ISSUES RELATED TO QUANTIFYING THE ENVIRONMENTAL IMPACTS OF TRANSPORTATION STRATEGIES USING GPS DATA

Bruce Hellinga, PhD, PEng
T. Nixon Chan, MASc Candidate

Department of Civil Engineering, University of Waterloo
Phone: 519-888-4567 Ext. 2630; Fax: 519-888-6197; Email: bhellinga@uwaterloo.ca

ABSTRACT

Policy makers, transportation engineers, and the public have become increasingly concerned about the impacts that transportation has on air quality. Concerns exist about the production of greenhouse-gases, such as CO_2, that contribute to global warming, and other tail-pipe emissions such as carbon monoxide, oxides of nitrogen, particulates, and hydrocarbons, which have negative health consequences.

Transportation professionals are being challenged to objectively evaluate the impact that competing traffic management strategies have on air quality. However, few tools currently exist that are able to efficiently provide this evaluation, and those that do exist, such as microscopic traffic simulation, often require specialised knowledge or extensive data collection and calibration efforts.

In this paper, we examine the feasibility and accuracy of quantifying the environmental impacts of traffic management strategies using second-by-second speed and acceleration data collected via a GPS-equipped floating car. A kernel smoothing technique is proposed to overcome errors in the GPS speed and acceleration data. A 3^{rd} order polynomial log linear energy and emissions model is used to estimate fuel consumption and tail-pipe emissions on the basis of these smoothed data. Application of this technique to 72 GPS runs along an urban arterial demonstrates that estimates of average fuel consumption and emission rates are significantly underestimated (by 20 to 60%) if these estimates are made on the basis of average vehicle speed profiles rather than individual vehicle profiles.

Two methods, namely two-stage sampling and 6-sigma, are examined for estimating the sample size required to estimate the mean fuel consumption and emission rate with some predetermined level of accuracy. It is found that the 6-sigma method is sufficiently accurate and overcomes several of the practical limitations inherent within the two-stage sampling method.

Introduction

Traditionally, data required for travel time and speed studies were collected manually via floating car studies. In the floating car method, the driver attempts to maintain average traffic stream speed by passing as many vehicles as the floating car is passed by. While the driver concentrates on the driving task, a passenger in the vehicle records the vehicle speed and cumulative travel time at predetermined locations.

The availability of low cost GPS (global positioning satellite) receivers enables transportation professionals to automatically collect continuous speed and position data from the floating car.

These data can be used to support traditional analyses, including travel time and speed studies, but they also provide the opportunity to estimate accurately vehicle fuel consumption and tail-pipe emissions.

Increasing public concern about degrading air quality, the depleting ozone layer, and global warming, is also increasing the pressure on transportation professionals to explicitly include air quality measures within impact analyses. While most transportation professionals have been willing to include air quality within the evaluation process, robust and practical methods for doing so have generally not been available.

This paper examines the feasibility of using GPS data to quantify fuel consumption and vehicle tail-pipe emissions. The paper identifies issues related to the use of GPS data and proposed a method for eliminating unwanted noise in the data. Two methods of estimating vehicle emissions are examined. The trade-off between computational effort and accuracy is identified. Finally, the issue of determining the necessary sample size is examined. A method for estimating required sample size is proposed and tested using actual GPS data.

The next section describes the microscopic fuel consumption and emissions model used within this paper. The following section describes the GPS data and identifies several key characteristics of GPS data that are undesirable when estimating emissions. A non-parametric smoothing method is proposed to eliminate the undesirable noise in the GPS data. Two methods of estimating average fuel consumption and emission rates for a particular corridor are identified and examined in turn. These methods are applied to 72 GPS runs along a 9.5km arterial corridor in Scottsdale, Arizona. Finally, conclusions and recommendations are made.

Fuel Consumption and Emission Model

A number of fuel consumption and emission models exist. The US EPA Mobile 5 model is an aggregate model that modifies rates developed for a predefined driving cycle using speed correction factors. Experience has shown that large estimation errors can occur when the driving conditions for which estimates are desired, differ significantly from the base driving cycle. The EPA is currently attempting to reduce these errors in the next release of Mobile (Mobile 6) by increasing from 2 to 13 the number of driving cycles considered (Sierra Research, 1999). However, even with this increased number of cycles, the user is still faced with the problem of deciding which of the predefined cycles most closely reflects the actual conditions experienced on the corridor of interest.

The Comprehensive Modal Emissions Model (CMEM) was developed by a team of researchers led by Dr. Barth from the University of California at Riverside (Barth et al, 1997; Barth et al, 2000; Barth et al, 1999; and An, et al., 1997). CMEM uses a physical, power-demand modal modelling approach that is an analytical representation of the emission production system. The physical process by which emissions are produced is separated into constituent components. Each of these components is modelled as an analytical representation consisting of various parameters that are characteristic of the physical process. The parameter values are dependent on the type and condition of the vehicle, engine, and emission controls. The model was calibrated on the basis of emission measurements collected from over 300 in-use vehicles tested on a vehicle dynamometer over predefined driving cycles.

Researchers at Virginia Tech Transportation Institute have also developed a microscopic fuel consumption and emission model (VTMicro[1]). This model is based on 8 vehicles tested on a dynamometer at Oak Ridge National Laboratory (Ahn et al., 1999; Rakha et al., 2000a; Rakha et al., 2000b). Unlike many energy and emission data collection efforts, data were collected for a matrix of speeds ranging from 0 to 120 km/h and for acceleration ranging from –1.5 to 3.7 m/s^2. No attempt was made to collect data over a driving cycle representative of contiguous field driving conditions. These data were then used to calibrate a log-linear 3rd order polynomial regression model (Equation 1).

$$\ln(F) = a_0 + a_1A + a_2A^2 + a_3A^3 + a_4S + a_5S^2 + a_6S^3 + a_7AS + a_8AS^2 + a_9AS^3 + a_{10}A^2S + a_{11}A^2S^2 + a_{12}A^2S^3 + a_{13}A^3S + a_{14}A^3S^2 + a_{15}A^3S^3 \quad (1)$$

where:

F	=	instantaneous fuel consumption or pollutant emission rate in litres/s or mg/s
A	=	instantaneous acceleration of vehicle in km/h/s
S	=	instantaneous speed of vehicle in km/h
a_0, a_1, ... a_{15}	=	vehicle specific calibration coefficients

The microscopic model does not consider high emitters, cold-starts, effects of vertical gradient of the road, or secondary loads on the engine, such as air conditioning. Furthermore, because the model considers only the instantaneous speed and acceleration of the vehicle, it cannot reflect history effects on fuel consumption and emissions. Coefficients for the model are provided in Table 1.

