM would also require knowledge of speed-time profiles on a transportation network. Currently, this is beyond current regional data collection ability. However, several research projects are currently underway to try to better understand the actual speed-time profiles that occur under a variety of traffic flow conditions and on different facility types. There is reason to suspect that if speed-time profiles on facility types under different levels of service (LOS) can be statistically separated, then current UTPS model outputs could be combined with LOS data and facility type data to estimate speed-time profiles in a region. These profiles could then be used as input to DITSEM, which would then estimate emissions in a region.

Of course there are still many uncertainties involved with the UTPS process and applying post hoc processing to already inaccurate outputs will not solve any of the current UTPS problems, but it is likely to improve emission estimates in the interim. The solution represents an incremental improvement to existing methodologies, instead of a complete revision. The long term solution however, will be a complete revision of the entire modeling 'chain'.

THE PERFORMANCE OF DITSEM

The new model was 'validated' by comparing its predictive ability to both the CALINE4 and EMFAC7F carbon monoxide estimation algorithms [see Washington, Guensler and Sperling, 1994]. We note from the start that both the CALINE4 and EMFAC7F models do much more than predict carbon monoxide emissions, and the comparison is not to the models themselves, but to the mathematical algorithms within the models that contribute to part of their overall purpose. It is important to note that the intended application of CALINE4 and EMFAC7F is very different. The CALINE4 model is used for local impact analyses or emission impact assessments, while the EMFAC7F model is used for regional assessments or emission inventory predictions. Regardless, the intention of the model algorithms are to accurately predict carbon monoxide emissions from motor vehicles.

An important operating characteristic of both the CALINE4 and EMFAC7F algorithms should be discussed. Both model's algorithms were developed using 'bag' collected data from motor vehicles, although from different samples. The equations contained in CALINE4 and EMFAC7F actually operate using *average* fleet emission rates. The average rates are updated yearly and reflect the average FTP Bag 2 emission rate for a fleet of vehicles. These mathematical algorithms have been shown to be drastically improved when *individual* FTP Bag 2 emission rates are employed in the calculus and prediction of carbon monoxide emissions [Washington, Guensler and Sperling, 1994]. For this reason, prediction comparisons will be made between the 'actual' operation of CALINE4 and EMFAC7F algorithms using *average* emission rates, and the 'theoretically' improved algorithms employing *individual* emission rates. In all subsequent discussions, therefore, the 'theoretical' algorithms will refer to the use of individual emission rates as inputs, whereas 'actual' will refer to the use of average emission rates as algorithm inputs.

In the subsequent analyses, the three models—EMFAC7F, CALINE4 and the newly estimated DITSEM model—will be assessed on two measures of predictive ability; Theil's U-Statistic and the Correlation Coefficient (four goodness of fit test results can be found in Washington, 1994). The tests are based on the ability of the algorithms to accurately predict the total amount of carbon monoxide in grams emitted from a sample of vehicles over a given testing cycle. The testing cycles represent unique profiles of acceleration, deceleration, cruise and idle, and therefore represent a good test bed to assess how well the models cope with modal differences.

It is clearly desirable to have a model that can perform equally well under different modal characteristics; in fact, this is the idea behind a modal model. One of the striking characteristics of CALINE4 and EMFAC7F is their inability to perform equally well on different cycles, despite the fact that CALINE4 contains some modal algorithms [Washington, Guensler and Sperling, 1994].

All of the analyses performed in this section were accomplished with BMDP statistical software, Microsoft Excel, Microsoft Visual Basic and Microsoft FORTRAN. In fact, a great portion of the analyses performed on CALINE4 and EMFAC7F required compilation of an extensive and detailed BASIC computer program [Washington, 1994] that performs the equivalent functions as the internal CALINE4 and EMFAC7F carbon monoxide prediction algorithms.

