United States
Environmental Protection
Agency

EPA/600/R-07/106
July 2007

Transit Bus Load-Based Modal Emission Rate Model Development

Chapters 9 through 11 elaborate the different emission models developed for idle, deceleration, acceleration and cruise driving modes.

Chapter 12, research results are verified.

Chapter 13 presents a discussion and conclusion on research results

Chapter 14 References

CHAPTER 9

9. DECELERATION MODE DEVELOPMENT

Chapter 7 introduced the concept of driving mode into the study and several sensitivity tests were performed for four different definitions, including comparison of modal average emission rate estimates, HTBR regression tree results, and residual mean deviance. After developing the idle mode definition and emission rate in Chapter 8, the next task is dividing the rest of the vehicle activity data into driving mode (deceleration, acceleration and cruise) for further analysis. The deceleration mode is examined first.

9.1 Critical Value for Deceleration Rates in Deceleration Mode

The first task related to analysis of emission rates in the deceleration mode is identifying critical values for deceleration. The literature indicates that critical values of -1 mph/s and -2 mph/s should be examined. Because the critical value of "acceleration < -1 mph/s" also includes all data that conform with a critical value of "acceleration < -2 mph/s", comparison of data that fall between these two potential cut points is first performed. In summary, these three deceleration bins for analysis include:

- Option 1: acceleration < -2 mph/s
- Option 2: acceleration ≥ -2 mph/s & acceleration < -1 mph/s
- Option 3: acceleration ≥ -1 mph/s & acceleration < 0 mph/s

If the critical value is set as -1 mph/s for deceleration mode, data falling into option 1 and option 2 will be classified as deceleration mode while data falling into option 3 will be classified as cruise mode. If the critical value is set as -2 mph/s for deceleration mode, data falling into option 1 will be classified as deceleration mode while data falling into option 2 and option 3 will be classified as cruise mode.

Figure 9-1 illustrates engine power distribution for these three options. Figures 9-2 to 9-4 compare engine power vs. emission rate for three pollutants for three options. Tables 9-1 and 9-2 provide the distribution for these three options in two ways: by number and percentage.

Table 9-1 Engine Power Distribution for Three Options for Three Pollutants

Deceleration	Pollutants	Engine Power (brake horsepower (bhp))					
		(0 20)	(20 30)	(30 40)	(40 50)	≥ 50	Total
Option 1	NO_x	9322	94	16	5	15	9452
	CO	9558	89	15	4	15	9681
	HC	9483	94	16	5	15	9613
Option 2	NO_x	6748	127	101	42	174	7192
	CO	6800	126	99	42	171	7238
	HC	6754	125	99	42	172	7192
Option 3	NO_x	6806	950	1062	562	4353	13733
	CO	6782	949	1061	558	4326	13676
	HC	6705	921	1044	541	4212	13423

Table 9-2 Percentage of Engine Power Distribution for Three Options for Three Pollutants

Deceleration	Pollutants	Engine Power (brake horsepower (bhp))					
		(0 20)	(20 30)	(30 40)	(40 50)	≥ 50	Total
Option 1	NO_x	98.6%	1.0%	0.2%	0.1%	0.2%	100.0%
	CO	98.7%	0.9%	0.2%	0.0%	0.2%	100.0%
	HC	98.6%	1.0%	0.2%	0.1%	0.2%	100.0%
Option 2	NO_x	93.8%	1.8%	1.4%	0.6%	2.4%	100.0%
	CO	93.9%	1.7%	1.4%	0.6%	2.4%	100.0%
	HC	93.9%	1.7%	1.4%	0.6%	2.4%	100.0%
Option 3	NO_x	49.6%	6.9%	7.7%	4.1%	31.7%	100.0%
	CO	49.6%	6.9%	7.8%	4.1%	31.6%	100.0%
	HC	50.0%	6.9%	7.8%	4.0%	31.4%	100.0%

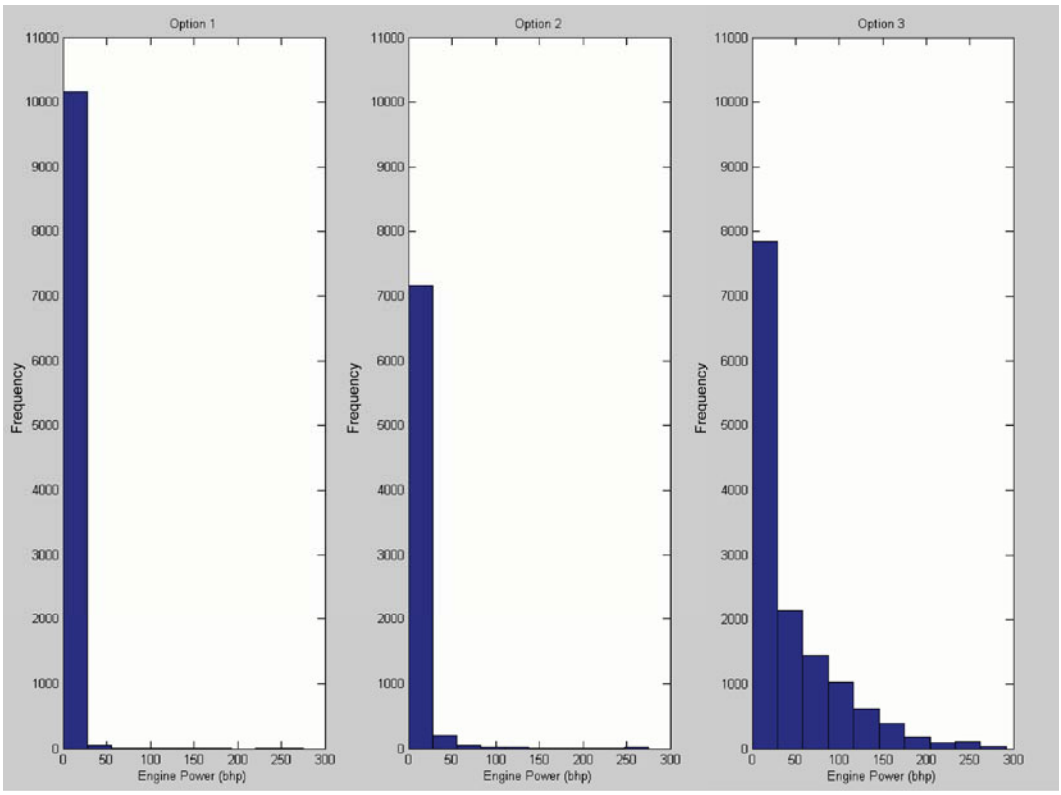


Figure 9-1 Engine Power Distribution for Three Options

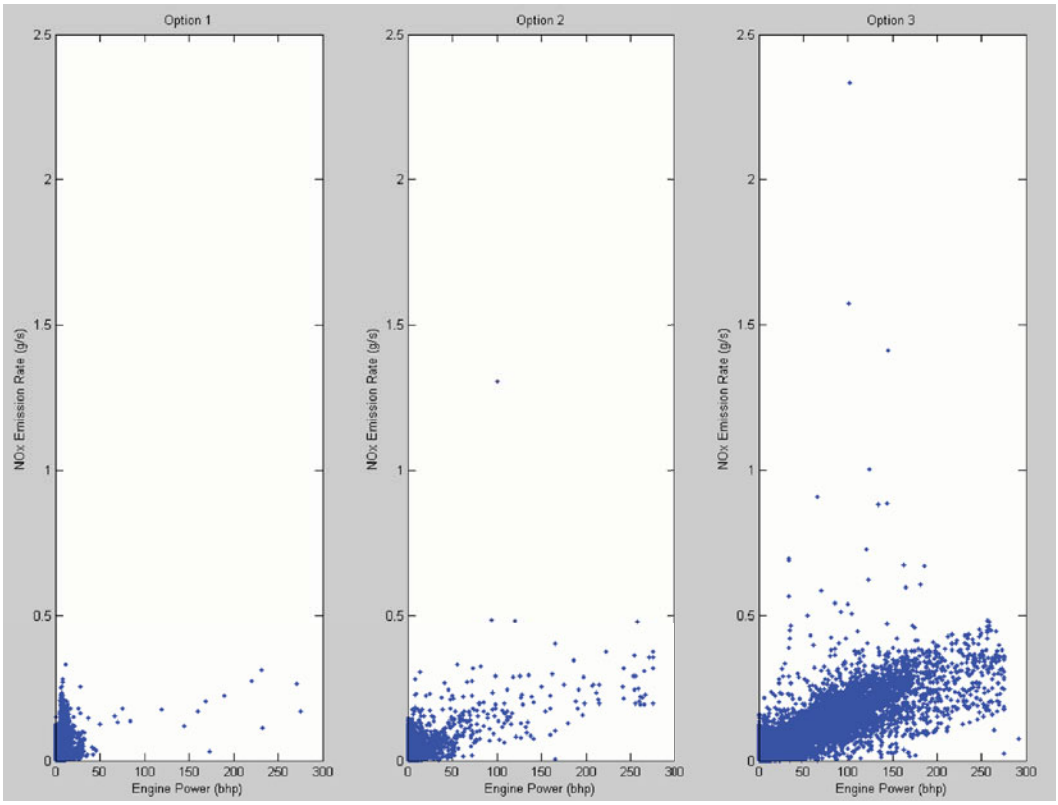


Figure 9-2 Engine Power vs. NO_x Emission Rate for Three Options

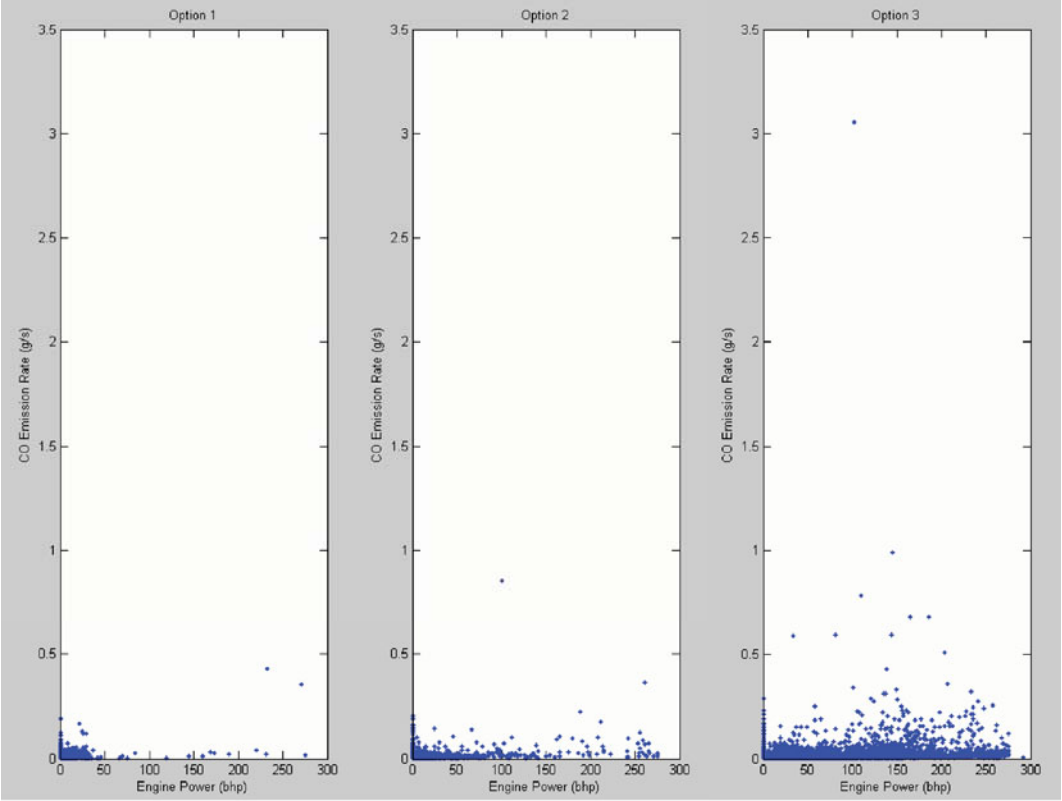


Figure 9-3 Engine Power vs. CO Emission Rate for Three Options

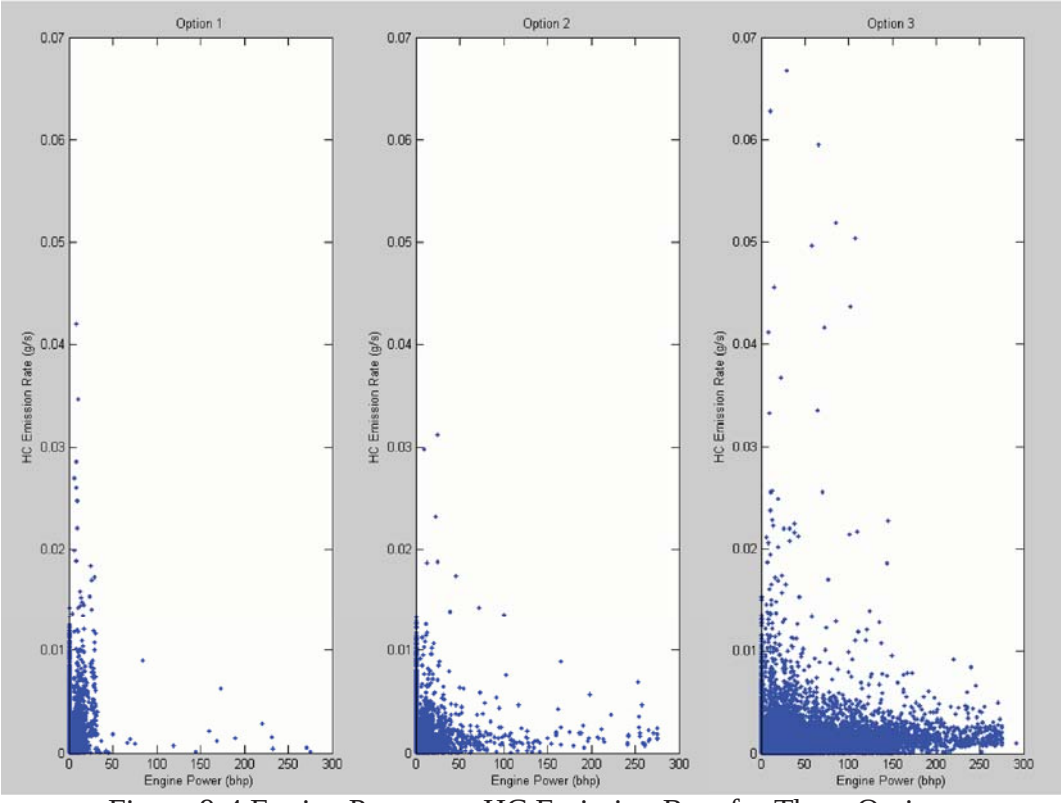


Figure 9-4 Engine Power vs. HC Emission Rate for Three Options

There is little difference in the engine power distributions noted for data falling into option 1 and option 2 while the power distribution for option 3 is obviously different from option 1 and option 2 in the above figures and tables. Tables 9-1 and 9-2 show that the engine power is more concentrated in the lower engine power regime (< 20 bhp) for data in deceleration mode. Tables 9-1 and 9-2 better reflect the power demand of the vehicle in real world in deceleration mode. Hence, the critical value is set to -1 mph/s for deceleration mode.

9.2 Analysis of Deceleration Mode Data

9.2.1 Emission Rate Distribution by Bus in Deceleration Mode

After defining vehicle activity data with "acceleration <-1 mph/s" as deceleration mode, emission rate histograms for each of the three pollutants for deceleration operations are presented in Figure 9-5. Figure 9-5 shows significant skewness for all three pollutants for deceleration mode. Inter-bus emission rate variability is illustrated by plotting median and mean NO_x, CO, and HC emission rates in deceleration mode for each bus in Figures 9-6 to 9-8 and Table 9-3. The difference between median and mean is also an indicator of skewness.

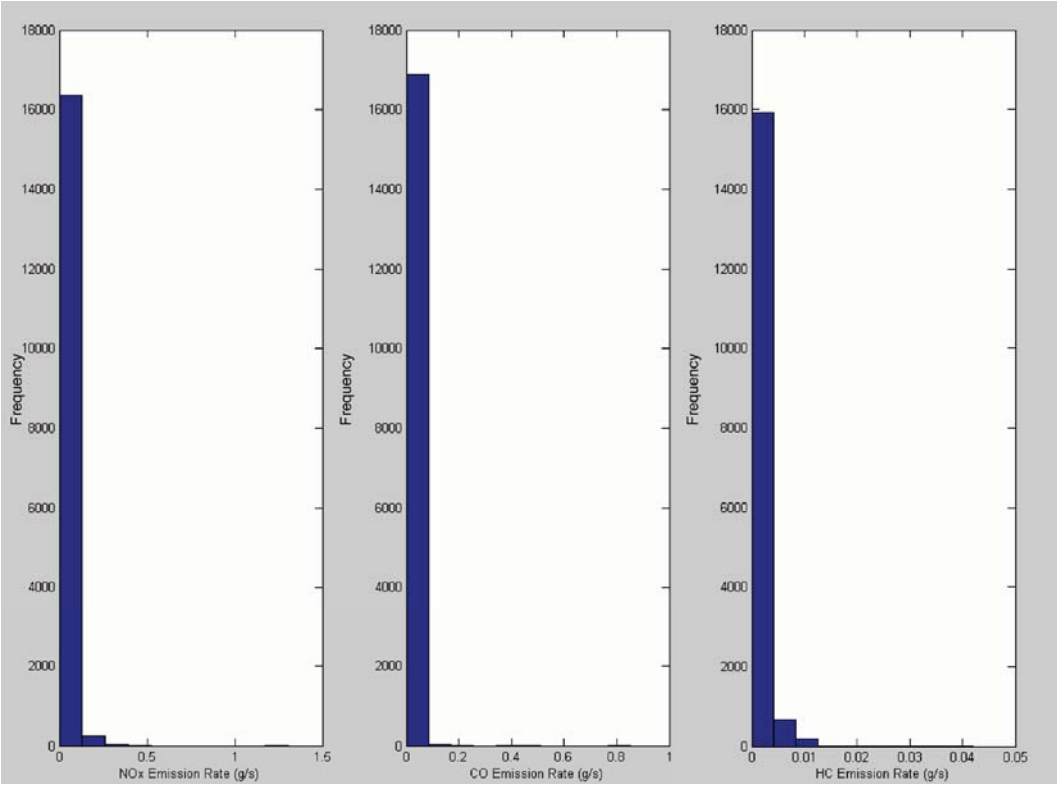


Figure 9-5 Histograms of Three Pollutants for Deceleration Mode

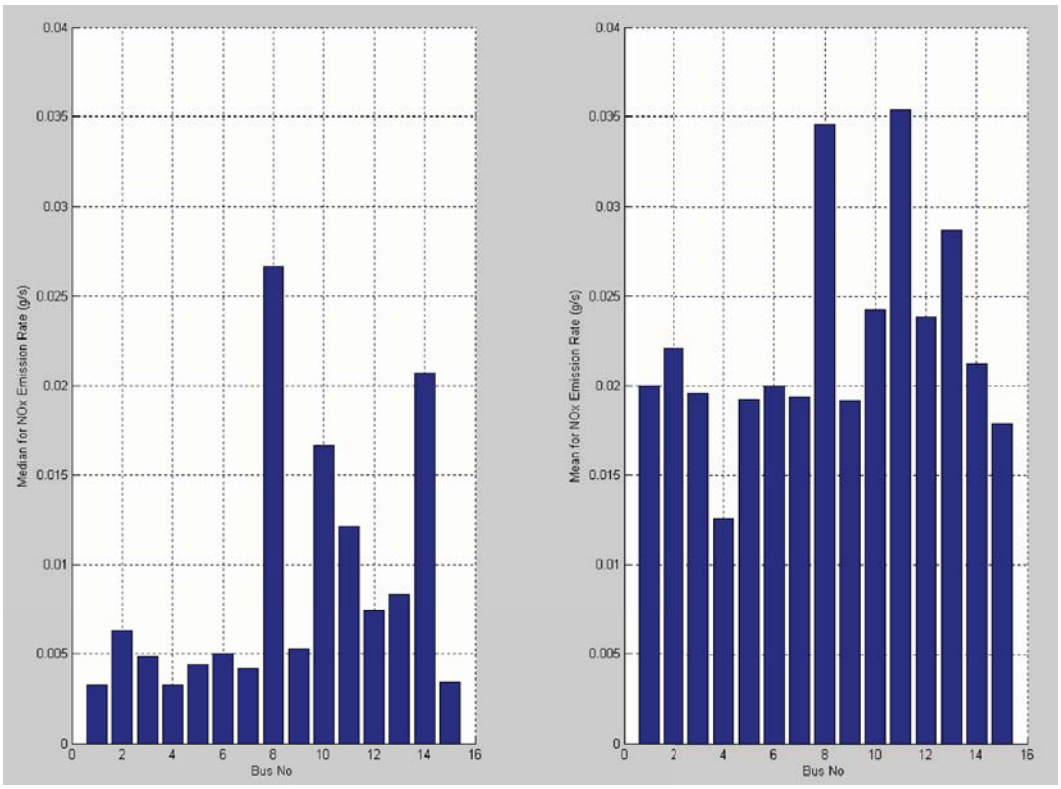


Figure 9-6 Median and Mean of NO_x Emission Rates in Deceleration Mode by Bus

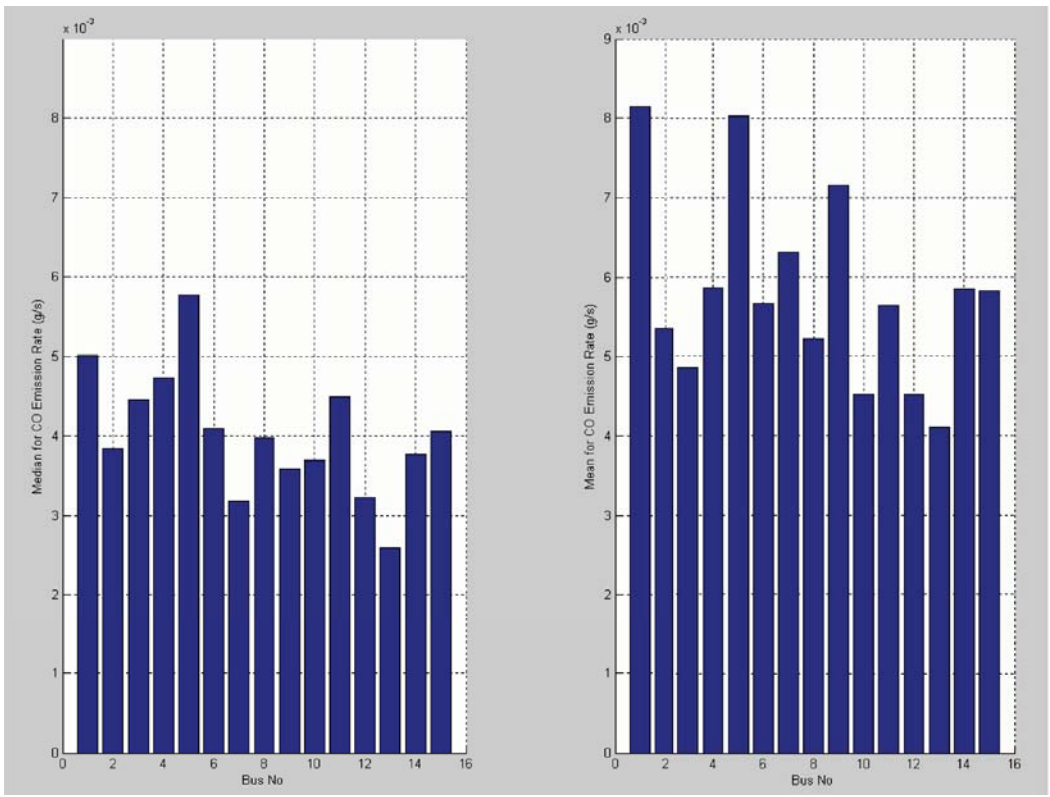


Figure 9-7 Median and Mean of CO Emission Rates in Deceleration Mode by Bus

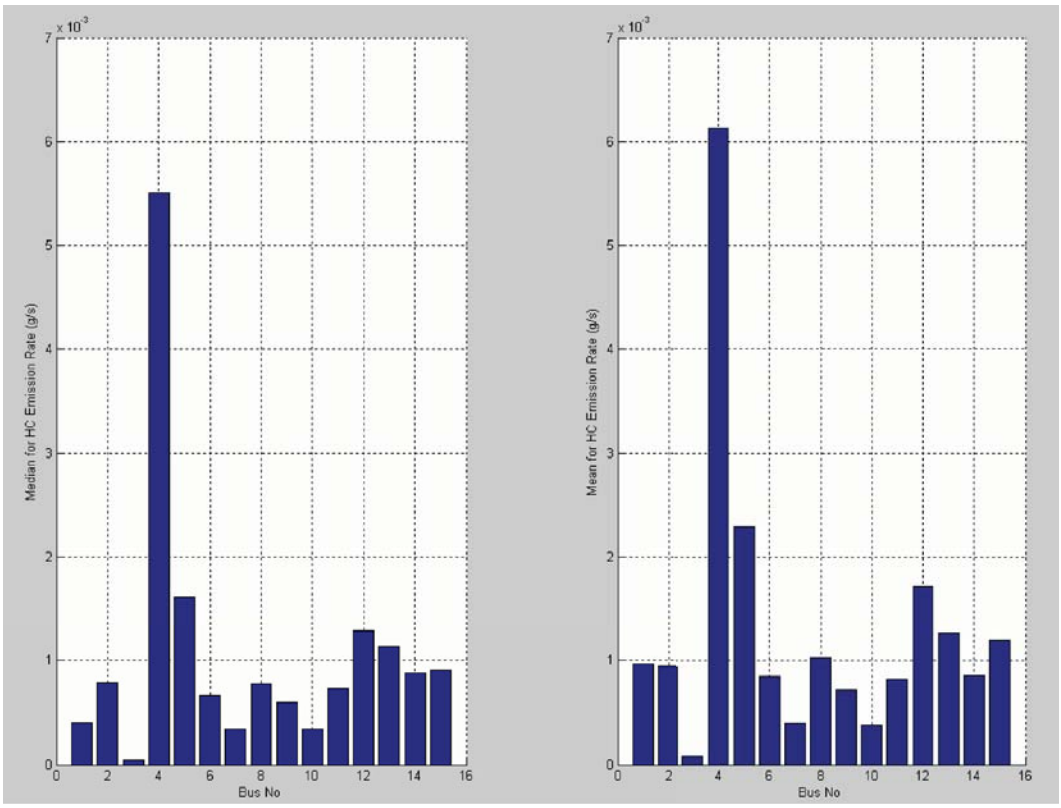


Figure 9-8 Median and Mean of HC Emission Rates in Deceleration Mode by Bus

Table 9-3 Median, and Mean for NO_x, CO, and HC in Deceleration Mode by Bus

	NO_x		CO		HC	
Bus ID	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Bus 360	0.00325	0.01998	0.00502	0.00814	0.00040	0.00097
Bus 361	0.00624	0.02206	0.00384	0.00535	0.00079	0.00095
Bus 363	0.00483	0.01952	0.00446	0.00486	0.00004	0.00008
Bus 364	0.00324	0.01255	0.00474	0.00586	0.00551	0.00613
Bus 372	0.00437	0.01924	0.00578	0.00803	0.00161	0.00229
Bus 375	0.00499	0.01997	0.00410	0.00567	0.00066	0.00085
Bus 377	0.00414	0.01940	0.00317	0.00630	0.00034	0.00040
Bus 379	0.02664	0.03457	0.00397	0.00522	0.00078	0.00103
Bus 380	0.00525	0.01914	0.00359	0.00716	0.00060	0.00072
Bus 381	0.01666	0.02420	0.00369	0.00452	0.00034	0.00038
Bus 382	0.01214	0.03541	0.00450	0.00564	0.00073	0.00083
Bus 383	0.00741	0.02385	0.00322	0.00452	0.00128	0.00172
Bus 384	0.00828	0.02869	0.00259	0.00411	0.00113	0.00127
Bus 385	0.02066	0.02118	0.00377	0.00585	0.00088	0.00086
Bus 386	0.00341	0.01786	0.00406	0.00583	0.00091	0.00120

Figures 9-6 to 9-8 and Table 9-3 illustrate that bus 379 has the largest median and the second largest mean for NO_x emissions, bus 372 has the largest median and the second largest mean for CO emissions, while bus 364 has the largest median and mean for HC emissions. At the same time, bus 382 has the largest mean for NO_x emissions, and bus 360 has the largest mean for CO emissions. The above figures and table demonstrate that although variability exists among buses, it is difficult to determine which, if any, bus is a high emitter (i.e., a bus that exhibits extremely high emission rates under all operating conditions, which also may exhibit significantly different emissions responses to operating activity than normal emitters).

The modeler notices that there is also a small number of some very high HC emissions events noted in deceleration mode. Based on definitions of "acceleration < -1 mph/s", 242/16237=1.49 % of data points in deceleration mode for HC are high emissions. This happened only for HC. This did not occur for NO_x and CO. All high HC emissions have been coded to determine if they are related to any other parameters. Tree analysis could be used for this screening analysis. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, no operating parameters appeared to be correlated to these high emissions events.

High HC emissions distribution by bus and trip are presented in Table 9-4. Unlike idle mode where high HC emissions occurred mainly in three idle segments (bus 360, trip 4, idle segment 1; bus 360, trip 4, idle segment 38; and bus 372, trip 1, idle segment 1), high HC emissions are dispersed among seven different buses and 18 different trips. Although there is not enough evidence to suggest a specific bus is a "high emitter", bus 364 is worthy of additional attention. There are 5284 data points for bus 364 and, among them, 887 data points classified as deceleration mode. There are 408 high HC emissions data points for bus 364 in deceleration mode. The percentage of high HC emission for bus 364 is 7.72% (408/5284), while the percentage of high HC emissions for bus 364 in deceleration mode is about 21% (193/887). Given the limited available data, no conclusion could be drawn about high HC emissions in deceleration mode. These potential outliers may simply reflect real-world emissions variability for these engines.

Emission rate behavior as a function of operating mode and power for high-emitting vehicles may differ significantly from normal-emitting vehicles. Since no high-emitting vehicle is identified in the AATA data set, it is impossible for the modeler to examine such a difference. To ensure that models are applicable to normal and high-emitters in the fleet, models have to have both normal and high-emitters available in the analytical data set. Thus it is important to identify high-emitting vehicles and bring them in for testing.

Table 9-4 High HC Emissions Distribution by Bus and Trip for Deceleration Mode

Bus ID	Number of High HC Events	Trip	Number of High HC Events
Bus 360	11	Bus 360, trip 3	3
		Bus 360, trip 4	8
Bus 361	1	Bus 361, trip 5	1
Bus 364	193	Bus 364, trip 1	46
		Bus 364, trip 2	61
		Bus 364, trip 3	86
Bus 372	19	Bus 372, trip 1	6
		Bus 372, trip 2	4
		Bus 372, trip 3	3
		Bus 372, trip 4	6
Bus 383	11	Bus 383, trip 1	3
		Bus 383, trip 2	3
		Bus 383, trip 3	2
		Bus 383, trip 4	3
Bus 384	1	Bus 384, trip 3	1
Bus 386	6	Bus 386, trip 1	1
		Bus 386, trip 2	2
		Bus 386, trip 4	3

9.2.2 Engine Power Distribution by Bus in Deceleration Mode

Engine power distribution by bus is shown in Figure 9-9 and Table 9-5. When the bus is decelerating, the engine typically absorbs energy, yielding low engine power, or even negative engine power. Table 9-5 reflects this characteristic of deceleration mode. According to Sensors, Inc. report (Ensfield 2001), negative engine power is recorded as zero power in the data, which explains the large number of zero power values in the deceleration mode. The emission rates under negative engine power conditions may be signficiantly different from those under positive engine power. Further analysis will examine this question. Moreover, bus 372 has the greatest 3rd Quartile engine power in deceleration mode, consistent with the finding in idle mode.

Table 9-5 Engine Power Distributions in Deceleration Mode by Bus

Bus No	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Bus 360	0	0	0	3.88	275.40
Bus 361	0	0	0	5.16	173.10
Bus 363	0	0	0	6.70	274.90
Bus 364	0	0	0	0	254.30
Bus 372	0	0	0	20.41	112.00
Bus 375	0	0	0	5.84	274.90
Bus 377	0	0	0	3.33	275.10
Bus 379	0	0	0	11.77	164.90
Bus 380	0	0	0	5.19	29.40
Bus 381	0	0	0	7.19	121.15
Bus 382	0	0	0	5.84	20.75
Bus 383	0	0	0	8.51	94.65
Bus 384	0	0	0	5.86	162.37
Bus 385	0	0	0	6.00	102.59
Bus 386	0	0	0	7.18	42.20

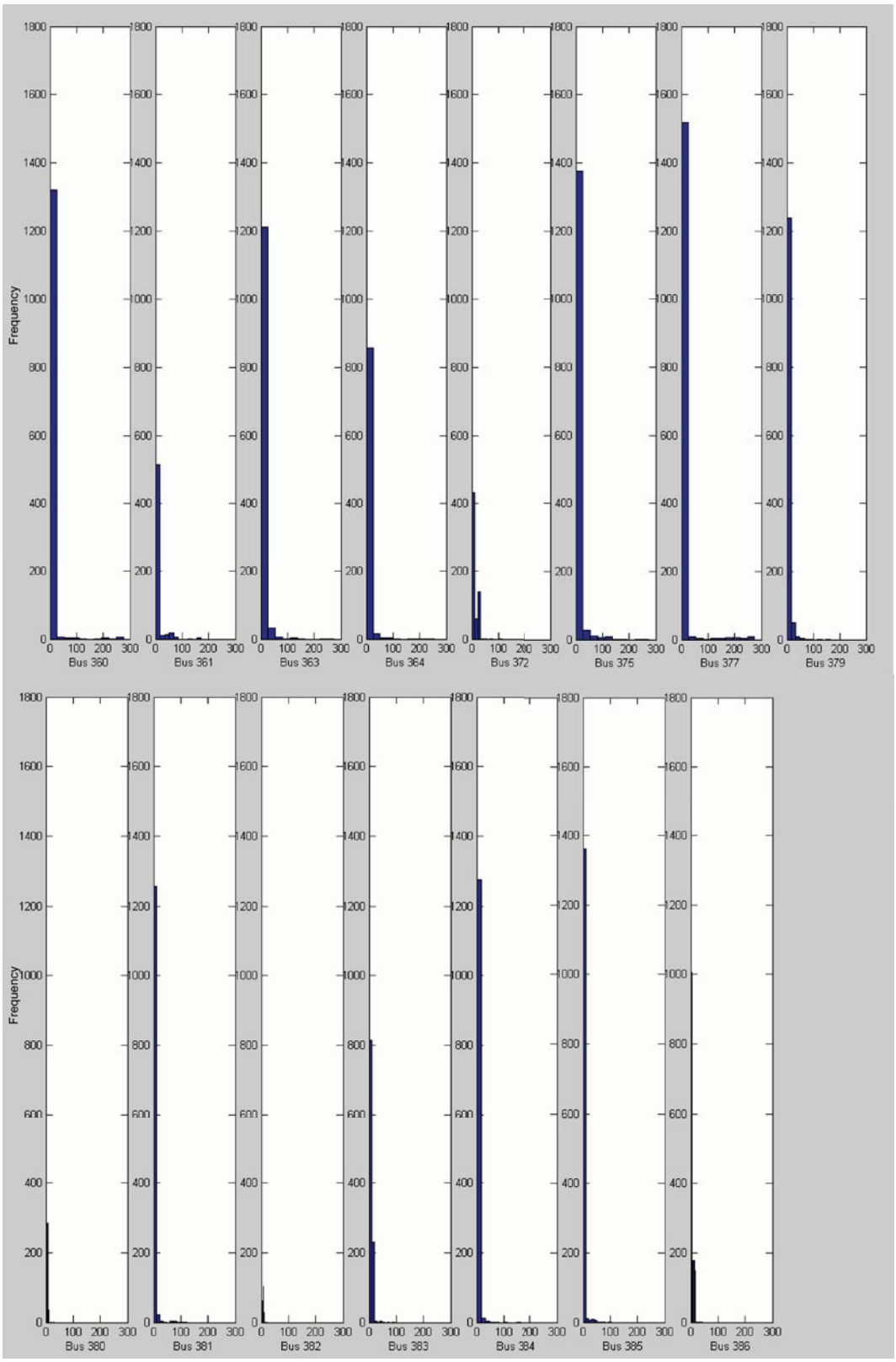


Figure 9-9 Histograms of Engine Power in Deceleration Mode by Bus

Based on definitions of “acceleration < -1 mph/s”, about 1% of data points with high engine power (≥50 bhp) fall in deceleration mode (Table 9-1). Figure 9-10 illustrates plots of engine power vs. vehicle speed, engine power vs. engine speed, and vehicle speed vs. engine speed. Figure 9-10 shows that higher engine power always occurred with higher vehicle speed and higher engine speed. These data points with higher engine power likely reflect the variability of the real world and are all retained in the data set and mode definition to avoid potentially biasing results.

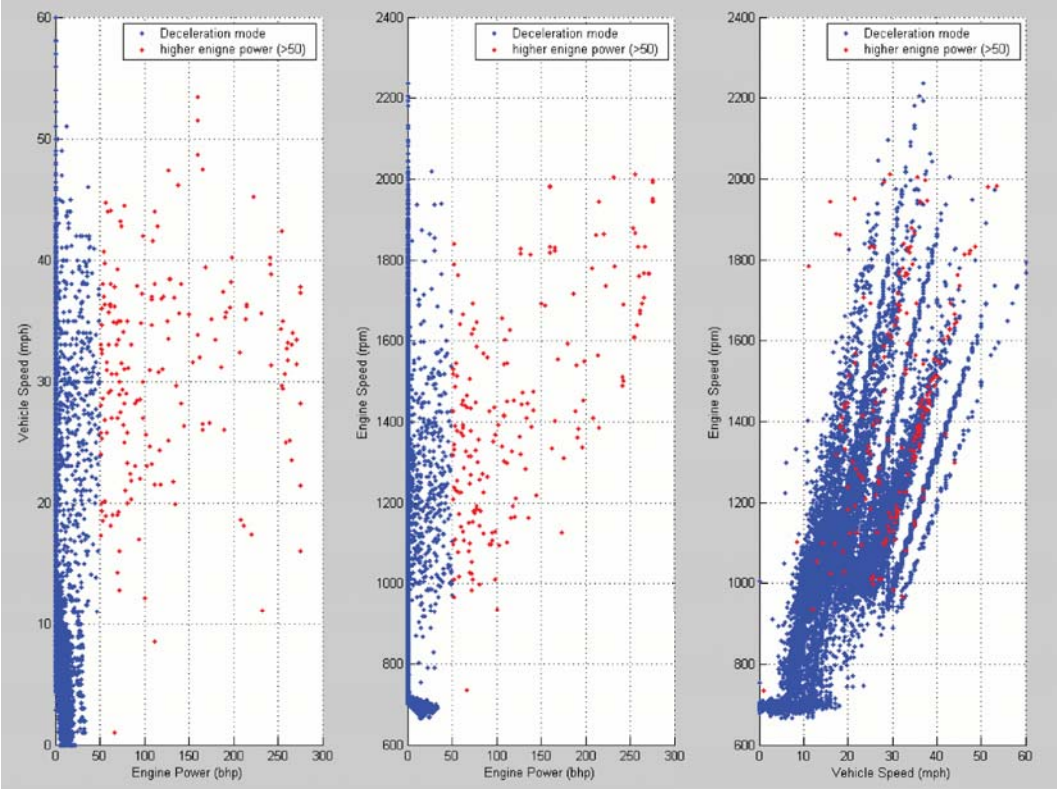
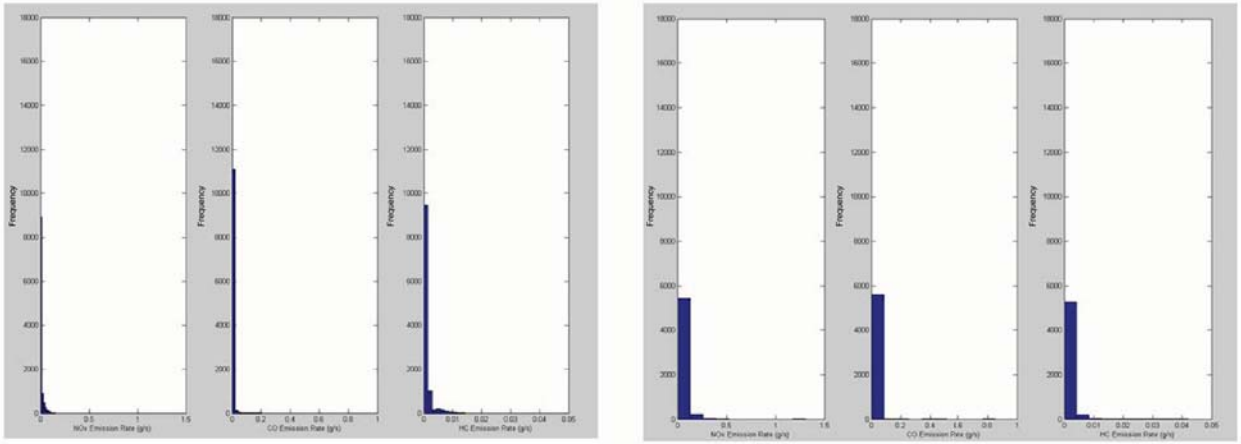


Figure 9-10 Engine Power vs. Vehicle Speed, Engine Power vs. Engine Speed, and Vehicle Speed vs. Engine Speed

9.3 The Deceleration Motoring Mode

Bus engines absorb energy during the deceleration mode, resulting in low or negative engine power. According to the Sensors, Inc. report (Ensfield 2001), such negative power was recorded as zero power. The emissions under these negative engine power conditions may be significantly different from those under positive engine power conditions, and therefore may need to be included in the modeling regime as a separate mode of operation. To examine this possibility, deceleration mode data were split into two mode bins for analysis. The first bin includes all data points with zero engine power in deceleration mode, termed ‘deceleration motoring mode.’ The

remaining data in the deceleration mode, which exhibit positive engine power, are classified as deceleration non-motoring mode. The analysis will begin as a comparison of histograms of three pollutants between deceleration motoring mode and deceleration non-motoring mode (Figure 9-11). Table 9-6 compares the mean, median, and skewness of emission distributions between these two modes for the three pollutants. The statistical results for all deceleration data are also presented as a reference. Figure 9-11 and Table 9-6 show that lower emission rates are more prevalent in the deceleration motoring mode than in the deceleration non-motoring mode. Skewness of emission distributions for deceleration motoring mode is also smaller.



(a) (b)

Figure 9-11 Histograms for Three Pollutants in Deceleration Motoring Mode (a) and Deceleration Non-Motoring Mode (b)

To test the differences between deceleration motoring mode and deceleration non-motoring mode, a Kolmogorov-Simirnov two-sample test was chosen rather than a standard *t*-test, because the normal distribution assumption was questionable. The Kolmogorov-Smirnov two-sample test is a test of the null hypothesis that two independent samples have been drawn from the same population (or from populations with the same distribution). The test uses the maximal difference between cumulative frequency distributions of two samples as the test statistic. Results of the Kolmogorov-Smirnov two-sample tests demonstrate that the differences in emission rates under deceleration motoring mode and deceleration non-motoring mode are statistically significant.

Table 9-6 Comparison of Emission Distributions between Deceleration Mode and Two Sub-Modes (Deceleration Motoring Mode and Deceleration Non-Motoring Mode)

	NO_x	CO	HC
	Deceleration Mode		
Number	16644	16919	16805
Minimum	0.00001	0.00001	0.00001
1st Quartile	0.00182	0.00249	0.00039
Median	0.00611	0.00398	0.00068
3rd Quartile	0.03155	0.00605	0.00120
Maximum	1.30640	0.85208	0.04200
Mean	0.02215	0.00580	0.00118
Skewness	6.02890	30.6459	5.76530
Sub-mode 1:Deceleration Motoring Mode			
Number	10925	11304	11240
Minimum	0.00001	0.00001	0.00001
1st Quartile	0.00124	0.00269	0.00041
Median	0.00272	0.00401	0.00067
3rd Quartile	0.00816	0.00567	0.00110
Maximum	0.14930	0.20366	0.01425
Mean	0.00978	0.00528	0.00111
Skewness	3.08780	12.27120	3.92760
Sub-mode 2: Deceleration Non-Motoring Mode			
Number	5719	5615	5565
Minimum	0.00002	0.00003	0.00001
1st Quartile	0.01973	0.00204	0.00034
Median	0.03431	0.00384	0.00069
3rd Quartile	0.05658	0.00741	0.00150
Maximum	1.30640	0.85208	0.04200
Mean	0.04576	0.00685	0.00131
Skewness	5.7018	26.8539	6.8026

9.4 Deceleration Emission Rate Estimations

Using the "acceleration < -1 mph/s" cutpoint, about 16% of total data collected are classified in the deceleration mode. While deceleration emission rates could simply be estimated directly by averaging all deceleration mode emission rates, the emission rate distribution is non-normal. Because lambdas identified by the Box-Cox procedure for the whole dataset and deceleration mode subsets are different, and because using a transformation to estimate the mean and construct confidence intervals will create other problems, the bootstrap (another class of general methods) was used for estimation of the mean and for construction of confidence intervals. The bootstrap function in this study resampled the emission rate data 1000 times and computed the mean, 2.5%, and 97.5% percentile of each sample.

The results of the bootstrap analyses indicate that splitting the deceleration mode into deceleration motoring mode and deceleration non-motoring mode using the zero engine power criteria is warranted. The bootstrap distributions of mean emission rates for deceleration mode, deceleration motoring mode, and deceleration non-motoring mode are presented in Figures 9-12 to 9-14 and Table 9-7. To illustrate the difference in emission rate estimation between deceleration motoring mode and deceleration non-motoring mode, Figure 9-15 presents bootstrap means and confidence intervals for the emission rates of all three pollutants. For reference purposes, deceleration mode emission rate estimations are also presented. Table 9-7 and Figure 9-15 show that the average emission rate for the deceleration motoring mode is much lower than that for deceleration non-motoring mode for all pollutants especially for NO_x.

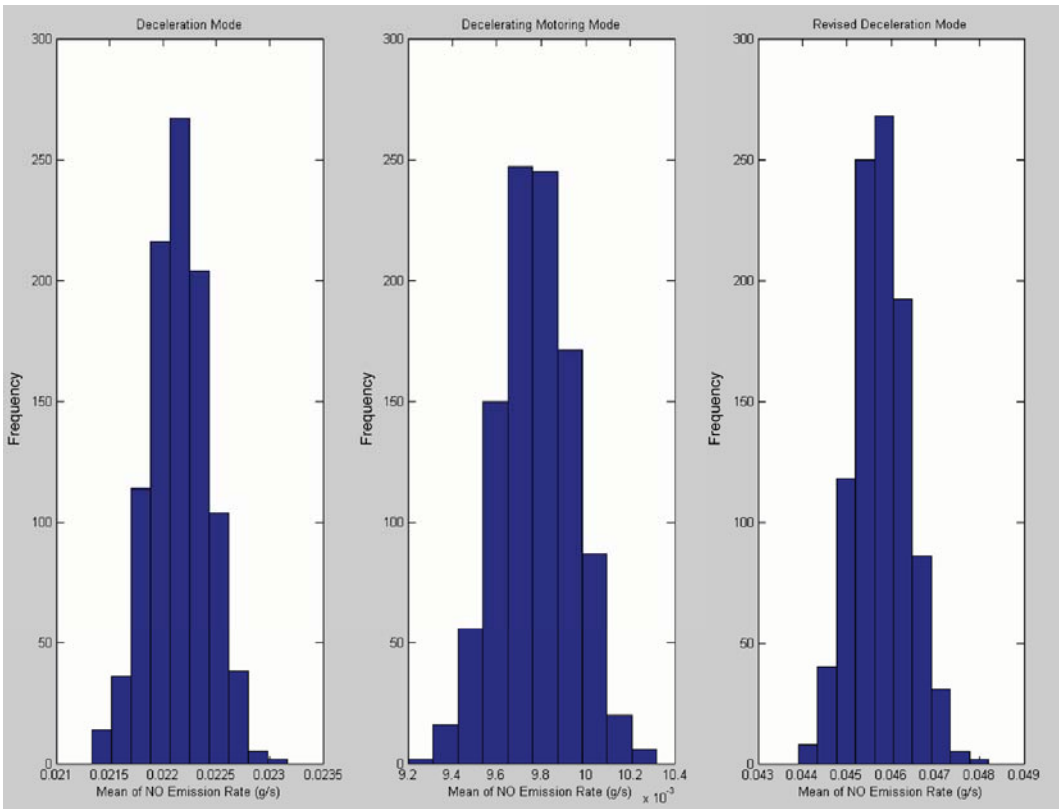


Figure 9-12 Bootstrap Results for NO_x Emission Rate Estimation in Deceleration Mode

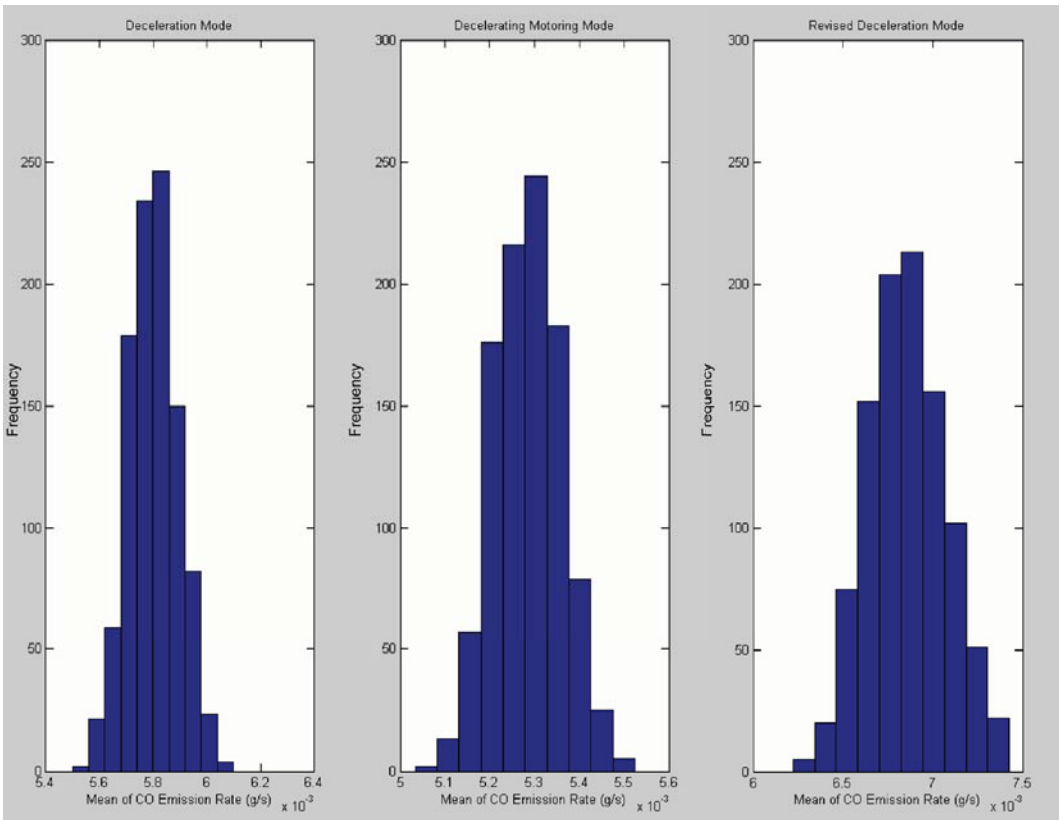


Figure 9-13 Bootstrap Results for CO Emission Rate Estimation in Deceleration Mode

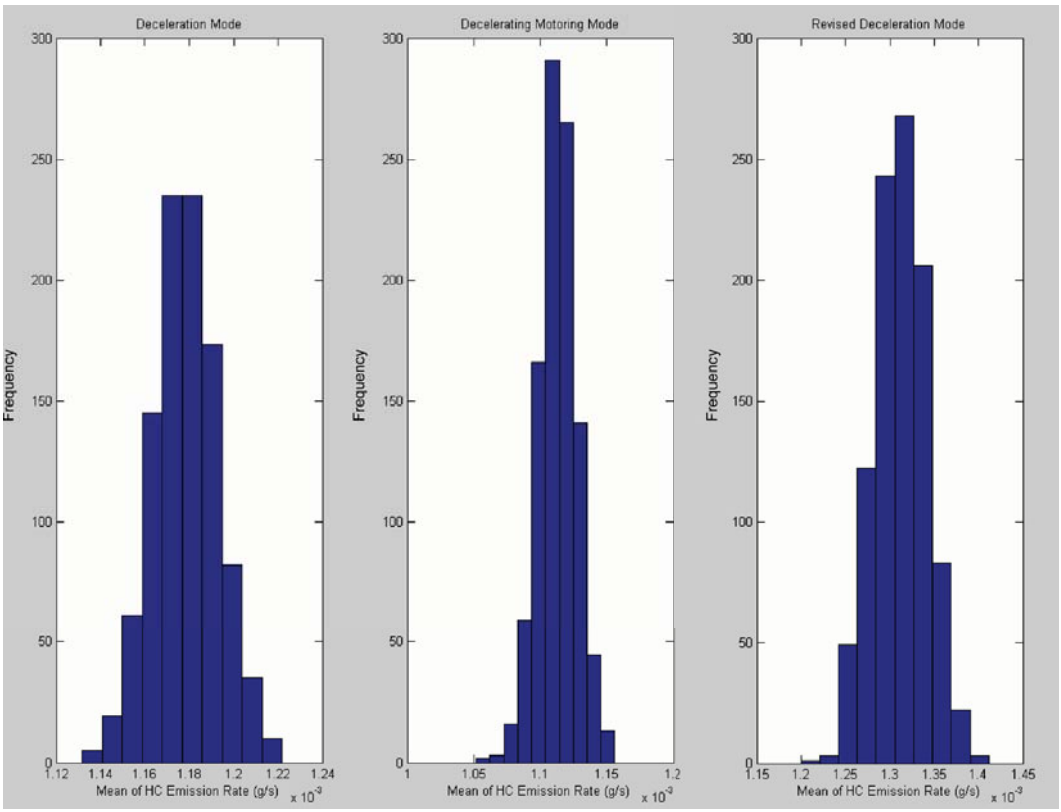


Figure 9-14 Bootstrap Results for HC Emission Rate Estimation in Deceleration Mode

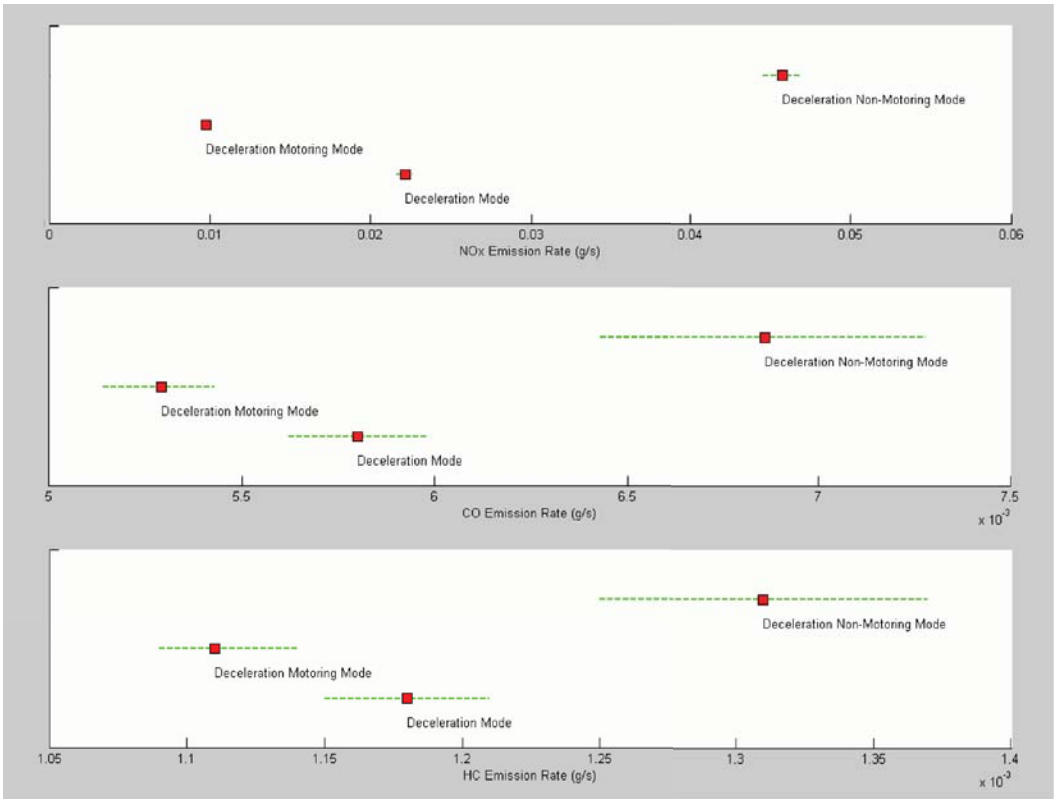


Figure 9-15 Emission Rate Estimation Based on Bootstrap for Deceleration Mode

Table 9-7 Emission Rate Estimation and 95% Confidence Intervals Based on Bootstrap for Deceleration Mode

		Average	2.5% Percentile	97.5% Percentile
Deceleration Mode				
NO_x	Estimation	0.02215	0.00024	0.10919
	Confidence Interval	0.02161	0.00022	0.10427
		0.02268	0.00027	0.11411
CO	Estimation	0.00580	0.00055	0.02191
	Confidence Interval	0.00562	0.00051	0.02067
		0.00598	0.00059	0.02314
HC	Estimation	0.00118	0.00004	0.00652
	Confidence Interval	0.00115	0.00004	0.00626
		0.00121	0.00004	0.00679
Deceleration Motoring Mode				
NO_x	Estimation	0.00978	0.00017	0.06540
	Confidence Interval	0.00945	0.00015	0.06306
		0.01010	0.00019	0.06774
CO	Estimation	0.00529	0.00072	0.01743
	Confidence Interval	0.00514	0.00068	0.01635
		0.00543	0.00075	0.01850
HC	Estimation	0.00111	0.00004	0.00652
	Confidence Interval	0.00109	0.00004	0.00621
		0.00114	0.00004	0.00683
Deceleration Non-Motoring Mode				
NO_x	Estimation	0.04578	0.00173	0.17187
	Confidence Interval	0.04457	0.00152	0.16343
		0.04698	0.00195	0.18031
CO	Estimation	0.00686	0.00037	0.02846
	Confidence Interval	0.00643	0.00033	0.02587
		0.00728	0.00040	0.03104
HC	Estimation	0.00131	0.00004	0.00650
	Confidence Interval	0.00125	0.00003	0.00594
		0.00137	0.00005	0.00706

Based on table 9-7, the deceleration emission rate for NO_x is set as 0.02215 g/s with 95% confidence interval (0.00024 to 0.10919), CO as 0.00580 g/s with 95% confidence interval (0.00055 to 0.02191), HC as 0.00118 g/s with 95% confidence interval (0.00004 to 0.00652). The deceleration motoring emission rate for NO_x is set as 0.00978 g/s with 95% confidence interval (0.00017 to 0.06540), CO as 0.00529 g/s with 95% confidence interval (0.00072 to 0.01743), HC as 0.00111 g/s with 95% confidence interval (0.00004 to 0.00652). The deceleration non-motoring mode emission rate for NO_x is set as 0.04578 g/s with 95% confidence interval (0.00173 to 0.17187), CO as 0.00686 g/s with 95% confidence interval (0.00037 to 0.02846), HC as 0.00131 g/s with 95% confidence interval (0.00004 to 0.00650).