Table 1: Parameter values for composite vehicle VTMicro Model

Coefficient	Fuel	HC	CO	NO_x
a_0	-7.5370E+00	-7.2804E-01	8.8745E-01	-1.0677E+00
a_1	1.2328E-01	0.0000E+00	1.3565E-01	2.3181E-01
a_2	1.3244E-02	1.9411E-02	2.5373E-02	7.3636E-03
a_3	-8.9997E-04	-7.0577E-05	-1.0203E-03	-7.1983E-04
a_4	2.7035E-02	2.2738E-02	6.4700E-02	4.2307E-02
a_5	-2.2923E-04	-1.7026E-04	-6.5281E-04	-1.4368E-04
a_6	1.1255E-06	1.4752E-06	3.4939E-06	4.3069E-07
a_7	3.9933E-03	8.4259E-03	3.2142E-03	1.2859E-02
a_8	-1.5543E-05	-7.7962E-05	7.0566E-05	-9.9156E-05
a_9	3.8216E-08	4.2630E-07	-4.8701E-07	2.2626E-07
a_{10}	6.3073E-05	-4.1555E-04	-7.0090E-04	2.1769E-03
a_{11}	6.4635E-07	2.5931E-05	3.3926E-05	-4.0464E-05
a_{12}	-1.5591E-08	-1.3390E-07	-1.9740E-07	1.5088E-07
a_{13}	-4.2300E-05	-7.7949E-05	0.0000E+00	-2.2143E-04
a_{14}	1.9111E-07	2.0809E-06	-8.8640E-07	1.2215E-06
a_{15}	-2.5610E-09	-9.9688E-09	0.0000E+00	-7.2188E-09

While the model explains a large portion of the variance in the observed data (i.e. $r^2 > 0.9$), model predictions are extremely sensitive to large absolute values of acceleration. For example, consider Figure 1, which illustrates the model's predictions of HC and CO emissions, in mg/second, as a function of acceleration at a constant speed of 40 km/h. It is clear from this

[1] The researchers at Virginia Tech have recently extended their model to a two-regime structure, with one regression for deceleration and one for acceleration conditions. The authors of the model claim increases in explanatory power of the extended model, however, the coefficients for this model have not yet been published.

figure that for large absolute values of acceleration, the model provides highly unrealistic estimates.

The model developers have attempted to avoid the problem of these highly unrealistic estimates by constraining the region of application of the model to within a feasible region of speed and acceleration combinations. This feasible region (Equations 2a and 2b) represents the speed and acceleration combinations that are physically possible for a typical automobile. However, as illustrated in Figure 1 for very large deceleration rates, the model may still provide vastly overestimated emission results for speed and acceleration combinations that are within the feasible region.

$$A_{\max} = \begin{cases} 3 - 0.025(S - 30) & S \geq 30 \\ 3 & S < 30 \end{cases} \quad (2a)$$

$$A_{min} = \text{-5.0 m/s}^2 \quad (2b)$$

where:

A_{min} = minimum instantaneous acceleration of vehicle in m/s^2
A_{max} = maximum instantaneous acceleration of vehicle in m/s^2
S = instantaneous speed of vehicle in km/h

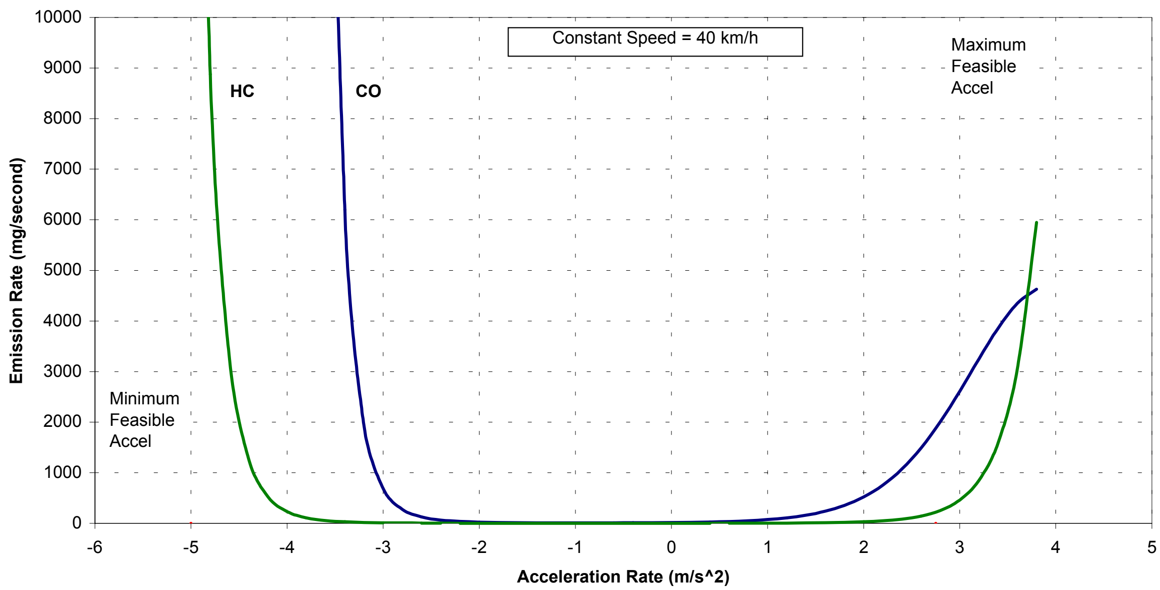

Figure 1: Illustration of unrealistic emission estimates produced by VTMicro for extreme acceleration/deceleration rates

Description of Field Data

The field data used in this study were collected in Scottsdale, Arizona as part of the Phoenix Metropolitan Model Deployment Initiative (Rakha et al, 2000c). The Scottsdale/Rural Road corridor is an arterial connecting the cities of Tempe and Scottsdale. The study area extended over 9.6-km and covered 21 signalized intersections. Four cars were equipped with GPS Placer 450 receivers and drove along the corridor over a three-day period in January 1999. Runs were conducted during the AM, PM and off-peak periods. Vehicle position and speed were collected each second. A total of 72 runs were conducted in the southbound direction and 69 in the northbound direction. Only the southbound data are used this paper. Speed data were not computed from finite difference of position data, but was computed directly by the Placer GPS unit. The specifications of the Placer 450 GPS unit indicate that speed accuracy is within 1 m/s (3.6 km/h). Table 2 provides descriptive characteristics of the data whereas Figure 2 illustrates the distribution of trip travel times for the 72 runs. It is evident that the runs have been collected over a broad range of traffic conditions, as the maximum trip duration is more than twice as long as the minimum.

Figure 3 illustrates the speed profiles of 3 typical runs. As expected, the individual vehicle profiles exhibit significant variation, as some vehicles experience delay at a signal and others do not.

Table 2: Selected characteristics of field data

	Mean	**Minimum**	**Maximum**
Trip Duration (seconds)	839	608	1433
Mean Trip Speed (km/h)	43.0	24.0	56.2
Std of Speed (km/h)	24.4	16.3	30.1