Theil's U-Statistic Comparisons

A proposed measure of model performance that is not subject to the scaling problems of the more commonly employed R-Square measure is Theil's U-Statistic [Greene, 1990; Fair, 1984]. The 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, while small values of U indicate good fit. The U-Statistic formula is given by:

$$U_j = \{ [(1/n_j) \Sigma_i(\Psi_{ij} - Y_{ij})^2] / [(1/n_j) \Sigma_i (\Psi_{ij})^2] \}^{0.5},$$

where;

U_j = Theil's U-Statistic for all vehicles on cycle j,
Σ_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,
n_j = number of vehicles tested on cycle j.

Theil's U-Statistic results are shown in Table 7. DITSEM is superior by this statistical measure on all test cycles. In some cases, the difference in U between DITSEM and the next 'best' competing model is an order of magnitude.

Again, inspection of the table shows that in most cases, the theoretically improved CALINE4 and EMFAC7F models perform better than do the 'actual' emission prediction algorithms.

Table 7 - Comparison of Theil's U-Statistic (Grams)

Cycle Name	U-Statistic New Model (Grams)	U-Statistic Theoretical CALINE4 (Grams)	U-Statistic Theoretical EMFAC7F (Grams)	U-Statistic Actual CALINE4 (Grams)	U-Statistic Actual EMFAC7F (Grams)
Highway Fuel Economy Test	0.033	0.605	0.537	0.967	0.966
High Speed Test Cycle # 1	**0.358**	1.107	1.524	0.760	0.968
High Speed Test Cycle # 2	**0.440**	1.524	1.930	0.922	1.125
High Speed Test Cycle # 3	**0.238**	0.799	0.935	0.752	0.720
High Speed Test Cycle # 4	**0.086**	0.921	1.054	0.940	0.936
Low Speed Test Cycle # 1	**0.094**	0.689	1.019	0.930	0.911
Low Speed Test Cycle # 2	**0.070**	0.655	0.964	0.943	0.923
Low Speed Test Cycle # 3	**0.061**	0.702	1.035	0.942	0.922
New York City Cycle	**0.061**	0.389	0.549	0.919	0.917
Speed Correction Factor 12	**0.075**	0.413	0.424	0.932	0.933
Speed Correction Factor 36	**0.034**	0.533	0.554	0.952	0.950
Unified Cycle	0.061	n/a	n/a	n/a	n/a

Bold = Smallest U-Statistic in emission estimate

Linear Correlation Coefficient Comparisons

As a final useful statistical comparison of the three models, the linear correlation coefficient is employed [Bevington and Robinson, 1992; Neter et al., 1990]. 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 we expect the linear correlation to be high; whereas, if a model predicts poorly the correlation coefficient will be low. The formula for the correlation coefficient is given by:

$$r_j = \Sigma_i (\Psi_{ij} - Y_{j\,(ave)}) (Y_{ij} - Y_{j\,(ave)}) \,/\, [\Sigma_i (\Psi_{ij} - Y_{j\,(ave)})^2\, \Sigma_i (Y_{ij} - Y_{j\,(ave)})^2]^{0.5},$$

where;

r_j = correlation coefficient between observed and predicted emissions for i vehicles on cycle j,

Σ_i = summation over I vehicles on cycle j,

Ψ_{ij} = observed emissions for vehicle i on cycle j in grams,

$Y_{j\,(ave)}$ = average observed emissions for all vehicles on cycle j in grams,

Y_{ij} = predicted emissions for vehicle i on cycle j in grams.

The correlation coefficients for the three models are compared in Table 8. Again we see that DITSEM outperforms the other models on all but one of the cycles. We also note that DITSEM is fairly constant across cycles, whereas both CALINE4 and EMFAC7F vary greatly across cycles. For example, the range of values for DITSEM, CALINE4 and EMFAC7F are 0.62 to 0.98, 0.20 to 0.91 and 0.27 to 0.91, respectively. A large range in r suggests that a model is performing better on some cycles than on others; not a desirable property, since there are infinitely many real-world driving cycles.