9.5 Conclusions and Further Considerations

In this research, deceleration mode is defined as "acceleration < -1 mph/s". However the emissions under negative engine power are different from those under positive engine power. Hence, the deceleration mode is split into deceleration motoring mode and deceleration non-motoring mode based on engine power.

Inter-bus variability analysis indicates that bus 372 has the largest 3rd Quartile value for engine power among 15 buses in deceleration mode, consistent with the finding in idle mode. At the same time, inter-bus variability analysis results show that bus 379 has the largest median and the second largest mean for NO_x emissions, bus 372 has the largest median and the second largest mean for CO emissions, while bus 364 has the largest median and mean for HC emissions. But it is difficult to conclude that these buses should be classified as high emitters or that there are any special modes that should be modeled separately as high-emitting modes.

Some high HC emissions events are noted in deceleration mode. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, these operating parameters could not be linked to these high emissions occurrences. Additional causal variables may be in play that are not included in the data available for analysis.

Based on definitions of "acceleration < -1 mph/s", about 1% of data points exhibit somewhat unusually high engine power ($\geq$ 50 bhp) in deceleration mode. Analysis shows that higher engine power always happened with higher vehicle speed and higher engine speed. These higher-power data points likely reflect the variability in real world power demand (perhaps associated with operations on grade, which could not be identified in the database). All of these data were retained in the model to avoid potentially biasing the results.

In summary, the deceleration non-motoring mode emission rate for NO_x is set as 0.04578 g/s, CO as 0.00686 g/s, and HC as 0.00131 g/s. The deceleration motoring emission rate for NO_x is set as 0.00978 g/s, CO as 0.00529 g/s, and HC as 0.00111 g/s. Emission rate estimation for the deceleration motoring mode is significantly lower than the deceleration non-motoring mode for all three pollutants, especially for NO_x.

CHAPTER 10

10. ACCELERATION MODE DEVELOPMENT

After developing the idle mode definition and emission rate in Chapter 8 and deceleration mode definitions and emission rates in Chapter 9, the next task is to divide the rest of the data into acceleration and cruise mode. This chapter examines the definition of acceleration activity and emission rates for acceleration activity.

10.1 Critical Value for Acceleration in Acceleration Mode

The first task related to analysis of emission rates in the acceleration mode is identifying a critical value for acceleration. Two values were tested: 1 mph/s and 2 mph/s. Since the critical value of "acceleration > 1 mph/s" will include all data under the critical value of "acceleration > 2 mph/s", comparison of data falling between these two potential cut points is conducted first. Once selected, the chosen critical value will be used to divide the data into acceleration mode and cruise mode. Thus "acceleration > 0 mph/s and acceleration ≤ 1 mph/s" will be another option. Similarly to analysis for deceleration mode, these three options will be:

- Option 1: acceleration > 2 mph/s
- Option 2: acceleration > 1 mph/s and acceleration ≤ 2 mph/s
- Option 3: acceleration > 0 mph/s and acceleration ≤ 1 mph/s

Figure 10-1 illustrates engine power distribution for these three options. Figures 10-2 to 10-4 compare engine power vs. emission rate for three pollutants for three options. Tables 10-1 and 10-2 provide the distribution for these three options in two ways: by number and percentage.

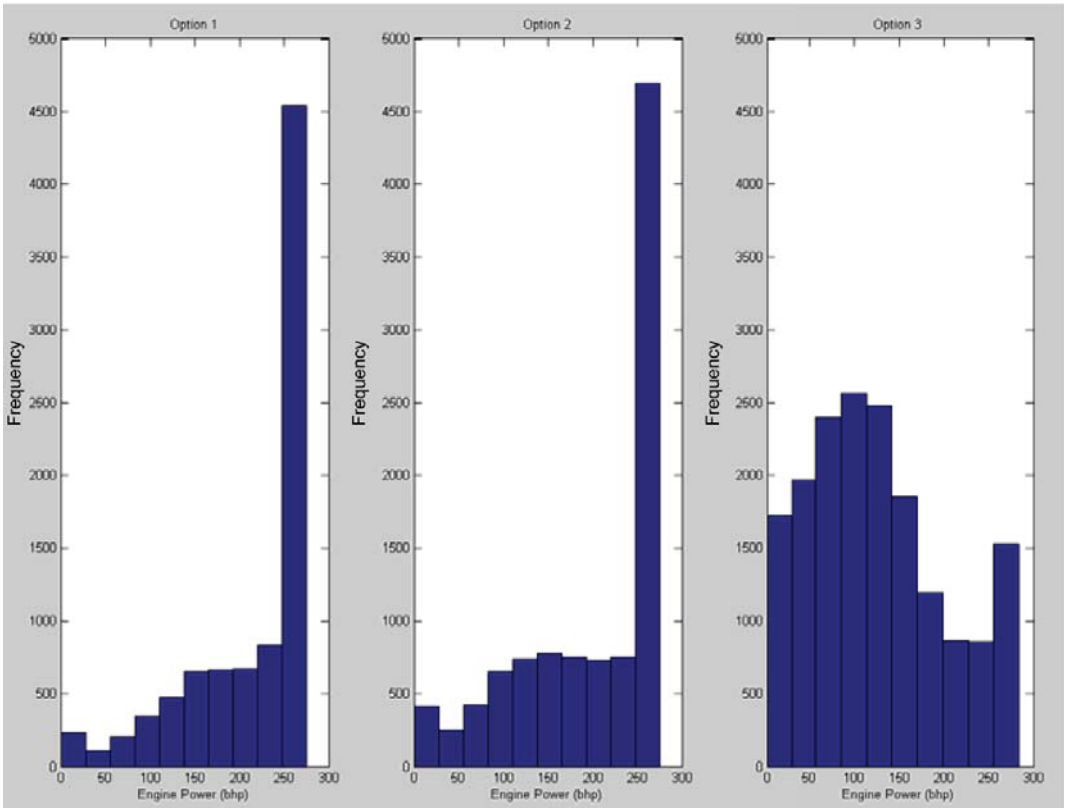


Figure 10-1 Engine Power Distribution for Three Options

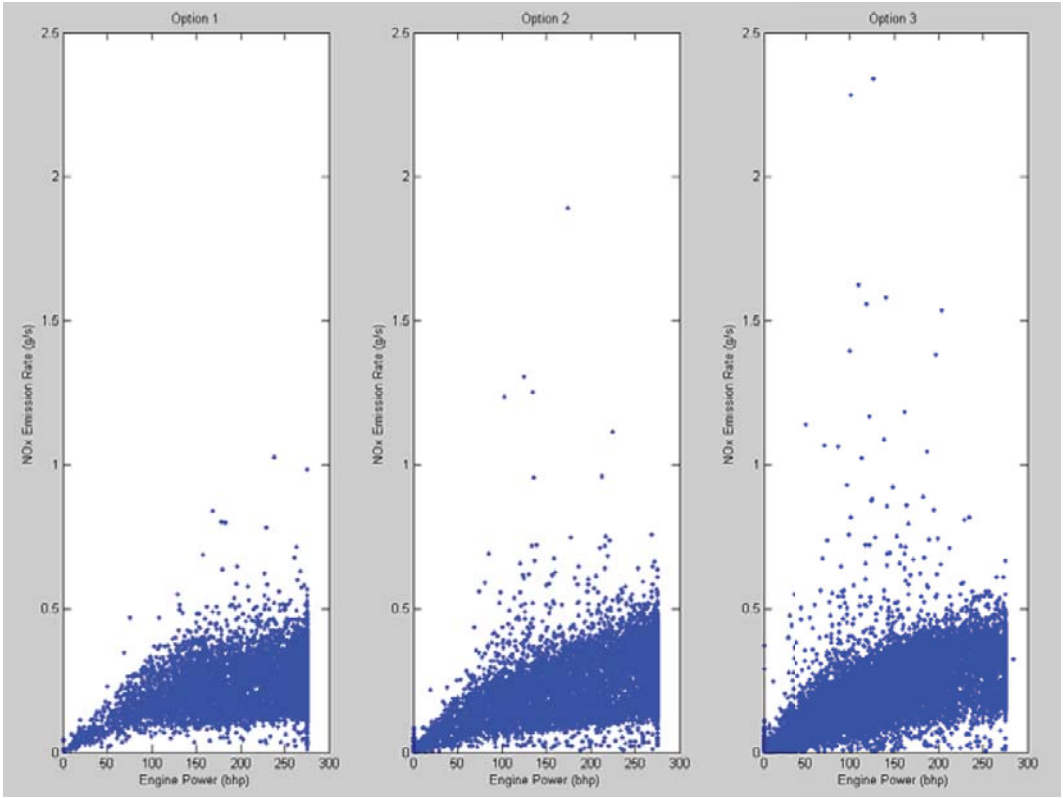


Figure 10-2 Engine Power vs. NO_x Emission Rate (g/s) for Three Options

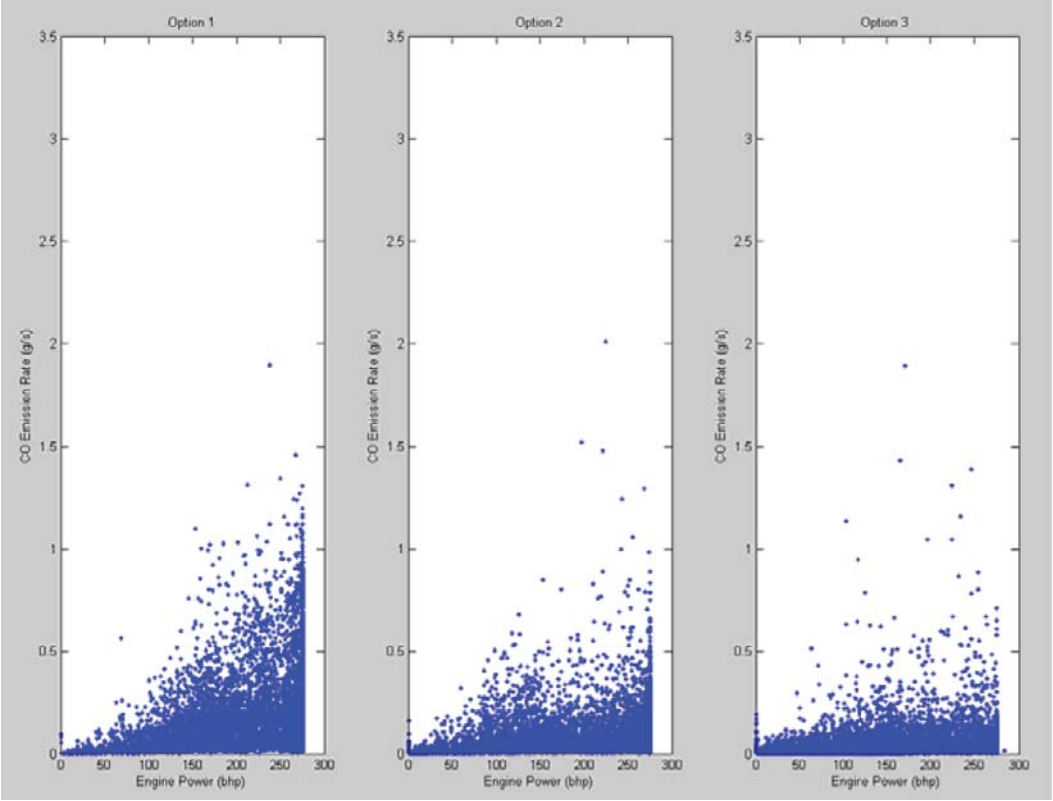


Figure 10-3 Engine Power vs. CO Emission Rate (g/s) for Three Options

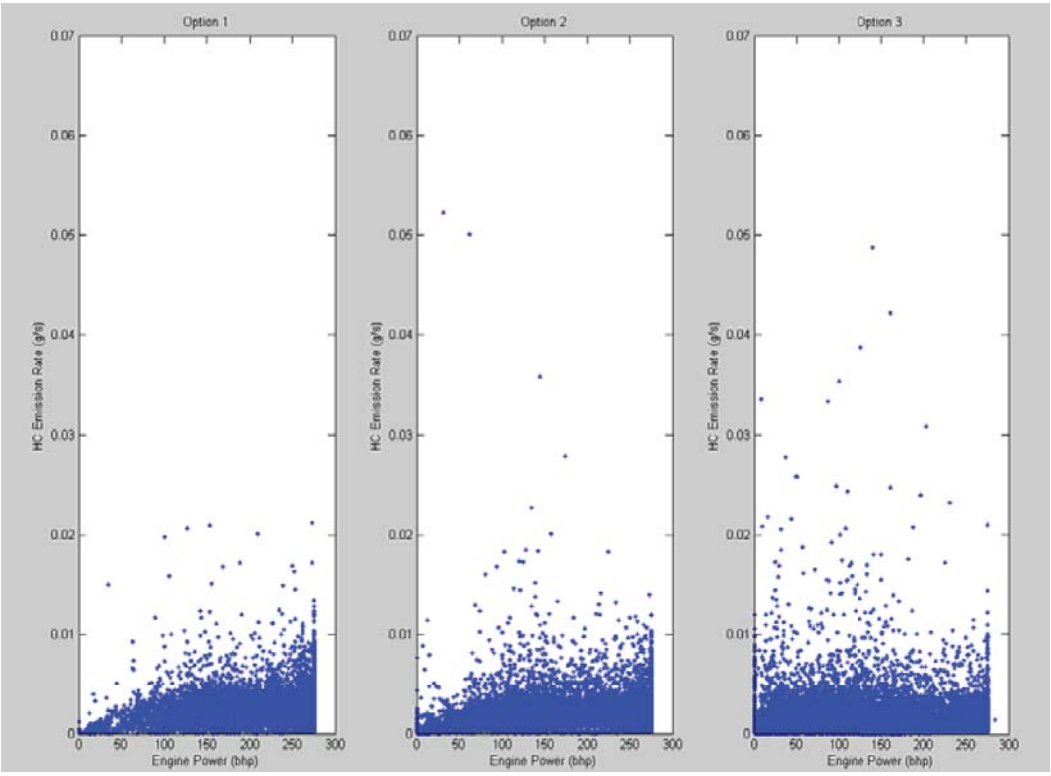


Figure 10-4 Engine Power vs. HC Emission Rate (g/s) for Three Options

Table 10-1 Engine Power Distribution for Three Options for Three Pollutants

Acceleration	Pollutants	Engine Power (brake horsepower (bhp))					
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	Total
Option 1	NO_x	322	446	852	1229	5870	8719
	CO	319	444	851	1228	5870	8712
	HC	318	440	833	1203	5649	8443
Option 2	NO_x	613	865	1358	1324	6015	10175
	CO	606	858	1355	1321	6012	10152
	HC	605	843	1328	1287	5824	9887
Option 3	NO_x	3208	4130	4378	2490	3205	17411
	CO	3190	4105	4362	2487	3185	17329
	HC	3104	3972	4195	2408	3131	16810

Table 10-2 Percentage of Engine Power Distribution for Three Options for Three Pollutants

Acceleration	Pollutants	Engine Power (brake horsepower (bhp))					
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	Total
Option 1	NO_x	3.7%	5.1%	9.8%	14.1%	67.3%	100.0%
	CO	3.7%	5.1%	9.8%	14.1%	67.4%	100.0%
	HC	3.8%	5.2%	9.9%	14.2%	66.9%	100.0%
Option 2	NO_x	6.0%	8.5%	13.3%	13.0%	59.1%	100.0%
	CO	6.0%	8.5%	13.3%	13.0%	59.2%	100.0%
	HC	6.1%	8.5%	13.4%	13.0%	58.9%	100.0%
Option 3	NO_x	18.4%	23.7%	25.1%	14.3%	18.4%	100.0%
	CO	18.4%	23.7%	25.2%	14.4%	18.4%	100.0%
	HC	18.5%	23.6%	25.0%	14.3%	18.6%	100.0%

If the critical value is set as 1 mph/s for acceleration mode, data falling into option 1 and option 2 will be classified as acceleration mode while data falling into option 3 will be classified as cruise mode. If the critical value is set as 2 mph/s for acceleration mode, data falling into option 1 will be classified as acceleration mode while data falling into option 2 and option 3 will be classified as cruise mode. There is little difference in the engine power distributions noted for data falling into option 1 and option 2 while the power distribution for option 3 is obviously different from option 1 and option 2 in the above figures and tables. Table 10-1 and 10-2 show that the engine power is more concentrated in higher engine power (≥200 bhp) for data in acceleration mode. Tables 10-1 and 10-2 better reflect the power demand of the vehicle in real world in acceleration mode. Hence, the critical value is set as 1 mph/s for acceleration mode.

After defining “acceleration > 1 mph/s” as acceleration mode, cruise mode data will consist of all of the remaining data in the database (i.e., data not previously classified into idle, deceleration, and now acceleration). Unlike idle and deceleration mode, there is a general relationship between engine power and emission rate for acceleration mode and cruise mode. Even though the engine power distribution for acceleration mode is different from that of cruise mode (Table 10-3), these two modes share a relationship between engine power and emission rate (Figure 10-5), although there are potentially some significant differences noted in the HC chart.

Table 10-3 Engine Power Distribution for Acceleration Mode and Cruise Mode

	Pollutants	Engine Power Distribution					
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	All
Acceleration mode							
Number	NO_x	935	1311	2210	2553	11885	18894
	CO	925	1302	2206	2549	11882	18864
	HC	923	1283	2161	2490	11473	18330
Percentage	NO_x	4.95%	6.94%	11.70%	13.51%	62.90%	100.00%
	CO	4.90%	6.90%	11.69%	13.51%	62.99%	100.00%
	HC	5.04%	7.00%	11.79%	13.58%	62.59%	100.00%
Cruise mode							
Number	NO_x	15885	8988	7173	3536	3792	39374
	CO	15834	8940	7145	3529	3770	39218
	HC	15481	8600	6830	3394	3715	38020
Percentage	NO_x	40.34%	22.83%	18.22%	8.98%	9.63%	100.00%
	CO	40.37%	22.80%	18.22%	9.00%	9.61%	100.00%
	HC	40.72%	22.62%	17.96%	8.93%	9.77%	100.00%

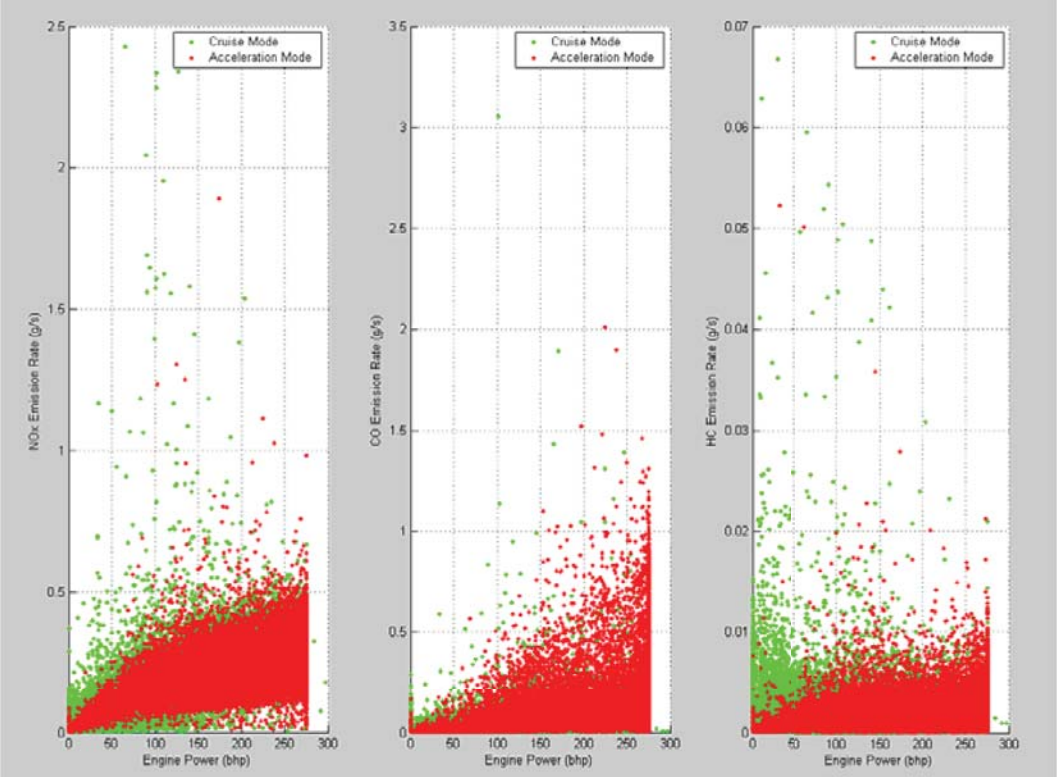


Figure 10-5 Engine Power vs. Emission Rate for Acceleration Mode and Cruise Mode

The relationships between emission rate and power for acceleration mode data will be explored in this chapter, while the relationships between emission rate and power for cruise mode data will be explored in the next chapter.

10.2 Analysis of Acceleration Mode Data

10.2.1 Emission Rate Distribution by Bus in Acceleration Mode

After defining vehicle activity data with "acceleration >1 mph/s" as acceleration mode, emission rate histograms for each of the three pollutants for acceleration operations are presented in Figure 10-6. Figure 10-6 shows significant skewness for all three pollutants for acceleration mode. There are also a small number of some very high HC emissions events noted in acceleration mode. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, no operating parameters appeared to be correlated with the high emissions events.

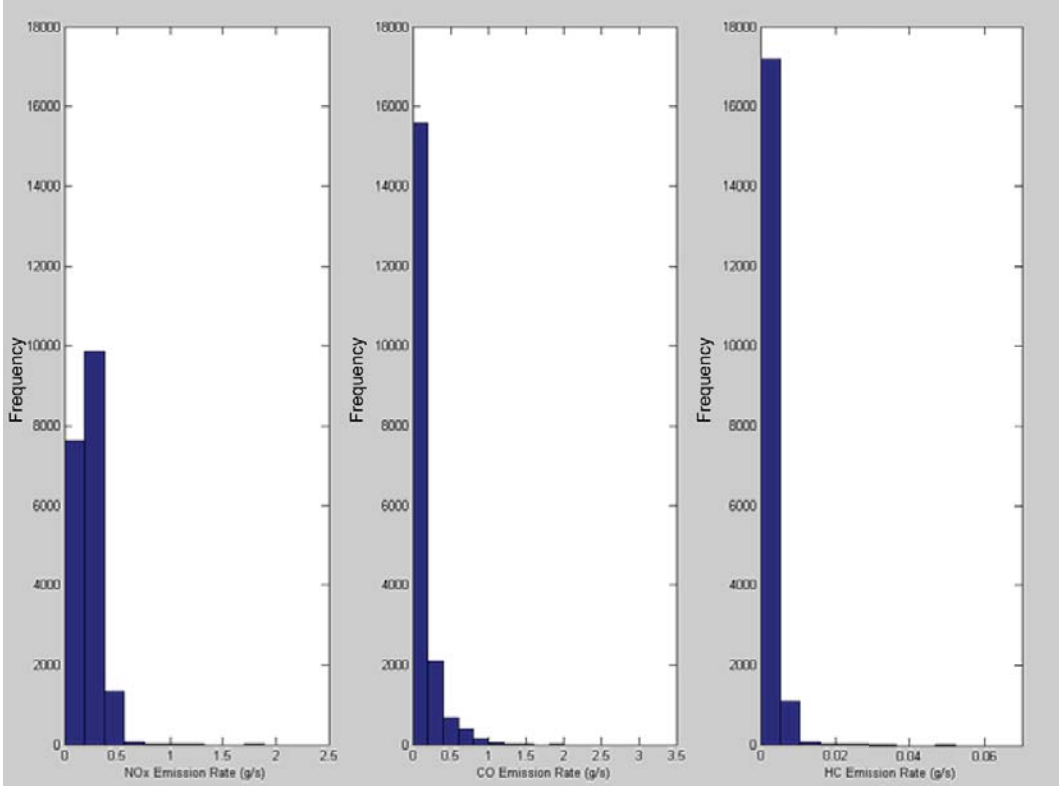


Figure 10-6 Histograms of Three Pollutants for Acceleration Mode

Inter-bus response variability for acceleration mode operations is illustrated in Figures 10-7 to 10-9 using median and mean of NO_x, CO, and HC emission rates. Table 10-4 presents the same information in tabular form. The difference between median and mean is also an indicator of skewness.

Table 10-4 Median and Mean of Three Pollutants in Acceleration Mode by Bus

	NO_x		CO		HC	
Bus ID	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Bus 360	0.27729	0.25957	0.06527	0.09217	0.00159	0.00182
Bus 361	0.30170	0.28125	0.05177	0.08001	0.00184	0.00228
Bus 363	0.14459	0.14058	0.03836	0.09012	0.00022	0.00039
Bus 364	0.28948	0.26033	0.03501	0.05650	0.00306	0.00363
Bus 372	0.17834	0.18627	0.02980	0.03475	0.00250	0.00279
Bus 375	0.31092	0.28991	0.05929	0.08619	0.00143	0.00176
Bus 377	0.17827	0.17335	0.04755	0.09612	0.00104	0.00112
Bus 379	0.17788	0.20883	0.08430	0.10346	0.00222	0.00276
Bus 380	0.26410	0.26620	0.08238	0.19149	0.00210	0.00253
Bus 381	0.18011	0.19806	0.07856	0.12646	0.00095	0.00106

	NO_x		CO		HC	
Bus ID	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Bus 382	0.28966	0.29152	0.09234	0.18179	0.00263	0.00272
Bus 383	0.24419	0.26739	0.05355	0.13112	0.00308	0.00368
Bus 384	0.18775	0.22139	0.07111	0.17389	0.00401	0.00429
Bus 385	0.17783	0.21706	0.05141	0.07893	0.00361	0.00384
Bus 386	0.22674	0.24673	0.10412	0.23806	0.00272	0.00282

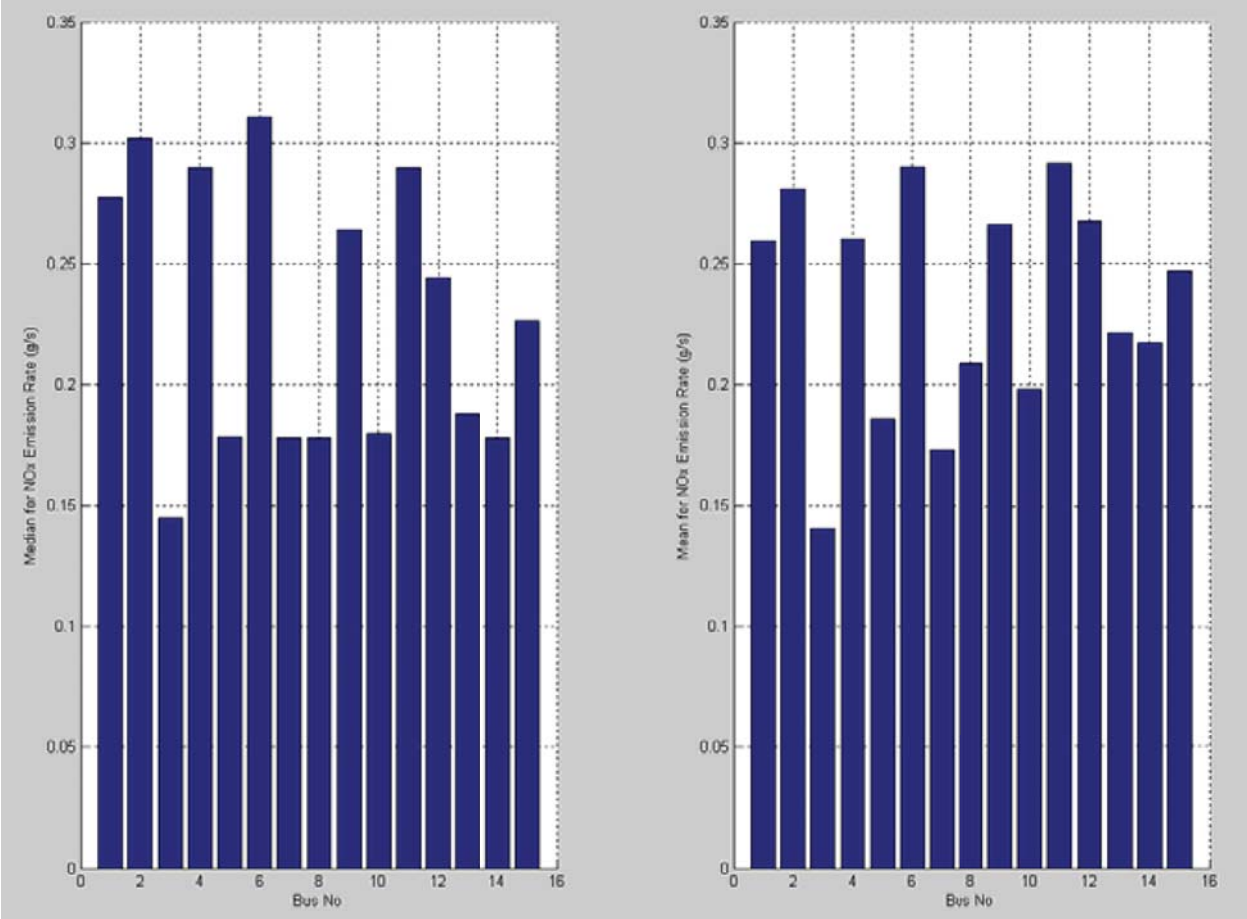


Figure 10-7 Median and Mean of NO_x Emission Rates in Acceleration Mode by Bus

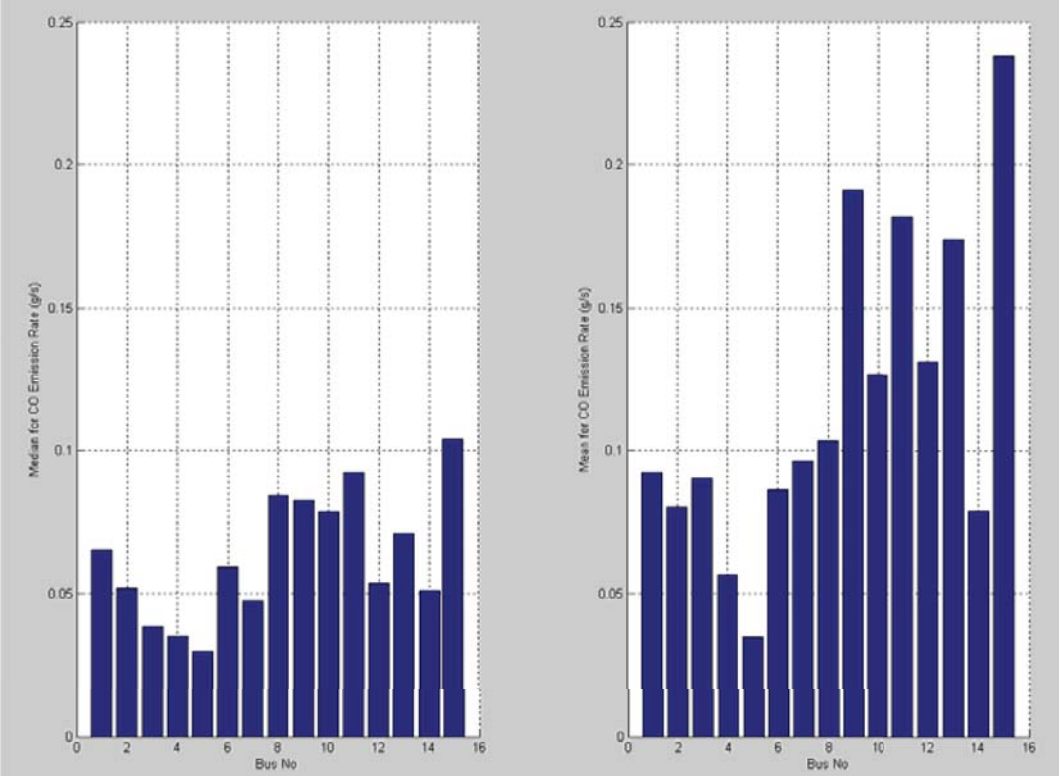


Figure 10-8 Median and Mean of CO Emission Rates in Acceleration Mode by Bus

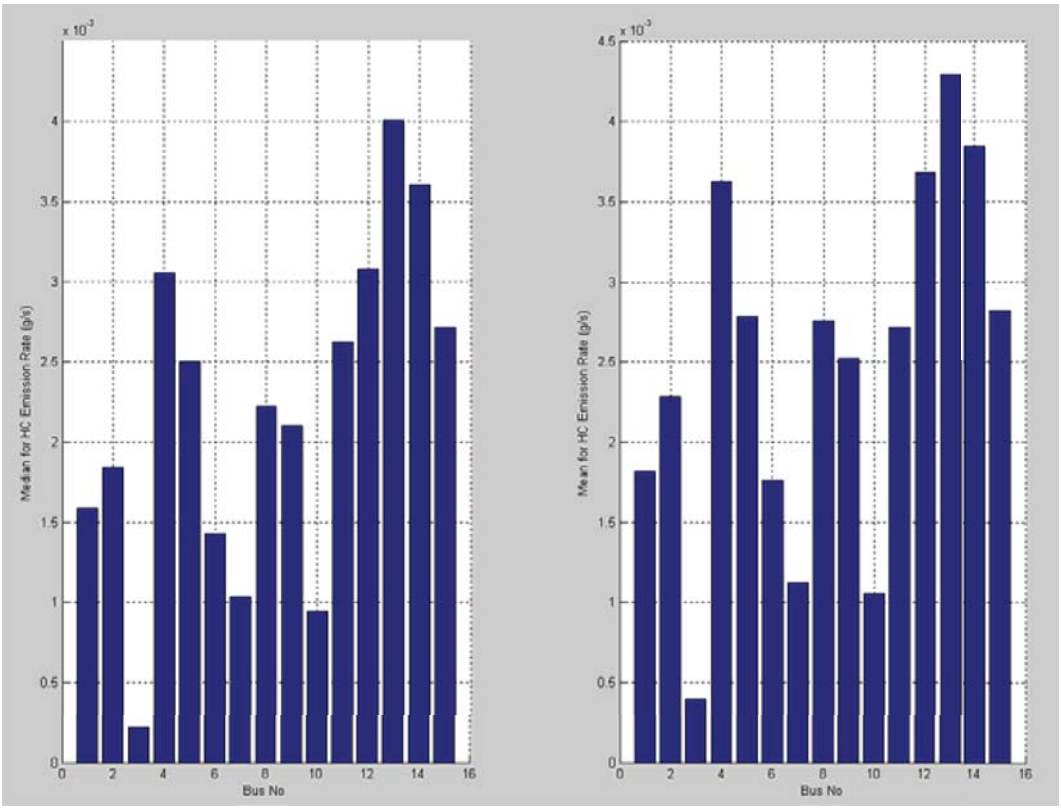


Figure 10-9 Median and Mean of HC Emission Rates in Acceleration Mode by Bus

Figures 10-7 to 10-9 and Table 10-4 illustrate that NO_x emissions are more consistent than CO and HC emissions. Across the 15 buses, Bus 386 has the largest median and mean for CO emissions, while Bus 384 has the largest median and mean for HC emissions. The above figures and table demonstrate that although variability exists across buses, it is difficult to conclude that there are any true "high emitters." That is, the emissions from these buses are not consistently more than one or two standard deviations from the mean under normal operating conditions. Meanwhile, Bus 363 has the smallest mean and median HC emissions compared to the other 14 buses.

10.2.2 Engine Power Distribution by Bus in Acceleration Mode

Engine power distribution in acceleration mode by bus is shown in Figure 10-10 and Table 10-5. When the bus is accelerating, the engine will be required to produce more power. Figure 10-10 and Table 10-5 reflect this characteristic of acceleration mode. The distribution of engine power in acceleration mode is significantly different from deceleration mode and idle mode. Bus 372 has the largest minimum engine power in acceleration mode, consistent with the finding for idle mode and deceleration mode. The maximum power values for each bus match well with the manufacturer's engine power rating. Although variability for engine power distribution exists across buses, it is difficult to conclude that such variability is affected by individual buses, bus routes, or other factors. The relationship between power and emissions appears consistent across the buses for acceleration mode.

Table 10-5 Engine Power Distribution in Acceleration Mode by Bus

Bus ID	Number	Min	1st Quartile	Median	3rd Quartile	Max	Mean
Bus 360	1507	0	162.96	255.57	275.05	275.59	212.04
Bus 361	545	7.16	131.96	199.58	261.51	275.54	184.46
Bus 363	1287	0	111.52	200.39	267.06	275.59	180.03
Bus 364	931	0	142.82	228.25	270.01	275.56	197.27
Bus 372	728	34.42	145.57	213.51	264.70	275.56	199.81
Bus 375	1599	0	140.92	259.45	275.13	275.57	205.56
Bus 377	1751	3.35	166.25	256.89	275.08	275.60	212.09
Bus 379	1427	0	204.15	264.54	275.18	275.58	233.71
Bus 380	1823	0	202.69	262.11	275.15	275.54	228.55
Bus 381	1362	0	139.86	220.00	272.21	275.60	199.20
Bus 382	691	0	173.36	250.90	275.05	275.58	218.82
Bus 383	1043	0	161.16	250.37	275.08	275.59	213.70
Bus 384	1292	0	144.10	213.87	269.50	275.60	198.80
Bus 385	1377	0	143.51	226.37	274.99	275.55	201.67
Bus 386	1532	13.81	164.27	244.80	275.06	275.60	215.95

Engine power distribution also shows that about 0.19% (36/18895) of data points show zero load in acceleration mode. For the 36 data points exhibiting zero indicated engine load, about 92% (33/36) occurred on roads reported to have zero or negative grade. Due to the inaccuracy of road grade values, it was not possible to simulate the engine power in this research. However, in the real world, linear acceleration with zero load can happen on downhill stretches. Application of load based emission rates to predicate engine load will be able to take grade into account in the overall modeling framework. Because only 36 data points with zero load were included in the acceleration data, it was unnecessary to develop a sub-model for them. Meanwhile, such zero loads in acceleration mode do reflect the variability of acceleration data in the real world.

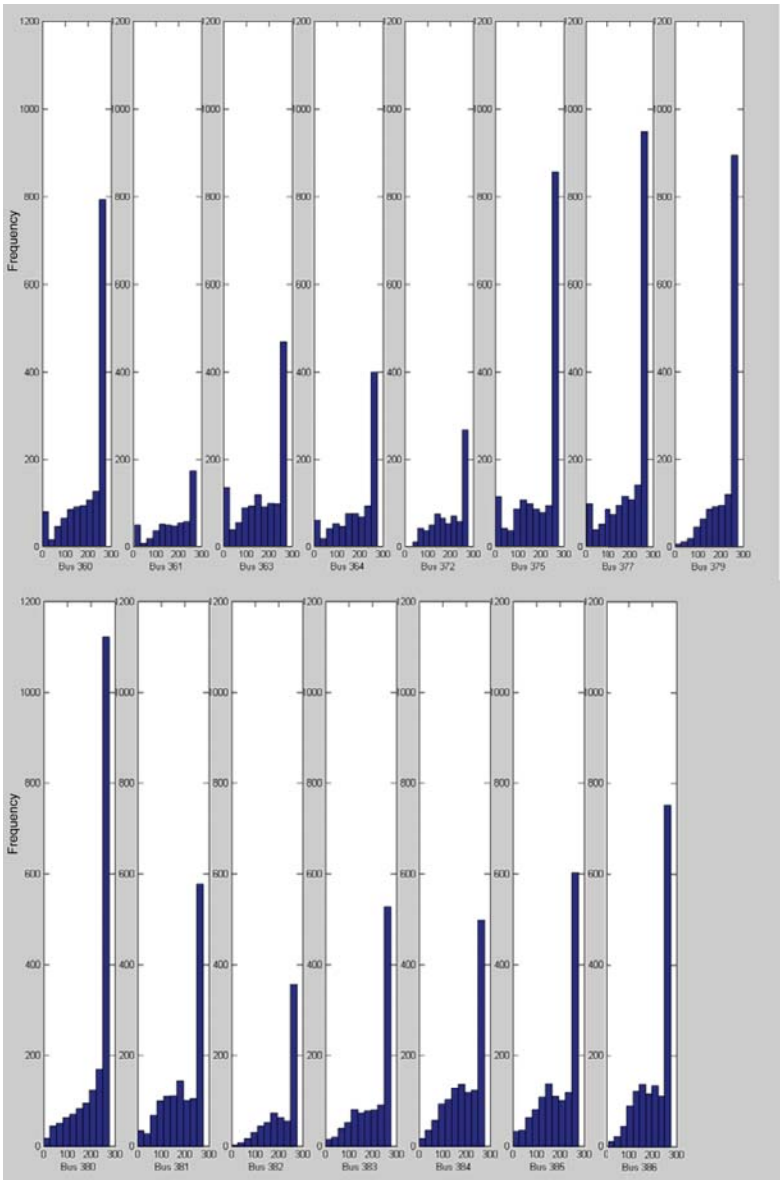


Figure 10-10 Histograms of Engine Power in Acceleration Mode by Bus

10.3 Model Development and Refinement

10.3.1 HTBR Tree Model Development

The potential explanatory variables included in the emission rate model development effort include:

- Vehicle characteristics: model year, odometer reading, bus ID (14 dummy variables);
- Roadway characteristics: dummy variable for road grade;
- Onroad load parameters: engine power (bhp), vehicle speed (mph), acceleration (mph/s);
- Engine operating parameters: engine oil temperature (deg F), engine oil pressure (kPa), engine coolant temperature (deg F), barometric pressure reported from ECM (kPa);
- Environmental conditions: ambient temperature (deg C), ambient pressure (mbar), ambient relative humidity (%).

The HTBR technique is used first to identify potentially significant explanatory variables; this analysis provides the starting point for conceptual model development. The HTBR model is used to guide the development of an OLS regression model, and not a model in its own right. HTBR can be used as a data reduction tool and for identifying potential interactions among the variables. Then OLS regression is used with the identified variables to estimate a preliminary "final" model.

These 27 variables were first offered to the tree model. To arrive at the "best" model, various regression tree models were created. The initial model was created by allowing the tree to grow unconstrained for the first cut. Once an initial model was created, the supervised technique in S-PLUS was used to simplify the model by removing the lower branches of the tree that explained the least deviance. For application purposes, the resulting tree was examined to ensure that the model's predictive ability was not compromised by allowing the overall amount of deviance to increase significantly.

The 27 variables include continuous, categorical, and dummy variables. Dummy variables for buses could be used to indicate the variability of buses. Like the analysis in Chapter 6, these 15 buses could be treated as a single group for purposes of analysis and model development. HTBR technique can examine the potential additional influence of road grade (i.e., above and beyond the contribution to power demand) using a dummy variable to represent a grade

category (the final model does not include this dummy variable due to the inaccuracy of road grade values). Analysis results in Chapter 6 indicate that all environmental characteristics, like temperature, humidity and barometric pressure, are moderately correlated with each other. On the other hand, engine operating parameters, like engine oil pressure, engine oil temperature, engine coolant temperature, and barometric pressure reported from ECM, are highly or moderately related to on-road operating parameters, like engine power, vehicle speed, and acceleration. The modeler should be aware of such correlations among explanatory variables.

Although evidence in the literature suggests that a logarithmic transformation is most suitable for modeling motor vehicle emissions (Washington 1994; Ramamurthy et al. 1998; Fomunung 2000; Frey et al. 2002), this transformation needs to be verified through the Box-Cox procedure. The Box-Cox function in MATLAB™ can automatically identify a transformation from the family of power transformations on emission data, ranging from -1.0 to 1.0. The lambdas chosen by Box-Cox procedure for acceleration mode are 0.683 for NO_x, 0.094438 for CO, 0.31919 for HC. The Box-Cox procedure is used only to provide a guide for selecting a transformation, so overly precise results are not needed (Neter et al. 1996). It is often reasonable to use a nearby lambda value that is easier to understand for the power transformation. Although the lambdas chosen by the Box-Cox procedure are different for acceleration and cruise mode, the nearby lambda values are same for these two modes. In summary, the lambda values used for transformations are ½ for NO_x, 0 for CO (indicating a log transformation), and ¼ for HC for acceleration mode. Figures 10-11 to 10-13 present histogram, boxplot, and probability plots of truncated emission rates in acceleration mode for NO_x, CO, and HC, while Figures 10-14 to 10-16 present the same plots for truncated transformed emission rates for NO_x, CO and HC, where a great improvement is noted.

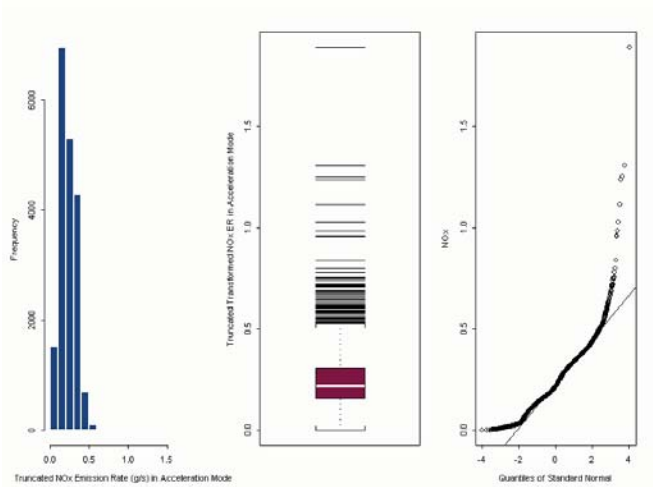


Figure 10-11 Histogram, Boxplot, and Probability Plot of Truncated NO_x Emission Rate in Acceleration Mode

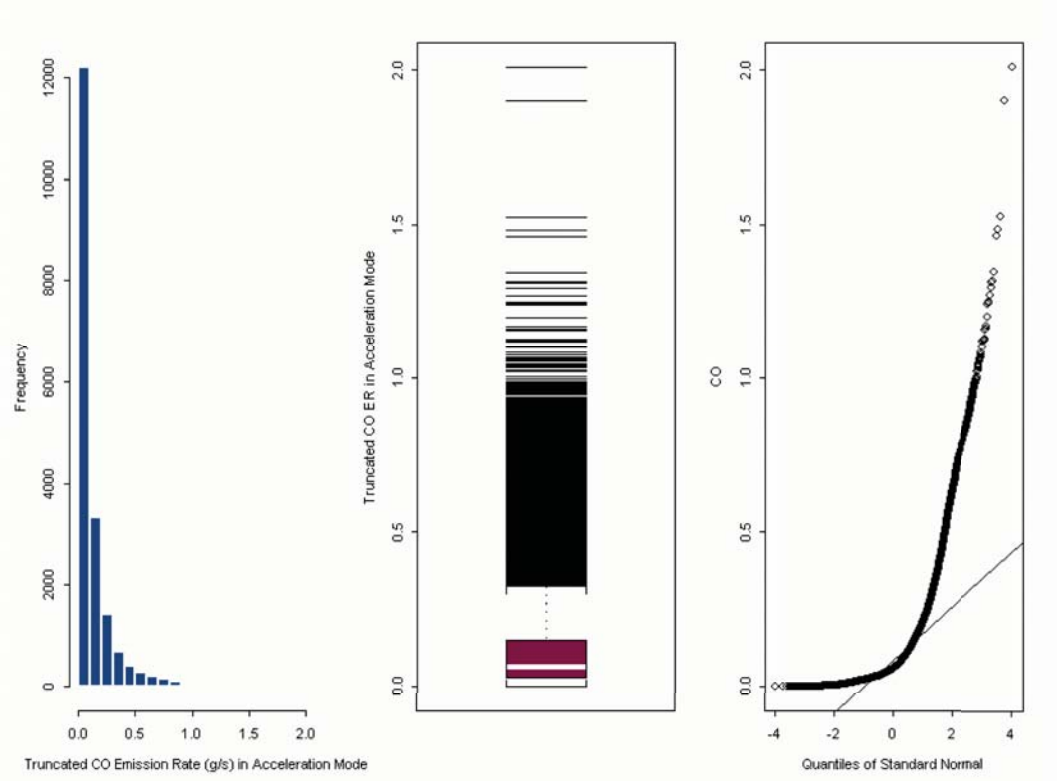


Figure 10-12 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Acceleration Mode

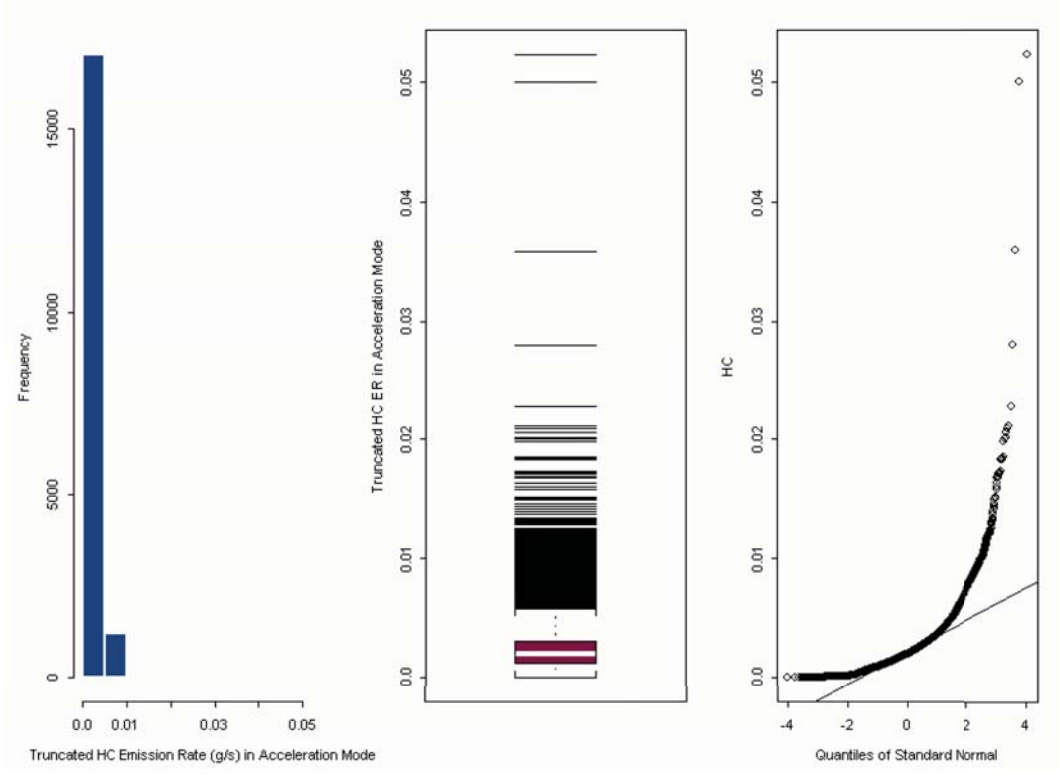


Figure 10-13 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Acceleration Mode

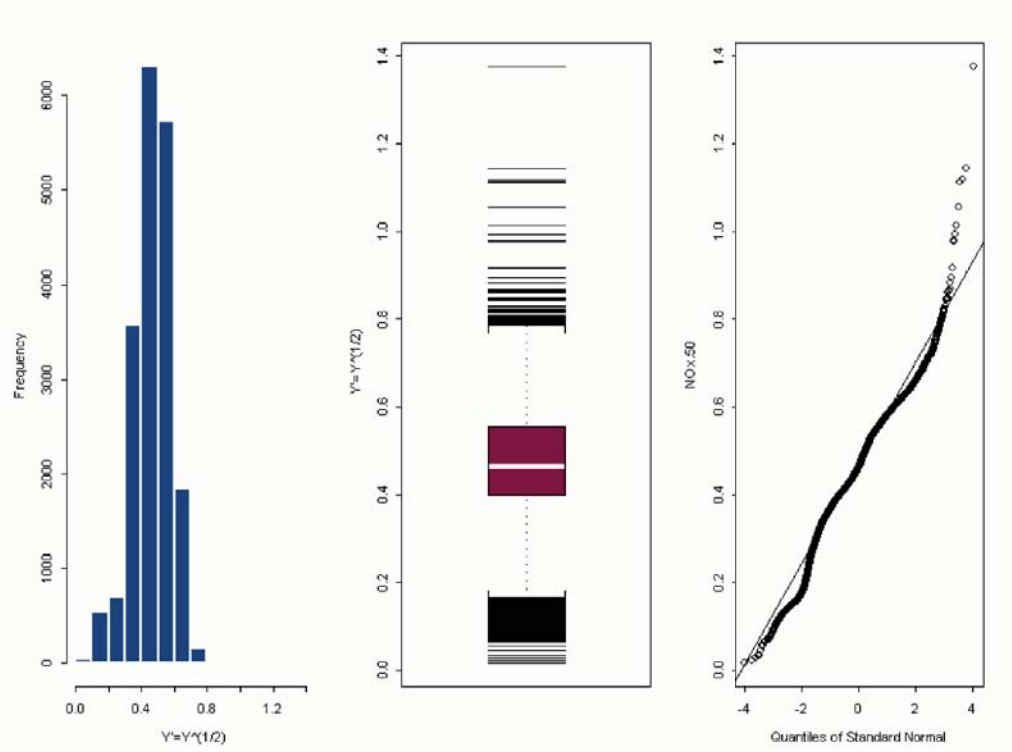


Figure 10-14 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO_x Emission Rate in Acceleration Mode

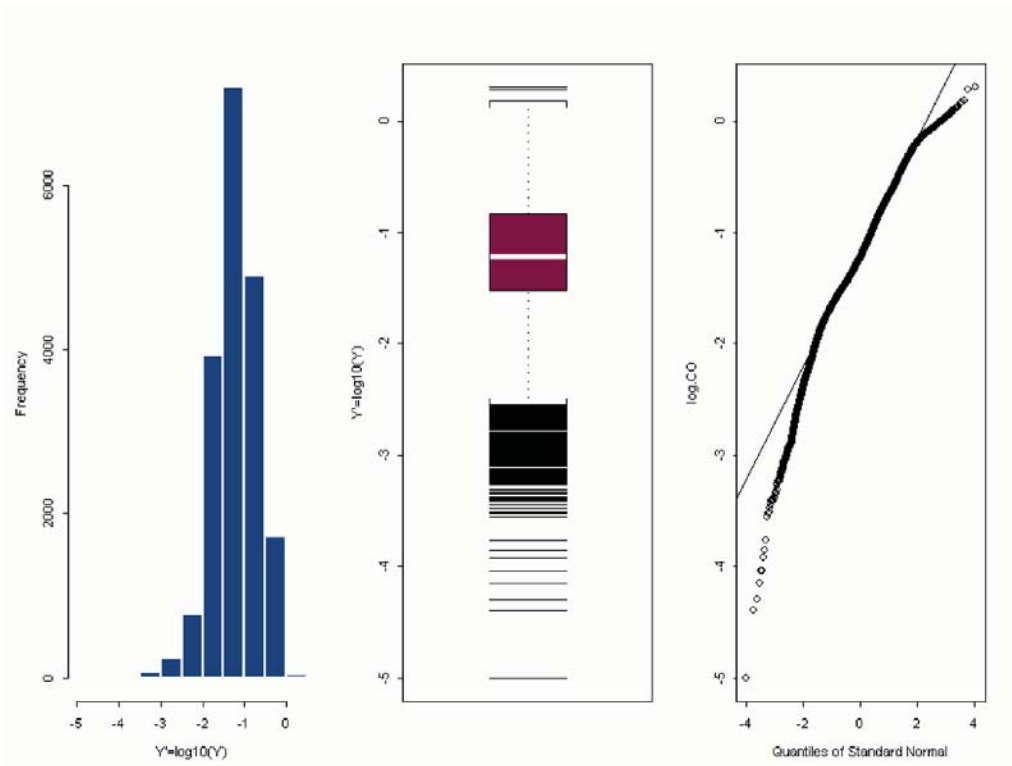


Figure 10-15 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate in Acceleration Mode

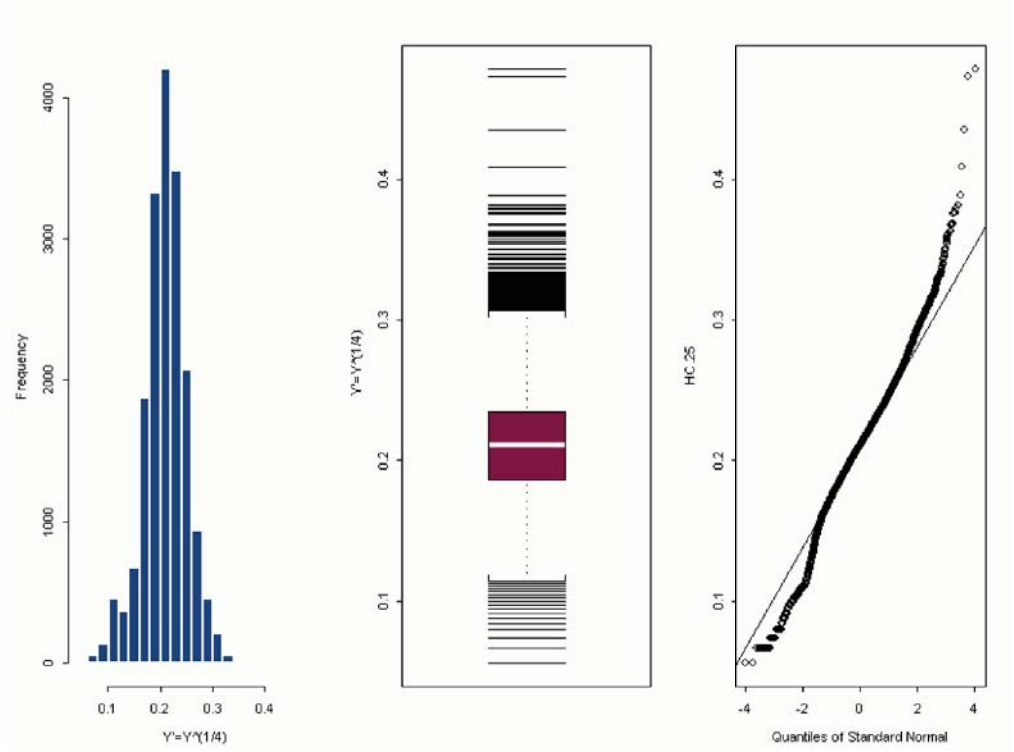


Figure 10-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate in Acceleration Mode

10.3.1.1 NO_x HTBR Tree Model Development

Figure 10-17 illustrates the initial tree model used for truncated transformed NO_x emission rate in acceleration mode. Results for the initial model are given in Table 10-6. The tree grew into a complex model, with a considerable number of branches and 36 terminal nodes. Figure 10-18 illustrates the amount of deviation explained corresponding to the number of terminal nodes.

