

# Transit Bus Load-Based Modal Emission Rate Model Development



EPA/600/R-07/106  
July 2007

**TRANSIT BUS LOAD-BASED MODAL  
EMISSION RATE MODEL DEVELOPMENT**

by

Chunxia Feng  
Randall Guensler  
Michael Rodgers

School of Civil and Environmental Engineering  
Georgia Institute of Technology  
Atlanta, GA

Contract No: EP-05C-000033

EPA Project Officer: Sue Kimbrough  
U.S. Environmental Protection Agency  
National Risk Management Research Laboratory  
Air Pollution Prevention and Control Laboratory  
Research Triangle Park, NC 27711

U.S. Environmental Protection Agency  
Office of Research and Development  
Washington, DC 20460

## ABSTRACT

Heavy-duty diesel vehicle (HDDVs) operations are a major source of oxides of nitrogen (NO<sub>x</sub>) and particulate matter (PM) emissions in metropolitan areas nationwide. Although HDDVs constitute a small portion of the onroad fleet, they typically contribute more than 45% of NO<sub>x</sub> and 75% of PM onroad mobile source emissions (U.S. EPA 2003). HDDV emissions are a large source of global greenhouse gas and toxic air containment emissions. Over the last several decades, both government and private industry have made extensive efforts to regulate and control mobile source emissions. The relative importance of emissions from HDDVs has increased significantly because today's gasoline powered vehicles are more than 95% cleaner than vehicles in 1968.

In current regional and microscale modeling conducted in every state except California, HDDV emissions rates are taken from the U.S. Environmental Protection Agency's (EPA's) MOBILE 6.2 model (U.S. EPA 2001a). The U.S. Environmental Protection Agency (U.S. EPA) is currently developing a new set of modeling tools for the estimation of emissions produced by onroad and off-road mobile sources. The new Multi-scale mOtor Vehicle & equipment Emission System, known as MOVES (U.S. EPA 2001a), is a modeling system designed to better predict emissions from onroad operations.

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology (Guensler, et al. 2006). HDDV-MEM differs from other proposed HDDV modal models (Barth, et al. 2004; Frey, et al. 2002; Nam 2003) in that the modeling framework first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions.

## FOREWORD

The U.S. Environmental Protection Agency (EPA) is charged by Congress with protecting the Nation's land, air, and water resources. Under a mandate of national environmental laws, the agency strives to formulate and implement actions leading to a compatible balance between human activities and the ability of natural systems to support and nurture life. To meet this mandate, EPA's research program is providing data and technical support for solving environmental problems today and building a science knowledge base necessary to manage our ecological resources wisely, understand how pollutants affect our health, and prevent or reduce environmental risks in the future.

The National Risk Management Research Laboratory (NRMRL) is the agency's center for investigation of technological and management approaches for preventing and reducing risks from pollution that threaten human health and the environment. The focus of the laboratory's research program is on methods and their cost-effectiveness for prevention and control of pollution to air, land, water, and subsurface resources; protection of water quality in public water systems; remediation of contaminated sites, sediments, and ground water; prevention and control of indoor air pollution; and restoration of ecosystems. NRMRL collaborates with both public and private sector partners to foster technologies that reduce the cost of compliance and to anticipate emerging problems. NRMRL's research provides solutions to environmental problems by: developing and promoting technologies that protect and improve the environment; advancing scientific and engineering information to support regulatory and policy decisions; and providing the technical support and information transfer to ensure implementation of environmental regulations and strategies at the national, state, and community levels.

This publication has been produced as part of the laboratory's strategic long-term research plan. It is published and made available by EPA's Office of Research and Development to assist the user community and to link researchers with their clients.

Sally Gutierrez, Director  
National Risk Management Research Laboratory

## **EPA REVIEW NOTICE**

This report has been peer and administratively reviewed by the U.S. Environmental Protection Agency and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

## TABLE OF CONTENTS

ABSTRACT .....	ii
FOREWORD .....	iii
EPA REVIEW NOTICE .....	iv
LIST OF ACRONYMS.....	xxii
SUMMARY .....	xxv
1. INTRODUCTION .....	1-1
1.1 Emissions from Heavy-Duty Diesel Vehicles .....	1-1
1.2 Current Heavy-Duty Vehicle Emissions Modeling Practices.....	1-2
1.3 Research Approaches and Objectives .....	1-2
1.4 Summary of Research Contributions .....	1-3
1.5 Report Organization .....	1-4
2. HEAVY-DUTY DIESEL VEHICLE EMISSIONS .....	2-1
2.1 How Diesel Engine Works .....	2-1
2.1.1 The Internal Combustion Engine .....	2-1
2.1.2 Comparison with the Gasoline Engine.....	2-3
2.2 Diesel Engine Emissions.....	2-4
2.2.1 Oxides of Nitrogen and Ozone Formation .....	2-4
2.2.2 Fine Particulate Matter (PM <sub>2.5</sub> ).....	2-5
2.3 Heavy-Duty Diesel Vehicle Emission Regulations .....	2-6
2.3.1 National Ambient Air Quality Standards .....	2-6
2.3.2 Heavy-Duty Engine Certification Standards.....	2-7
2.3.3 Heavy-Duty Engine Emission Regulations.....	2-8
2.4 Heavy-Duty Diesel Vehicle Emission Modeling.....	2-8
3. HEAVY-DUTY DIESEL VEHICLE EMISSIONS MODELING .....	3-1
3.1 VMT-Based Vehicle Emission Models .....	3-1
3.1.1 MOBILE .....	3-1

3.1.2	EMFAC .....	3-5
3.1.3	Summary .....	3-6
3.2	Fuel-Based Vehicle Emission Models .....	3-7
3.3	Modal Emission Rate Models .....	3-8
3.3.1	CMEM.....	3-8
3.3.2	MEASURE.....	3-9
3.3.3	MOVES.....	3-10
3.3.4	HDDV-MEM.....	3-11
3.3.4.1	Model Development Approaches.....	3-12
3.3.4.2	Vehicle Activity Module .....	3-13
3.3.4.3	Engine Power Module.....	3-14
3.3.4.4	Emission Rate Module.....	3-18
3.3.4.5	Emission Outputs .....	3-19
4.	EMISSION DATASET DESCRIPTION AND POST-PROCESSING PROCEDURE....	4-1
4.1	Transit Bus Dataset.....	4-1
4.1.1	Data Collection Method .....	4-2
4.1.2	Transit Bus Data Parameters .....	4-4
4.1.3	Sensors, Inc. Data Processing Procedure .....	4-5
4.1.4	Data Quality Assurance/Quality Check .....	4-6
4.1.5	Database Formation .....	4-11
4.1.6	Data Summary.....	4-12
4.2	Heavy-duty Vehicle Dataset .....	4-14
4.2.1	Data Collection Method .....	4-14
4.2.2	Heavy-duty Vehicle Data Parameters.....	4-16
4.2.3	Data Quality Assurance/Quality Control Check .....	4-17
4.2.4	Database Formation .....	4-20
4.2.5	Data Summary.....	4-20
5.	METHODOLOGICAL APPROACH .....	5-1
5.1	Modeling Goal and Objectives.....	5-1
5.2	Statistical Method.....	5-2
5.2.1	Parametric Methods .....	5-2
5.2.1.1	The <i>t</i> -Test .....	5-2
5.2.1.2	Ordinary Least Squares Regression .....	5-3
5.2.1.3	Robust Regression .....	5-5
5.2.2	Nonparametric Methods.....	5-5
5.2.2.1	Chi-Square Test.....	5-5

5.2.2.2 Kolmogorv-Smirnov Two-Sample Test .....	5-6
5.2.2.3 Wilcoxon Mann-Whitney Test .....	5-7
5.2.2.4 Analysis of Variance (ANOVA).....	5-7
5.2.2.5 HTBR.....	5-8
5.3 Modeling Approach.....	5-11
5.4 Model Validation .....	5-13
6. DATA SET SELECTION AND ANALYSIS OF EXPLANATORY VARIABLES .....	6-1
6.1 Data Set Used for Model Development .....	6-1
6.2 Representative Ability of the Transit Bus Data Set.....	6-3
6.3 Variability in Emissions Data .....	6-5
6.3.1 Inter-bus Variability.....	6-5
6.3.2 Descriptive Statistics for Emissions Data .....	6-8
6.4 Potential Explanatory Variables .....	6-18
6.4.1 Vehicle Characteristics .....	6-19
6.4.2 Roadway Characteristics.....	6-22
6.4.3 Onroad Load Parameters.....	6-23
6.4.4 Environmental Conditions .....	6-23
6.4.5 Summary .....	6-24
6.5 Selection of Explanatory Variables .....	6-24
7. MODAL ACTIVITY DEFINITIONS DEVELOPMENT .....	7-1
7.1 Overview of Current Modal Activity Definitions .....	7-1
7.2 Proposed Modal Activity Definitions and Validation.....	7-3
7.3 Conclusions .....	7-11
8. IDLE MODE DEVELOPMENT .....	8-1
8.1 Critical Value for Speed in Idle Mode .....	8-1
8.2 Critical Value for Acceleration in Idle Mode .....	8-4
8.3 Emission Rate Distribution by Bus in Idle Mode .....	8-8
8.4 Discussions .....	8-13
8.4.1 High HC Emissions.....	8-13
8.4.2 High Engine Operating Parameters.....	8-15
8.5 Idle Emission Rates Estimation .....	8-16
8.6 Conclusions and Further Considerations .....	8-19
9. DECELERATION MODE DEVELOPMENT .....	9-1
9.1 Critical Value for Deceleration Rates in Deceleration Mode .....	9-1
9.2 Analysis of Deceleration Mode Data .....	9-5
9.2.1 Emission Rate Distribution by Bus in Deceleration Mode .....	9-5

9.2.2	Engine Power Distribution by Bus in Deceleration Mode .....	9-9
9.3	The Deceleration Motoring Mode .....	9-12
9.4	Deceleration Emission Rate Estimations .....	9-15
9.5	Conclusions and Further Considerations.....	9-19
10.	ACCELERATION MODE DEVELOPMENT .....	10-1
10.1	Critical Value for Acceleration in Acceleration Mode .....	10-1
10.2	Analysis of Acceleration Mode Data.....	10-6
10.2.1	Emission Rate Distribution by Bus in Acceleration Mode .....	10-6
10.2.2	Engine Power Distribution by Bus in Acceleration Mode .....	10-10
10.3	Model Development and Refinement .....	10-12
10.3.1	HTBR Tree Model Development.....	10-12
10.3.1.1	NO <sub>x</sub> HTBR Tree Model Development.....	10-16
10.3.1.2	CO HTBR Tree Model Development .....	10-20
10.3.1.3	HC HTBR Tree Model Development .....	10-22
10.3.2	OLS Model Development and Refinement .....	10-29
10.3.2.1	NO <sub>x</sub> Emission Rate Model Development for Acceleration Mode .....	10-29
10.3.2.1.1	Linear Regression Model with Engine Power .....	10-29
10.3.2.1.2	Linear Regression Model with Engine Power and Vehicle Speed .....	10-34
10.3.2.1.3	Linear Regression Model with Dummy Variables.....	10-36
10.3.2.1.4	Model Discussions .....	10-38
10.3.2.2	CO Emission Rate Model Development for Acceleration Mode.....	10-42
10.3.2.2.1	Linear Regression Model with Engine Power .....	10-42
10.3.2.2.2	Linear Regression Model with Engine Power and Vehicle Speed .....	10-46
10.3.2.2.3	Linear Regression Model with Dummy Variables.....	10-47
10.3.2.3	HC Emission Rate Model Development for Acceleration Mode.....	10-54
10.3.2.3.1	Linear Regression with Engine Power.....	10-55
10.3.2.3.2	Linear Regression Model with Dummy Variables.....	10-59
10.3.2.3.3	Model Discussions .....	10-61
10.4	Conclusions and Further Considerations .....	10-63
11.	CRUISE MODE DEVELOPMENT .....	11-1
11.1	Analysis of Cruise Mode Data.....	11-1
11.1.1	Engine Rate Distribution by Bus in Cruise Mode .....	11-2
11.1.2	Engine Power Distribution by Bus in Cruise Mode .....	11-5
11.2	Model Development and Refinement .....	11-7

11.2.1	HTBR Tree Model Development .....	11-7
11.2.1.1	NO <sub>x</sub> HTBR Tree Model Development.....	11-11
11.2.1.2	CO HTBR Tree Model Development .....	11-15
11.2.1.3	HC HTBR Tree Model Development .....	11-19
11.2.2	OLS Model Development and Refinement .....	11-25
11.2.2.1	NO <sub>x</sub> Emission Rate Model Development for Cruise Mode.....	11-25
11.2.2.1.1	Linear Regression Model with Engine Power .....	11-25
11.2.2.1.2	Linear Regression Model with Dummy Variables .....	11-30
11.2.2.1.3	Model Discussion.....	11-32
11.2.2.2	CO Emission Rate Model Development for Cruise Mode .....	11-35
11.2.2.2.1	Linear Regression Model with Engine Power .....	11-35
11.2.2.2.2	Linear Regression Model with Dummy Variables .....	11-39
11.2.2.2.3	Model Discussion.....	11-41
11.2.2.3	HC Emission Rate Model Development for Cruise Mode .....	11-43
11.2.2.3.1	Linear Regression Model with Engine Power .....	11-43
11.2.2.3.2	Linear Regression Model with Dummy Variables .....	11-47
11.2.2.3.3	Model Discussion.....	11-49
11.3	Conclusions and Further Considerations .....	11-51
12.	MODEL VERIFICATION .....	12-1
12.1	Engine Power vs. Surrogate Power Variables .....	12-1
12.2	Mean Emission Rates vs. Linear Regression Model.....	12-4
12.3	Mode-specific Load Based Modal Emission Rate Model vs. Emission Rate Models as a Function of Engine Load.....	12-6
12.4	Separation of Acceleration and Cruise Modes .....	12-11
12.5	MOBILE6.2 vs. Load-Based Modal Emission Rate Model.....	12-12
12.6	Conclusions .....	12-13
13.	CONCLUSIONS.....	13-1
13.1	Transit Bus Emission Rate Models .....	13-3
13.2	Model Limitations .....	13-4
13.3	Lessons Learned.....	13-5
13.4	Contributions.....	13-6
13.5	Recommendation for Further Studies .....	13-6
14.	REFERENCES .....	14-1

## TABLE OF FIGURES

Figure 2.1 Actions of a four-stroke gasoline internal combustion engine -- Adapted from (HowStuffWorks 2005) .....	2-2
Figure 2.2 Actions of a four-stroke diesel engine (HowStuffWorks 2005) .....	2-3
Figure 3-1 FTP Transient Cycle (DieselNet 2006) .....	3-3
Figure 3-2 Urban Dynamometer Driving Schedule Cycle for Heavy-Duty Vehicle (DieselNet 2006) .....	3-3
Figure 3-3 CARB's Four Mode Cycles (CARB 2002).....	3-6
Figure 3-4 A Framework of Heavy-Duty Diesel Vehicle Modal Emission Model (Guensler et al. 2005).....	3-12
Figure 3-5 Primary Elements in the Drivetrain (Gillespie 1992) .....	3-14
Figure 4-1 Bus Routes Tested for U. S. EPA (Ensfield 2002).....	4-3
Figure 4-2 SEMTECH-D in Back of Bus (Ensfield 2002) .....	4-4
Figure 4-3 Bus 380 GPS vs. ECM Vehicle Speed (Ensfield 2002) .....	4-6
Figure 4-4 Example Check for Erroneous GPS Data for Bus 360 (Ensfield 2002).....	4-8
Figure 4-5 Example Check for Synchronization Errors for Bus 360.....	4-9
Figure 4-6 Histograms of Engine Power for Zero Speed Data Based on Three Different Time Delays .....	4-10
Figure 4-7 General Criteria for Maximum Grades (Roess et al. 2004) .....	4-11
Figure 4-8 Onroad Diesel Emissions Characterization Facility (U.S. EPA 2001c).....	4-14
Figure 4-9 Example Check for Erroneous Measured Horsepower for Test 3DRI2-2 .....	4-18
Figure 4-10 Vehicle Speed Correlation (U.S. EPA 2001c) .....	4-19
Figure 4-11 Vehicle Speed Error for Different Speed Ranges (U.S. EPA 2001c) .....	4-19
Figure 6-1 HTBR Regression Tree Result for NO <sub>x</sub> Emission Rate for All Data Sets .....	6-2
Figure 6-2 HTBR Regression Tree Result for CO Emission Rate for All Data Sets.....	6-2
Figure 6-3 HTBR Regression Tree Result for HC Emission Rate for All Data Sets.....	6-3
Figure 6-4 Transit Bus Speed-Acceleration Matrix .....	6-4

