

# Development of VT-Micro Model for Estimating Hot Stabilized Light Duty Vehicle and Truck Emissions

Hesham Rakha<sup>1</sup>, Kyoungho Ahn<sup>2</sup>, and Antonio Trani<sup>3</sup>

*Department of Civil and Environmental Engineering and Virginia Tech Transportation Institute, Virginia Tech, Blacksburg, Virginia 24061-0536, USA*

---

## Abstract

The paper presents and applies a framework for developing microscopic emission models (VT-Micro model version 2.0) for assessing the environmental impacts of transportation projects. The original VT-Micro model was developed using chassis dynamometer data on nine light duty vehicles. In this paper, the VT-Micro model is expanded by including data from 60 light duty vehicles and trucks. Statistical clustering techniques are applied to group vehicles into homogenous categories. Specifically, Classification and Regression Tree algorithms are utilized to classify the 60 vehicles into 5 LDV and 2 LDT categories. In addition, the framework accounts for temporal lags between vehicle operational variables and measured vehicle emissions. The VT-Micro model is validated by comparing against laboratory measurements with prediction errors within 17 percent.

*Keywords:* VT-Micro, Vehicle emission, Microscopic, Modeling.

---

## 1. Introduction

Estimating accurate mobile source emissions has gained interest among transportation professionals as a result of increasing environmental problems in large metropolitan urban areas. While current emission inventory models, such as MOBILE and EMPAC, are capable of estimating large scale inventories, they are

---

<sup>1</sup> Associate Professor. Charles Via Department of Civil and Environmental Engineering. Virginia Tech Transportation Institute, 3500 Transportation Research Plaza (0536), Blacksburg VA, 24061. E-mail: hrakha@vt.edu.

<sup>2</sup> Senior Research Associate. Virginia Tech Transportation Institute, 3500 Transportation Research Plaza (0536), Blacksburg VA, 24061. E-mail: kahn@vt.edu.

<sup>3</sup> Associate Professor. Charles Via Department of Civil and Environmental Engineering. Patton Hall, Blacksburg VA, 24061. E-mail: vuel@vt.edu.

unable to estimate accurate vehicle emissions that result from operational-level projects. Alternatively, microscopic emission models are capable of assessing the impact of transportation projects on the environment and performing project-level analyses. This paper presents an extension to the original VT-Micro model that was developed using chassis dynamometer data on nine light duty vehicles. In this paper, the VT-Micro model is expanded by including data from 60 light duty vehicles and trucks and considers non steady-state emission data in developing the models.

The objective of the paper is to develop a framework for estimating vehicle emissions using non-steady state emission data. The framework develops emission models by aggregating data based on vehicle and operational variables. Specifically, statistical techniques for aggregating vehicles into homogenous categories are utilized as part of the framework. In addition, the framework adjusts for temporal lags between vehicle operational variables and vehicle emissions for non-steady state emission measurement. This framework is then utilized to develop the VT-Micro model version 2.0.

The research that is described in this paper provides several significant contributions. First, the paper introduces a comprehensive microscopic energy and emission modeling framework which can be utilized for developing microscopic energy and emission models using second-by-second fuel consumption and emission data. This framework presents the use of statistical tools to categorize emission data based on vehicle-specific variables, develops procedures for operational variable binning, and develops techniques for adjusting for the temporal lag between operational variables and their associated vehicle emissions. Secondly, the research develops a microscopic energy and emission model, the VT-Micro version 2.0, which can be easily incorporated within microscopic traffic simulation software to evaluate the environmental of operational-level projects.

## **2. State-of-the-Art Microscopic Vehicle Emission Models**

This section briefly describes and reviews the state-of-the-art microscopic mobile source emission models in North America. Specifically, microscopic models estimate vehicle pollutants at a second-by-second level of resolution using either vehicle engine or vehicle speed/acceleration data. Two models that are emerging

include the Comprehensive Modal Emissions Model (CMEM) and the Virginia Tech Microscopic (VT-Micro) model. These models are briefly described in terms of their structure, logic, and validity.

### *2.1 Comprehensive Modal Emission Model*

The Comprehensive Modal Emissions Model (CMEM), which is one of the newest power demand-based emission models, was developed by researchers at the University of California, Riverside (Barth et al., 2000). The CMEM model estimates LDV and LDT emissions as a function of the vehicle's operating mode. The term "comprehensive" is utilized to reflect the ability of the model to predict emissions for a wide variety of LDVs and LDTs in various operating states (e.g., properly functioning, deteriorated, malfunctioning).

The development of the CMEM model involved extensive data collection for both engine-out and tailpipe emissions of over 300 vehicles, including more than 30 high emitters. These data were measured at a second-by-second level of resolution on three driving cycles, namely: the Federal Test Procedure (FTP), US06, and the Modal Emission Cycle (MEC). The MEC cycle was developed by the UC Riverside researchers in order to determine the load at which a specific vehicle enters into fuel enrichment mode. CMEM predicts second-by-second tailpipe emissions and fuel consumption rates for a wide range of vehicle/technology categories. The model is based on a simple parameterized physical approach that decomposes the entire emission process into components corresponding to the physical phenomena associated with vehicle operation and emission production. The model consists of six modules that predict engine power, engine speed, air-to-fuel ratio, fuel use, engine-out emissions, and catalyst pass fraction. Vehicle and operation variables (such as speed, acceleration, and road grade) and model calibrated parameters (such as cold start coefficients, engine friction factor) are utilized as input data to the model.

Vehicles were categorized in the CMEM model based on a vehicle's total emission contribution. Twenty-eight vehicle categories were constructed based on a number of vehicle variables. These vehicle variables included the vehicle's fuel and emission control technology (e.g. catalyst and fuel injection), accumulated mileage, power-to-weight ratio, emission certification level (tier0 and tier1), and emitter level category (high

and normal emitter). In total 24 normal vehicle and 4 high emitter categories were considered (Barth *et al.*, 2000).

## 2.2 The Virginia Tech Microscopic Energy and Emission Model (VT-Micro Model)

The VT-Micro model was developed from experimentation with numerous polynomial combinations of speed and acceleration levels. Specifically, linear, quadratic, cubic, and quartic terms of speed and acceleration were tested using chassis dynamometer data collected at the Oak Ridge National Laboratory (ORNL). The final regression model included a combination of linear, quadratic, and cubic speed and acceleration terms because it provided the least number of terms with a relatively good fit to the original data ( $R^2$  in excess of 0.92 for all Measures of Effectiveness (MOE)). The ORNL data consisted of nine normal emitting vehicles including six light-duty automobiles and three light duty trucks. These vehicles were selected in order to produce an average vehicle that was consistent with average vehicle sales in terms of engine displacement, vehicle curb weight, and vehicle type. The data collected at ORNL contained between 1,300 to 1,600 individual measurements for each vehicle and MOE combination depending on the envelope of operation of the vehicle, which has a significant advantage against emission data collected from few driving cycles since it is impossible to cover the entire vehicle operational regime with only a few driving cycles. Typically, vehicle acceleration values ranged from  $-1.5$  to  $3.7 \text{ m/s}^2$  at increments of  $0.3 \text{ m/s}^2$  ( $-5$  to  $12 \text{ ft/s}^2$  at  $1 \text{ ft/s}^2$  increments). Vehicle speeds varied from  $0$  to  $33.5 \text{ m/s}$  ( $0$  to  $121 \text{ km/h}$  or  $0$  to  $110 \text{ ft/s}$ ) at increments of  $0.3 \text{ m/s}$ .

Detailed descriptions of the model derivation are provided in the literature (Rakha *et al.*, 2000b, Ahn *et al.*, 2002). The first regression model that was tested included an enhancement to the Post and Akcelik models by introducing more variables, as demonstrated in Equation 1:

$$MOE_e = \sum_{i=0}^3 \sum_{j=0}^3 (K_{i,j}^e \times u^i \times a^j) \quad [1]$$

Where  $MOE_e$  is the instantaneous fuel consumption or emission rate (l/s in the case of fuel consumption and mg/s in the case of vehicle emissions),  $K_{i,j}^e$  is the regression model coefficient for MOE “e” at speed power

“ $i$ ” and acceleration power “ $j$ ”,  $u$  is the instantaneous vehicle speed (km/h), and  $a$  is the instantaneous vehicle acceleration (km/h/s).

The model produced reasonable fits to the original data except in a few instances where the model produced negative dependent variable values. To address this problem, a data transformation technique using the natural logarithm was adopted resulting in the new log-transformed model that is presented in Equation 2:

$$MOE_e = e^{\sum_{i=0}^3 \sum_{j=0}^3 (K_{i,j}^e \times u^i \times a^j)} \quad [2]$$

The coefficient of determination of the MOE estimates using Equation 2 ranged from 0.69 to 0.99. The coefficient of determination was computed from the correlation coefficient between the original data and the model predicted values, not from the regression itself. The statistical results indicated a good fit for fuel consumption ( $R^2 = 0.995$ ) and  $NO_x$  estimates ( $R^2 = 0.960$ ) and a relatively poor fit for HC and CO emission estimates ( $R^2 = 0.689$  and  $0.717$ , respectively).

In order to isolate and identify the shortcomings of the log-transformed polynomial regression models, Figure 1 illustrates graphically the quality of fit between the regression models and the ORNL data.

Figure 1 also demonstrates that the errors in the HC and CO model estimates are significant at high acceleration levels (overestimates HC emissions by up to 25 percent and CO emissions by 100 percent). These errors in the regression model estimates are caused by the significant sensitivity of the dependent variable to the independent variables at high accelerations compared with the marginal sensitivity of the dependent variable in the negative acceleration range. Differences in behavior for positive versus negative accelerations can be attributed to the fact that in positive accelerations the vehicle exerts power, while in the negative acceleration range the vehicle does not exert power.

Consequently, separate regression models were developed for positive and negative accelerations, as demonstrated in Equation 3:

$$MOE_e = \begin{cases} e^{\sum_{i=0}^3 \sum_{j=0}^3 (L_{i,j}^e \times u^i \times a^j)} & \text{for } a \geq 0 \\ e^{\sum_{i=0}^3 \sum_{j=0}^3 (M_{i,j}^e \times u^i \times a^j)} & \text{for } a < 0 \end{cases} \quad [3]$$

Where  $L_{i,j}^e$  and  $M_{i,j}^e$  represent model regression coefficients for MOE “e” at speed power “i” and acceleration power “j”. It should be noted that the intercept at zero speed and zero acceleration was estimated using the positive acceleration model and fixed in order to ensure a continuous function between the two regression regimes. The final models that were developed resulted in good fits to the ORNL data, as demonstrated in Figure 2 ( $R^2$  in excess of 0.92 for all MOEs). Figure 2 further illustrates the effectiveness of the hybrid log-transformed model in predicting vehicle fuel consumption and emission rates as a function of a vehicle’s instantaneous speed and acceleration levels. A comparison of Figure 1 and Figure 2 clearly demonstrates the enhancement in model predictions as a result of separating the positive and negative acceleration regimes.

### 2.3 Model Comparison and Validation

A comprehensive study compared the MOBILE5a, VT-Micro, and CMEM models for estimating hot stabilized light duty vehicle emissions (Brzezinski et al., 1999). The study concluded that the original power based models, namely the Post and Akcelik models, do not provide good Measure of Effectiveness (MOE) estimates when compared to field data. The paper also demonstrated that the log-transformed dual-regime 3<sup>rd</sup> order polynomial model structure, which is utilized within the VT-Micro model, predicts fuel consumption and emissions within a small margin of error with respect to field data (coefficient of determination greater than 92 percent for all MOEs). The study also demonstrated that the CMEM model exhibits some abnormal behaviors. First, the model estimates identical MOE estimates for speeds of 0 km/h regardless of the acceleration rate. Second, the CMEM model estimates constant MOE estimates during deceleration maneuvers. Third, the CMEM model generally underestimates MOEs for acceleration maneuvers when compared to laboratory data. Forth, the CMEM model CO emission estimates exhibit strange behavior at low speeds and high acceleration levels (sudden drops of emissions). Finally, the CMEM model NO<sub>x</sub> emissions do not exhibit the typical decay in emission rates at high engine loads. Consequently, the paper

concluded that the VT-Micro model appeared to be accurate in terms of absolute light-duty hot stabilized normal vehicle tailpipe emissions. Specifically, the emission estimates were found to be within the 5<sup>th</sup> and 95<sup>th</sup> percentile confidence limits of field data and within the same level of magnitude as the MOBILE5a model estimates. Furthermore, the VT-Micro model was found to reflect differences in drive cycles in a fashion that was consistent with laboratory measurements, which was not the case for the CMEM model.