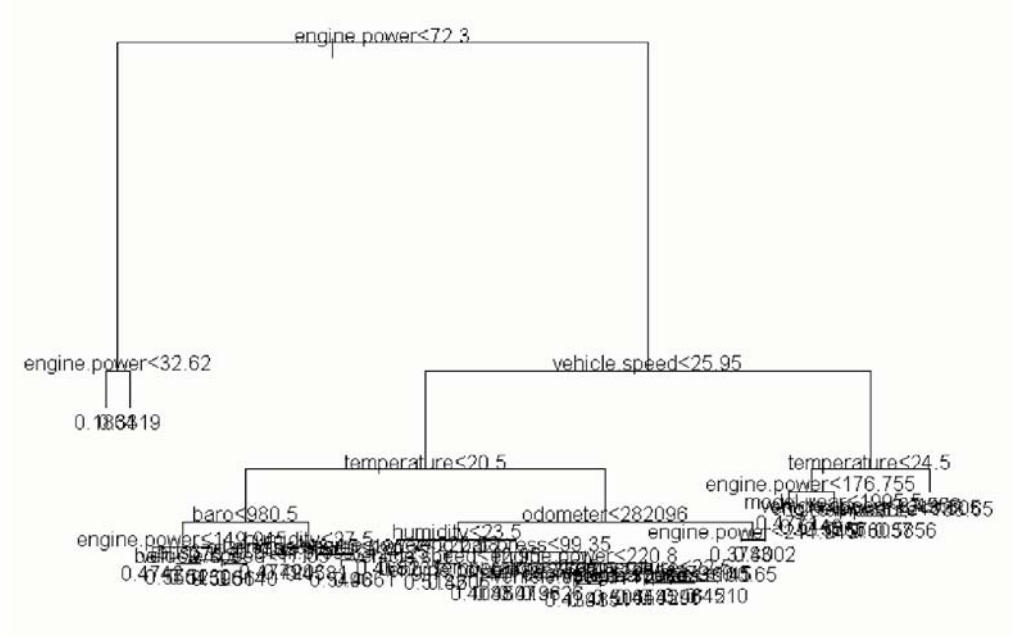


Figure 10-17 Original Untrimmed Regression Tree Model for Truncated Transformed NO_x Emission Rate in Acceleration Mode

Table 10-6 Original Untrimmed Regression Tree Results for Truncated Transformed NO_x Emission Rate in Acceleration Mode

```
Regression tree:
tree(formula = NOx.50 ~ model.year + odometer + temperature + baro + humidity + ve-
hicle.speed + oil.temperture + oil.press + cool.temperature + eng.bar.press + engine.
power + acceleration + bus360 + bus361 + bus363 + bus364 + bus372 + bus375 + bus377
+ bus379 + bus380 + bus381 + bus382 + bus383 + bus384 + bus385 + dummy.grade, data =
busdata10242006.1.3,
na.action = na.exclude, mincut = 400, minsize = 800, mindev = 0.01)
Variables actually used in tree construction:
 [1] “engine.power”   “vehicle.speed”  “temperature”    “baro”
 [5] “bus375”         “humidity”       “oil.press”      “odometer”
 [9] “eng.bar.press”  “bus379”         “model.year”     “oil.temperture”
Number of terminal nodes:  36
Residual mean deviance:  0.005538 = 104.4 / 18860
Distribution of residuals:
        Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
 -3.769e-001 -4.176e-002 -4.298e-003  3.661e-017  3.957e-002  8.965e-001
```

For model application purposes, it is desirable to select a final model specification that balances the model's ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Figure 10-18 indicates that reduction in deviation with addition of nodes after 4, although potentially statistically significant, is very small. A simplified tree model was derived which ends in 4 terminal nodes as compared to the 36 terminal nodes in the initial model. The residual mean deviation only increased from 104.4 to 151.2 and yielded a much more efficient model. Results are shown in Table 10-7 and Figure 10-19. Based on above analysis, an NO_x acceleration emission rate model will be developed based upon these results.

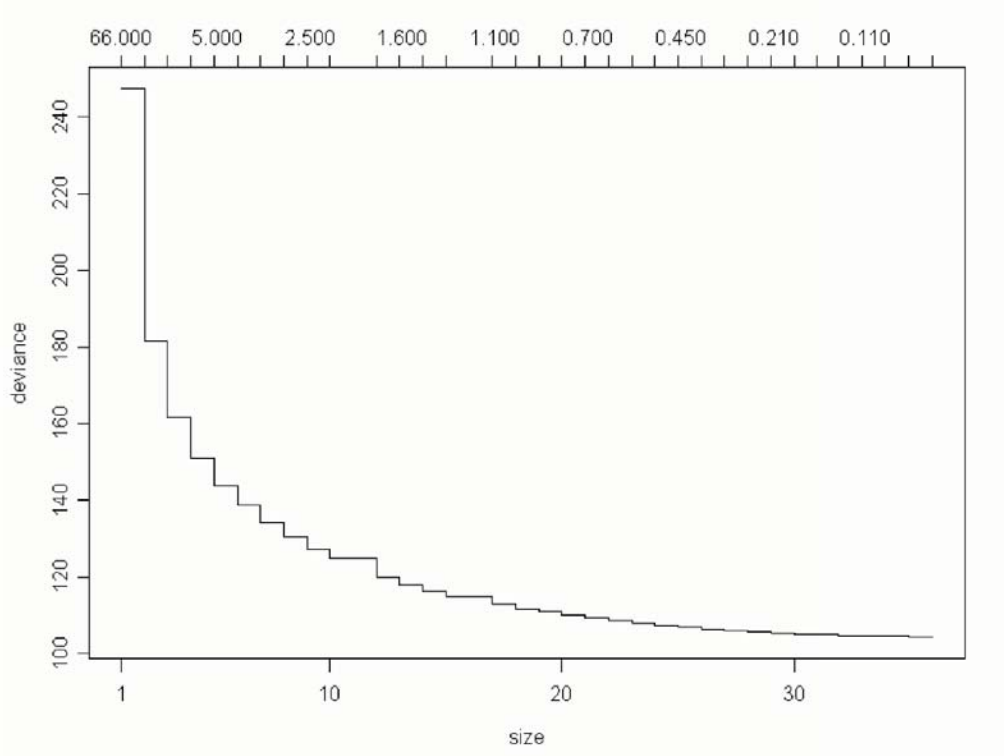


Figure 10-18 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed NO_x Emission Rate in Acceleration Mode

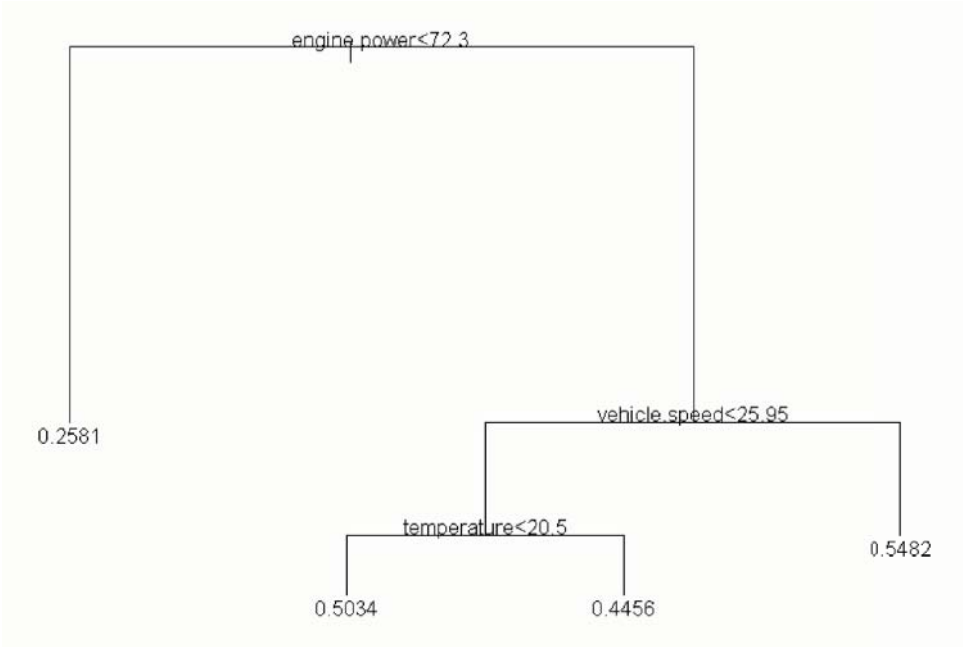


Figure 10-19 Trimmed Regression Tree Model for Truncated Transformed NO_x Emission Rate in Acceleration Mode

Table 10-7 Trimmed Regression Tree Results for Truncated Transformed NO_x Emission Rate in Acceleration Mode

```
Regression tree:
snip.tree(tree = tree(formula = NOx.50 ~ model.year + odometer + temperature +
       baro + humidity + vehicle.speed + oil.temperture + oil.press +
       cool.temperature + eng.bar.press + engine.power + acceleration +
       bus360 + bus361 + bus363 + bus364 + bus372 + bus375 + bus377 + bus379 +
       bus380 + bus381 + bus382 + bus383 + bus384 + bus385 + dummy.grade,
       data = busdata10242006.1.3, na.action = na.exclude, mincut = 400,
       minsize = 800, mindev = 0.01), nodes = c(13., 7., 12., 2.))
Variables actually used in tree construction:
[1] “engine.power”  “vehicle.speed” “temperature”
Number of terminal nodes:  4
Residual mean deviance:  0.008002 = 151.2 / 18890
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -4.265e-001 -5.813e-002 -7.517e-004  8.861e-016  5.810e-002  8.710e-001
node), split, n, deviance, yval
      * denotes terminal node

 1) root 18894 247.20 0.4669
   2) engine.power<72.3 1397  13.67 0.2581 *
   3) engine.power>72.3 17497 167.70 0.4836
     6) vehicle.speed<25.95 13777 121.40 0.4662
      12) temperature<20.5 4902  42.44 0.5034 *
      13) temperature>20.5 8875  68.45 0.4456 *
     7) vehicle.speed>25.95 3720  26.60 0.5482 *
```

This tree model suggests that engine power is the most important explanatory variable for NO_x emissions. This result is consistent with previous research results which verified the important effect of engine power on NO_x emissions (Ramamurthy et al. 1998; Clark et al. 2002; Barth et al. 2004). Analysis in the previous chapter also indicates that engine power is correlated with not only on-road load parameters such as vehicle speed, acceleration, and grade, but also engine operating parameters such as throttle position and engine oil pressure. On the other hand, engine power in this research is derived from engine speed, engine torque and percent engine load. Therefore engine power can correlate on-road modal activity with engine operating conditions to that extent. This fact strengthens the importance of introducing engine power into the conceptual model and the need to improve the ability to simulate engine power for regional inventory development.

HTBR results suggest that temperature may be an important predictive variable for NO_x emissions under certain conditions. Temperature effects may need to be integrated into new models in the form of a temperature correction factor. But adequate data are not yet available for this purpose. For the time being, temperature is removed from consideration in further linear regression model development, but the effect is probably significant and should be examined when more comprehensive emission rate data collected under a wider variety of temperature conditions are available for analysis.

10.3.1.2 CO HTBR Tree Model Development

Figure 10-20 illustrates the initial tree model used for truncated transformed CO emission rate in acceleration mode. Results from the initial model are given in Table 10-8. The tree grew into a complex model with a considerable number of branches and 33 terminal nodes. Figure 10-21 illustrates the amount of deviation explained corresponding to the number of terminal nodes.

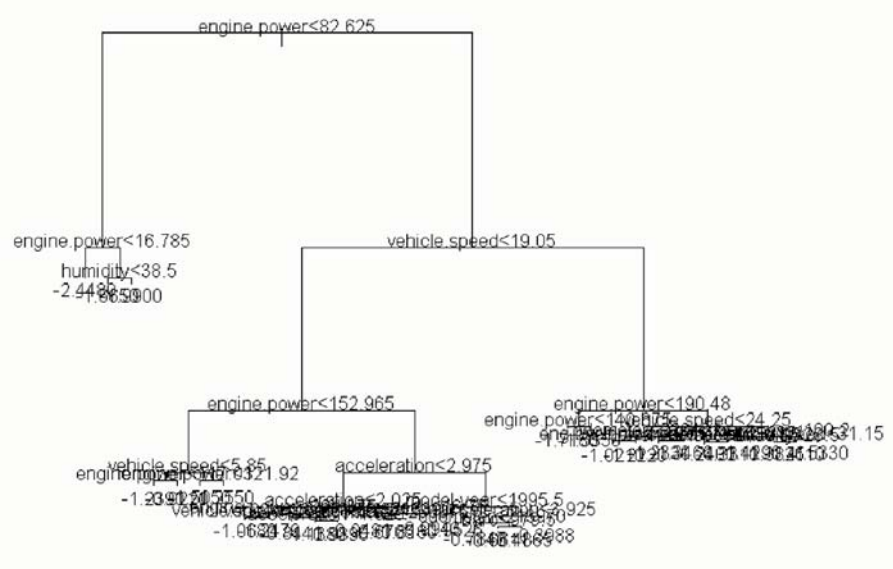


Figure 10-20 Original Untrimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode

Table 10-8 Original Untrimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Acceleration Mode

```
Regression tree:
tree(formula = log.CO ~ model.year + odometer + temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + bus360 + bus361 + bus363 +
       bus364 + bus372 + bus375 + bus377 + bus379 + bus380 + bus381 + bus382 +
       bus383 + bus384 + bus385 + dummy.grade, data = busdata10242006.1.3,
       na.action = na.exclude, mincut = 400, minsize = 800, mindev = 0.01)
Variables actually used in tree construction:
[1] “engine.power”  “humidity”      “vehicle.speed” “acceleration”
[5] “odometer”      “model.year”    “baro”          “eng.bar.press”
Number of terminal nodes:  33
Residual mean deviance:  0.1184 = 2229 / 18830
Distribution of residuals:
        Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
 -2.552e+000 -2.001e-001 -1.285e-002  3.025e-017  1.981e-001  1.653e+000
```

For model application purposes, it is desirable to select a final model specification that balances the model’s ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Figure 10-21 indicated that the reduction in deviation with addition of nodes after four, although potentially statistically significant, is very small. A simplified

tree model was derived which ends in four terminal nodes as compared to the 33 terminal nodes in the initial model. The residual mean deviation only increased from 2229 to 3093 and yielded a much cleaner model than the initial one. Results are shown in Table 10-9 and Figure 10-22. The CO acceleration emission rate model will be developed based upon these results.

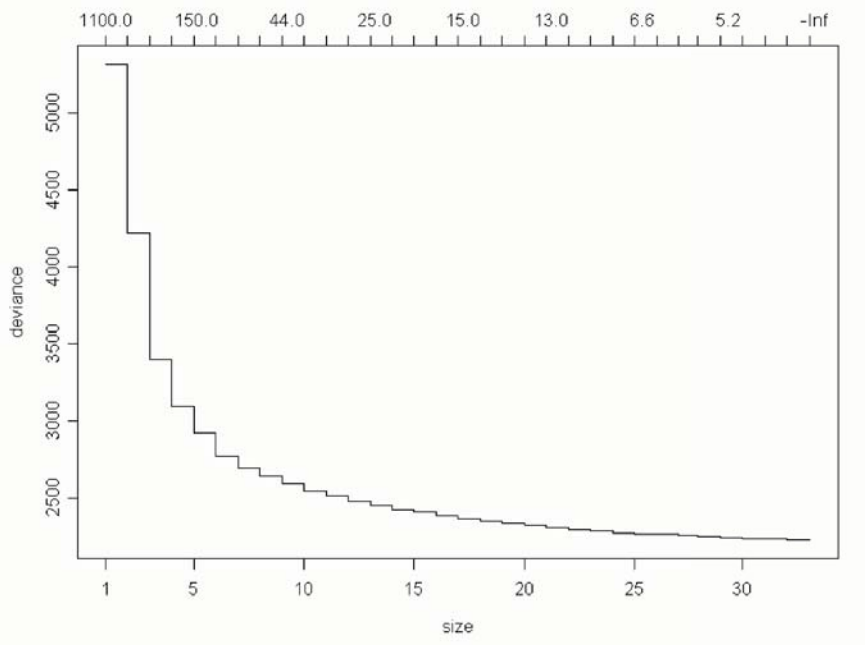


Figure 10-21 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed CO Emission Rate in Acceleration Mode

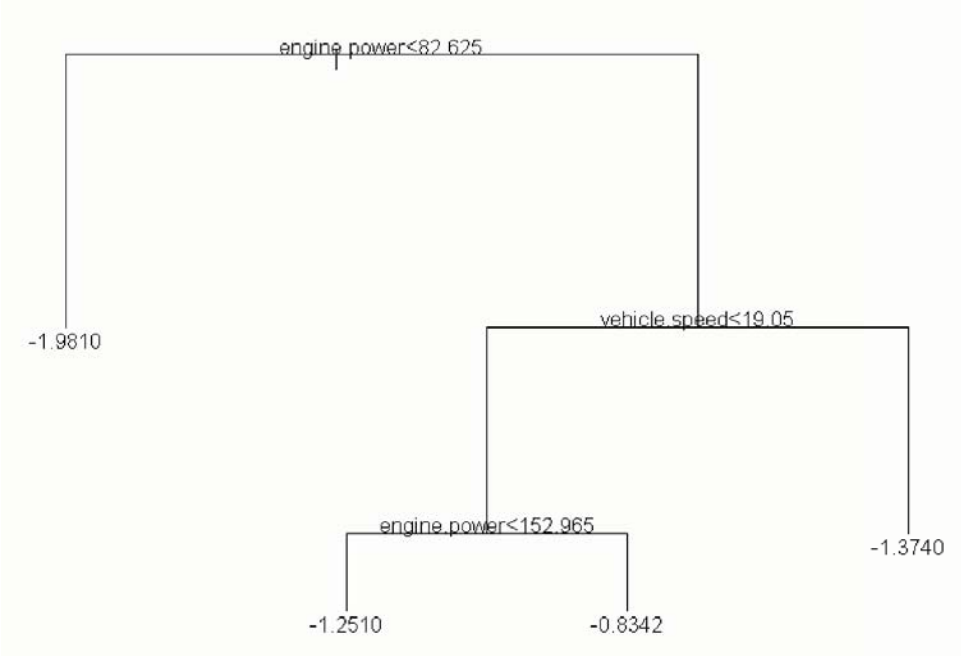


Figure 10-22 Trimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode

Table 10-9 Trimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Acceleration Mode

```
Regression tree:
snip.tree(tree = tree(formula = log.CO ~ model.year + odometer + temperature +
       baro + humidity + vehicle.speed + oil.temperture + oil.press +
       cool.temperature + eng.bar.press + engine.power + acceleration +
       bus360 + bus361 + bus363 + bus364 + bus372 + bus375 + bus377 + bus379 +
       bus380 + bus381 + bus382 + bus383 + bus384 + bus385 + dummy.grade,
       data = busdata10242006.1.3, na.action = na.exclude, mincut = 400,
       minsize = 800, mindev = 0.01), nodes = c(12., 7., 2., 13.))
Variables actually used in tree construction:
[1] “engine.power”  “vehicle.speed”
Number of terminal nodes:  4
Residual mean deviance:  0.164 = 3093 / 18860
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -3.019e+000 -2.450e-001 -1.062e-002 -9.774e-017  2.430e-001  1.735e+000
node), split, n, deviance, yval
      * denotes terminal node

 1) root 18864 5309.0 -1.1990
   2) engine.power<82.625 1624  560.0 -1.9810 *
   3) engine.power>82.625 17240 3662.0 -1.1250
     6) vehicle.speed<19.05 9752 1994.0 -0.9339
      12) engine.power<152.965 2335  522.6 -1.2510 *
      13) engine.power>152.965 7417 1163.0 -0.8342 *
     7) vehicle.speed>19.05 7488  847.2 -1.3740 *
```

This tree model suggested that engine power is the most important explanatory variable for CO emissions, consistent with NO_x emissions. This tree will be used as reference for linear regression model development.

10.3.1.3 HC HTBR Tree Model Development

Figure 10-23 illustrates the initial tree model used for the truncated transformed HC emission rate in acceleration mode. Results for the initial model are given in Table 10-10. The tree grew into a complex model with a considerable number of branches and 30 terminal nodes.

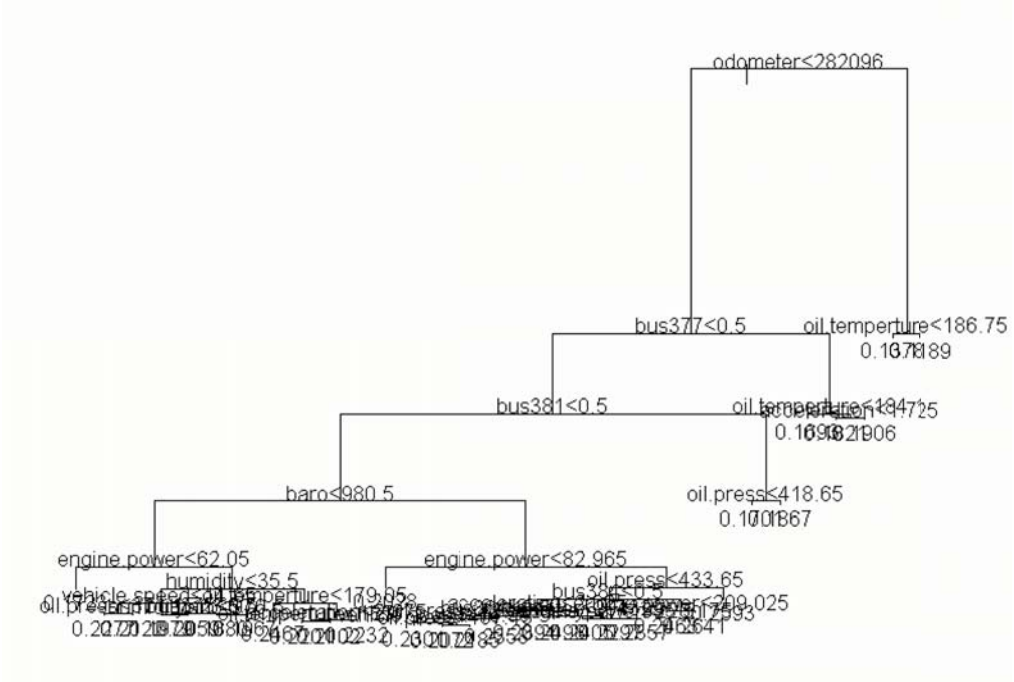


Figure 10-23 Original Untrimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Acceleration Mode

Table 10-10 Original Untrimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Acceleration Mode

```
Regression tree:
tree(formula = HC.25 ~ model.year + odometer + temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + bus360 + bus361 + bus363 +
       bus364 + bus372 + bus375 + bus377 + bus379 + bus380 + bus381 + bus382 +
       bus383 + bus384 + bus385 + dummy.grade, data = busdata10242006.1.3,
       na.action = na.exclude, mincut = 400, minsize = 800, mindev = 0.01)
Variables actually used in tree construction:
 [1] “odometer”       “bus377”         “bus381”         “baro”
 [5] “engine.power”   “humidity”       “vehicle.speed”  “oil.press”
 [9] “bus375”         “oil.temperture” “acceleration”   “bus384”
[13] “bus364”         “model.year”
Number of terminal nodes:  31
Residual mean deviance:  0.0005694 = 10.42 / 18300
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.004e-001 -1.347e-002 -2.222e-003  1.386e-016  1.091e-002  2.755e-001
```

Figure 10-23 and Table 10-12 suggest that the tree analysis of HC emission rates identified a number of buses that appear to exhibit significantly different emission rates under all load conditions than the other buses (i.e., some of the bus dummy variables appeared significant in the initial tree splits). Two bus dummy variables split the data pool at the top levels of the HC tree model. The first cut point of “odometer > 282096” in the HC tree model could be directly replaced by “bus 363 > 0.5”, because only bus 363 has an odometer reading larger than 282096.

There were three bus dummy variables that split the first three levels of the HC tree model. Although higher emissions were noted for all three pollutants for some of the 15 buses, the division was even more obvious for HC emissions (see Figure 10-9 and Table 10-4), consistent with the findings in idle and deceleration mode. Although it is tempting to develop different emission rates for these buses to reduce emission rate deviation in the sample pool, it is difficult to justify doing so. Unless there is an obvious reason to classify these three buses as high emitters (i.e., significantly higher than normal emitting vehicles, perhaps by as much as a few standard deviations from the mean), and unless there are enough data to develop separate emission rate models for high emitters, one cannot justify removing the data from the data set. Until data exist to justify treating these buses as high emitters, the bus dummy variables for individual buses are removed from the analyses and all 15 buses are treated as part of the whole data set.

Another tree model was generated excluding the bus dummy variables, model year, and odometer. This new tree model is illustrated in Figure 10-25 and Table 10-11. The tree model is then trimmed for application purposes, as was done for the NO_x and CO models.

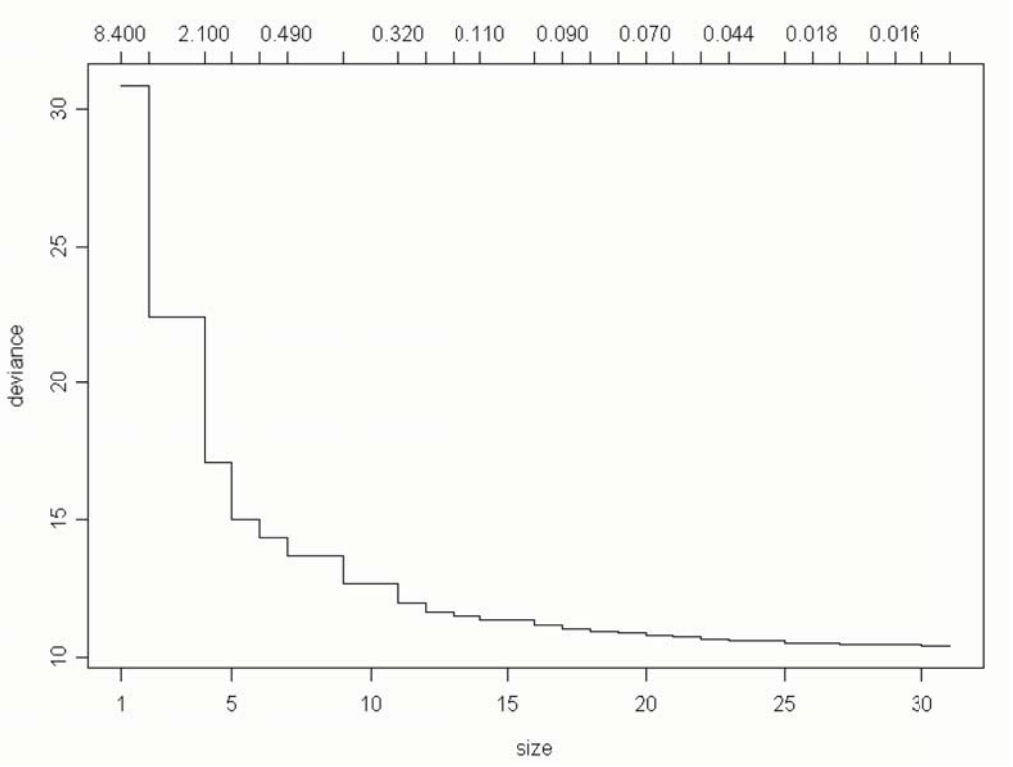


Figure 10-24 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed HC Emission Rate in Acceleration Mode

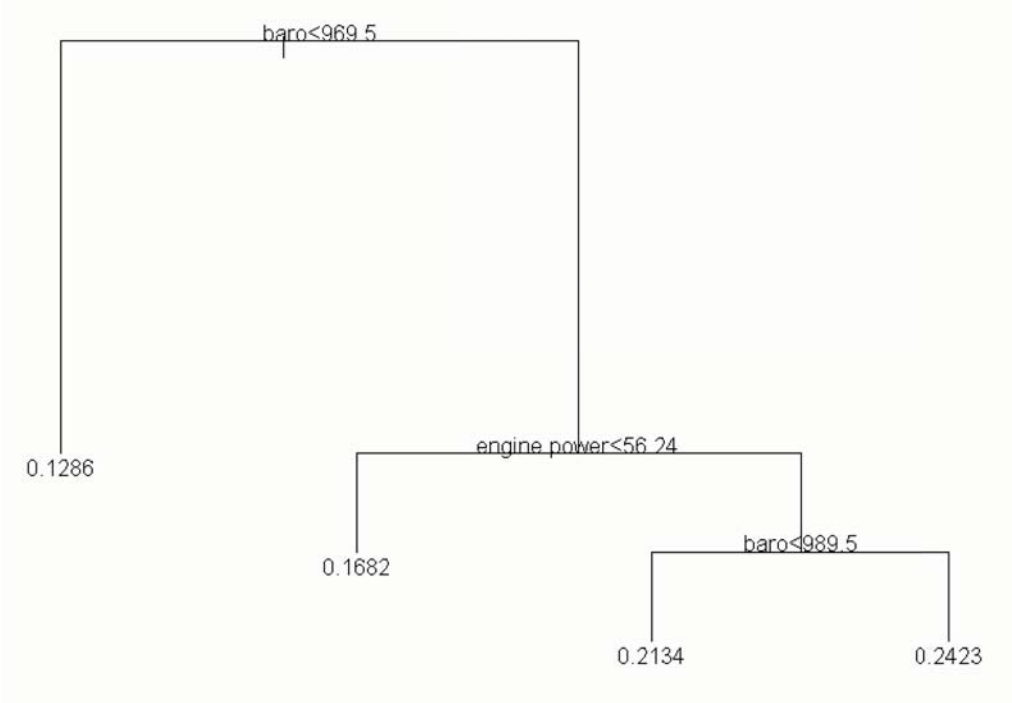


Figure 10-25 Trimmed Regression Tree Model for Truncated Transformed HC in Acceleration Mode

Table 10-11 Trimmed Regression Tree Results for Truncated Transformed HC in Acceleration Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + dummy.grade, data =
       busdata10242006.1.3, na.action = na.exclude, mincut = 400, minsize =
       800, mindev = 0.01), nodes = c(2., 6., 15., 14.))
Variables actually used in tree construction:
[1] “baro”         “engine.power”
Number of terminal nodes:  4
Residual mean deviance:  0.001018 = 18.65 / 18330
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -9.502e-002 -2.174e-002 -2.213e-003  9.390e-016  1.844e-002  3.100e-001
node), split, n, deviance, yval
      * denotes terminal node

 1) root 18330 30.840 0.2099
   2) baro<969.5 1189  1.239 0.1286 *
   3) baro>969.5 17141 21.210 0.2155
     6) engine.power<56.24 850  1.069 0.1682 *
     7) engine.power>56.24 16291 18.140 0.2180
      14) baro<989.5 13717 13.970 0.2134 *
      15) baro>989.5 2574  2.372 0.2423 *
```

The new tree model suggests that barometric pressure is the most important explanatory variable for HC emission rates. However, this finding is challenged by this fact: among

those 1189 data points (baro < 969.5) in the first left branch, 1187 data points belong to bus 363. Although this dataset was collected under a wide variety of environmental conditions, the scope of barometric pressures was limited for individual buses tested. As reported earlier, Bus 363 exhibited significantly lower HC emissions that the other buses (see Figure 10-9); the reason is not clear at this time. To develop a reasonable tree model given the limited data collected, the environmental parameters are excluded from the model until a greater distribution of environmental conditions can be represented in a test data set. With data collected from a more comprehensive testing program, environmental variables can be integrated into the model directly, or perhaps correction factors for the emission rates can be developed. The secondary trimmed tree is presented in Figure 10-26 and Table 10-12.

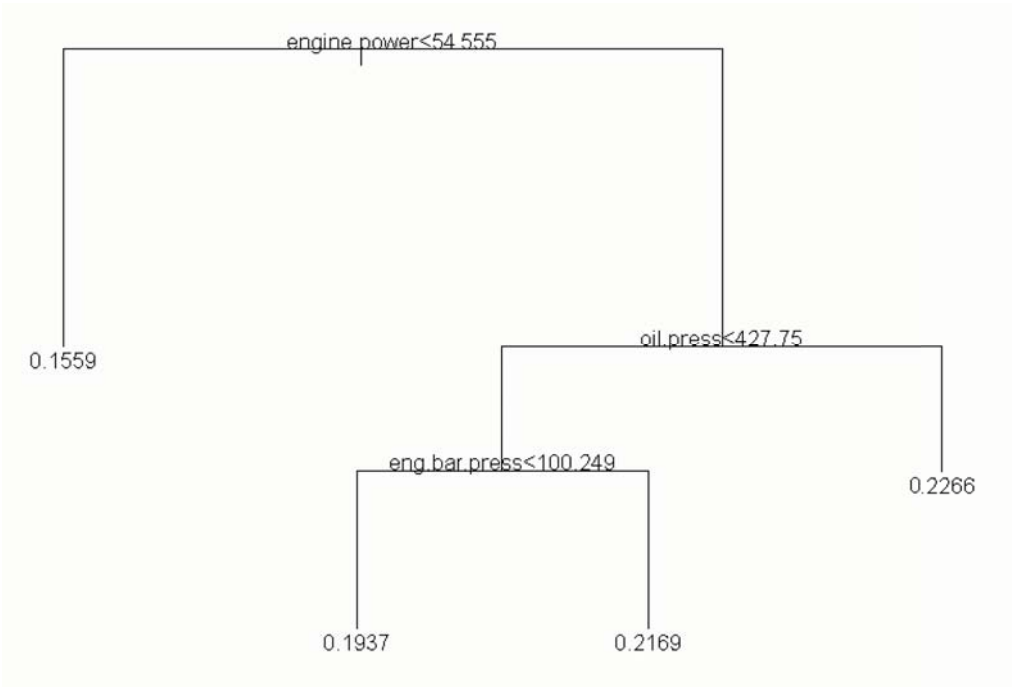


Figure 10-26 Secondary Trimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Acceleration Mode

Table 10-12 Secondary Trimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Acceleration Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ engine.power + vehicle.speed +
       acceleration + oil.temperture + oil.press + cool.temperature +
       eng.bar.press, data = busdata10242006.1.3, na.action = na.exclude,
       mincut = 400, minsize = 800, mindev = 0.1), nodes = c(7., 13., 12.))
Variables actually used in tree construction:
[1] “engine.power”  “oil.press”     “eng.bar.press”
Number of terminal nodes:  4
Residual mean deviance:  0.00136 = 24.92 / 18330
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.178e-001 -2.378e-002  6.119e-004 -4.275e-017  2.231e-002  3.223e-001
node), split, n, deviance, yval
      * denotes terminal node

 1) root 18330 30.840 0.2099
   2) engine.power<54.555 988  1.779 0.1559 *
   3) engine.power>54.555 17342 26.020 0.2130
     6) oil.press<427.75 12457 18.610 0.2076
      12) eng.bar.press<100.249 4989  9.241 0.1937 *
      13) eng.bar.press>100.249 7468  7.763 0.2169 *
     7) oil.press>427.75 4885  6.136 0.2266 *
```

This tree model suggests that engine power is the most important explanatory variable for HC emissions, consistent with analysis of NO_x and CO emission rates. HTBR results also suggest that oil pressure and engine barometric pressure may be important predictive variables for HC emissions under certain conditions. After excluding engine barometric pressure and oil pressure from the tree model, leaving engine power only, the residual mean deviation increased slightly from 24.92 to 27.34. While engine operating parameters such as oil pressure and engine barometric pressure may impact emissions, such variables are not easy to include in real-world models. The final HTBR tree for HC emissions is shown in Figure 10-27 and Table 10-13. An HC acceleration emission rate model will be developed based upon these results.

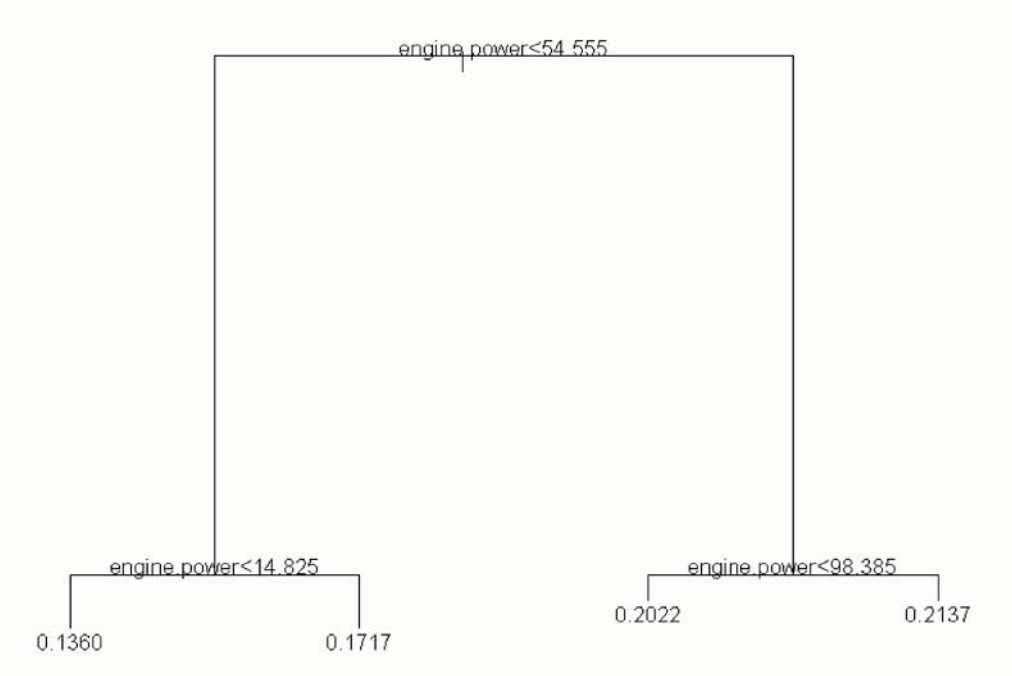


Figure 10-27 Final Regression Tree Model for Truncated Transformed HC and Engine Power in Acceleration Mode

Table 10-13 Final Regression Tree Results for Truncated Transformed HC and Engine Power in Acceleration Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ engine.power, data = busdata10242006.1.3,
na.action = na.exclude, mincut = 5, minsize = 10,mindev = 0.01), nodes = c(7., 6.,
4., 5.))
Number of terminal nodes:  4
Residual mean deviance:  0.001492 = 27.34 / 18330
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.296e-001 -2.277e-002  8.001e-005  4.271e-016  2.298e-002  3.065e-001
node), split, n, deviance, yval
      * denotes terminal node

1) root 18330 30.8400 0.2099
  2) engine.power<54.555 988  1.7790 0.1559
    4) engine.power<14.825 438  0.6518 0.1360 *
    5) engine.power>14.825 550  0.8171 0.1717 *
  3) engine.power>54.555 17342 26.0200 0.2130
    6) engine.power<98.385 1177  1.8580 0.2022 *
    7) engine.power>98.385 16165 24.0100 0.2137 *
```

10.3.2 OLS Model Development and Refinement

Once a manageable number of modal variables have been identified through regression tree analysis, the modeling process moves into the phase where ordinary least squares techniques are used to obtain a final model. The research objective here is to identify the extent to which the identified factors influence emission rates in acceleration mode. Modelers rely on previous research, a priori knowledge, educated guesses, and stepwise regression procedures to identify acceptable functional forms, to determine important interactions, and to derive statistically and theoretically defensible models. The final model will be our best understanding about the functional relationship between independent variables and dependent variables.

10.3.2.1 NO_x Emission Rate Model Development for Acceleration Mode

Based on previous analysis, truncated transformed NO_x will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing the performance of statistical models. HTBR tree model results suggest that engine power is the best one to begin with. Linear regression model with engine power will be developed first, followed by a combined power and vehicle speed model.

10.3.2.1.1 Linear Regression Model with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1(engine.power) + Error \qquad (1.1)$$

The regression run yields the results shown in Table 10-14.

Table 10-14 Regression Result for NO_x Model 1.1

```
Call: lm(formula = NOx.50 ~ engine.power, data = busdata10242006.1.3, na.action =
na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.4093 -0.08133 0.005414 0.07084 0.9344

Coefficients:
                Value Std. Error  t value Pr(>|t|)
 (Intercept)   0.3054   0.0021   147.9391   0.0000
engine.power   0.0008   0.0000    83.3557   0.0000

Residual standard error: 0.09781 on 18892 degrees of freedom
Multiple R-Squared: 0.2689
F-statistic: 6948 on 1 and 18892 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.9387

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1   66.4763 66.47630 6948.175     0
   Residuals 18892  180.7482  0.00957
```

These results suggest that engine power explains about 27% of the variance in truncated transformed NO_x. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, residual normality is checked by examining quantile-quantile (QQ) plot and checking constancy of variance by examining residuals vs. fitted values.

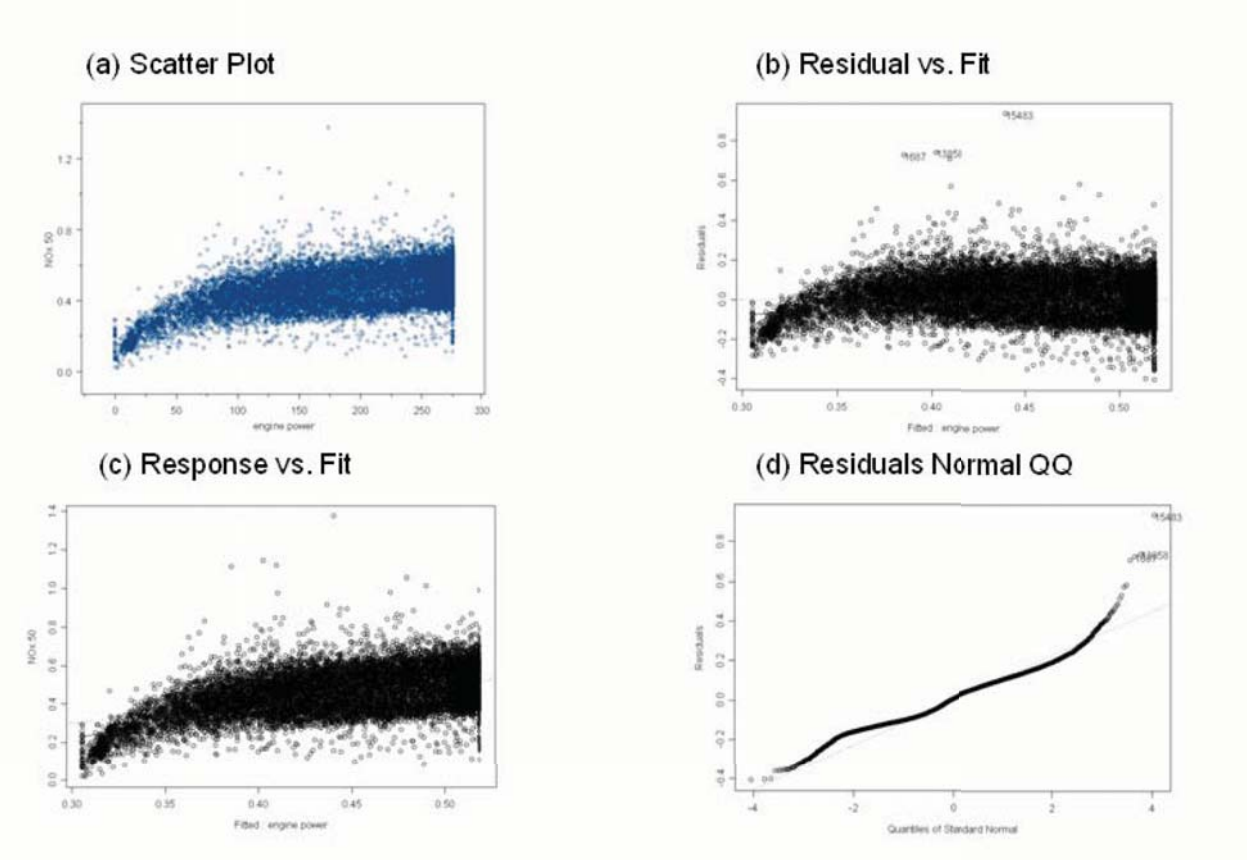


Figure 10-28 QQ and Residual vs. Fitted Plot for NO_x Model 1.1

The residual plot in Figure 10-28 shows a slight departure from linear regression assumptions indicating a need to explore a curvilinear regression function. Since the variability at the different X levels appears to be fairly constant, a transformation on X is considered. The reason to consider transformation first is to avoid multicollinearity brought about by adding the second-order of X. Based on the prototype plot in Figure 10-28, the square root transformation and logarithmic transformation are tested. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to determine which transformation is most effective.

$$Y = \beta_0 + \beta_1 engine.power^{(1/2)} + Error \qquad (1.2)$$

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + Error \qquad (1.3)$$

The result for Model 1.2 will be shown in Table 10-15 and Figure 10-29, while the result for Model 1.3 will be shown in Table 10-16 and Figure 10-30.

Table 10-15 Regression Result for NO_x Model 1.2

```
Call: lm(formula = NOx.50 ~ engine.power^(1/2), data = busdata10242006.1.3,
na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.4106 -0.07981 0.004093 0.06858 0.9248

Coefficients:
                        Value Std. Error t value Pr(>|t|)
          (Intercept)  0.1912  0.0030    63.2141  0.0000
I(engine.power^(1/2))  0.0196  0.0002    93.5953  0.0000

Residual standard error: 0.09455 on 18892 degrees of freedom
Multiple R-Squared: 0.3168
F-statistic: 8760 on 1 and 18892 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                       (Intercept)
I(engine.power^(1/2)) -0.9738

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                         Df Sum of Sq  Mean Sq  F Value Pr(F)
I(engine.power^(1/2))     1   78.3199 78.31986 8760.082     0
            Residuals 18892  168.9047  0.00894
```

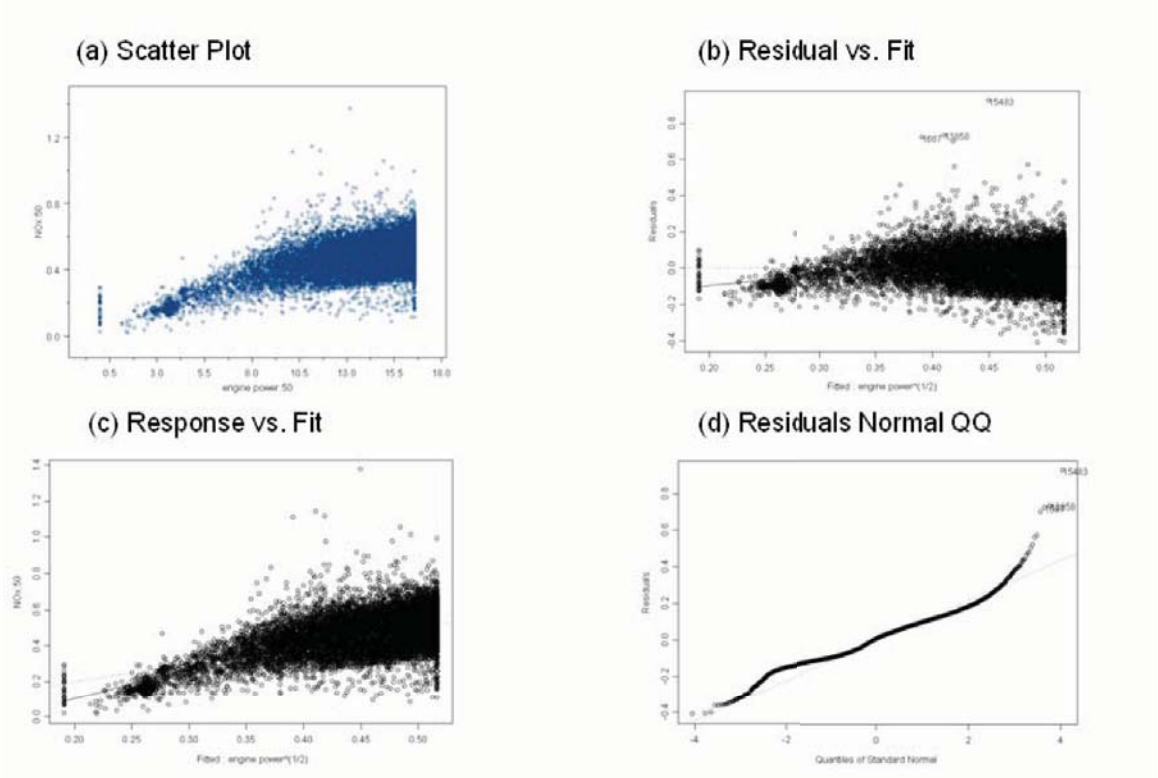


Figure 10-29 QQ and Residual vs. Fitted Plot for NO_x Model 1.2

Table 10-16 Regression Result for NO_x Model 1.3

```
*** Linear Model ***

Call: lm(formula = NOx.50 ~ log10(engine.power + 1), data = busdata10242006.1.3,
na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.4109 -0.07485 0.001841 0.06716 0.9119

Coefficients:
                           Value Std. Error  t value Pr(>|t|)
            (Intercept)  -0.0514   0.0052    -9.7873   0.0000
log10(engine.power + 1)   0.2291   0.0023    99.6000   0.0000

Residual standard error: 0.09263 on 18892 degrees of freedom
Multiple R-Squared: 0.3443
F-statistic: 9920 on 1 and 18892 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                         (Intercept)
log10(engine.power + 1) -0.9917