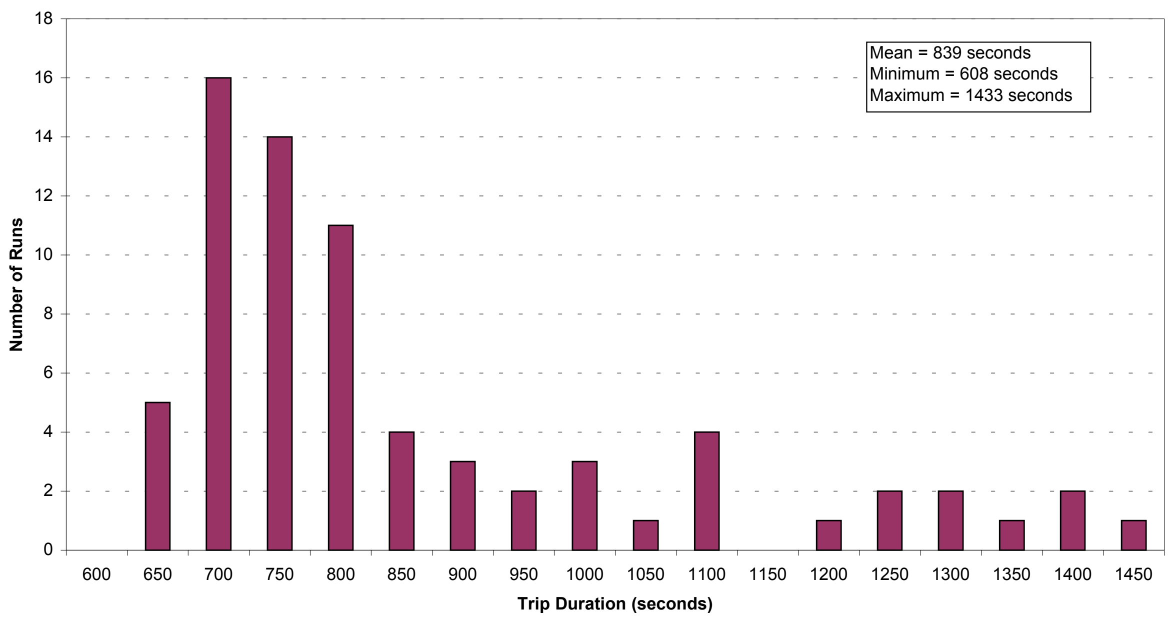

Figure 2: Distribution of GPS trip run duration

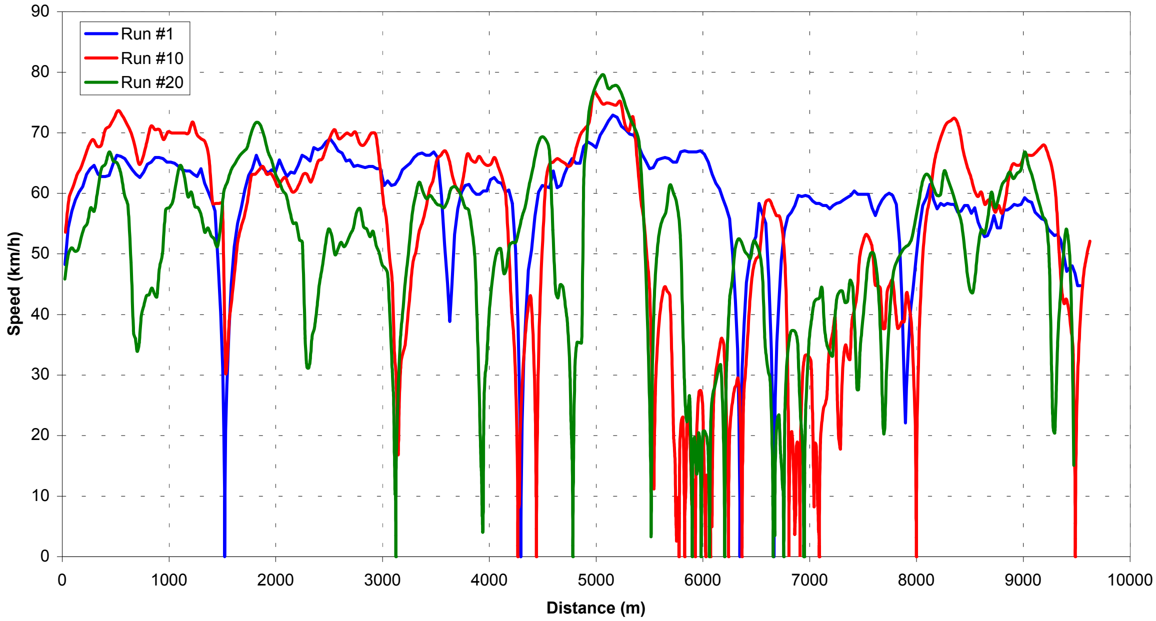

Figure 3: Speed profiles obtained from GPS data for a typical sample of runs

GPS Data Filtering and Smoothing

It is possible to directly compute acceleration on the basis of the measured speed data using numerical differentiation (Equation 3).

$$a_t = \frac{S_t - S_{t-1}}{\Delta} \tag{3}$$

where:

a_t	=	estimate of instantaneous acceleration at time *t* (km/h/s)
S_t	=	measured speed of vehicle at time *t* (km/h)
S_{t-1}	=	measured speed of vehicle at time *t-1* (km/h)
Δ	=	duration between speed observations at time *t* and *t*-1 (seconds)

However, in practice, the measured speed data contain error, which can lead to large errors in the estimated instantaneous acceleration rate. For example, consider Figure 4, which illustrates the instantaneous acceleration rate computed using Equation 3 from the raw speeds measured for run 61. It is clear that during a number of time intervals, the computed acceleration rate falls outside the feasible acceleration limits (i.e. Equation 2). For all 72 runs, the acceleration computed on the basis of the raw speed data falls outside the feasible region for 166 of the 60,360 observations or 0.275%. While this does not appear to be a large problem in terms of the number of outliers, some of the computed accelerate rates fall far outside the feasible limits. For example, the maximum acceleration rate is 16.98 m/s^2 and the minimum acceleration is -16.62 m/s^2. If these acceleration values, along with their associated speeds, are used directly in the microscopic emission model, the emission model provides nonsensical results. For example, the acceleration rate of 16.98 m/s^2 is associated with a speed of 68.12 km/h. For this conditions, the emission model predicts the instantaneous HC emission to be 1.11×10^{233} mg/second! Since the model is so sensitive to speed and acceleration rates that fall near to or outside the feasible region, even a small number of outliers can radically alter the mean fuel consumption and emission rate. Therefore, it is necessary to filter the raw speed data to reduce the impact of errors in the measured speed data.

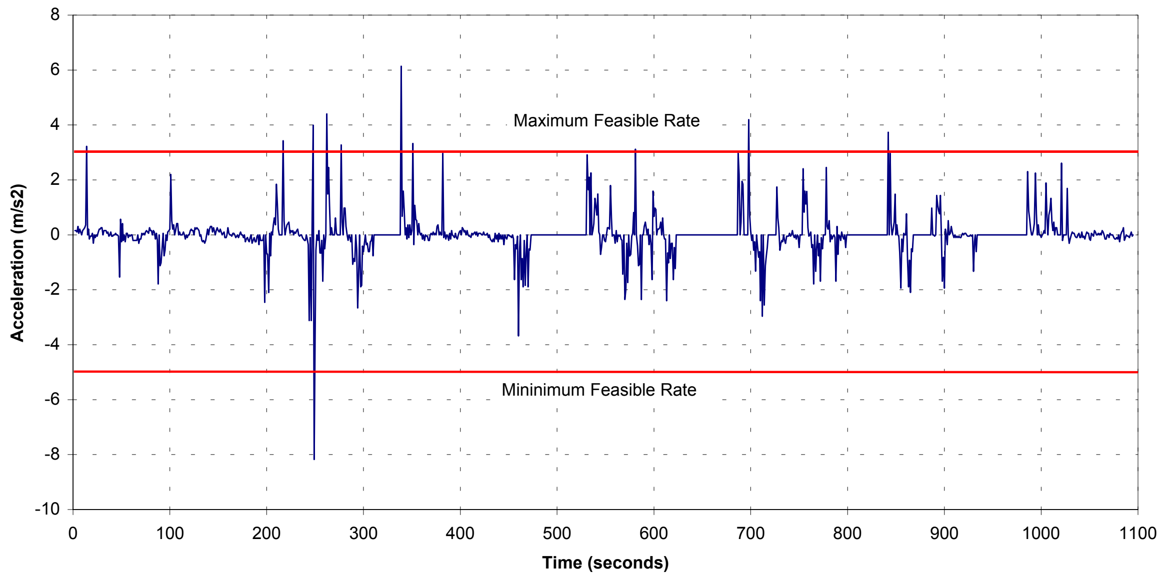

Figure 4: Acceleration rate computed from the raw GPS speed data (Run #61)

There exists a wide range of smoothing methods that can be applied to this specific problem. Parametric approaches assume that the distribution of the errors be known *a priori* and then make use of the characteristics of the distribution to make decisions regarding outliers. These methods can work quite well as long as the data do in fact follow the assumed distribution. Non-parametric approaches differ in that do not make any assumption about the structure of the underlying process. Examples of non-parametric smoothing approaches include weighting moving average (or more generally termed density kernels).