If we consider that the results for High Speed test cycle #2 are based on a sample of 25 vehicles, the effect of one or two 'outlying' vehicles could highly skew the results shown in the table. If we omit this value, then the range of r for DITSEM becomes 0.83 to 0.98, a fairly consistent and narrow range of values.

Table 8 - Comparison of Correlation Coefficients, r

Cycle Name	r DITSEM (Grams)	r Theoretical CALINE4 (Grams)	r Theoretical EMFAC7F (Grams)	r Actual CALINE4 (Grams)[1]	r Actual EMFAC7F (Grams)[2]
Highway Fuel Economy Test	**0.978**	0.792	0.835	0	0.006
High Speed Test Cycle # 1	**0.853**	0.843	0.836	0	0.126
High Speed Test Cycle # 2	0.622	**0.786**	0.774	0	0.028
High Speed Test Cycle # 3	**0.836**	0.310	0.361	0	0.328
High Speed Test Cycle # 4	**0.944**	0.201	0.266	0	0.149
Low Speed Test Cycle # 1	**0.937**	0.702	0.635	0	0.128
Low Speed Test Cycle # 2	**0.941**	0.734	0.634	0	0.141
Low Speed Test Cycle # 3	**0.966**	0.684	0.515	0	0.161
New York City Cycle	**0.949**	0.911	0.878	0	0.092
Speed Correction Factor 12	**0.949**	0.900	0.913	0	0.102
Speed Correction Factor 36	**0.960**	0.832	0.828	0	0.068
Unified Cycle	0.861	n/a	n/a	n/a	n/a

Bold = Greatest correlation coefficient between observed and predicted emissions
[1] 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).
[2] The correlation coefficient for EMFAC7F is non-zero since different within-cycle predictions result from the differences brought about by different technology groupings

CONCLUSIONS

This research first substantiates a functional relationship between vehicular modes of activity and CO emissions. Using second-by-second emissions data collected from an Australian vehicle, the importance of discrete modes of vehicle activity (including acceleration, cruise, idle, and deceleration) are discussed. The findings confirm other beliefs and findings regarding vehicle activity and emissions: extreme modes of activity, such as hard accelerations, contribute disproportionately to the overall inventory of CO emissions from a motor vehicle. The research again points modelers down the road toward modal emission models; but another question arises. Can we use the existing collection of emissions data to develop a modally based emissions model?

To answer this question, we turned to the existing emissions models and the data they are based upon. We showed that the existing data can be used to derive modally based emissions models. Furthermore, the 'new' models possess superior predictive ability over currently employed emission prediction algorithms contained in the EMFAC7F model and the CALINE4 models.

The new model, dubbed DITSEM, actually consists of two linear regression models, one for 'high' emitting vehicles and one for 'normal' emitting vehicles. When we compare and contrast DITSEM, based on 'bag' collected data and the elemental model based on 'second-by-second' collected data, the results are enlightening. Both models possess similar linear additive structure with inclusion of similar explanatory variables. This result is not all that surprising, since we expect vehicle emissions profiles to generally behave similarly to the same stimulus; in this case vehicle modal activities. We do note however, that this finding raises a question as to the utility and purpose for a comprehensive second-by-second data collection effort, especially when the existing data can be used to drastically improve our current ability to cope with microscopic flow changes. The answer lies in the fact that the current SCF data set is not representative of all vehicle technologies and vintages of the on-road fleet and that many other specific modeling issues cannot be addressed with the SCF data. For instance, grade effects, load effects and the closely related commanded enrichment events that occur in vehicles is poorly understood, and additionally is not adequately represented in the SCF data. These issues need to be specifically and poignantly addressed in any second-by-second data collection effort.

The results also suggest that for an interim improvement effort, the SCF data set could be revisited to produce far superior activity-based model algorithms. New activity-based algorithms could be employed to replace those within currently employed models, both regional and project level. Of course, we would need to see parallel and complementary improvements to transportation activity models, and enhanced data collection efforts to obtain the necessary modal activity data on transportation networks. Work currently underway, and funded by Caltrans, aims to quantify modal activity on different facility types under different levels of congestion in the Los Angeles, California region. A model improvement effort of this type could realistically be ready within a 2-year time frame, whereas alternative model improvement efforts based on second-by-second data will not be ready for about 5 years.