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1   85.1206 85.12056 9920.161     0
              Residuals 18892  162.1040  0.00858
```

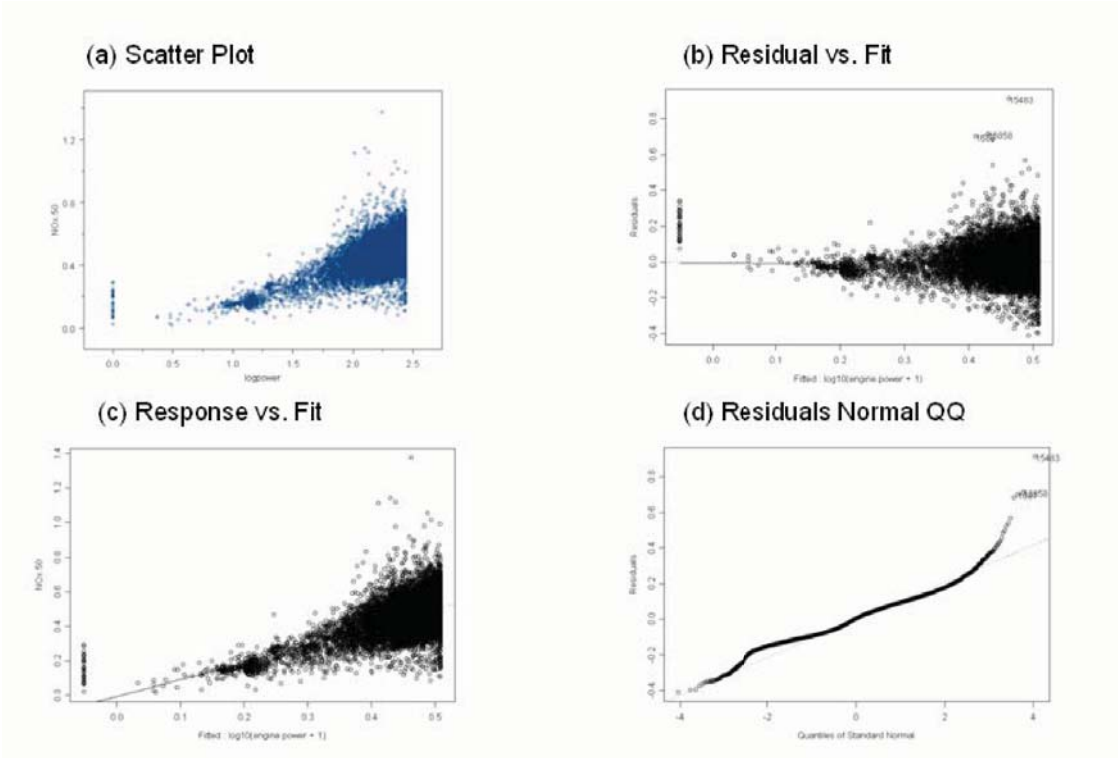


Figure 10-30 QQ and Residual vs. Fitted Plot for NO_x Model 1.3

The results suggest that by using square root transformed engine power, the model increases the amount of variance explained in truncated transformed NO_x from about 27% (Model 1.1) to about 32% (Model 1.2), while the increase is about 34% (Model 1.3) by using log transformed engine power.

Model 1.3 improves the R^2 more than does Model 1.2. The residuals scatter plot for Model 1.3 (Figure 10-30) shows a more reasonably linear relationship than Model 1.2 (Figure 10-29). Figure 10-30 also shows that Model 1.3 does a better job in improving the pattern of variance. QQ plot shows general normality with the exceptions arising in the tails.

10.3.2.1.2 Linear Regression Model with Engine Power and Vehicle Speed

HTBR tree model results also suggest that vehicle speed may be an important predictive variable for emissions under certain conditions. After developing a linear regression model with engine power, adding vehicle speed might improve the model predictive ability. The new model is proposed as:

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + \beta_2 vehicle.speed + Error \qquad (1.4)$$

The result for Model 1.4 is shown in Table 10-17 and Figure 10-31.

Table 10-17 Regression Result for NO_x Model 1.4

```
Call: lm(formula = NOx.50 ~ log10(engine.power + 1) + vehicle.speed, data =
       busdata10242006.1.3, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.4133 -0.07416 0.004219 0.06303 0.9019

Coefficients:
                           Value Std. Error  t value Pr(>|t|)
            (Intercept)  -0.0195   0.0053    -3.6693   0.0002
log10(engine.power + 1)   0.2007   0.0025    79.3288   0.0000
          vehicle.speed   0.0019   0.0001    25.1554   0.0000

Residual standard error: 0.09112 on 18891 degrees of freedom
Multiple R-Squared: 0.3656
F-statistic: 5442 on 2 and 18891 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                        (Intercept) log10(engine.power + 1)
log10(engine.power + 1) -0.9681
          vehicle.speed  0.2383     -0.4470

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1   85.1206 85.12056 10251.92     0
          vehicle.speed     1    5.2540  5.25404   632.80     0
              Residuals 18891  156.8499  0.00830
```

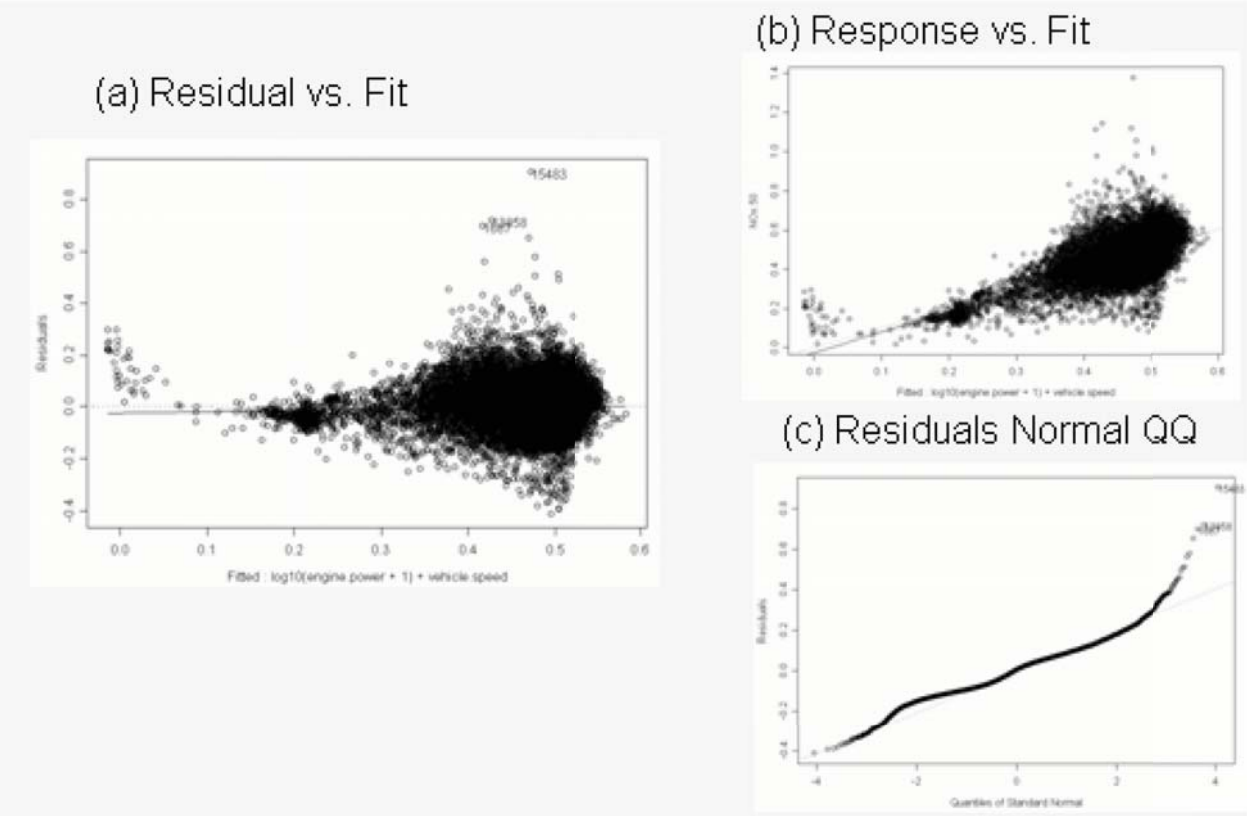


Figure 10-31 QQ and Residual vs. Fitted Plot for NO_x Model 1.4

The results suggest that by using vehicle speed and transformed engine power, the model increases the amount of variance explained in truncated transformed NO_x from about 34% (Model 1.3) to about 37% (Model 1.4). The residuals scatter plot for Model 1.4 (Figure 10-31) shows a more reasonably linear relationship. Figure 10-31 also shows that model 1.4 does a better job in improving the pattern of variance. QQ plot shows general normality, with deviation at the tails.

10.3.2.1.3 Linear Regression Model with Dummy Variables

Figure 10-19 suggests that the relationship between NO_x and engine power may be somewhat different across the engine power ranges identified in the tree analysis. That is, there may be higher or lower NO_x emissions in different engine power operating ranges. One dummy variable is created to represent different engine power ranges identified in Figure 10-19 for use in linear regression analysis as illustrated below:

Engine power (bhp)	dummy1
< 72.30	1
≥ 72.30	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variables and interactions can help improve the model:

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + \beta_2 vehicle.speed + \beta_3\, dummy1 + \beta_4 dummy1\ log_{10}(engine.power+1) + \beta_5\, dummy1 vehicle.speed + Error \quad (1.5)$$

The result for Model 1.5 is shown in Table 10-18 and Figure 10-32.

Table 10-18 Regression Result for NO_x Model 1.5

```
Call: lm(formula = NOx.50 ~ log10(engine.power + 1) + vehicle.speed + dummy1 *
log10( engine.power + 1) + dummy1:vehicle.speed, data = busdata10242006.1.3,
       na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.4124 -0.07157 0.003012 0.06319 0.8924

Coefficients:
                                 Value Std. Error  t value Pr(>|t|)
                   (Intercept)   0.1439   0.0115    12.4979   0.0000
       log10(engine.power + 1)   0.1281   0.0052    24.8261   0.0000
                 vehicle.speed   0.0023   0.0001    28.9191   0.0000
                        dummy1  -0.1492   0.0148   -10.0783   0.0000
dummy1:log10(engine.power + 1)   0.0609   0.0081     7.4995   0.0000
          dummy1:vehicle.speed  -0.0035   0.0003   -10.4883   0.0000

Residual standard error: 0.09022 on 18888 degrees of freedom
Multiple R-Squared: 0.3781
F-statistic: 2297 on 5 and 18888 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                                 Df Sum of Sq  Mean Sq  F Value
       log10(engine.power + 1)    1   85.1206 85.12056 10456.89
                 vehicle.speed    1    5.2540  5.25404   645.45
                        dummy1    1    1.9017  1.90166   233.62
dummy1:log10(engine.power + 1)    1    0.3018  0.30180    37.08
          dummy1:vehicle.speed    1    0.8955  0.89546   110.01
                     Residuals 18888  153.7510  0.00814

                                     Pr(F)
       log10(engine.power + 1) 0.000000e+000
                 vehicle.speed 0.000000e+000
                        dummy1 0.000000e+000
dummy1:log10(engine.power + 1) 1.158203e-009
          dummy1:vehicle.speed 0.000000e+000
                     Residuals
```

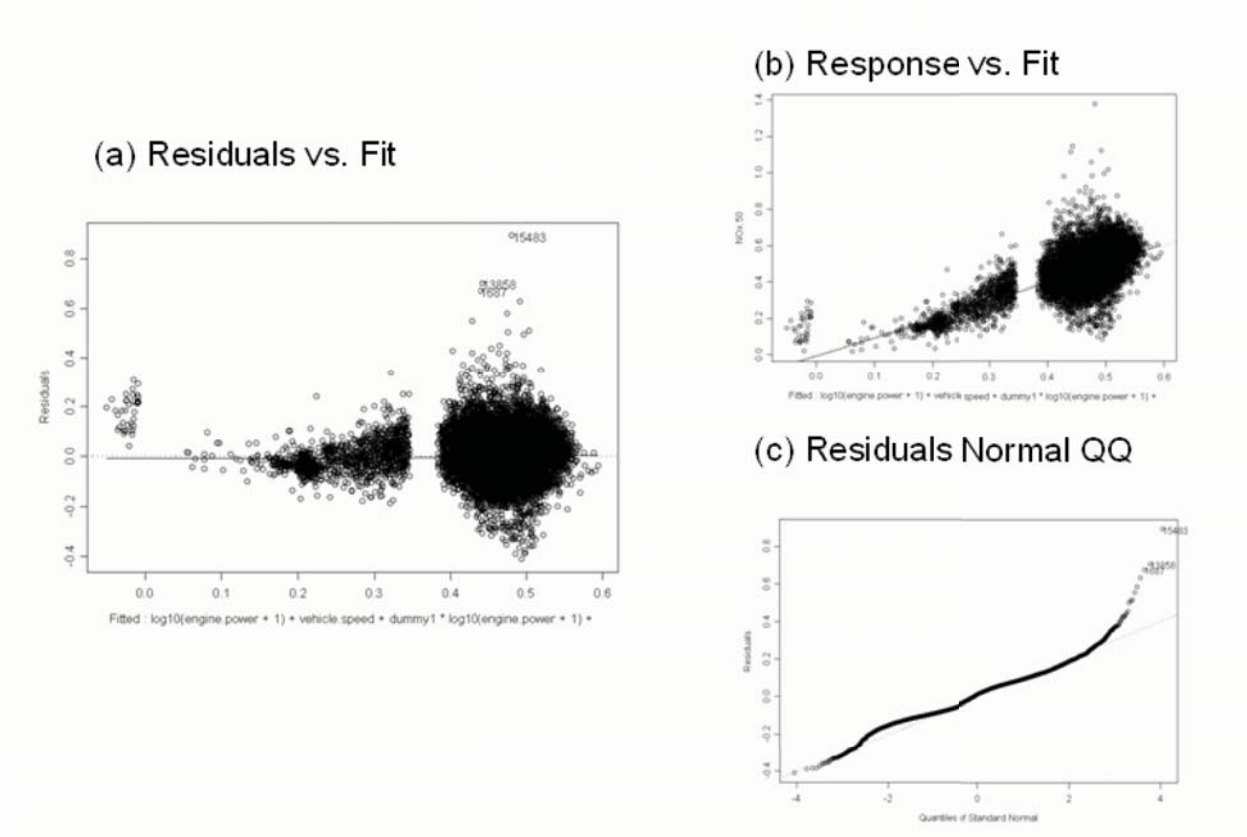


Figure 10-32 QQ and Residual vs. Fitted Plot for NO_x Model 1.5

The results suggest that by using dummy variables and interactions with transformed engine power and vehicle speed, the model slightly increases the amount of variance explained in truncated transformed NO_x from about 37% (Model 1.4) to about 38% (Model 1.5).

Model 1.5 slightly improves the R^2 compared to Model 1.4. The residuals scatter plot for Model 1.5 (Figure 10-32) shows a slightly more linear relationship. Figure 10-32 also shows that Model 1.4 may also do a slightly better job in improving the pattern of variance. The QQ plot shows general normality with the exceptions arising in the tails. However, it is important to note that the model improvement, in terms of amount of variance explained by the model, is marginal at best.

10.3.2.1.4 Model Discussions

The performance of alternative models can be evaluated by comparing model predictions and actual observations for emission rates. The performance of the model can be evaluated in terms of precision and accuracy (Neter et al. 1996). The R^2 value is an indication of precision. Usually, higher R^2 values imply a higher degree of precision and less unexplained variability in

model predictions than lower R^2 values. The slope of the trend line for the observed versus predicted values is an indication of accuracy. A slope of one indicates an accurate prediction, in that the prediction of the model corresponds to an observation. Since the R^2 and slope are derived by comparing model predictions and actual observations for emission rates, these numbers will be different from those observed in linear regression models.

The models' predictive ability is also evaluated using the root mean square error (RMSE) and the mean prediction error (MPE) (Neter et al. 1996). The RMSE is a measure of prediction error. When comparing two models, the model with a smaller RMSE is a better predictor of the observed phenomenon. Ideally, mean prediction error is close to zero. RMSE and MPE are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$ Equation (10-1)

$$MPE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$ Equation (10-1)

where:

RMSE:	=	root mean square error
n:	=	number of observations
y_i:	=	observaton y
$\bar{y}_i$:	=	mean of observation y
MPE:	=	mean predictive error

Previous sections provide the model development process from one model to another model. To test whether the linear regression with power was a beneficial addition to the regression tree model, the mean ERs at HTBR end nodes (single value) are compared to the predictions from the linear regression function with engine power. The results of the performance evaluation are shown in Table 10-19. The improvement in R^2 associated with moving toward a linear function of engine power is large. Hence, the use of the linear regression function will provide a significant improvement in spatial and temporal model prediction capability. But this linear regression function might still be improved. Since the R^2 and slope in Table 10-19 are derived by comparing model predictions and actual observations for emission rates (untransformed y), these numbers are different from those obtained from linear regression models.

Two transforms of engine power were tested: square root transformation and log transformation. The results of the performance evaluation are shown in Table 10-19. Results suggest that linear regression function with log transformation performs slightly better.

The addition of vehicle speed was also tested. The results of the performance evaluation are shown in Table 10-19. Analysis results suggest that a linear regression function for engine power and vehicle speed also performs slightly better.

Since the regression tree modeling exercise indicated that a number of power cutpoints may play a role in the emissions process, an additional modeling run was performed. The results of the performance evaluation are also shown in Table 10-19. Analysis results suggest that a linear regression function with dummy variables performs slightly better than the model without the power cutpoints.

Table 10-19 Comparative Performance Evaluation of NO_x Emission Rate Models

	Coefficient of determination (R^2)	**Slope (β_1)**	**RMSE**	**MPE**
Mean ERs	0.00026	1.000	0.10455	0.00001
Linear Regression (Power)	0.190	0.838	0.09463	0.00428
Linear Regression ($Power^5$)	0.215	0.901	0.09321	0.00898
Linear Regression (log(Power))	0.236	1.012	0.09178	0.00872
Linear Regression (log(Power)+Speed)	0.268	1.001	0.08982	0.00837
Linear Regression (log(Power)+Speed+Dummy)	0.280	1.036	0.08912	0.00834

Although the linear regression function with dummy variables works slightly better than the linear regression function with engine power and vehicle speed, it introduces more explanatory variables (dummy variables and the interaction with engine power) and increases the complexity of the regression model. There is only one regression function for Model 1.4 while there are two regression functions for Model 1.5. There is also no obvious reason why the engine may be performing slightly differently within these power regimes, yielding different regression slopes and intercepts. The fuel injection systems in these engines may operate slightly differently under low load (near-idle) and high load conditions. This fuel injection system may be controlled by the engine computer. There may be a sufficient number of low power cruise operations and high power cruise operations that are incorrectly classified, and that may be better classified as idle or acceleration events (perhaps due to GPS speed data errors). In any case, because the model with dummy variables does not perform appreciably better than the model without the dummy variables, the dummy variables are not included in the final model selection at this time. These

dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis. Model 1.4 is selected as the preliminary 'final' model.

The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform with a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have uniform variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying any patterns in the residuals. Figure 10-31(c) shows this plot for NO_x model. Without considering variance due to high emission points and zero load data, there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 10-31 plot (c) shows the normal quantile plot of the NO_x model. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm \sqrt{MSE}$ and about 90 percent fall between $\pm 1.645 \sqrt{MSE}$. Actually, 72.64% of residuals fall within the first limits, while 93.79% of residuals fall within the second limits. Thus, the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but are due to the nature of the data set. For example, even after the transformation, the response variable is not a true normal distribution.

Based on the above analysis, the final NO_x emission model for cruise mode is:

$$NO_x = [-0.0195 + 0.201\log_{10}(\text{engine.power}+1) + 0.0019\text{vehicle.speed}]^2$$

Analysis results support the observation that the final NO_x emission model is significantly better at explaining variability without making the model too complex. Since there is only one engine type, complexity may not be valid in terms of transferability. This model is specific to the engine classes employed in the transit bus operations. Different models may need to be developed for other engine classes and duty cycles.

10.3.2.2 CO Emission Rate Model Development for Acceleration Mode

Based on previous analysis, truncated transformed CO will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing statistical models. HTBR tree model results suggest that engine power is best to begin with.

10.3.2.2.1 Linear Regression Model with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1 engine.power + Error \qquad (2.1)$$

The regression run yields the results shown in Table 10-20.

Table 10-20 Regression Result for CO Model 2.1

```
Call: lm(formula = log.CO ~ engine.power, data = busdata10242006.1.3, na.action =
na.exclude)
Residuals:
    Min      1Q   Median     3Q   Max
 -3.151 -0.3515 -0.05231 0.3448 1.453

Coefficients:
                Value Std. Error   t value  Pr(>|t|)
 (Intercept)   -1.8549    0.0100  -185.2318    0.0000
engine.power    0.0031    0.0000    69.7761    0.0000

Residual standard error: 0.473 on 18862 degrees of freedom
Multiple R-Squared: 0.2052
F-statistic: 4869 on 1 and 18862 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.939

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1  1089.300 1089.300 4868.698     0
   Residuals 18862  4220.097    0.224
```

The results suggest that engine power explains about 21% of the variance in truncated transformed CO. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, the normality is examined in the QQ plot and constancy of variance is checked by examining residuals vs. fitted values.

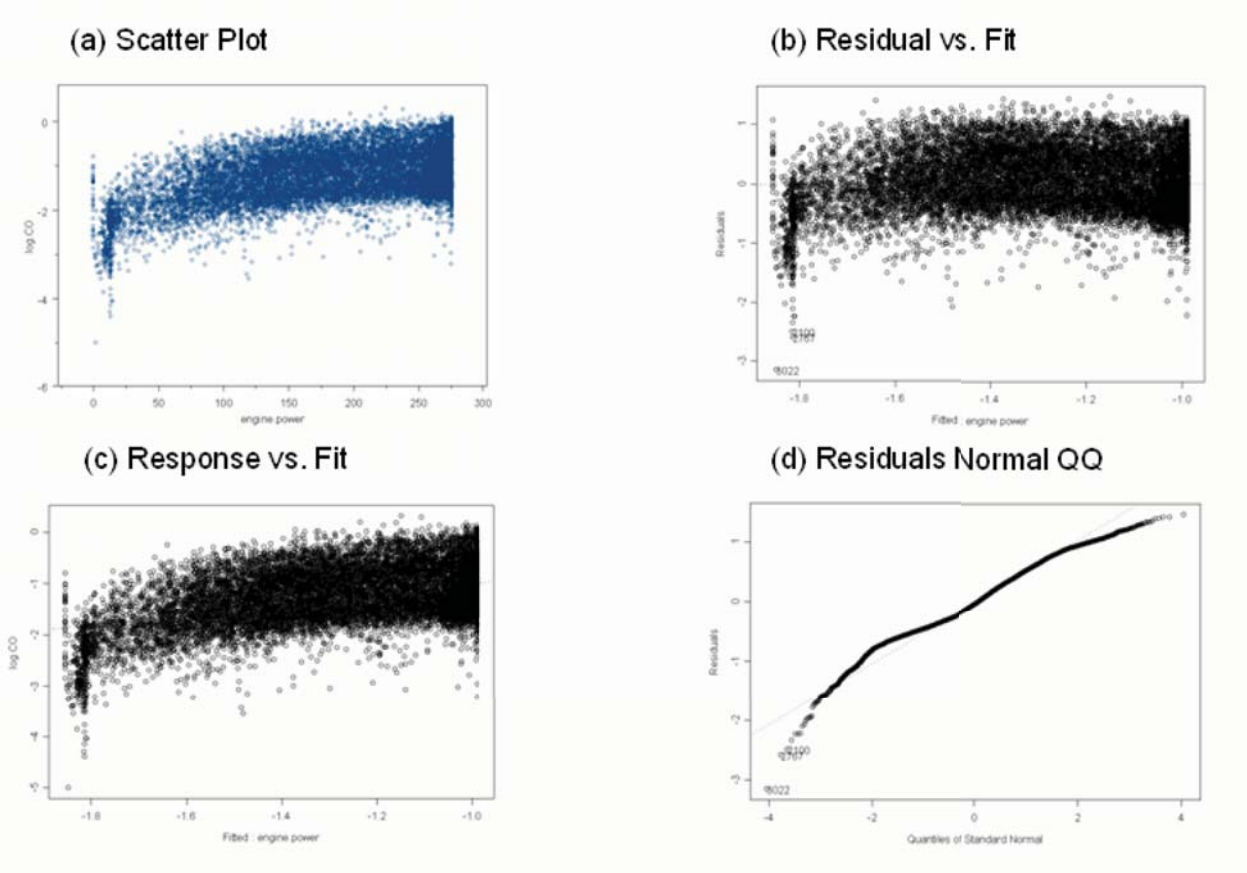


Figure 10-33 QQ and Residual vs. Fitted Plot for CO Model 2.1

The residual plot in Figure 10-33 shows a slight departure from linear regression assumptions indicating a need to explore a curvilinear regression function. Since the variability at the different X levels appears to be fairly constant, a transformation on X is considered. The reason to consider transformation first is avoiding multicollinearity brought about by adding the second-order of X. Based on the prototype plot in Figure 10-33, the square root transformation and logarithmic transformation were tested. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to determine which transformation is most effective.

$$Y = \beta_0 + \beta_1 engine.power\text{^}(1/2) + Error \qquad (2.2)$$

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + Error \qquad (2.3)$$

The result for Model 2.2 is shown in Table 10-21 and Figure 10-34, while the result for Model 2.3 is shown in Table 10-22 and Figure 10-35.

Table 10-21 Regression Result for CO Model 2.2

```
Call: lm(formula = log.CO ~ engine.power^(1/2), data = busdata10242006.1.3,
na.action = na.exclude)
Residuals:
    Min      1Q  Median     3Q  Max
 -2.798 -0.3492 -0.0529 0.3381 1.52

Coefficients:
                          Value Std. Error   t value  Pr(>|t|)
         (Intercept)   -2.3146    0.0149  -155.8023    0.0000
I(engine.power^(1/2))    0.0793    0.0010    77.1161    0.0000

Residual standard error: 0.4626 on 18862 degrees of freedom
Multiple R-Squared: 0.2397
F-statistic: 5947 on 1 and 18862 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                       (Intercept)
I(engine.power^(1/2)) -0.974

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                          Df Sum of Sq  Mean Sq  F Value Pr(F)
I(engine.power^(1/2))     1  1272.706 1272.706 5946.896     0
            Residuals 18862  4036.691    0.214
```

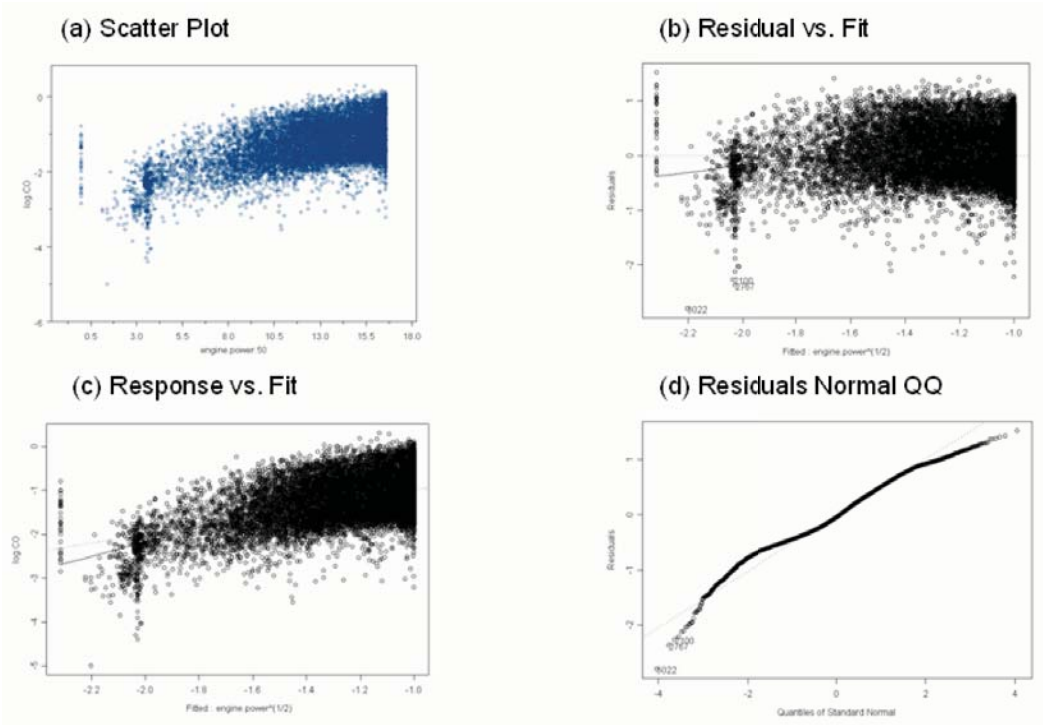


Figure 10-34 QQ and Residual vs. Fitted Plot for CO Model 2.2

Table 10-22 Regression Result for CO Model 2.3

```
Call: lm(formula = log.CO ~ log10(engine.power + 1), data = busdata10242006.1.3,
na.action = na.exclude)
Residuals:
    Min      1Q   Median     3Q   Max
 -2.187 -0.3475 -0.05182 0.3313 2.475

Coefficients:
                            Value Std. Error   t value  Pr(>|t|)
            (Intercept)   -3.2695    0.0261  -125.3639    0.0000
log10(engine.power + 1)    0.9152    0.0114    80.0560    0.0000

Residual standard error: 0.4584 on 18862 degrees of freedom
Multiple R-Squared: 0.2536
F-statistic: 6409 on 1 and 18862 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                          (Intercept)
log10(engine.power + 1) -0.9918

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                             Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1  1346.515 1346.515 6408.966     0
              Residuals 18862  3962.882    0.210
```

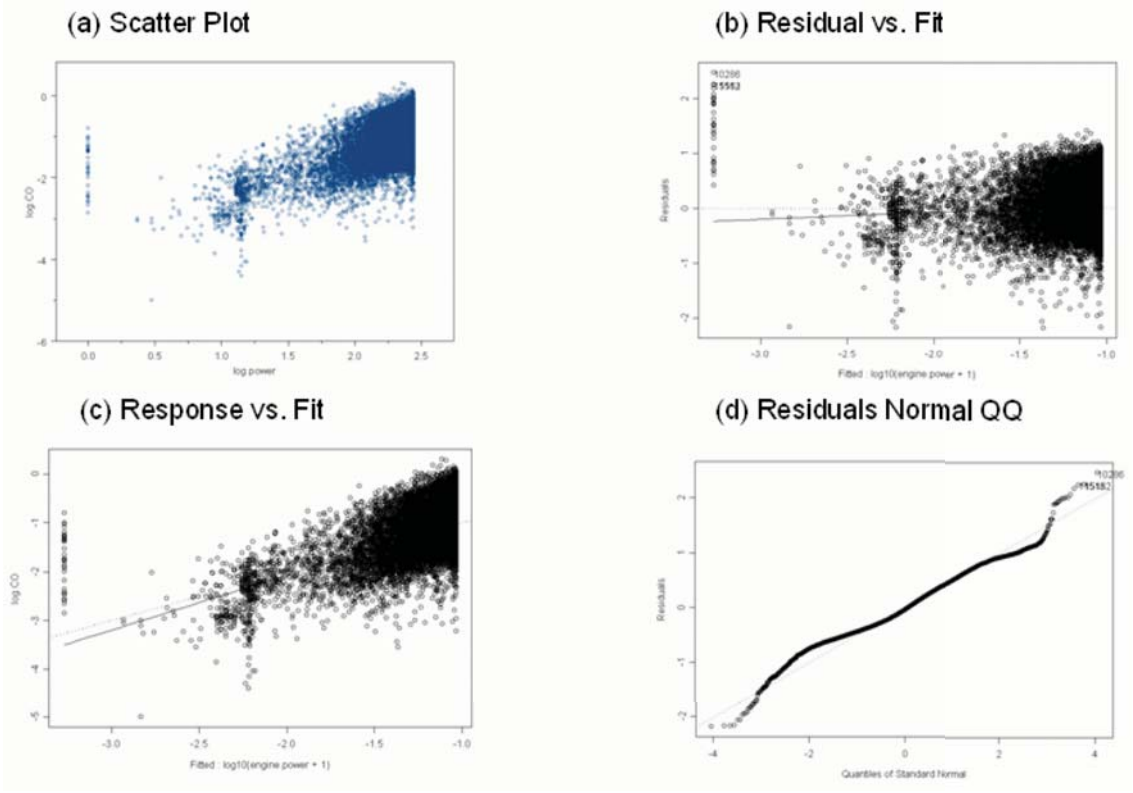


Figure 10-35 QQ and Residual vs. Fitted Plot for CO Model 2.3

The results suggest that by using transformed engine power, the model increases the amount of variance explained in truncated transformed CO from about 21% to about 25%.

Model 2.3 improves the R^2 more than does Model 2.2. The residuals scatter plot for Model 2.3 (Figure 10-35) shows a more reasonably linear relationship than Model 2.2 (Figure 10-34). Figure 10-35 also shows that Model 2.3 does a better job in improving the pattern of variance. QQ plot shows general normality with the exceptions arising in the tails.

10.3.2.2.2 Linear Regression Model with Engine Power and Vehicle Speed

HTBR tree model results also suggest that vehicle speed may be an important predictive variable for emissions under certain conditions. After developing a linear regression model with engine power, adding vehicle speed might improve the model predictive ability. The new model is proposed as:

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + \beta_2 vehicle.speed + Error \qquad (2.4)$$

The result for Model 2.4 will be shown in Table 10-23 and Figure 10-36.

Table 10-23 Regression Result for CO Model 2.4

```
Call: lm(formula = log.CO ~ log10(engine.power + 1) + vehicle.speed, data =
       busdata10242006.1.3, na.action = na.exclude)
Residuals:
    Min     1Q   Median     3Q   Max
 -2.299 -0.236 -0.02889 0.2281 3.209

Coefficients:
                            Value Std. Error   t value  Pr(>|t|)
            (Intercept)   -3.7472    0.0225  -166.3169    0.0000
log10(engine.power + 1)    1.3412    0.0107   125.1282    0.0000
          vehicle.speed   -0.0285    0.0003   -89.0585    0.0000

Residual standard error: 0.3846 on 18861 degrees of freedom
Multiple R-Squared: 0.4746
F-statistic: 8517 on 2 and 18861 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                         (Intercept) log10(engine.power + 1)
log10(engine.power + 1) -0.9683
          vehicle.speed  0.2380     -0.4463

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1  1346.515 1346.515 9103.577     0
          vehicle.speed     1  1173.140 1173.140 7931.415     0
              Residuals 18861  2789.742    0.148
```

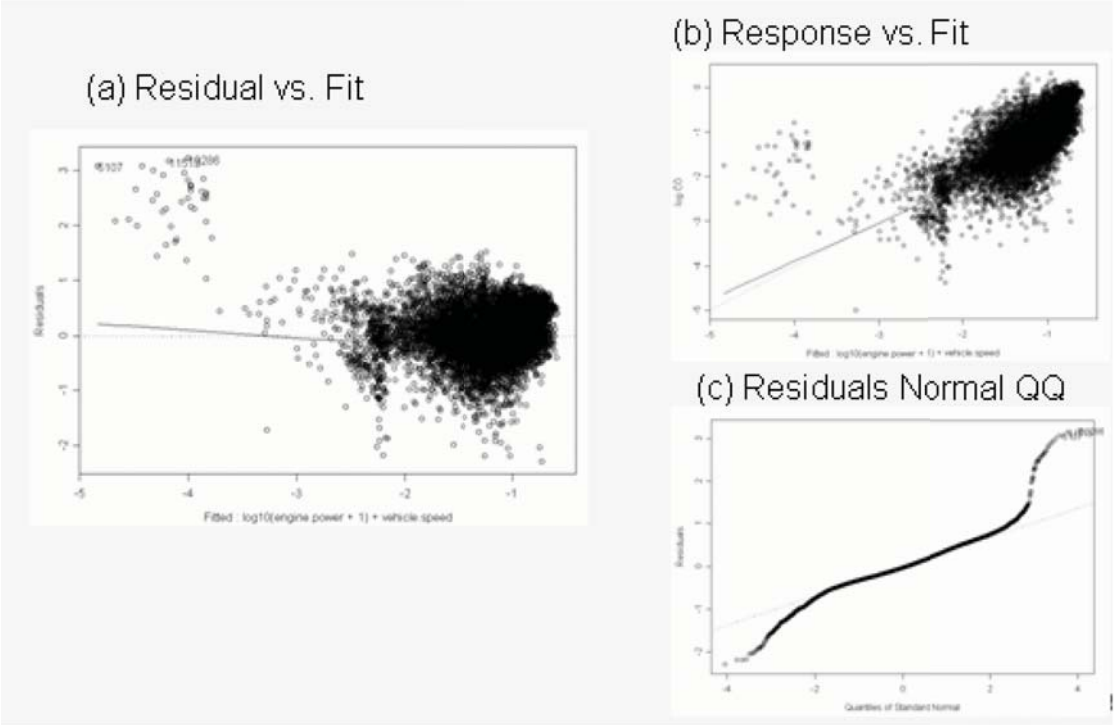


Figure 10-36 QQ and Residual vs. Fitted Plot for CO Model 2.4

The results suggest that by using vehicle speed and transformed engine power, the model increases the amount of variance explained in truncated transformed CO from about 25% to about 47%.

Model 2.4 tremendously improves the R^2 achieved in Model 2.3. The residuals scatter plot for Model 2.4 (Figure 10-36) shows a reasonably linear relationship. Figure 10-36 also shows that Model 2.4 does a slightly better job in improving the pattern of variance. QQ plot shows general normality with the exceptions arising in the tails.

10.3.2.2.3 Linear Regression Model with Dummy Variables

Figure 10-22 suggests that the relationship between CO and engine power may be somewhat different across the engine power ranges identified in the tree analysis. That is, there may be higher or lower CO emissions in different engine power operating ranges. One dummy variable is created to represent different engine power ranges identified in Figure 10-22 for use in linear regression analysis as illustrated below:

Engine power (bhp)	Dummy1
< 82.625	1
≥ 82.625	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variable and interactions can help improve the model.

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + \beta_2 vehicle.speed + \beta_3\, dummy1 + \beta_4 dummy1\ log_{10}(engine.power+1) + \beta_5\, dummy1 vehicle.speed + Error \quad (2.5)$$

The result for Model 2.5 are shown in Table 10-24 and Figure 10-37.

Table 10-24 Regression Result for CO Model 2.5

```
Call: lm(formula = log.CO ~ log10(engine.power + 1) + vehicle.speed + dummy1 * log10(
      engine.power + 1) + dummy1 * vehicle.speed, data = busdata10242006.1.3,
      na.action = na.exclude)
Residuals:
    Min     1Q   Median     3Q   Max
 -2.383 -0.233 -0.02602 0.2235 2.124

Coefficients:
                                  Value Std. Error   t value  Pr(>|t|)
                    (Intercept)   -4.4320    0.0498   -89.0217    0.0000
       log10(engine.power + 1)    1.6746    0.0222    75.4956    0.0000
                  vehicle.speed   -0.0333    0.0003  -102.3796    0.0000
                         dummy1    1.4402    0.0614    23.4537    0.0000
dummy1:log10(engine.power + 1)   -1.0349    0.0321   -32.2634    0.0000
          dummy1:vehicle.speed    0.0414    0.0013    32.8802    0.0000

Residual standard error: 0.3655 on 18858 degrees of freedom
Multiple R-Squared: 0.5255
F-statistic: 4177 on 5 and 18858 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                                (Intercept) log10(engine.power + 1)
       log10(engine.power + 1) -0.9926
                  vehicle.speed  0.3000     -0.4020
                         dummy1 -0.8108      0.8047
dummy1:log10(engine.power + 1)  0.6864     -0.6915
          dummy1:vehicle.speed -0.0774      0.1038

                                vehicle.speed  dummy1
       log10(engine.power + 1)
                  vehicle.speed
                         dummy1 -0.2432
dummy1:log10(engine.power + 1)  0.2780       -0.9559
          dummy1:vehicle.speed -0.2581        0.0018

                                dummy1:log10(engine.power + 1)
       log10(engine.power + 1)
                  vehicle.speed
                         dummy1
dummy1:log10(engine.power + 1)
          dummy1:vehicle.speed -0.1467

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                                   Df Sum of Sq  Mean Sq  F Value Pr(F)
       log10(engine.power + 1)      1  1346.515 1346.515 10079.07     0
                  vehicle.speed     1  1173.140 1173.140  8781.31     0
                         dummy1     1    23.180   23.180   173.51     0
dummy1:log10(engine.power + 1)     1   102.793  102.793   769.44     0
          dummy1:vehicle.speed     1   144.430  144.430  1081.10     0
                     Residuals 18858  2519.338    0.134
```

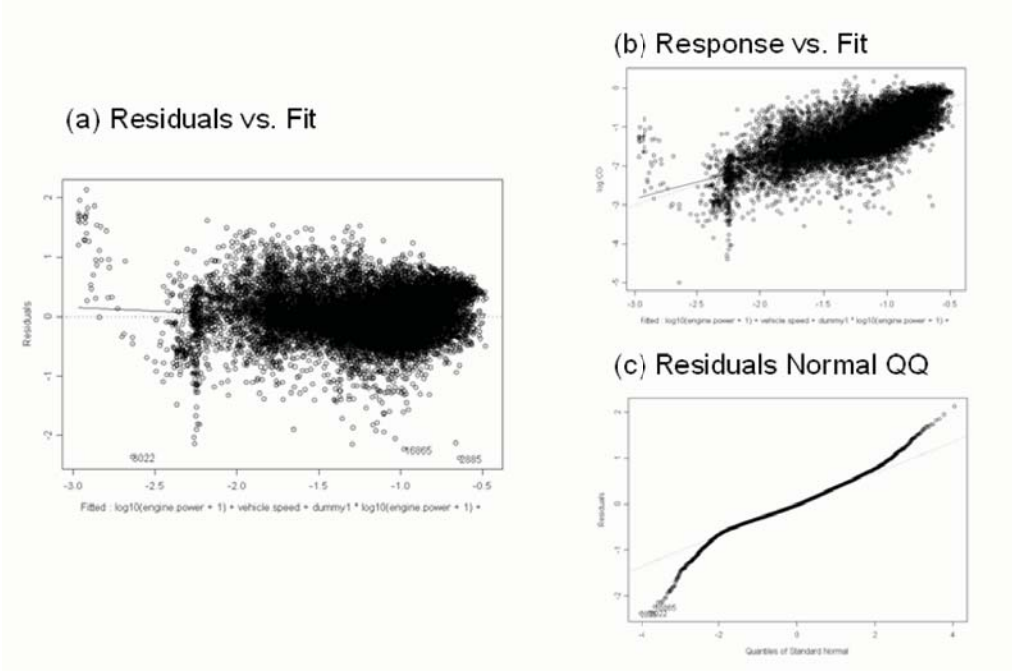


Figure 10-37 QQ and Residual vs. Fitted Plot for CO Model 2.5

Model 2.5 does improve R^2 from around 0.47 to around 0.52 by adding the dummy variables. The residuals scatter plot for Model 2.5 (Figure 10-37) shows a slightly more linear relation. Figure 10-37 also shows that Model 2.5 perhaps may improve the pattern of variance. The QQ plot again shows general normality with the exceptions arising in the tails. However, it is important to note that the model improvement, in terms of amount of variance explained by the model, is not large.

Then three more dummy variables will be created to represent different engine power and vehicle speed ranges in Figure 10-22 and are shown as follow:

Thresholds	Dummy21	Dummy22	Dummy23
engine.power < 82.625	1	0	0
engine.power [82.625, 152.96] & vehicle.speed < 19.05	0	1	0
engine.power ≥ 152.96 & vehicle.speed < 19.05	0	0	1
engine.power ≥ 82.625 & vehicle.speed ≥ 19.05	0	0	0

These three dummy variables and the interaction between dummy variables and engine power and vehicle speed are added to improve the model. This model will be:

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 \log_{10}(\text{engine.power}+1) + \beta_2\ \text{vehicle.speed} + \beta 3\text{dummy21} + \\ & \beta_4\ \text{dummy21}\ \log_{10}(\text{engine.power}+1) + \beta_5\ \text{dummy21}\ \text{vehicle.speed} + \beta_6\ \text{dummy22} + \\ & \beta_7\ \text{dummy22}\ \log_{10}(\text{engine.power}+1) + \beta_8\ \text{dummy22}\ \text{vehicle.speed} + \beta_9 \text{dummy23} + \\ & \beta_{10}\text{dummy23}\log_{10}(\text{engine.power}+1) + \beta_{11}\text{dummy23}\ \text{vehicle.speed} + \text{Error} \end{aligned} \quad (2.6)$$

The results for Mode. 2.6 are shown in Table 10-25 and Figure 10-35.

Table 10-25 Regression Result for CO Model 2.6

```
*** Linear Model ***
Call: lm(formula = log.CO ~ log10(engine.power + 1) + vehicle.speed + dummy21 *
       log10(engine.power + 1) + dummy21 * vehicle.speed + dummy22 * log10(
       engine.power + 1) + dummy22 * vehicle.speed + dummy23 * log10(
       engine.power + 1) + dummy23 * vehicle.speed, data =
       busdata10242006.1.3, na.action = na.exclude)
Residuals:
    Min      1Q   Median     3Q   Max
 -2.562 -0.2086 -0.02372 0.2012 2.124
Coefficients:                          Value Std. Error  t value Pr(>|t|)
                     (Intercept)  -3.5895   0.0945   -37.9720   0.0000
        log10(engine.power + 1)   1.1014   0.0389    28.3316   0.0000
                  vehicle.speed  -0.0150   0.0007   -21.0912   0.0000
                        dummy21   0.5978   0.1007     5.9384   0.0000
                        dummy22  -1.4856   0.2216    -6.7035   0.0000
                        dummy23  -2.3863   0.1632   -14.6202   0.0000
dummy21:log10(engine.power + 1)  -0.4617   0.0448   -10.3020   0.0000
          dummy21:vehicle.speed   0.0231   0.0014    16.8659   0.0000
dummy22:log10(engine.power + 1)   0.8643   0.1048     8.2494   0.0000
          dummy22:vehicle.speed  -0.0194   0.0016   -12.1421   0.0000
dummy23:log10(engine.power + 1)   1.3505   0.0701    19.2614   0.0000
          dummy23:vehicle.speed  -0.0387   0.0012   -30.9943   0.0000
Residual standard error: 0.3517 on 18852 degrees of freedom
Multiple R-Squared: 0.5609
F-statistic: 2189 on 11 and 18852 degrees of freedom, the p-value is 0
Analysis of Variance Table
Response: log.CO
Terms added sequentially (first to last)
                                  Df Sum of Sq  Mean Sq  F Value
        log10(engine.power + 1)    1  1346.515 1346.515 10887.89
                  vehicle.speed    1  1173.140 1173.140  9485.98
                        dummy21    1    23.180   23.180   187.44
                        dummy22    1    67.463   67.463   545.50
                        dummy23    1   100.345  100.345   811.39
dummy21:log10(engine.power + 1)    1    35.491   35.491   286.98
          dummy21:vehicle.speed    1    93.450   93.450   755.63
dummy22:log10(engine.power + 1)    1     3.681    3.681    29.76
          dummy22:vehicle.speed    1     3.564    3.564    28.82
dummy23:log10(engine.power + 1)    1    12.318   12.318    99.61
          dummy23:vehicle.speed    1   118.804  118.804   960.65
                      Residuals 18852  2331.445    0.124
                                        Pr(F)
        log10(engine.power + 1) 0.000000e+000
                  vehicle.speed 0.000000e+000
                        dummy21 0.000000e+000
                        dummy22 0.000000e+000
                        dummy23 0.000000e+000
dummy21:log10(engine.power + 1) 0.000000e+000
          dummy21:vehicle.speed 0.000000e+000
dummy22:log10(engine.power + 1) 4.942365e-008
          dummy22:vehicle.speed 8.032376e-008
dummy23:log10(engine.power + 1) 0.000000e+000
          dummy23:vehicle.speed 0.000000e+000
                      Residuals
```

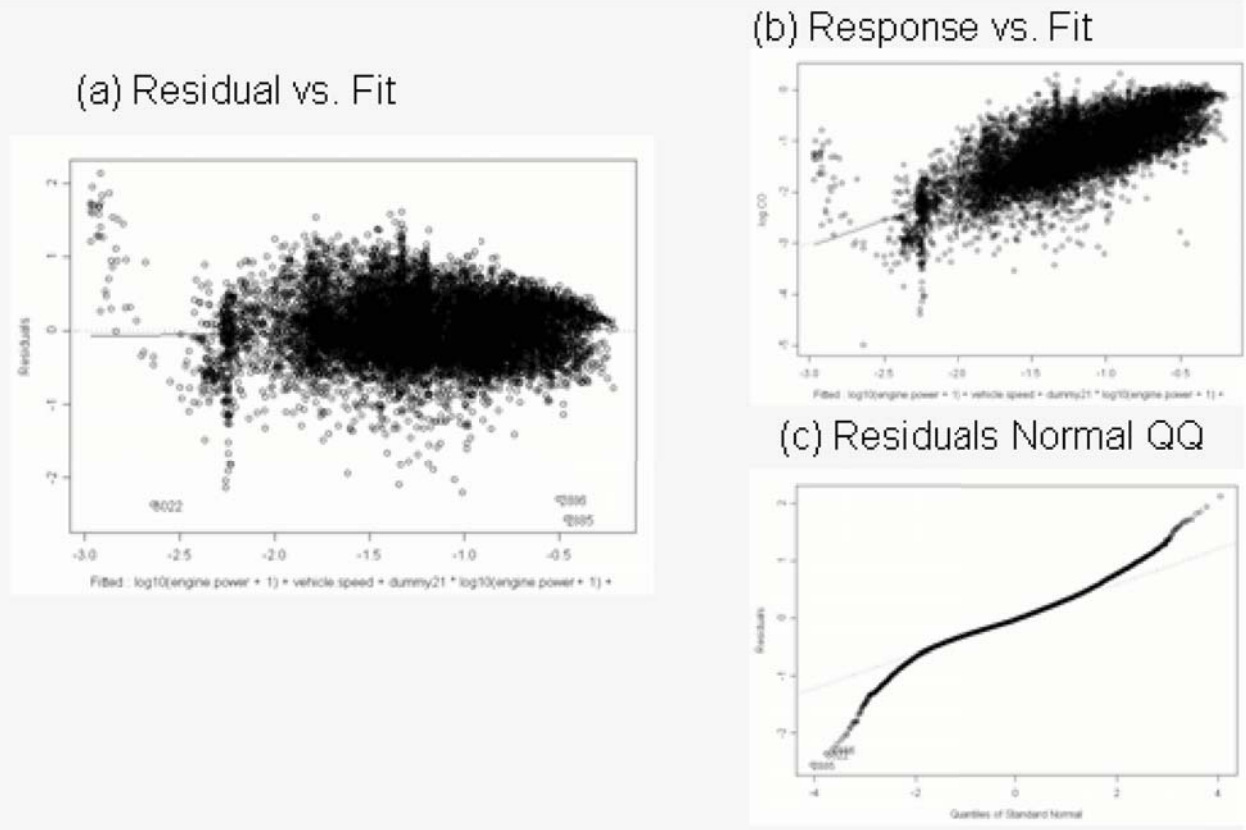


Figure 10-38 QQ and Residual vs. Fitted Plot for CO Model 2.6

Model 2.6 does improve the ability to explain variance by another 4% (R^2 increases from from 0.47 to 0.52 and then to 0.56 by adding the dummy variables). Model 2.6 slightly improves R^2 compared to Model 2.5. The residuals scatter plot for Model 2.6 (Figure 10-38) shows a more reasonably linear relation. Figure 10-38 also shows that Model 2.6 does a better job in improving the pattern of variance. The QQ plot again shows general normality with the exceptions arising in the tails. However, it is important to note that the model improvement, in terms of amount of variance explained by the model, is small.

10.3.2.2.4 Model Discussions

The previous sections outline the model development process from the regression tree model, to a simple OLS model, to more complex OLS models. Since the performance of the models is evaluated by comparing model predictions and actual observations for emission rates, the R^2 and slope are different from those in previous linear regression models. The results of each step in the model improvement process are presented in Table 10-26. The mean emission rates at HTBR end nodes (single value) are compared to the results of various linear regression functions with engine power. Since the R^2 values and slopes in Table 10-26 are derived by

comparing model predictions and actual observations for emission rates (untransformed y), these numbers will be different from those obtained from linear regression models.

Table 10-26 Comparative Performance Evaluation of CO Emission Rate Models

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
Mean ERs	0.00003	1.000	0.16032	-0.00002
Linear Regression (Power)	0.0462	1.180	0.16516	0.05229
Linear Regression ($Power^{0.5}$)	0.0502	1.227	0.16420	0.05006
Linear Regression (log(Power))	0.0553	1.534	0.16455	0.05120
Linear Regression (log(Power)+Speed)	0.392	2.161	0.14252	0.04211
Linear Regression (log(Power)+Speed+Dummy Set 1)	0.406	1.765	0.13632	0.03689
Linear Regression (log(Power)+Speed+Dummy Set 2)	0.437	1.242	0.12565	0.03003

The improvement in R^2 associated with moving toward a linear function of engine power is significant. Hence, the use of the linear regression function will provide a significant improvement on spatial and temporal model prediction capability. However this linear regression function might still be improved.

Results suggest that a linear regression function with log transformation performs slightly better than the others and that the use of dummy variables can further improve model performance. Although the linear regression function with dummy variables performs slightly better than the linear regression function with log transformation, the introduction of more explanatory variables (dummy variables and the interaction with engine power) increases the complexity of the regression model. As discussed in Section 10.3.2.1.4, there is no compelling reason to include the dummy variables in the model since: 1) the models with dummy variables are more complex without significantly improving model performance, and 2) there is no compelling engineering reason at this time to support the difference in model performance within these specific power regions. Yet, given the explanatory power of the power cutpoint dummy variables (a 10% increase in explained variance), additional investigation into why these values are turning out to be significant is definitely warranted. It may be wise to include such cutpoints in on-road models for various engine technology groups. Such dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis.

It can be argued that inclusion of the dummy variables for power is warranted. However, Model 2.4 is chosen as the preliminary 'final' model based solely upon ease of implementation. The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform

to a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have uniform variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying patterns in the residuals. Figure 10-36 plot (a) shows this plot for CO Model 2.4. Without considering variance due to high emission points and zero load data, there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error Terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 10-36 plot (c) shows the normal quantile plot of CO Model 2.4. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm \sqrt{MSE}$ and about 90 percent to fall between $\pm 1.645 \sqrt{MSE}$. Actually, 87.35% of residuals fall within the first limits, while 92.19% of residuals fall within the second limits. Thus, the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but are due to the nature of the data set. For example, even after the transformation, the response variable is not the real normal distribution.

Based on above analysis, final CO emission model for cruise mode is:

$$CO = 10^{[-3.747+1.341\log10(engine.power+1) - 0.0285vehicle.speed]}$$

Analysis results support the observation that the final CO emission model (2.4) is significantly better at explaining variability without making the model too complex. Since there is only one engine type, complexity may not be valid in terms of transferability. This model is specific to the engine classes employed in the transit bus operations. Different models may need to be developed for other engine classes and duty cycles.

10.3.2.3 HC Emission Rate Model Development for Acceleration Mode

Based on previous analysis, truncated transformed HC will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing statistical models. HTBR tree model results suggest that engine power is the best one to begin with.

10.3.2.3.1 Linear Regression with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1 engine.power + Error \qquad (3.1)$$

The regression run yields the results shown in Table 10-27 and Figure 10-39.

Table 10-27 Regression Result for HC Model 3.1

```
Call: lm(formula = HC.25 ~ engine.power, data = busdata10242006.1.3, na.action =
na.exclude)
Residuals:
     Min       1Q      Median      3Q    Max
 -0.1285 -0.02417 -0.00003173 0.02467 0.2904

Coefficients:
               Value Std. Error  t value Pr(>|t|)
 (Intercept)   0.1840   0.0009   216.4203   0.0000
engine.power   0.0001   0.0000    32.4947   0.0000

Residual standard error: 0.03989 on 18328 degrees of freedom
Multiple R-Squared: 0.05447
F-statistic: 1056 on 1 and 18328 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.938