### **3. Emission Data Description**

This section describes the data that were utilized to develop and test the VT-Micro framework for emission modeling. The data are described in terms of the data collection procedures, the vehicles tested, and the drive cycles that were utilized for data collection.

The data were gathered by EPA on a chassis dynamometer at the Automotive Testing Laboratories, Inc. (ATL), in Ohio and EPA's National Vehicle and Fuels Emission Laboratory (NVREL), in Ann Arbor, Michigan in the spring of 1997. All vehicles at ATL were drafted as a stratified random sample at Inspection and Maintenance lanes utilized by the State of Ohio. The vehicles tested at the EPA laboratory were recruited randomly. All vehicles were tested under as-received condition (without repairs). Of the total 101 vehicles 62 vehicles were tested at ATL and 39 vehicles were tested at NVREL. Unfortunately, the 101 vehicle sample size was reduced to 96 vehicles because of insufficient data for 5 vehicles. All vehicles were tested at FTP under ambient conditions using the standard vehicle certification test fuel. Vehicle emission tests were performed in random order to offset any possible order bias that could result in different ambient conditions for the tested cycles. The HC, CO, NO<sub>x</sub>, and CO<sub>2</sub> emissions were measured as composite "bags" and in grams on a second-by-second basis (Brzezinski et al., 1999).

The 96 vehicles that were tested included model years that ranged from 1986 through 1996. These vehicles were initially screened in order to separate normal from high emitting vehicles using a threshold that was set at two times the manufacturer standards for HC and NO<sub>x</sub> and three times the manufacturer standards for CO. The details of the procedures that were utilized to screen the high emitting vehicles are beyond the scope of this paper but will be described in a forthcoming publication. Of the total sample size of 96 vehicles, 60

vehicles were classified as normal vehicles and 36 were classified as high emitting vehicles. Also, among the 60 normal vehicles, 43 vehicles were LDVs and the remaining 17 vehicles were LDTs, as illustrated in Figure 3. The 60 vehicles included 42 vehicles with automatic transmission and 18 vehicles with manual transmission. All 60 vehicles used fuel injection gasoline engines that ranged in size from 1.0 to 5.8 liters, with the majority of vehicles in the 2.0 to 4.0 liter range. The majority of vehicles had a mileage less than 160,000 kilometers (100,000 miles).

The EPA developed new facility-specific and area-wide drive cycles based on real-world driving studies. These drive cycles have been incorporated in the MOBILE6 model (Brzezinski et al., 1999). In order to represent real-world driving conditions, the driving behavior in each drive cycle was developed using observed speed-acceleration profiles and specific power frequency distributions of chase car driving data from a range of roadway types during periods of various congestion levels. These drive cycles cover four roadway types that include: Freeways, Arterial/Collectors, Freeway Ramps, and Local Roadways. The roadways are further classified based on Level of Service (LOS) measures, similar to the transportation congestion index, from "A" to "G." Table 1 provides a brief description of the new EPA cycles and some additional drive cycles used for emission testing. The additional cycles include the LA4 (urban dynamometer driving cycle), the California Air Resources Board (CARB) area-wide Unified Cycle (LA92), the New York City Cycle (a low speed cycle which was previously used for speed correction factors in the MOBILE5 model), and ST01 (the first 258 seconds of the vehicle certification air conditioning cycle). Of the total 17 drive cycles, each vehicle was tested over 14 to 16 drive cycles. These emission data collected from 14 to 16 drive cycles represent a wide range of vehicle driving behaviors on real roads. It should be noted that the ST01 drive cycle was not utilized for model validation because it includes cold start emissions. The maximum speed of all cycles was 119.5 km/h on the Freeway High Speed Cycle, while the maximum acceleration rate was 11.04 km/h/s ( $3.07 \text{ m/s}^2$ ) on the Freeway, LOS F and LA92 cycles. It should be noted that the New York cycle had the lowest average speed of all drive cycles (11.4 km/h).

## **4. Model Development**

This section describes the framework that was utilized to develop the VT-Micro model. Specifically, the framework involves two major classification efforts, as illustrated in Figure 4. The first effort aggregates vehicles into categories that are similar in characteristics and technologies. The second effort aggregates data based on operation-based variables. The VT-Micro approach is to utilize instantaneous vehicle speed and acceleration levels as operation-based variables. The use of speed and acceleration as operation-based variables allows the model to be utilized in conjunction with Global Positioning System (GPS) data for the field evaluation of environmental impacts of operational-level transportation projects (Rakha et al., 2000a). Furthermore, these emission models can be incorporated within microscopic traffic simulation software (Rakha et al., 2000b). The use of vehicle dynamics models within microscopic simulation software is critical in order to ensure that vehicle accelerations are accurate. These models can also account for grade effects on vehicle emissions by accounting for the additional acceleration in the direction of motion of the vehicle as a result of the grade.

The use of second-by-second emission data requires that data be temporally offset in order to ensure that there is no time lag between speed/acceleration observations and emission measurements. The proposed framework develops a procedure for offsetting the data and demonstrates how this procedure was applied to the EPA data. Finally, the section describes how the models were developed using the EPA data.

### *4.1 Vehicle Classification*

Vehicle classification is a fundamental procedure required in developing mobile source emission models. For example, depending on vehicle characteristics, such as vehicle model year, engine technology, engine size, and vehicle mileage, the amount and pattern of vehicle emissions can vary significantly. Traditional emission models such as MOBILE and EMFAC categorize vehicles into common vehicle groups that include light-duty gasoline vehicles (LDGV), light-duty trucks (LDT), and heavy-duty trucks (HDT). While this classification is important, it only serves as a first step in categorizing vehicles, as illustrated in Figure 4. The proposed framework includes three levels of vehicle classifications. The first level involves categorizing

vehicles based on whether a vehicle is a heavy duty truck, light duty truck, or light duty vehicle. The second level involves classifying vehicles based on whether the vehicle is diesel or gasoline powered. The final classification involves the use of statistical Classification and Regression Tree (CART) algorithms to further group vehicles into categories that are similar in their emission characteristics.

The CART algorithm is a data-mining technique that uses a regression tree method that automatically searches for important patterns and relationships and quickly finds hidden structures in highly complex data. Tree structured classifiers or binary tree structured classifiers are built by repeating splits at active nodes. An active node is divided into two sub-nodes based on a split criterion and a split value. The splitting process is generally continued until (a) the number of observations in a child node has met minimum population criteria or (b) minimum deviance criteria at a node are met, where the deviance criteria  $D$  is defined as the Sum of Squared Error (SSE), as computed in Equations 4 and 5.

The CART algorithm was utilized to classify the 60 normal test vehicles into a number of categories that were similar in emission behavior. The dependent variable ( $Y$ ) was a 60-by-4 matrix that included 4 dependent variables for 60 normal vehicles. The dependent variables included HC, CO, CO<sub>2</sub>, and NO<sub>x</sub> emissions averaged over all 15 drive cycles. Similarly, the independent variable ( $X$ ) was a 60-by- $n$  matrix that included a number of vehicle attributes, including the vehicle model year, engine technology, engine size, and vehicle mileage. Alternatively, the  $X$  matrix can be thought of as a set of vectors  $X_k$ , each composed of 60 elements, where  $k$  is the vehicle attribute index under consideration in the CART algorithm.

Within the CART algorithm, the observations in  $Y$  are divided on an independent variable  $X_k$  resulting in two children nodes,  $a_1$  and  $a_2$ , each containing  $n_1$  and  $n_2$  observations of the original  $n$  observations ( $n_1+n_2 = n$ ) using the deviance criterion that is presented in Equation 6. The SSE for all observations at node  $a$  is calculated using Equation 7, while the error at each of the sub-nodes is calculated using Equations 8 and 9. The problem is then formulated as an optimization problem in which the objective function is to maximize the deviance reduction function (Equation 6) by solving for the variable index  $k$  and the value of  $X_k$  for partitioning  $Y$ . Once this splitter is found, the original data at node  $a$  are separated into two children nodes  $a_1$

and  $a_2$  each having minimal combined deviance compared with all other possible nodes. Several numerical search procedures are used to maximize the deviance reduction. The most common splitting function is the ‘‘Gini’’ followed by the ‘‘Twoing’’ function. While a detailed description of the CART algorithm is beyond the scope of this paper, the reader is referred to a number of sources in the literature (Breiman, 1984, Wolf et al., 1998, Insightful, 2001).

$$D = \sum_{i=1}^n (Y_i - \mu)^2 \quad [4]$$

$$\mu = \frac{1}{n} \sum_{i=1}^n Y_i \quad [5]$$

$$\Delta = D_a - (D_{a_1} + D_{a_2}) \quad [6]$$

$$D_a = \sum_{i=1}^n [Y_i - \mu_a]^2 \quad \forall i \in a \quad [7]$$

$$D_{a_1} = \sum_{i=1}^{n_1} [Y_i - \mu_{a_1}]^2 \quad \forall i \in a_1 \quad [8]$$

$$D_{a_2} = \sum_{i=1}^{n_2} [Y_i - \mu_{a_2}]^2 \quad \forall i \in a_2 \quad [9]$$

Where  $D$  is the total deviance of  $Y$ , or the sum of squared errors (SSE),  $Y_i$  is  $i^{\text{th}}$  observation in  $Y$ ,  $n$  is the sample size over which  $D$  is calculated ( $n = N$  for a sample equal to the population),  $\mu$  is the arithmetic mean of  $Y$ , and  $\Delta$  is deviance reduction when parent node  $a$  is partitioned on  $X_i$  to obtain children nodes  $a_1$  and  $a_2$ .

The CART algorithm was applied to the data using the S-PLUS software, which allows the user to input a set of dependent variables, a set of independent variables, and a minimum number of observations within a category. In conducting the vehicle categorization, all test vehicles were initially divided into LDV and LDT vehicles because, as was discussed earlier, LDVs and LDTs have significantly different emitting characteristics. LDV vehicles were further categorized using the CART algorithm considering a number of independent variables that included the vehicle model year, engine size, vehicle mileage, vehicle power-to-weight ratio, and federal emission standard (tier0 and tier1). After running the S-PLUS software, it was

found that only three vehicle characteristics were selected for vehicle aggregation, namely: vehicle model year, engine size, and vehicle mileage, as summarized in Table 2 and Figure 5. These three variables resulted in five vehicle categories. It should be noted that the minimum number of vehicles within a category was set at 5. As indicated in Table 2 two model year breakpoints were selected by the CART algorithm, namely 1990 and 1995. It is interesting to note that all post 1990 model year vehicles were equipped with fuel injection technology, which suggests, as would be expected, that fuel injection technology reduced vehicle emissions. Also, federal emission standard, tier1 was introduced in 1994 (60% tier0 and 40% tier1) and became more prevalent in 1995 (20% tier0 and 80% tier1, 100% tier1 in 1996). This again demonstrates that the model year that was selected by the CART algorithm corresponded to the introduction of tier1 vehicles into the vehicle fleet.

In the case of LDT categorization, the independent variables that were considered included the vehicle model year, engine size, vehicle mileage, vehicle gross weight, and federal emission standard (tier0 and tier1). The use of vehicle gross weight instead of power-to-weight ratio was considered in order to ensure consistency with the MOBILE and CMEM model procedures. After running the software, it was found that only one vehicle characteristic, namely vehicle model year was chosen for LDT classification, generating two LDT vehicle groups, as summarized in Table 2 and Figure 6.