Figure 6-5 Test Environmental Conditions .....	6-5
Figure 6-6 Median and Mean of NO <sub>x</sub> Emission Rates by Bus .....	6-6
Figure 6-7 Median and Mean of CO Emission Rates by Bus .....	6-7
Figure 6-8 Median and Mean of HC Emission Rates by Bus .....	6-7
Figure 6-9 Empirical Cumulative Distribution Function Based on Bus Based Median Emission Rates for Transit Buses .....	6-8
Figure 6-10 Histogram, Boxplot, and Probability Plot of NO <sub>x</sub> Emission Rate .....	6-9
Figure 6-11 Histogram, Boxplot, and Probability Plot of CO Emission Rate .....	6-10
Figure 6-12 Histogram, Boxplot, and Probability Plot of HC Emission Rate .....	6-10
Figure 6-13 Histogram, Boxplot, and Probability Plot of Truncated NO <sub>x</sub> Emission Rate .....	6-12
Figure 6-14 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate .....	6-13
Figure 6-15 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate .....	6-13
Figure 6-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO <sub>x</sub> Emission Rate .....	6-15
Figure 6-17 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate .....	6-15
Figure 6-18 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate .....	6-16
Figure 6-19 The X Classes and Typical Vehicle Configurations .....	6-20
Figure 6-20 Throttle Position vs. Engine Power for Transit Bus Data Set .....	6-27
Figure 6-21 Scatter plots for environmental parameters .....	6-29
Figure 7-1 Average NO <sub>x</sub> Modal Emission Rates for Different Activity Definitions .....	7-5
Figure 7-2 Average CO Modal Emission Rates for Different Activity Definitions .....	7-5
Figure 7-3 Average HC Modal Emission Rates for Different Activity Definitions .....	7-6
Figure 7-4 HTBR Regression Tree Result for NO <sub>x</sub> Emission Rate .....	7-8
Figure 7-5 HTBR Regression Tree Result for CO Emission Rate .....	7-8
Figure 7-6 HTBR Regression Tree Result for HC Emission Rate .....	7-9
Figure 8-1 Engine Power vs. NO <sub>x</sub> Emission Rate for Three Critical Values .....	8-2
Figure 8-2 Engine Power vs. CO Emission Rate for Three Critical Values .....	8-2
Figure 8-3 Engine Power vs. HC Emission Rate for Three Critical Values .....	8-3
Figure 8-4 Engine Power Distribution for Three Critical Values based on NO <sub>x</sub> Emissions ..	8-3
Figure 8-5 Engine Power vs. NO <sub>x</sub> Emission Rate for Four Options .....	8-5

Figure 8-6 Engine Power vs. CO Emission Rate for Four Options .....	8-6
Figure 8-7 Engine Power vs. HC Emission Rate for Four Options .....	8-6
Figure 8-8 Engine Power Distribution for Four Options based on NO <sub>x</sub> Emission Rates .....	8-7
Figure 8-9 Histograms of Three Pollutants for Idle Mode.....	8-9
Figure 8-10 Median and Mean of NO <sub>x</sub> Emission Rates in Idle Mode by Bus.....	8-9
Figure 8-11 Median and Mean of CO Emission Rates in Idle Mode by Bus .....	8-10
Figure 8-12 Median and Mean of HC Emission Rates in Idle Mode by Bus .....	8-10
Figure 8-13 Histograms of Engine Power in Idle Mode by Bus .....	8-12
Figure 8-14 Tree Analysis Results for High HC Emission Rates by Bus and Trip .....	8-14
Figure 8-15 Time Series Plot for Bus 360 Trip 4 Idle Segment 1 (130 Seconds).....	8-14
Figure 8-16 Time Series Plot for Bus 360 Trip 4 Idle Segment 38 (516 Seconds).....	8-15
Figure 8-17 Time Series Plot for Bus 372 Trip 1 Idle Segment 1 (500 Seconds).....	8-15
Figure 8-18 Time Series Plot for Bus 383 Trip 1 Idle Segment 12 (1258 Seconds).....	8-16
Figure 8-19 Graphical Illustration of Bootstrap (Adopted from Li 2004)) .....	8-17
Figure 8-20 Bootstrap Results for Idle Emission Rate Estimation .....	8-18
Figure 9-1 Engine Power Distribution for Three Options .....	9-3
Figure 9-2 Engine Power vs. NO <sub>x</sub> Emission Rate for Three Options.....	9-3
Figure 9-3 Engine Power vs. CO Emission Rate for Three Options .....	9-4
Figure 9-4 Engine Power vs. HC Emission Rate for Three Options .....	9-4
Figure 9-5 Histograms of Three Pollutants for Deceleration Mode .....	9-5
Figure 9-6 Median and Mean of NO <sub>x</sub> Emission Rates in Deceleration Mode by Bus .....	9-6
Figure 9-7 Median and Mean of CO Emission Rates in Deceleration Mode by Bus.....	9-6
Figure 9-8 Median and Mean of HC Emission Rates in Deceleration Mode by Bus.....	9-7
Figure 9-9 Histograms of Engine Power in Deceleration Mode by Bus .....	9-11
Figure 9-10 Engine Power vs. Vehicle Speed, Engine Power vs. Engine Speed, and Vehicle Speed vs. Engine Speed .....	9-12
Figure 9-11 Histograms for Three Pollutants in Deceleration Motoring Mode (a) and Deceleration Non-Motoring Mode (b) .....	9-13
Figure 9-12 Bootstrap Results for NO <sub>x</sub> Emission Rate Estimation in Deceleration Mode .....	9-16
Figure 9-13 Bootstrap Results for CO Emission Rate Estimation in Deceleration Mode .....	9-16
Figure 9-14 Bootstrap Results for HC Emission Rate Estimation in Deceleration Mode .....	9-17
Figure 9-15 Emission Rate Estimation Based on Bootstrap for Deceleration Mode .....	9-17

Figure 10-1 Engine Power Distribution for Three Options .....	10-2
Figure 10-2 Engine Power vs. NO <sub>x</sub> Emission Rate (g/s) for Three Options .....	10-2
Figure 10-3 Engine Power vs. CO Emission Rate (g/s) for Three Options .....	10-3
Figure 10-4 Engine Power vs. HC Emission Rate (g/s) for Three Options .....	10-3
Figure 10-5 Engine Power vs. Emission Rate for Acceleration Mode and Cruise Mode....	10-6
Figure 10-6 Histograms of Three Pollutants for Acceleration Mode.....	10-7
Figure 10-7 Median and Mean of NO <sub>x</sub> Emission Rates in Acceleration Mode by Bus.....	10-8
Figure 10-8 Median and Mean of CO Emission Rates in Acceleration Mode by Bus .....	10-9
Figure 10-9 Median and Mean of HC Emission Rates in Acceleration Mode by Bus .....	10-9
Figure 10-10 Histograms of Engine Power in Acceleration Mode by Bus.....	10-11
Figure 10-11 Histogram, Boxplot, and Probability Plot of Truncated NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-13
Figure 10-12 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Acceleration Mode .....	10-14
Figure 10-13 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Acceleration Mode .....	10-14
Figure 10-14 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-15
Figure 10-15 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate in Acceleration Mode .....	10-15
Figure 10-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate in Acceleration Mode .....	10-16
Figure 10-17 Original Untrimmed Regression Tree Model for Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-17
Figure 10-18 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode.	10-18
Figure 10-19 Trimmed Regression Tree Model for Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-18
Figure 10-20 Original Untrimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode .....	10-20
Figure 10-21 Reduction in Deviation with the Addition of Nodes of Regression Tree for Truncated Transformed CO Emission Rate in Acceleration Mode ..	10-21
Figure 10-22 Trimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode .....	10-21

Figure 10-23 Original Untrimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Acceleration Mode .....	10-23
Figure 10-25 Trimmed Regression Tree Model for Truncated Transformed HC in Acceleration Mode.....	10-25
Figure 10-26 Secondary Trimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Acceleration Mode .....	10-26
Figure 10-27 Final Regression Tree Model for Truncated Transformed HC and Engine Power in Acceleration Mode.....	10-28
Figure 10-28 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.1.....	10-31
Figure 10-29 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.2.....	10-32
Figure 10-30 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.3.....	10-33
Figure 10-31 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.4.....	10-36
Figure 10-32 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.5.....	10-38
Figure 10-33 QQ and Residual vs. Fitted Plot for CO Model 2.1 .....	10-43
Figure 10-34 QQ and Residual vs. Fitted Plot for CO Model 2.2 .....	10-44
Figure 10-35 QQ and Residual vs. Fitted Plot for CO Model 2.3 .....	10-45
Figure 10-36 QQ and Residual vs. Fitted Plot for CO Model 2.4 .....	10-47
Figure 10-37 QQ and Residual vs. Fitted Plot for CO Model 2.5 .....	10-50
Figure 10-38 QQ and Residual vs. Fitted Plot for CO Model 2.6 .....	10-52
10-55	
Figure 10-39 QQ and Residual vs. Fitted Plot for HC Model 3.1 .....	10-56
Figure 10-40 QQ and Residual vs. Fitted Plot for HC Model 3.2 .....	10-57
Figure 10-41 QQ and Residual vs. Fitted Plot for HC Model 3.3 .....	10-58
Figure 10-42 QQ and Residual vs. Fitted Plot for HC Model 3.4 .....	10-61
Figure 11-1 Histograms of Three Pollutants for Cruise Mode .....	11-2
Figure 11-2 Median and Mean of NO <sub>x</sub> Emission Rates in Cruise Mode by Bus .....	11-3
Figure 11-3 Median and Mean of CO Emission Rates in Cruise Mode by Bus .....	11-3
Figure 11-4 Median and Mean of HC Emission Rates in Cruise Mode by Bus .....	11-4
Figure 11-5 Histograms of Engine Power in Cruise Mode by Bus .....	11-6
Figure 11-6 Histogram, Boxplot, and Probability Plot of Truncated NO <sub>x</sub> Emission Rates in Cruise Mode .....	11-8
Figure 11-7 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Cruise Mode.....	11-9
Figure 11-8 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Cruise Mode.....	11-9

Figure 11-9 Histogram, Boxplot, and Probability Plot of Truncated Transformed	
NO <sub>x</sub> Emission Rate in Cruise Mode.....	11-10
Figure 11-10 Histogram, Boxplot, and Probability Plot of Truncated Transformed	
CO Emission Rate in Cruise Mode .....	11-10
Figure 11-11 Histogram, Boxplot, and Probability Plot of Truncated Transformed	
HC Emission Rate in Cruise Mode .....	11-11
Figure 11-12 Original Untrimmed Regression Tree Model for Truncated	
Transformed NO <sub>x</sub> Emission Rate in Cruise Mode .....	11-12
Figure 11-13 Reduction in Deviation with the Addition of Nodes of Regression	
Tree for Truncated Transformed NO <sub>x</sub> Emission Rate in Cruise Mode .....	11-12
Figure 11-14 Trimmed Regression Tree Model for Truncated Transformed NO <sub>x</sub>	
Emission Rate in Cruise Mode.....	11-14
Figure 11-15 Original Untrimmed Regression Tree Model for Truncated	
Transformed CO Emission Rate in Cruise Mode.....	11-16
Figure 11-16 Reduction in Deviation with the Addition of Nodes of Regression	
Tree for Truncated Transformed CO Emission Rate in Cruise Mode.....	11-16
Figure 11-17 Trimmed Regression Tree Model for Truncated Transformed CO	
Emission Rate in Cruise Mode.....	11-18
Figure 11-18 Original Untrimmed Regression Tree Model for Truncated	
Transformed HC Emission Rate in Cruise Mode.....	11-19
Figure 11-19 Trimmed Regression Tree Model for Truncated Transformed HC	
Emission Rate in Cruise Mode.....	11-21
Figure 11-20 Secondary Trimmed Regression Tree Model for Truncated	
Transformed HC in Cruise Mode .....	11-22
Figure 11-21 Final Regression Tree Model for Truncated Transformed HC	
and Engine Power in Cruise Mode.....	11-24
Figure 11-22 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.1 .....	11-27
Figure 11-23 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.2.....	11-28
Figure 11-24 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.3.....	11-29
Figure 11-25 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.4.....	11-32
Figure 11-26 QQ and Residual vs. Fitted Plot for CO Model 2.1 .....	11-36
Figure 11-27 QQ and Residual vs. Fitted Plot for CO Model 2.2 .....	11-37
Figure 11-28 QQ and Residual vs. Fitted Plot for CO Model 2.3 .....	11-38
Figure 11-29 QQ and Residual vs. Fitted Plot for CO Model 2.4 .....	11-40
Figure 11-30 QQ and Residual vs. Fitted Plot for HC Model 3.1 .....	11-44

Figure 11-31 QQ and Residual vs. Fitted Plot for HC Model 3.2 .....	11-45
Figure 11-32 QQ and Residual vs. Fitted Plot for HC Model 3.3 .....	11-46
Figure 11-33 QQ and Residual vs. Fitted Plot for HC Model 3.4 .....	11-48
Figure 12-1 QQ and Residual vs. Fitted Plot for NO <sub>x</sub> Model 1.....	12-4
Figure 12-2 Trimmed Regression Tree Model for Truncated Transformed NO <sub>x</sub> .....	12-7
Figure 12-3 QQ and Residual vs. Fitted Plot for Load-Based Only NO <sub>x</sub> Emission Rate Model .....	12-9

## TABLE OF TABLES

Table 2-1. National Ambient Air Quality Standards (U.S. EPA 2006) .....	2-6
Table 2-2. Heavy-Duty Engine Emissions Standards (U.S. EPA 1997) .....	2-8
Table 3-1. Heavy-Duty Vehicle NO <sub>x</sub> Emission Rates in MOBILE6 .....	3-4
Table 3-2 Heavy-Duty Vehicle CO Emission Rates in MOBILE6.....	3-4
Table 3-3 Heavy-Duty Vehicle HC Emission Rates in MOBILE6.....	3-4
Table 4-1 Buses Tested for U.S. EPA (Ensfield 2002) .....	4-2
Table 4-2 Transit Bus Parameters Given by the U.S. EPA (Ensfield 2002).....	4-4
Table 4-3 List of Parameters Used in Explanatory Analysis for Transit Bus .....	4-12
Table 4-4 Summary of Transit Bus Database .....	4-13
Table 4-5 Onroad Tests Conducted with Pre-Rebuild Engine .....	4-15
Table 4-6 Onroad Tests Conducted with Post-Rebuild Engine.....	4-16
Table 4-7 List of Parameters Given in Heavy-duty Vehicle Dataset Provided by U.S. EPA .....	4-17
Table 4-8 List of Parameters Used in Explanatory Analysis for HDDV .....	4-20
Table 4-9 Summary of Heavy-Duty Vehicle Data U.S. EPA 2001c).....	4-21
Table 5-1 ANOVA Table for Single-Factor Study (Neter et al. 1996).....	5-8
Table 6-1 Basic Summary Statistics for Emissions Rate Data for Transit Bus .....	6-9
Table 6-2 Basic Summary Statistics for Truncated Emissions Rate Data .....	6-12
Table 6-4 Percent of High Emission Points by Bus .....	6-18
Table 6-5 Correlation Matrix for Transit Bus Data Set.....	6-25
Table 7-1 Comparison of Modal Activity Definition.....	7-3

Table 7-2 Four Different Mode Definitions and Modal Variables .....	7-4
Table 7-3 Results for Pairwise Comparison for Modal Average Estimates In Terms of P-value .....	7-7
Table 7-4 Sensitivity Test Results for Four Mode Definition .....	7-9
Table 8-1 Engine Power Distribution for Three Critical Values for Three Pollutants .....	8-4
Table 8-2 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants .....	8-4
Table 8-3 Engine Power Distribution for Four Options for Three Pollutants.....	8-7
Table 8-4 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants .....	8-8
Table 8-5 Median, and Mean of Three Pollutants in Idle Mode by Bus.....	8-11
Table 8-6 Engine Power Distribution in Idle Mode by Bus.....	8-13
Table 8-7 Idle Mode Statistical Analysis Results for NO <sub>x</sub> , CO, and HC.....	8-17
Table 8-8 Idle Emission Rates Estimation and 95% Confidence Intervals Based on Bootstrap .....	8-18
Table 9-1 Engine Power Distribution for Three Options for Three Pollutants .....	9-2
Table 9-2 Percentage of Engine Power Distribution for Three Options for Three Pollutants .....	9-2
Table 9-3 Median, and Mean for NO <sub>x</sub> , CO, and HC in Deceleration Mode by Bus .....	9-7
Table 9-4 High HC Emissions Distribution by Bus and Trip for Deceleration Mode .....	9-9
Table 9-5 Engine Power Distributions in Deceleration Mode by Bus.....	9-10
Table 9-6 Comparison of Emission Distributions between Deceleration Mode and Two Sub-Modes (Deceleration Motoring Mode and Deceleration Non-Motoring Mode) .....	9-14
Table 9-7 Emission Rate Estimation and 95% Confidence Intervals Based on Bootstrap for Deceleration Mode .....	9-18
Table 10-1 Engine Power Distribution for Three Options for Three Pollutants.....	10-4
Table 10-2 Percentage of Engine Power Distribution for Three Options for Three Pollutants .....	10-4
Table 10-3 Engine Power Distribution for Acceleration Mode and Cruise Mode.....	10-5
Table 10-4 Median and Mean of Three Pollutants in Acceleration Mode by Bus.....	10-7
Table 10-5 Engine Power Distribution in Acceleration Mode by Bus.....	10-10