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1   1.67991 1.679912 1055.908     0
   Residuals 18328  29.15918 0.001591
```

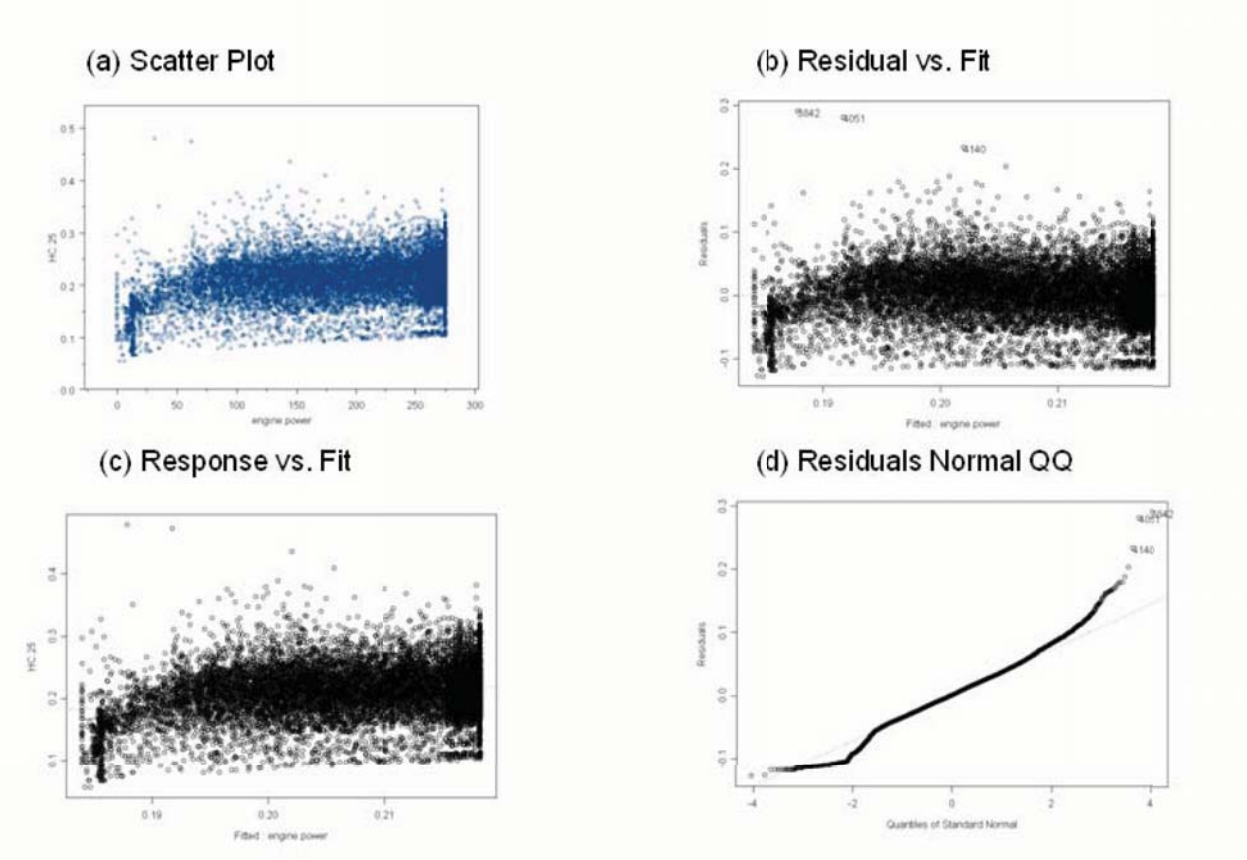


Figure 10-39 QQ and Residual vs. Fitted Plot for HC Model 3.1

The results suggest that engine power explains about 5% of the variance in truncated transformed HC. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, the normality is examined in the QQ plot and constancy of variance is checked by examining residuals vs. fitted values.

The residual plot in Figure 10-39 shows a slight departure from linear regression assumptions indicating a need to explore a curvilinear regression function. Since the variability at the different X levels appears to be fairly constant, a transformation on X is considered. The reason to consider transformation first is to avoid multicollinearity brought about by adding the second-order of X. Based on the prototype plot in Figure 10-39, the square root transformation and logarithmic transformation are tested. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to determine which transformation is most effective.

$$Y = \beta_0 + \beta_1 engine.power^{(1/2)} + Error \qquad (3.2)$$

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + Error \qquad (3.3)$$

The result for Model 3.2 is shown in Table 10-28 and Figure 10-40, while the result for Model 3.3 is shown in Table 10-29 and Figure 10-41.

Table 10-28 Regression Result for HC Model 3.2

```
Call: lm(formula = HC.25 ~ engine.power^(1/2), data = busdata10242006.1.3, na.action
= na.exclude)
Residuals:
     Min       1Q     Median     3Q    Max
 -0.1173 -0.02389 -0.0002473 0.0244 0.2969

Coefficients:
                         Value Std. Error  t value Pr(>|t|)
          (Intercept)   0.1625   0.0013   127.4341   0.0000
I(engine.power^(1/2))   0.0034   0.0001    38.2005   0.0000

Residual standard error: 0.03948 on 18328 degrees of freedom
Multiple R-Squared: 0.07375
F-statistic: 1459 on 1 and 18328 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                       (Intercept)
I(engine.power^(1/2)) -0.9735

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                         Df Sum of Sq  Mean Sq F Value Pr(F)
I(engine.power^(1/2))     1   2.27433 2.274333 1459.28     0
            Residuals 18328  28.56475 0.001559
```

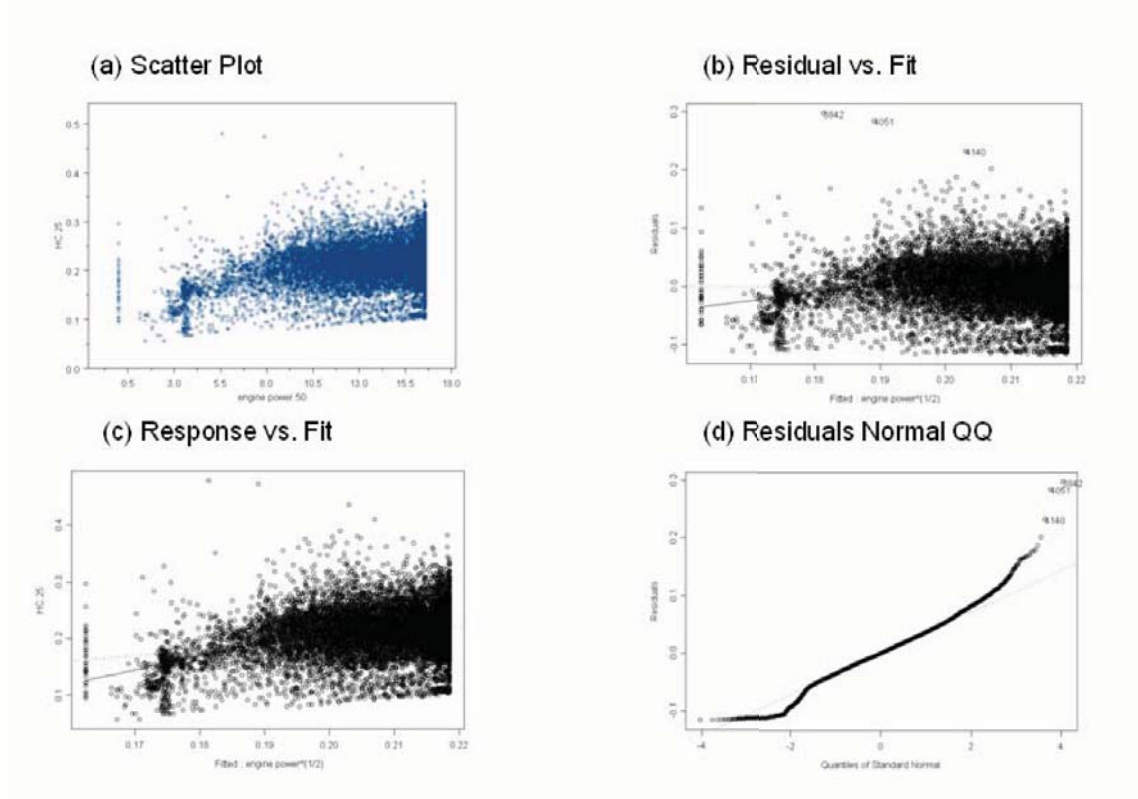


Figure 10-40 QQ and Residual vs. Fitted Plot for HC Model 3.2

Table 10-29 Regression Result for HC Model 3.3

```
Call: lm(formula = HC.25 ~ log10(engine.power + 1), data = busdata10242006.1.3,
na.action = na.exclude)
Residuals:
     Min       1Q      Median      3Q     Max
 -0.1186 -0.02345 -0.00007336 0.02386 0.3004

Coefficients:
                          Value Std. Error t value Pr(>|t|)
            (Intercept)  0.1136  0.0022    50.8911  0.0000
log10(engine.power + 1)  0.0426  0.0010    43.4726  0.0000

Residual standard error: 0.03906 on 18328 degrees of freedom
Multiple R-Squared: 0.09347
F-statistic: 1890 on 1 and 18328 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                         (Intercept)
log10(engine.power + 1) -0.9916

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1   2.88268 2.882681 1889.863     0
              Residuals 18328  27.95641 0.001525
```

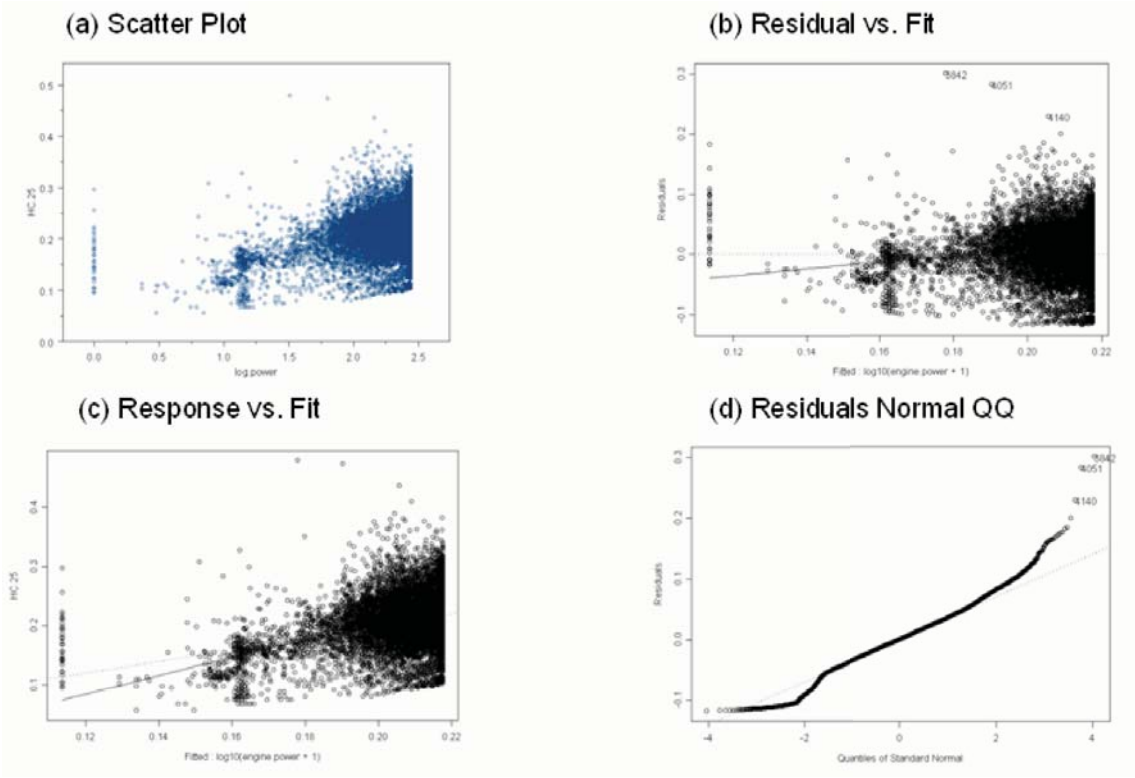


Figure 10-41 QQ and Residual vs. Fitted Plot for HC Model 3.3

The results suggest that by using transformed engine power, the model increases the amount of variance explained in truncated transformed HC from about 5% to about 9%.

Model 3.3 improves R^2 relative to Model 3.2. The residuals scatter plot for Model 3.3 (Figure 10-41) also shows a more reasonably linear relation than Model 2.2 (Figure 10-40). Figure 10-41 also shows that Model 3.3 does a better job in improving the pattern of variance. QQ plot shows general normality with the exceptions arising in the tails.

10.3.2.3.2 Linear Regression Model with Dummy Variables

Figure 10-26 suggests that the relationship between HC and engine power may differ across the engine power ranges. One dummy variable is created to represent different engine power ranges identified in Figure 10-26 for use in linear regression analysis as illustrated below:

Engine power (bhp)	Dummy1
< 54.555	1
≥ 54.555	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variable and interaction can help improve the model.

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + \beta_2\, dummy1 + \beta_3 dummy1\ log_{10}(engine.power+1) + Error \quad (3.4)$$

The results for Model 3.4 are shown in Table 10-30 and Figure 10-42.

Table 10-30 Regression Result for HC Model 3.4

```
Call: lm(formula = HC.25 ~ log10(engine.power + 1) + dummy1 * log10(engine.power +
1), data = busdata10242006.1.3, na.action = na.exclude)
Residuals:
     Min       1Q    Median     3Q   Max
 -0.1278 -0.02305 0.0002278 0.0231 0.314

Coefficients:
                                  Value Std. Error  t value Pr(>|t|)
                    (Intercept)   0.1734   0.0042    41.4191   0.0000
       log10(engine.power + 1)   0.0171   0.0018     9.4715   0.0000
                        dummy1  -0.0643   0.0062   -10.3151   0.0000
dummy1:log10(engine.power + 1)   0.0195   0.0039     4.9731   0.0000

Residual standard error: 0.03873 on 18326 degrees of freedom
Multiple R-Squared: 0.1084
F-statistic: 742.8 on 3 and 18326 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                                   Df Sum of Sq  Mean Sq  F Value
       log10(engine.power + 1)      1   2.88268 2.882681 1921.331
                        dummy1      1   0.42377 0.423774  282.449
dummy1:log10(engine.power + 1)      1   0.03711 0.037107   24.732
                     Residuals  18326  27.49553 0.001500

                                      Pr(F)
       log10(engine.power + 1) 0.000000e+000
                        dummy1 0.000000e+000
dummy1:log10(engine.power + 1) 6.647205e-007
                     Residuals
```

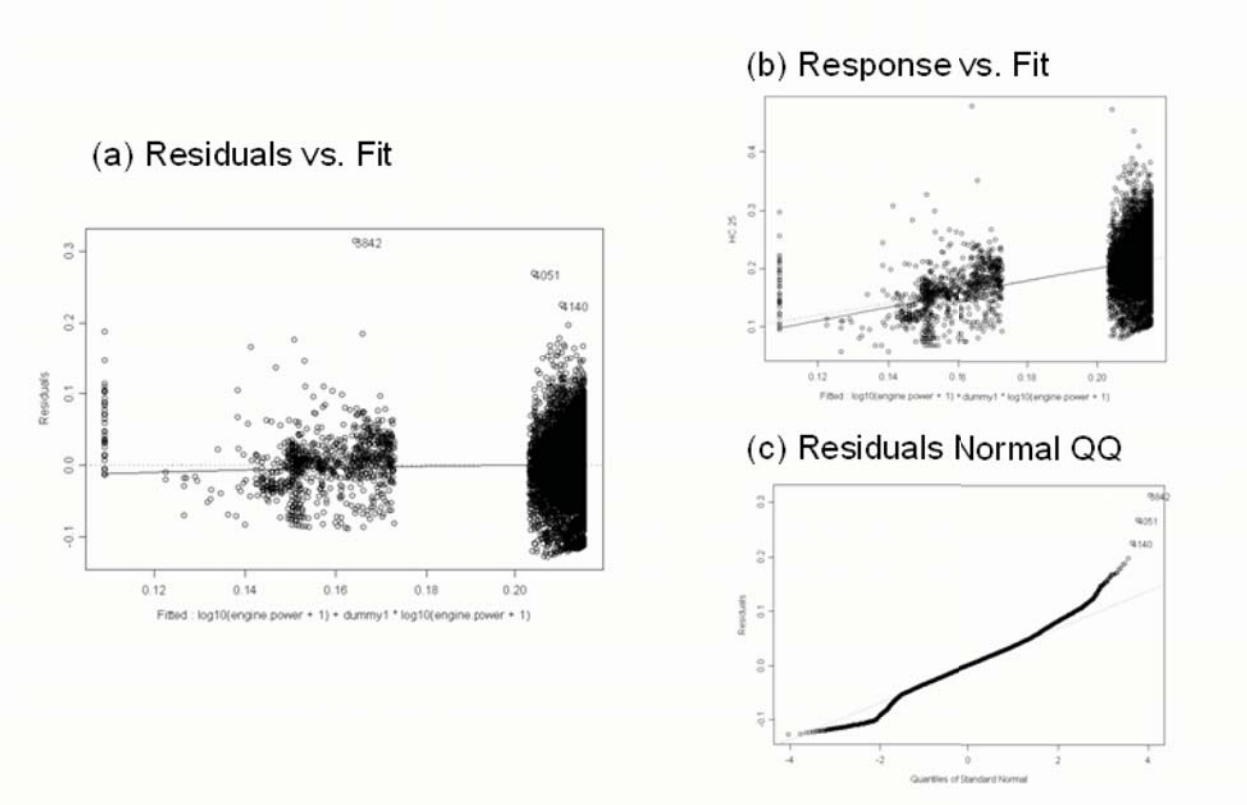


Figure 10-42 QQ and Residual vs. Fitted Plot for HC Model 3.4

The results suggest that by using transformed engine power and speed, the model only increases the amount of variance explained in truncated transformed HC from about 9% to about 10%.

Model 3.4 slightly improves R^2 relative to Model 3.3. The residuals scatter plot for Model 3.4 (Figure 10-42) is not appreciably better nor does Model 3.4 do a better job in improving the pattern of variance. The QQ plot still shows general normality with the exceptions arising in the tails.

10.3.2.3.3 Model Discussions

The previous sections outline the model development process from regression tree model, to a simple OLS model, to more complex OLS models. To test whether the linear regression with power was a beneficial addition to the regression tree model, the mean ERs at HTBR end nodes (single value) were compared to the predictions from the linear regression function with engine power. The results of the performance evaluation are shown in Table 10-31. The improvement in R^2 associated with moving toward a linear function of engine power is nearly imperceptible. Hence, the use of the linear regression function will provide almost no significant improvement over spatial and temporal model prediction capability. This linear regression function might still be improved. Since the R^2 and slope in Table 10-31 are derived by compar-

ing model predictions and actual observations for emission rates, these numbers will be different from those obtained from linear regression models.

Table 10-31 Comparative Performance Evaluation of HC Emission Rate Models

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
Mean ERs	0.000090	1.000	0.0019072	0.00000022
Linear Regression (Power)	0.0166	0.979	0.0019879	0.00061206
Linear Regression ($Power^{0.5}$)	0.0214	0.749	0.0019311	0.00040055
Linear Regression (log(Power))	0.0281	0.864	0.0019249	0.00040884
Linear Regression (log(Power) + Dummy)	0.0367	1.060	0.0019151	0.00040366

Results suggest that the linear regression function with log transformation performs slightly better than the others and that the use of dummy variables can further improve model performance, but again there is almost no perceptible change in terms of explained variance. Although the linear regression function with log transformation and dummy variables performs slightly better than linear regression function with log transformation alone, the revised model introduces additional explanatory variables (dummy variables and the interaction with engine power) and increases the complexity of regression model without significantly improving the model. As discussed in Section 10.3.2.1.4, there is no compelling reason to include the dummy variables in the model, given that: 1) the second model is more complex without significantly improving model performance, and 2) there is no compelling engineering reason at this time to support the difference in model performance within these specific power regions. These dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis.

Model 3.3 is recommended as the preliminary 'final' model (although one might argue that using the regression tree results directly would also probably be acceptable). The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform to a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have the same variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying any patterns in the residuals. Figure 10-41 plot (b) is residuals vs. fit for HC Model 3.3. Without considering variance due to high emission points and zero load data, it can be seen that there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 10-40 plot (d) shows the normal quantile plot of of HC Model 3.2. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm\sqrt{MSE}$ and about 90 percent to fall between $\pm$ 1.645 $\sqrt{MSE}$. Actually, 84.83% of residuals fall within the first limits, while 93.60% of residuals fall within the second limits. Thus, the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but this is due to the nature of the data set. For example, even after the transformation, the response variable is not the real normal distribution.

Based on above analysis, final HC emission model for cruise mode is:

$$HC = [0.114 + 0.0426\log_{10}(\text{engine.power}+1)]^4$$

10.4 Conclusions and Further Considerations

In this research, acceleration mode is defined as “acceleration >1 mph/s”. Data not considered to be in idle, deceleration or acceleration mode will be deemed to be in cruise mode. Compared to cruise mode activity, the engine power is more concentrated in higher engine power ranges ($\geq$ 200 bhp) for acceleration mode activity.

Inter-bus variability analysis indicated that some of the 15 buses are higher emitters than others (especially noted for HC emissions). However, none of the buses appears to qualify as a traditional high-emitter, which would exhibit emission rates of two to three standard deviations above the mean. Hence, it is difficult to classify any of these 15 buses as high emitters for modeling purposes. At this moment, these 15 buses are treated as a whole for model development. Modelers should keep in mind that although no true high-emitters are present in the database, such vehicles may behave significantly different than the vehicles tested. Hence, data from high-emitting vehicles should be collected and examined in future studies.

Some high HC emissions events are noted in acceleration mode. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, no variables were identified that could be linked to these high emissions events. These events may represent natural variability in on-road emissions, or

some other variable (such as grade or an engine variable that is not measured) may be linked to these events.

Engine power is selected as the most important variable for three pollutants based on HTBR tree models. This finding is consistent with previous research results which verified the important role of engine power (Ramamurthy et al. 1998; Clark et al. 2002; Barth et al. 2004). The HC relationship is significant but fairly weak. Analysis in previous chapters also indicates that engine power is correlated with not only on-road load parameters such as vehicle speed, acceleration, and grade, but also potentially with engine operating parameters such as throttle position and engine oil pressure. On the other hand, engine power in this research is derived from engine speed, engine torque and percent engine load.

The regression tree models suggest that some other variables, like oil pressure and engine barometric pressure, may also impact the HC emissions. Further analysis demonstrates that by using engine power alone one might be able to achieve explanatory ability similar to using engine power and other variables. To develop models that are efficient and easy to implement, only engine power is used to develop emission models. However, additional investigation into these variables is warranted as additional detailed data from engine testing become available for analysis.

Given the relationships noted between engine indicated HP and emission rates, it is imperative that data be collected to develop solid relationships in engine power demand models (estimating power demand as a function of speed/acceleration, grade, vehicle characteristics, surface roughness, inertial losses, etc.) for use in regional inventory development and microscale impact assessment.

In summary, the modeler recommends the following acceleration emission models:

$$NO_x = [-0.0195 + 0.201\log_{10}(\text{engine.power}+1) + 0.0019\text{vehicle.speed}]^2$$

$$CO = 10^{[-3.747 + 1.341\log10(\text{engine.power}+1) - 0.0285\text{vehicle.speed}]}$$

$$HC = [0.114 + 0.0426\log_{10}(\text{engine.power}+1)]^4$$

CHAPTER 11

11. CRUISE MODE DEVELOPMENT

After developing idle mode definition and emission rate in Chapter 8, deceleration mode definition and emission rate in Chapter 9, and acceleration emission model in Chapter 10, the next task will be to develop cruise mode.

11.1 Analysis of Cruise Mode Data

After dividing the database into idle mode, deceleration mode, and acceleration mode, cruise mode data will be all of the remaining data in the database (i.e., data not previously classified into idle, deceleration, and acceleration). Unlike the idle and deceleration modes, there is a general relationship between engine power and emission rate for acceleration mode and cruise mode. The engine power distribution for data collected in the cruise mode is provided in Table 11-1.

Table 11-1 Engine Power Distribution for Cruise Mode

	Pollutants	Engine Power Distribution					
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	All
Number	NO_x	15885	8988	7173	3536	3792	39374
	CO	15834	8940	7145	3529	3770	39218
	HC	15481	8600	6830	3394	3715	38020
Percentage	NO_x	40.34%	22.83%	18.22%	8.98%	9.63%	100.00%
	CO	40.37%	22.80%	18.22%	9.00%	9.61%	100.00%
	HC	40.72%	22.62%	17.96%	8.93%	9.77%	100.00%

Emission rate histograms for each of the three pollutants for cruise operations are presented in Figure 11-1. Figure 11-1 shows significant skewness for all three pollutants for cruise mode. Some high HC emissions events are noted in cruise mode. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM

pressure, and other parameters, no operating parameters appeared to correlate with the high emissions events.

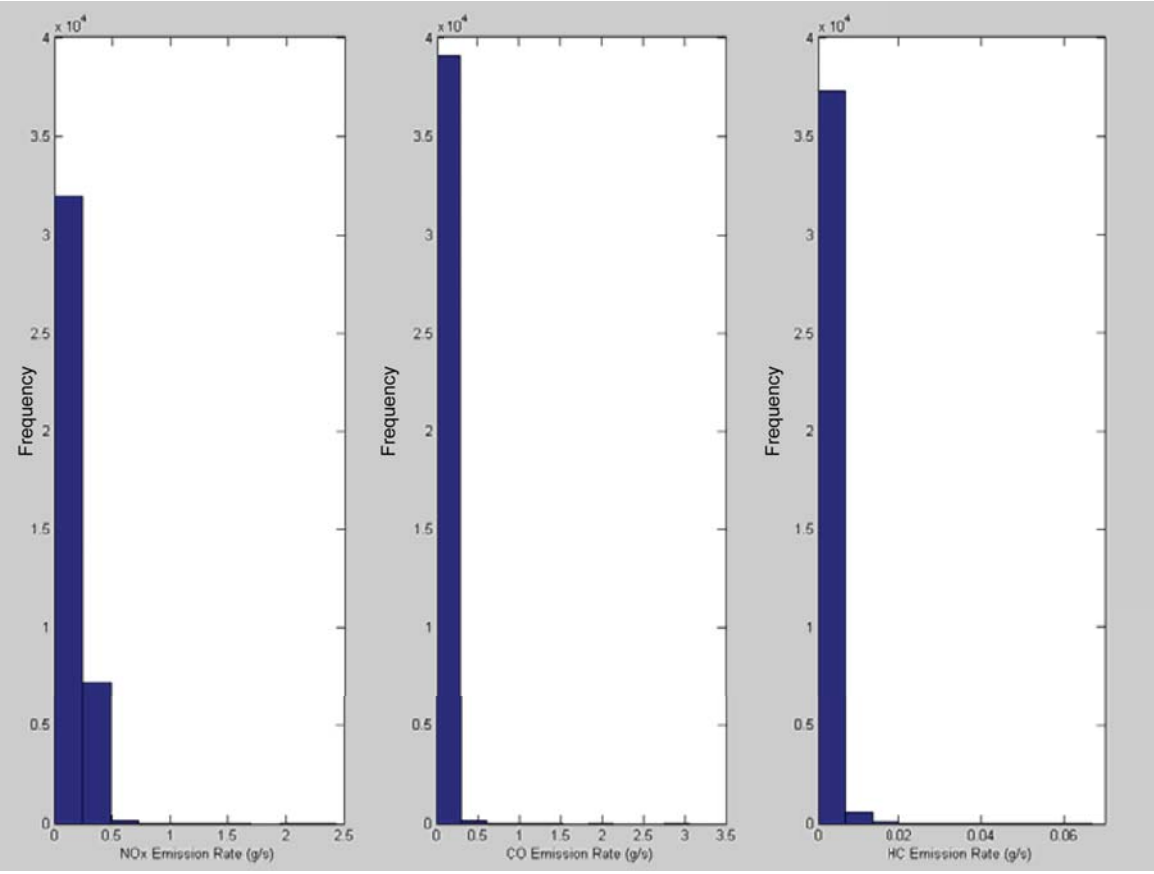


Figure 11-1 Histograms of Three Pollutants for Cruise Mode

11.1.1 Engine Rate Distribution by Bus in Cruise Mode

Inter-bus response variability for cruise mode operations is illustrated in Figures 11-2 to 11-4 using median and mean of NO_x, CO, and HC emission rates. Table 11-2 presents the same information in tabular form. The difference between median and mean is also an indicator of skewness.

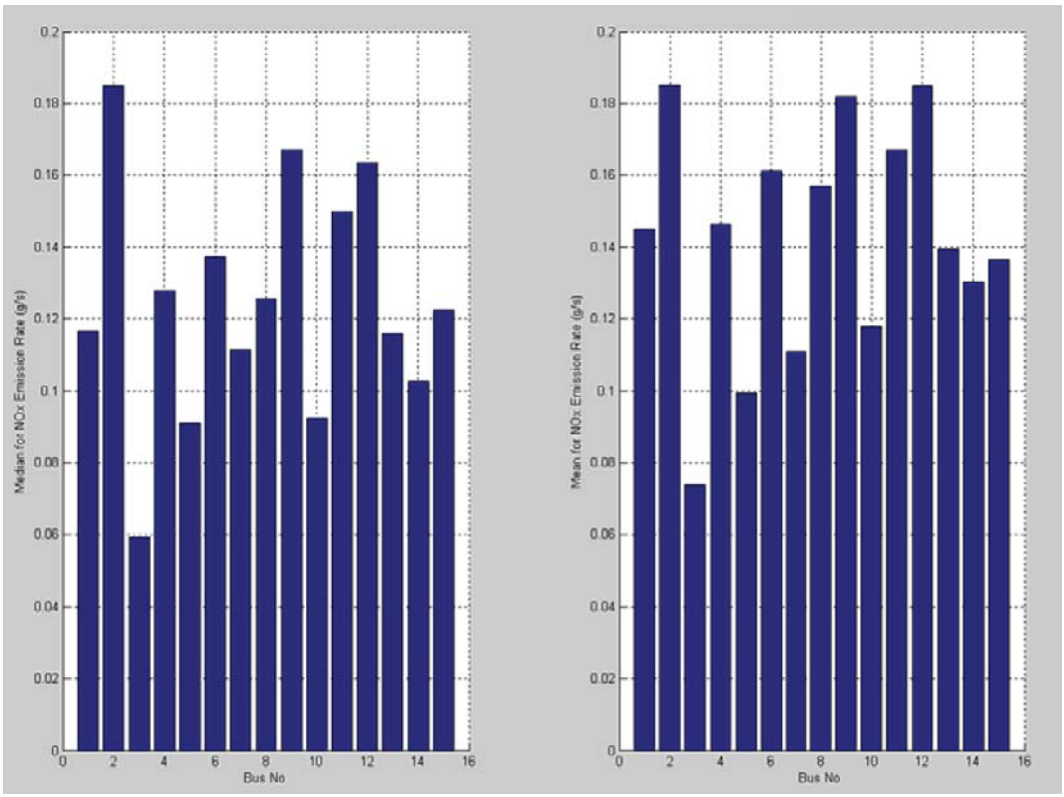


Figure 11-2 Median and Mean of NO_x Emission Rates in Cruise Mode by Bus

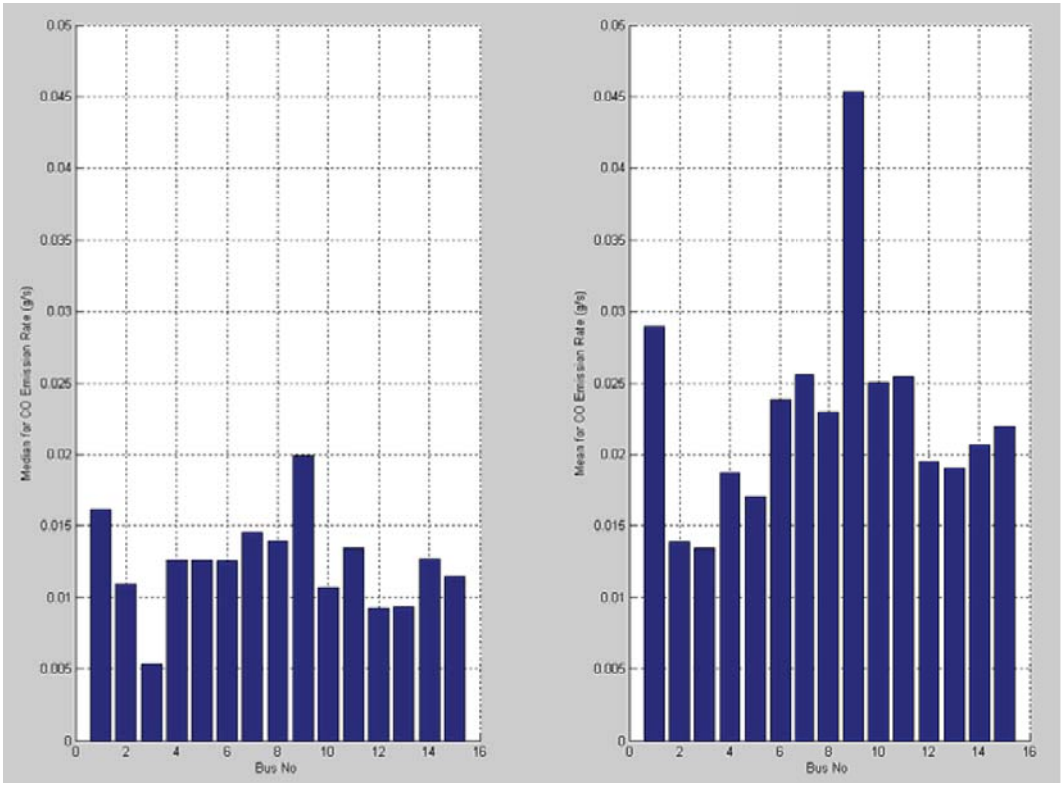


Figure 11-3 Median and Mean of CO Emission Rates in Cruise Mode by Bus

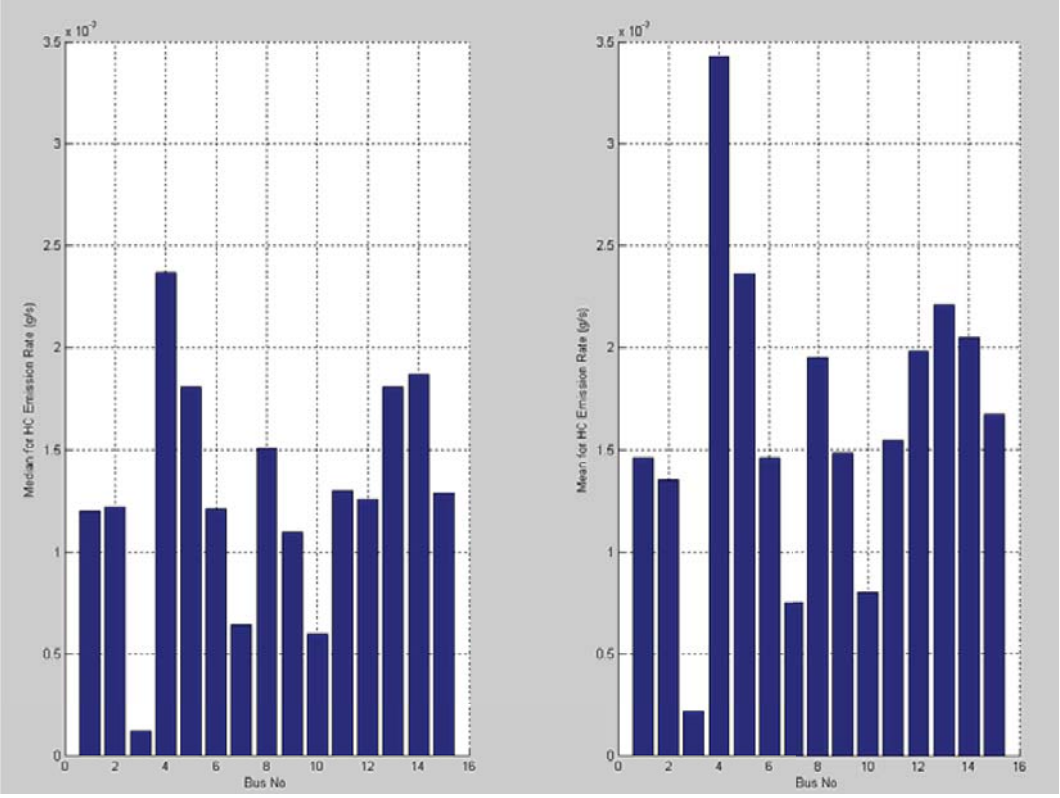


Figure 11-4 Median and Mean of HC Emission Rates in Cruise Mode by Bus

Table 11-2 Median and Mean of Three Pollutants in Cruise Mode by Bus

	NOx		CO		HC	
Bus ID	**Median**	**Mean**	**Median**	**Mean**	**Median**	**Mean**
Bus 360	0.11666	0.14506	0.01618	0.02891	0.00120	0.00146
Bus 361	0.18479	0.18507	0.01091	0.01389	0.00122	0.00135
Bus 363	0.05924	0.07384	0.00534	0.01341	0.00012	0.00021
Bus 364	0.12779	0.14644	0.01259	0.01875	0.00237	0.00343
Bus 372	0.09092	0.09936	0.01262	0.01704	0.00181	0.00236
Bus 375	0.13714	0.16103	0.01254	0.02383	0.00121	0.00146
Bus 377	0.11139	0.11094	0.01454	0.02559	0.00064	0.00075
Bus 379	0.12570	0.15673	0.01394	0.02298	0.00151	0.00195
Bus 380	0.16713	0.18183	0.01994	0.04532	0.00110	0.00148
Bus 381	0.09227	0.11789	0.01074	0.02505	0.00060	0.00080
Bus 382	0.14987	0.16698	0.01342	0.02544	0.00130	0.00155
Bus 383	0.16355	0.18468	0.00921	0.01949	0.00126	0.00198
Bus 384	0.11597	0.13933	0.00934	0.01903	0.00181	0.00221
Bus 385	0.10244	0.13024	0.01266	0.02066	0.00187	0.00205
Bus 386	0.12254	0.13632	0.01147	0.02197	0.00129	0.00167

Figures 11-2 to 11-4 and Table 11-2 illustrate that NO_x emissions are more consistent than CO and HC emissions. Across the 15 buses, Bus 380 has the largest median and mean for CO emissions, while Bus 364 has the largest median and mean for HC emissions. The above figures and table demonstrate that although variability exists across buses, it is difficult to conclude that there are any true "high emitters" in the database. This conclusion is consistent with the result for the other three modes. As was also noted in the acceleration mode data, Bus 363 has the smallest mean and median HC emissions compared to the other 14 buses.

11.1.2 Engine Power Distribution by Bus in Cruise Mode

Engine power distribution in cruise mode by bus is shown in Figure 11-5 and Table 11-3. Bus 361 has the largest 1st quartile engine power in cruise mode while Bus 377 has the largest median and 3rd quartile engine power in cruise mode. The maximum power values for each bus match well with the manufacturer's engine power rating. Although variability for engine power distribution exists across buses, it is difficult to conclude that such variability is affected by individual buses, bus routes, or other factors. The relationship between power and emissions appears consistent across the buses for acceleration mode.

Table 11-3 Engine Power Distribution in Cruise Mode by Bus

Bus ID	Number	Min	1st Quartile	Median	3rd Quartile	Max	Mean
Bus 360	1653	0	14.68	71.25	169.03	275.46	97.70
Bus 361	3140	0	70.13	108.12	140.28	296.91	107.16
Bus 363	3286	0	10.46	47.19	112.37	275.55	71.45
Bus 364	2575	0	14.47	64.30	130.62	275.51	85.56
Bus 372	2278	0	30.13	68.23	118.10	275.49	79.77
Bus 375	2890	0	23.19	72.09	142.47	275.54	94.36
Bus 377	1647	0	17.93	118.01	210.27	275.50	121.33
Bus 379	2544	0	43.51	102.68	165.04	275.57	110.84
Bus 380	1242	0	18.85	91.07	187.71	275.56	109.41
Bus 381	2537	0	6.72	49.18	113.81	275.46	70.68
Bus 382	1208	0	32.39	81.02	124.97	275.55	89.42
Bus 383	3062	0	29.42	77.95	141.19	275.53	90.85
Bus 384	3638	0	21.82	61.20	115.75	275.46	72.69
Bus 385	3327	0	11.86	48.80	102.91	275.47	68.20
Bus 386	4539	0	19.24	53.43	94.38	275.30	61.66

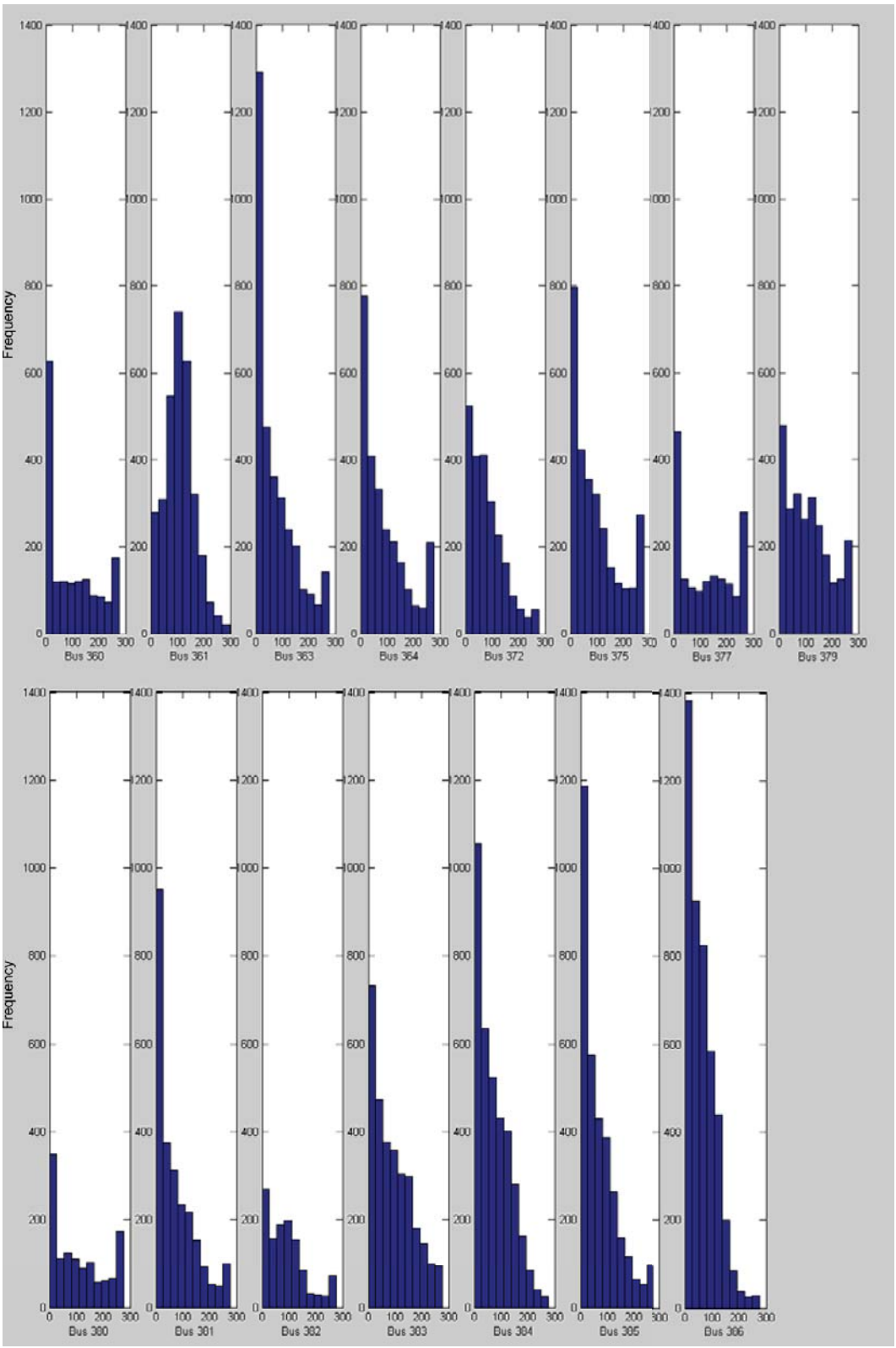


Figure 11-5 Histograms of Engine Power in Cruise Mode by Bus

11.2 Model Development and Refinement

11.2.1 HTBR Tree Model Development

The potential explanatory variables included in the emission rate model development effort include:

Vehicle characteristics: model year, odometer reading, bus ID (14 dummy variables)

Roadway characteristics: dummy variable for road grade;

Onroad load parameters: engine power (bhp), vehicle speed (mph), acceleration (mph/s);

Engine operating parameters: engine oil temperature (deg F), engine oil pressure (kPa), engine coolant temperature (deg F), barometric pressure reported from ECM (kPa);

Environmental conditions: ambient temperature (deg C), ambient pressure (mbar), ambient relative humidity (%).

HTBR technique is used first to identify potentially significant explanatory variables and this analysis provides the starting point for conceptual model development. The HTBR model is used to guide the development of an OLS regression model, rather than as a model in its own right. HTBR can be used as a data reduction tool and for identifying potential interactions among the variables. Then OLS regression is used with the identified variables to estimate a preliminary "final" model.

Although evidence in the literature suggests that a logarithmic transformation is most suitable for modeling motor vehicle emissions (Washington 1994; Ramamurthy et al. 1998; Fomunung 2000; Frey et al. 2002), this transformation needs to be verified through the Box-Cox procedure. The Box-Cox function in MATLAB™ can automatically identify a transformation from the family of power transformations on emission data, ranging from -1.0 to 1.0. The lambdas chosen by the Box-Cox procedure for cruise mode are 0.40619 for NO_x, 0.012969 for CO, 0.241 for HC. The Box-Cox procedure is only used to provide a guide for selecting a transformation, so overly precise results are not needed (Neter et al. 1996). It is often reasonable to use a nearby lambda value that is easier to understand for the power transformation. Although the lambdas chosen by the Box-Cox procedure are different for acceleration and cruise modes, the nearby lambda values are same for these two modes. In summary, the lambda values used for transformations are ½ for NO_x, 0 for CO (indicating a log transformation), and ¼ for HC for cruise mode. Figures 11-6 to 11-8 present the histogram, boxplot, and probability plots of truncated emission rates in cruise mode for NO_x, CO, and HC, while Figures 11-9 to 11-11 pres-

ent the same plots for truncated transformed emission rates for NO_x, CO and HC, where a great improvement is noted.

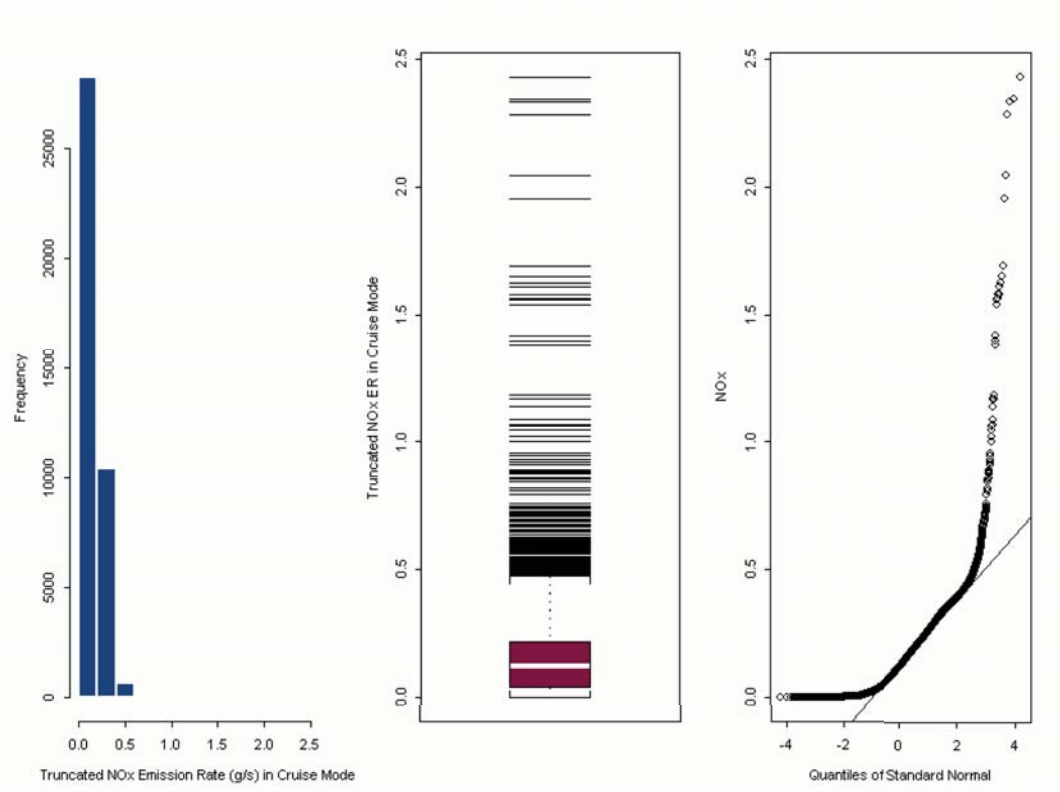


Figure 11-6 Histogram, Boxplot, and Probability Plot of Truncated NO_x Emission Rates in Cruise Mode

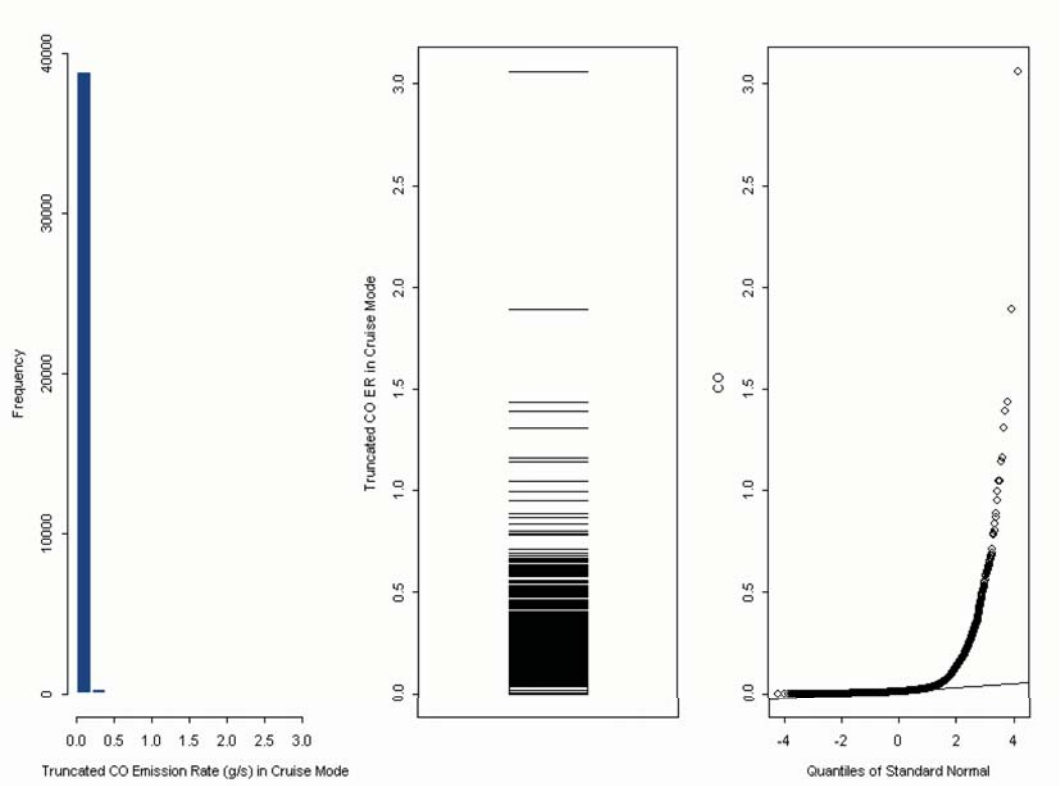


Figure 11-7 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Cruise Mode

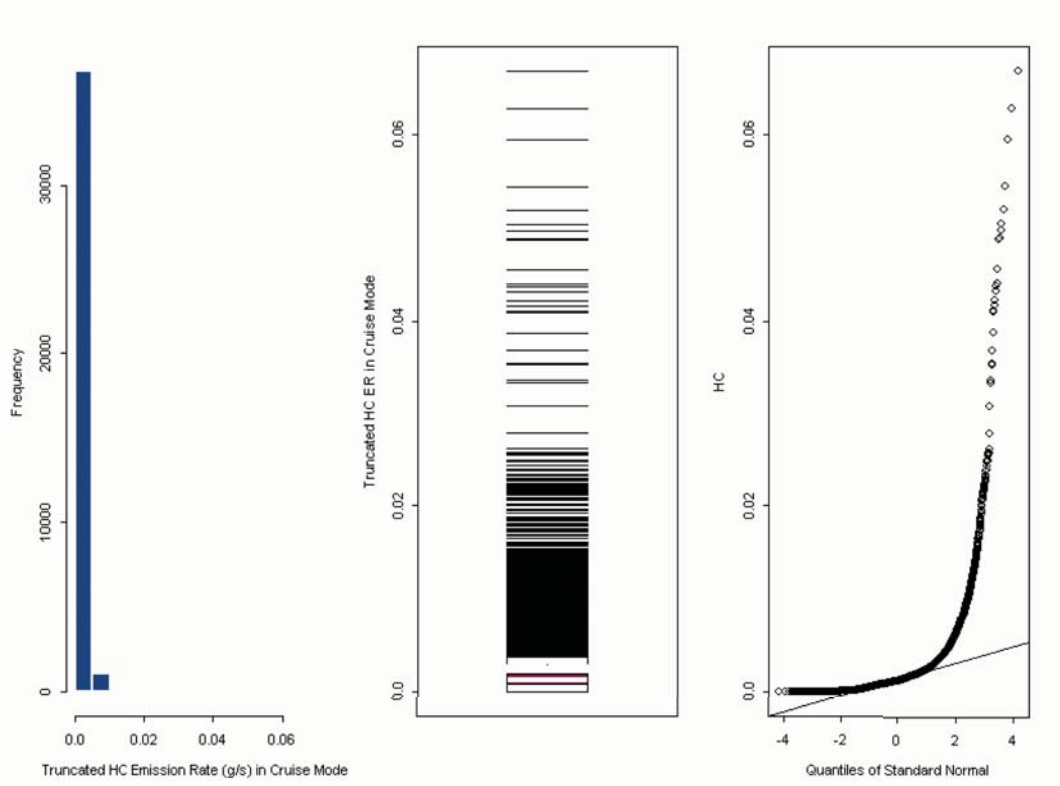


Figure 11-8 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Cruise Mode

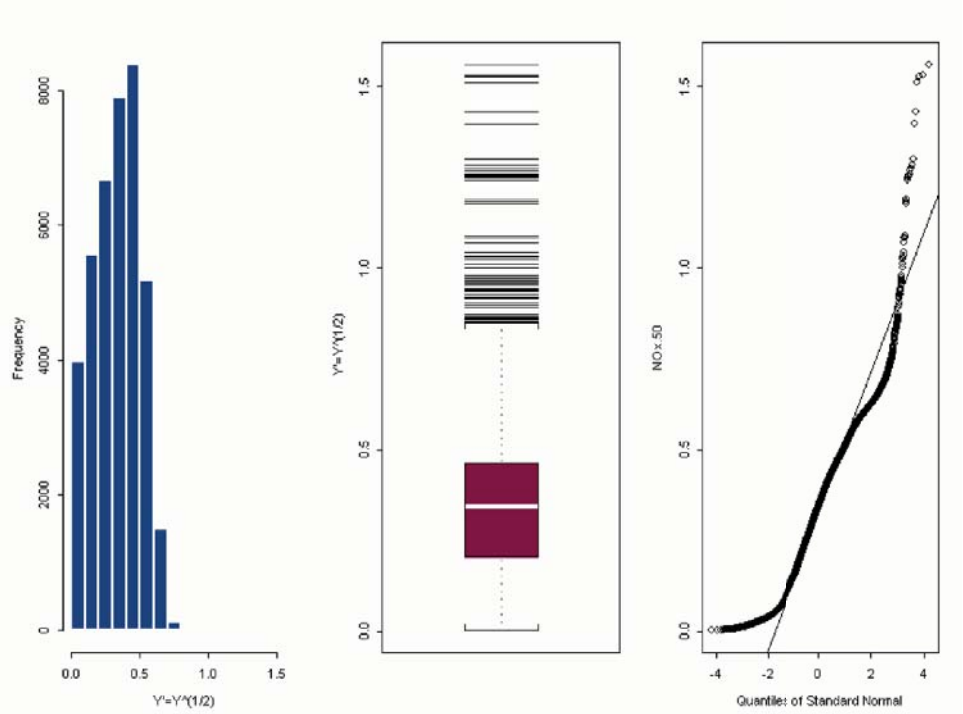


Figure 11-9 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO_x Emission Rate in Cruise Mode

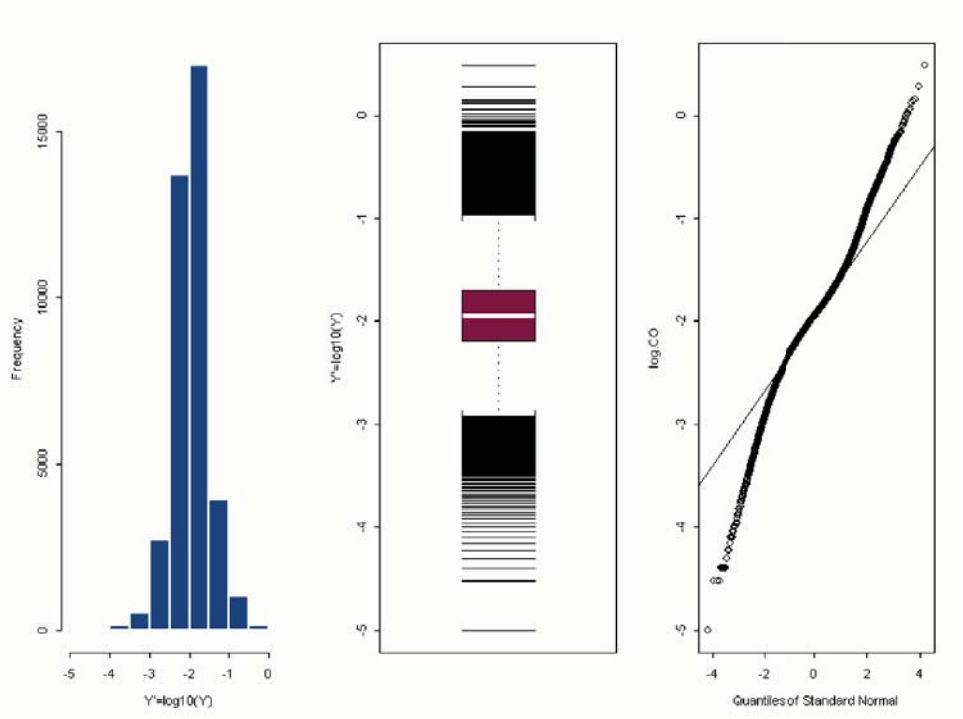


Figure 11-10 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate in Cruise Mode

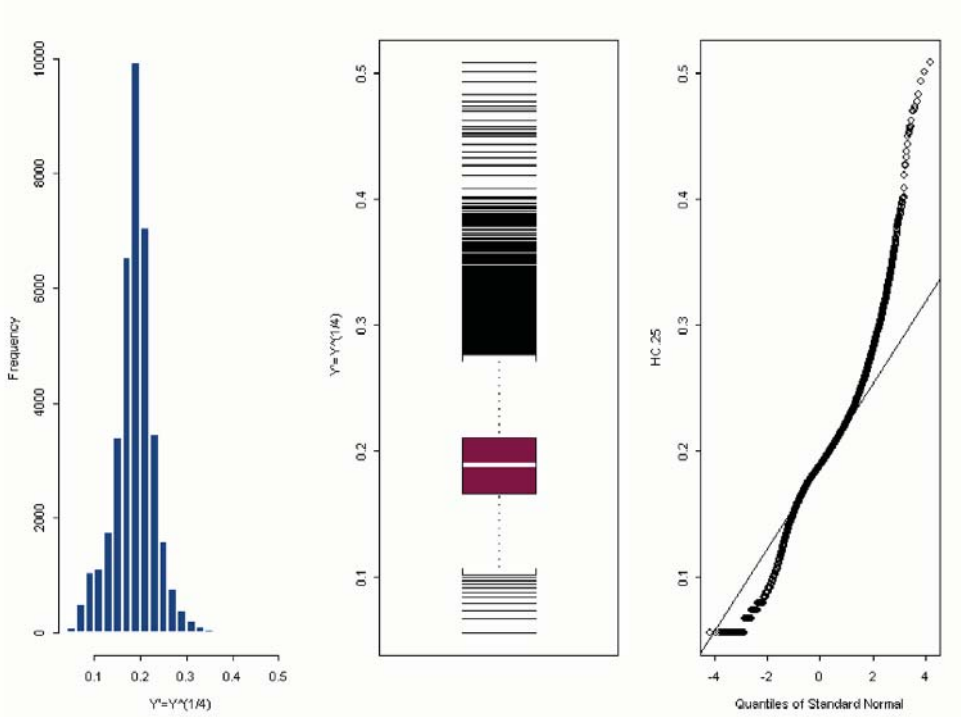


Figure 11-11 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate in Cruise Mode

11.2.1.1 NO_x HTBR Tree Model Development

Figure 11-12 illustrates the initial tree model used for the truncated transformed NO_x emission rate in cruise mode. Results for the initial model are given in Table 11-4. The tree grew into a complex model, with a considerable number of branches and 32 terminal nodes. Figure 11-13 illustrates the amount of deviation explained corresponding to the number of terminal nodes.

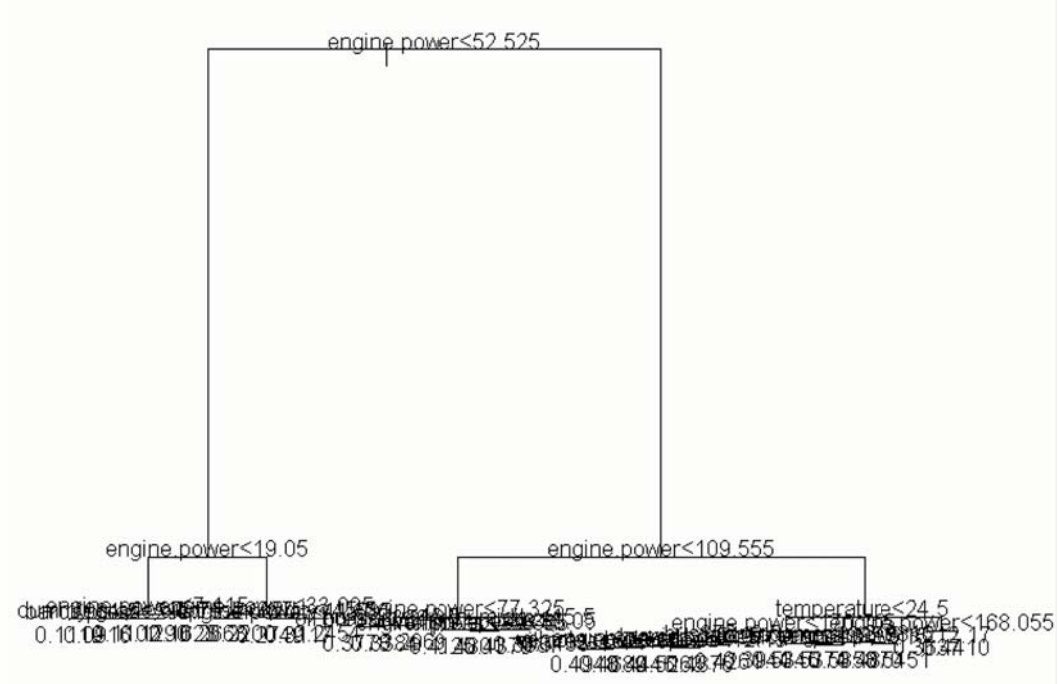


Figure 11-12 Original Untrimmed Regression Tree Model for Truncated Transformed NO_x Emission Rate in Cruise Mode

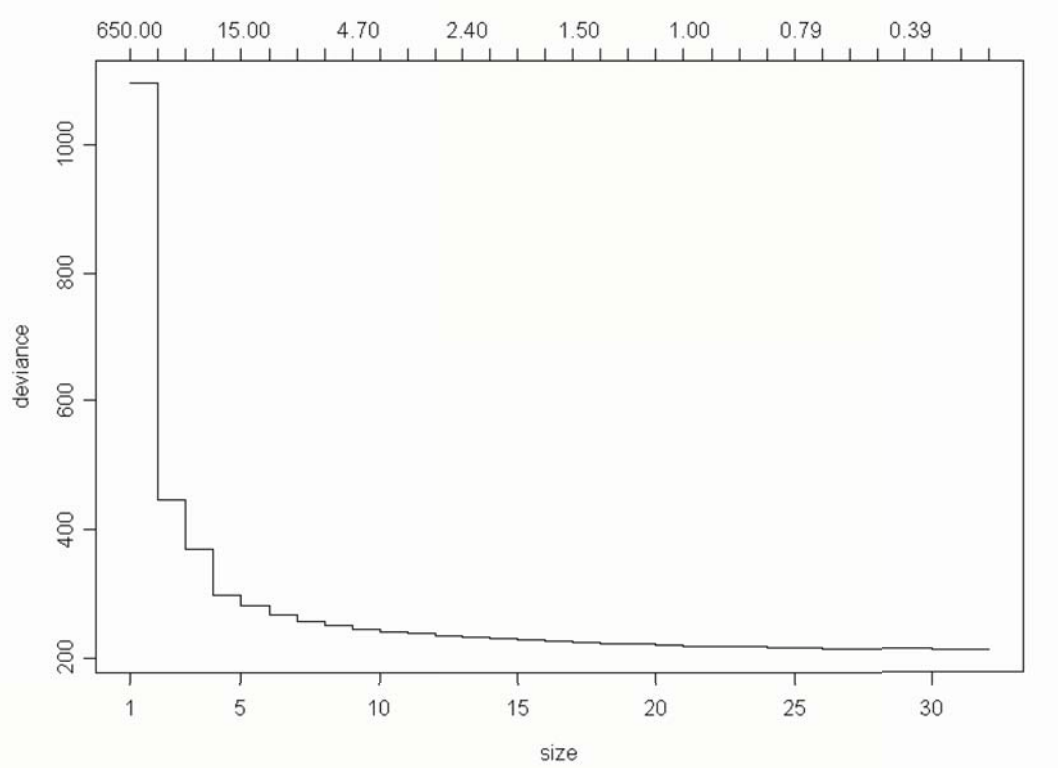


Figure 11-13 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed NO_x Emission Rate in Cruise Mode

Table 11-4 Original Untrimmed Regression Tree Results for Truncated Transformed NO_x Emission Rate in Cruise Mode