Table 3 summarizes the mean, standard deviation, and Coefficient of Variation (COV), computed as the ratio of standard deviation to mean, for all 15 drive cycles across all vehicle classes that were established by the CART algorithm. The results indicate that the variations within a class are reasonably small. Alternatively, Table 4 summarizes the mean, standard deviation, and COV based on the CMEM model categories for the same 60 vehicles. The CMEM classification includes 6 LDV categories (category 4, 5, 6, 7, 8, and 11) and 3 LDT categories (category 16, 17, and 18). Comparing Table 3 and Table 4 a number of conclusions can be drawn. First, the CMEM model categories can result in categories that vary considerably in sample size (ranging from 1 to 21 vehicles within a category), whereas the CART algorithm ensures that the sample size is not less than a user specified minimum (in this case it was set at 5 vehicles). Second, it is noted that

CMEM classification has slightly higher COVs in comparison with the CART algorithm classifications. For example, Category 5 (the largest dataset in the CMEM classification) has HC, CO, NO<sub>x</sub>, and CO<sub>2</sub> COVs that are 0.95, 0.78, 0.95, and 0.16 while the CART algorithm LDV2 (the largest dataset in the CART classification) has COV values of 0.56, 0.40, 0.65, and 0.10, respectively.

Consequently, it was concluded that the CART algorithm is an effective tool for aggregating vehicles into homogeneous categories while maintaining a minimum number of vehicles within a category.

#### *4.2 Temporal Normalization of Data*

The use of second-by-second emission data requires that data be temporally offset in order to ensure that there is no time lag between speed/acceleration observations and emission measurements. Analysis of second-by-second laboratory and on-board field data by the authors has revealed lags between operational variables and vehicle emissions in the range of 5 to 10 seconds. This section describes how the emission measurements were offset in order to remove any temporal lags between operation variables and emission measurements.

The EPA data were in the format of a database containing approximately 1,283,000 second-by-second records. Each record included the test site unique ID, the vehicle ID, the driving cycle, the time stamp since the start of the driving cycle, the instantaneous speed, and the instantaneous emission rate of HC, CO, CO<sub>2</sub>, and NO<sub>x</sub>, and other related data. As mentioned earlier, these data covered between 14 and 16 drive cycles.

The time lag is a very important factor in developing instantaneous emission models in order to accurately associate vehicle emissions with their corresponding operational variable values. In identifying the temporal lag, the ORNL composite vehicle model (original VT-micro model) was utilized in conjunction with the second-by-second speed measurements to estimate instantaneous vehicle emission levels. It should be noted that the ORNL data did not include a time-lag, given that the operation variables (vehicle speed and acceleration measurements) were held constant for 15 seconds. A sum of squared error between the estimated second-by-second emissions and the measured emissions was utilized to compute the optimum

temporal shift in the emission data. Therefore, the smallest sum of squared error between 0 second and 15 seconds was selected as the time lag for the vehicle-drive cycle combination. The optimum temporal shift was found to range from 6 to 8 seconds for the NVREL data while the offset ranged from 0 to 1 second for the ATL data, which suggests that the ATL data had been temporally normalized as part of the data collection effort.

#### *4.3 Model Construction*

This section describes how the emission models were developed using regression procedures. As was described earlier, a nonlinear multi-dimensional dual-regime polynomial model structure was utilized. This multiple regression model relates the dependent variable (instantaneous emission measurement) to a set of independent variables that include the vehicle's instantaneous speed and acceleration levels. In total 5 LDV and 2 LDT models were developed using the EPA data.

A first step in developing the VT-Micro model was to create a binning procedure in order to reduce the noise in the data and to reduce the number of raw data points. Specifically, emissions of all vehicles within a vehicle category and in a speed/acceleration bin were averaged to generate a single average emission estimate. The speed bins included vehicle speeds ranging from 0 to 120 km/h at increments of 1 km/h, while acceleration bins included accelerations that ranged from -6 to 10 km/h/s ( $-1.7$  to  $2.8$  m/s<sup>2</sup>) at increments of 1 km/h/s. Figure 7 illustrates a sample speed/acceleration frequency distribution for one of the 60 test vehicles over all 15 drive cycles with 5 km/h and 1 km/h/s increments. As illustrated in the figure, the majority of the speed and acceleration data occur during steady-state operation (acceleration ranging between -1 and 1 km/h/s).

A sample data set for LDT1 is presented in Figure 8 for illustration purposes. The figure clearly demonstrates the large nonlinear behavior in all emissions as a function of the vehicle speed and acceleration. In addition, it is evident that, as vehicle acceleration and speed levels increase, the emissions generally tend to increase. Furthermore, it is noted that the gradient of the emissions in the negative acceleration regime is generally smaller than that in the positive acceleration regime. Furthermore, the figure

clearly demonstrates a high level of emission variability at high speed and high acceleration levels, which could be attributed to the small sample size of data points at these high engine loads or, alternatively, could be a result of variability in vehicles reaching fuel enrichment mode.

The regression model uses the VT-micro model structure, which combines linear, quadratic, and cubic speed and acceleration terms, as expressed in Equation 3. The emission model utilizes a logarithmic transformation of the second-by-second emission data in order to ensure that emission estimates are non-negative. Another reason for the use of a logarithmic transformation is to enhance the model accuracy in the low speed and low acceleration regime, which generally results in low emission rates. The logarithmic transformation allows the model to respond rapidly to changes in emissions as a function of speed/acceleration levels. It should also be noted that the emission model employs a hybrid model that separates positive acceleration from negative acceleration regimes.

The majority of microscopic emission models assume a constant emission rate when a vehicle is decelerating. However, as shown in Figure 9, emission rates typically increase as vehicle speeds increase, even in a vehicle deceleration mode of operation. The proposed dual-regime model enhances the accuracy of the emission model in deceleration mode, which is typically neglected and overlooked by other models. A detailed comparison of single-regime and dual-regime is described in the literature (Ahn et al., 2002). Since the dual-regime model introduces a discontinuity between the acceleration and deceleration modes of operation, an attempt was made to reduce this level of discontinuity by ensuring the regression model constant is identical in both regimes. It should be noted that the models were confined to speed and acceleration levels within the envelope of the EPA data. This limitation resulted from the inherent limitation of any model to extrapolate response values beyond the boundaries used in the model calibration procedure. While most vehicles can travel faster than 121 km/h (upper limit of the testing boundary), it is impossible to establish a reliable forecasting pattern for energy and emission rates at high speeds due to the heavy nonlinear nature of the response curves. It should be noted that all speed and acceleration profiles in the EPA driving cycles used in the modeling were generally sufficient for typical vehicle operations. However,

in cases in which a vehicle exceeds the boundary, it is recommended to use the boundary speed and acceleration levels in order to ensure realistic vehicle MOE estimates. Furthermore, it should be noted that these models have been successfully applied to Global Positioning System (GPS) speed measurements after applying robust smoothing techniques in order to ensure feasible speed measurements (Rakha et al., 2000a).

EPA data did not include second-by-second fuel consumption data. However, the fuel consumption was computed using carbon balance equations, as demonstrated in Equation 10. Given that ambient air does not include carbon, whatever carbon enters the engine as fuel will leave the engine as emissions such as HC (g/s), CO (g/s), and CO<sub>2</sub> (g/s). Given that the molecular weight of carbon is 12 g/molecule and the molecular weight of oxygen is 16 g/molecule the molecular weight of CO<sub>2</sub> can be calculated to be 44 g/molecule (12+16x2). Therefore, CO<sub>2</sub> contains 27.3 percent (12/44) carbon. Similarly, the molecular weight of CO is 28 g/molecule (12+16) and there is 42.9 percent carbon in CO. Also, according to the Code of Federal Regulations Title 40 Part 86 (40 CFR 86), HC emissions contain 86.6 percent carbon by weight. Thus, the emissions of carbon, in grams per second can be computed using Equation 10. In addition, recognizing that average gasoline sold in US contains 86.4 percent of carbon, and has a density of 738.8 g/liter (or 2800 g/gallon); there are 638.31 (0.864x738.79) grams of carbon in a liter of gasoline. Consequently, the fuel consumption rate can be computed using Equation 11.

$$C = 0.866 HC + 0.429 CO + 0.273 CO_2 \quad [10]$$

$$F = \frac{0.866 HC + 0.429 CO + 0.273 CO_2}{638.31} \quad [11]$$

Where  $C$  is instantaneous carbon emissions rate (g/s) and  $F$  is the instantaneous fuel consumption rate (l/s).

Figure 9 and Figure 10 illustrate sample model emission predictions superimposed on the raw data for Category LDT1 and LDV2. As illustrated in the figures, the variability in emission estimates typically increases at high speed and acceleration levels. The figures may demonstrate that the model tends to underestimate CO emissions at the high speeds at first glance, however, it should be noted that significant observations of low emission measurements are also observed. Noteworthy is the fact that the LDT1

emission rates are higher than the LDV2 emission rates, as would be expected. Sample model coefficients for estimating HC emission rates for the LDT1 vehicle category are summarized in Table 5.

Figure 11 illustrates how the predicted fuel consumption compared to the field data for vehicle category LDT2. Noteworthy is the fact that the shape of the fuel consumption chart is extremely similar in the appearance to the CO<sub>2</sub> behavior.

## **5. Model Validation**

The next step in the analysis was to validate the newly developed emission models against bag trip measurements and against instantaneous second-by-second measurements. Furthermore, the validation effort included the applicability of the proposed VT-Micro model with sample GPS speed data. The details of the validation effort are described in this section.

### *5.1 Aggregate Model Validation*

In an attempt to validate the model using aggregate emission data, the EPA data that were utilized in developing the models were utilized for validation purposes because other independent data were not available at the time of the study. While the use of the EPA data for model validation is not ideal, it does offer a number of benefits. First, the database includes many off-cycle (non-FTP) emission results over different facility types and therefore provides a good assessment of the quality of model estimates for different roadway types and different levels of congestion. Second, as was mentioned earlier, the EPA database was utilized for the development of EPA's MOBILE6 model. Consequently, such a comparison ensures that the VT-Micro model estimates are consistent with MOBILE6 emissions across different facility types and different levels of congestion.

Figure 12 and Figure 13 compare the VT-Micro model emission predictions against laboratory measurements for all 15 drive cycles. The 5<sup>th</sup> and 95<sup>th</sup> percentile emission rates are computed based on differences in vehicle emissions within a vehicle category. For example, the LDT1 category includes 11 vehicles while the LDV2 category includes 15 vehicles. The bar plots represent the predicted emission

estimates for composite vehicles LDT1 and LDV2, while the vertical line and small horizontal bar represent the 95<sup>th</sup> percentile, 5<sup>th</sup> percentile, and mean value of measured EPA emission data, respectively. The predicted emissions are computed as the sum of instantaneous predicted vehicle emissions along the entire drive cycle. Figure 12 and Figure 13 clearly illustrate a good fit between the model estimates and the laboratory measurements. Specifically, all predictions lie within the 95<sup>th</sup> and 5<sup>th</sup> percentile limits. Furthermore, the model estimates generally follow the mean category laboratory measurements. Specifically, the error in trip emissions does not exceed 14 percent across all 15 drive cycles for all four emissions.

### *5.2 Instantaneous Emission Model Validation*

The next step in validating the proposed models was to compare second-by-second HC, CO, CO<sub>2</sub>, and NO<sub>x</sub> measurements against instantaneous model estimates with the objective of identifying any shortcomings in the proposed models. Two vehicle categories, namely LDT1 and LDV2 were selected for comparison purposes. These vehicles encompass a light duty truck and a light duty vehicle category. Figure 14 illustrates the speed and acceleration profiles of the Arterial LOS A-B (ARTA) drive cycle, which involves several full and partial stops in addition to travel at a fairly high speed (in the range of 100 km/h). The ARTA drive cycle was developed by EPA within their facility-specific and area-wide drive cycles to be incorporated in the MOBILE6 model as was described earlier in section 3.3. The figure clearly demonstrates that the ARTA drive cycle involves a more aggressive and realistic driving behavior compared to the old driving cycles.