Table 10-6 Original Untrimmed Regression Tree Results for Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-17
Table 10-7 Trimmed Regression Tree Results for Truncated Transformed NO <sub>x</sub> Emission Rate in Acceleration Mode .....	10-19
Table 10-8 Original Untrimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Acceleration Mode .....	10-20
Table 10-9 Trimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Acceleration Mode .....	10-22
Table 10-10 Original Untrimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Acceleration Mode .....	10-23
Table 10-11 Trimmed Regression Tree Results for Truncated Transformed HC in Acceleration Mode.....	10-25
Table 10-12 Secondary Trimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Acceleration Mode .....	10-27
Table 10-13 Final Regression Tree Results for Truncated Transformed HC and Engine Power in Acceleration Mode .....	10-28
Table 10-14 Regression Result for NO <sub>x</sub> Model 1.1 .....	10-30
Table 10-15 Regression Result for NO <sub>x</sub> Model 1.2 .....	10-32
Table 10-16 Regression Result for NO <sub>x</sub> Model 1.3 .....	10-33
Table 10-17 Regression Result for NO <sub>x</sub> Model 1.4 .....	10-35
Table 10-18 Regression Result for NO <sub>x</sub> Model 1.5 .....	10-37
Table 10-19 Comparative Performance Evaluation of NO <sub>x</sub> Emission Rate Models .....	10-40
Table 10-20 Regression Result for CO Model 2.1 .....	10-42
Table 10-21 Regression Result for CO Model 2.2.....	10-44
Table 10-22 Regression Result for CO Model 2.3.....	10-45
Table 10-23 Regression Result for CO Model 2.4.....	10-46
Table 10-24 Regression Result for CO Model 2.5.....	10-49
Table 10-25 Regression Result for CO Model 2.6 .....	10-51
Table 10-26 Comparative Performance Evaluation of CO Emission Rate Models.....	10-53
Table 10-27 Regression Result for HC Model 3.1 .....	10-55
Table 10-28 Regression Result for HC Model 3.2.....	10-57
Table 10-29 Regression Result for HC Model 3.3.....	10-58

Table 10-31 Comparative Performance Evaluation of HC Emission Rate Models.....	10-62
Table 11-1 Engine Power Distribution for Cruise Mode .....	11-1
Table 11-2 Median and Mean of Three Pollutants in Cruise Mode by Bus.....	11-4
Table 11-3 Engine Power Distribution in Cruise Mode by Bus.....	11-5
Table 11-4 Original Untrimmed Regression Tree Results for Truncated Transformed NO <sub>x</sub> Emission Rate in Cruise Mode .....	11-13
Table 11-5 Trimmed Regression Tree Results for Truncated Transformed NO <sub>x</sub> Emission Rate in Cruise Mode .....	11-14
Table 11-6 Original Untrimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Cruise Mode.....	11-17
Table 11-7 Trimmed Regression Tree Results for Truncated Transformed CO Emission Rate in Cruise Mode .....	11-18
Table 11-8 Original Untrimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Cruise Mode.....	11-20
Table 11-9 Trimmed Regression Tree Results for Truncated Transformed HC Emission Rate in Cruise Mode .....	11-21
Table 11-10 Trimmed Regression Tree Results for Truncated Transformed HC in Cruise Mode .....	11-23
Table 11-11 Final Regression Tree Results for Truncated Transformed HC and Engine Power in Cruise Mode.....	11-24
Table 11-12 Regression Result for NO <sub>x</sub> Model 1.1 .....	11-26
Table 11-13 Regression Result for NO <sub>x</sub> Model 1.2 .....	11-28
Table 11-15 Regression Result for NO <sub>x</sub> Model 1.4 .....	11-31
Table 11-16 Comparative Performance Evaluation of NOx Emission Rate Models.....	11-33
Table 11-17 Regression Result for CO Model 2.1 .....	11-35
Table 11-18 Regression Result for CO Model 2.2.....	11-37
Table 11-19 Regression Result for CO Model 2.3.....	11-38
Table 11-20 Regression Result for CO Model 2.4.....	11-40
Table 11-21 Comparative Performance Evaluation of CO Emission Rate Models .....	11-41
Table 11-22 Regression Result for HC Model 3.1 .....	11-43
Table 11-23 Regression Result for HC Model 3.2.....	11-45
Table 11-24 Regression Result for HC Model 3.3.....	11-46

Table 11-25 Regression Result for HC Model 3.4.....	11-48
Table 11-26 Comparative Performance Evaluation of HC Emission Rate Models .....	11-49
Table 12-1 Regression Result for NO <sub>x</sub> Model 1 .....	12-3
Table 12-2 Comparative Performance Evaluation between Mode-Only Models and Linear Regression Models .....	12-6
Table 12-3 Trimmed Regression Tree Results for Truncated Transformed NO <sub>x</sub> .....	12-8
Table 12-4 Regression Result for NO <sub>x</sub> Load-Based Only Emission Rate Model.....	12-9
Table 12-5 Comparative Performance Evaluation Between Load-Based Only Emission Rate (ER) Model and Load-Based Modal Emission Rate Model.....	12-10
Table 12-6 Comparative Performance Evaluation between Linear Regression with Combined Mode and Linear Regression with Acceleration and Cruise Modes .....	12-12
Table 12-7 Comparative Performance Evaluation between MOBILE 6.2 and Load-Based Modal ER Model.....	12-13
Table 13-1 Load Based Modal Emission Models .....	13-3

## LIST OF ACRONYMS

%	percent
AADT	annual average daily traffic
AATA	Ann Arbor Transit Authority
Acc	acceleration
ANOVA	analysis of variance
APPCD	Air Pollution Prevention and Control Division
bhp	brake horsepower
BSFC	brake specific fuel consumption
C	Celsius
CARB	California Air Resources Board
CART	classification and regression testing
CE-CERT	College of Engineering - Center for Environmental Research and Technology
CMEM	Comprehensive Modal Emissions Model
CO	carbon monoxide
deg	degree
df	degrees of freedom
DPS	drag power surrogate
DVD	digital video disc
ECM	electronic control module
EMFAC	CARB's mobile source emission factor model
E(MS)	expected mean square
EPA	Environmental Protection Agency
F	Fahrenheit
FHWA	Federal Highway Administration
FR	Federal Register
FTP	Federal Test Procedure
g/bhp-hr	grams per brake-horsepower-hour
g/h	grams per hour
g/s	grams per second
GIS	geographic information system
GPS	global positioning system
GVWR	gross vehicle weight rating
HC	hydrocarbon
HDD	heavy-duty diesel
HDDV	heavy-duty diesel vehicle

HDDV-MEM	Heavy-Duty Diesel Vehicle-Modal Emission Model
HDV	heavy-duty vehicle
HDV8B	heavy-duty vehicle 8B
HDV-UDDS	heavy-duty vehicle urban dynamometer driving schedule
Hg	mercury
HHDDE	heavy-heavy duty diesel engine
HTBR	hierarchical tree-based regression
Hz	hertz
IC	internal combustion
IPS	inertial power surrogate
kPa	kilopascal
K/S	Kolmogorov-Smirnov
LAFY	Los Angeles freeway
LANF	Los Angeles non-freeway
lb	pound
lb-ft	pound-feet
LDV	light-duty vehicle
LHDDE	light-heavy duty diesel engine
MARTA	Metropolitan Atlanta Rapid Transit Authority
MDPV	medium duty passenger vehicle
MEASURE	Mobile Emissions Assessment System for Urban and Regional Scale Emissions
MM	method of moments
mg/m <sup>3</sup>	milligrams per cubic meter
MHDDE	medium-heavy duty diesel engine
MOBILE	EPA's mobile source emission rate model
MOBILE6	EPA's mobile source emission rate model
MOVES	Motor Vehicle Emission Simulator
MPE	mean prediction error
mpg	miles per gallon
mph	miles per hour
mph/s	miles per hour per second
MS	mean square
N <sub>2</sub>	nitrogen
NAAQS	National Ambient Air Quality Standards
NCSU	North Carolina State University
NGM	EPA's Next Generation Model (mobile sources)
NIPER	National Institute for Petroleum and Energy Research
NIST	National Institute of Standards and Technology
NO	nitrogen oxide
NO <sub>2</sub>	nitrogen dioxide
NONROAD	EPA's emission rate model for non-road sources

NO <sub>x</sub>	nitrogen oxides
NRMRL	National Risk Management Research Laboratory
NYNF	New York non-freeway
O <sub>2</sub>	oxygen
O <sub>3</sub>	ozone
ODEC	Onroad Diesel Emissions Characterization
OLS	ordinary least squares
OTAQ	Office of Transportation and Air Quality
Pb	lead
PCV	positive crankcase ventilation
PERE	Physical Emission Rate Estimator
PM	particulate matter
PM <sub>10</sub>	particulate matter ≤10 microns
PM <sub>2.5</sub>	particulate matter ≤ 2.5 microns
ppmv	parts per million by volume
QA/QC	quality assurance/quality control
QQ	quantile-quantile
RARE	Regional Applied Research Effort
RMSE	root mean square error
RPM	revolutions per minute
SCFM	standard cubic feet per minute
SO <sub>2</sub>	sulfur dioxide
SS	sum of squares
SSE	sum of squares due to errors
SSTO	total sum of squares
SUV	sport utility vehicle
TB-EPDS	transit bus engine power demand simulator
TIUS	Truck Inventory and Use Survey
TRB	Transportation Research Board
UCR-CERT	University of California Riverside – Center for Environmental Research and Technology
UDDS	urban dynamometer driving schedule
U.S. EPA	U.S. Environmental Protection Agency
UV	ultraviolet
VIF	variance inflation factor
VMT	vehicle miles traveled
VOCs	volatile organic compounds
VSP	vehicle specific power
µg/m <sup>3</sup>	micrograms per cubic meter
µm	micron

## SUMMARY

Heavy-duty diesel vehicle (HDDV) operations are a major source of pollutant emissions in major metropolitan areas. Accurate estimation of heavy-duty diesel vehicle emissions is essential in air quality planning efforts because highway and non-road heavy-duty diesel emissions account for a significant fraction of the oxides of nitrogen ( $\text{NO}_x$ ) and particulate matter (PM) emissions inventories. MOBILE6 (U.S. EPA 2002a), EPA's mobile source emission rate model, uses an "average trip-based" approach to modeling as opposed to a more fundamental and robust modal modeling approach.

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology. HDDV-MEM first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions.

To provide better estimates of microscale level emissions, this modeling approach is designed to predict second-by-second emissions from on-road vehicle operations. This research statistically analyzes the database provided by EPA and yields a model for prediction of emissions at a microscale level based on engine power demand and driving mode. Research results demonstrate the importance of including the influence of engine power demand vis-à-vis emissions and simulating engine power in real world applications. The modeling approach provides a significant improvement in HDDV emissions modeling compared to the current average speed cycle-based emissions models.

This page left blank deliberately.

## CHAPTER 1

### 1. INTRODUCTION

#### 1.1 Emissions from Heavy-Duty Diesel Vehicles

Heavy-duty diesel vehicles (HDDVs) operations are a major source of oxides of nitrogen ( $\text{NO}_x$ ) and particulate matter (PM) emissions in metropolitan areas nationwide. Although HDDVs constitute a small portion of the on-road fleet, they typically contribute more than 45% of  $\text{NO}_x$  and 75% of PM on-road mobile source emissions (U.S. EPA 2003). HDDV emissions are a large source of global greenhouse gas and toxic air contaminant emissions. According to Environmental Defense Report in 2002,  $\text{NO}_x$  causes many environmental problems including acid rain, haze, global warming and nutrient overloading leading to water quality degradation (CEDF 2002). HDDV emissions are also harmful to human health and the environment (SCAQMD 2000). Groundbreaking long-term studies of children's health conducted in California have demonstrated that particle pollution may significantly reduce lung function growth in children (Avol 2001, Gauderman 2002, Peters 1999). Previous studies have stressed the significance of emissions from HDVs, in urban non-attainment areas especially for ozone (for which nitrogen oxides are a precursor) and  $\text{PM}_{2.5}$  (Gautam and Clark 2003, Lloyd and Cackette 2001).

Over the last several decades, both government and private industry have made extensive efforts to regulate and control mobile source emissions. In 1961, the first automotive emissions control technology in the nation, Positive Crankcase Ventilation (PCV), was mandated by the California Motor Vehicle State Bureau of Air Sanitation to control hydrocarbon crankcase emissions, and PCV Requirement went into effect on domestic passenger vehicles for sale in California in 1963 (CARB 2004). At the same time, first Federal Clean Air Act was enacted. Although this act only dealt with reducing air pollution by setting emissions standards for stationary sources such as power plants and steel mills at the beginning, amendments of 1965, 1966 and 1967 focused on establishing standards for automobile emissions (AMS 2005). Emission control was first required on light-duty gasoline vehicles (LDVs) by U.S. EPA in the 1968 model year. Developed and refined over a period of more than 30 years, these controls have become more effective at reducing LDV emissions (FCAP 2004).

The relative importance of emissions from HDDVs has increased significantly because today's gasoline powered vehicles are more than 95% cleaner than vehicles in 1968. Considering that HDDVs typically have a life cycle of over one million miles, may be on the road as long as 30 years, and will continue to play a major emission inventory role with increases in goods movement with their high durability and reliability, modeling of HDDV emissions is going to become increasingly important in air quality planning.

## **1.2 Current Heavy-Duty Vehicle Emissions Modeling Practices**

In current regional and microscale modeling conducted in every state except California, HDDV emissions rates are taken from the U.S. Environmental Protection Agency's (EPA's) MOBILE 6.2 model (U.S. EPA 2001b). MOBILE 6.2<sup>1</sup> emission rates were derived from base-line emission rates (gram/brakehorsepower-hour) developed in the laboratory using engine dynamometer test cycles. While different driving cycles have been developed over the years, dynamometer testing is conceptually designed to obtain a "representative sample" of vehicle operations. These work-based emission rates are then modified through a series of conversion and correction factors to obtain approximate emission rates in units of grams/mile that can be applied to on-road vehicle activity (vehicle miles traveled), as a function of temperature, humidity, altitude, average vehicle speed, etc. (Guensler 1993). The conversion process used to translate laboratory emission rates to on-road emission rates employs fuel density, brake specific fuel consumption, and fuel economy for each HDDV technology class. However, the emission rate conversion process does not appropriately account for the impacts of roadway operating conditions on brake specific fuel consumption and fuel economy (Guensler, et al. 1991).

The U.S. Environmental Protection Agency (U.S. EPA) is currently developing a new set of modeling tools for the estimation of emissions produced by on-road and off-road mobile sources. The new **Motor Vehicle Emissions Simulator**, known as MOVES<sup>2</sup> (Koupal, et al. 2004), is a modeling system designed to better predict emissions from on-road operations. The philosophy behind MOVES is to develop a model that is as directly data-driven as possible, meaning that emission rates are developed from second-by-second or binned emission rate data.

## **1.3 Research Approaches and Objectives**

The major effort of this research is to develop a new heavy-duty vehicle load-based modal emission rate model that overcomes some of the limitations of existing models and emission

---

<sup>1</sup>MOBILE = Current mobile source emissions model used for State Implementation Plan emission inventories.

<sup>2</sup>MOVES = Mobile Vehicle Emissions Estimator, next generation mobile source emissions model. The model will be used for State Implementation Plan emission inventories and will replace the current MOBILE model.

rates prediction methods. This model is part of the proposed Heavy-Duty Diesel Vehicle Modal Emission Modeling (HDDV-MEM) which was developed by Georgia Institute of Technology (Guensler, et al. 2006). HDDV-MEM differs from other proposed HDDV modal models (Barth, et al. 2004, Frey, et al. 2002, Nam 2003) in that the modeling framework first predicts second-by-second engine power demand as a function of vehicle operating conditions and then applies brake-specific emission rates to these activity predictions. This means that HDDV emission rates are predicted as a function of engine horsepower loads for different driving modes. Hence, the basic algorithm and matrix calculation in the HDDV-MEM should be transferable to MOVES. The new model implementation is similar in general structure to previous model emission rate model known as Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE<sup>1</sup>) model developed by Georgia Institute of Technology several years ago (Bachman 1998, Guensler, et al. 1998, Bachman, et al. 2000).

The major effort of this research consists of a number of specific objectives outlined below:

- Develop a new load-based modal emission rate model to improve spatial/temporal emissions modeling;
- Develop a HDDV modal emission rate model to more accurately estimate on-road HDDV emissions;
- Develop a modal model that can be verified at multiple levels;
- Develop a HDDV modal emission rate model that can be integrated into the MOVES.

#### **1.4 Summary of Research Contributions**

There are four major contributions developed by this research. First, a framework for emission rate modeling suitable for predicting emissions at different scales (microscale, mesoscale, and macroscale) is established. Since this model is developed using on-board emissions data which are collected under real-world conditions, this model will provide capabilities for integrating necessary vehicle activity data and emission rate algorithms to support second-by-second and link-based emissions prediction. Combined with GIS framework, this model will improve spatial/temporal emissions modeling.