Rakha et al (2001) investigated the suitability of a number of smoothing techniques for application to GPS data. They applied these smoothing methods to the instantaneous acceleration rates computed using Equation 3 and found that the Epanechnikov density kernel performed well in comparison to other smoothing methods. In this paper, we apply the Epanechnikov Kernel smoothing method to the raw speed data and compute the instantaneous acceleration rates on the basis of the smoothed speeds.

The Epanechnikov Kernel is computed as

$$K(z_{ij}) = \begin{cases} 0.75(1 - z_{ij}{}^2) & |z_{ij}| \leq 1 \\ 0 & otherwise \end{cases} \tag{4}$$

where:

z_{ik} = $\dfrac{i-j}{\alpha}$

i = current time interval (seconds)

j = time interval j for which density kernel is being computed (seconds)

α = duration of time period in the past and in the future considered within the smoothing (bandwidth)

When computing the smoothed speed at time interval i ($\hat{S}_i$), the weight associated with each speed in time interval j (W_{ij}), is computed by

$$W_{ij} = \frac{K(z_{ij})}{\sum_{k=i-\alpha}^{i+\alpha} K(z_{ik})} \quad (5)$$

Finally, the smoothed speed in interval i ($\hat{S}_i$) is computed by,

$$\hat{S}_i = \sum_{j=i-\alpha}^{i+\alpha} \left(W_{ij} S_i\right) \quad (6)$$

and acceleration is computed as,

$$a_i = \frac{\hat{S}_i - \hat{S}_{i-\Delta}}{\Delta} \quad (7)$$

As indicated by Figure 5, the Epanechnikov kernel smoothing function provides parabolic weights that are dependent on the bandwidth (α).

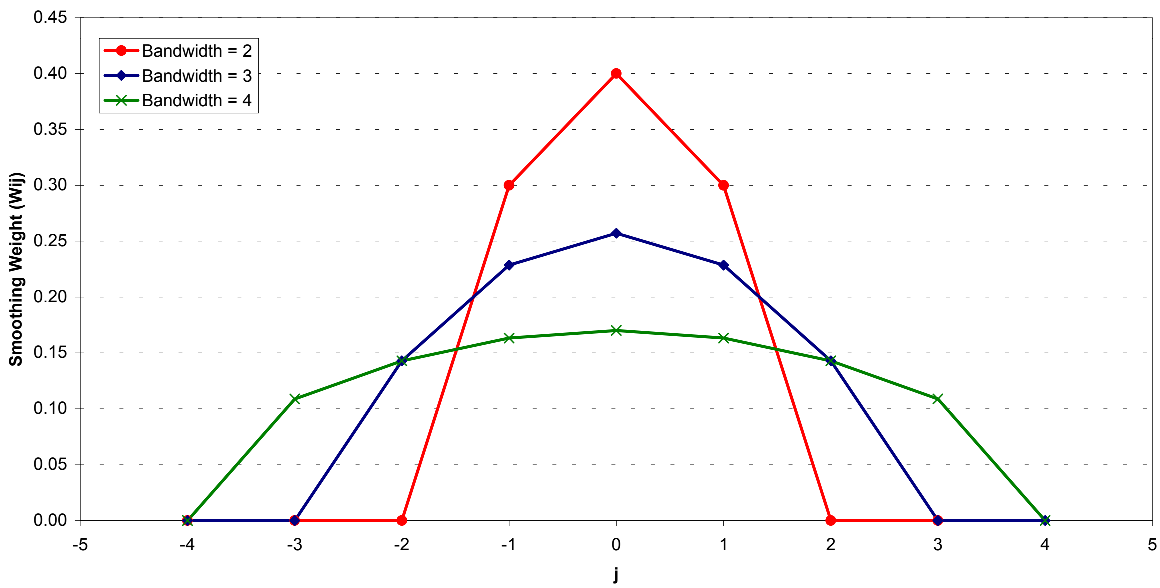

Figure 5: Epanechnikov Kernel smoothing weights for different bandwidths

Figure 6 illustrates the application of Epanechnikov kernel smoothing with α = 3 seconds to a portion of a typical speed profile. The smoothed speeds follow the measured speeds quite closely, with only some moderate flattening of the peaks and valleys. Note that the proposed smoothing function does not guarantee that the computed acceleration falls within the feasible region. However, application of the smoothing function to all 72 runs resulted in only 5 acceleration observations, out of the total of 60,360, violating the feasible region constraints (Table 3).

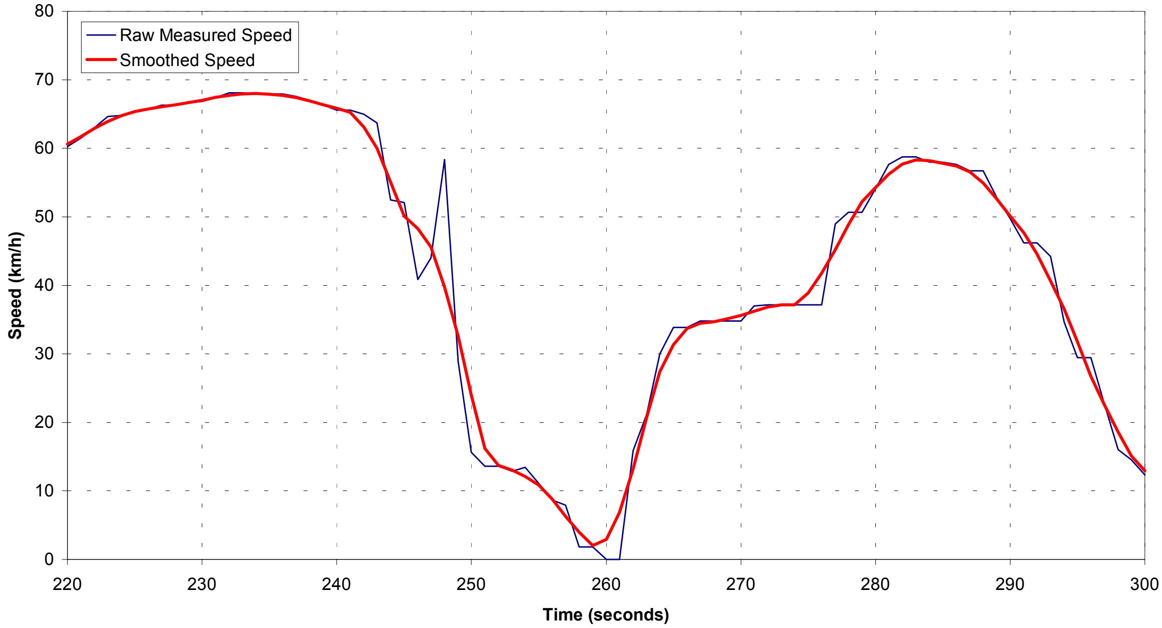

Figure 6: Raw speed and Epanechnikov kernel smoothed speed data (α = 3 seconds)

Table 3: Acceleration data falling outside of feasible region after application of Epanechnikov Kernel smoothing

Run #	Time Interval (sec)	Speed (t-1) (km/h)	Speed (t) (km/h)	Acceleration (m/s^2)	Max Accel (m/s^2)	Min Accel (m/s^2)
59	788	60.28	68.02	2.15	2.05	-5.00
60	611	15.82	26.97	3.10	3.00	-5.00
63	1240	17.03	29.66	3.51	3.00	-5.00
63	1241	29.66	41.67	3.34	2.71	-5.00
66	139	59.64	68.52	2.47	2.04	-5.00

Estimating Fuel-Consumption and Vehicle Tail-Pipe Emissions

Once the smoothed speed and acceleration data have been obtained, it is necessary to apply the fuel consumption and emissions model. However, two alternate methods exist by which this can be done, as described below.