Figure 15 and Figure 16 illustrate the variation in the maximum, minimum, and mean value of the instantaneous vehicle emissions as measured on a dynamometer as it travels through the drive cycle. The maximum, minimum, and mean is computed based on all vehicles within a category (11 vehicles for LDT1 and 15 vehicles for LDV2) for all four pollutants. The dotted lines represent the mean measurements of vehicle emissions while the shaded grey region shows the max/min limits of emission values. The solid lines represent the model estimates of vehicle emissions based on the vehicle's instantaneous speed and acceleration levels. Figure 15 and Figure 16 illustrate that, in general, the model prediction lines follow the dotted lines (mean values) of the EPA vehicle emission measurements, and follow all the valleys and peaks

in vehicle emissions. Furthermore, the total vehicle emissions of HC, CO, CO<sub>2</sub>, and NO<sub>x</sub> for LDT1 as measured in the laboratory were 0.489, 10.805, 2.384, and 2162.7 grams. The estimated emissions based on the proposed hybrid model were 0.565, 11.188, 2.437, and 2155.4 grams, which correspond to a 16, 4, 2, and 0.3 percent difference in overall emissions for the entire cycle. Vehicle category LDV2 showed similar results, producing 17, 5, 7, and 0.4 percent errors in overall emissions.

It should be noted that the sensitivity of the model to instantaneous vehicle speed and acceleration levels offers a unique tool for assessing the environmental impacts of traffic improvement projects, including Intelligent Transportation System (ITS) technologies. Contrary to other state-of-the-art models that estimate vehicle fuel consumption and emission rates based on average roadway speeds, the proposed models capture the transient changes in a vehicle's speed and acceleration as it travels on a highway network.

### *5.3 GPS Speed Data Application*

As was mentioned earlier, the proposed model can be utilized in conjunction with GPS data to evaluate the environmental impacts of operational-level transportation projects. This section presents a sample GPS data application for purposes of illustration, while more detailed descriptions of such applications can be found in the literature (Rakha et al., 2000b, Rakha et al., 2000a).

The sample GPS instantaneous speed measurements and the corresponding computed instantaneous accelerations were utilized to evaluate the Scottsdale/Rural Road Signal Coordination Project, and more specifically, to evaluate the impacts of re-timing three traffic signals at the boundary between the cities of Tempe and Scottsdale as a part of the Phoenix Metropolitan Model Deployment Initiative. It should be noted that robust smoothing techniques were applied to the GPS speed measurements in order to remove erroneous speed measurements. The details of the smoothing techniques are described in the literature (Rakha et al., 2000b). Figure 17 illustrates the speed and acceleration profiles of the sample GPS data, which involves three full stops on a typical urban arterial. Figure 17 also illustrates the instantaneous HC emissions as estimated using the proposed model as the vehicle travels along the arterial. The figure clearly

demonstrates the uniqueness of the model for applications with GPS technologies to evaluate the environmental impacts of transportation projects. Specifically, the proposed models fully reflect differences in MOEs for trips with similar average speeds but involving different speed profiles. A more detailed description of the VT-Micro model application with GPS data is provided in the literature (Rakha et al., 2000a).

## **6. Study Conclusions**

The paper presents a framework for developing microscopic emission models (VT-Micro model version 2.0) for assessing the environmental impacts of transportation projects. The framework develops emission models by aggregating data using vehicle and operational variables. Specifically, statistical CART algorithms for aggregating vehicles into homogenous categories are utilized as part of the framework. In addition, the framework accounts for temporal lags between vehicle operational variables and vehicle emissions. Finally, the framework is utilized to develop the VT-Micro model version 2.0 utilizing second-by-second chassis dynamometer data for a total of 60 light duty vehicles and trucks. A total of 5 LDV and 2 LDT categories are developed as part of this research effort.

The ultimate expansion of this model is its implementation within microscopic traffic simulation software. These models can then be utilized to evaluate the environmental impacts of microscopic vehicle behaviors, such as ramp metering, traffic signal coordination, and alternative ITS strategies. Currently, the VT-Micro model has been implemented in the INTEGRATION software. Also, the model can be applied to estimate vehicle emissions using instantaneous GPS speed measurements.

## **Acknowledgements**

This research effort was sponsored by the Intelligent Transportation System (ITS) Implementation Center.

## References

- Ahn, K., Rakha, H., Trani, A., and Van Aerde, M., 2002. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *Journal of Transportation Engineering*, 128(2), 182-190.
- Barth, M., An, F., Younglove, T., Scora, G., Levine, C., Ross, M., and Wenzel, T., 2000. *Comprehensive modal emission model (CMEM), version 2.0 user's guide*, Riverside: University of California, Riverside.
- Breiman, L. (1984). *Classification and regression trees*, Belmont: Wadsworth International Group.
- Brzezinski, D., Enns, P., and Hart, C., 1999. Facility-Specific Speed Correction Factors. *MOBILE6 Stakeholder Review Document (M6.SPD.002)*, Ann Arbor: US EPA.
- Insightful, 2001. *S-Plus 6 guide to statistics volume 1 and volume2*, Seattle: Insightful Corporation.
- Rakha, H., Medina, A., Sin, H., Dion, F., Van Aerde, M., and Jenq, J., 2000a. Traffic signal coordination across jurisdictional boundaries: Field evaluation of efficiency, energy, environmental, and safety impacts. *Transportation Research Record*. No. 1727, 42-51.
- Rakha, H., Van Aerde, M., Ahn, K., and Trani, A. A., 2000b. Requirements for evaluating traffic signal control impacts on energy and emissions based on instantaneous speed and acceleration measurements. *Transportation Research Record*. No. 1738, 56-67.
- Rakha, H., Ahn, K., and Trani, A., 2003. Comparison of MOBILE5a, VT-MICRO, and CMEM models for hot-stabilized light duty gasoline vehicles. *Canadian Journal of Civil Engineering*. In Press.
- Wolf, J., Guensler, R., Washington, S., and Bachman, W. 1998. High-emitting vehicle characterization using regression tree analysis. *Transportation Research Record*, No.1641, 58-65.

## **List of Tables**

*Table 1: EPA's New Facility-Specific Drive Cycle Characteristics*

*Table 2: CART Based Vehicle Classification*

*Table 3: Variation of Vehicle Classification*

*Table 4: CMEM Vehicle Classification*

*Table 5: Sample Coefficients of Hybrid Regression Model (HC Emissions for LDT1 Vehicle)*

## **List of Figures**

*Figure 1: Regression Model Predictions (Composite Vehicle – Log-Transformed Polynomial Model)*

*Figure 2: Regression Model Predictions (Composite Vehicle – Log-Transformed Hybrid Polynomial Model)*

*Figure 3: EPA Test Vehicle Characteristics*

*Figure 4: Second-by-second Emission Data Binning Framework*

*Figure 5: Light Duty Vehicle Classification using CART Algorithm*

*Figure 6: Light Duty Truck Classification using CART Algorithm*

*Figure 7: Speed and Acceleration Distribution for Sample Vehicle (15 Drive Cycles)*

*Figure 8: Sample Normalized Emission Data (LDT1)*

*Figure 9: Model Prediction (LDT1)*

*Figure 10: Model Prediction (LDV2)*

*Figure 11: Fuel Consumption Prediction (LDT2)*

*Figure 12: Model Validation for 15 Driving Cycles (LDT1)*

*Figure 13: Model Prediction for 15 Driving Cycles (LDV2)*

*Figure 14: Speed and Acceleration Profiles for ARTA Driving Cycle*

*Figure 15: Instantaneous Model Validation for ARTA Cycle (LDT1)*

*Figure 16: Instantaneous Model Validation for ARTA Cycle (LDV2)*

*Figure 17: GPS Speed Data Application*

Table 1  
EPA's New Facility-Specific Drive Cycle Characteristics

Cycle	Avg. Speed (km/h)	Max. Speed (km/h)	Max. Accel. (km/h/s)	Duration (s)	Length (km)
Freeway, High Speed	101.12	119.52	4.32	610	17.15
Freeway, LOS A-C	95.52	116.96	5.44	516	13.68
Freeway, LOS D	84.64	112.96	3.68	406	9.54
Freeway, LOS E	48.8	100.8	8.48	456	6.18
Freeway, LOS F	29.76	79.84	11.04	442	3.66
Freeway, LOS G	20.96	57.12	6.08	390	2.27
Freeway Ramps	55.36	96.32	9.12	266	4.10
Arterial/Collectors LOS A-B	39.68	94.24	8	737	8.11
Arterial/Collectors LOS C-D	30.72	79.2	9.12	629	5.38
Arterial/Collectors LOS E-F	18.56	63.84	9.28	504	2.59
Local Roadways	20.64	61.28	5.92	525	2.99
Non-Freeway Area-Wide Urban Travel	31.04	83.68	10.24	1348	11.60
LA04	31.36	90.72	5.28	1368	11.92
Running 505	40.96	90.72	5.28	505	5.744
LA 92	39.36	107.52	11.04	1435	15.696
ST01	32.32	65.6	8.16	248	2.224
New York Cycle	11.36	44.32	9.6	600	1.888

Table 2  
 CART Based Vehicle Classification

Vehicle Category	Number of Vehicles in Category
<i>Category for Light Duty Vehicles</i>	
LDV1: Model Year < 1990	6
LDV2: 1990<=Model Year<1995, Engine Size < 3.2 liters, Mileage < 83653,	15
LDV3: Model Year >= 1995, Engine Size < 3.2 liters, Mileage < 83653,	8
LDV4: Model Year >=1990, Engine Size < 3.2 liters, Mileage >= 83653	8
LDV5: Model Year >=1990, Engine Size >= 3.2 liters	6
LDV High Emitters	24
<i>Category for Light Duty Trucks</i>	
LDT1: Model Year >= 1993	11
LDT2: Model Year < 1993	6
LDT High Emitters	12
<i>Total Vehicles</i>	96

Table 3  
Variation of Vehicle Classification

Category	Measure	HC	CO	NOx	CO2
LDV1 (6 vehicles)	Standard Deviation	0.123	2.167	0.367	26.958
	Mean (g/km)	0.185	3.351	0.555	181.848
	Coefficient of Variation	0.663	0.647	0.661	0.148
LDV2 (15 vehicles)	Standard Deviation	0.029	0.529	0.190	19.521
	Mean (g/km)	0.051	1.319	0.293	190.844
	Coefficient of Variation	<b>0.564</b>	<b>0.401</b>	<b>0.649</b>	<b>0.102</b>
LDV3 (8 vehicles)	Standard Deviation	0.013	0.385	0.125	35.601
	Mean (g/km)	0.023	0.664	0.129	201.097
	Coefficient of Variation	0.556	0.580	0.966	0.177
LDV4 (8 vehicles)	Standard Deviation	0.209	1.466	0.317	41.181
	Mean (g/km)	0.154	3.203	0.341	188.847
	Coefficient of Variation	1.358	0.458	0.929	0.218
LDV5 (6 vehicles)	Standard Deviation	0.055	2.061	0.426	24.146
	Mean (g/km)	0.120	3.660	0.602	245.131
	Coefficient of Variation	0.457	0.563	0.707	0.099
LDT1 (11 vehicles)	Standard Deviation	0.039	1.529	0.284	49.146
	Mean (g/km)	0.077	2.117	0.362	269.938
	Coefficient of Variation	0.502	0.722	0.785	0.182
LDT2 (6 vehicles)	Standard Deviation	0.118	2.352	0.171	40.960
	Mean (g/km)	0.157	5.404	0.589	247.262
	Coefficient of Variation	0.750	0.435	0.290	0.166
Average Coefficient of Variation			0.526		