```
Regression tree:
tree(formula = NOx.50 ~ model.year + odometer + temperature + baro + humidity +
vehicle.speed + oil.temperture + oil.press + cool.temperature + eng.bar.press + en-
gine.power + acceleration + bus360 + bus361 + bus363 + bus364 + bus372 + bus375 +
bus377 + bus379 + bus380 + bus381 + bus382 + bus383 + bus384 + bus385 + dummy.grade,
data = busdata10242006.1.4, na.action = na.exclude, mincut = 400, minsize = 800,
mindev = 0.01)
Variables actually used in tree construction:
 [1] “engine.power”  “dummy.grade”   “baro”          “oil.press”
 [5] “humidity”      “vehicle.speed” “temperature”   “bus372”
 [9] “odometer”      “model.year”
Number of terminal nodes:  32
Residual mean deviance:  0.005398 = 212.4 / 39340
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -4.634e-001 -4.130e-002 -1.265e-003 -1.315e-016  3.646e-002  1.180e+000
```

For model application purposes, it is desirable to select a final model specification that balances the model's ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Figure 11-7 indicates that reduction in deviation with addition of nodes after four, although potentially statistically significant, is very small. A simplified tree model was derived which ends in four terminal nodes as compared to the 37 terminal nodes in the initial model. The residual mean deviation only increased from 210.2 to 298.9 and yielded a much cleaner model than the initial one. Results are shown in Table 11-5 and Figure 11-14. Based on above analysis, NO_x cruise model will be developed based on this result.

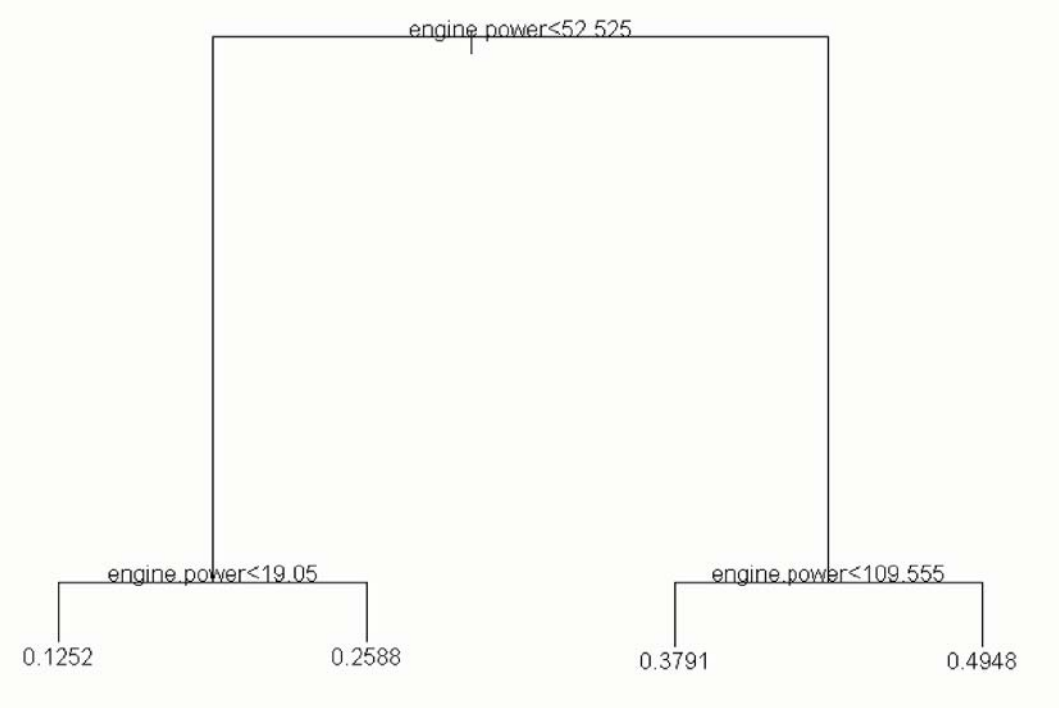


Figure 11-14 Trimmed Regression Tree Model for Truncated Transformed NO_x Emission Rate in Cruise Mode

Table 11-5 Trimmed Regression Tree Results for Truncated Transformed NO_x Emission Rate in Cruise Mode

```
Regression tree:
snip.tree(tree = tree(formula = NOx.50 ~ model.year + odometer + temperature + baro
+ humidity + vehicle.speed + oil.temperture + oil.press + cool.temperature + eng.
bar.press + engine.power + acceleration + bus360 + bus361 + bus363 + bus364 + bus372
+ bus375 + bus377 + bus379 + bus380 + bus381 + bus382 + bus383 + bus384 + bus385 +
dummy.grade, data = busdata10242006.1.4, na.action = na.exclude, mincut = 400,minsize
= 800, mindev = 0.01), nodes = c(5., 4., 6., 7.))
Variables actually used in tree construction:
[1] “engine.power”
Number of terminal nodes:  4
Residual mean deviance:  0.007591 = 298.9 / 39370
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -4.643e-001 -5.592e-002  3.280e-004 -4.143e-016  5.370e-002  1.179e+000
node), split, n, deviance, yval
      * denotes terminal node

1) root 39374 1095.00 0.3360
  2) engine.power<52.525 16280  160.50 0.1831
    4) engine.power<19.05 9222   47.70 0.1252 *
    5) engine.power>19.05 7058   41.36 0.2588 *
  3) engine.power>52.525 23094  285.90 0.4438
    6) engine.power<109.555 10186   81.41 0.3791 *
    7) engine.power>109.555 12908  128.40 0.4948 *
```

This tree model suggests that engine power is the most important explanatory variable for NO_x emissions. This finding is consistent with previous research results which verified the important effect of engine power on NO_x emissions (Ramamurthy et al. 1998; Clark et al. 2002; Barth et al. 2004). Analysis in previous chapter also indicates that engine power is correlated not only with onroad load parameters such as vehicle speed, acceleration, and grade, but also with engine operating parameters such as throttle position and engine oil pressure. On the other hand, engine power in this research is derived from engine speed, engine torque and percent engine load. So engine power can connect onroad modal activity with engine operating conditions to that extent. This fact strengthens the importance of introducing engine power into the conceptual model and the need to improve the ability to simulate engine power for regional inventory development.

11.2.1.2 CO HTBR Tree Model Development

Figure 11-15 illustrates the initial tree model used for truncated transformed CO emission rate in cruise mode. Results for initial model are given in Table 11-6. The tree grew into a complex model with a considerable number of branches and 65 terminal nodes. Figure 11-16 illustrates the amount of deviation explained corresponding to the number of terminal nodes.

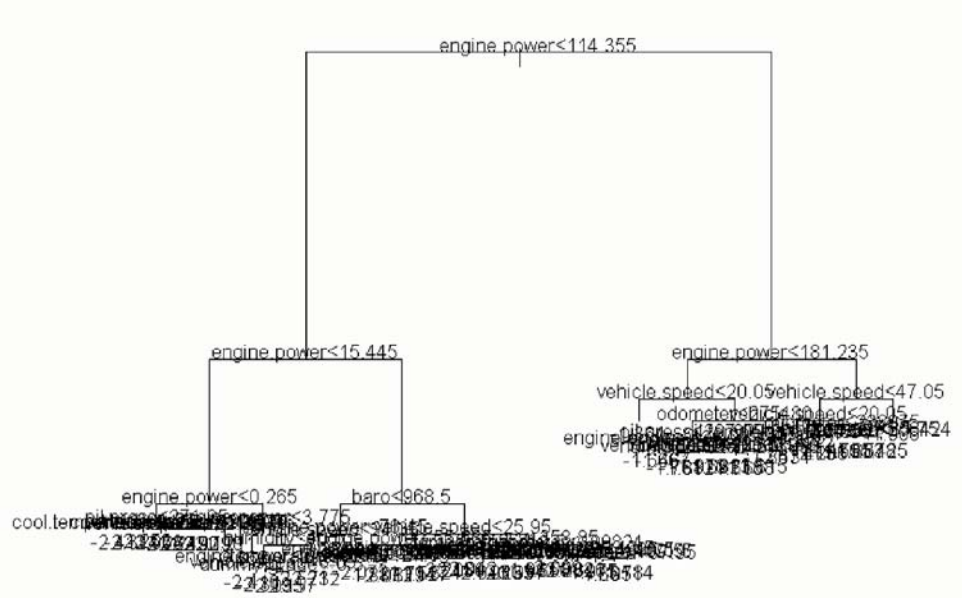


Figure 11-15 Original Untrimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Cruise Mode

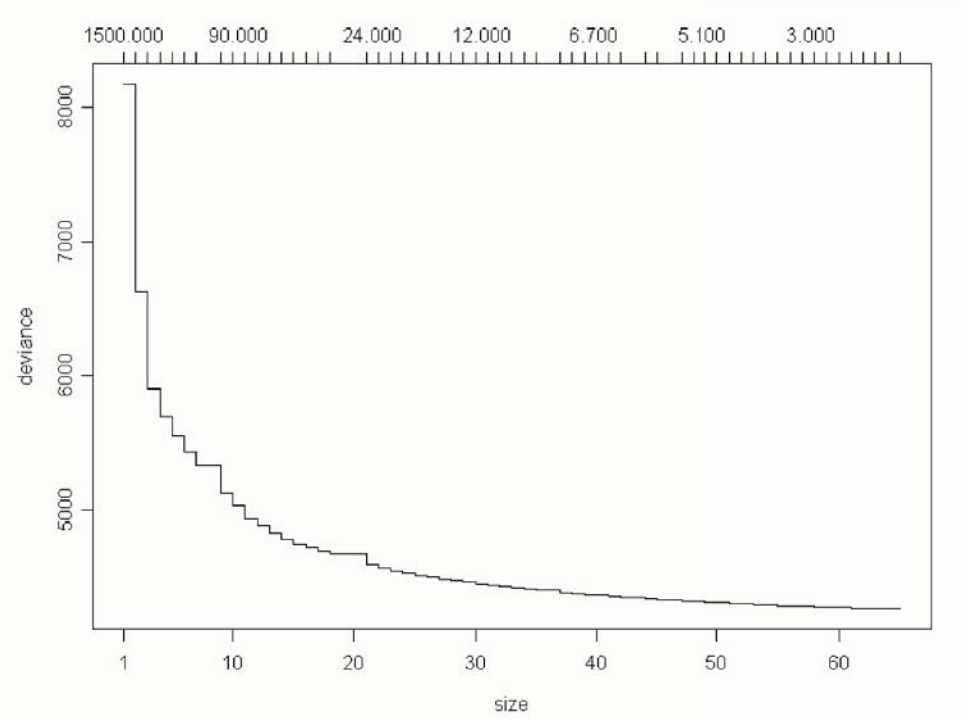


Figure 11-16 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed CO Emission Rate in Cruise Mode

Table 11-6 Original Untrimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Cruise Mode

```
Regression tree:
tree(formula = log.CO ~ model.year + odometer + temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + bus360 + bus361 + bus363 +
       bus364 + bus372 + bus375 + bus377 + bus379 + bus380 + bus381 + bus382 +
       bus383 + bus384 + bus385 + dummy.grade, data = busdata10242006.1.4,
       na.action = na.exclude, mincut = 400, minsize = 800, mindev = 0.01)
Variables actually used in tree construction:
 [1] “engine.power”     “oil.press”        “baro”
 [4] “cool.temperature” “vehicle.speed”    “acceleration”
 [7] “humidity”         “odometer”         “dummy.grade”
[10] “temperature”      “eng.bar.press”    “model.year”
[13] “oil.temperture”
Number of terminal nodes:  65
Residual mean deviance:  0.1089 = 4265 / 39150
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -2.335e+000 -1.783e-001 -1.233e-002  1.869e-016  1.691e-001  2.013e+000
```

For model application purposes, it is desirable to select a final model specification that balances the model’s ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Figure 11-16 indicates that reduction in deviation with addition of nodes after 4, although potentially statistically significant, is very small. A simplified tree model was derived which ends in 4 terminal nodes as compared to the 67 terminal nodes in the initial model. The residual mean deviation only increased from 4265 to 5698 and yielded a much more efficient model. Results are shown in Table 11-7 and Figure 11-17. The CO cruise emission rate model will be based upon these results.

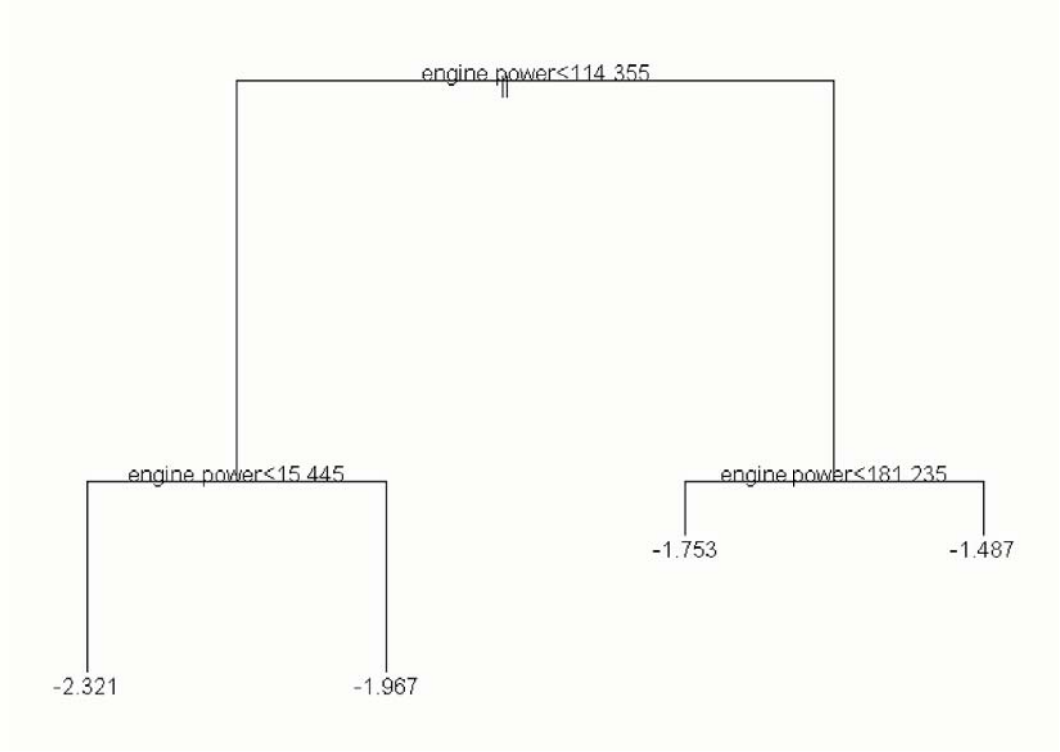


Figure 11-17 Trimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Cruise Mode

Table 11-7 Trimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Cruise Mode

```
Regression tree:
snip.tree(tree = tree(formula = log.CO ~ model.year + odometer + temperature +
       baro + humidity + vehicle.speed + oil.temperture + oil.press +
       cool.temperature + eng.bar.press + engine.power + acceleration +
       bus360 + bus361 + bus363 + bus364 + bus372 + bus375 + bus377 + bus379 +
       bus380 + bus381 + bus382 + bus383 + bus384 + bus385 + dummy.grade,
       data = busdata10242006.1.4, na.action = na.exclude, mincut = 400,
       minsize = 800, mindev = 0.01), nodes = c(4., 6., 7., 5.))
Variables actually used in tree construction:
[1] “engine.power”
Number of terminal nodes:  4
Residual mean deviance:  0.1453 = 5698 / 39210
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -2.679e+000 -2.065e-001 -7.150e-003 -4.942e-015  2.041e-001  2.452e+000
node), split, n, deviance, yval
      * denotes terminal node

1) root 39218 8170.0 -1.944
  2) engine.power<114.355 27187 4482.0 -2.076
    4) engine.power<15.445 8414 1639.0 -2.321 *
    5) engine.power>15.445 18773 2115.0 -1.967 *
  3) engine.power>114.355 12031 2147.0 -1.646
    6) engine.power<181.235 7220 1146.0 -1.753 *
    7) engine.power>181.235 4811  797.8 -1.487 *
```

This tree model suggested that engine power is the most important explanatory variable for CO emissions. This finding is consistent with NO_x emissions. This tree will be used as reference for linear regression model development.

11.2.1.3 HC HTBR Tree Model Development

Figure 11-18 illustrates the initial tree model used for truncated transformed HC emission rate in cruise mode. Results for initial model are given in Table 11-8. The tree grew into a complex model with a considerable number of branches and 61 terminal nodes.

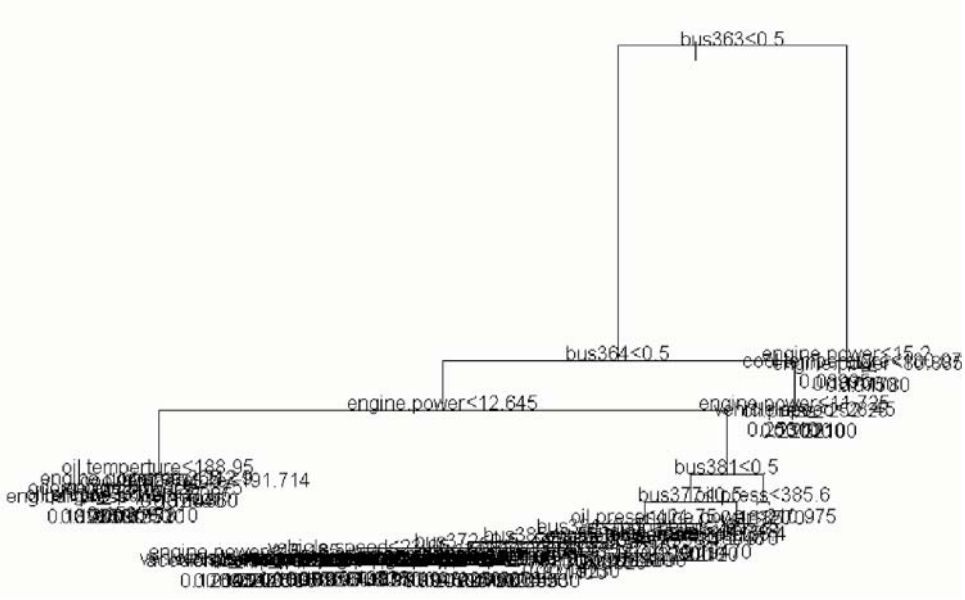


Figure 11-18 Original Untrimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Cruise Mode

Table 11-8 Original Untrimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Cruise Mode

```
Regression tree:
tree(formula = HC.25 ~ model.year + odometer + temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + bus360 + bus361 + bus363 +
       bus364 + bus372 + bus375 + bus377 + bus379 + bus380 + bus381 + bus382 +
       bus383 + bus384 + bus385 + dummy.grade, data = busdata10242006.1.4,
       na.action = na.exclude, mincut = 400, minsize = 800, mindev = 0.01)
Variables actually used in tree construction:
 [1] "bus363"           "bus364"           "engine.power"
 [4] "oil.temperture"   "odometer"         "oil.press"
 [7] "humidity"         "cool.temperature" "bus381"
[10] "bus377"           "baro"             "temperature"
[13] "bus372"           "vehicle.speed"    "dummy.grade"
[16] "bus385"
Number of terminal nodes:  56
Residual mean deviance:  0.0008147 = 30.93 / 37960
Distribution of residuals:
        Min.      1st Qu.       Median         Mean      3rd Qu.         Max.
 -1.862e-001 -1.595e-002 -3.021e-003 -1.297e-018  1.230e-002  2.886e-001
```

Figure 11-18 and Table 11-8 suggest that the tree analysis of HC emission rates identified a number of buses that appear to exhibit significantly different emission rates under all load conditions than the other buses (i.e., some of the bus dummy variables appeared as significant in the initial tree splits). Two bus dummy variables split the data pool at the first two levels of the HC tree model. This same result was noted for these buses in the acceleration mode. Although variability exists for three pollutants across 15 buses, the division was even more obvious for HC emissions (see Figure 11-4 and Table 11-2). Although it is tempting to develop different emission rates for these buses to reduce emission rate deviation in the sample pool, it is difficult to justify doing so. Unless these is an obvious reason to classify these three buses as high emitters (i.e., significantly higher than normal emitting vehicles, perhaps by as much as a few standard deviations from the mean), and unless there are enough data to develop separate emission rate models for high emitters, one cannot justify removing the data from the data set. Until such data exist to justify treating these buses as high emitters, the bus dummy variables for individual buses are removed from the analyses and all 15 buses are treated as part of the whole data set.

Another tree model was generated excluding the bus dummy variables. However, odometer reading also had to be excluded because the previous "Bus 363<0.5" tree cutpoint was replaced by "odometer>282096" (i.e., was identically correlated to the same bus). This new tree model is illustrated in Figure 11-19 and Table 11-9. The tree model is then trimmed for application purposes, as was done for the NO_x and CO models.

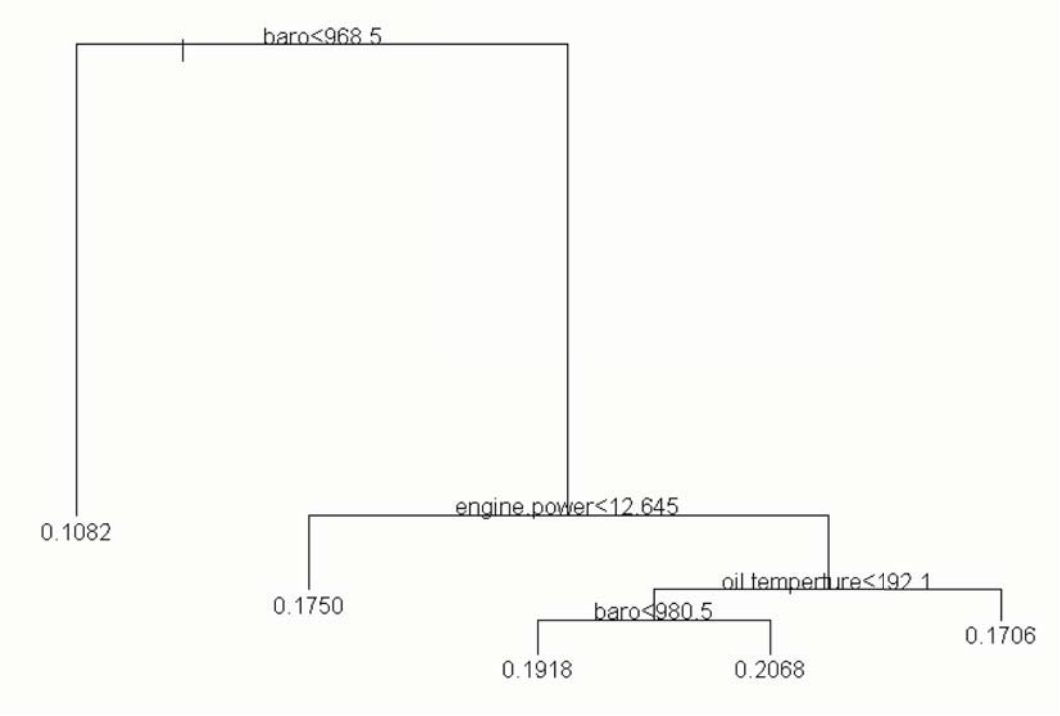


Figure 11-19 Trimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Cruise Mode

Table 11-9 Trimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Cruise Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ temperature + baro + humidity +
       vehicle.speed + oil.temperture + oil.press + cool.temperature +
       eng.bar.press + engine.power + acceleration + dummy.grade, data =
       busdata10242006.1.4, na.action = na.exclude, mincut = 400, minsize =
       800, mindev = 0.01), nodes = c(15., 28., 2., 29., 6.))
Variables actually used in tree construction:
[1] “baro”           “engine.power”   “oil.temperture”
Number of terminal nodes:  5
Residual mean deviance:  0.001207 = 45.87 / 38020
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.328e-001 -2.037e-002 -3.530e-003  1.177e-015  1.609e-002  3.256e-001
node), split, n, deviance, yval
      * denotes terminal node

 1) root 38020 71.970 0.1876
   2) baro<968.5 2957  2.349 0.1082 *
   3) baro>968.5 35063 49.420 0.1943
     6) engine.power<12.645 6821 13.850 0.1750 *
     7) engine.power>12.645 28242 32.420 0.1989
      14) oil.temperture<192.1 26727 29.900 0.2005
        28) baro<980.5 11265  9.610 0.1918 *
        29) baro>980.5 15462 18.820 0.2068 *
      15) oil.temperture>192.1 1515  1.244 0.1706 *
```

The new tree model suggests that barometric pressure is the most important explanatory variable for HC emission rates. However, this finding is challenged by the fact that all the 2957 data points in the first left hand branch of the tree (barometric pressure < 968.5) belong to Bus 363. Although this dataset was collected under a wide variety of environmental conditions, the scope of barometric pressure was limited for individual buses tested. As reported earlier, Bus 363 exhibited significantly lower HC emissions than the other buses (see Figure 11-4), but the reason is not clear at this time. To develop a reasonable tree model given the limited data collected, the environmental parameters are excluded from the model until a greater distribution of environmental conditions can be represented in a test data set. With data collected from a more comprehensive testing program, environmental variables can be integrated into the model directly, or perhaps correction factors for the emission rates can be developed. The secondary trimmed tree is presented in Figure 11-20 and Table 11-10.

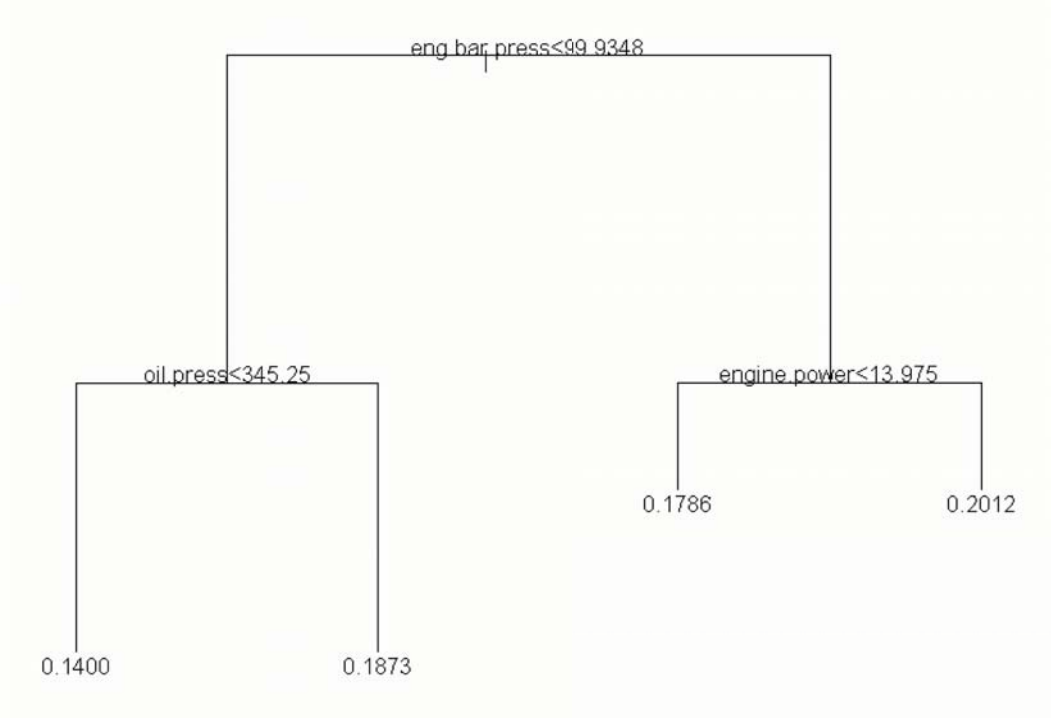


Figure 11-20 Secondary Trimmed Regression Tree Model for Truncated Transformed HC in Cruise Mode

Table 11-10 Trimmed Regression Tree Results for Truncated Transformed HC in Cruise Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ engine.power + vehicle.speed +
       acceleration + oil.temperture + oil.press + cool.temperature +
       eng.bar.press, data = busdata10242006.1.4, na.action = na.exclude,
       mincut = 400, minsize = 800, mindev = 0.01), nodes = c(6., 5., 7.,
       4.))
Variables actually used in tree construction:
[1] “eng.bar.press” “oil.press”     “engine.power”
Number of terminal nodes:  4
Residual mean deviance:  0.00148 = 56.27 / 38020
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.310e-001 -2.290e-002 -2.164e-003  1.281e-015  1.942e-002  3.220e-001
node), split, n, deviance, yval
      * denotes terminal node

1) root 38020 71.970 0.1876
  2) eng.bar.press<99.9348 10827 24.640 0.1656
    4) oil.press<345.25 4965 10.870 0.1400 *
    5) oil.press>345.25 5862  7.754 0.1873 *
  3) eng.bar.press>99.9348 27193 40.010 0.1963
    6) engine.power<13.975 5879 12.660 0.1786 *
    7) engine.power>13.975 21314 24.990 0.2012 *
```

The tree model excluding bus dummy variables, odometer readings, and environmental conditions is shown in Figure 11-20 and Table 11-11. This final tree model suggests that engine power is the most important explanatory variable for HC emissions. This finding is consistent with analysis of NO_x and CO emission rates. Although engine operating parameters such as oil pressure might impact emissions, such variables are not easy to implement in real-world models. After excluding engine barometric pressure and oil pressure from the tree model, leaving engine power only, the residual mean deviation increased slightly from 56.27 to 65.56. The final HTBR tree for HC emissions is shown in Figure 11-21 and Table 11-11. HC cruise emission rate model will be developed based upon these results.

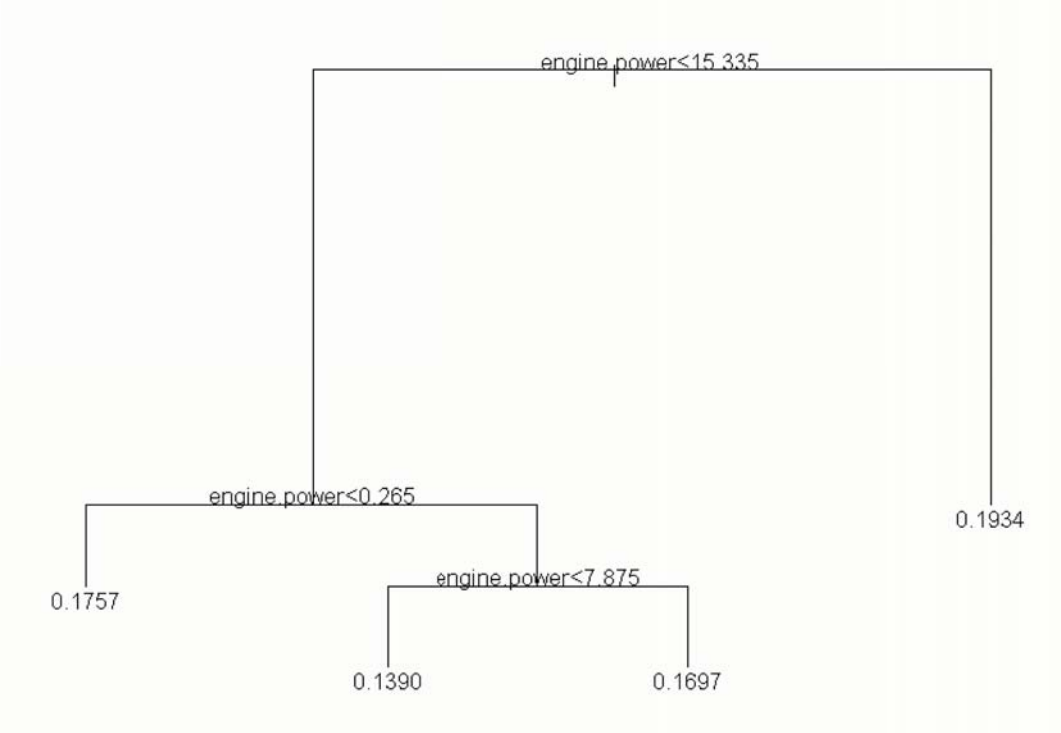


Figure 11-21 Final Regression Tree Model for Truncated Transformed HC and Engine Power in Cruise Mode

Table 11-11 Final Regression Tree Results for Truncated Transformed HC and Engine Power in Cruise Mode

```
Regression tree:
snip.tree(tree = tree(formula = HC.25 ~ engine.power, data =
       busdata10242006.1.4, na.action = na.exclude, mincut = 400, minsize =
       800, mindev = 0.01), nodes = c(11., 10., 3.))
Number of terminal nodes:  4
Residual mean deviance:  0.001725 = 65.56 / 38020
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -1.372e-001 -2.070e-002 -6.875e-004  1.742e-015  2.090e-002  3.309e-001
node), split, n, deviance, yval
      * denotes terminal node

 1) root 38020 71.970 0.1876
   2) engine.power<15.335 8298 21.630 0.1666
     4) engine.power<0.265 4617  9.741 0.1757 *
     5) engine.power>0.265 3681 11.020 0.1551
      10) engine.power<7.875 1746  3.849 0.1390 *
      11) engine.power>7.875 1935  6.311 0.1697 *
   3) engine.power>15.335 29722 45.660 0.1934 *
```

11.2.2 OLS Model Development and Refinement

Once a manageable number of modal variables have been identified through regression tree analysis, the modeling process moves into the phase in which ordinary least squares techniques are used to obtain a final model. The research objective here is to identify the extent to which the identified factors influence emission rate in cruise mode. Modelers rely on previous research, a priori knowledge, educated guesses, and stepwise regression procedures to identify acceptable functional forms, to determine important interactions, and to derive statistically and theoretically defensible models. The final model will be our best understanding about the functional relationship between independent variables and dependent variables.

11.2.2.1 NO_x Emission Rate Model Development for Cruise Mode

Based on previous analysis, truncated transformed NO_x will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing the performance of statistical models. HTBR tree model results suggest that engine power is the best one to begin with.

11.2.2.1.1 Linear Regression Model with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1(engine.power) + Error \qquad (1.1)$$

The regression run yields the results shown in Table 11-12 and Figure 11-22.

Table 11-12 Regression Result for NO_x Model 1.1

```
Call: lm(formula = NOx.50 ~ engine.power, data = busdata10242006.1.4, na.action =
na.exclude)
Residuals:
     Min       1Q   Median      3Q   Max
 -0.5717 -0.06302 0.006377 0.06653 1.259

Coefficients:
                Value Std. Error  t value Pr(>|t|)
 (Intercept)   0.1815   0.0007   242.8528   0.0000
engine.power   0.0018   0.0000   274.7573   0.0000

Residual standard error: 0.09765 on 39372 degrees of freedom
Multiple R-Squared: 0.6572
F-statistic: 75490 on 1 and 39372 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.7526

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1  719.8396 719.8396 75491.58     0
   Residuals 39372  375.4263   0.0095
```

The results suggest that engine power explains about 66% of the variance in truncated transformed NO_x. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, residual normality is examined in the QQ plot and constancy of variance is checked by examining residuals vs. fitted values.

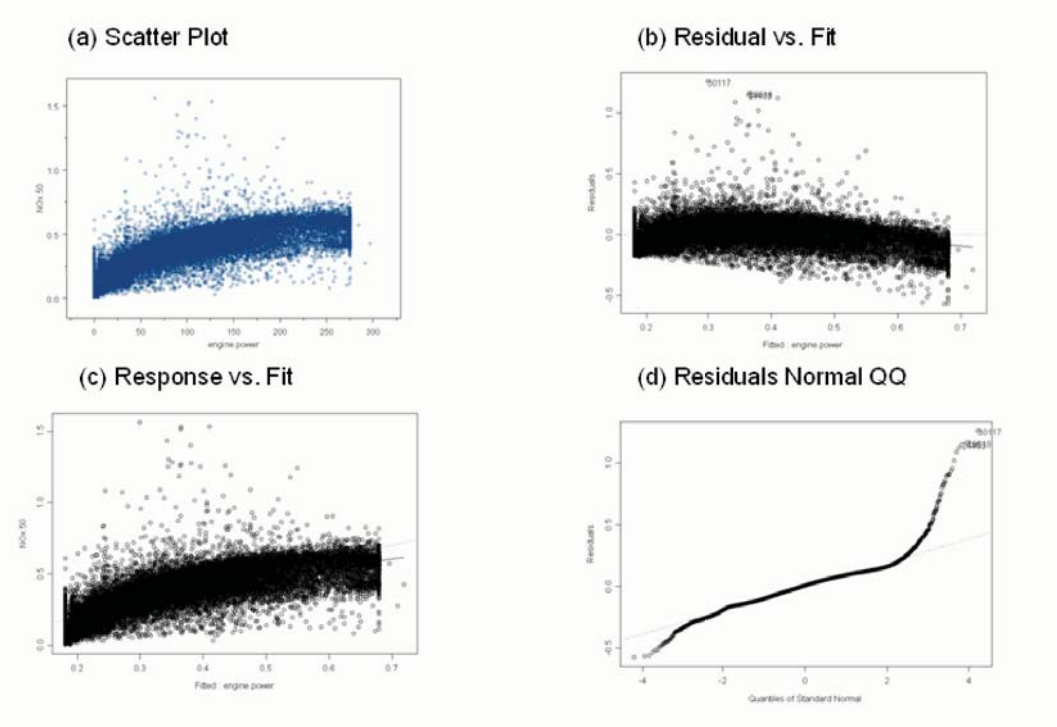


Figure 11-22 QQ and Residual vs. Fitted Plot for NO_x Model 1.1

The residual plot in Figure 11-22 shows a departure from linear regression assumptions indicating a need to explore a curvilinear regression function. Since the variability at the different X levels appears to be fairly constant, a transformation on X is considered. The reason to consider transformation first is to avoid multicollinearity brought about by adding the second-order of X. Based on the prototype plot in Figure 11-22, the square root transformation and logarithmic transformation are tested. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to determine which transformation is most effective.

$$Y = \beta_0 + \beta_1 engine.power^{(1/2)} + Error \qquad (1.2)$$

$$Y = \beta_0 + \beta_1 log_{10}(engine.power+1) + Error \qquad (1.3)$$

The result for Model 1.2 is shown in Table 11-13 and Figure 11-23, while the result for Model 1.3 is shown in Table 11-14 and Figure 11-24.

Table 11-13 Regression Result for NO_x Model 1.2

```
Call: lm(formula = NOx.50 ~ engine.power^(1/2), data = busdata10242006.1.4,
na.action = na.exclude)
Residuals:
     Min       1Q     Median      3Q  Max
 -0.5007 -0.04881 -0.0008896 0.05047 1.22

Coefficients:
                          Value Std. Error  t value Pr(>|t|)
          (Intercept)   0.0874   0.0008   104.1024   0.0000
I(engine.power^(1/2))   0.0311   0.0001   342.3056   0.0000

Residual standard error: 0.08364 on 39372 degrees of freedom
Multiple R-Squared: 0.7485
F-statistic: 117200 on 1 and 39372 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                      (Intercept)
I(engine.power^(1/2)) -0.8649

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                         Df Sum of Sq  Mean Sq  F Value Pr(F)
I(engine.power^(1/2))     1  819.8002 819.8002 117173.2     0
            Residuals 39372  275.4656   0.0070
```

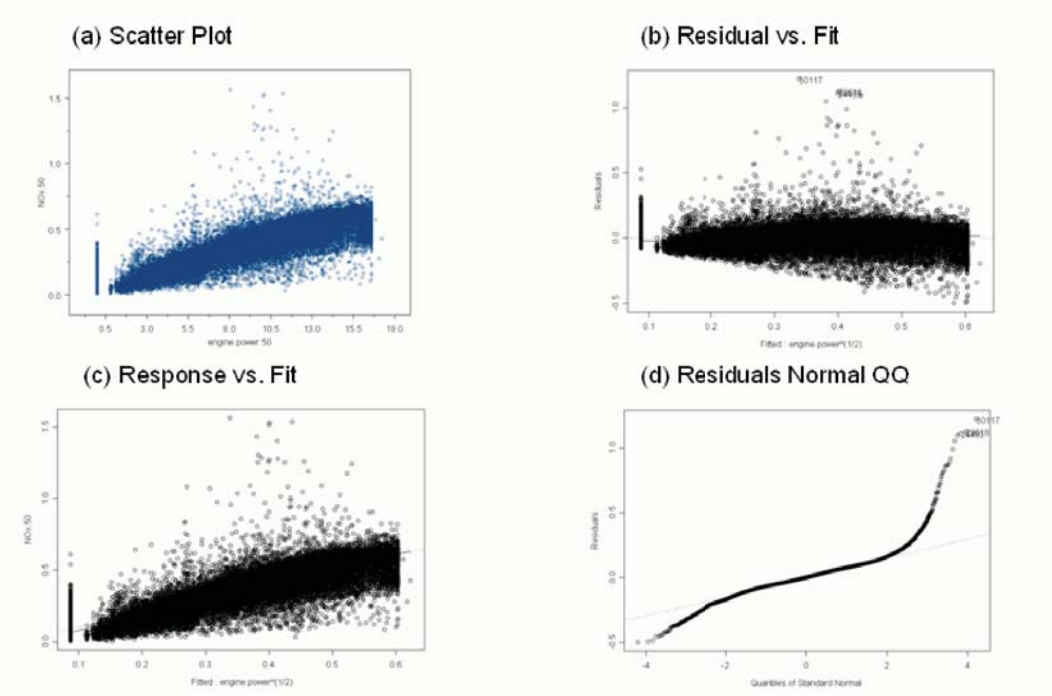


Figure 11-23 QQ and Residual vs. Fitted Plot for NO_x Model 1.2

Table 11-14 Regression Result for NO_x Model 1.3

```
Call: lm(formula = NOx.50 ~ log10(engine.power + 1), data = busdata10242006.1.4,
na.action = na.exclude)
Residuals:
     Min       1Q    Median      3Q   Max
 -0.4047 -0.06677 -0.002155 0.06107 1.182

Coefficients:
                           Value Std. Error  t value Pr(>|t|)
            (Intercept)   0.0306   0.0012    25.5525   0.0000
log10(engine.power + 1)   0.1895   0.0007   279.4403   0.0000

Residual standard error: 0.09656 on 39372 degrees of freedom
Multiple R-Squared: 0.6648
F-statistic: 78090 on 1 and 39372 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                         (Intercept)
log10(engine.power + 1) -0.9135

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1  728.1347 728.1347 78086.87     0
              Residuals 39372  367.1311   0.0093
```

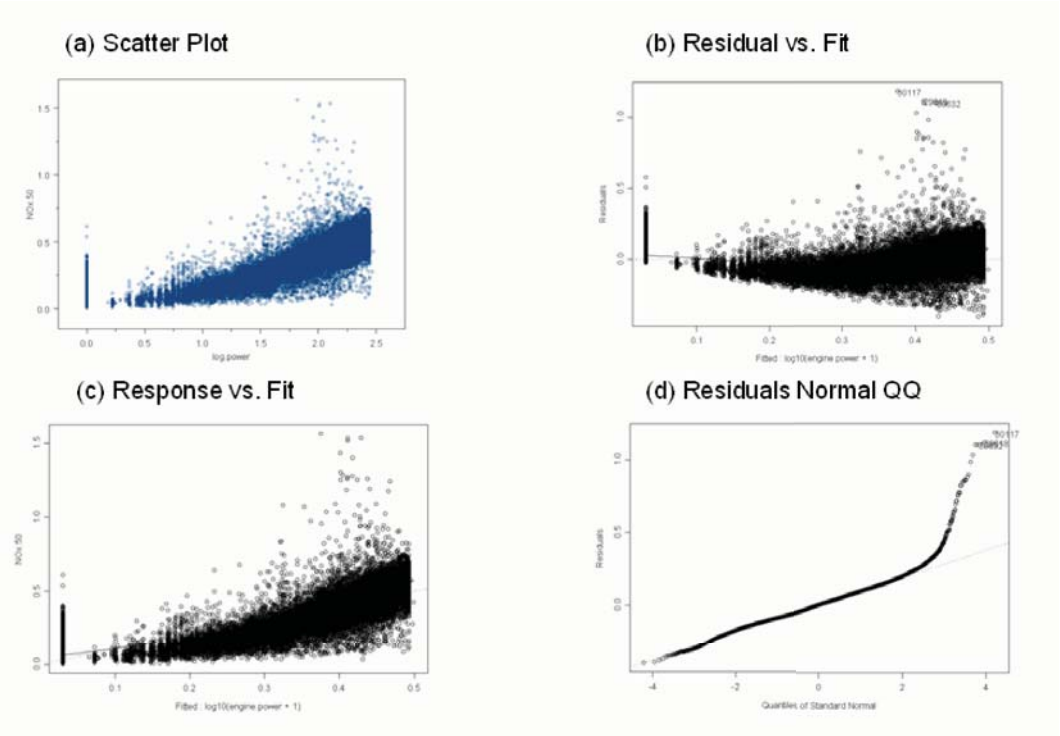


Figure 11-24 QQ and Residual vs. Fitted Plot for NO_x Model 1.3

The results suggest that by using square root transformed engine power, the model increases the amount of variance explained in truncated transformed NO_x from about 66% (Model 1.1) to about 75% (Model 1.2), while remaining about 66% (Model 1.3) by using log transformed engine power.

Model 1.2 improves the R^2 more than does Model 1.3. The residuals scatter plot for Model 1.2 (Figure 11-23) shows a more reasonably linear relation than Model 1.3 (Figure 11-24). Figure 11-23 also shows that Model 1.2 does a better job in improving the pattern of variance. QQ plot shows a kind of normality except two tails.

11.2.2.1.2 Linear Regression Model with Dummy Variables

Figure 11-14 suggests that the relationship between NO_x and engine power may be somewhat different across the engine power ranges identified in the tree analysis. That is, there may be higher or lower NO_x emissions in different engine power operating ranges. One dummy variable is created to represent different engine power ranges identified in Figure 11-14 for use in linear regression analysis as illustrated below:

Engine power (bhp)	Dummy1
< 52.525	1
≥ 52.525	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variables and interactions can help improve the model.

$$Y = \beta_0 + \beta_1\, engine.power^{(1/2)} + \beta_2\, dummy1 + \beta_3\, dummy1 engine.power^{(1/2)} + Error \qquad (1.4)$$

The result for Model 1.4 is shown in Table 11-15 and Figure 11-25.

Table 11-15 Regression Result for NO_x Model 1.4

```
Call: lm(formula = NOx.50 ~ engine.power^(1/2) + dummy1 * engine.power^(1/2), data =
busdata10242006.1.4, na.action = na.exclude)
Residuals:
     Min       1Q    Median      3Q   Max
 -0.4812 -0.04778 0.0001059 0.04843 1.195

Coefficients:
                                Value Std. Error   t value  Pr(>|t|)
                 (Intercept)    0.1581    0.0024    65.9078    0.0000
       I(engine.power^(1/2))    0.0254    0.0002   122.2468    0.0000
                      dummy1   -0.0682    0.0026   -25.9438    0.0000
I(engine.power^(1/2)):dummy1    0.0020    0.0003     6.1264    0.0000

Residual standard error: 0.08224 on 39370 degrees of freedom
Multiple R-Squared: 0.7569
F-statistic: 40850 on 3 and 39370 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                            (Intercept) I(engine.power^(1/2))  dummy1
       I(engine.power^(1/2)) -0.9742
                      dummy1 -0.9123      0.8888
I(engine.power^(1/2)):dummy1  0.6175     -0.6339               -0.8171