Method 1: Individual GPS Runs

We can apply the model to each run independently to obtain the total emission for each trip (Equation 8). The average emission rate in units of g/veh-km is computed by summing the total trip emission across all 72 trips and dividing by the cumulative distance travelled by all 72 runs (Equation 9).

$$E_R = \sum_{t=1}^{D_R} F_{t,R} \tag{8}$$

$$\overline{E} = \frac{\frac{1}{N}\sum_{R=1}^{N} E_R}{\sum_{R=1}^{N} L_R} \tag{9}$$

where:

R	=	run number
N	=	total number of runs (in this study N=72)
t	=	time interval (seconds)
D_R	=	duration of run R (seconds)
E_R	=	total fuel consumption/emission associated with run R (mg for emissions and litres for fuel consumption)
L_R	=	length of run R (km)
$F_{t,R}$	=	instantaneous fuel consumption/emission at time t for run R which is estimated using Equation 1.

Table 4 provides the mean, maximum, minimum, standard deviation, and coefficient of variation for each of the tail-pipe emissions and for fuel consumption that result from the application of Equations 8 and 9.

Table 4: Descriptive statistics of fuel consumption and emission rates estimated using Method 1

	Fuel (ml/veh-km)	Emissions (g/veh-km)		
		HC	CO	NOx
Average	128.69	0.140	2.515	0.363
Max	168.82	0.18	3.66	0.49
Min	105.78	0.11	1.96	0.27
Std	16.18	0.02	0.35	0.05
COV	0.13	0.13	0.14	0.15

Method 2: Average Speed Profile

In many cases, such as calibration of simulation models, GPS data are collected and average speed versus distance profiles are determined. In such cases, it may be computationally attractive to apply the fuel consumption and emissions model directly to the mean speed profile rather than applying it to each individual run. Of course, since the emissions model is based on time rather than distance, the mean speed versus distance profile must be transformed into a mean speed versus time profile. Acceleration is computed using Equation 3 as before, except this is done using the mean speeds rather than the speeds of the individual runs.

The instantaneous mass of emission produced and fuel consumed at time t (Ft) is computed on the basis of the mean speed and acceleration at time t. The mean emission and fuel consumption rate is computed using Equation 10. The numerator in Equation 10 represents the weighted total emission produced or fuel consumed by all GPS vehicles as they traverse the roadway. The denominator represents the total combined distance travelled by all GPS vehicles.

$$\overline{E} = \frac{\sum_{t=1}^{D} n_t F_t}{\frac{1}{3600}\sum_{t=1}^{D} n_t \overline{S}_t} \tag{10}$$

where:

$\overline{E}$	=	average fuel consumption or emission rate (l/veh-km for fuel and mg/veh-km for emissions)
n_t	=	number of runs for which a smoothed speed exists at time t (in this study $n_t \leq 72$)
t	=	time interval (seconds)
D	=	duration of average speed profile (seconds)
$\overline{S}_t$	=	mean smoothed speed at time t (km/h)
F_t	=	instantaneous fuel consumed or emission produced at time t (litres for fuel and mg for emissions)

Results

Table 5 provides the average fuel consumption and emission rates associated with the 2 estimation methods. It is evident that estimating average fuel consumption or emission rate on the basis of averaged speed and acceleration data (i.e. Method 2) introduces significant errors. Fuel consumption is underestimated by 25% and tail-pipe emissions are underestimated by 34% to 46%.

Table 5: Comparison of two methods for estimating fuel consumption and emission rates

	Fuel (l/veh-km)	**Estimated Average Emissions (g/veh-km)**			**Estimation Error (%)**[1]			
		HC	CO	NOx	Fuel	HC	CO	NOx
Method 1 - individual runs	0.129	0.140	2.515	0.363				
Method 2- average profile	0.096	0.092	1.579	0.198	-25.1%	-34.6%	-37.2%	-45.5%

[1] Error computed as ((Method 2 - Method 1)/Method 1 x 100%

These results are obtained because averaging speed profiles tends to produce speed profiles that have less severe acceleration and deceleration events. To illustrate, consider Figure 7, which provides the relative frequency distribution of acceleration as well as descriptive statistics for both methods. The range and standard deviation of instantaneous acceleration rate are the greatest for Method 1. Since emission rates are sensitive to acceleration rate, the reductions in the severity of acceleration that results from application of Method 2 also produces a corresponding reduction in the estimated average emission rate.

From these results, it can be concluded that the benefits of reduced computational load and convenience associated with Method 2 are not sufficiently large to warrant the acceptance of the associated reductions in accuracy.

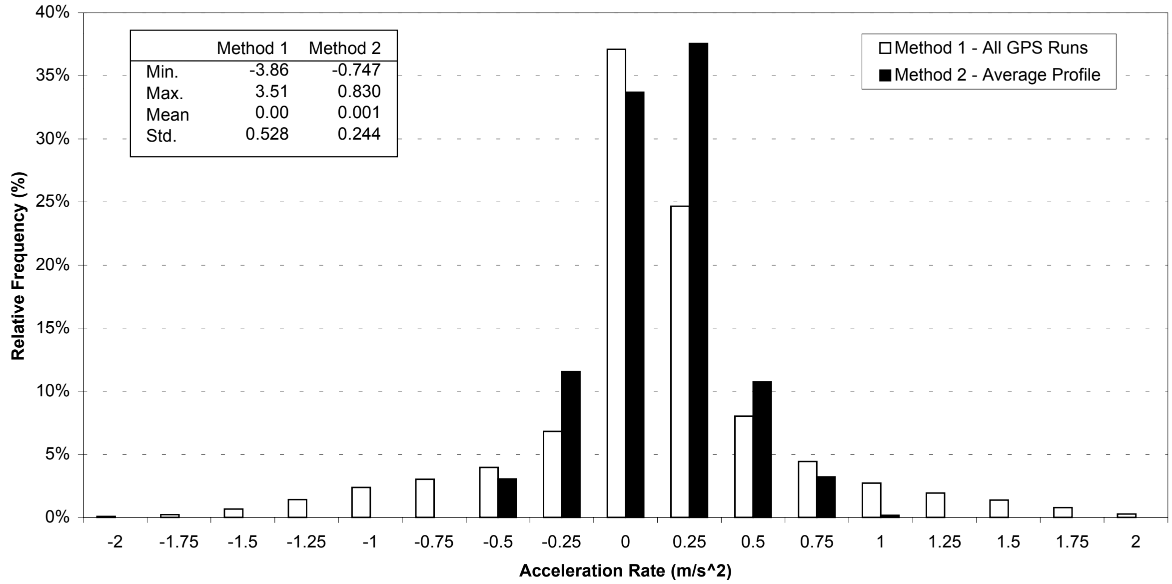

Figure 7: Distribution of acceleration rates for the two computational methods

Effect of Sample Size

The number of GPS runs required to accurately determine fuel consumption and emission rates is an important consideration, since collecting GPS data and subsequently processing it can be time consuming and costly. For any data collection effort, it is necessary to determine the sample size required to provide population estimates with some desired degree of reliability. In this case, we assume we are interested in estimating the mean emission rate, in terms of mass of pollutant emitted per vehicle per unit distance for vehicles travelling along a particular corridor. If we assume a large (infinite) population, then Equation 11 expresses the confidence we have that the sample mean emission rate will lie within plus or minus some distance of the population mean emission rate (μ).