Table 4  
CMEM Vehicle Classification

Category		HC	CO	NOx	CO2
Category 4 (7 vehicles)	Standard Deviation	0.077	1.578	0.212	34.593
	Mean (g/km)	0.122	2.919	0.439	179.109
	Coefficient of Variation	0.628	0.541	0.483	0.193
Category 5 (21 vehicles)	Standard Deviation	0.079	1.881	0.337	33.840
	Mean (g/km)	0.083	2.412	0.356	209.585
	Coefficient of Variation	<b>0.949</b>	<b>0.780</b>	<b>0.947</b>	<b>0.161</b>
Category 6 (3 vehicles)	Standard Deviation	0.340	1.223	0.458	27.855
	Mean (g/km)	0.268	2.606	0.563	179.199
	Coefficient of Variation	1.266	0.469	0.814	0.155
Category 7 (3 vehicles)	Standard Deviation	0.062	1.870	0.346	20.270
	Mean (g/km)	0.090	2.130	0.425	203.507
	Coefficient of Variation	0.693	0.878	0.814	0.100
Category 8 (1 vehicle)	Standard Deviation	N/A	N/A	N/A	N/A
	Mean (g/km)	0.043	1.246	0.362	159.877
	Coefficient of Variation	N/A	N/A	N/A	N/A
Category 11 (8 vehicles)	Standard Deviation	0.013	0.385	0.125	35.601
	Mean (g/km)	0.023	0.664	0.129	201.097
	Coefficient of Variation	0.556	0.580	0.966	0.177
Category 16 (12 vehicles)	Standard Deviation	0.090	2.532	0.271	41.963
	Mean (g/km)	0.125	3.651	0.527	260.195
	Coefficient of Variation	0.719	0.694	0.513	0.161
Category 17 (3 vehicles)	Standard Deviation	0.009	0.856	0.139	28.846
	Mean (g/km)	0.045	1.942	0.219	225.858
	Coefficient of Variation	0.192	0.441	0.636	0.128
Category 18 (2 vehicles)	Standard Deviation	0.057	3.721	0.100	35.080
	Mean (g/km)	0.076	3.036	0.265	326.485
	Coefficient of Variation	0.750	1.226	0.376	0.107
Average Coefficient of Variation			0.565		

Table 5  
Sample Coefficients of Hybrid Regression Model (HC Emissions for LDT1 Vehicle)

Positive Acceleration Coefficients				
	Constant	Acceleration	Acceleration <sup>2</sup>	Acceleration <sup>3</sup>
Constant	-8.27978	0.36696	-0.04112	0.00139
Speed	0.06229	-0.02143	0.00245	3.71E-06
Speed <sup>2</sup>	-0.00124	0.000518	6.77E-06	-7.4E-06
Speed <sup>3</sup>	7.72E-06	-2.3E-06	-5E-07	1.05E-07
Negative Acceleration Coefficients				
	Constant	Acceleration	Acceleration <sup>2</sup>	Acceleration <sup>3</sup>
Constant	-8.27978	-0.27907	-0.05888	-0.00477
Speed	0.06496	0.03282	0.00705	0.000434
Speed <sup>2</sup>	-0.00131	-0.00066	-0.00013	-7.6E-06
Speed <sup>3</sup>	8.23E-06	3.54E-06	6.48E-07	3.98E-08

(Speed: km/h, Acceleration: km/h/s, HC Emission Rate: g/s)

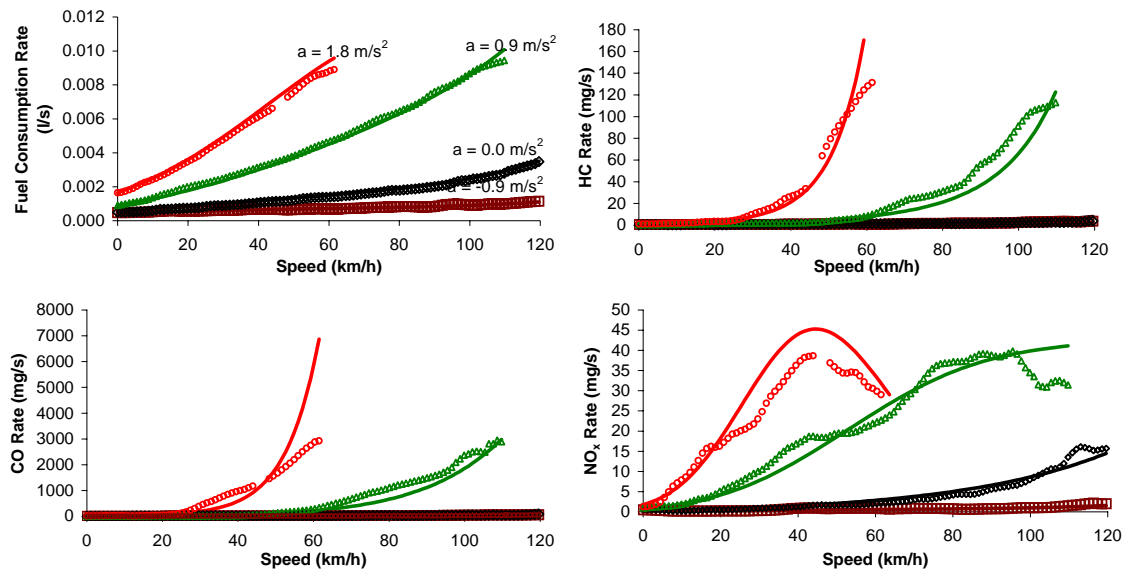


Figure 1. Regression Model Predictions (Composite Vehicle – Log-Transformed Polynomial Model)

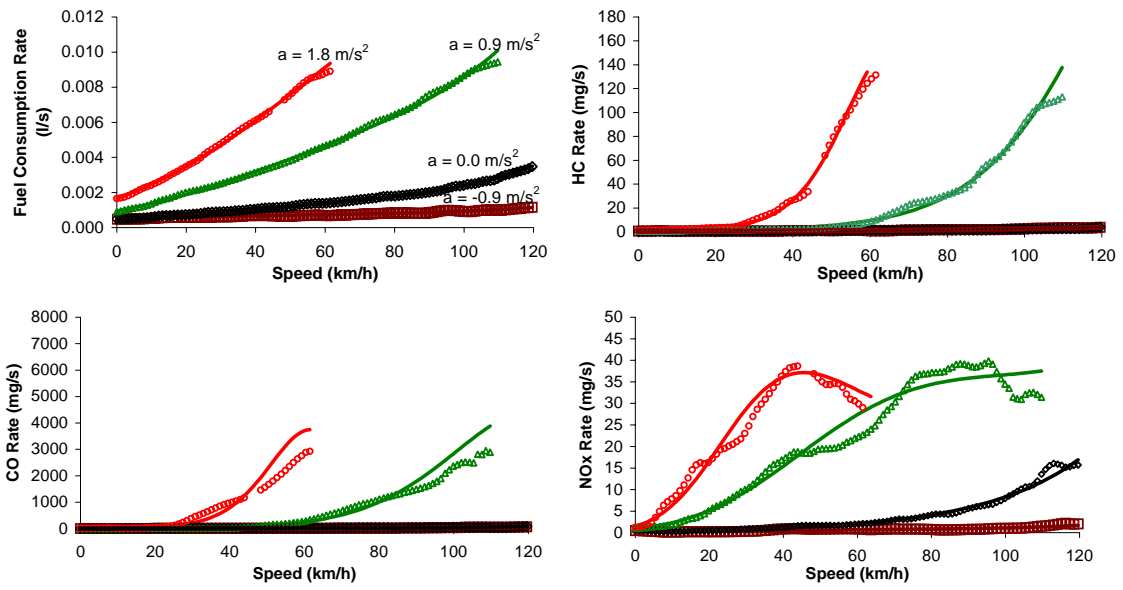


Figure 2. Regression Model Predictions (Composite Vehicle – Log-Transformed Hybrid Polynomial Model)