---

<sup>1</sup>MEASURE = Mobile Emissions Assessment System for Urban and Regional Evaluation Model. This model is a prototype GIS-based modal emissions model.

Second, the relationship between engine power and emissions is explored and integrated into the modeling framework. Research results indicate that engine power is more powerful than surrogate variables to present load data in the proposed model. Based on the important role of engine power in explaining the variability of emissions, it is better to include the load data measurement during emission data collection procedure. Meanwhile, development of methods to simulate real world engine power is equally important.

Third, this research verifies that vehicle emission rates are highly correlated with modal vehicle activity. To get better understanding of driving modes, it is important to examine not only emission distributions, but also engine power distributions.

Finally, a dynamic framework is created for further improvement. As more databases become available, this approach could be re-run to obtain a more reliable load-based modal emission model based on the same philosophy.

## **1.5 Report Organization**

Chapter 2 examines the diesel fuel combustion process and its relationship to diesel engine emissions formation. Chapter 3 overviews the existing heavy-duty vehicle emission models and presents the proposed heavy-duty diesel vehicle modal emission model (HDDV-MEM). Chapter 4 provides an overview of the emission rate testing databases provided by U.S. EPA, the quality assurance and quality control (QA/QC) procedures to review the validity of the data, and the methods used to post-process these databases to correct data deficiencies. In Chapter 5, the various statistical models considered for data analysis are discussed. Chapter 6 selects the database used to develop the conceptual model and discusses the influence of explanatory variables on emissions. Chapter 7 covers sensitivity tests of driving mode definitions and outlines the potential impacts on derived models. Chapters 8 to 11 elaborate the different emission models developed for idle, deceleration, acceleration and cruise driving modes. In Chapter 12, research results are verified. Finally, Chapter 13 presents a discussion and conclusion on research results.

## **CHAPTER 2**

### **2. HEAVY-DUTY DIESEL VEHICLE EMISSIONS**

Diesel engines differ from gasoline engines in terms of the combustion processes and engine size, giving rise to their different emission properties and therefore different emissions standards. This chapter examines the diesel fuel combustion process and its relationship to diesel engine emissions formation followed by a summary of the emission regulations for diesel engines.

#### **2.1 How Diesel Engine Works**

By far the predominant engine design for transportation vehicles is the reciprocating internal combustion (IC) engine which operates either on a four-stroke or a two-stroke cycle. The two-stroke engine is commonly found in lower-power applications such as snowmobiles, lawnmowers, mopeds, outboard motors and motorcycles, while both gasoline and diesel automotive engines are classified as four-stroke engines. To understand the formation and control of emissions, it is necessary to first develop an understanding of the operation of the internal combustion engine.

##### **2.1.1 The Internal Combustion Engine**

Internal combustion engines generate power by converting the chemical energy stored in fuels into mechanical energy. The engine is termed “internal combustion” because combustion occurs in a confined space called a combustion chamber. Combustion of the fuel charge inside a chamber causes a rapid rise in temperature and pressure of the gases in the chamber, which are permitted to expand. The expanding gases are used to move a piston, turbine blades, rotor, or the engine itself.

The four-stroke gasoline engine cycle is also called Otto cycle, in honor of Nikolaus Otto, who is credited with inventing the process in 1867. The four piston strokes are illustrated in Figure 2-1. The following processes take place during one cycle of operation:

**1. Intake stroke:** the piston starts at the top, the intake valve opens, and the piston moves down to let the engine take in a fresh charge composed of a mixture of fuel and air (for spark-ignition or gasoline engine) or air only (for auto-ignition or diesel engine). (Part 1 of the figure.)

**2. Compression stroke:** then the piston moves back up to compress this fuel/air mixture (gasoline engines) or the air only (diesel engines). In gasoline engines combustion is started by ignition from a spark plug, in diesel engines auto-ignition occurs when fuel is injected into the compressed air which has achieved a high temperature through compression such that the temperature is high enough to cause self-ignition. (Part 2 of the figure.)

**3. Expansion stroke:** when the piston reaches the top of its stroke, the combustion process results in a substantial increase in the gas temperature and pressure and drives the piston down. (Part 3 of the figure.)

**4. Exhaust stroke:** once the piston hits the bottom of its stroke, the exhaust valve opens and the exhaust leaves the cylinder into the exhaust manifold and then into the tail pipe. Discharge of the burnt gases (exhaust) from the cylinder occurs to make room for the next cycle. (Part 4 of the figure.)

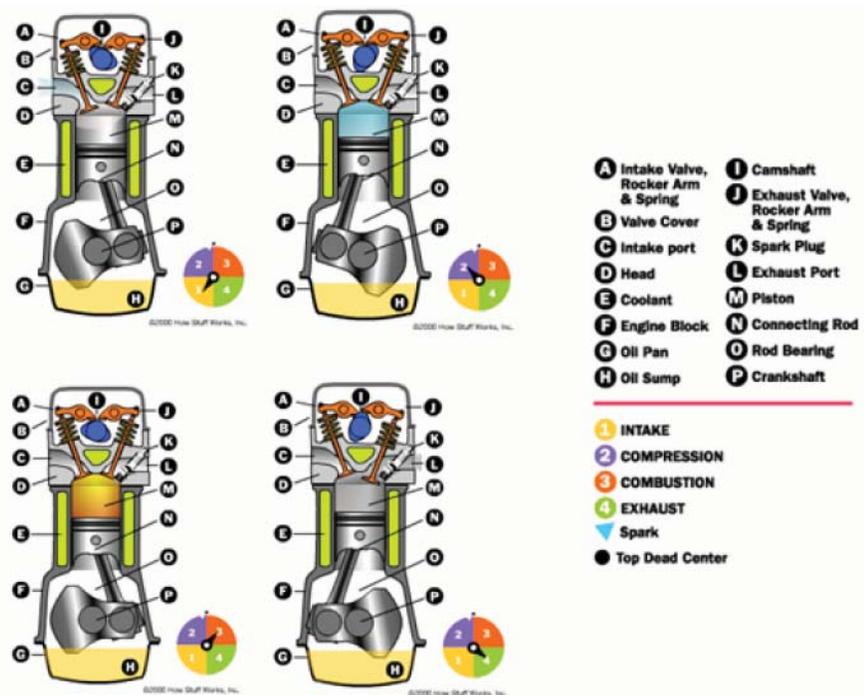


Figure 2.1 Actions of a four-stroke gasoline internal combustion engine -- Adapted from (HowStuff-Works 2005)

Figure 2-1 is a diagrammatic representation of the four strokes of an internal combustion engine. The upper end of the cylinder consists of a clearance space in which ignition and combustion occur. The expanding medium pushes against the piston head inside the cylinder, causing the piston to move; this straight line motion of the piston is converted into the desired rotary motion of the wheels by means of a drivetrain consisting of a connecting rod and crankshaft. Figure 2-1 illustrates that the only stroke that delivers useful work is the expansion stroke; the other three strokes are thus termed idle strokes. The reader interested in a detailed description

of the internal combustion engine is referred to specialized texts, such as Heywood (Heywood 1998) and Newton et al. (Newton, et al. 1996).

### 2.1.2 Comparison with the Gasoline Engine

The diesel engine employs the compression ignition cycle. German engineer Rudolf Diesel developed the idea for the diesel engine and received the patent on February 23, 1893. His goal was to create an engine with high efficiency. Figure 2-2 is a diagrammatic representation of the four strokes of a diesel engine. The main differences between the gasoline engine and the diesel engine are:

- A gasoline engine compresses at a ratio of 8:1 to 12:1, while a diesel engine compresses at a ratio of 14:1 to as high as 25:1. The higher compression ratio of the diesel engine leads to higher peak combustion temperatures and better fuel efficiency.
- Unlike a gasoline engine, which takes in a mixture of gas and air, compresses it and ignites the mixture with a spark, a diesel engine takes in just air, compresses it and then injects fuel into the compressed air. The heat of the compressed air spontaneously ignites the fuel.

- Gasoline engines generally use either carburetion, in which the air and fuel is mixed long before the air enters the cylinder, or port fuel injection, in which the fuel is injected just prior to the intake stroke (outside the cylinder), while diesel engines use direct fuel injection – the diesel fuel is injected directly into the cylinder.

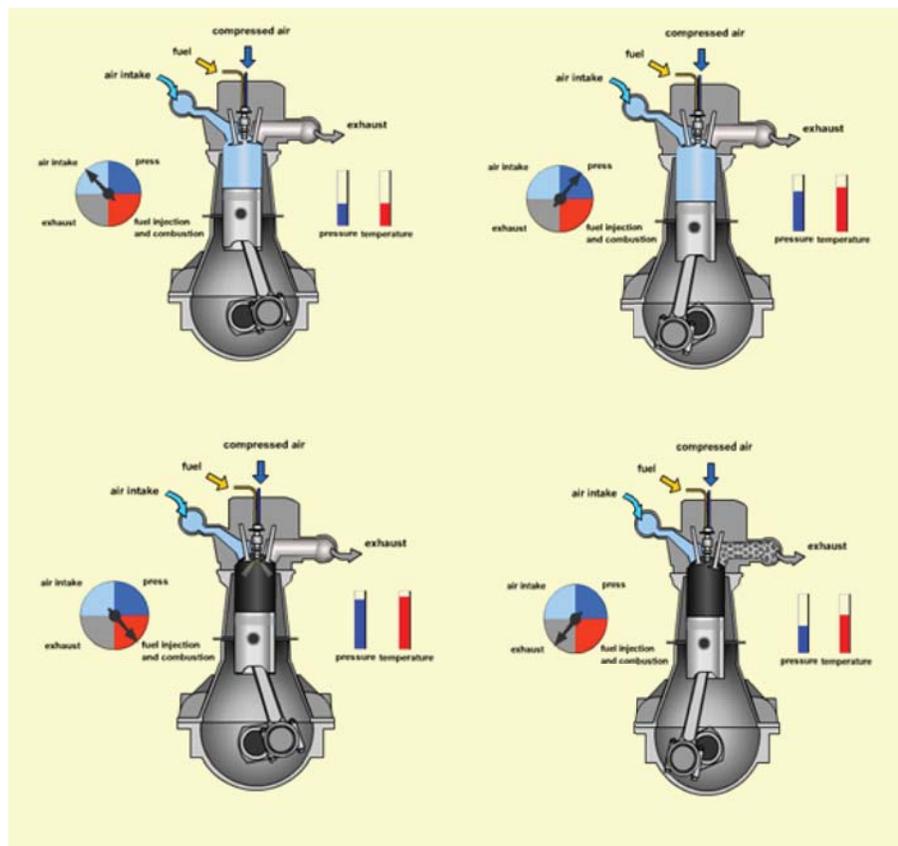


Figure 2.2 Actions of a four-stroke diesel engine (HowStuffWorks 2005)

## 2.2 Diesel Engine Emissions

Like any other internal combustion engine, diesel engines convert the chemical energy contained in diesel fuel into mechanical power. Diesel fuel is injected under pressure into the engine cylinder, where it mixes with air and combustion occurs. Diesel fuel is heavier and oilier than gasoline. Diesel fuel evaporates much more slowly than gasoline, with a boiling point that is actually higher than that of water. The lean nature of the diesel-air mixture results in a combustion environment that produces lower emission rates of carbon monoxide (CO) and hydrocarbons (HC) compared to gasoline-powered engines. However, diesel engines do produce relatively high level emissions of oxides of nitrogen (NO<sub>x</sub>) and particulate matter (PM), especially fine particulate matter. This section will discuss oxides of nitrogen and particulate emissions in detail.

### 2.2.1 Oxides of Nitrogen and Ozone Formation

Oxides of nitrogen, a mixture of nitric oxide (NO) and nitrogen dioxide (NO<sub>2</sub>), are produced from the destruction of atmospheric nitrogen (N<sub>2</sub>) during the combustion process. Atmospheric air generally consists of 80% N<sub>2</sub> and 20% O<sub>2</sub>, and these elements are stable because of the moderate temperatures and pressures. However, during high temperature and pressure conditions of combustion, excess oxygen in the combustion chamber reacts with N<sub>2</sub> to create NO which is quickly transformed into NO<sub>2</sub>. The role of nitrogen contained in the air in NO formation was initially postulated by Zeldovich (Zeldovich, et al. 1947). In near-stoichiometric or lean systems the mechanisms associated with NO formation (as many as 30 or so independent chemical reactions that also involve participation of hydrocarbon species) can generally be simplified to the following:



In near-stoichiometric and fuel-rich mixtures, where the concentration of OH radicals can be high, the following reaction also takes place:



Reaction 4, together with reactions 1, 2 and 3, are known as the extended Zeldovich mechanism. It is also important to note that emitted nitric oxide (NO) will oxidize to nitrogen dioxide (NO<sub>2</sub>) in the atmosphere over a period of a few hours.

Oxides of nitrogen (NO<sub>x</sub>) are reactive gases that cause a host of environmental concerns impacting adversely on human health and welfare. Nitrogen dioxide (NO<sub>2</sub>), in particular, is a brownish gas that has been linked with higher susceptibility to respiratory infection, increased airway resistance in asthmatics, and decreased pulmonary function. Most importantly, NO<sub>x</sub> emitted from heavy-duty vehicles plays a major role in the formation of ground level ozone pollution, which causes wide-ranging damage to human health and the environment (U.S. EPA 1995). Ozone is a colorless, highly reactive gas with a distinctive odor. Naturally, ozone is formed by electrical discharge (lightning) and in the upper atmosphere at altitudes between 15 and 35 km. Stratospheric ozone protects the Earth from harmful ultraviolet radiation from the sun. However, ground level ozone is formed by chemical reactions involving NO<sub>x</sub> and volatile organic compounds (VOCs) combining in the presence of heat and sunlight. These two categories of pollutants are also referred to as ozone precursors. The production of photochemical oxidants usually occurs over several hours which means that the highest concentrations of ozone normally occur on summer afternoons, in areas downwind of major sources of ozone precursors. The simplified reaction processes are illustrated as:



At ground level, elevated ozone concentrations can cause health and environmental problems. Ozone can affect the human cardiac and respiratory systems, irritating the eyes, nose, throat, and lungs. Symptoms of ozone exposure include itchy and watery eyes, sore throats, swelling within the nasal passages and nasal congestion. Effects from ozone are experienced only for the period of exposure to elevated levels. EPA promulgated 8-hour ozone standards in 1997 and designated an area as nonattainment if it has violated, or has contributed to violations of, the national 8-hour ozone standard over a three-year period.

### **2.2.2 Fine Particulate Matter (PM<sub>2.5</sub>)**

Particulate matter (PM) is a complex mixture of solid and liquid particles (excluding water) that are suspended in air. These particles typically consist of a mixture of inorganic and organic chemicals, including carbon, sulfates, nitrates, metals, acids, and semivolatile compounds. The size of PM in air ranges from approximately 0.005 to 100 micrometers (µm) in aerodynamic diameter -- the size of just a few atoms to about the thickness of a human hair. U.S. EPA defined three general categories for PM as coarse (10 to 2.5 µm), fine (2.5 µm or smaller), and ultrafine (0.1 µm or smaller).

Heavy-duty diesel vehicles are known to emit large quantities of small particles (Kittelson, et al. 1978). A majority of the PM found in diesel exhaust is in the nanometer size range.

Lloyd found that more than 90% of fine particles from heavy-duty vehicles are smaller than 1µm in diameter (Lloyd and Cackette 2001).

Fine PM can cause not only human health problems and property damage, but also adversely impact the environment through visibility reduction and retard plant growth (Davis, et al. 1998). Health studies have shown a significant association between exposure to fine particles and premature death from heart or lung diseases. Other important effects include aggravation of respiratory and cardiovascular disease, lung disease, decreased lung function, or asthma attacks. Individuals particularly sensitive to fine particle exposure include older adults, people with heart and lung disease, and children (U.S. EPA 2005). EPA promulgated the PM2.5 standard in 1997 and included a 24-hour standard for PM2.5 set at 65 micrograms per cubic meter (µg/m<sup>3</sup>), and an annual standard of 15 µg/m<sup>3</sup>.

### 2.3 Heavy-Duty Diesel Vehicle Emission Regulations

#### 2.3.1 National Ambient Air Quality Standards

The Clean Air Act, which was last amended in 1990, requires the U.S. EPA to set National Ambient Air Quality Standards (NAAQS) to safeguard public health against six common air pollutants: ozone (O<sub>3</sub>), particulate matter (PM), sulfur dioxide (SO<sub>2</sub>), carbon monoxide (CO), nitrogen dioxide (NO<sub>2</sub>) and lead (Pb). The Clean Air Act established two types of national air quality standards. Primary standards set limits to protect public health, including the health of “sensitive” populations such as asthmatics, children, and the elderly. Secondary standards set limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation, and buildings (CFR 2004a). Table 2-1 illustrates the current NAAQS for ambient concentrations of various pollutants. Units of measure for the standards are parts per million by volume (ppmv), milligrams per cubic meter of air (mg/m<sup>3</sup>), and micrograms per cubic meter of air (µg/m<sup>3</sup>).