$$P\left[\mu - z_{\alpha/2}\sigma_{\mu} \le \overline{E} \le \mu + z_{\alpha/2}\sigma_{\mu}\right] = 1-\alpha \quad (11)$$

where

μ	=	population mean emission rate
z	=	normal standard deviate
α	=	probability associated with sample mean falling outside of confidence limits
$\overline{E}$	=	sample mean emission rate
σ_{μ}	=	standard deviation of the population mean = $\frac{\sigma}{\sqrt{n}}$
σ	=	standard deviation of individual observations about the population mean
n	=	number of observations in the sample

If we define d as the maximum error we are willing to tolerate in our estimate of the population mean emission rate, then

$$d = z_{\alpha/2}\sigma_{\mu} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad (12)$$

and by rearranging we can solve for n

$$n = \frac{z_{\alpha/2}^{2}\sigma^{2}}{d^{2}} \quad (13)$$

Thus, if we know that σ = 1 g/veh-km, and we are willing to accept a maximum error of 0.25 g/veh-km in our estimate of the population mean with a reliability of 95% (i.e. α = 5%), then z = 1.96 and

$$n = \frac{(1.96)^2(1)^2}{(0.25)^2} = 61 \text{ GPS runs are required.}$$

Of course, in practice, we do not know the standard deviation of individual vehicle average emission rates for the population (σ), and therefore we can not actually use Equation 13.

There exist several choices for dealing with this problem. In this paper we examine two approaches, namely the two-stage sampling method, and the 6-sigma method. Each of these methods is described in the following sections.

Two-Stage Sample Method

In the two-stage sampling method, an initial sample is taken, from which the sample standard deviation ($\hat{\sigma}$) is computed. This sample standard deviation is used as a point estimate of the population standard deviation (σ). On the basis of the estimated population standard deviation, the sample size required to satisfy some prescribed maximum error is computed using Equation 13.

Of course, any errors in the estimate of the population standard deviation will result in the square of the error in the estimate of the sample size required. For example, estimating the population standard deviation as 1.25, when the true population standard deviation is actually 1 (an error of 25%) will result in an estimate of the required sample size that is 56% higher than is necessary.

To quantify the error associated with using this two stage sampling approach, we applied this method to the field data. A range of initial sample sizes was considered. For each, a random sample of the GPS runs was chosen from the 72 runs. A total of 20 samples (repetitions) were selected for each initial sample size examined. For each sample, the sample standard deviation was computed and used in Equation 13 as a point estimate of the population standard deviation. It was assumed that the standard deviation and mean computed from all 72 GPS runs represented the population. A reliability of 95% (i.e. $z_{\alpha/2}$ = 1.96) was chosen, and the maximum error was specified as 10% of the population mean. Thus, for CO, d = 0.25 g/veh-km.

Figure 8 illustrates the mean, maximum, and minimum estimates of the required sample size as a function of the size of the initial sample for CO. Similar results were obtained for HC, NO_x, and fuel. Using the population standard deviation computed from all 72 runs in Table 4 (rather than a point estimate from a sample) indicates that 7.6 runs are required to meet the maximum error constraint for estimates for CO. However, it is evident from Figure 8 that using the 2-stage sampling approach with an initial sample size up to even 30 runs can result in estimates of the required sample size that exceeds 10 runs.

As is evident from Figure 8, there is a trade-off between the size of the initial sample and the reliability with which it is possible to predict the sample size required to satisfy the desired maximum error in the population estimate. If the initial sample consists of 6 runs, then using Equation 13 will result in an estimate of the required sample size that will range from 3 to 21 runs. As the initial sample size increases, the range in the estimate of the required sample size decreases.

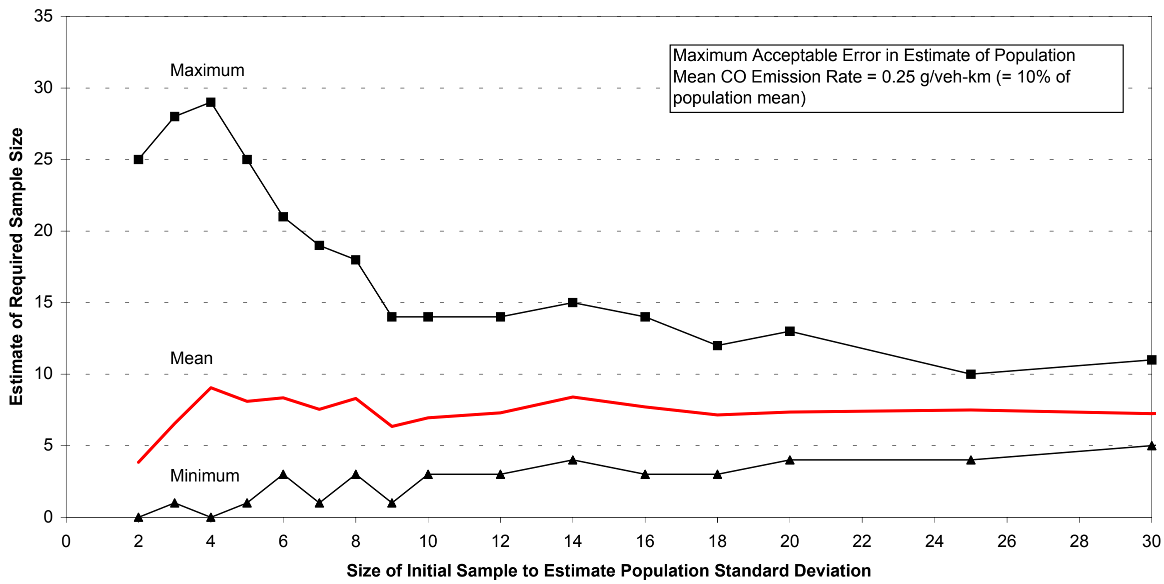

Figure 8: Range of estimates of required sample size as a function of size of initial sample size

On the basis of the results obtained, it can be concluded that the two-stage sampling method suffers from three undesirable characteristics. First, the method cannot be used to estimate the required sample size without first collecting the initial sample. In many cases, such as estimating data collection costs for proposals, or developing a data collection schedule, taking an initial sample is not feasible. Second, for the corridor examined in this paper, a relatively large initial sample size is required to estimate the required sample size with even a reasonable degree of accuracy. Third, errors in the estimate of the population standard deviation produce a square of this error in the estimate of the required sample size.

Six-Sigma Method

The 6-sigma method is an alternative to the two-stage sampling method for estimating the sample size required to achieve some desired level of accuracy in estimating average fuel consumption rate or average emission rate. In the 6-sigma method, it is assumed that the population of mean fuel consumption and emission rates is Normally distributed. On the basis of this assumption, it is known that 99.7% of individual observations (i.e. mean fuel consumption and emission rates) will fall within $\pm 3\sigma$ of the population mean. We define the upper and lower bounds as

$$B_U = \mu + 3\sigma$$
$$B_L = \mu - 3\sigma$$

where:

B_U	=	99.85% upper bound on mean fuel consumption and emission rate
B_L	=	0.15% lower bound on mean fuel consumption and emission rate
σ	=	population standard deviation of mean fuel consumption and emission rate
μ	=	Population average of individual mean fuel consumption and emission rate

Subtracting the lower bound from the upper bound is equal to 6σ (thus the name of the method). The population standard deviation is then equal to

$$\sigma = \frac{B_U - B_L}{6} \tag{14}$$

Of course, in order to use Equation 14, we must be able to estimate values for B_U and B_L. We cannot compute these values directly as they depend on the unknown population mean and standard deviation. However, we can estimate these values by attempting to estimate the minimum and maximum expected mean fuel consumption and emission rate.