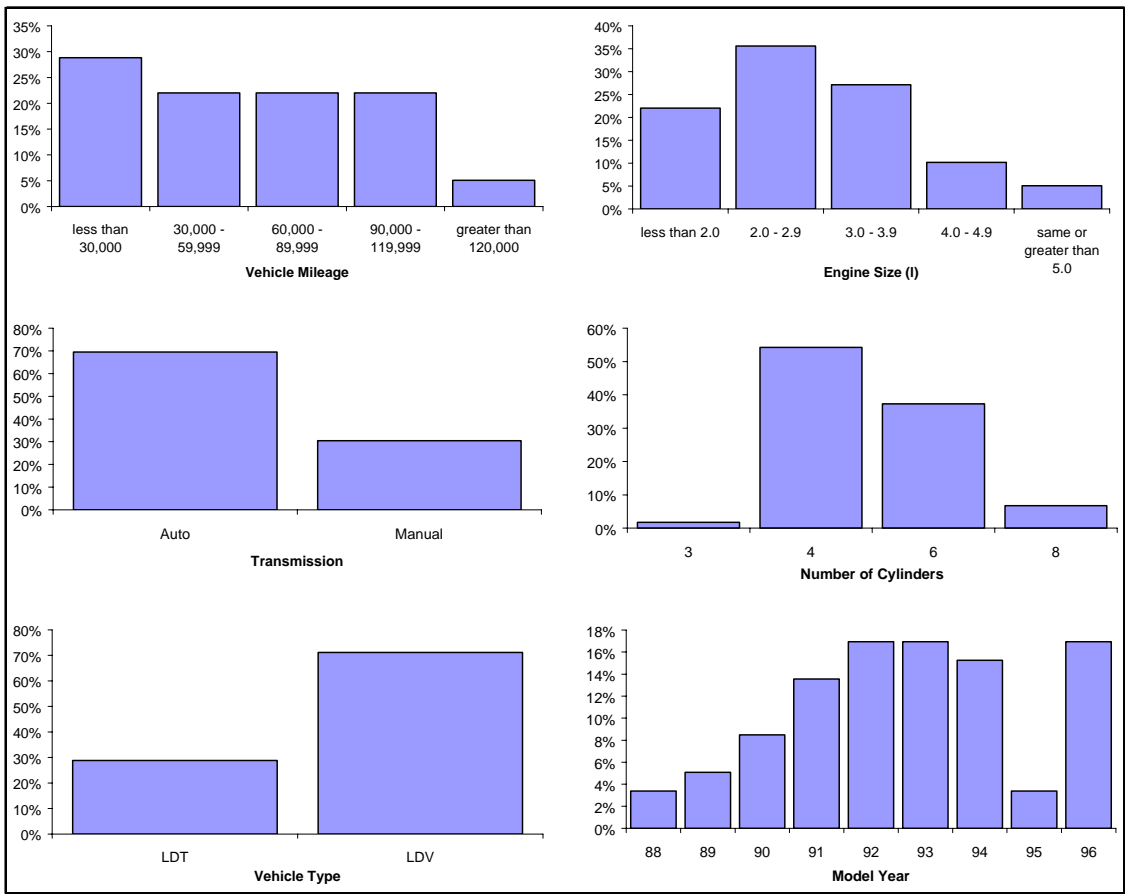


Figure 3. EPA Test Vehicle Characteristics

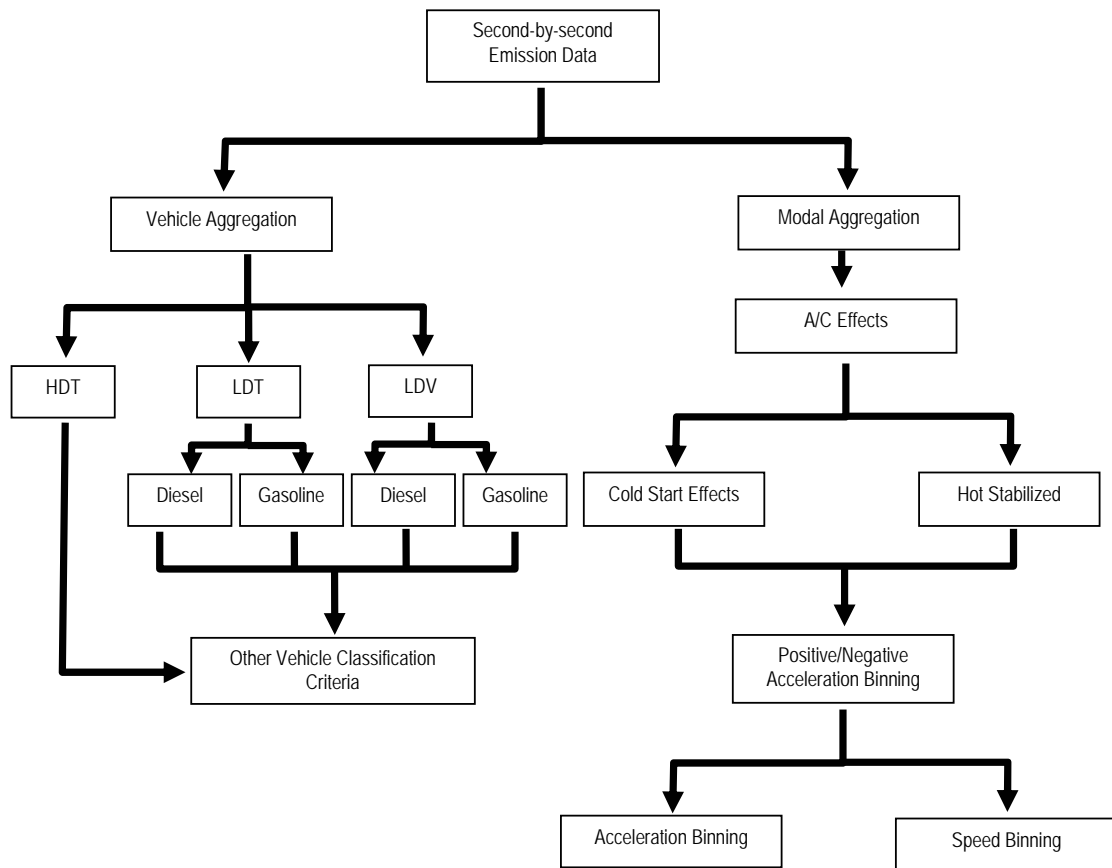


Figure 4: Second-by-second Emission Data Binning Framework

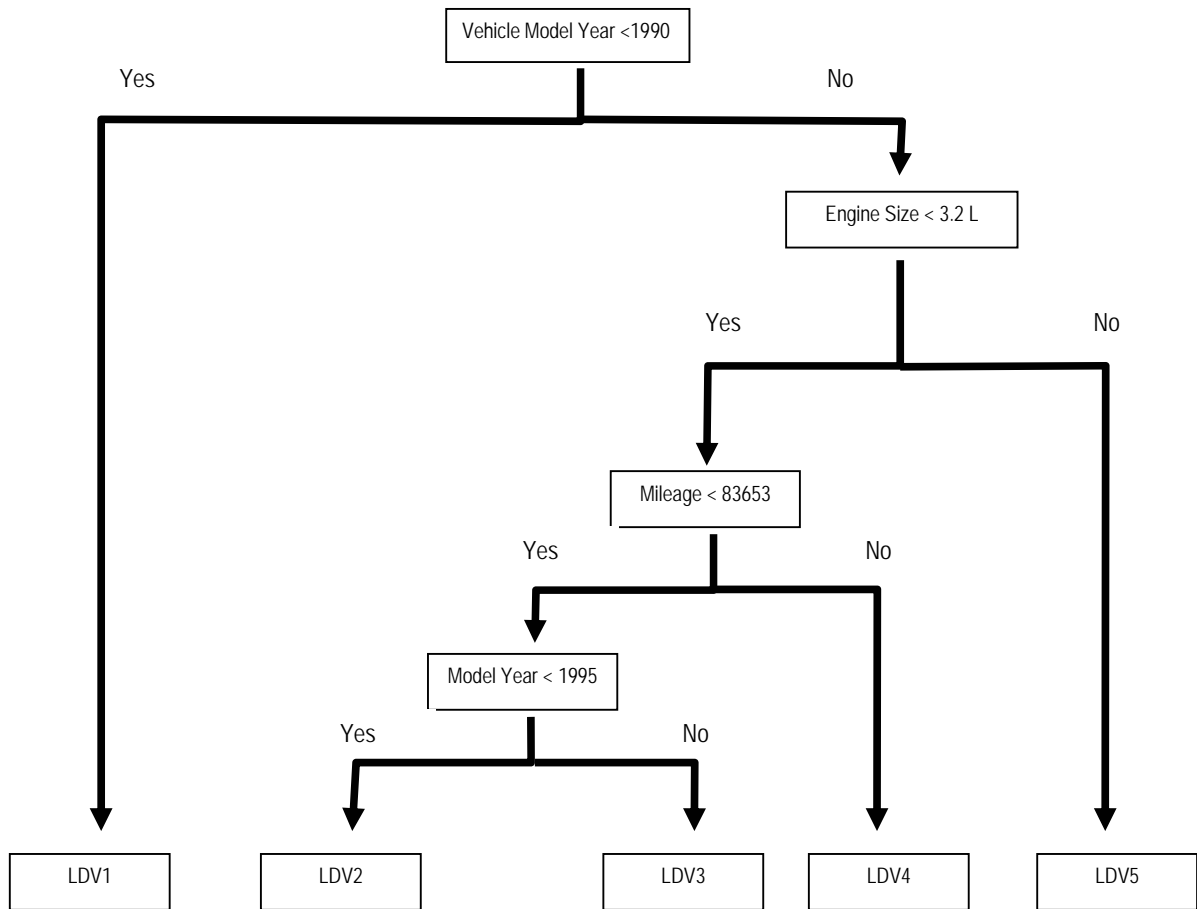


Figure 5. Light Duty Vehicle Classification using CART Algorithm

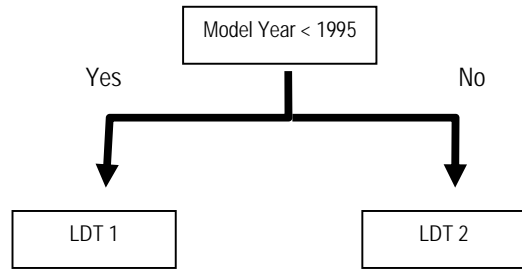


Figure 6. Light Duty Truck Classification using CART Algorithm

		Acceleration (m/s <sup>2</sup> )															
		-1.7	-1.4	-1.1	-0.8	-0.6	-0.3	0.0	0.3	0.6	0.8	1.1	1.4	1.7	1.9		2.2
Speed (km/h)	5			5	22	55	82	1322	59	55	32	31	12	21	8	3	1707
	10	3	18	20	26	37	71	84	36	25	24	21	11	13	9	8	406
	15	6	16	30	27	33	56	59	39	35	24	20	12	13	8	3	381
	20	9	16	27	35	26	50	81	64	55	36	27	11	13	5		455
	25	11	19	22	26	29	44	48	55	56	53	37	11	2	1		414
	30	15	15	24	20	33	45	72	63	37	33	31	23	7	1		419
	35	9	18	20	29	28	43	59	40	52	19	30	14	5	4		370
	40	6	15	23	22	31	58	93	83	40	34	25	3	3		1	437
	45	6	6	23	27	30	48	92	88	44	29	17		1			411
	50	4	4	13	15	36	70	96	81	42	23	11	3				398
	55	4	4	15	18	23	42	98	86	33	14	11	2				350
	60	4	3	10	11	16	75	123	87	26	15	4	2				376
	65	1	3	6	12	14	33	101	51	27	8	6	1				263
	70			10	7	20	38	81	58	14	14	1					243
	75			2	5	13	34	74	44	12	10						194
	80	2		2	5	9	32	50	47	9	11						167
	85		2	4	4	7	37	78	53	11	4						200
	90	1		3	4	10	41	107	63	9	3						241
	95				1	13	52	178	69	5	2	1					321
	100		1		3	5	79	210	90	4		1					393
105			2	1	5	63	149	66	1							287	
110			1		4	48	124	46	1							224	
115				1	2	22	29	18								72	
120						4	11	3								18	
		81	140	262	321	479	1167	3419	1389	593	388	274	105	78	36	15	8747
																Frequency	
																60<=	
																40<=and<60	
																20<=and<40	
																0< and <20	

Figure 7. Speed and Acceleration Distribution for Sample Vehicle (15 Drive Cycles)

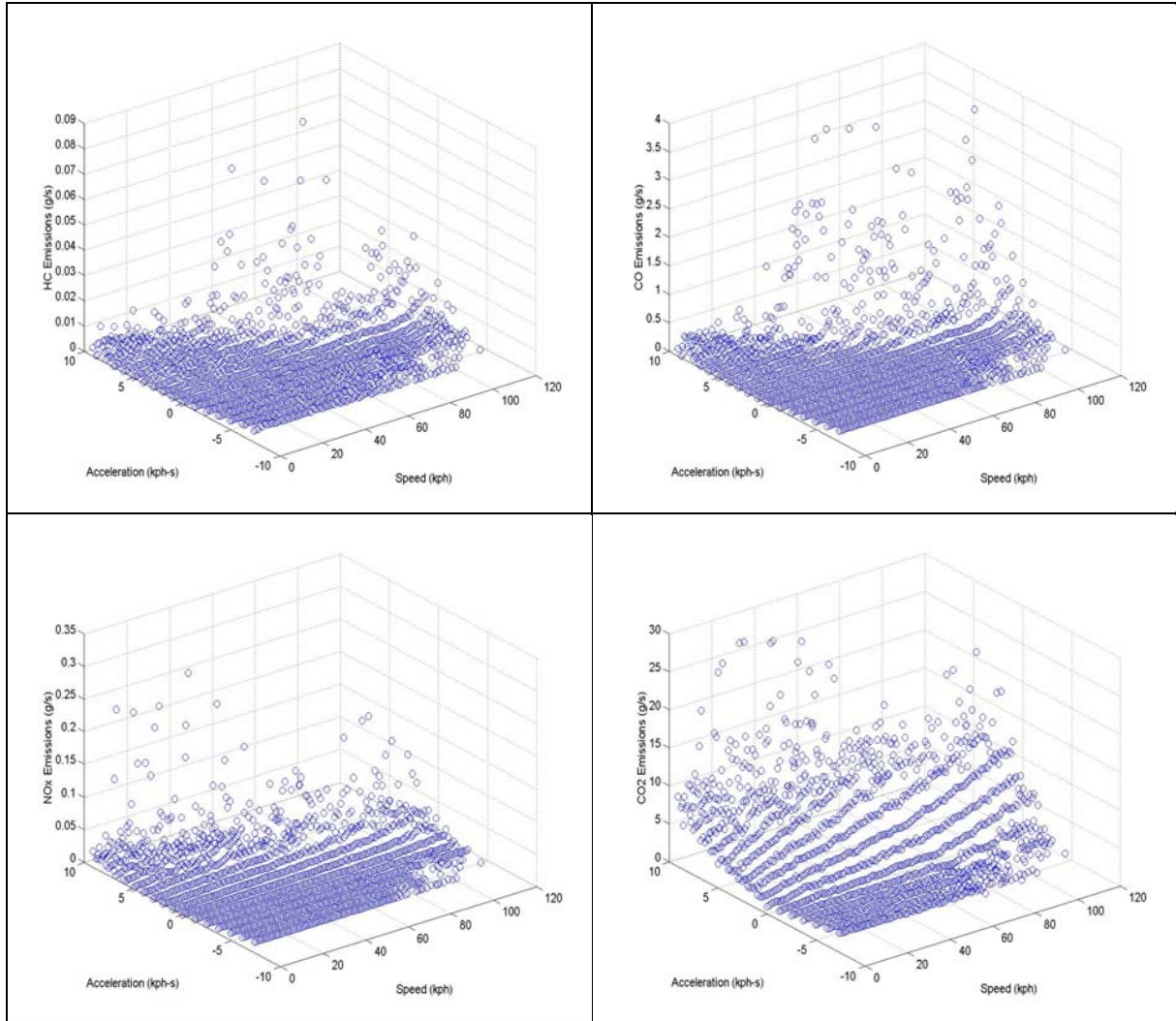


Figure 8. Sample Normalized Emission Data (LDT1)