Table 2-1. National Ambient Air Quality Standards (U.S. EPA 2006)

Pollutant	Average Times	Standard Value	Standard Type
Carbon Monoxide (CO)	8-hour Average	9 ppmv (10 mg/m <sup>3</sup> )	Primary
	1-hour Average	35 ppmv (40 mg/m <sup>3</sup> )	Primary
Nitrogen Dioxide (NO <sub>2</sub> )	Annual Arithmetic Mean	0.053 ppmv (100 µg/m <sup>3</sup> )	Primary & Secondary
Ozone (O <sub>3</sub> )	1-hour Average	0.12 ppmv (235 µg/m <sup>3</sup> )	Primary & Secondary
	8-hour Average	0.08 ppmv (157 µg/m <sup>3</sup> )	Primary & Secondary

<b>Pollutant</b>	<b>Average Times</b>	<b>Standard Value</b>	<b>Standard Type</b>
Lead (Pb)	Quarterly Average	1.5 $\mu\text{g}/\text{m}^3$	Primary & Secondary
Particulate (PM10)	Annual Arithmetic Mean	50 $\mu\text{g}/\text{m}^3$	Primary & Secondary
	24-hour Average	150 $\mu\text{g}/\text{m}^3$	Primary & Secondary
Particulate (PM2.5)	Annual Arithmetic Mean	15 $\mu\text{g}/\text{m}^3$	Primary & Secondary
	24-hour Average	65 $\mu\text{g}/\text{m}^3$	Primary & Secondary
Sulfur Dioxide (SO <sub>2</sub> )	Annual Arithmetic Mean	0.030 ppmv (80 $\mu\text{g}/\text{m}^3$ )	Primary
	24-hour Average	0.14 ppmv (365 $\mu\text{g}/\text{m}^3$ )	Primary
	3-hour Average	0.50 ppmv (1300 $\mu\text{g}/\text{m}^3$ )	Secondary

### 2.3.2 Heavy-Duty Engine Certification Standards

Heavy-duty vehicles are defined as vehicles of GVWR (gross vehicle weight rating) above 8,500 lbs in the federal jurisdiction and above 14,000 lbs in California (model year 1995 and later). Diesel engines used in heavy-duty vehicles are further divided into service classes by GVWR, as follows:

- Light heavy-duty diesel engines: 8,500<LHDDE<19,500 (14,000<LHDDE<19,500 in California, 1995+)
- Medium heavy-duty diesel engines: 19,500≤MHDDE≤33,000
- Heavy heavy-duty diesel engines (including urban bus): HHDDE>33,000

Under the federal light-duty Tier 2 regulation (phased in beginning 2004), vehicles of GVWR up to 10,000 lbs used for personal transportation have been re-classified as “medium-duty passenger vehicles” (MDPV – primarily larger SUVs and passenger vans) and are subject to the light-duty vehicle legislation. Thus, the same diesel engine model used for the 8,500-10,000 lbs vehicle category may be classified as either light- or heavy-duty and certified to different standards, depending on the manufacturer-defined application (CFR 2004b). Except for the heavy-duty vehicles classified as LDVs, all heavy-duty vehicle emissions standards are established using the engine dynamometer certification process.

### 2.3.3 Heavy-Duty Engine Emission Regulations

EPA regulates heavy-duty vehicle emissions for compliance with emissions standards over the useful life of the engine. Useful life is defined as follows (U.S. EPA and California) (CFR 2004c):

- LHDDE – 8 years/110,000 miles (whichever occurs first)
- MHDDE – 8 years/185,000 miles
- HHDE – 8 years/290,000 miles

Federal useful life requirements were later increased to 10 years, with no change to the above mileage numbers, for the urban bus PM standard (1994+) and for the NO<sub>x</sub> standard (1998+). The emission warranty period is 5 years/100,000 miles (5 years/100,000 miles/3,000 hours in California), but no less than the basic mechanical warranty for the engine family. Table 2-2 shows the heavy-duty engine emissions standards by model year group.

Table 2-2. Heavy-Duty Engine Emissions Standards (U.S. EPA 1997)

Year	HC (g/bhp-hr)	CO (g/bhp-hr)	NO <sub>x</sub> (g/bhp-hr)	PM (g/bhp-hr)
Heavy-Duty Diesel Truck Engines				
1988	1.3	15.5	10.7	0.60
1990	1.3	15.5	6.0	0.60
1991	1.3	15.5	5.0	0.25
1994	1.3	15.5	5.0	0.10
1998	1.3	15.5	4.0	0.10
Urban Bus Engines				
1991	1.3	15.5	5.0	0.25
1993	1.3	15.5	5.0	0.10
1994	1.3	15.5	5.0	0.07
1996	1.3	15.5	5.0	0.05*
1998	1.3	15.5	4.0	0.05*
* in-use PM standard 0.07				

### 2.4 Heavy-Duty Diesel Vehicle Emission Modeling

There are several models currently used to estimate emissions from heavy-duty vehicles. The most common emission rate models are VMT-based or cycle-based models, developed from laboratory test facility driving cycle data. Due to lack of available data representing real world conditions, all previous models were developed based upon engine dynamometer data. The following chapter will address this issue in detail.

## **CHAPTER 3**

### **3. HEAVY-DUTY DIESEL VEHICLE EMISSIONS MODELING**

Several models are currently used to estimate emissions from heavy-duty vehicles. A comprehensive review of the existing heavy-duty vehicle emission models will help modelers understand the different approaches and how they can contribute to the development of enhanced emission rate modeling techniques.

The most common emission rate models are VMT-based or cycle-based developed from laboratory test facility driving cycle data. Fuel-based models model emissions as a function of fuel usage rate as well as other parameters. In the 1990s, even the proposed enhanced modal models, designed to predict emissions as a function of speed and acceleration profiles of vehicles, were still based upon statistical analysis of cycle-based data (Bachman 2000; Fomunung 2000). More recent emission rate modeling frameworks are proposing to model modal emission rates on a second-by-second basis directly from the vehicle operating mode.

#### **3.1 VMT-Based Vehicle Emission Models**

The current emission rate models used by state and federal agencies include the Mobile Source Emission Model (MOBILE) series of models developed by the U.S. Environmental Protection Agency (U.S. EPA) and the Emission Factor Emission Inventory Model (EMFAC) series developed by California Air Resources Board (CARB).

##### **3.1.1 MOBILE**

MOBILE (U.S. EPA 1993), developed by the US EPA in the late 1970s to estimate vehicle emission, has since become the nation's standard in assessing the emission impacts of various transportation inputs. MOBILE uses the method of base emission rates and correction factors. This model has undergone significant expansion and improvements over the years. The latest version is MOBILE6 released in February 2002 (U.S. EPA 2002a).

MOBILE is based on engine dynamometer test data from selected driving cycles. The Federal Test Procedure (FTP) transient cycle is composed of a unique profile of stops, starts, constant speed cruises, accelerations and decelerations. Different driving cycles are developed to simulate both urban and freeway driving. A concern with driving cycles is that they may not be sufficiently representative of real-world emissions (Kelly and Groblicki 1993; Denis et al. 1994). For HDV emission rates, MOBILE uses the method of base emission rates and conversion factors which convert the g/bhp-hr emissions estimates observed in the laboratory to g/mile emission rates, to be consistent with available travel information. Conversion factors are used to convert the g/bhp-hr emissions estimates to grams per mile traveled. These conversion factors contribute a large source of uncertainty to the MOBILE model since the BSFC (brake specific fuel consumption) data are aggregated for the fleet and may not represent in-use vehicle characteristics (Guensler et al. 1991). Conversion factors have improved accuracy in MOBILE6 due to improved data, but fundamental flaws remain (Guensler et al. 2006).

#### 3.1.1.1 Diesel Engine Test Cycles

EPA currently uses the transient Federal Test Procedure (FTP) engine dynamometer cycle, which includes both engine cold and warm start operations, for heavy-duty vehicles (CFR Title 40, Part 86.1333). Unlike the chassis dynamometer test for light-duty vehicle, the engine is removed from the vehicle's chassis, mounted on the engine dynamometer test stands, and operated in the transient FTP test cycle. The transient cycle (Figure 3-1) consists of four phases: the first is a NYNF (New York Non Freeway) phase typical of light urban traffic with frequent stops and starts, the second is LANF (Los Angeles Non Freeway) phase typical of crowded urban traffic with few stops, the third is a LAFY (Los Angeles Freeway) phase simulating crowded expressway traffic in Los Angeles, and the fourth phase repeats the first NYNF phase. This cycle consists of a cold start after parking overnight, followed by idling, acceleration and deceleration phases, and a wide variety of different speeds and loads sequenced to simulate the running of the vehicle that corresponds to the engine being tested. There are few stabilized running conditions, and the average load factor is about 20 to 25% of the maximum horsepower available at a given speed.

Emission and operation parameters are measured while the engine operates during the test cycle. The engine torque is determined by applying performance percentages with an engine lug curve (maximum torque curve). Engine torque is then converted to engine brake horsepower using engine revolution per minute (RPM). Brake specific emissions rates are reported in g/bhp-hr and then converted to g/mile using pre-defined conversion factors (CFR Title 40, Part 86.1342-90).

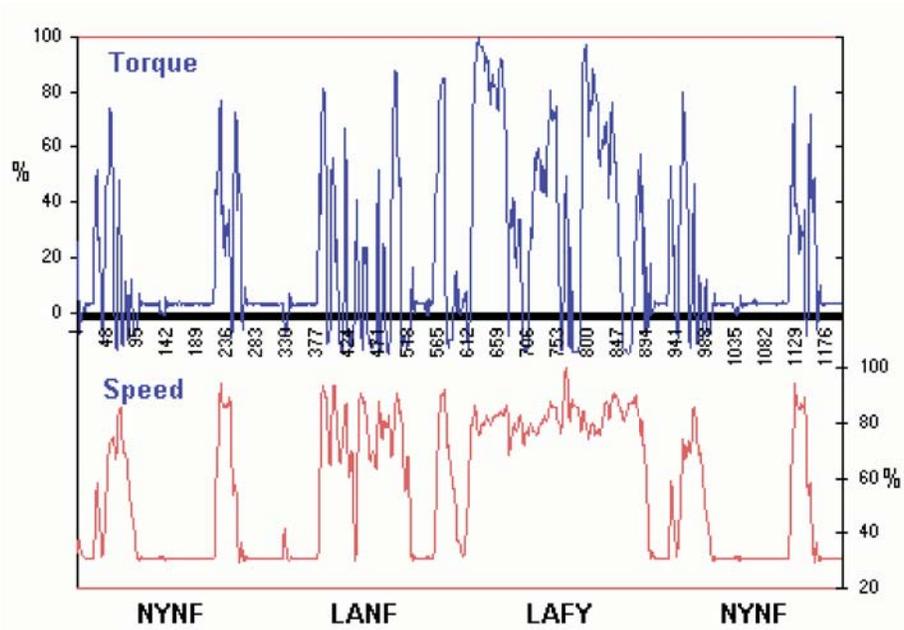


Figure 3-1 FTP Transient Cycle (DieselNet 2006)

Because the engine dynamometer test procedure does not directly account for the impacts from load and grade changes, a chassis dynamometer test procedure and the cycle known as the HDV urban dynamometer driving schedule (HDV-UDDS) was developed [CFR Title 40, Part 86, App. I], sometimes referred to as “cycle D”. This cycle is different from the UDDS cycle for light-duty vehicles (FTP-72). This HDV cycle lasts 1060 seconds and covers 5.55 miles. The average speed for HDV UDDS is 18.86 mph while the maximum speed is 58 mph. Figure 3-2 shows the speed profile for the chassis UDDS test.

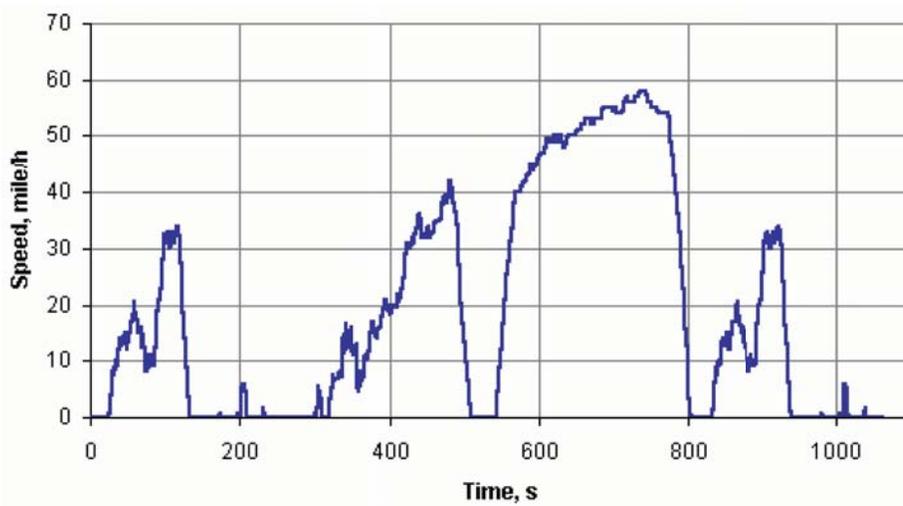


Figure 3-2 Urban Dynamometer Driving Schedule Cycle for Heavy-Duty Vehicle (DieselNet 2006)

### 3.1.1.2 Baseline Emission Rates

Baseline emission rates (g/bhp-hr) for heavy-duty vehicles are obtained from the engine dynamometer test results collected during U.S. EPA's cooperative test program with engine manufacturers. The zero mile levels and deterioration rates for NO<sub>x</sub>, CO, and HC are presented in the following tables for heavy-duty gasoline and diesel engines. All the emission rates are available from "Update of Heavy-Duty Emission Levels (Model Years 1998-2004+) for Use in MOBILE6" (Lindhjem and Jackson 1999).

Table 3-1. Heavy-Duty Vehicle NO<sub>x</sub> Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	4.96	6.28	6.43	4.34	0.044	0.01	0.009	0.002
1990	3.61	4.85	4.85	4.85	0.026	0.004	0.006	0.011
1991-1993	3.24	4.56	4.53	1.38	0.038	0.004	0.007	0.003
1994-1997	3.24	4.61	4.61	1.08	0.038	0.003	0.001	0.001
1998-2003	2.59	3.68	3.69	3.26	0.038	0.003	0.001	0.001
2004+	2.59	1.84	1.84	1.63	0.038	0.003	0.001	0.001

Table 3-2 Heavy-Duty Vehicle CO Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	13.84	1.34	1.70	1.21	0.246	0.008	0.018	0.022
1990	6.89	1.81	1.81	1.81	0.213	0.005	0.007	0.012
1991-1993	7.10	1.82	1.26	0.40	0.255	0.003	0.010	0.004
1994-1997	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003
1998-2003	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003
2004+	7.10	1.07	0.85	1.19	0.255	0.004	0.009	0.003

Table 3-3 Heavy-Duty Vehicle HC Emission Rates in MOBILE6

Model Year Class	Zero Mile Level (g/bhp-hr)				Deterioration (g/bhp-hr/10,000 miles)			
	Gasoline Engine	Diesel Engine			Gasoline Engine	Diesel Engine		
		Heavy	Med.	Light		Heavy	Med.	Light
1988-1989	0.62	0.47	0.66	0.64	0.023	0.001	0.002	0.002
1990	0.35	0.52	0.52	0.52	0.023	0.000	0.001	0.001
1991-1993	0.33	0.30	0.40	0.47	0.021	0.000	0.001	0.001
1994-1997	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001
1998-2003	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001
2004+	0.33	0.22	0.31	0.26	0.021	0.001	0.001	0.001

### 3.1.1.3 Conversion Factors

Because emission standards for both gasoline and diesel heavy-duty vehicles are expressed in terms of grams per brake-horsepower hour (g/bhp-hr), the MOBILE6.2 model employs conversion factors of brake horsepower-hour per mile (bhp-hr/mile) to convert the emission certification data from engine testing to grams per mile. Conversion factors are a function of fuel density, brake-specific fuel consumption (BSFC), and fuel economy for each HDV class (U.S. EPA 2002b). The conversion factors were calculated using Equation 3-1:

$$\text{Conversion Factor (bhp-hr/mi)} = \frac{\text{Fuel Density (lb/gal)}}{\text{BSFC (lb/bhp-hr)} \times \text{Fuel Economy (mi/gal)}} \quad (\text{Equation 3-1})$$

To calculate BSFC, U.S. EPA first obtained data from model year 1987 through 1996 supplied by six engine manufacturers (U.S. EPA 2002d). U.S. EPA then performed regression analysis for BSFCs by model year for each weight class and used a logarithmic curve to extrapolate values prior to 1988 and after 1995, since sales data were only available for model years 1988 through 1995 (U.S. EPA 2002d).