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                               Df Sum of Sq  Mean Sq  F Value         Pr(F)
       I(engine.power^(1/2))    1  819.8002 819.8002 121203.8 0.000000e+000
                      dummy1    1    8.9202   8.9202   1318.8 0.000000e+000
I(engine.power^(1/2)):dummy1    1    0.2539   0.2539     37.5 9.073785e-010
                    Residuals 39370  266.2915   0.0068
```

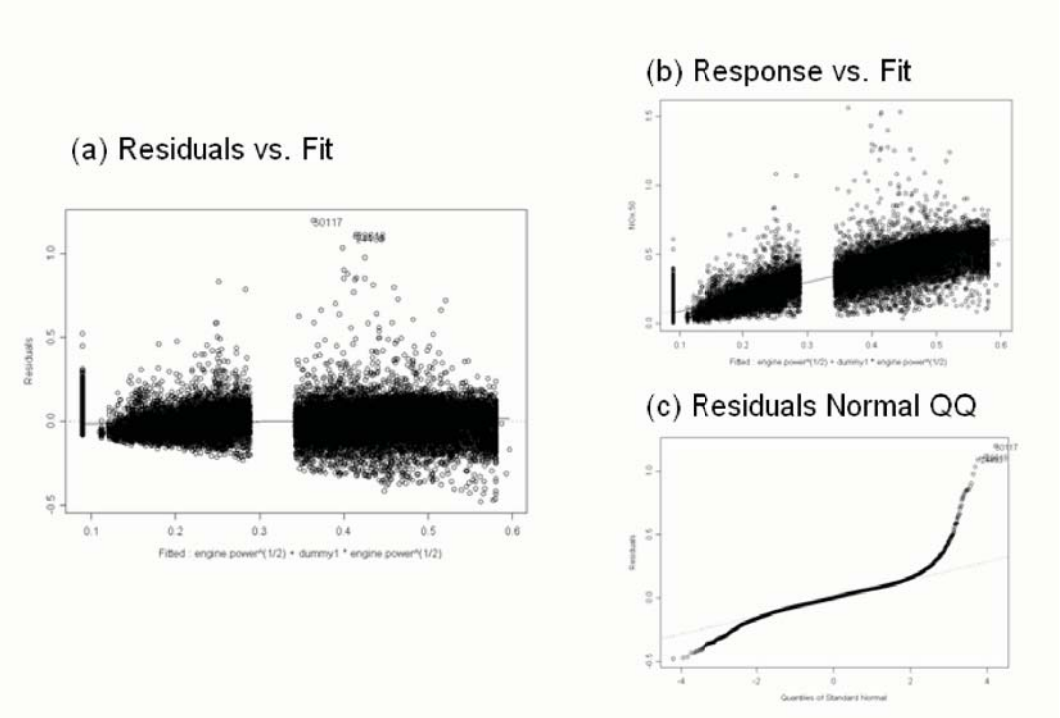


Figure 11-25 QQ and Residual vs. Fitted Plot for NO_x Model 1.4

The results suggest that by using dummy variables and interactions with transformed engine power, the model increases the amount of variance explained in truncated transformed NO_x from about 75% (Model 1.2) to about 77% (Model 1.4).

Model 1.4 slightly improves the R^2 more than does Model 1.2. The residuals scatter plot for Model 1.4 (Figure 11-25) shows a slightly more reasonably linear relation. Figure 11-25 shows that Model 1.4 may also do a slightly better job in improving the pattern of variance. The QQ plot shows general normality with the exceptions arising in the tails. However, it is important to note that the model improvement, in terms of amount of variance explained by the model, is marginal at best.

11.2.2.1.3 Model Discussion

Previous sections provide the model development process from one OLS model to another OLS model. To test whether the linear regression with power was a beneficial addition to the regression tree model, the mean ERs at HTBR end nodes (single value) are compared to the predictions from the linear regression function with engine power. The results of the performance evaluation are shown in Table 11-16. The improvement in R^2 associated with moving toward a linear function of engine power is tremendous. Hence, the use of the linear regression

function will provide a significant improvement in spatial and temporal model prediction capability. However this linear regression function might still be improved. Since the R^2 and slope in Table 11-16 are derived by comparing model predictions and actual observations for emission rates (untransformed y), these numbers are different in linear regression models.

Two transforms of engine power were tested: square root transformation and log transformation. The results of the performance evaluation are shown in Table 11-16. These results suggest that linear regression function with square root transformation performs slightly better.

Given that the regression tree modeling exercise indicated that a number of power cutpoints may play a role in the emissions process, an additional modeling run was performed. The results of the performance evaluation are shown in Table 11-16. Analysis results suggest that the linear regression function with dummy variable performs slightly better than the model without the power cutpoints.

Table 11-16 Comparative Performance Evaluation of NOx Emission Rate Models

	Coefficient of determination (R^2)	**Slope (β_1)**	**RMSE**	**MPE**
Mean ERs	0.00003	1.000	0.12008	-0.000006
Linear regression (power)	0.529	0.814	0.08542	0.01031
Linear regression (power^0.5)	0.614	0.975	0.07494	0.00707
Linear regression (log(power))	0.587	1.287	0.08043	0.00933
Linear regression (power^0.5) w/dummy variables	0.627	1.011	0.07372	0.00704

Although the linear regression function with dummy variables performs slightly better than linear regression function with square root transformation, more explanatory variables (dummy variable and the interaction with engine power) are introduced and the complexity of the regression model increases. There is only one regression function for Model 1.2 while there are two regression functions for Model 1.4. There is also no obvious reason why the engine may be performing slightly differently within these power regimes, yielding different regression slopes and intercepts. The fuel injection systems in these engines may operate slightly differently under low load (near-idle) and high load conditions. The fuel injection system may be controlled by the engine computer, or there may be a sufficient number of low power cruise operations and high power cruise operations that are incorrectly classified, and may be better classified as idle or acceleration events (perhaps due to GPS speed data errors). In any case, because the model with dummy variables does not perform appreciably better than the model without the dummy variables, the dummy variables are not included in the final model selection at this time.

These dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis. Model 1.2 is selected as the preliminary 'final' model.

The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform to a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have uniform variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying any patterns in the residuals. Figure 11-23 plot (b) shows this plot for NO_x model 1.2. Without considering variance due to high emission points and zero load data, there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 11-23 plot (d) shows the normal quantile plot of NO_x model 1.2. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm\sqrt{MSE}$ and about 90 percent to fall between $\pm$ 1.645 $\sqrt{MSE}$. Actually, 81.79% of residuals fall within the first limits, while 94.05% of residuals to fall within the second limits. Thus the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but are due to the nature of the data set. For example, even after the transformation, the response variable is not a true normal distribution.

Based on the above analysis, the final NO_x emission rate model selected for cruise mode is:

$$NO_x = (0.087 + 0.0311(\text{engine.power})^{(1/2)})^2$$

Analysis results support the observation that the final NO_x emission model is significantly better at explaining variability without making the model too complex. Since there is only one engine type, complexity may not be valid in terms of transferability. This model is specific to the engine classes employed in the transit bus operations. Different models may need to be developed for other engine classes and duty cycles.

11.2.2.2 CO Emission Rate Model Development for Cruise Mode

Based on previous analysis, truncated transformed CO will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing statistical models. HTBR tree model results suggest that engine power is the best one to begin with.

11.2.2.2.1 Linear Regression Model with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1 engine.power + Error \qquad (2.1)$$

The regression run yields the results shown in Table 11-17 and Figure 11-26.

Table 11-17 Regression Result for CO Model 2.1

```
Call: lm(formula = log.CO ~ engine.power, data = busdata10242006.1.4, na.action =
na.exclude)
Residuals:
    Min      1Q   Median     3Q   Max
 -2.779 -0.2088 -0.01417 0.2153 2.376

Coefficients:
                 Value Std. Error   t value  Pr(>|t|)
 (Intercept)   -2.2230    0.0030  -751.4277    0.0000
engine.power    0.0033    0.0000   125.1304    0.0000

Residual standard error: 0.3859 on 39216 degrees of freedom
Multiple R-Squared: 0.2853
F-statistic: 15660 on 1 and 39216 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.7525

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1  2331.251 2331.251 15657.62     0
   Residuals 39216  5838.839    0.149
```

These results suggest that engine power explains about 29% of the variance in truncated transformed CO. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, the normality is examined in the QQ plot and constancy of variance is checked by examining residuals vs. fitted values.

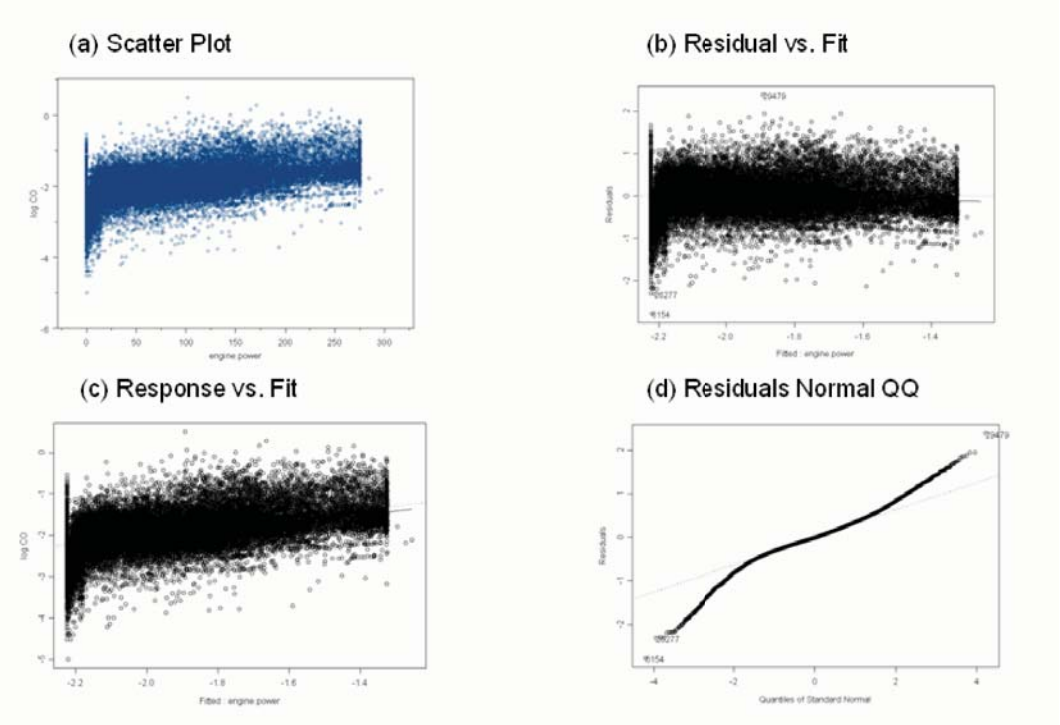


Figure 11-26 QQ and Residual vs. Fitted Plot for CO Model 2.1

Although the residual plot in Figure 11-26 shows a linear relationship between engine power and truncated transformed CO, square root transformation and logarithmic transformation are tested to see whether transformation would be useful to improve the model. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to decide which transformation is most effective.

$$Y = \beta_0 + \beta_1 \text{engine.power}^{(1/2)} + \text{Error} \qquad (2.2)$$

$$Y = \beta_0 + \beta_1 \log_{10}(\text{engine.power}+1) + \text{Error} \qquad (2.3)$$

The results for Model 2.2 are shown in Table 11-18 and Figure 11-27, while the results for Model 2.3 are shown in Table 11-19 and Figure 11-28.

Table 11-18 Regression Result for CO Model 2.2

```
Call: lm(formula = log.CO ~ engine.power^(1/2), data = busdata10242006.1.4,
       na.action = na.exclude)
Residuals:
    Min      1Q   Median     3Q   Max
 -2.679 -0.2124 -0.01769 0.2178 2.319

Coefficients:
                          Value Std. Error   t value  Pr(>|t|)
         (Intercept)   -2.3645    0.0039  -610.0636    0.0000
I(engine.power^(1/2))    0.0526    0.0004   125.3638    0.0000

Residual standard error: 0.3857 on 39216 degrees of freedom
Multiple R-Squared: 0.2861
F-statistic: 15720 on 1 and 39216 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                       (Intercept)
I(engine.power^(1/2)) -0.8646

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                          Df Sum of Sq  Mean Sq  F Value Pr(F)
I(engine.power^(1/2))     1  2337.466 2337.466 15716.09     0
            Residuals 39216  5832.624    0.149
```

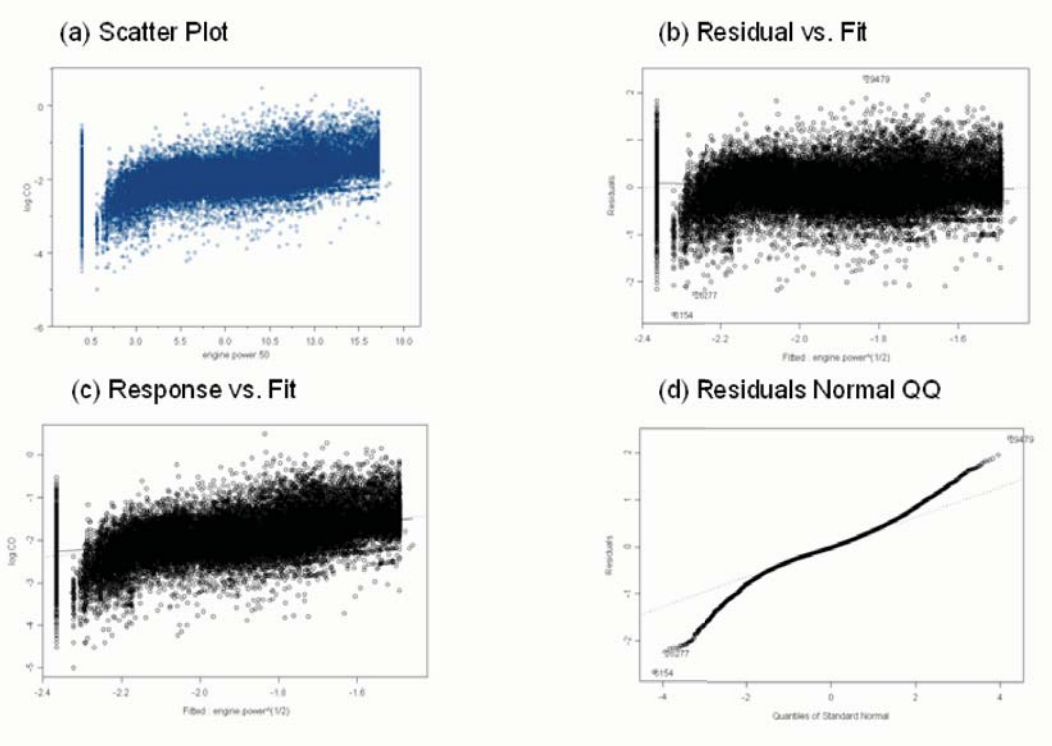


Figure 11-27 QQ and Residual vs. Fitted Plot for CO Model 2.2

Table 11-19 Regression Result for CO Model 2.3

```
Call: lm(formula = log.CO ~ log10(engine.power + 1), data = busdata10242006.1.4,
na.action = na.exclude)
Residuals:
    Min      1Q  Median     3Q   Max
 -2.636 -0.2225 -0.0167 0.2193 2.308

Coefficients:
                           Value Std. Error   t value  Pr(>|t|)
            (Intercept)   -2.4326    0.0050  -489.4690    0.0000
log10(engine.power + 1)    0.3031    0.0028   107.5567    0.0000

Residual standard error: 0.4011 on 39216 degrees of freedom
Multiple R-Squared: 0.2278
F-statistic: 11570 on 1 and 39216 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                        (Intercept)
log10(engine.power + 1) -0.9132

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)     1  1861.106 1861.106 11568.45     0
              Residuals 39216  6308.983    0.161
```

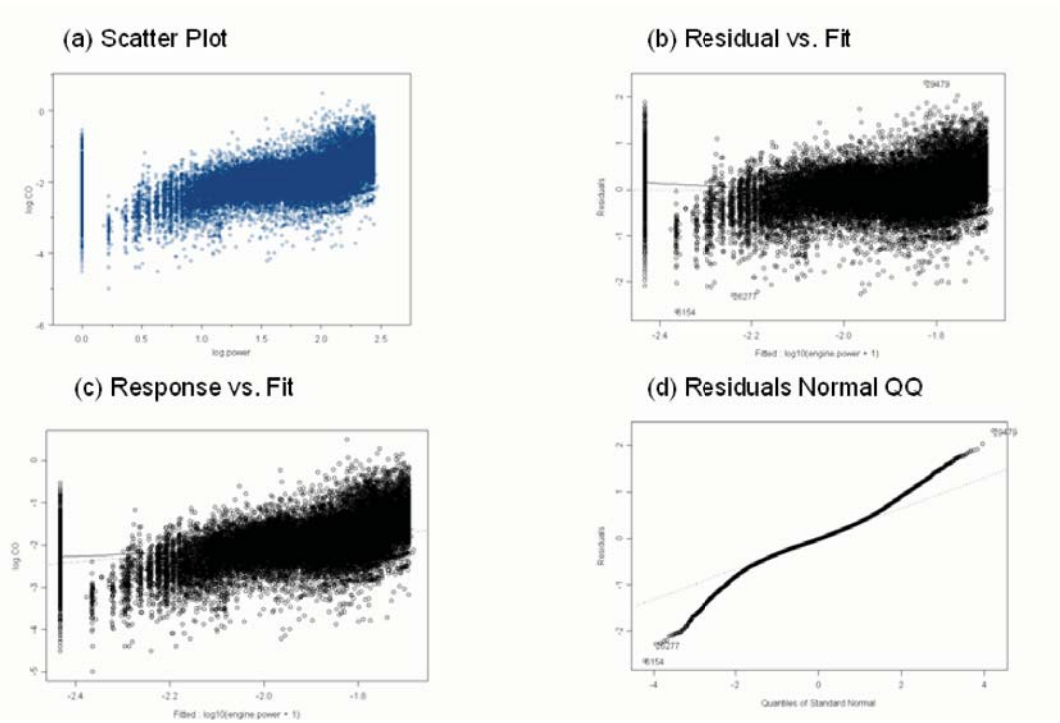


Figure 11-28 QQ and Residual vs. Fitted Plot for CO Model 2.3

The results suggest that by using transformed engine power, the model retains the amount of variance explained in truncated transformed CO at about 29% (Model 2.2), and even decreases to 23% (Model 2.3).

Considering two kinds of transformation, Model 2.2 improves the R^2 more than does Model 2.3. The residuals scatter plot for Model 2.2 (Figure 11-27) shows a more reasonably linear relationship than Model 2.3 (Figure 11-28). Figure 11-27 also shows that Model 2.2 does a better job of improving the pattern of variance comparing with Model 2.3. The QQ plot shows a kind of normality except for the two tails. Model 2.1 and Model 2.2 are both acceptable at this point.

11.2.2.2.2 Linear Regression Model with Dummy Variables

Figure 11-17 suggests that the relationship between CO and engine power may be somewhat different across the engine power ranges identified in the tree analysis. That is, there may be higher or lower CO emissions in different engine power operating ranges. One dummy variable is created to represent different engine power ranges identified in Figure 11-17 for use in linear regression analysis as illustrated below:

Engine power (bhp)	Dummy1
<114.355	1
≥114.355	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variable and interactions can help improve the model.

$$Y = \beta_0 + \beta_1\, engine.power^{(1/2)} + \beta_2\, dummy1 + \beta_3\, dummy1\ engine.power^{(1/2)} + Error \qquad (2.4)$$

The regression yields the results shown in Table 11-20 and Figure 11-29.

Table 11-20 Regression Result for CO Model 2.4

```
*** Linear Model ***

Call: lm(formula = log.CO ~ engine.power^(1/2) + dummy1 * engine.power^(1/2), data =
busdata10242006.1.4, na.action = na.exclude)
Residuals:
    Min      1Q   Median     3Q  Max
 -2.714 -0.2081 -0.01473 0.2136 2.37

Coefficients:
                                Value Std. Error   t value  Pr(>|t|)
                (Intercept)   -2.6690    0.0250  -106.5896    0.0000
      I(engine.power^(1/2))    0.0772    0.0019    41.2399    0.0000
                     dummy1    0.3472    0.0254    13.6516    0.0000
I(engine.power^(1/2)):dummy1   -0.0338    0.0020   -17.0016    0.0000

Residual standard error: 0.3836 on 39214 degrees of freedom
Multiple R-Squared: 0.2936
F-statistic: 5432 on 3 and 39214 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: log.CO

Terms added sequentially (first to last)
                                Df Sum of Sq  Mean Sq  F Value Pr(F)
      I(engine.power^(1/2))      1  2337.466 2337.466 15881.03     0
                     dummy1      1    18.325   18.325   124.50     0
I(engine.power^(1/2)):dummy1     1    42.545   42.545   289.05     0
                   Residuals 39214  5771.754    0.147
```

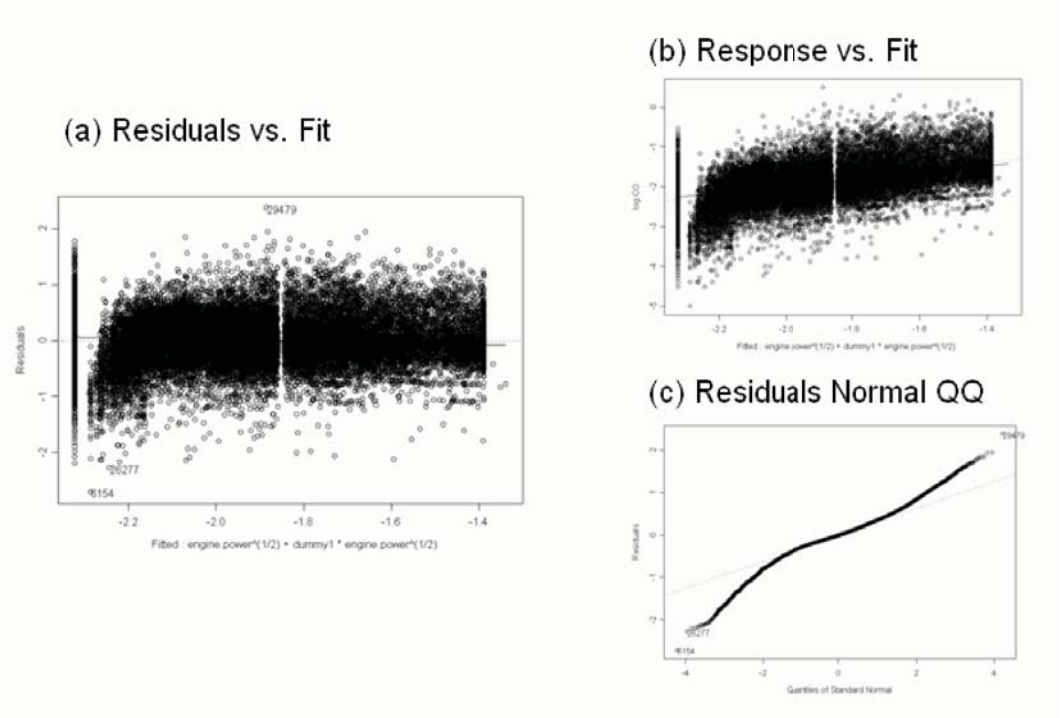


Figure 11-29 QQ and Residual vs. Fitted Plot for CO Model 2.4

Model 2.4 improves R^2 only marginally and retains the amount of variance explained in truncated transformed CO at about 29%, same as Model 2.1 and Model 2.2. Model 2.4 slightly improves R^2 more than does Model 2.2. The residuals scatter plot for Model 2.4 (Figure 11-29) shows a reasonably linear relationship. Figure 11-29 also shows that Model 2.4 does a good job of improving the pattern of variance. QQ plot shows general normality with the exceptions arising in the tails. These three models (Model 2.1, Model 2.2, and Model 2.4) are all acceptable.

11.2.2.2.3 Model Discussion

The previous sections outline the model development process from a regression tree model, to a simple OLS model, to more complex OLS models. Since the performance of the models is evaluated by comparing model predictions and actual observations for emission rates, the R^2 and slope are different from those in previous linear regression models. The results of each step in the model improvement process are presented in Table 11-21. The mean emission rates at HTBR end nodes (single value) are compared to the results of various linear regression functions with engine power. Since the R^2 and slope in Table 11-21 are derived by comparing model predictions and actual observations for emission rates (untransformed y), these numbers are different from those encountered in linear regression models.

Table 11-21 Comparative Performance Evaluation of CO Emission Rate Models

	Coefficient of determination (R^2)	**Slope (β_1)**	**RMSE**	**MPE**
Mean ERs	0.000005	1.000	0.047559	0.0000002
Linear regression (power)	0.0880	1.422	0.04622	0.00749
Linear regression ($power^{0.5}$)	0.0899	1.984	0.04662	0.00804
Linear regression (log(power))	0.0659	2.560	0.04736	0.00866
Linear regression ($power^{0.5}$) w/dummy variables	0.0915	1.657	0.04634	0.00777

The improvement in R^2 associated with moving toward a linear function of engine power is significant. Hence, the use of the linear regression function will provide a significant improvement in spatial and temporal model prediction capability. However, this linear regression function might still be improved.

Results suggest that a linear regression function with square root transformation performs slightly better than the others and that the use of dummy variables can further improve model performance. However, given the marginal improvement in R^2, one could argue that use of the engine power may be just as reasonable considering the slope, RMSE, and MPE. Although the

linear regression function with dummy variables performs slightly better than other linear regression models, more explanatory variables (dummy variables and the interaction with engine power) are introduced and the complexity of regression model increases. As discussed in Section 11.2.2.1, there is no compelling reason to include the dummy variables in the model, given that: 1) the second model is more complex without significantly improving model performance, and 2) there is no compelling engineering reason at this time to support the difference in model performance within these specific power regions. These dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis.

Considering all four parameters together, Model 2.1 is recommended as the preliminary 'final' model. The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform to a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have uniform variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying patterns in the residuals. Figure 11-26 plot (b) shows this plot for CO Model 2.1. Without considering variance due to high emission points and zero load data, there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error Terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 11-26 plot (c) shows the normal quantile plot of CO model 2.1. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm \sqrt{MSE}$ and about 90 percent to fall between $\pm 1.645 \sqrt{MSE}$. Actually, 95.20% of residuals fall within the first limits, while 96.97% of residuals fall within the second limits. Thus the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but these tails are due to the nature of the data set. For example, even after the transformation, the response variable is not the real normal distribution.

Based on the above analysis, the final CO emission rate model for the cruise mode is:

$$CO = 10^{(-2.223+0.0033 engine.power)}$$

11.2.2.3 HC Emission Rate Model Development for Cruise Mode

Based on previous analysis, truncated transformed HC will serve as the independent variable. However, modelers should keep in mind that the comparisons should always be made on the original untransformed scale of Y when comparing statistical models. Previous analysis results suggest that engine power is the best one to begin with.

11.2.2.3.1 Linear Regression Model with Engine Power

Let's select engine power to begin with, and estimate the model:

$$Y = \beta_0 + \beta_1 engine.power + Error \qquad (3.1)$$

The regression run shows the results in Table 11-22 and Figure 11-30.

Table 11-22 Regression Result for HC Model 3.1

```
Call: lm(formula = HC.25 ~ engine.power, data = busdata10242006.1.4, na.action =
na.exclude)
Residuals:
    Min      1Q     Median      3Q    Max
 -0.123 -0.0212 0.00002295 0.02228 0.3279

Coefficients:
                Value Std. Error  t value Pr(>|t|)
 (Intercept)   0.1769   0.0003   537.0480   0.0000
engine.power   0.0001   0.0000    43.0656   0.0000

Residual standard error: 0.04248 on 38018 degrees of freedom
Multiple R-Squared: 0.04651
F-statistic: 1855 on 1 and 38018 degrees of freedom, the p-value is 0

Correlation of Coefficients:
             (Intercept)
engine.power -0.7501

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                Df Sum of Sq  Mean Sq  F Value Pr(F)
engine.power     1   3.34748 3.347484 1854.647     0
   Residuals 38018  68.61934 0.001805
```

The results suggest that engine power explains about 5% of the variance in truncated transformed HC. F-statistic shows that $\beta_1 \neq 0$, and the linear relationship is statistically significant. To evaluate the model, the normality is examined in the QQ plot and constancy of variance is checked by examining residuals vs. fitted values.

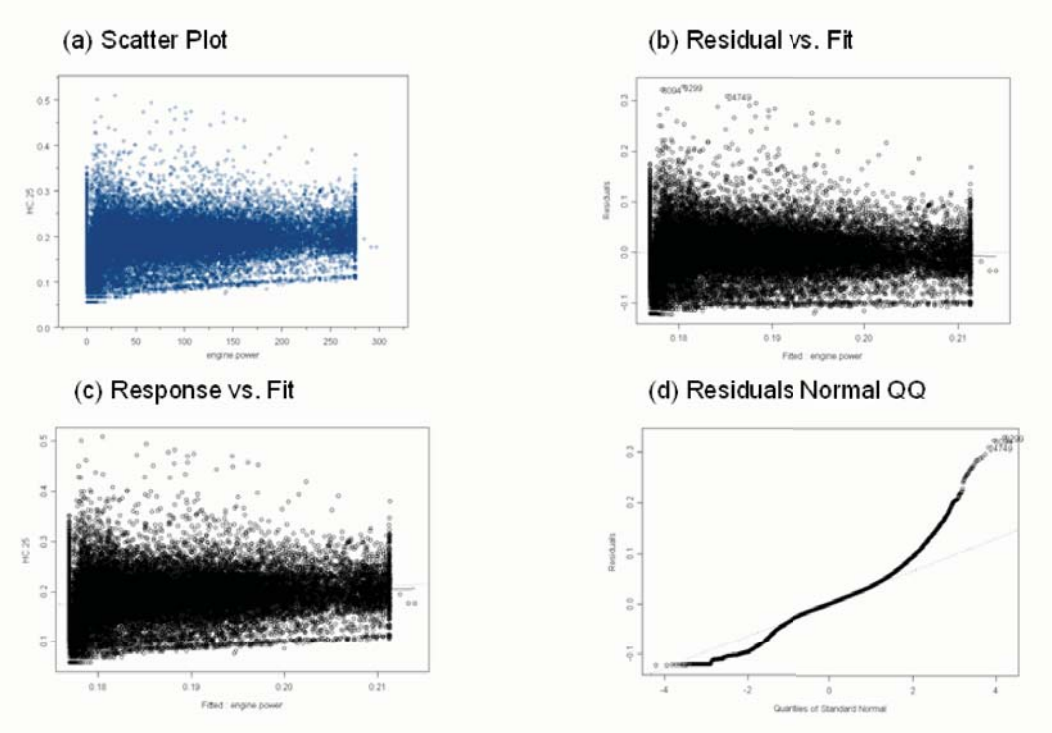


Figure 11-30 QQ and Residual vs. Fitted Plot for HC Model 3.1

The residual plot in Figure 11-30 shows a slight departure from linear regression assumptions indicating a need to explore a curvilinear regression function. Since the variability at the different X levels appears to be fairly constant, a transformation on X is considered. The reason to consider transformation first is to avoid multicollinearity brought about by adding the second-order of X. Based on the prototype plot in Figure 11-30, the square root transformation and logarithmic transformation are tested. Scatter plots and residual plots based on each transformation should then be prepared and analyzed to determine which transformation is most effective.

$$Y = \beta_0 + \beta_1 engine.power^{(1/2)} + Error \qquad (3.2)$$

$$Y = \beta_0 + \beta_1 log10(engine.power+1) + Error \qquad (3.3)$$

The results for Model 3.2 are shown in Table 11-23 and Figure 11-31, while the results for Model 3.3 are shown in Table 11-24 and Figure 11-32.

Table 11-23 Regression Result for HC Model 3.2

```
Call: lm(formula = HC.25 ~ engine.power^(1/2), data = busdata10242006.1.4, na.action
       = na.exclude)
Residuals:
     Min       1Q     Median      3Q    Max
 -0.1233 -0.02113 -0.0002419 0.02195 0.3266

Coefficients:
                          Value Std. Error  t value Pr(>|t|)
         (Intercept)   0.1700   0.0004   396.7451   0.0000
I(engine.power^(1/2))   0.0022   0.0000    47.6385   0.0000

Residual standard error: 0.04227 on 38018 degrees of freedom
Multiple R-Squared: 0.05633
F-statistic: 2269 on 1 and 38018 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                       (Intercept)
I(engine.power^(1/2)) -0.8625

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                          Df Sum of Sq  Mean Sq  F Value Pr(F)
I(engine.power^(1/2))     1   4.05395 4.053948 2269.422     0
            Residuals 38018  67.91288 0.001786
```

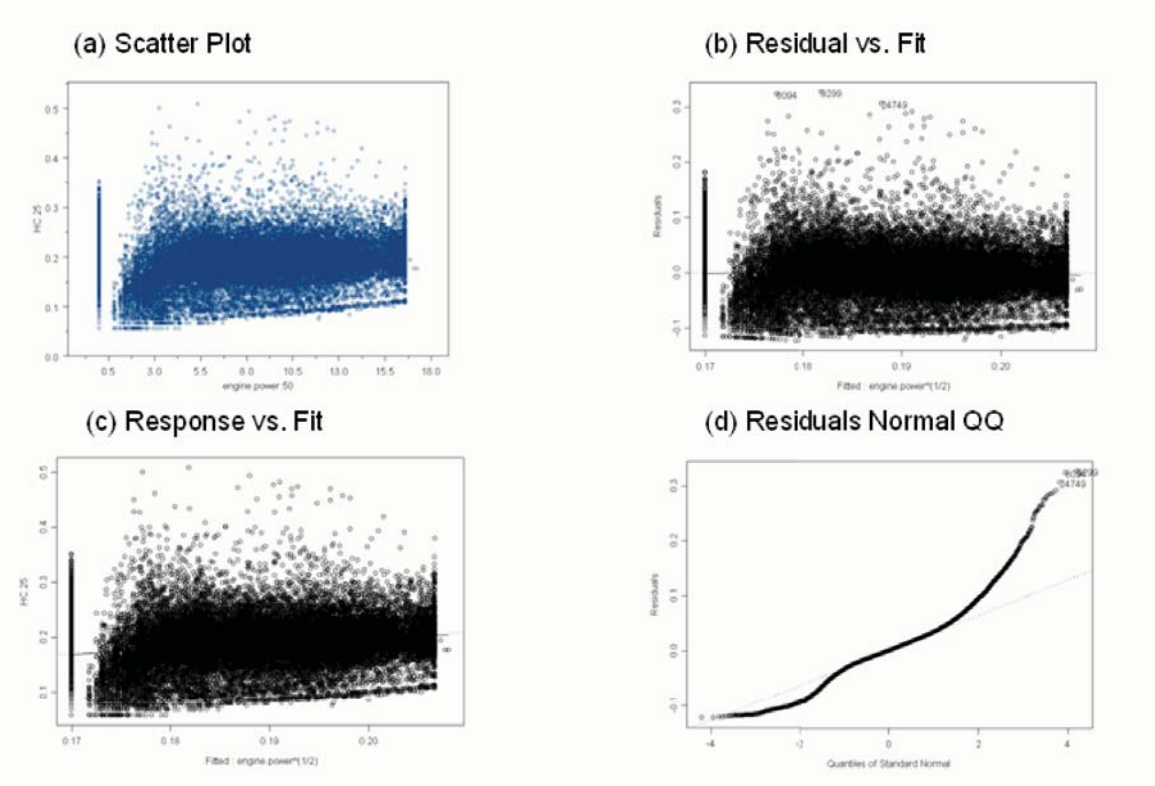


Figure 11-31 QQ and Residual vs. Fitted Plot for HC Model 3.2

Table 11-24 Regression Result for HC Model 3.3

```
Call: lm(formula = HC.25 ~ log10(engine.power + 1), data = busdata10242006.1.4,
       na.action = na.exclude)
Residuals:
    Min       1Q     Median      3Q    Max
 -0.127 -0.02073 -0.0003198 0.02203 0.3226

Coefficients:
                           Value Std. Error  t value Pr(>|t|)
            (Intercept)   0.1653   0.0005   313.2136   0.0000
log10(engine.power + 1)   0.0139   0.0003    46.4046   0.0000

Residual standard error: 0.04233 on 38018 degrees of freedom
Multiple R-Squared: 0.05361
F-statistic: 2153 on 1 and 38018 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                          (Intercept)
log10(engine.power + 1) -0.9114

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                           Df Sum of Sq  Mean Sq F Value Pr(F)
log10(engine.power + 1)     1   3.85779 3.857786 2153.39     0
              Residuals 38018  68.10904 0.001791
```

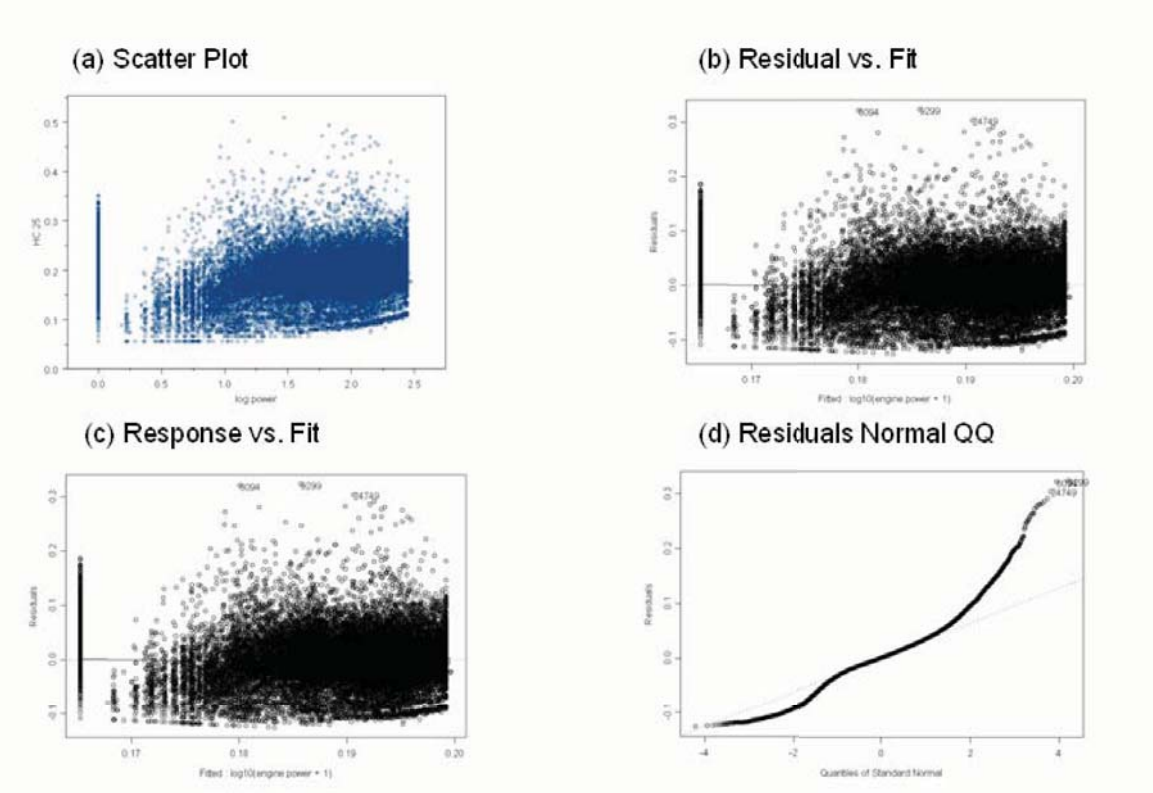


Figure 11-32 QQ and Residual vs. Fitted Plot for HC Model 3.3

The results suggest that by using transformed engine power, the model retains the amount of variance explained in truncated transformed HC at about 5% (Model 2.2 and Model 2.3). The improvement is very small.

Model 3.2 improves R^2 relative to Model 3.3. The scatter plot for Model 3.2 (Figure 11-31) also shows a better linear relationship than Model 3.3 (Figure 11-32). Figure 11-31 also shows that Model 3.2 does a good job of improving the pattern of variance. The QQ plot shows general normality with the exceptions arising in the tails.

11.2.2.3.2 Linear Regression Model with Dummy Variables

Figure 11-21 suggests that the relationship between HC and engine power may differ across the engine power ranges. One dummy variable is created to represent different engine power ranges identified in Figure 11-21 for use in linear regression analysis as illustrated below:

Engine power (bhp)	Dummy1
< 15.335	1
≥ 15.335	0

This dummy variable and the interaction between dummy variable and engine power are then tested to determine whether the use of the variable and interaction can help improve the model.

$$Y = \beta_0 + \beta_1 \log_{10}(engine.power+1) + \beta_2\, dummy1 + \beta_3\, dummy1 \log_{10}(engine.power+1) + Error \qquad (3.4)$$

The regression run shows the results in Table 11-25 and Figure 11-33.

Table 11-25 Regression Result for HC Model 3.4

```
Call: lm(formula = HC.25 ~ log10(engine.power + 1) + dummy1 * log10(engine.power +
1), data = busdata10242006.1.4, na.action = na.exclude)
Residuals:
     Min      1Q     Median      3Q     Max
 -0.1292 -0.0209 -0.0007262 0.02123 0.3423

Coefficients:
                                  Value Std. Error   t value  Pr(>|t|)
                    (Intercept)    0.1695    0.0015   109.7632    0.0000
       log10(engine.power + 1)    0.0124    0.0008    15.7058    0.0000
                        dummy1    0.0022    0.0017     1.3388    0.1807
dummy1:log10(engine.power + 1)   -0.0249    0.0012   -20.1153    0.0000

Residual standard error: 0.04184 on 38016 degrees of freedom
Multiple R-Squared: 0.07514
F-statistic: 1030 on 3 and 38016 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: HC.25

Terms added sequentially (first to last)
                                  Df Sum of Sq  Mean Sq  F Value Pr(F)
       log10(engine.power + 1)     1   3.85779 3.857786 2203.411     0
                        dummy1     1   0.84128 0.841276  480.503     0
dummy1:log10(engine.power + 1)     1   0.70843 0.708425  404.624     0
                     Residuals 38016  66.55934 0.001751
```

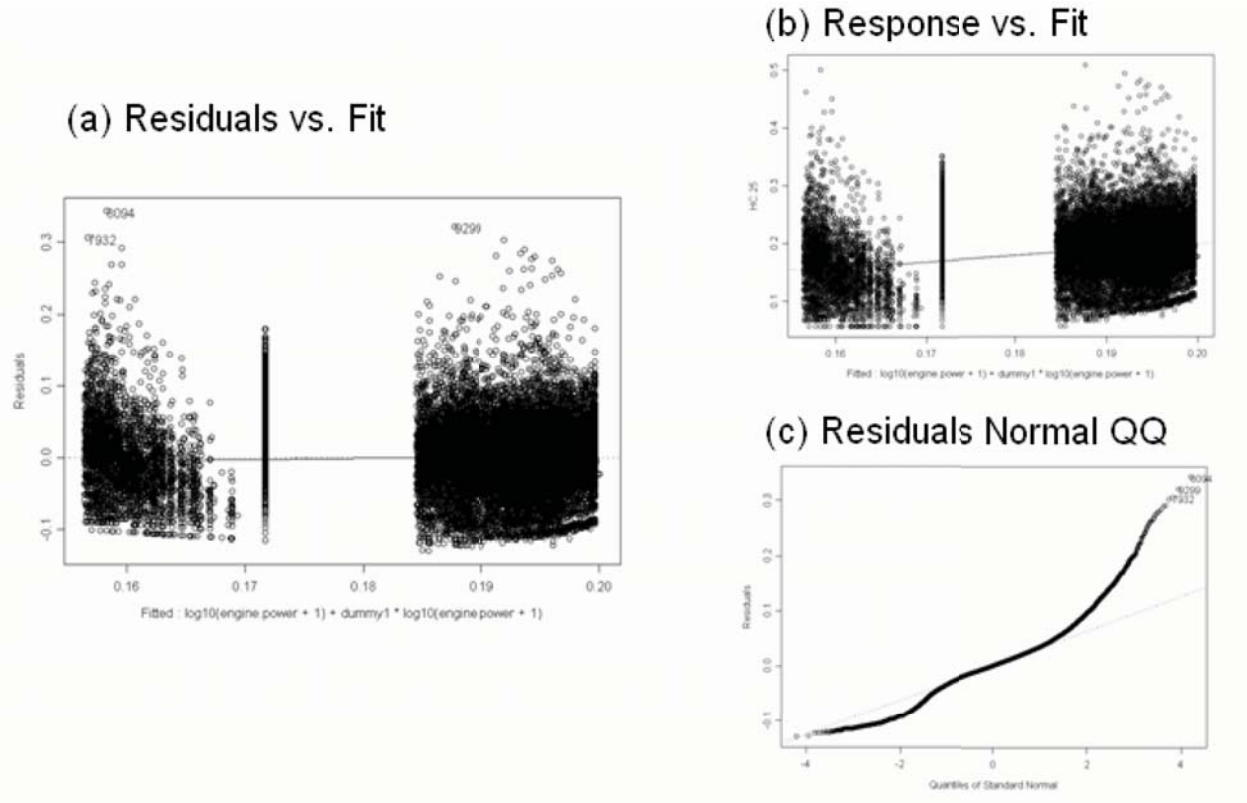


Figure 11-33 QQ and Residual vs. Fitted Plot for HC Model 3.4

The results suggest that by using dummy variables and interactions with transformed engine power, the model only increases the amount of variance explained in truncated transformed HC from about 5% to about 8%.

Model 3.4 slightly improved R^2 relative to Model 3.2. The F-statistic shows that all β values are not equal to zero, and the linear relationship is statistically significant. The gap in the residuals plot may be shifted regarding the intercept and slope by the difference of two regression functions.

11.2.2.3.3 Model Discussion

The previous sections outline the model development process from regression tree model, to a simple OLS model, to more complex OLS models. Since the performance of the models is evaluated by comparing model predictions and actual observations for emission rates, the R^2 and slope are different from those in previous linear regression models. To test whether the linear regression with power was a beneficial addition to the regression tree model, the mean ERs at HTBR end nodes (single value) are compared to the predictions from the linear regression function with engine power. The results of the performance evaluation are shown in Table 11-26. The improvement in R^2 associated with moving toward a linear function of engine power is nearly imperceptible. Hence, the use of the linear regression function will provide almost no significant improvement in spatial and temporal model prediction capability. This linear regression function might still be improved. Since the R^2 and slope in Table 11-26 are derived by comparing model predictions and actual observations for emission rates (untransformed y), these numbers are different from the results obtained from linear regression models.

Table 11-26 Comparative Performance Evaluation of HC Emission Rate Models

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
Mean ERs	0.00002	1.000	0.0020519	0.0000003
Linear regression (power)	0.00766	0.886	0.0020984	0.00047397
Linear regression ($power^{0.5}$)	0.00912	0.724	0.0020845	0.00040936
Linear regression (log(power))	0.00950	0.820	0.0020831	0.00040857
Linear regression (log(power)) w/dummy variables	0.00939	-1.142	0.0022933	0.00097449

Results suggest that the linear regression function with log transformation performs slightly better than the others and that the use of dummy variables can further improve model performance, but again there is almost no perceptible change in terms of explained variance. Although the linear regression function with log transformation and dummy variables performs slightly better than the linear regression function with square root transformation alone, the

revised model introduces additional explanatory variables (dummy variables and the interaction with engine power) and increases the complexity of the regression model without significantly improving the model. As discussed in Section 11.2.2.1, there is no compelling reason to include the dummy variables in the model, given that: 1) the second model is more complex without significantly improving model performance, and 2) there is no compelling engineering reason at this time to support the difference in model performance within these specific power regions. These dummy variables are, however, worth exploring when additional data from other engine technology groups become available for analysis.

Model 3.2 is recommended as the preliminary "final" model (although one might argue that using the regression tree results directly would also probably be acceptable). The next step in model evaluation is to once again examine the residuals for the improved model. A principal objective was to verify that the statistical properties of the regression model conform ta a set of properties of least squares estimators. In summary, these properties require that the error terms be normally distributed, have a mean of zero, and have uniform variance.

Test for Constancy of Error Variance

A plot of the residuals versus the fitted values is useful in identifying any patterns in the residuals. Figure 11-31 plot (c) shows this plot for HC Model 3.2. Without considering variance due to high emission points and zero load data, there is no obvious pattern in the residuals across the fitted values.

Test of Normality of Error terms

The first informal test normally reserved for the test of normality of error terms is a quantile-quantile plot of the residuals. Figure 11-31 plot (d) shows the normal quantile plot of the HC model. The second informal test is to compare actual frequencies of the residuals against expected frequencies under normality. Under normality, we expect 68 percent of the residuals to fall between $\pm\sqrt{MSE}$ and about 90 percent to fall between $\pm$ 1.645 $\sqrt{MSE}$. Actually, 95.20% of residuals fall within the first limits, while 96.99% of residuals fall within the second limits. Thus, the actual frequencies here are reasonably consistent with those expected under normality. The heavy tails at both ends are a cause for concern, but are due to the nature of the data set. For example, even after the transformation, the response variable is not the real normal distribution.

The final HC emission rate model selected for cruise mode is:

$$HC = [0.170 + 0.0022(\text{engine.power})^{(1/2)}]^4$$

11.3 Conclusions and Further Considerations

In this research, engine power is used as the main explanatory variable to develop cruise emission rate models. The explanatory ability of engine power varies by pollutant. In general, the relationship between NOx and engine power is more highly correlated than the other two pollutants.

Inter-bus variability analysis indicated that some of the 15 buses are higher emitters that others (especially noted for HC emissions). However, none of the buses appear to qualify as traditional high-emitters, which would exhibit emission rates of two to three standard deviations above the mean. Hence, it is difficult to classify any of these 15 buses as high emitters for modeling purposes. At this point, these 15 buses are treated as a whole data set for model development. Modelers should keep in mind that although no true high-emitters are present in the database, such vehicles may behave significantly differently than the vehicles tested. Hence, data from high-emitting vehicles should be collected and examined in future studies.

Some high HC emissions events are noted in cruise mode. After screening engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature, ECM pressure, and other parameters, no variables were identified that could be linked to these high emissions events. These events may represent natural variability in onroad emissions, or some other variable (such as grade or an engine variable that is not measured) may be linked to these events.

Engine power is selected as the most important variable for three pollutants based on HTBR tree models. This finding is consistent with previous research results which verified the important role of engine power (Ramamurthy et al. 1998; Clark et al. 2002; Barth et al. 2004). The noted HC relationship is significant but fairly weak. Analysis in previous chapters also indicates that engine power is correlated with not only onroad load parameters such as vehicle speed, acceleration, and grade, but also potentially correlated with engine operating parameters such as throttle position and engine oil pressure. On the other hand, engine power in this research is derived from engine speed, engine torque and percent engine load.

The regression tree models still suggest that some other variables, like oil pressure and engine barometric pressure, may also impact the HC emissions. Further analysis demonstrates that by using engine power alone one might be able to achieve similar explanatory ability as opposed to using engine power and other variables. To develop models that are efficient and easy to implement, only engine power is used to develop emission models. However, additional investigation into these variables is warranted as additional detailed data from engine testing become available for analysis.

Given the relationships noted between engine indicated HP and emission rates, it is imperative that data be collected to develop solid relationships in engine power demand models (estimating power demand as a function of speed/acceleration, grade, vehicle characteristics, surface roughness, inertial losses, etc.) for use in regional inventory development and microscale impact assessment.

In summary, the cruise emission rate models selected for implementation are:

$$NOx = [0.0087+0.0311\ (engine.power)^{(1/2)}]^2$$

$$CO = 10\text{^}(-2.223+0.0033 engine.power)$$

$$HC = [0.170+0.0022\ (engine.power)^{(1/2)}]^4$$

CHAPTER 12

12. MODEL VERIFICATION

In the previous chapters, three statistically-derived modal emission rate models were developed for use in predicting emissions of NO_x, CO and HC from transit buses. This chapter discusses the reasons for using engine power instead of surrogate power variables in emission rate modeling, the necessity of developing a linear regression model rather than using mean emission rates, the need to introduce driving mode with load modeling, the possibility of combining acceleration and cruise modes, and other issues.

12.1 Engine Power vs. Surrogate Power Variables

The first step towards verifying the model is to compare the explanatory power of real load data and surrogate power variables. Different approaches have been proposed by several researchers. The MOVES model employs vehicle specific power (VSP), defined as instantaneous power per unit mass of the vehicle (Jimenez-Palacios 1999).

VSP is a measure of the road load on a vehicle, defined as the power per unit mass to overcome road grade, rolling and aerodynamic resistance, and inertial acceleration (Jimenez-Palacios 1999; U.S. EPA 2002b; Nam 2003; Younglove et al. 2005):

$$VSP = v*(a*(1+\gamma)+g*grade+g*C_R)+0.5\rho*C_D*A*v^3/m$$

where:

v:	vehicle speed (assuming no headwind) in m/s
a:	vehicle acceleration in m/s2
γ:	mass factor accounting for the rotational masses (~0.1)
g:	acceleration due to gravity
grade:	road grade
C_R:	rolling resistance (~0.0135)
ρ:	air density (1.2)
C_D:	aerodynamic drag coefficient
A:	the frontal area
M:	vehicle mass in metric tons

Using typical values for coefficients, in SI units the equation becomes (CDA/m ~ 0.0005) (Younglove et al. 2005):

$$VSP\left(kW/metricTon\right)=V\times(1.1\times a+9.81\times grade(\%)+0.132)+0.001208\times v^3$$

The VSP approach to emission characterization was developed by several researchers (Jimenez-Palacios 1999; U.S. EPA 2002b; Nam 2003; Younglove et al. 2005) and further developed as part of the MOVES model. The coefficients used to estimate VSP were different in previous research because of the choice of typical values of coefficients. However, the coefficients given in the above equation are specific for light-duty vehicles. For example, a mass factor of 0.1 is not suitable to describe the transit bus characteristics of inertial loss. This surrogate power variable (VSP) is not suitable to compare with engine load data for this study. First, the implementation approach that is used in MOVES is based upon VSP bins, and not on instantaneous VSP. Second, the coefficients given in the above equation are specific for light-duty vehicles, not for transit buses.

Other research efforts have used surrogate power variables such as the inertial power surrogate, defined as acceleration times velocity, and drag power surrogate, defined as acceleration times velocity squared (Fomunung 2000). Barth and Frey also used acceleration times velocity for power demand estimation (Barth and Norbeck 1997; Frey et al. 2002). Both surrogate variables for power demand can be used to compare NO_x in cruise mode. Using surrogate variables instead of real load data, the model is:

$$Y = \beta_0 + \beta_1\, acceleration + \beta_2\, vehicle.speed + \beta_3\, vehicle.speed*acceleration + \beta_4\, vehicle.speed^2*acceleration + Error \quad (1)$$

The regression run shows the results in Table 12-1 and Figure 12-1.

Table 12-1 Regression Result for NOx Model 1

```
Call: lm(formula = NOx.50 ~ vehicle.speed * acceleration + vehicle.speed^2:
       acceleration, data = busdata10242006.1.4, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q   Max
 -0.4779 -0.08625 0.001824 0.08759 1.338

Coefficients:
                                    Value Std. Error   t value  Pr(>|t|)
                     (Intercept)    0.1996    0.0018   113.0559    0.0000
                   vehicle.speed    0.0043    0.0001    77.4369    0.0000
                    acceleration    0.0738    0.0052    14.2957    0.0000
     vehicle.speed:acceleration     0.0066    0.0004    15.5704    0.0000
acceleration:I(vehicle.speed^2)    -0.0001    0.0000   -13.7590    0.0000

Residual standard error: 0.1323 on 39369 degrees of freedom
Multiple R-Squared: 0.3708
F-statistic: 5801 on 4 and 39369 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                                (Intercept) vehicle.speed acceleration
                   vehicle.speed -0.9243
                    acceleration  0.0796     -0.0590
     vehicle.speed:acceleration -0.0825      0.0569       -0.9114
acceleration:I(vehicle.speed^2)  0.0782     -0.0593        0.7978

                                vehicle.speed:acceleration
                   vehicle.speed
                    acceleration
     vehicle.speed:acceleration
acceleration:I(vehicle.speed^2) -0.9678