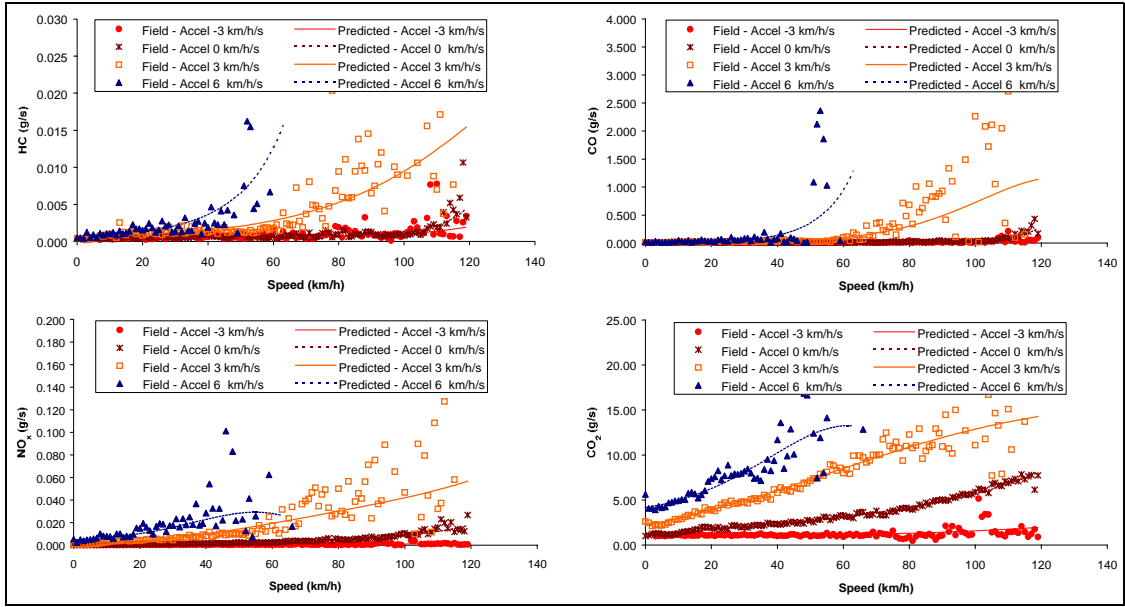


Figure 9. Model Prediction (LDT1)

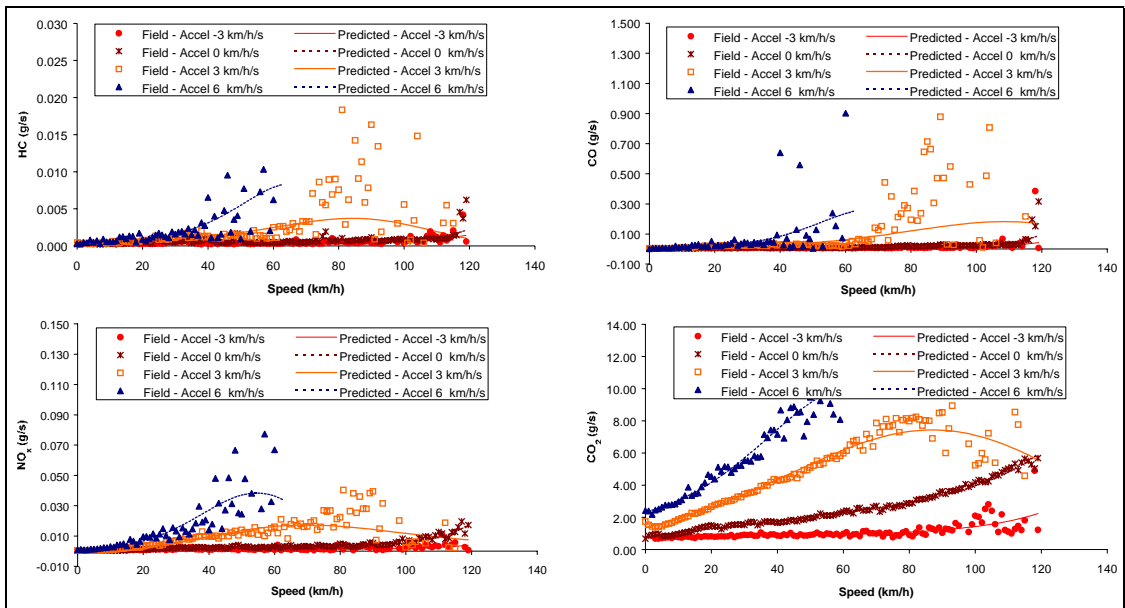


Figure 10. Model Prediction (LDV2)

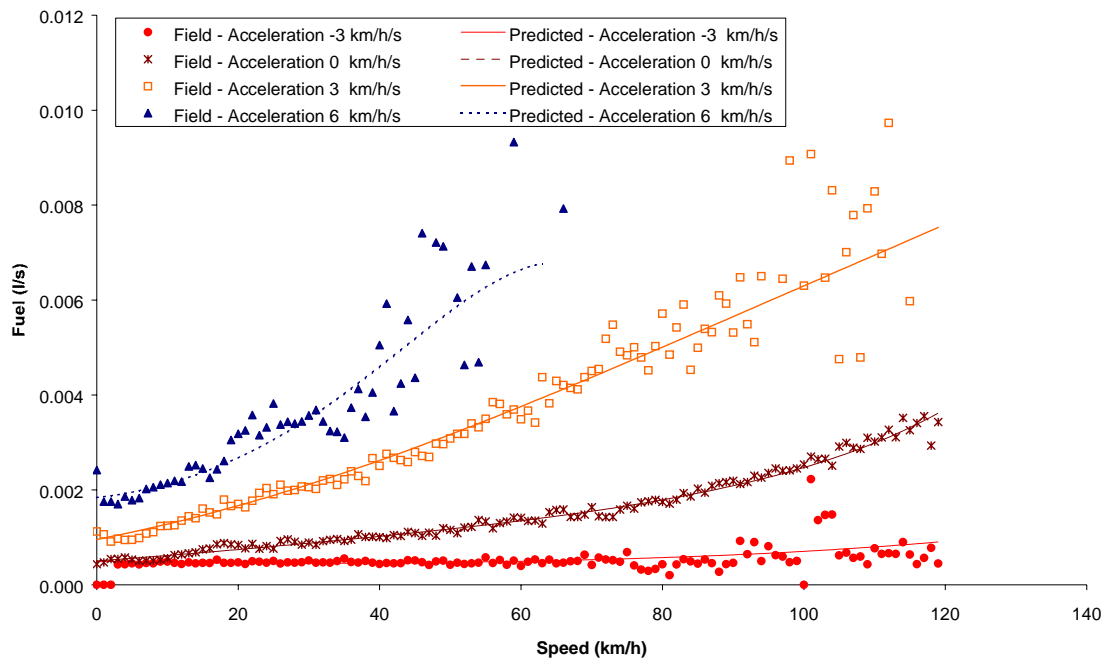


Figure 11: Fuel Consumption Prediction (LDT2)

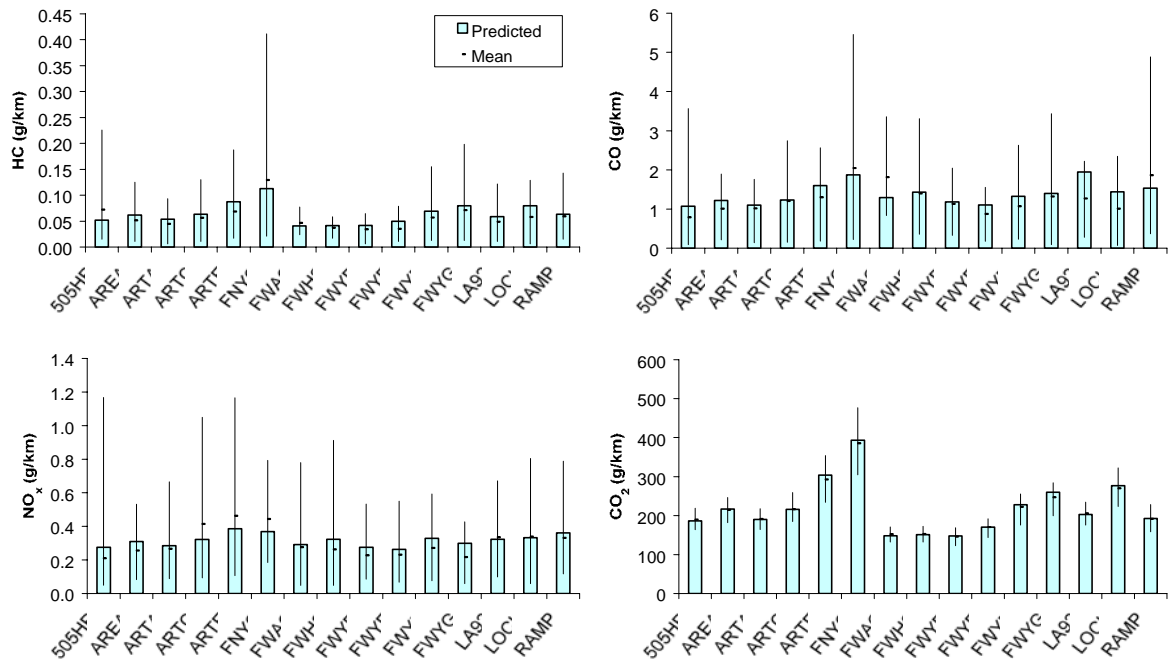


Figure 12. Model Validation for 15 Driving Cycles (LDT1)

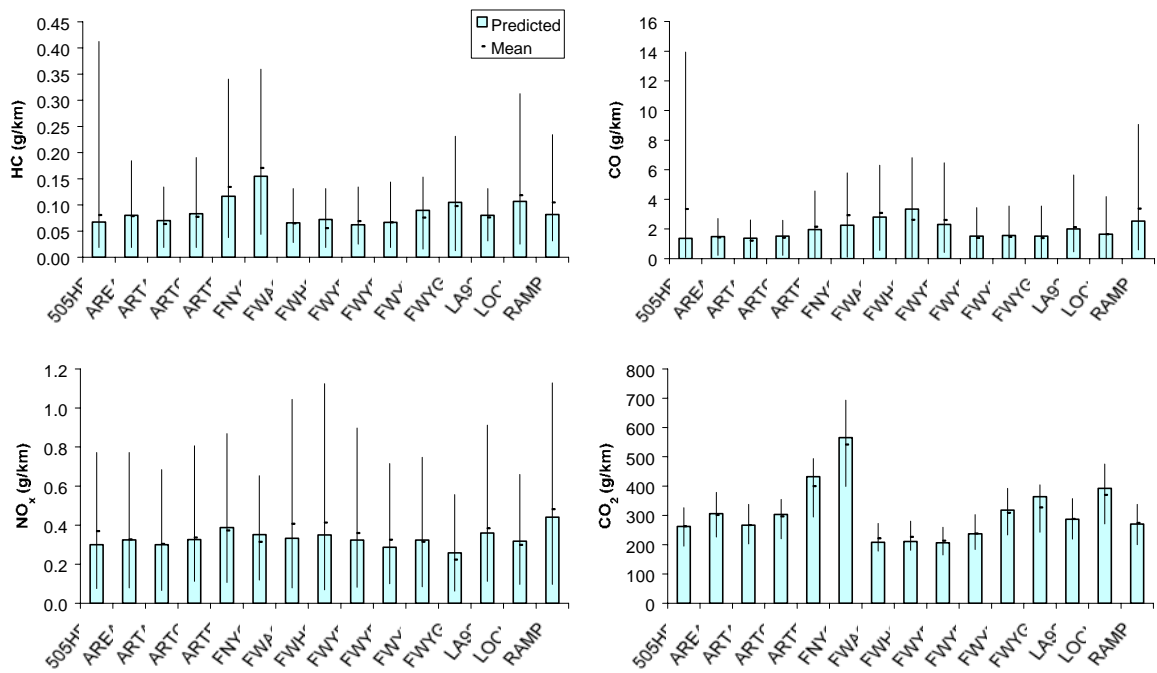


Figure 13. Model Prediction for 15 Driving Cycles (LDV2)