Fuel economy was calculated using a regression curve derived from the 1992 Truck Inventory and Use Survey (TIUS) conducted by the U.S. Census Bureau. Fuel densities were determined from National Institute for Petroleum and Energy Research (NIPER) publications for both gasoline and diesel (Browning 1998). Using the equation defining the conversion factor together with the data described above, weight class specific conversion factors were calculated for gasoline and diesel vehicles for model years 1987 through 1996 (U.S. EPA 2002c).

### 3.1.2 EMFAC

EMFAC (CARB 2007) was developed by CARB separately from MOBILE based upon the presence of vehicle technologies in the on-road fleet that would be subject to more stringent standards and fuels used in California. The latest version, EMFAC 2002, was released in September 2002. EMFAC can estimate emissions for calendar years 1970 to 2040.

EMFAC abandoned the use of conversion factors from EMFAC 2000 and used chassis dynamometer data collected for 70 trucks tested over the Urban Dynamometer Driving Schedule (UDDS). Although the use of UDDS test data marked a significant improvement, it is hard to say that UDDS adequately represented the full range of heavy duty diesel operation. Although the cycle was constructed from actual truck activity data, it lacks extended cruises known to cause many trucks to default to a high NO<sub>x</sub> emitting, fuel saving mode referred to as “Off-Cycle”

NOx. The cycle also lacks hard accelerations known to result in high emissions of particulate matter (CARB 2002).

CARB continues to develop more mode test cycles designed to better depict the emissions of HDDVs under real world conditions, including emissions from engine programming to go “off-cycle” at certain speeds. Activity data from instrumented truck studies conducted by Battelle and Jack Faucett Associates for CARB (CARB 2002) have been used to develop a four mode heavy-heavy-duty diesel cycle. Figure 3-3 shows these four mode cycles developed by CARB. The creep mode produced the greatest gram per mile results followed by the transient and the cruise mode. The transient and cruise modes produced higher and lower emissions, respectively, than the UDDS (CARB 2002).

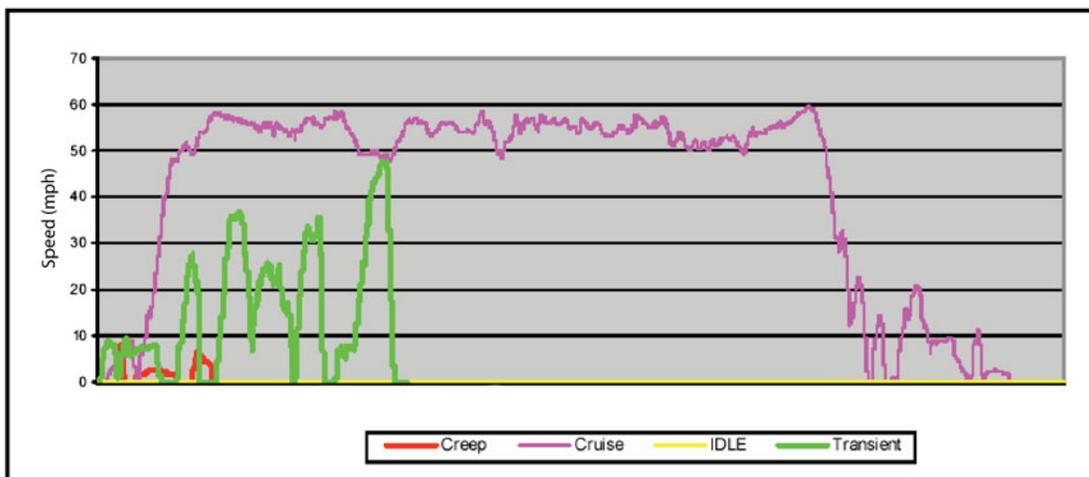


Figure 3-3 CARB's Four Mode Cycles (CARB 2002)

### 3.1.3 Summary

EPA's MOBILE series models have significantly improved through the series of model revisions from 1970s. However, the MOBILE series of models still has major modeling defects for the heavy-duty components. These defects have been widely recognized for more than 10 years (Guensler et al. 1991). One of the most frequently stated defects is that fleet average speed, which aggregates other vehicle activity factors that may yield significant bias in emissions characterization, is used to characterize vehicle emission rates.

In developing emissions inventories using the MOBILE and EMFAC (CARB 2007) emission rate models, vehicle activity is estimated using travel demand models. The estimation of VMT was based on EPA's fleet characterization study (U.S. EPA 1998). It is common to estimate heavy-duty travel as a fixed percentage of predicted traffic volumes (TRB 1995). This

estimate is not correct since heavy-duty truck travel does not follow the same spatial and temporal patterns as light-duty vehicle travel (Schlappi et al. 1993).

### 3.2 Fuel-Based Vehicle Emission Models

The fuel-based emission inventory models for heavy-duty diesel trucks combine vehicle activity data (i.e., volume of diesel fuel consumed) with emission rates normalized to fuel consumption (i.e., mass of pollutant emitted per unit volume of fuel burned) to estimate emissions within a region of interest (Dreher and R. Harley 1998). This approach was proposed to increase accuracy of truck VMT estimation by combining state level truck VMT with statewide fuel sales to estimate total heavy-duty truck activity, using the amount of fuel consumed as a measure of activity.

In California, fuel consumption data are available through tax records at the statewide level and this statewide fuel consumption can be apportioned to provide emission estimates for an individual air basin by month, day of week, and time of day. At the same time, emission rates are normalized to fuel consumption using Equation 3-2:

$$EI_p = \frac{S_p}{BSFC} \quad \text{(Equation 3-2)}$$

- where
- $EI_p$  : emission index for pollutant P, in units of mass of pollutant emitted per unit mass of fuel burned;
  - $S_p$  : brake specific pollutant emission rate obtained from the dynamometer test, expressed in g/bhp-hr units;
  - BSFC : brake specific fuel consumption of the engine being tested, also in g/bhp.

Exhaust emissions are estimated by multiplying vehicle activity, as measured by the volume of fuel used, by emission rates which are normalized to fuel consumption and expressed as grams of pollutant emitted per gallon of diesel fuel burned instead of grams of pollutant per mile (Dreher and R. Harley 1998). Average emission rates for subgroups of vehicles are weighted by the fraction of total fuel used by each vehicle subgroup to obtain an overall fleet-average emission rate. The fleet-average emission rate is multiplied by regional fuel sales to compute pollutant emissions (Singer and Harley 1996).

The advantages of the fuel-based approach include the fact that fuel-use data are available from tax records in California. Furthermore, emission rates normalized to fuel consumption vary considerably less over the full range of driving conditions than travel-normalized emission

factors (Singer and Harley 1996). The disadvantage is obvious, too. Tax records are not available for other states. It is difficult to get input data outside of California, limiting the scope of the modeling approach. Furthermore, the users first have to run two models to predict fuel used and then predict emission rates, which is not statistically efficient.

### **3.3 Modal Emission Rate Models**

Modal emission rate models work on the premise that emissions are better modeled as a function of specific modes of vehicle operation (idle, steady-state cruise, various levels of acceleration/deceleration, etc.), than as a function of average vehicle speed (Bachman 1998; Ramamurthy et al. 1998; U.S. EPA 2001b). Emissions of heavy-duty vehicles powered by diesel cycle engines are more likely to be a function of brake work output of engine than normal gasoline vehicles, because instantaneous emissions levels of diesel engine are highly correlated with the instantaneous work output of the engine (U.S. EPA 2001b).

With the consideration of vehicle modal activity, EPA and various research communities have been developing modal activity-based emission models. The report published by National Research Council (NRC 2000) comprehensively reviewed the modeling of mobile source emissions and provided recommendations for the improvement of future mobile source emission models. The following sections will introduce the most representative modal emission models one by one.

#### **3.3.1 CMEM**

The Comprehensive Modal Emissions Model (CMEM) (Barth et al. 2000) was developed by the Center for Environmental Research and Technology at University of California Riverside (UCR-CERT). Development of CMEM was first funded by National Cooperative Highway Research Program Project (1995-2000) and then is being enhanced and improved with EPA funding (2000-present). From 2001, CE-CERT created a modal-based inventory at the micro- (intersection), meso- (highway link), and macro- (region) scale levels for light-duty vehicles (LDV) and heavy-duty diesel (HDD) vehicles. The CMEM model derives a fuel rate from road-load and a simple powertrain model. Emissions rates are then derived empirically from the fuel rate. Fuel rate, or fuel consumption per unit time, forms the basis for CMEM.

The CMEM HDD emissions model (Barth et al. 2004) accepted the same approach as the light-duty vehicle model. In that model, second-by-second tailpipe emissions are modeled as the product of three components: fuel rate (FR), engine-out emission indices (grams of emissions/gram of fuel), and an emission after-treatment pass fraction. The model is composed of six mod-

ules: 1) engine power demand; 2) engine speed; 3) fuel-rate; 4) engine control unit; 5) engine-out emissions; and 6) after-treatment pass fraction. The vehicle power demand is determined based on operating variables [second-by-second vehicle speed (from which acceleration can be derived; note that acceleration can be input as a separate input variable), grade, and accessory use (such as air conditioning)] and specific vehicle parameters (vehicle mass, engine displacement, cross-sectional area, aerodynamics, vehicle accessory load, transmission efficiency, and drive-train efficiency, and so on). The core of the model is the fuel rate calculation which is a function of power demand and engine speed. Engine speed is determined based on vehicle velocity, gear shift schedule and power demand (Barth et al. 2004). The model uses a total of 35 parameters to estimate vehicle tailpipe emissions.

### **3.3.2 MEASURE**

The Mobile Emissions Assessment System for Urban and Regional Evaluation (MEASURE) (Bachman et al. 2000) model was developed by Georgia Institute of Technology in the late 1990s. The MEASURE model is developed within a geographic information system (GIS) and employs modal emission rates, varying emissions according to vehicle technologies and modal operation (cruise, acceleration, deceleration, idle). The model emission rate database consists of more than 13,000 laboratory tests conducted by the EPA and CARB using standardized test cycle conditions and alternative cycles (Bachman 1998). The aggregate modal model within MEASURE employs emission rates based on theoretical engine-emissions relationships. The relationships are dependent on both modal and vehicle technology variables, and they are “aggregate” in the sense that they rely on bag data to derive their modal activities (Washington et al. 1997a). Emission rates were statistically derived from the emission rate data as a function of operating mode power demand surrogates. The model uses statistical techniques to predict emission rates using a process that utilizes the best aspects of hierarchical tree-based regression (HTBR) and ordinary least squares regression (OLS) (Breiman et al. 1984). HTBR is used to reduce the number of predictor variables to a manageable number, and to identify useful interactions among the variables; then OLS regression techniques are applied until a satisfactory model is obtained (Fomunung et al. 2000). Vehicle activity variables include average speed, acceleration rates, deceleration rates, idle time, and surrogates for power demand. The MEASURE model for light-duty vehicles was completed in 2000.

MEASURE provides the following benefits since it has been developed under the GIS platform (Bachman et al. 2000): 1) manages topographical parameters that affect emissions; 2) calculates emissions from vehicle modal activities; 3) allows a ‘layered’ approach to indi-

vidual vehicle activity estimation; and 4) aggregates emission estimates into grid cells for use in photochemical air quality models.

### 3.3.3 MOVES

To keep pace with new analysis needs, modeling approaches, and data, the U.S. EPA's Office of Transportation and Air Quality (OTAQ) is developing a modeling system termed MOVES (Koupal et al. 2004, U.S. EPA 2001a). This new system will estimate emissions for on-road and non-road sources, cover a broad range of pollutants, and allow multiple scale analysis, from fine-scale analysis to national inventory estimation. In the future, MOVES will serve as the replacement for MOBILE6 and NONROAD (U.S. EPA 2001a). This project was previously known as the New Generation Mobile Source Emissions Model (NGM) (U.S. EPA 2001a).

The current plan for MOVES is to use vehicle specific power (VSP) as a variable on which emission rates can be based (Koupal et al. 2002). The VSP approach to emissions characterization was developed by Jimenez-Palacios (Jimenez-Palacios 1999). VSP is a function of speed, acceleration, road grade, etc., as shown in Equation 3-3:

$$VSP = v \times (a \times (1 + \epsilon) + g \times \text{grade} + g \times C_R) + 0.5\rho \times C_D \times A \times v^3 / m \quad (\text{Equation 3-3})$$

- where:
- v: vehicle speed (assuming no headwind) (m/s)
  - a: vehicle acceleration (m/s<sup>2</sup>)
  - ε: mass factor accounting for the rotational masses (~0.1) - constant
  - g: acceleration due to gravity (m/s<sup>2</sup>)
  - grade: road grade (ratio of rise to run)
  - C<sub>R</sub>: rolling resistance (~0.0135)
  - μ: air density (1.2)
  - C<sub>D</sub>: aerodynamic drag coefficient (dimensionless)
  - A: the frontal area (m<sup>2</sup>)
  - m: vehicle mass (metric tons)

The basic concept of MOVES starts with the characterization of vehicle activity and the development of relationships between characterized vehicle activity and energy consumption, and between energy consumption and vehicle emission (Nam 2003). The U.S. EPA established a modal binning approach, developed using VSP, to characterize the relationship between vehicle activity and energy consumption. Originally, a total of 14 modal bins were developed based on different VSP ranges (U.S. EPA 2001a). This approach was revised in two different ways. U.S. EPA refined the VSP binning approach by the association of second-by-second speed, engine

rpm, and acceleration rates, and the original 14 VSP binning approaches are revised with the combination of five different speed operating modes and redirected to a total of 37 VSP bins (Koupal et al. 2004). Researchers at North Carolina State University (NCSU) divided each bin into four strata representing two engine sizes and two odometer reading categories, and this approach was referred to as the “56-bin” approach. (U.S. EPA 2002b).

Another important conceptual model for MOVES was developed by NCSU in 2002 (Frey et al. 2002). Dr. Frey summarized the conceptual analytical methodology in the report “Recommended Strategy for On-Board Emission Data Analysis and Collection for the New Generation Model” (Frey et al. 2002). This method uses power demand estimate (P) as a variable on which emission rates can be based (Frey et al. 2002) as shown in Equation 3-4.

$$P = v \times a \quad \text{(Equation 3-4)}$$

where: P : power demand (mph<sup>2</sup>/sec)  
v : vehicle speed (mph)  
a : vehicle acceleration in (mph/s)

This method uses on-board emissions data where data are collected under real-world conditions to develop a modal emission model which can estimate emissions at different scales such as microscale, mesoscale, and macroscale. The philosophy is similar to MEASURE (Fomung 2000), which first segregated the data into four modes based on suitable modal definitions, then developed an OLS regression model for each mode using explanatory variables selected by HTBR techniques. These explanatory variables include model year, humidity, temperature, altitude, grade, pressure, and power. Second and third powers of speed and acceleration were also included in the regression analysis.

### 3.3.4 HDDV-MEM

The researchers in Georgia Institute of Technology have developed a beta version of HDDV-MEM, which is based on vehicle technology groups, engine emission characteristics, and vehicle modal activity (Guensler et al. 2005). The HDDV-MEM first predicts second-by-second engine power demand as a function of on-road 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 vehicle technology group), an engine power module, and an emission rate module. The model framework is illustrated in Figure 3-4.