For example, we can safely assume that the lowest average fuel consumption and emission rate for a vehicle trip along the Scottsdale corridor would be associated with a vehicle travelling at a constant speed (i.e. zero acceleration). For zero acceleration, Equation 1 becomes,

$$\ln(F) = a_0 + a_4 S + a_5 S^2 + a_6 S^3 \tag{15}$$

From this we can say that the lowest average CO emission rate we would expect could be determined by examining a plot of emission (mg/veh-km) as a function of speed. Figure 9 indicates that the minimum expected CO emission is 1240 mg/veh-km (i.e. for CO B_L = 1240 mg/veh-km).

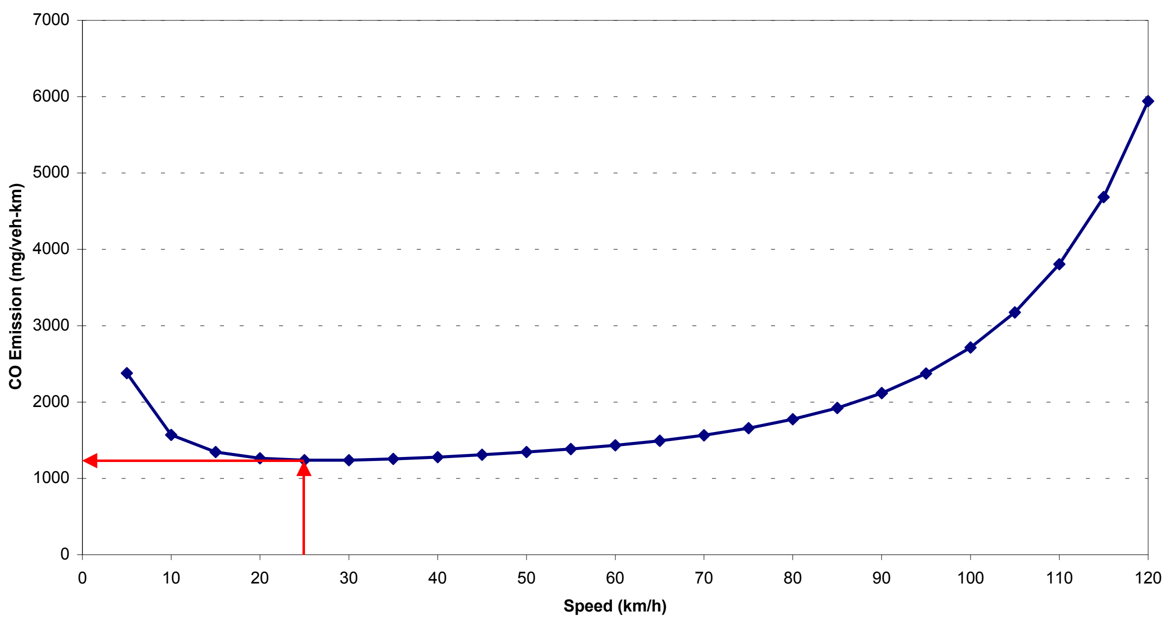

Figure 9: CO Emission as a function of speed for zero acceleration (from Equation 15)

The maximum emission rate is more difficult to estimate. However, we can examine some candidate boundary values by applying Equation 1 to a set of predefined driving cycles. These driving cycles have been developed for use in the US EPA energy and emissions model Mobile 6. Equation 1 was applied to a total of 10 predefined driving cycles, 4 for arterial roadways and 6 for freeways. In each case the average fuel consumption and emission rate computed over the driving cycle (Table 6).

Table 6: Average fuel consumption and emission rates for selected predefined driving cycles

Driving Cycle	Emission (g/veh-km)			Fuel (ml/veh-km)
	CO	HC	NOx	
Arterial LOS A-B	2.35	0.14	0.35	129.60
Arterial LOS C-D	2.47	0.16	0.37	148.74
Arterial LOS E-F	2.78	0.20	0.47	204.30
FTP Urban	2.08	0.15	0.29	150.68
Frwy High-Speed	3.85	0.19	0.36	95.60
Frwy LOS A-C	3.47	0.16	0.35	95.47
Frwy LOS D	3.27	0.15	0.34	97.53
Frwy LOS E	2.46	0.13	0.32	115.28
Frwy LOS F	2.32	0.15	0.35	151.87
Frwy LOS G	1.74	0.15	0.25	166.74

Since the Scottsdale corridor is an arterial roadway, it is likely most appropriate to consider only those driving cycles associated with arterials (i.e. first 4 rows in Table 6). The driving cycle having the largest fuel consumption and emission rate is the Arterial LOS E-F driving cycle. This driving cycle (Figure 10) is representative of arterial driving under congested traffic conditions.

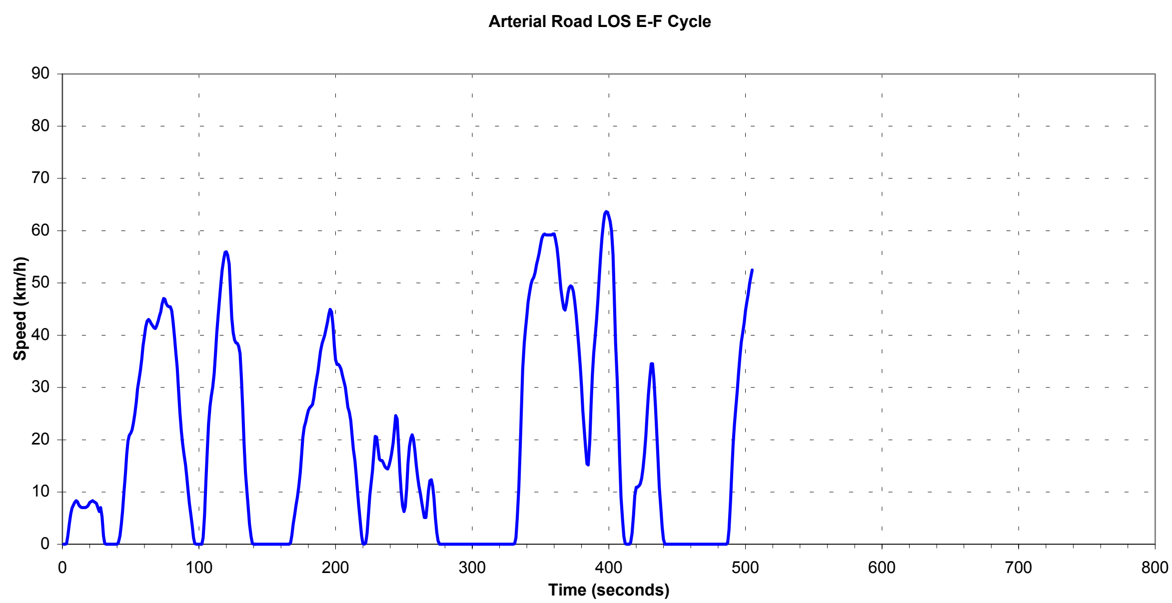

Figure 10: Arterial LOS E-F driving cycle

Having made estimates for B_U and B_L for each emission and for fuel consumption, Equation 14 can be used to estimate the associated population standard deviation. For example, for CO, the population standard deviation is estimated as σ = (2780 - 1240)/6 = 257 mg/veh-km. In comparison, the true population standard deviation (Table 4) is 350 mg/veh-km.

Figure 11 illustrates the estimated number of sample runs required to meet a specified maximum allowable error in the estimate of the population mean CO emission rate. In this figure, the maximum allowable error is expressed as a function of the population mean CO emission rate (2.515 g/veh-km). Two curves are shown in Figure 11. The upper curve represents the sample size required when the true population standard deviation is actually known. The lower curve represents the sample size estimated using the 6-sigma method. The difference between the estimates of the two methods is quite large for very small maximum estimation errors (i.e. < 5%). However, when the maximum allowable error exceeds approximately 10% of the population mean, the curves provide rather similar estimates of the required sample size.