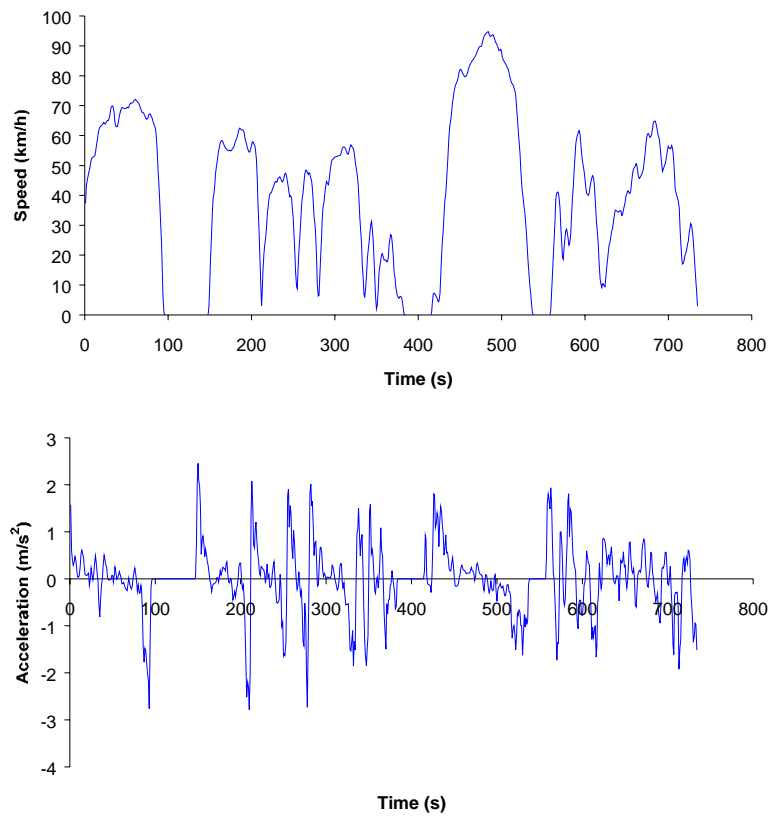


Figure 14. Speed and Acceleration Profiles for ARTA Driving Cycle

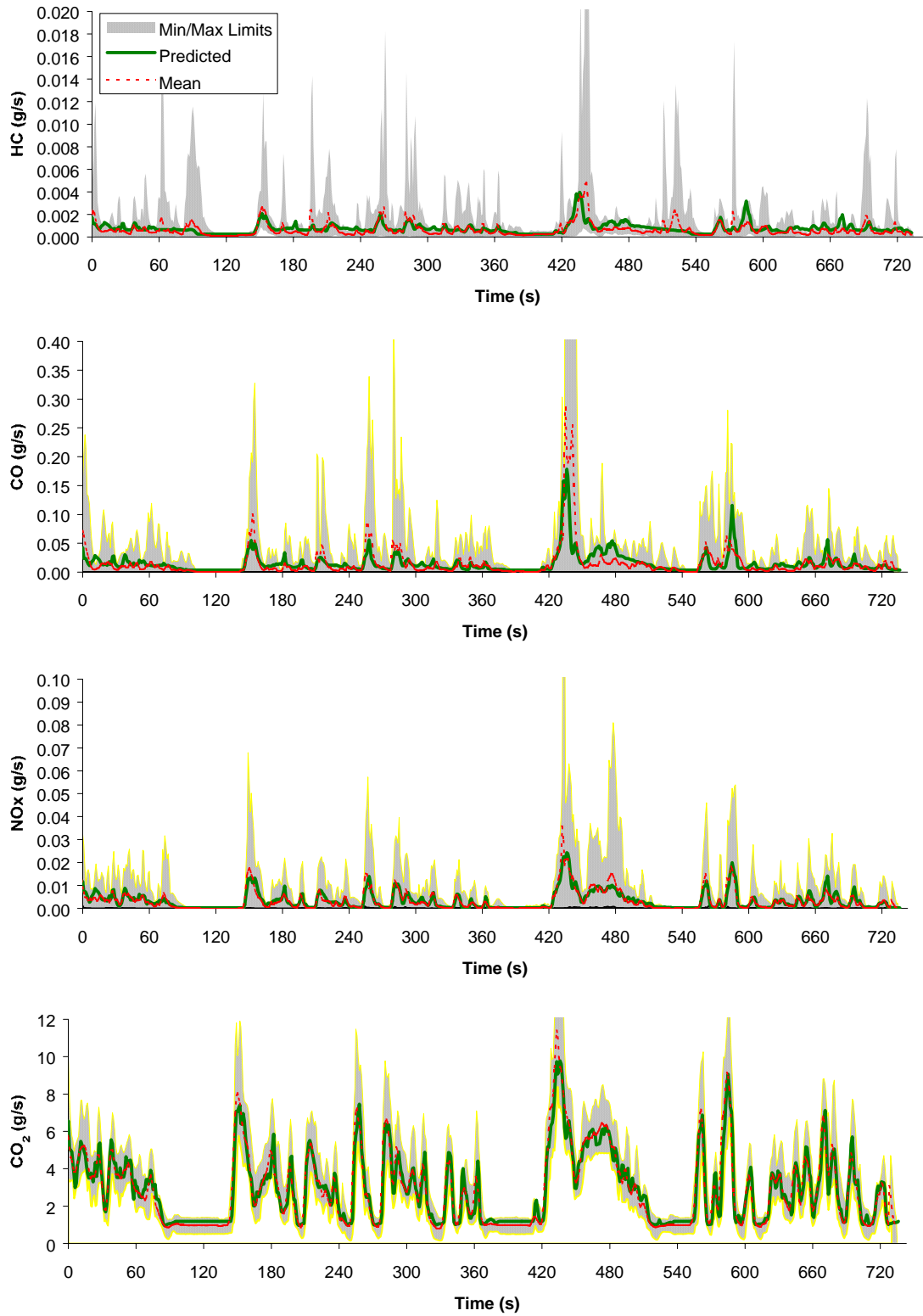


Figure 15. Instantaneous Model Validation for ARTA Cycle (LDT1)

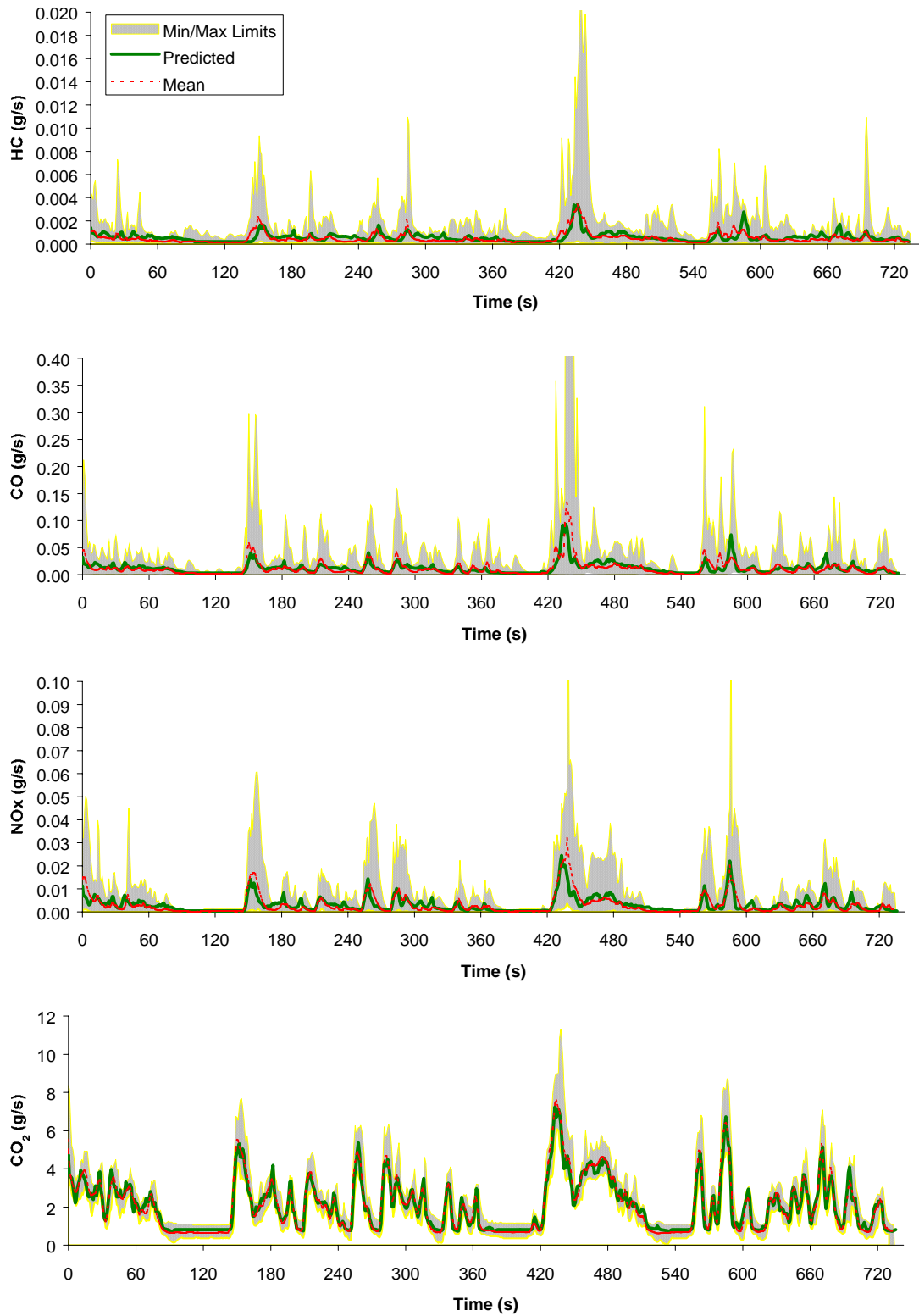


Figure 16. Instantaneous Model Validation for ARTA Cycle (LDV2)

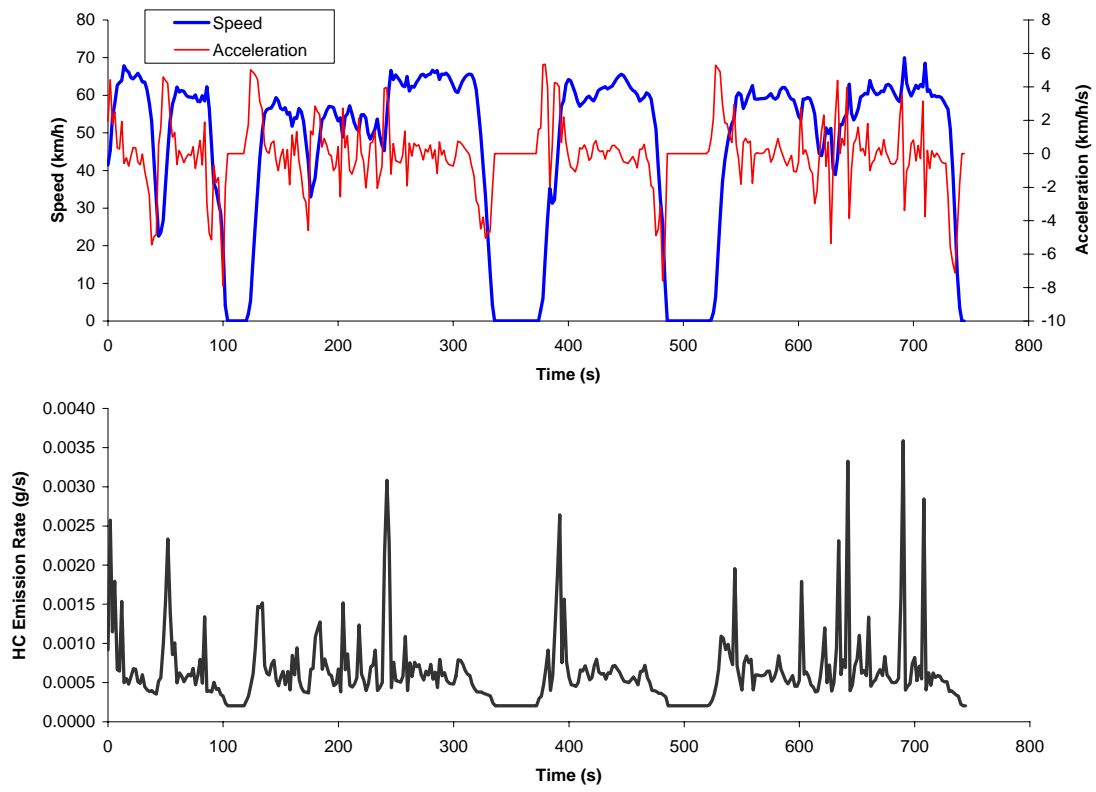


Figure 17. GPS Speed Data Application