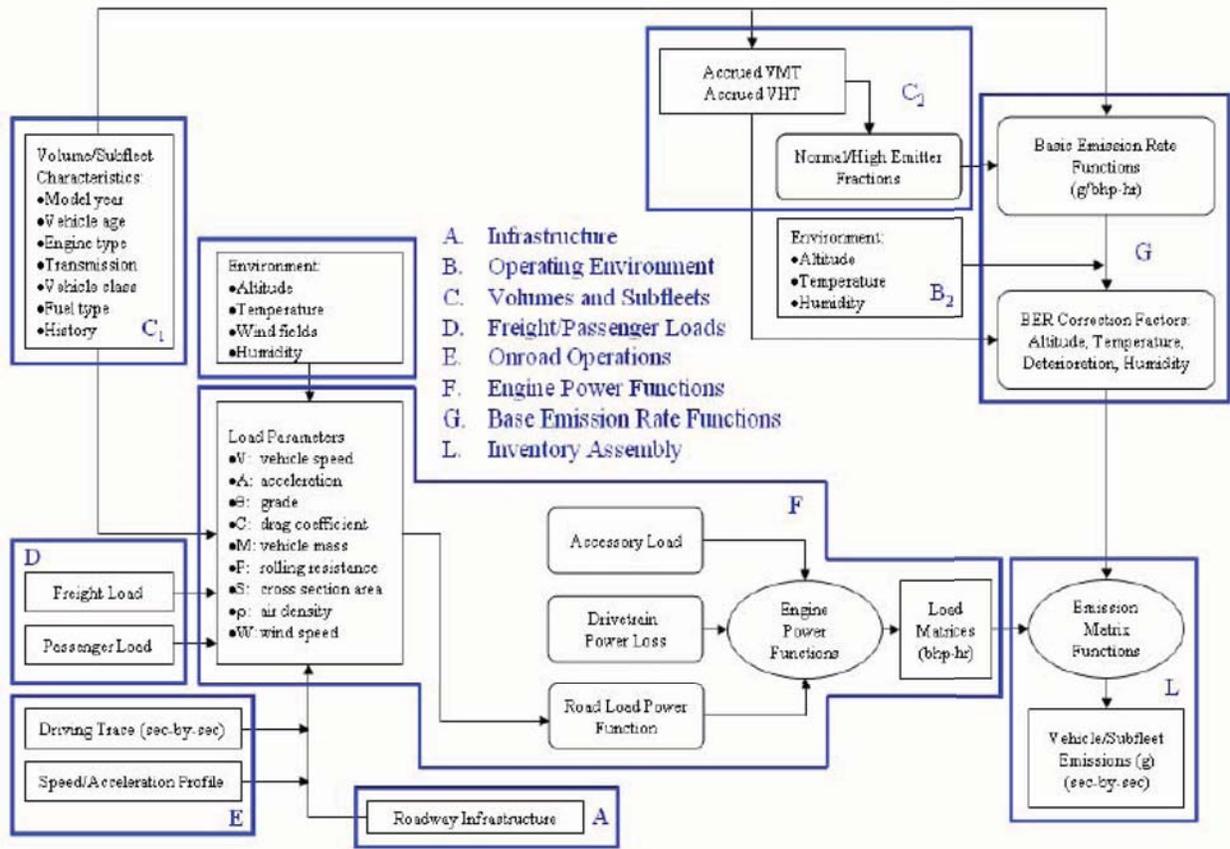


Figure 3-4 A Framework of Heavy-Duty Diesel Vehicle Modal Emission Model (Guensler et al. 2005)

### 3.3.4.1 Model Development Approaches

The HDDV-MEM modeling framework is designed for transportation infrastructure implementation on link-by-link basis. While the modeling routines are actually amenable to implementation on a vehicle-by-vehicle basis, the large number of vehicles operating on infrastructure links precludes practical application of the model in this manner. As such, the model framework capitalizes upon previous experience gained in development of the MEASURE modeling framework, in which vehicle technology groups were employed. A new heavy-duty vehicle visual classification scheme, which is an EPA and Federal Highway Administration (FHWA) hybrid vehicle classification scheme developed by Yoon et al. (Yoon et al. 2004b), classified vehicle technology groups by engine horsepower ratings, vehicles GVWR, vehicle configurations, and vehicle travel characteristics (Yoon 2005c). On the other hand, the MEASURE model employs load surrogates for the implementation of a light-duty modal modeling regime. This new modeling framework directly implements heavy-duty vehicle operating loads and uses these load predictions in the emission prediction process. An engine power module is designed for this task.

Emission rates are first established for various heavy-duty technology groups (engine and vehicle family, displacement, certification group, drivetrain, fuel delivery system, emission control system, etc.) based upon statistical analysis of standard engine dynamometer certification data, or on-road emission rate data when available (Wolf et al. 1998; Fomunung et al. 2000). The following subsets will discuss three main modules in the HDDV-MEM.

### 3.3.4.2 Vehicle Activity Module

The vehicle activity module provides hourly vehicle volumes for each vehicle technology group on each transportation link in the modeled transportation system. The annual average daily traffic (AADT) estimate for each road link is processed to yield vehicle-hours of operation per hour for each technology group (using truck percentages, VMT fraction by vehicle technology group, diesel fraction, hourly volume apportionment of daily travel, link length, and average vehicle speed) (Guensler et al. 2005; Yoon 2005c), as shown in Equation 3-5.

$$VA_{v,h,s|f} = (AADT_s \times (NL_s / TNL) \times HVF_{v,h} \times VF_v \times DF_v) \times (SL_s / AS_v) \quad (\text{Equation 3-5})$$

where: VA: the estimated vehicle activity (veh-hr/hr):  
v: the vehicle technology group  
h: the hour of day  
s: the transportation link  
f: the facility type for the link  
AADT<sub>s</sub>: the annual average daily traffic for the link (number of vehicles)  
NL<sub>s</sub>: the number of lanes in the specific link direction  
TNL: the total number of lanes on the link  
HVF<sub>v,h</sub>: the hourly vehicle fraction  
VF<sub>v</sub>: the VMT fraction for each vehicle technology group  
DF<sub>v</sub>: the diesel vehicle fraction for each technology group  
SL<sub>s</sub>: the link length (miles)  
AS<sub>v</sub>: the link average speed of the technology group (mph)

To estimate on-road running emissions from each link, two sets of calculations are performed. On-road vehicle activity (vehicle-hr) for each hour is multiplied by engine power demand for observed link operations (positive tractive power demand plus auxiliary power demand), and then by baseline emission rates (g/bhp-hr). These calculations are processed separately for each speed/acceleration matrix cell (Yoon et al. 2005b). Emissions from motoring/idling activity are calculated by the determination of the vehicle-hours of motoring/idling activity on each link for each hour and the multiplication of the baseline idle emission rate (g/hr).

### 3.3.4.3 Engine Power Module

Internal combustion engines translate linear piston work (force through a distance) to a crankshaft, rotating the crankshaft and creating engine output torque (work performed in angular rotation). The crankshaft rotation speed (engine speed in revolutions per minute) is a function of engine combustion and physical design parameters (mean effective cylinder pressure, stroke length, connecting rod angle, etc.). The torque available at the crankshaft (engine output shaft) is less than the torque generated by the pistons, in that there are torque losses inside the engine associated with operating a variety of internal engine components. Torque is transferred from the engine output shaft to the driveshaft via the transmission (sometimes through a torque-converter, i.e., fluid coupling) and through a series of gears that allows the drive shaft to rotate at different speeds relative to engine crankshaft speed. The drive shaft rotation is then transferred to the drive axle via the rear differential. The ring and pinion gears in the rear differential translate the rotation of the drive shaft by 90 degrees from the drive shaft running along the vehicle to the drive axle that runs across the vehicle. Torque available at the drive axle is now delivered directly to the drive wheels. This process generates the tractive force used to overcome road friction, wind resistance, road grade (gravity), and other resistive forces, allowing the vehicle to accelerate on the roadway. Figure 3-5 illustrates the primary components of concern.

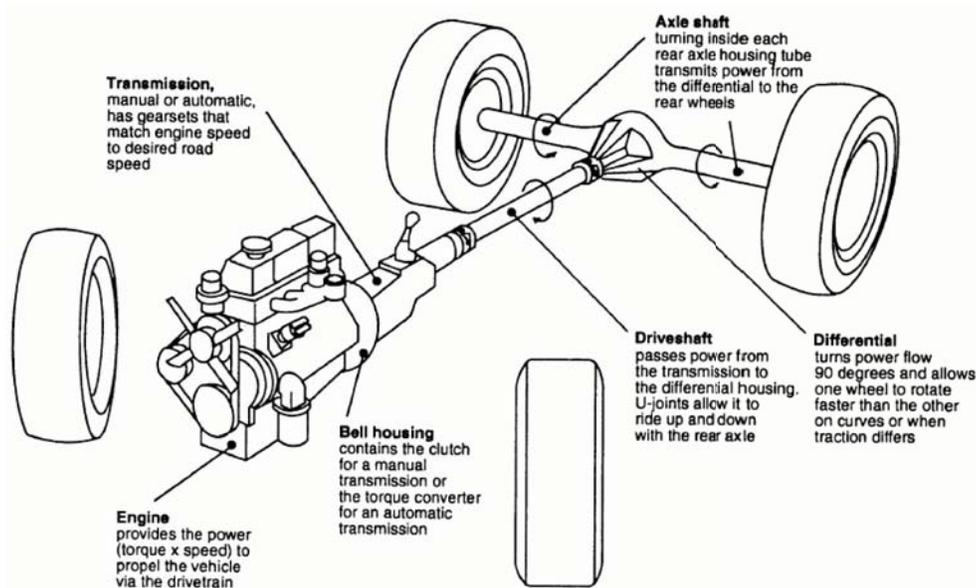


Figure 3-5 Primary Elements in the Drivetrain (Gillespie 1992)

The vehicle drivetrain (engine, torque converter, transmission, drive shaft, rear differential, axles, and wheels) is designed as a system to convert engine torque into useful tractive force

at the wheel-to-pavement interface. When the tractive force is greater than the sum of forces acting against the vehicle, the vehicle accelerates in the direction of travel. Given that on-road speed/acceleration patterns for HDDVs can be observed (or empirically modeled), the modal modeling approach works backwards from observed speed and acceleration to estimate the tractive force (and power) that was available at the wheels to meet the observed conditions. Then, working backwards from tractive force, the model accounts for additional power losses that occurred between the engine and the wheels to predict the total brake-horsepower output of the engine. Force components that reduce available wheel torque and tractive force include:

- Aerodynamic drag, which depends on the frontal area, the drag coefficient, and the square of the vehicle speed;
- Tire rolling resistance, which is determined by the coefficient of rolling resistance, vehicle mass, and road grade (where the coefficient of rolling resistance is a function of tire construction and size; tire pressure; axle geometry, i.e., caster and camber; and whether the wheels are driven or towed);
- Grade load, which is determined by the roadway grade and vehicle mass; and
- Inertial load, which is determined by the vehicle's mass and acceleration.

The tractive force required at the interface between the tires and the road to overcome these resistive forces and provide vehicle acceleration can be described by (Gillespie 1992), as shown in Equation 3-6:

$$F_T = F_D + F_R + F_W + F_I + ma \quad (\text{Equation 3-6})$$

- where:
- $F_T$ : the tractive force available at the wheels (lbf)
  - $F_D$ : the force necessary to overcome aerodynamic drag (lbf)
  - $F_R$ : the force required to overcome tire rolling resistance (lbf)
  - $F_W$ : the force required to overcome gravitational force (lbf)
  - $F_I$ : the force required to overcome inertial loss (lbf)
  - $m$ : the vehicle mass (lbm)
  - $a$ : the vehicle acceleration (ft/sec<sup>2</sup>)

Load prediction models could employ a wide variety of aerodynamic drag (Wolf-Heinrich 1998) and rolling resistance functional forms, some of which may be more appropriate for certain vehicle designs and at certain vehicle speeds. Note that vehicle mass is a critical parameter that must be included in the load-based modeling approach. Therefore, estimates of gross

vehicle weight must be included in any transit (vehicle weight plus passenger loading) or heavy-duty truck (vehicle weight plus cargo payload) application. The following subsections describe each force in Equation 3-6, taken from Yoon et al. (Yoon et al. 2005a).

### Aerodynamic Drag Force ( $F_D$ )

As a vehicle moves forward through the atmosphere, drag forces are created at the interface of the front of the vehicle and by the vacuum generated at the tail of the vehicle. The flow of the air around the vehicle creates a very complex set of forces providing both resistance to forward motion and vehicle lift. The net aerodynamic drag force is a function of air density, aerodynamic drag coefficient, vehicle frontal area, and effective vehicle velocity, as shown in Equation 3-7 (Yoon et al. 2005a).

$$D = \left(\frac{\rho}{2g}\right) \times C_d \times A_f \times V_e^2 \quad (\text{Equation 3-7})$$

where:  $F_D$ : aerodynamic drag force  
 $\rho$ : the air density (lb/ft<sup>3</sup>)  
 $g$ : the acceleration of gravity (32.2 ft/sec<sup>2</sup>)  
 $C_d$ : the aerodynamic drag coefficient  
 $A_f$ : the vehicle frontal area (ft<sup>2</sup>)  
 $V_e$ : the effective vehicle velocity (ft/sec)

### Rolling Resistance Force ( $F_R$ )

Rolling resistance force is the sum of the forces required to overcome the combined friction resistance at the tires. Tires deform at their contact point with the ground as they roll along the roadway surface. Rolling resistance is caused by contact friction, the tires' resistance to deformation, aerodynamic drag at the tire, etc. The force required to overcome rolling resistance can be expressed with rolling resistance coefficient, vehicle weight, and road grade, as shown in Equation 3-8 (Yoon et al. 2005a).

$$F_R = C_r \times m \times g \times \cos(\theta) \quad (\text{Equation 3-8})$$

where:  $F_R$ : force required to overcome rolling resistance  
 $C_r$ : the rolling resistance coefficient  
 $\theta$ : the road grade (degrees)  
 $m$ : vehicle mass in metric tons  
 $g$ : acceleration due to gravity

### Gravitational Weight Force ( $F_w$ )

The gravitational force components account for the effect of gravity on vehicle weight when the vehicle is operating on a grade. The grade angle is positive on uphill grades (generating a positive resistance) and negative on downgrades (creating a negative resistance), as shown in Equation 3-9 (Yoon et al. 2005a).

$$F_w = m \times g \times \sin(\theta) \quad (\text{Equation 3-9})$$

where:  $F_w$ : gravitational weight force  
 $m$ : vehicle mass in metric tons  
 $g$ : acceleration due to gravity  
 $\theta$ : the road grade (degrees)

### Drivetrain Inertial Loss ( $F_I$ )

The engine, transmission, drive shaft, axles and wheels are all in rotation. The rotational speed of each component depends upon the transmission gear ratio, the final drive ratio, and the location of the component in the drive train (i.e., the total gear ratio between each component and the wheels). The rotational moment of inertia of the various drivetrain components constitutes a resistance to change in motion. The torque delivered by each rotating component to the next component in the power chain (engine to clutch/torque converter, clutch/torque converter to transmission, transmission to drive shaft, drive shaft to axle, axle to wheel) is reduced by the amount necessary to increase angular rotation of the spinning mass during vehicle acceleration. Given the torque loss at each component, the reduction in motive force available at the wheels due to inertial losses along the drivetrain can be modeled (Wolf-Heinrich 1998). This model term is most significant under low speed acceleration conditions, such as vehicle operation in truck and rail yards where vehicles are lugging heavy loads over short distances. However, as will be discussed later, significant new data will be required to incorporate the inertial loss effects into modal models, as shown in Equation 3-10 (Yoon et al. 2005a).

$$F_I = \frac{a \times I_{EFF}}{r^2} = \frac{a \times [(I_w + (G_d^2 \times I_D)) + (G_t^2 \times G_d^2) \times (I_E + I_t)]}{r^2} \quad (\text{Equation 3-10})$$

where:  $a$ : the acceleration in the direction of vehicle motion (ft/sec<sup>2</sup>)  
 $I_{EFF}$ : the effective moment of inertia (ft- lbf -sec<sup>2</sup>)

$I_W$ :	the rotational moment of inertia of the wheels and axles (ft-lbf -sec <sup>2</sup> )
$I_D$ :	the rotational moment of inertia of the drive shaft (ft-lbf -sec <sup>2</sup> )
$I_T$ :	the rotational moment of inertia of the transmission (ft-lbf -sec <sup>2</sup> )
$I_E$ :	the rotational moment of inertia of the engine (ft-lbf -sec <sup>2</sup> )
$G_t$ :	the gear ratio at the engine transmission
$G_d$ :	the gear ratio in the differential
$r$ :	wheel radius (ft)

### Power Demand

Using the equations outlined above, the total engine power demand, which is the combination of tractive power and auxiliary power demands, can be expressed in Equation 3-11 (Yoon et al. 2005a):

$$P = \left[ \left( \frac{V}{550} \right) \times (F_D + F_R + F_W + F_I + ma) \right] + AP \quad \text{(Equation 3-11)}$$

where	$P$ :	total engine power demand
	$V$ :	the vehicle speed (ft/s)
	$F_D$ :	the force necessary to overcome aerodynamic drag (lbf)
	$F_R$ :	the force required to overcome tire rolling resistance (lbf)
	$F_W$ :	the force required to overcome gravitational force (lbf)
	$F_I$ :	the force required to overcome inertial loss (lbf)
	$m$ :	the vehicle mass (lbm)
	$a$ :	the vehicle acceleration (ft/sec <sup>2</sup> )
	$AP$ :	the auxiliary power demand (bhp)
	550:	the conversion factor to bhp

#### 3.3.4.4 Emission Rate Module

The emission rate module provides work-related emission rates (g/bhp-hr) and idle emission rates (g/hr) for each technology group. The basic application of the HDDV-MEM incorporates a simple emission rate modeling approach. The predicted engine power demand (bhp) for each second of vehicle operation is multiplied by emission rates in gram/bhp-sec for a given bhp load. Technology groups (i.e., vehicles that perform similarly on the certification tests) are established based upon the engine and control system characteristics and each technology group is assigned a constant g/bhp-sec emission rate based upon regression tree and other statistical analysis of certification data. Under the assumption that testing cycles represent the typical modal activities undertaken by on-road activities, such emission rates are applied to on-road activity data. Given the large repository of certification data, detailed statistical analysis of the certification

test results can be used to obtain applicable emission rates for these statistically derived vehicle technology groups. The data required for analysis must come from chassis dynamometer (the engine remains in the vehicle and the vehicle is tested on a heavy-duty treadmill) and on-road test programs in which second-by-second grams/second emission rate data have been collected concurrently with axle-hp loads.