The 6-sigma method is straight-forward to implement, however, its accuracy depends entirely on the accuracy of the estimates of B_U and B_L. And since these values are not computed on the basis of sample data, the analyst's judgement of what constitutes appropriate conditions for the upper and lower bounds significantly influences the estimates of the required sample size. Nevertheless, the 6-sigma method does provide a practical means of estimating required sample size without having to take any initial samples.

We can compare the 2-stage sampling method with the 6-sigma method. Assuming we specify a maximum error of 10% of the population mean, then the 6-sigma method indicates the required sample size is 4 runs. The actual number of required runs, based on use of the population standard deviation, is 8. Thus the use of only 4 runs will result in a maximum error of 14% rather than 10%. The results of the 2-stage sampling method depend on the size of the initial sample. If we assume the initial sample is 6, then the estimates of the required sample size may range from 3 to 21 runs. Obviously, as 6 runs are already taken in the 1-stage of the sampling, any estimate for the required sample size that is less than or equal to 6 would mean no additional runs are to be performed. Therefore, with an initial sample of 6, the maximum error would be 11%. If a sample of 21 runs were used, the maximum error would only be 6%, but 5 times more runs would be required than using the 6-sigma method.

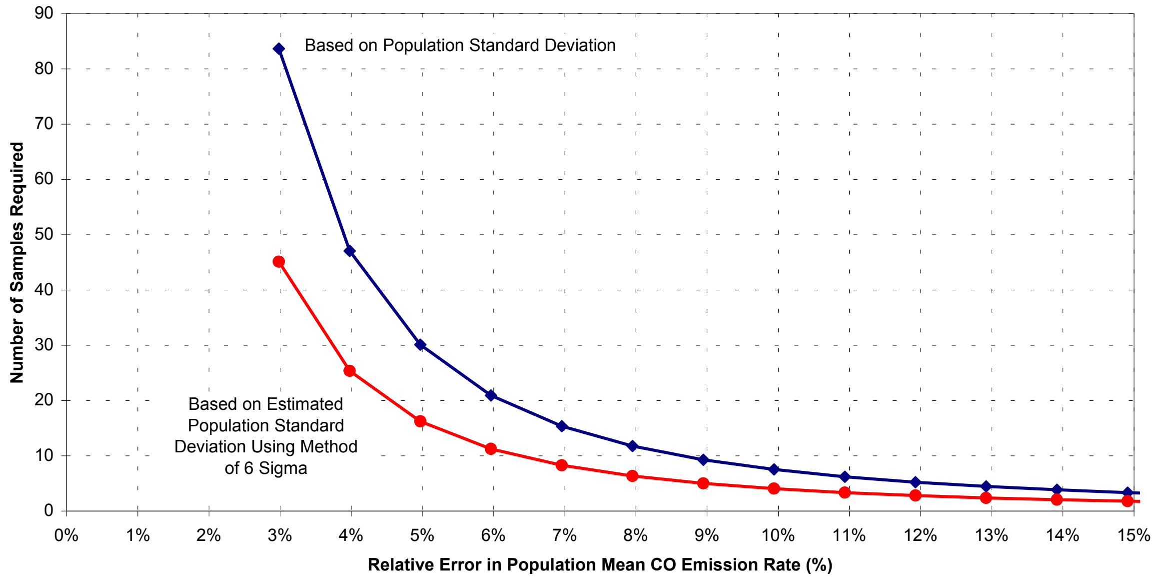

Figure 11: Required sample size for estimating the population mean CO emission rate

Conclusions and Recommendations

The availability of low cost GPS receivers provide a mechanism for collecting detailed vehicle speed and position data cost effectively. In addition to other uses, these data can be used to estimate fuel consumption and vehicle tail-pipe emissions. On the basis of the study described in this paper, the following conclusions can be made:

1. As a result of sources of error inherent within the GPS system, errors in speed measurements occur. Since instantaneous acceleration is computed on the basis of these speed measurements, errors in the speed data create errors in the computed acceleration rates.

2. Fuel consumption and emissions models having a polynomial structure (e.g. VTMicro Model) can be particularly sensitive to speed or acceleration errors. Consequently, a robust method is required to filter the GPS speed and computed acceleration data to ensure values are within feasible vehicle operating conditions.

3. The Epanechnikov Kernel (EK) method of smoothing, applied to the raw GPS speed data, appears to be quite effective in producing speed and acceleration data within the defined feasible region. Furthermore, the EK method is simple to implement and computationally efficient.

4. The use of an average speed profile, rather than individual vehicle profiles, results in estimates of average fuel consumption and emission rates that are under estimated by 25% to 45%. These errors are quite large, and consequently, it is recommended that average emission and fuel consumption rates be estimated on the basis of individual trip emission rates, rather than mean vehicle speed profiles.

5. The 6-sigma method is a statistically valid method that can be practically implemented for determining the sample size required for estimating average fuel consumption and emission rates.

6. The two-stage sampling method is statistically valid; however, the method is not as attractive as the 6-sigma method for practical reasons. For example, the 2-stage sampling method cannot be used to estimate the cost of data collection or devise a data collection schedule unless the first set of data are collected.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Natural Science and Engineering Research Council of Canada.

References

Ahn, K., A. A. Trani, H. Rakha, and M. Van Aerde (1999) "Microscopic Fuel consumption and emission models" Presented at the 78th annual meeting of the Transportation Research Board, Washington D.C.

An, F., M. Barth, M. Ross and J. Norbeck. (1997) "The Development of a Comprehensive Modal Emissions Model: Operating Under Hot-Stabilized Conditions". *Transportation Research Record 1587*. TRB, National Research Council, Washington, D. C. pp. 52-62

Barth, M., T. Younglove, T. Wenzel, G. Scora, F. An, M. Ross, and J. Norbeck (1997) "Analysis of Modal Emissions from a Diverse In Use Vehicle Fleet," *Transportation Research Record 158*7, TRB, National Research Council, Washington, D. C. pp. 73-84.

Barth, M., F. An, T. Younglove, C. Levine, G. Scora, M. Ross, and T. Wenzel. (1999) "The development of a comprehensive modal emissions model". (Draft) Final report to the National Cooperative Highway Research Program, November, 1999, 255 pages

Barth, M., F. An, T. Younglove, G. Scora, C. Levine, M. Ross, and T. Wenzel. (2000) "Comprehensive Model Emissions Model (CMEM), Version 2 User's Guide" University of California at Riverside.

Brzezinski, D., P. Enns, and C. Hart (1999) "Facility-Specific Speed Correction Factors" Report Number M6.SPD.002 (EPA420-P-99-002), published by the Assessment and Modelling Division, Office of Mobile Sources, U.S. Environmental Protection Agency

Rakha, H., F. Dion, and H.G. Sin (2002) "Field evaluation of energy and emission impacts of traffic flow improvement projects using GPS data: Issues and proposed solutions". Presented at the 81st annual meeting of the Transportation Research Board, Washington, DC.

Rakha, H., M. Van Aerde, K. Ahn, and A. A. Trani (2000b) "Requirements for evaluating traffic signal control impacts on energy and emissions based on instantaneous speed and acceleration measurements." Presented at the 79th annual meeting of the Transportation Research Board, Washington D.C.

Rakha, H., A. Medina, H. Sin, F. Dion, and M. Van Aerde (2000c) "Coordination of traffic signals across jurisdictional boundaries: field and modelling results." Presented at the 79th annual meeting of the Transportation Research Board, Washington D.C.

Sierra Research (1997) "Development of Speed Correction Cycles", US EPA Document number M6.SPD.001