REFERENCES

Austin, T.C., Carlson, T.R., DiGenova, F.J., Lee, J.M. and Carlock, M. (1992). An Analysis of Driving Patterns in Los Angeles During 1992, *Presented at the Third Annual CRC-APRAC On-Road Vehicle Emissions Workshop*, Coordinating Research Council, Atlanta, GA, September 1992.

Benson, P. (1989). "CALINE4 - A Dispersion Model for Predicting Pollutant Concentrations Near Roadways". (FHWA/CA/TL-84/15); State of California Department of Transportation, Division of New Technology and Research; Sacramento, CA; November 1984, Revised June 1989.

Bevington, P. and Robinson, D. (1992). "Data Reduction and Error Analysis for the Physical Sciences; Second Edition". McGraw-Hill, Inc.

Bowyer, D.P., Akçelik, R. and Biggs, D.C. (1985). "Guide to Fuel Consumption Analyses for Urban Traffic Management", Australian Road Research Board, Special Report No. 32.

CARB, California Air Resources Board (1992b). Methodology to Calculate Emission Factors for On-Road Motor Vehicles. Technical Support Division, Sacramento CA.

Darlington, T.L., Korsog, P.E. and Strassburger, R. (1992). Real World and Engine Operation: Results of the MVMA/AIAM Instrumented Vehicle Pilot Study, *Proceedings of the 85th Annual Meeting of the Air and Waste Management Association*, AWMA, Pittsburgh, PA, June 1992.

Fair, R. (1984). "Specification and Analysis of Macroeconomic Models". Cambridge, Massachusetts: Harvard University Press.

Greene, W.H. (1990). "Econometric Analysis". Macmillan Publishing Company, 866 Third Avenue, New York.

Guensler, R. (1993). "Vehicle Emission Rates and Average Vehicle Operating Speeds". A dissertation submitted to the University of California at Davis Graduate School, Davis CA 95616.

Heywood, J.B. (1988). Internal Combustion Engine Fundamentals; McGraw-Hill Publishing Company, New York.

Kelly, N.A and Groblicki, P.J. (1993). Real World Emissions from a Modern Production Vehicle Driven in Los Angeles, Air and Waste Management, 43, October 1993, pp. 1351-1357.

LeBlanc, D.C., Meyer, M.D., Saunders, F.M. and Mullholland, J.A. (1994). Carbon Monoxide Emissions from Road Driving: Evidence of Emissions Due to Power Enrichment, *Proceedings of the 1994 Annual Meeting of the Transportation Research Board*, Paper 561, Georgia Institute of Technology, Atlanta, GA, January 1994.

Neter, Wasserman and Kutner (1990). *Applied Linear Models*. Richard D. Irvin, Inc.

Smith, M., Guensler, R. and Washington, S. (1994). Statistical Representation of the Speed Correction Factor Database: More Bad News. Draft Report, Institute of Transportation Studies, University of California at Davis.

USDOT, US Department of Transportation, Federal Highway Administration, Federal Transit Administration (1993). 23 Code of Federal Regulations, Part 450 and 49 Code of Federal Regulations, Part 613, "Statewide Planning; Metropolitan Planning; Rule". Published in the Federal Register, Vol. 58, No 207, October 28, 1993.

Washington, S. (1994). "Estimation of a Vehicular Carbon Monoxide Modal Emissions Model and Assessment of an Intelligent Transportation Technology". Ph.D. Dissertation. University of California at Davis, Institute of Transportation Studies, Report # UCD-ITS-RR-94-29.

Washington, S., Guensler, R. and Sperling, D. (1994). "Modeling IVHS Emission Impacts, Volume II: Assessment of the CALINE 4 Line Source Dispersion Model". PATH Research Report, July 28, 1994, Institute of Transportation Studies, University of California at Davis.