Analysis of Variance Table

Response: NOx.50

Terms added sequentially (first to last)
                                   Df Sum of Sq  Mean Sq  F Value Pr(F)
                   vehicle.speed    1  122.5215 122.5215  6999.67     0
                    acceleration    1  278.9165 278.9165 15934.55     0
     vehicle.speed:acceleration     1    1.4036   1.4036    80.19     0
acceleration:I(vehicle.speed^2)     1    3.3136   3.3136   189.31     0
                       Residuals 39369  689.1106   0.0175
```

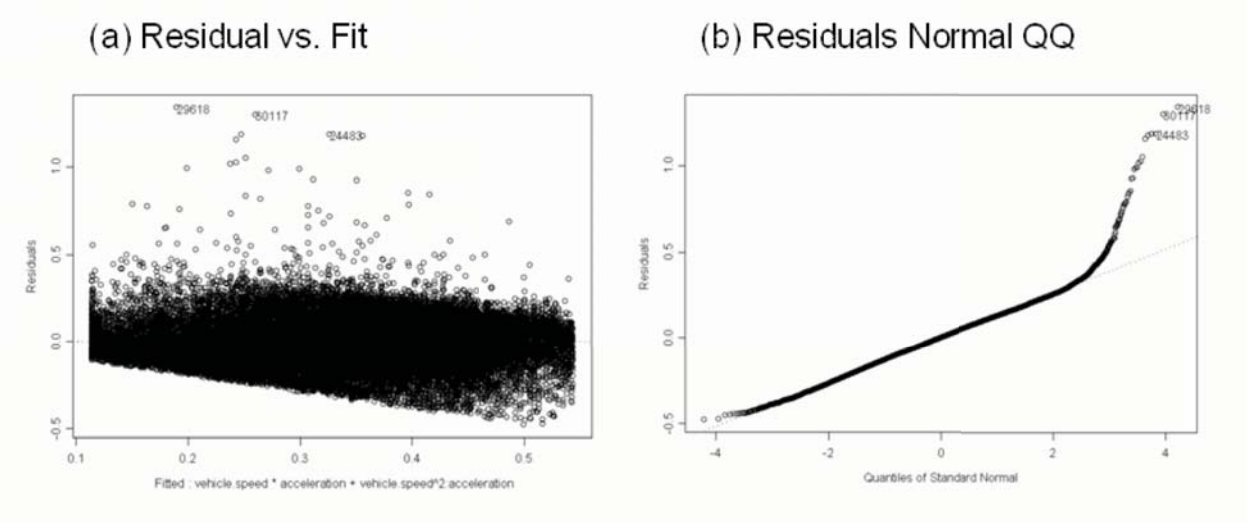


Figure 12-1 QQ and Residual vs. Fitted Plot for NO_x Model 1

The results suggest that the surrogate variable model can explain about 37 % of the variance in truncated transformed NO_x, whereas the OLS model developed in Chapter 10 explained more than 75% of the cruise mode variance. Considering the theoretical equation of engine power presented much earlier in Chapter 3, the surrogate variables can only represent some, and not all, of the components of engine power. Given the importance of engine power in explaining the variability of emissions, it is essential that field data collection efforts include the measurement of indicated load data as well as all of the operating conditions necessary to estimate bhp load when second-by-second emission rate data are collected.

12.2 Mean Emission Rates vs. Linear Regression Model

The modeling approach employed in this research involved the separation of data into separate driving modes for analysis and then applying modeling techniques to derive emission rates as a function of engine load. Although constant emission rates in grams/second were adequate for idle, motoring, and non-motoring deceleration modes, modeling efforts in Chapters 10 and 11 demonstrated that a linear regression function should improve spatial and temporal model prediction capability significantly for acceleration and cruise modes. However, one verification comparison that should be undertaken is on the overall benefit of introducing engine load into the modeling regime vs. simply using average emission rate values for each operating mode. This comparison will provide insight into the overall effect of introducing engine load (even though it is only introduced into acceleration and cruise modes).

There are a number of model goodness-of-fit criteria that can be used to assess the difference between the emissions predicted by the load-based modal emission rate model and the mode-only emission rate models. Normally, one would compare the alternative model perfor-

mance for an independent set of data collected from similar vehicles, which is currently not available. Alternatively, model developers would set aside a significant subset of the data in the model development data set so that the data are not used in model development and instead used in model comparisons. However, there were not enough data available to do this. Hence, at this time, the only comparisons that can be made are for alternative model performance using the same data that were used to develop the models presented in this research effort.

The performance of the models is first evaluated by comparing model predictions and actual observations for emission rates. The performance of the model can be evaluated in terms of precision and accuracy (Neter et al. 1996). The R^2 value is an indication of precision. Usually, higher R^2 values imply a higher degree of precision and less unexplained variability in model predictions than lower R^2 values. The slope of the trend line for the observed versus predicted values is an indication of accuracy. A slope of one indicates an accurate prediction, in that the prediction of the model corresponds to an observation.

The model's predictive ability is also evaluated using the root mean square error (RMSE) and the mean prediction error (MPE) (Neter et al. 1996). The RMSE is a measure of prediction error. When comparing two models, the model with a smaller RMSE is a better predictor of the observed phenomenon. Ideally, mean prediction error is close to zero. RMSE and MPE are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{Equation (12-1)}$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \qquad \text{Equation (12-2)}$$

where:

RMSE:	=	root mean square error
n:	=	number of observations
y_i:	=	observaton y
$\bar{y}_i$:	=	mean of observation y
MPE:	=	mean predictive error

To test whether the linear regression with power was a beneficial addition to the regression tree model, the mean ERs at HTBR end nodes (single value) are compared to the predictions from the linear regression function with engine power. The results of the performance evaluation are shown in Table 12-2.

Table 12-2 Comparative Performance Evaluation between Mode-Only Models and Linear Regression Models

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
NO_x				
Mean ERs	0.438	1.000	0.08725	0.000002
Linear Regression	0.665	1.102	0.07122	0.021463
CO				
Mean ERs	0.248	1.000	0.07406	-0.000004
Linear Regression	0.491	1.749	0.06691	0.010285
HC				
Mean ERs	0.0686	1.000	0.00190	0.0000005
Linear Regression	0.0677	1.213	0.00192	0.000223

For NO_x and CO, the R^2 values indicate that load based modal emission model performs slightly better than mean emission rates and the use of linear regression function can further improve model performance. The results shown in Table 12-2 reinforce the importance of introducing linear regression functions in acceleration and cruise mode. For HC, there is no discernible difference in model performance. Combining this finding with the performance results for HC noted in Chapters 8 through 11, using constant emission rates for each operating mode could be justified for this data set. When additional data are collected, researchers should compare mean emission rates approaches to power-based approaches to ensure that power demand models for HC are necessary.

12.3 Mode-specific Load Based Modal Emission Rate Model vs. Emission Rate Models as a Function of Engine Load

Modal modeling approaches are becoming widely accepted as more accurate in making realistic estimates of mobile source contributions to local and regional air quality. Research at Georgia Tech has clearly identified that modal operation is a better indicator of emission rates than average speed (Bachman 1998). The analysis of emissions with respect to driving modes, also referred to as modal emissions, has been performed in recent research studies (Barth et al. 1996; Bachman 1998; Fomunung et al. 1999; Frey et al. 2002; Nam 2003; Barth et al. 2004). These studies indicated that driving modes might have the ability to explain a certain portion of the variability in emissions data. In Chapters 10 and 11, emission rates were derived as a function of driving mode (cruise, idle, acceleration, and deceleration operations) and engine power because previous research efforts had separately suggested that vehicle emission rates were

highly correlated with modal activity and engine power. In this research, five driving modes are introduced in total: idle mode, deceleration motoring mode, revised deceleration mode, acceleration mode, and cruise mode.

Chapters 10 and 11 did not compare the combined modal and engine power models to models that use power alone to predict emission rates. To test the effect of adding driving modes in the emission rate model, the derivation of a load-only model for NO_x emissions is illustrated in detail. Load-only CO emissions models and HC emissions models are also derived for comparison purposes and presented in final form (however, the detailed regression plots and tables are omitted for the purposes of brevity).

As in previous chapters, the first step for a load based only model is to select the most important variable for NO_x emissions. When using the entire database at once (data are not broken into mode subsets for this derivation), the appropriate transformation for NO_x is ¼ based on Box-Cox results, rather than the ½ value used in developing models for acceleration and cruise mode (see Chapters 10 and 11). The trimmed HTBR tree models for NO_x are illustrated in Figure 12-2 and Table 12-3.

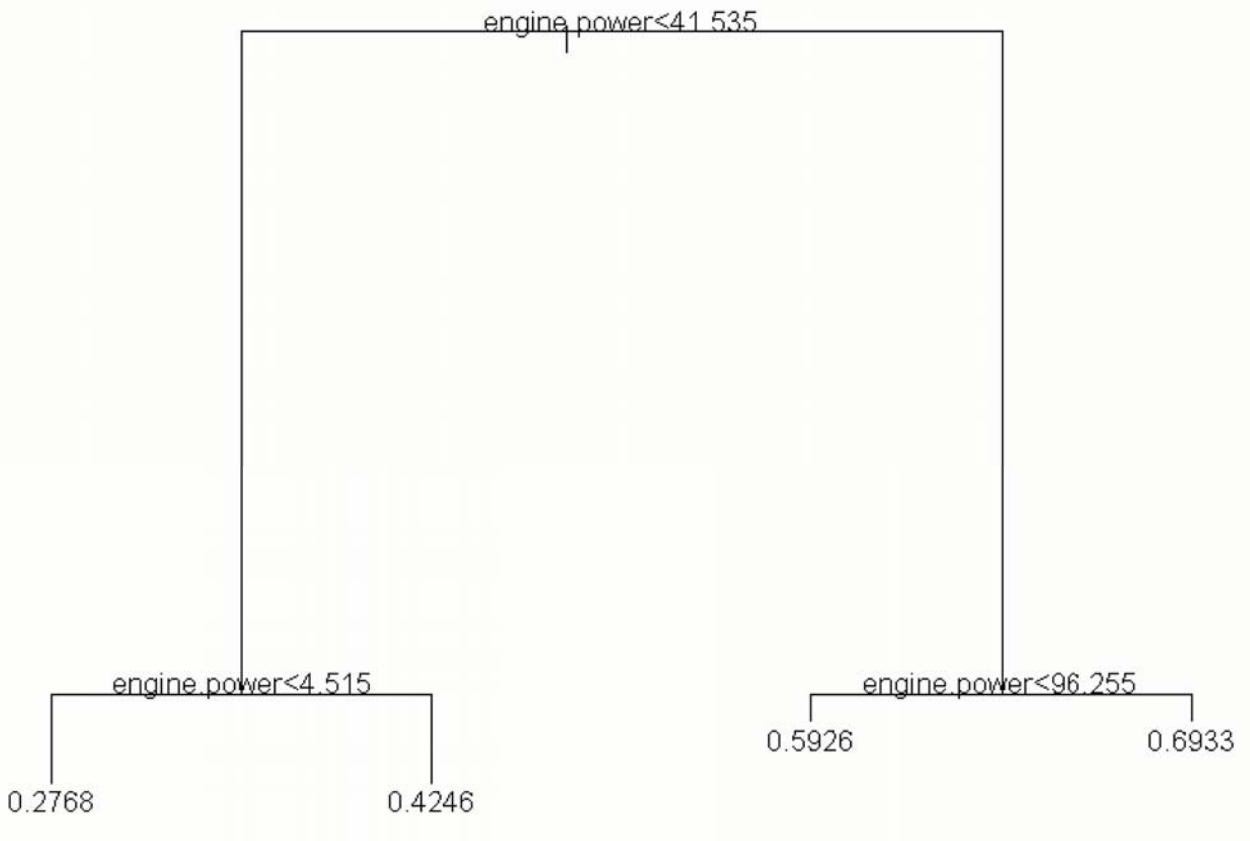


Figure 12-2 Trimmed Regression Tree Model for Truncated Transformed NO_x

Table 12-3 Trimmed Regression Tree Results for Truncated Transformed NO_x

```
Regression tree:
tree(formula = NOx.25 ~ engine.power + vehicle.speed + acceleration +
       oil.temperture + oil.press + cool.temperature + eng.bar.press +
       model.year + odometer + bus360 + bus361 + bus363 + bus364 + bus372 +
       bus375 + bus377 + bus379 + bus380 + bus381 + bus382 + bus383 + bus384 +
       bus385 + dummy.grade, data = busdata10242006.1, na.action = na.exclude,
       mincut = 3000, minsize = 6000, mindev = 0.1)
Variables actually used in tree construction:
[1] “engine.power”
Number of terminal nodes:  4
Residual mean deviance:  0.005837 = 618.6 / 106000
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
 -5.187e-001 -4.510e-002 -9.204e-003  3.768e-016  5.004e-002  6.557e-001
node), split, n, deviance, yval
      * denotes terminal node

1) root 105976 3058.00 0.4991
  2) engine.power<41.535 62441  666.60 0.3823
    4) engine.power<4.515 17897  195.50 0.2768 *
    5) engine.power>4.515 44544  192.20 0.4246 *
  3) engine.power>41.535 43535  316.60 0.6667
    6) engine.power<96.255 11504   61.56 0.5926 *
    7) engine.power>96.255 32031  169.20 0.6933 *
```

After testing different transformations for Y and adding dummy variables according to HTBR results, Table 12-4 and Figure 12-3 show that a load based only model for NO_x emissions is a fairly good model, considering the constancy of error variance and normality of error terms. So, the final load based only model for NO_x emissions is:

$$NO_x = [0.230 + 0.195\log10(\text{engine.power}+1)]^4$$

The regression run shows the results in Table 12-4 and Figure 12-3.

Table 12-4 Regression Result for NO_x Load-Based Only Emission Rate Model

```
Call: lm(formula = NOx.25 ~ log10(engine.power + 1), data = busdata10242006.1,
       na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q   Max
 -0.4683 -0.04297 -0.01329 0.04138 0.663

Coefficients:
                           Value Std. Error  t value Pr(>|t|)
            (Intercept)   0.2303   0.0005   489.9131   0.0000
log10(engine.power + 1)   0.1950   0.0003   657.2170   0.0000

Residual standard error: 0.0754 on 105974 degrees of freedom
Multiple R-Squared: 0.803
F-statistic: 431900 on 1 and 105974 degrees of freedom, the p-value is 0

Correlation of Coefficients:
                          (Intercept)
log10(engine.power + 1) -0.8702

Analysis of Variance Table

Response: NOx.25

Terms added sequentially (first to last)
                            Df Sum of Sq  Mean Sq  F Value Pr(F)
log10(engine.power + 1)      1  2455.676 2455.676 431934.2     0
              Residuals 105974   602.494    0.006
```

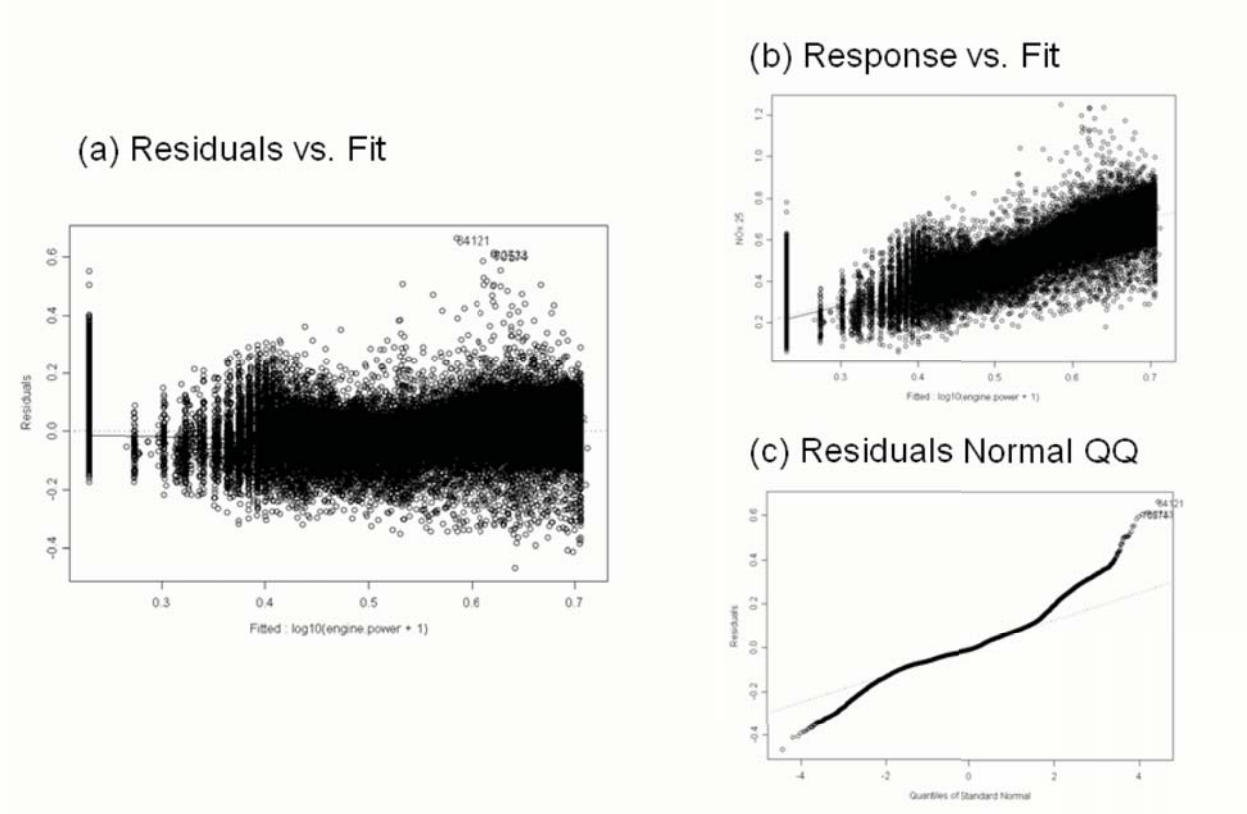


Figure 12-3 QQ and Residual vs. Fitted Plot for Load-Based Only NO_x Emission Rate Model

Following the same derivation techniques, the final load-only model for CO emissions is:

$$CO = 10^{\wedge}[-2.659 + 0.0899(engine.power)^{(1/2)}]$$

Following the same derivation techniques, the final load-only model for HC emissions is:

$$HC = 10^{\wedge}[-3.306 + 0.0382(engine.power)^{(1/2)}]$$

The performance of the load-only models relative to the combined mode and load models developed in Chapters 8 through 11 is presented in Table 12-5.

Table 12-5 Comparative Performance Evaluation Between Load-Based Only Emission Rate (ER) Model and Load-Based Modal Emission Rate Model

	Coefficient of determination (R^2)	**Slope (β_1)**	**RMSE**	**MPE**
NO_x				
Load-Only Emission Rate Model	0.715	1.181	0.06494	0.011382
Mode/Load Emission Rate Models	0.665	1.102	0.07122	0.021463
CO				
Load-Only Emission Rate Model	0.246	2.071	0.07886	0.015568
Mode/Load Emission Rate Models	0.490	1.749	0.06691	0.010285
HC				
Load-Only Emission Rate Model	0.0672	0.982	0.00197	0.000499
Mode/Load Emission Rate Models	0.0677	1.213	0.00192	0.000223

For NO_x, both models perform well in explaining the variance of emission rates, reinforcing the importance of including engine power as a variable in explaining the variance of NOx emission rates. Results suggest that a mode/load modal emission modeling approach performs slightly better than load-only emission rate models for CO. For HC, there is no discernible difference in model performance. Combining this finding with the performance results for HC noted in Chapters 8 through 11, using constant emission rates for each operating mode could be justified for this data set. When additional data are collected, researchers should compare mode-only approaches to power-based approaches to ensure that power demand models for HC are necessary.

12.4 Separation of Acceleration and Cruise Modes

In this research effort, separate models were developed for acceleration and cruise modes (Chapters 10 and 11). However, it may be possible to combine acceleration and cruise mode activity into a new "combined driving" mode. As noted in Chapter 10, although engine power distribution for acceleration mode is different from cruise mode, these two modes share a similar pattern. A quick analysis of the impact of combining acceleration and cruise mode is presented in this section.

After examining HTBR results, selecting the important explanatory variables, testing different transformations for X and Y, and adding dummy variables according to HTBR results, the final NO_x emission model for combined driving mode is:

$$NO_x = [0.113 + 0.0266(engine.power^{(1/2)}]^2$$

The final CO emission model for combined driving mode is:

$$CO = 10\text{^}[-2.238 + 0.0043(engine.power)]$$

while the final HC emission model for combined driving mode is:

$$HC = [0.167 + 0.0028(engine.power^{(1/2)}]^4$$

To test whether combining acceleration and cruise modes would benefit the load-based modal emission model, the predictions from the linear regression function for combined driving mode are compared to the predictions from sub-models for acceleration and cruise mode in the load-based modal emission model. Since the other elements are the same for two models, they will be excluded from test. The results of the performance evaluation are shown in Table 12-6.

Table 12-6 Comparative Performance Evaluation between Linear Regression with Combined Mode and Linear Regression with Acceleration and Cruise Modes

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
NO_x				
Combined Driving Mode	0.531	0.921	0.08488	0.00840
Acceleration & Cruise Mode	0.527	0.953	0.09312	0.03904
CO				
Combined Driving Mode	0.177	1.594	0.10395	0.02305
Acceleration & Cruise Mode	0.452	1.775	0.08966	0.01873
HC				
Combined Driving Mode	0.0338	0.907	0.00204	0.00042
Acceleration & Cruise Mode	0.0410	0.905	0.00203	0.00041

Results shown in Table 12-6 suggest that separate linear regression functions for acceleration and cruise modes perform significantly better than linear regression functions with combined driving mode for CO. For NO_x and HC, both models perform similarly with respect to explaining the variance of emission rates. In general, these results support introducing acceleration and cruise mode into the conceptual model. However, as new data become available for testing, researchers should examine whether it is reasonable to simply separate idle and deceleration modes from other driving modes and then apply a simple power-based model to the remaining combined driving activity for NO_x.

12.5 MOBILE6.2 vs. Load-Based Modal Emission Rate Model

The final step undertaken in the model verification process was a comparison of prediction results from MOBILE6.2 and the load-based modal emission rate model developed in this research. Comparisons are based upon the Ann Arbor transit vehicle test data. These data were used to develop the modal emission rates for this report, but were not used in developing the MOBILE6.2 model. Normally, one would compare alternative model performance using an independent set of data collected from similar vehicles, which is currently not available. Hence, the comparisons that will be presented are far from unbiased. When new data from an independent test fleet become available, these comparisons should be performed again.

To facilitate the emission rate prediction comparison, lookup tables for MOBILE6.2 transit bus emission rates on arterial roads were first created for average speeds from 2.5 mph to 65 mph. The MOBILE6.2 calendar year was set to January 2002 since the data set was collected during October 2001. The temperature was set as 75 ^oF, since the emission rates for transit buses

in MOBILE6.2 do not change with temperature. Emissions predictions from MOBILE6.2 were then obtained by combining lookup tables and corresponding speed values in the AATA data set. The results of the performance evaluation are shown in Table 12-7.

Table 12-7 Comparative Performance Evaluation between MOBILE 6.2 and Load-Based Modal ER Model

	Coefficient of determination (R^2)	Slope (β_1)	RMSE	MPE
NO_x				
MOBILE 6.2	0.172	0.706	0.10825	0.011217
Load-Based Modal ER Model	0.665	1.102	0.07122	0.021463
CO				
MOBILE 6.2	0.0195	1.690	0.08516	0.013399
Load-Based Modal ER Model	0.491	1.749	0.06691	0.010285
HC				
MOBILE 6.2	0.0408	0.584	0.00194	0.000173
Load-Based Modal ER Model	0.0677	1.213	0.00192	0.000223

Results suggest that load-based modal emission rate model performs significantly better than MOBILE6.2 for NO_x and CO, and slightly better for HC. The performance of the load-based modal emission rate model is not surprising because the same data used to develop the model are used in the comparison. Results suggest that the load-based modal emission model performs well vis-à-vis explaining the variance of NO_x and CO emission rates on a microscopic level. The slight differences in RMSE and MPE indicate that both models (MOBILE6.2 and the load-based modal emission model) perform well at the macroscopic level, and should perform similarly when used in regional inventory development.

12.6 Conclusions

In general, the results provided here are encouraging for the load based modal emission model. The comparison between engine power and surrogate power variables confirms the important role of engine power in explaining the variability of emissions. The comparison between the load-only emission rate model and the load-based modal emission rate model shows that the impact of driving mode on emissions is signficiant for NO_x and CO emissions while no such trend is discernible for HC. The comparison between acceleration and cruise modes and combined driving mode indicates that the relationships between engine power and emissions are slightly different for acceleration and cruise modes. Splitting the database into five modes (idle mode, decelerating motoring mode, deceleration non-motoring mode, acceleration mode, and cruise mode) appears warranted.

The data used to develop the load based modal emission model in this research are very limited since the data set contained only 15 transit buses. Inter-bus variability is more obvious for HC emissions since Bus 363 has the lowest HC emissions compared with the other 14 buses. This kind of variability might influence the explanatory variables of the modal emission model for HC emissions. When new data become available, these models should be re-derived to obtain further improved performance in applications to the transit bus fleet.

CHAPTER 13

13. CONCLUSIONS

The goal of this research is to provide emission rate models that fill the gap between existing models and ideal models for predicting emissions of NO_x, CO, and HC from heavy-duty diesel vehicles. The researchers at Georgia Institute of Technology have developed a beta version of HDDV-MEM (Guensler et al. 2005), which is based upon vehicle technology groups, engine emission characteristics, and vehicles modal activity. The HDDV-MEM first predicts second-by-second engine power demand as a function of onroad vehicle operating conditions and then applies brake-specific emission rates to these activity predictions. The HDDV-MEM consists of three modules: a vehicle activity module (with vehicle activity tracked by a vehicle technology group), an engine power module, and an emission rate module.

Using second-by-second data collected from onroad vehicles, the research effort reported herein developed models to predict emission rates as a function of onroad operating conditions that affect vehicle emissions. Such models should be robust and ensure that assumptions about the underlying distribution of the data are verified and that assumptions associated with applicable statistical methods are not violated. Due to the general lack of data available for development of heavy-duty vehicle modal emission rate models, this study focuses on development of an analytical methodology that is repeatable with different data sets collected across space and time. The only acceptable second-by-second data set in which emission rate and applicable load and vehicle activity data had been collected in parallel was the AATA bus emissions database collected by Sensors, Inc., for use by the U.S. EPA.

The models developed in this report are applicable to transit buses only, and are not applicable to all transit buses (see limitations discussion in Section 13.2). However, a significant contribution of the research is in the development of the analytical framework established for analysis of second-by-second emission rate data collected in parallel with engine load and other onroad operating parameters, and in the development of applicable processes for developing statistical models using such data. To demonstrate the capability of the modeling framework, three

modal emission rate models have been developed for prediction of NOx, CO and HC emissions from mid-1990s transit buses.

The AATA transit bus data set was first post-processed through a quality control/quality assurance process. Data problems were identified and corrected during this stage of the research effort. The types of errors checked include: loss of data, erroneous ECM data, GPS dropouts, and synchronization errors. Data records for which all data elements were not collected were removed to avoid any bias to the results. No erroneous ECM data were identified. Six buses experienced GPS dropouts and synchronization errors and these problems were treated as described in chapter 4. Emission rate variability was also assessed across the sample of buses to identify any potential high-emitters that may behave differently than other buses under normal operating conditions and therefore warrant separate model development. However, no high-emitters were identified. To find the true 'high-emitters', modelers need to include a representative sample of buses to try to ensure that mean emissions and response rates to operating variables are represented in the data. Since there are only 15 buses in the data set, modelers could not exclude buses that showed higher emission rates than the others.

Model development then proceeded through a structured series of steps. Transformations of emission rates (NO_x, CO, and HC) were verified through a Box-Cox procedure to improve the specific modeling assumptions, such as linearity or normality. HTBR regression tree results were used to identify the most important explanatory variables for emission rates. OLS regression models were developed for transformed emission rates using chosen explanatory variables. Dummy variables were created to represent the cut points identified in HTBR trees. Interaction effects for identified explanatory variables were also tested to see whether they could improve the model. The models were comparatively evaluated and the most efficient models for each pollutant were selected. By demonstrating statistical "robustness" and sufficiency in previous chapters, the main goal of this research, that of "developing new load-based models with significant improvement", was achieved.

This chapter will review the key accomplishments of this research. The chapter provides the final models selected for implementation and begins with a summary of the final models developed for the transit buses, followed immediately by a discussion on the limitations of these models. The chapter concludes with the lessons learned and recommendations on further research.

13.1 Transit Bus Emission Rate Models

The goal of this research was to develop a methodology for creating load-based emission rate models designed to predict emission rates of NO_x, CO, and HC from transit buses as a function of onroad operating conditions. The models should be robust and ensure that statistical assumptions in model development are not violated. With limited available data, this study developed a methodology that is repeatable with a different data set from across space and across time. The final estimated models are presented in Table 13-1.

Table 13-1 Load Based Modal Emission Models

Driving Mode	
NO_x	
Idle Mode	0.033415 g/s
Decelerating Motoring Mode	0.0097768 g/s
Deceleration Non-Motoring Mode	0.045777 g/s
Acceleration Mode	NOx = (-0.0195 + 0.201log10(engine.power + 1) + 0.0019vehicle.speed)^2
Cruise Mode	NOx = (0.0087 + 0.0311 (engine.power)$^{(1/2)}$)^2
CO	
Idle Mode	0.0059439 g/s
Decelerating Motoring Mode	0.0052857 g/s
Deceleration Non-Motoring Mode	0.0068557 g/s
Acceleration Mode	CO = 10^(-3.747 + 1.341log10(engine.power + 1) - 0.0285vehicle.speed)
Cruise Mode	CO = 10$^{(-2.223+0.0033engine.power)}$
HC	
Idle Mode	0.00091777 g/s
Decelerating Motoring Mode	0.001113 g/s
Revised Deceleration Mode	0.001312 g/s
Acceleration Mode	HC = (0.114 + 0.0426log10(engine.power + 1))^4
Cruise Mode	HC = (0.170 + 0.0022 (engine.power)$^{(1/2)}$)^4

The transformations employed for the three pollutants in acceleration and cruise modes are different. The predictive capabilities of each of the models for three pollutants are also different. The R^2 value is high for NO_x and CO emission rates, but very low for HC emission rates. HC models are not much better than simply using HTBR mean ERs. The relatively poor performance of the HC models is not an inherent limitation of the modal modeling approach. Instead,

it is a result of the lack of availability of a suitable explanatory variable for model development purposes. Although the model with dummy variables and interactions works better, the final model is not necessarily the best fit, but is one that can be readily implemented.

The three models include all of those significant variables identified as affecting gram/second emissions rates, with the exception of those variables that are highly correlated with individual bus ID. Although a few of the vehicles behaved differently from other vehicles, modelers could not reasonably include bus ID as a variable, nor environmental parameters of testing since all low barometric pressure tests were conducted on one or two vehicles. Additional exploration of environmental conditions should be conducted by collecting data for a larger fleet under a wider variety of environmental conditions over a longer time.

The new modal emission rate models all indicate that engine power has a significant impact on the acceleration and cruise emission rates. This observation strengthens the importance of using load based emission data to develop new emission models and simulate engine power in real world applications. All three models were shown to be robust by use of several statistical measures. Although some departures from accepted norms were noted, these departures were judged not so serious as to compromise the usefulness of the models. Hence, no remedial measures were taken.

13.2 Model Limitations

There are several limitations in the models estimated and presented in this work. Theoretically, the models cannot be used to forecast emissions beyond the domain of variables used in estimating the models. These models were developed from 15 buses equipped with same fuel injection type, catalytic converter type, transmission type, and so on, so the models could not consider the effect of variation in vehicle technologies on emissions. Another limitation is the consideration of the effect of emission control technology deterioration on emission levels since all buses were only 5 or 6 years old at the time testing was conducted. Although the speed/acceleration profiles between the AATA data set and the Atlanta buses are similar, there is no way to estimate the effect of changes in vehicle technologies and deterioration on emissions in the current and future fleet in Atlanta. Such a limitation introduces obvious uncertainties in the use of the model to make predictions for other fleets.

The predictive models are derived from a research effort conducted by other parties. Modeling at this time cannot control for those variables for which data were not collected. This inability to control the variables may yield several uncertainties in the models. First, important or useful variables relevant to the effect of emission rates may not have been observed at all, so it

may be difficult to derive a model with sufficient explanatory power, or variables that are selected may simply be correlated to the true causal variables that are affecting instantaneous emission rates. Second, the interpretation of the effects of individual variables effects might be limited. For example, the ability of negative load to explain the variability on emissions is limited due to the negative loads recorded as zero.

An additional limitation imposed by the data is the uncertainty introduced by the actual data collection process. The uncertainty in the GPS position will introduce significant instantaneous error in grade computation (grade should be collected by means other than GPS). Although filter limits were imposed on the rate of change of engine speed (RPM), fuel flow, and vehicle speed data, data could yield unreasonable instantaneous vehicle acceleration or deceleration rates, and still be within reasonable absolute limits. This uncertainty may bias predictions.

The possible presence of outliers has the potential to cause a misleading fit by disproportionately pulling the fitted regression line away from the majority of the data points (Neter et al. 1996). Cook's distance plots indicated that some points do have influence over the regression fit. However, none of these points is indicative of obvious errors in data. It is difficult to determine whether those extreme values were actually outliers or not. Since the data passed through EPA's rigorous QA/QC procedures and no "true" outliers exist, and these high-emission events are assumed to be representative of events that occur in the real world. Therefore, all of these data were retained in model development. When additional data become available, researchers should make it a priority to examine these high emissions events to identify the underlying causal factors.

13.3 Lessons Learned

Because driving mode definitions varied across previous research efforts, findings from these efforts are not directly comparable. This study independently developed driving mode definitions through comparison across critical values. Suitable modal activity definition can divide the data into several homogeneous groups according to emission rates and driving conditions. Unlike previous research efforts which only present pairwise comparisons of modal average estimates or HTBR regression tree analyses, this study compared distributions of engine operating characteristics under proposed vehicle mode definitions by defining applicable vehicle modes.

A representative data set is the most critical issue for development the final version of the proposed model. This issue plays an important role no matter which modeling approach is employed. The representative data set should reflect the real world with respect to vehicle emissions and activity patterns. The data set used for the proposed model consists of EPA AATA data

and includes 15 buses. At the time this research was conducted, the AATA data were the only applicable data set that contained all required data (second-by-second emission rates, engine load, and applicable operating variables) all collected in parallel. New data sets will improve model performance in future.

A combination of tree and OLS regression methods was used to estimate NO_x, CO and HC emission models from EPA's transit bus database tested by Sensors, Inc. The HTBR technique was used as a tool to reveal underlying data structure and identify useful explanatory variables and was demonstrated as a powerful tool that will allow researchers to deal with large multivariate data sets with mixed mode (discrete and continuous) variables.

13.4 Contributions

This research verifies that vehicle emission rates are highly correlated with modal vehicle activity. Furthermore, the relationship between engine power and emissions is also significant and is quantified for the available data. Research results indicate that engine power is more powerful than surrogate variables in predicting second-by-second grams/second emission rates. Hence, to improve our understanding of emission rates, it is important to examine not only vehicle operating modes, but also engine power distributions. Based upon the important role of engine power in explaining the variability of emissions, it is critical to include the load data measurement (and collection of all onroad operating parameters to estimate load, such as grade) during the emission data collection procedure.

Another major contribution of the work is the establishment of a framework for emission rate model development suitable for predicting emissions at microscopic level. As more databases become available, the model development steps can be re-run to develop a more robust load-based modal emission model based on the same philosophy. This living modeling framework provides the ability to integrate necessary vehicle activity data and emission rate algorithms to support second-by-second and link-based emissions prediction. Combined with a GIS framework, models derived through this methodology will improve spatial/temporal emissions modeling.

13.5 Recommendation for Further Studies

The methodology developed and applied in this research can, and should, be used to estimate similar models for the on-road fleet consisting of transit buses and heavy-duty vehicles. Since emissions of these vehicles are heavily dependent on vehicle dynamics (that is, load and power), a successful validation will provide further evidence of the "correctness" of the method employed here. When new data become available and these models are re-derived, modelers

can expect further improved performance in applications to the transit bus fleet and eventually to other heavy-duty vehicle fleets.

Given the important role of engine power in explaining the variability of emissions, engine load data should be measured during the emission data collection procedure and all parameters necessary to estimate onroad load (such as grade and vehicle payload) should be included in the data collection efforts. Similarly, simulation of engine power demand for onroad operations becomes important in the implementation of emission inventory modeling for heavy-duty transit buses. Refinement of roadway characteristic data (grade, etc.) for urban areas is paramount and research efforts that can quantify drive train inertial losses under various operating conditions will help enhance modal model development.

Because all buses tested were of the same model with the same engine, the test data were valuable from the perspective of controlling potential explanatory variables related to vehicle characteristics. However, these data simultaneously constrain the ability to explain the effect of vehicle technology groups and deterioration of emission control technologies on emissions data. Expanded data collection efforts should focus on identification of appropriate vehicle technology groups and high-emitting vehicle groups. In these test programs, it will also be important to test buses under their real-world operating conditions (on a variety of routes, road types and grades, onroad operating conditions, environmental conditions, passenger loadings, etc.) to better reflect real world conditions. These high-resolution data collection efforts will provide the data needed by modelers to develop new and enhanced modal emission rate models for a variety of heavy-duty vehicle classes.

14. REFERENCES

Ahanotu, D. (1999). Heavy-Duty Vehicle Weight and Horsepower Distributions: Measurement of Class-Specific Temporal and Spatial Variability. School of Civil and Environmental Engineering. Atlanta, GA, Georgia institute of Technology. Ph.D. dissertation.

AMS. (2005). "A look at U.S. air pollution laws and their amendments." Retrieved July 30, 2005, from http://www.ametsoc.org/sloan/cleanair/cleanairlegisl.html

Avol, et. al. (2001). "Respiratory effects of relocating to areas of differing air pollution levels." Am. J. Respir. Crit. Care. Med. 164: 2067-2072.

Bachman, W. (1998). A GIS-Based Modal Model of Automobile Exhaust Emissions Final Report. Atlanta, GA, Prepared by Georgia Institute of Technology for U.S. Environmental Protection Agency. EPA-600/R-98-097.

Bachman, W., W. Sarasua, et al. (2000). "Modeling Regional Mobile Source Emissions in a GIS Framework." Transportation Research C 8(1-6): 205-229.

Barth, M., F. An, et al. (1996). "Modal Emission Modeling: A Physical Approach." Transportation Research Record 1520: 81-88.

Barth, M., F. An, et al. (2000). "Comprehensive Modal Emissions Model (CMEM), Version 2.0 User's Guide." http://pah.cert.ucr.edu/cmem/cmem_users_guide.pdf. January 2000.

Barth, M., G. Gcora, et al. (2004). A Modal Emission Model for Heavy Duty Diesel Vehicles. Proceedings of the 83rd Transportation Research Board Annual Meeting Proceedings (CD-ROM), Washington, DC.

Barth, M. and J. Norbeck (1997). NCHRP Project 25-11: The Development of a Comprehensive Modal Emission Model. Proceedings of the 7th CRC On-Road Vehicle Emissions Workshop, Coordinating Research Council, Atlanta, GA.

Breiman, L., J. Friedman, et al. (1984). Classification and Regression Trees. Wadsworth International Group, Belmont. CA.

Brown, J. Edward, et al. (2001).” Heavy Duty Diesel Fine Particulate Matter Emissions: Development and Application of On-Road Measurement Capabilities.” Research Triangle Park, NC, Prepared by ARCADIS Geraghty & Miller, Inc. for U.S. Environmental Protection Agency. EPA-600/R-01-079.

Browning, L. (1998). Update of Heavy-Duty Engine Emission Conversion Factors -- Analysis of Fuel Economy, Non-Engine Fuel Economy Improvements and Fuel Densities, U.S. Environmental Protection Agency.

CARB (1991). Modal Acceleration Testing. Mailout No. 91-12; Mobile Source Division; El Monte, CA.

CARB (2002). “Heavy-Duty Diesels Compression Ignition Engine Emissions and Testing.” California Air Resources Board Emissions Inventory Series 1(10).

CARB (2004). “ California’s Air Quality History Key Events.” California Air Resources Board Retrieved July 2, 2004, from http://www.arb.ca.gov/html/brochure/history.htm

CARB (2007) “EMFAC” California Air Resources Board Retrieved July 20, 2007, from http://www.arb.ca.gov/msei/onroad/latest_version.htm

Carlock, M. A. (1994). An Analysis of High Emitting Vehicles in the On-road Vehicle Fleet. Proceedings of the 87th Air and Waste Management Association Annual Meeting Proceeding Pittsburgh, PA.

CEDF (2002). “Nitrogen Oxides: How NOx Emissions Affect Human Health and the Environment.” Environmental Defense.

CFR (2007a). Calculations: exhaust emissions (40CFR86.1342-90). Code of Federal Regulations. National Archives and Records Administration.

CFR (2007b). Urban Dynamometer Schedules (40CFR86. Appendix I). Code of Federal Regulations. National Archives and Records Administration.

CFR (2004a). National Primary and Secondary Ambient Air Quality Standards (40CFR50). Code of Federal Regulations. National Archives and Records Administration.

CFR (2004b). Gross Vehicle Weight Rating (40CFR86.1803). Code of Federal Regulations. National Archives and Records Administration.

CFR (2004c). Useful Lift (40CFR86.1805). Code of Ferderal Regulations. National Archives and Records Adminisration.

Chakravart, L. and Roy (1967). Handbook of Methods of Applied Statistics, Volume I, John Wiley.

Clark, N. N., J. M. Kern, et al. (2002). “Factors Affecting Heavy-Duty Diesel Vehicle Emissions.” Journal of the Air & Waste Management Association 52: 84-94.

Clark, N. N., A. S. Khan, et al. (2005). Idle Emissions from Heavy-Duty Diesel Vehicles, Center for Alternative Fuels, Engines, and Emissions (CAFEE), Department of Mechanical and Aerospace Engineering, West Virginia University (WVU).

Conover, W. J. (1980). Practical Non-parametric Statistics, John Wiley and Sons; New York, NY.

Copt, S. and S. Heritier (2006). Robust MM-Estimation and Inference in Mixed Linear Models. Department of Econometrics, Working Papers, University of Sydney.

Davis, W., K. Wark, et al., Eds. (1998). Air Pollution Its Origin and Control. 3rd Edition, 2003 Special Studies. Addison Wesley Longman, Inc. Menlo Park, California.

Denis, M. J. S., P. Cicero-Fernandez, et al. (1994). “Effects of In-Use Driving Conditions and Vehicle/Engine Operating Parameters on “Off-Cycle” Events: Comparison with Federal Test Procedure Conditions.” Journal of the Air & Waste Management Association 44(1): 31-38.

DieselNet. (2006). “Heavy-Duty FTP Transient Cycle.” Retrieved December 20, 2006, from http://www.dieselnet.com/standards/cycles/ftp_trans.html

Dreher, D. and R. Harley (1998). “A Fuel-Based Inventory for Heavy-Duty Diesel Truck Emissions.” Journal of the Air & Waste Management Association 48: 352-358.

Easton, V. J. and J. H. McColl. (2005). “Statistics Glossary.” Retrieved March, 28, 2005, from http://www.stats.gla.ac.uk/steps/glossary/index.html.

Ensfield, C. (2002). On-Road Emissions Testing of 18 Tier 1 Passenger Cars and 17 Diesel Powered Public Transport Buses. Saline, MIchigan, Sensors, inc.

FCAP (2004). “ Ambient Air Quality Trends: An Analysis of Data Collected by the U.S. Environmental Protection Agency.” Foundation for Clean Air Progress.

Feng, C., S. Yoon, et al. (2005). Data Needs for a Proposed Modal Heavy-Duty Diesel Vehicle Emission Model. Proceedings of the 98th Air and Waste Management Association Annual Meeting Proceeding (CD-ROM), Pittsburgh, PA.

Fomunung, I., S. Washington, et al. (1999). “A Statistical Model for Estimating Oxides of Nitrogen Emissions from Light-Duty Motor Vehicles.” Transportation Research D 4D(5): 333-352.

Fomunung, I., S. Washington, et al. (2000). "Validation of the MEASURE Automobile Emissions Model: A Statistical Analysis." Journal of Transportation Statistics 3(2): 65-84.

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

Frey, H. C., A. Unal, et al. (2002). Recommended Strategy for On-Board Emission Data Analysis and Collection for the New Generation Model. Raleigh, NC, Prepared by Computational Laboratory for Energy, Air, and Risk, Department of Civil Engineering, North Carolina State University, Prepared for Office of Transportation and Air Quality, U.S. Environmental Protection Agency. http://www.epa.gov/otaq/models/ngm/ncsu.pdf.

Frey, H. C. and J. Zheng (2001). Methods and Example Case Study for Analysis of Variability and Uncertainty in Emissions Estimation (AUVEE). Research Triangle Park, NC, Prepared by North Carolina State University for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency.

Gajendran, P. and N. N. Clark (2003). "Effect of Truck Operating Weight on Heavy-Duty Diesel Emissions." Environment Science and Technology 37: 4309-4317.

Gauderman, et. al. (2002). "Association between air pollution and lung function growth in Southern California children: Results from a second cohort." Am J Resp Crit Care Med 166(1): 74-84.

Gautam, M. and N. Clark (2003). Heavy-Duty Vehicle Chassis Dynamometer Testing for Emissions Inventory, Air Quality Modeling, Source Apportionment and Air Toxics Emission Inventory; Phase I Report. Coordinating Research Council, Project No. E-55/E-59.

Gillespie, T. (1992). Fundamentals of Vehicle Dynamics. Warrendale, PA, Society of Automotive Engineers, Inc.

Granell, J. L., R. Guensler, et al. (2002). Using Locality-Specific Fleet Distributions in Emissions Inventories: Current Practice, Problems, and Alternatives. Proceedings of the 81st Transportation Research Board Annual Meeting (CD-ROM), Washington, DC.

Grant, C., R. Guensler, et al. (1996). Variability of Heavy-Duty Vehicle Operating Mode Frequencies for Prediction of Mobile Emissions. Proceedings of the 89th Air and Waste Management Association Annual Meeting Proceeding (CD-ROM), Pittsburgh, PA.

Guensler, R. (1993). “Data Needs for Evolving Motor Vehicle Emission Modeling Approaches.” In: Transportation Planning and Air Quality II, Paul Benson, Ed.; American Society of Civil Engineers: New York, NY; 1993.

Guensler, R., S. Yoon., et al. (2005). Heavy-Duty Diesel Vehicle Modal Emissions Modeling Framework. Regional Applied Research Effort (RARE) Project. Presented to U.S. Environmental Protection Agency, Georgia Institute of Technology.

Guensler, R., S. Yoon., et al. (2006). Heavy-Duty Diesel Vehicle Modal Emissions Modeling Framework. Regional Applied Research Effort (RARE) Project. Presented to U.S. Environmental Protection Agency, Georgia Institute of Technology.

Guensler, R., D. Sperling, et al. (1991). Uncertainty in the Emission Inventory for Heavy-Duty Diesel-Powered Trucks. Proceedings of the 84th Air and Waste Management Association Annual Meeting Proceedings (CD-ROM), Pittsburgh, PA.

Guensler, R., S. Washington, et al. (1998). “Overview of MEASURE Modeling Framework.” Proc. Conf. Transport Plan Air Quality A: 51-70.

Hallmark, S. L. (1999). Analysis and Prediction of Individual Vehicle Activity for Microscopic Traffic Modeling. Civil and Environmental Engineering. Atlanta, Georgia Institute of Technology. Ph.D. dissertation.

Heywood, J. B. (1998). Internal Combustion Engine Fundamentals. New York, The McGraw Hill Publishing Company

HowStuffWorks (2005). Retrieved December 30, 2005, from http://www.howstuffworks.com

Jimenez-Palacios, J. (1999). Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing. Cambridge, MA, Massachusetts Institute of Technology. Ph. D. dissertation.

Kelly, N. A. and P. J. Groblicki (1993). “Real-World Emissions from a Modern Production Vehicle Driven in Los Angeles.” Journal of the Air & Waste Management Association 43(10): 1351-1357.

Kittelson, D. B., D. F. Dolan, et al. (1978). “Diesel Exhaust Particle Size Distribution - Fuels and Additive Effects.” SAE Paper No. 780787.

Koehler, K. J. and K. Larnz (1980). “An empirical investigation of goodness-of-fit statistics for sparse multinomials.” Journal of the American Statistical Association 75: 336-344.

Koupal, J., M. Cumberworth, et al. (2002). Draft Design and Implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emissions System (MOVES), U. S. Environmental Protection Agency. EPA-420/P-02-006.

Koupal, J., N. E. Nam, et al. (2004). The MOVES Approach to Modal Emission Modeling. Proceedings of the 14th CRC On-Road Vehicle Emissions Workshop, Coordinating Research Council, San Diego, CA.

Li, L. (2004). Calculating the Confidence Intervals Using Bootstrap, Department of Statistics, University of Toronto, Presented for a project of Ontario Power Generation on October 28, 2004.

Lindhjem, C. and T. Jackson (1999). Update of Heavy-Duty Emission Levels (Model Years 1998-2004+) for Use in MOBILE6, U.S. Environmental Protection Agency.

Lloyd, A. C. and T. A. Cackette (2001). "Diesel Engines: Environmental Impact and Control." Journal of Air and Waste Management Association 51: 809-847.

MOBILE6. (2007) Access http://www.epa.gov/otaq/m6.htm

Nam, E. K. (2003). Proof of Concept Investigation for the Physical Emission Rate Estimator (PERE) to be Used in MOVES, Ford Research and Advanced Engineering.

Neter, J., M. H. Kutner, et al. (1996). Applied Linear Statistical Models, McGraw-Hill: Chicago IL.

Newton, K., W. Steeds, et al. (1996). The Motor Vehicle. Warrendale, PA, Society of Automotive Engineers, Inc.

NRC (2000). Modeling Mobile-Source Emissions. Washington, D.C., National Academy Press, National Research Council.

Peters, J. M., et al. (1999). "A study of twelve Southern California communities with differing levels and types of air pollution. II. Effects on pulmonary function." Am. J. Respir. Crit. Care Med. 159: 768-775.

Prucz, J. C., N. N. Clark, et al. (2001). "Exhaust Emissions from Engines of the Detroit Diesel Corporation in Transit Buses: A Decade of Trends." Environment Science and Technology 35: 1755-1764.

Ramamurthy, R. and N. Clark (1999). "Atmospheric Emissions Inventory Data for Heavy-Duty Vehicles." Environmental Science and Technology 33: 55-62.

Ramamurthy, R., N. N. Clark, et al. (1998). “Models for Predicting Transient Heavy Duty Vehicle Emissions.” Society of Automotive Engineers SAE 982652.

Roess, R. P., E. S. Prassas, et al. (2004). Traffic Engineering, Pearson Education, Inc.

SCAQMD (2000). Multiple Air Toxics Exposure Study (MATES-II), South Coast Air Quality Management District Governing Board.

Schlappi, M. G., R. G. Marshall, et al. (1993). “Truck travel in the San Francisco Bay Area.” Transportation Research Record 1383: 85-94.

Siegel S, and N. Castellan. (1988). Non-parametric Statistics for the Behavioural Sciences 2nd Edition. McGraw-Hill. January 1988.

Singer, B. C. and R. A. Harley (1996). “A fuel-based motor vehicle emission inventory.” Journal of the Air & Waste Management Association 46: 581-593.

StatsDirect. (2005). “Statistical Help.” http://www.statsdirect.com/ Retrieved May 30, 2005.

TRB (1995). Expanding Metropolitan Highways: Implications for Air Quality and Energy Use. Washington, DC, Transportation Research Board, National Academy Press.

U.S. EPA (1993). User’s Guide to Mobile5a. http://www.epa.gov/otaq/models/mobile5/mob5ug.pdf

U.S. EPA (1995). National Air Quality and Emission Trends Report 1995, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. http://www.epa.gov/air/airtrends/aqtrnd95/report/

U.S. EPA (1997). Emissions Standards Reference Guide for Heavy-Duty and Nonroad Engines. http://www.epa.gov/otaq/cert/hd-cert/stds-eng.pdf. EPA-420-F-97-014.

U.S. EPA (1998). Update of Fleet Characterization Data for Use in MOBILE6 - Final Report, U.S. Environmental Protection Agency. EPA-420/P-98-016.

U.S. EPA (2001a). EPA’s New Generation Mobile Source Emissions Model: Initial Proposal and Issues, U.S. Environmental Protection Agency. EPA-420/R-01-007.

U.S. EPA (2001b). Update of Heavy-duty Emission Levels (Model Years 1988-2004) for Use in MOBILE6, U.S. Environmental Protection Agency. EPA-420/R-99-010.

U.S. EPA (2001c). Heavy Duty Diesel Fine Particulate Matter Emissions: Development And Application Of On-Road Measurement Capabilities. Research Triangle Park, NC. Prepared by National Risk Management Research Laboratory for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. EPA-600/R-01-079.

U.S. EPA (2002a) "MOBILE6 Vehicle Emission Modeling Software" Retrieved July 20, 2007 from http://www.epa.gov/otaq/m6.htm.

U.S. EPA (2002b). Update Heavy-Duty Engine Emission Conversion Factors for MOBILE6, Analysis of Fuel Economy, Non-Engine Fuel Economy Improvements and Fuel Densities. EPA-420/P-98-014.

U.S. EPA (2002c). Methodology for Developing Modal Emission Rates for EPA's Multi-Scale Motor Vehicle and Equipment Emission System. Raleigh, NC, Prepared by North Carolina State University for Office of Transportation and Air Quality, U.S. Environmental Protection Agency. EPA-420/R-01-027.

U.S. EPA (2002d). Update Heavy-duty Engine Emission Conversion Factors for MOBILE6: Analysis of BSFCs and Calculation of Heavy-duty Engine Emission Conversion Factors. EPA-420/P-98-015.

U.S. EPA (2003). National Air Quality and Emissions Trends Report, 2003 Special Studies Edition. Research Triangle Park, NC, Office of Air Quality and Standards, U.S. Environmental Protection Agency. EPA-454/R-03-005.

U.S. EPA (2004c). Technical Guidance on the Use of MOBIEL6 for Emissions Inventory Preparation. Publication No. EPA420-R-04-013. U.S. Environmental Protection Agency.

U.S. EPA (2005). "Fine Particle (PM2.5) Designations." Retrieved October 20, 2005, from http://www.epa.gov/pmdesignations/index.htm.

U.S. EPA (2006). "National Ambient Air Quality Standards (NAAQS)." Retrieved October 30, 2006, from http://www.epa.gov/air/criteria.html

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

Washington, S., J. Leonard, et al. (1997b). Forecasting Vehicle Modes of Operation needed as Input to 'Model' Emissions Models. Proceedings of the 4th International Scientific Sym-

posium on Transport and Air Pollution, Lyon, France.

Washington, S., L. F. Mannering, et al. (2003). Statistical and Econometric Methods for Transportation Data Analysis, CRC Pr I Llc.

Washington, S., J. Wolf, et al. (1997a). "Binary Recursive Partitioning Method for Modeling Hot-Stabilized Emissions from Motor Vehicles." Transportation Research Record 1587: 96-105.

Whitley, E. and J. Ball (2002). "Statistics review 6: Nonparametric methods." Critical Care 6: 509-513.

Wolf-Heinrich, H. (1998). Aerodynamics of Road Vehicles, Society Of Locomotive Engineers Inc., USA.

Wolf, J., R. Guensler, et al. (1998). "High Emitting Vehicle Characterization Using Regression Tree Analysis." Transportation Research Record 1641: 58-65.

Yoon, S. (2005c). A New Heavy-Duty Vehicle Visual Classification and Activity Estimation Method For Regional Mobile Source Emissions Modeling. School of Civil and Environmental Engineering. Atlanta, Georgia Institute of Technology. Ph.D. dissertation.

Yoon, S., H. Li, et al. (2005a). Transit Bus Engine Power Simulation: Comparison of Speed-Acceleration-Road Grade Matrices to Second-by-Second Speed, Acceleration, and Road Grade Data. Proceedings of the 98th Air and Waste Management Association Annual Meeting Proceeding (CD-ROM), Pittsburgh, PA.

Yoon, S., H. Li, et al. (2005b). A Methodology for Developing Transit Bus Speed-Acceleration Matrices to be used in Load-Based Mobile Source Emissions Models. Proceedings of the 84th Transportation Research Board Annual Meeting Proceedings (CD-ROM), Washington, DC.

Yoon, S., M. Rodgers, et al. (2004b). "Engine and Vehicle Characteristics of Heavy-Duty Diesel Vehicles in the Development of Emissions Inventories: Model Year, Engine Horsepower and Vehicle Weight." Transportation Research Record(1880): 99-107.

Yoon, S., P. Zhang, et al. (2004a). A Heavy-Duty Vehicle Visual Classification Scheme: Heavy-Duty Vehicle Reclassification Method for Mobile Source Emissions Inventory Development. Proceedings of the 97th Air and Waste Management Association Annual Meeting Proceeding (CD-ROM), Pittsburgh, PA.

Younglove, T., G. Scora, et al. (2005). Designing On-road Vehicle Test Programs for Effective Vehicle Emission Model Development. Proceedings of the 84th Transportation Research Board Annual Meeting Proceedings (CD-ROM), Washington, DC.

Zeldovich, Y. B., P. Y. Sadonikov, et al. (1947). “The oxidation of nitrogen in combustion and explosions.” Acta Physicochimica USSR 21(4): 577-628.