At this moment, HDDV-MEM accepts EPA's baseline running emission rate data as work-related emission rates and EMFAC2002 idling emission rate test data as idle emission rates. Diesel vehicle registration fractions and annual mileage accumulation rates are employed to develop calendar year emission rates for each technology group. In the future, a constant emission rate need not be used as more refined testing data become available. Linear, polynomial, or generalized relationships can be established between gram/second emission rate and tractive horsepower (axle horsepower) and other variables. Sufficient testing data are required to establish statistically significant samples for each technology group.

#### 3.3.4.5 Emission Outputs

HDDV-MEM outputs link-specific emissions in grams per hour (g/hr) for VOCs, CO, NO<sub>x</sub>, and PM for each vehicle type. Toxic air contaminant emission rates (benzene, 1, 3-butadiene, formaldehyde, acetaldehyde, and acrolein) are also estimated in grams/hour for each vehicle type using the MOBILE6.2-modeled ratios of air toxics to VOC for each calendar year. HDDV-MEM provides not only hourly emissions, but also aggregated total daily emissions (in accordance with input command options). The structure of output files, which provide link-specific hourly emissions, can be directly incorporated with roadway network features in a GIS environment for use in interactive air quality analysis in various spatial scales, i.e., national, regional, and local scales (Guensler et al. 2005; Yoon 2005c).

## CHAPTER 4

### 4. EMISSION DATASET DESCRIPTION AND POST-PROCESSING PROCEDURE

Using second-by-second data collected from on-road vehicles (Brown et al. 2001, Ensfield 2002), the research effort reported here developed models to predict emission rates as a function of on-road 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 datasets collected across space and time. There are two second-by-second data sets in which emission rate and applicable load and vehicle activity data have been collected in parallel (Brown et al. 2001, Ensfield 2002). One database was a transit bus dataset, collected on diesel transit buses operated by Ann Arbor Transit Authority (AATA) in 2001 (Ensfield 2002), and another dataset was heavy HDV (HDV8B) dataset prepared by National Risk Management Research Laboratory (NRMRL) in 2001 (Brown et al. 2001). Each is summarized in the following sections.

#### 4.1 Transit Bus Dataset

Transit bus emissions dataset was prepared by Sensors, Inc. (Ensfield 2002). Sensors, Inc. has supplied gas analyzers and portable emissions testing systems worldwide for over three decades. Their products, SEMTECH-G for gasoline powered vehicles, and SEMTECH-D for diesel powered vehicles, are commercially available for on-vehicle emission test applications. In October 2001, Sensors, Inc. conducted real-world, on-road emissions measurements of 15 heavy-duty transit buses for U.S. EPA (Ensfield 2002). Transit buses were provided by the AATA and all of them were New Flyer models with Detroit Diesel Series 50 engines. Table 4-1 summarizes the buses tested for U.S. EPA.

Table 4-1 Buses Tested for U.S. EPA (Ensfield 2002)

Bus #	Bus ID	Model Year	Odometer	Engine series	Displacement (liters)	Peak Torque (lb-ft)	Test Date
1	BUS360	1995	270476	SERIES 50 8047 GK40	8.5	890	10/25/2001
2	BUS361	1995	280484	SERIES 50 8047 GK38	8.5	890	10/25/2001
3	BUS363	1995	283708	SERIES 50 8047 GK37	8.5	890	10/24/2001
4	BUS364	1995	247379	SERIES 50 8047 GK42	8.5	890	10/24/2001
5	BUS372	1995	216278	SERIES 50 8047 GK41	8.5	890	10/26/2001
6	BUS375	1996	211438	SERIES 50 8047 GK39	8.5	890	10/25/2001
7	BUS377	1996	252253	SERIES 50 8047 GK36	8.5	890	10/24/2001
8	BUS379	1996	260594	SERIES 50 8047 GK35	8.5	890	10/23/2001
9	BUS380	1996	223471	SERIES 50 8047 GK28	8.5	890	10/23/2001
10	BUS381	1996	200459	SERIES 50 8047 GK29	8.5	890	10/22/2001
11	BUS382	1996	216502	SERIES 50 8047 GK30	8.5	890	10/17/2001
12	BUS383	1996	199188	SERIES 50 8047 GK31	8.5	890	10/19/2001
13	BUS384	1996	222245	SERIES 50 8047 GK32	8.5	890	10/17/2001
14	BUS385	1996	209470	SERIES 50 8047 GK33	8.5	890	10/18/2001
15	BUS386	1996	228770	SERIES 50 8047 GK34	8.5	890	10/19/2001

#### 4.1.1 Data Collection Method

A total of 15 files were provided for the purpose of model development (Ensfield 2002). Each file represents data collected from different transit buses. Five of these buses were 1995 model year and the rest were 1996 model year. All of the bus test periods lasted approximately two hours. The buses operated along standard Ann Arbor bus routes and stopped at all regular stops although the buses did not board or discharge any passengers. The routes were mostly different for each test, and were selected for a wide variety of driving conditions. All of the bus routes for the test are shown in Figure 4-1.



Figure 4-1 Bus Routes Tested for U. S. EPA (Ensfield 2002).

Sensors, Inc. engineers performed the instrument setup and data collection for all the buses. Test equipment, SEMTECH-D analyzer, is shown in Figure 4-2. Because engine computer vehicle interface (SAE J1708) data were collected at 10 Hz, Sensors, Inc. engineers manually started and stopped data collections at approximately 30 minute intervals to keep file size manageable. A total of four trip files were generated per bus. Zero drift was checked between data collections. Then four files for each bus were combined into one file after post-processing. The time for each bus is thus sometimes not continuous. To derive other variables easily, like acceleration, and keep data manageable or other purposes, data for each bus were separated into trips based on continuous time. After this processing, there were 62 “trips” in the transit bus database.



Figure 4-2 SEMTECH-D in Back of Bus (Ensfield 2002)

#### 4.1.2 Transit Bus Data Parameters

Each of the 15 data files share the same format. The data fields included in each file are summarized in Table 4-2.

Table 4-2 Transit Bus Parameters Given by the U.S. EPA (Ensfield 2002)

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	License number; Engine size; Instrument configuration number
Roadway Characteristics	GPS Latitude (degree); GPS Longitude (degree); GPS Altitude (feet); Grade (%)
Onroad Load Parameters	Vehicle speed (mph); Engine speed (rpm); Torque (lb-ft); Engine power (bhp)
Engine Operating Parameters	Engine load (%); Throttle position (0 – 100%); Fuel volumetric flow rate (gal/s); Fuel specific gravity; Fuel mass flow rate (g/s); Calculated instantaneous fuel economy (mpg); Engine Oil temperature(deg F); Engine oil pressure (kPa); Engine warning lamp (Binary); Engine coolant temperature (deg F); Barometric pressure reported from ECM (kPa); Calculated exhaust flow rate (SCFM)
Environment Conditions	Ambient temperature (deg C); Ambient pressure (mbar); Ambient relative humidity (%); Ambient absolute humidity (grains/lb air)
Vehicle Emission	HC, CO, NO <sub>x</sub> , CO <sub>2</sub> emission (in ppm, g/sec, g/ke-fuel, g/bhp-hr units)

### 4.1.3 Sensors, Inc. Data Processing Procedure

It is helpful to understand how Sensors, Inc. processed the dataset after data collection. This information is very important for data quality assurance and quality control. This section is adapted primarily from the Sensor's field data collection report (Ensfield 2002).

*Data Synchronization:* According to Sensor's report, the analytical instruments, vehicle interface, and global positioning system (GPS) equipment reported data individually to the SEMTECH data logger asynchronously and at differing rates, but with a timestamp at millisecond precision. The first step of the post-processing procedure is to eliminate the extra data by interpolating and synchronizing all the data to 1 Hz. With all the raw data synchronized to the same data rate, it is then time-aligned so that engine data corresponds to emissions data in real time.

*Mass Emissions Calculations:* Mass emissions (gram/second) are calculated by fuel flow method. With access to real-time, second-by-second fuel flow rates, a value for transient mass emissions is computed as shown by the equation below. Using NO as an example, NO mass emissions are calculated on a second-by-second basis (Ensfield 2002).

$$NO_{g/sec} = NO_{fs} \times Fuelflow \quad \text{Equation 4-1}$$

where  $NO_{(g/sec)}$  : NO emissions (grams/second)  
 $NO_{fs}$  : NO emission rate (grams of NO per gram of fuel)  
Fuelflow : flow of fuel per unit time (grams per second).

Fuel specific emissions are the ratios of the mass of each pollutant to the fuel in the combusted air/fuel mixture. The mass fuel flow rate is converted from fuel volumetric flow rate using fuel specific gravity.

*Brake Specific Emissions Calculations:* Engine torque is first computed by applying the engine load parameter, which represents the ratio between current engine torque and maximum engine torque, to the engine lug curve (maximum torque curve). Engine horsepower is then converted from engine torque using engine speed data. Work (bhp-hr) is computed for each second of the test, and brake specific emissions are reported as the sum of the grams of pollutant emitted over the desired interval (one second) divided by the total work.

*Vehicle Speed Validation:* Vehicle speed is a critical parameter that influences the derived parameters, acceleration and emission rates. It is important for researchers to understand

the method of measurement and data accuracy. Sensors, Inc. measured vehicle speed using two methods: vehicle Electronic Control Module (ECM) and Global Positioning System (GPS). Figure 4-3 shows the GPS vs. ECM comparison for Bus 380. The regression analysis shows that the ECM data are around 10% higher than the GPS data, according to Sensors report (Ensfield 2002). Sensors, Inc. researchers believe that this comparison suggests that GPS data may be more reliable for on-road testing. Buses of model year 1995 were equipped with an earlier version ECM that did not provide vehicle speed and GPS velocity data were used in place of the ECM data. Buses of model year 1996 were equipped with the current version ECM that can provide vehicle speed and vehicle speed was reported after validation with the GPS data. GPS data were within 1% accuracy based upon analysis of 10 miles of data (Ensfield 2002).

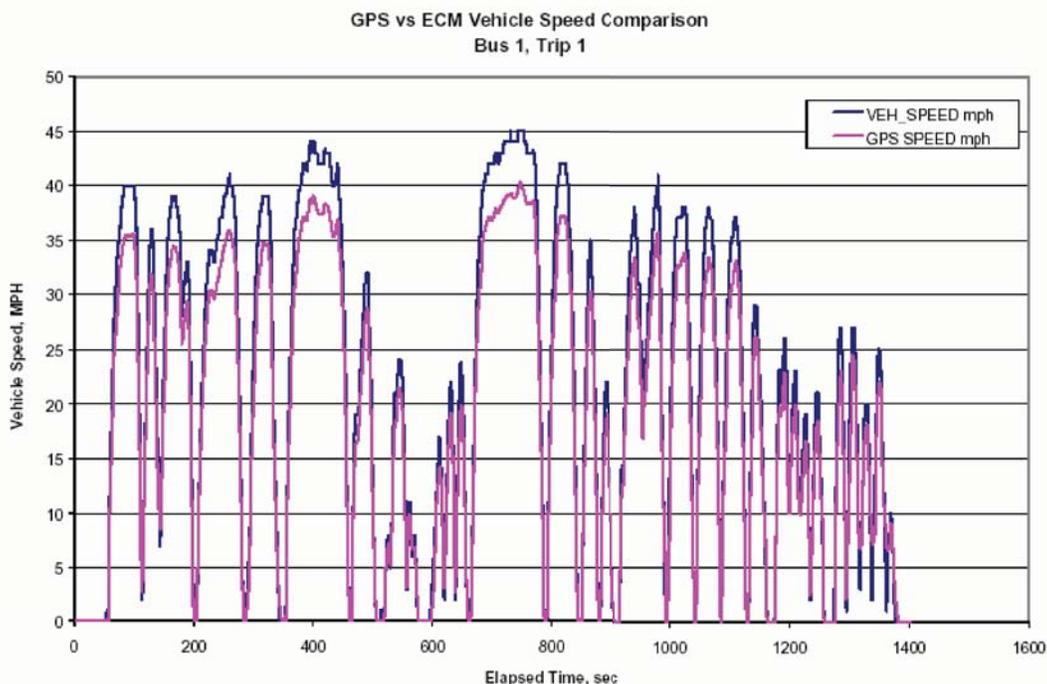


Figure 4-3 Bus 380 GPS vs. ECM Vehicle Speed (Ensfield 2002)

#### 4.1.4 Data Quality Assurance/Quality Check

After understanding the manner in which Sensors, Inc. processed the reported data set, the data set for each bus was screened to check for errors or possible problems. Possible sources of errors associated with data collection should be considered before undertaking data analysis for the development of a model. The types of errors checked are listed below.

*Loss of Data:* Emission data are missing for some buses. For example, bus 382 had missing HC data for 343 seconds. Buses 361, 377 and 384 have similar problems. There might be several reasons for loss of data. Communication between instruments might be lost or a particu-

lar vehicle may have failed to report a particular variable. These records are removed from the test database and not employed in development of HC models because the instantaneous emission values will be recorded as zero, introducing significant bias to the result. Similarly, calculated fuel economy data are missing for some buses.

*Erroneous ECM Data:* There were some cases where certain engine parameters were well outside physical limits, and these erroneous ECM data were filtered out with pre-defined filter limits. The following filter limits (Ensfield 2002) were imposed on the rate of change of RPM, fuel flow, and vehicle speed data:

- Rate of change limit for RPM = 10,000 (RPM)/sec
- Rate of change limit for Fuel flow = 0.003 (gal/sec)/sec
- Rate of change limit for Vehicle speed = 21 (mph)/sec

According to Sensors, Inc. report, these filters remove the data outside the defined limits. The SEMTECH post-processor automatically interpolates between the remaining data, and produces results at 1Hz as before (Ensfield 2002). Because this procedure was finished by manually plotting the ECM parameters and computed mass results, all the buses' data were screened again to check any remaining data spikes for data quality assurance purposes. No such errors were identified for this kind of problem. But the modeler should keep in mind that data could be erroneous because "unreasonable" engine acceleration or deceleration was removed that could have been within reasonable absolute limits.

*GPS Dropouts:* There were a few instances when the GPS lost communication with the satellite for unknown reasons, and these erroneous GPS data were removed manually (Ensfield 2002). To guarantee data quality, the modeler screened all GPS data again to check any remaining erroneous cases. The principles for screening erroneous GPS data are based on the consistency between GPS data and engine parameters. The secondary screening identified that bus 360 data still contained some erroneous GPS data. The questionable area covers the beginning 434 seconds of the whole trip (see Figure 4-4). Their GPS data are shown as red in the left figure. The right figure illustrates the time series plot for checked area. Although GPS signals are reported as some fixed positions in the left figure while vehicle speed data are reported as zero in the right figure, engine speed and engine power in the right figure shows that bus 360 did move during that period. This error might due to GPS dropouts.

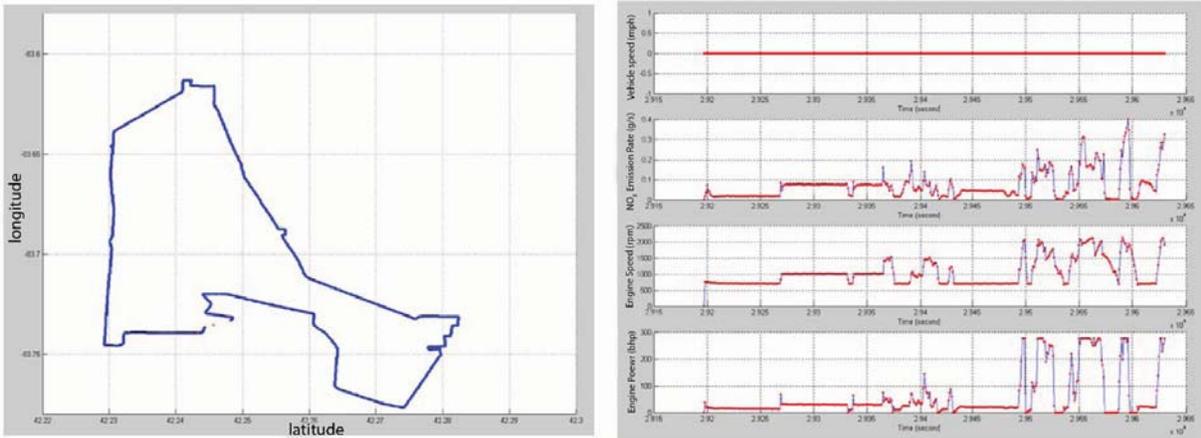


Figure 4-4 Example Check for Erroneous GPS Data for Bus 360 (Ensfled 2002)

Due to GPS dropouts, the GPS signals were reported as some fixed positions. At the same time, the vehicle speed might be reported as zero while other ECM data, such as engine speed and engine power, would show that the bus did move during that period. If the modeler fails to screen and remove such data, these data will be classified as idle mode. Further, these data will cause erroneous analysis result for idle mode. The modeler screened all buses manually and found that six buses had such problems (buses 360,361, 363, 364, 375 and 377). Usually, this type of error was prevalent during the beginning of the bus trip. All erroneous data were removed manually. The correction of the database to remove these erroneous data is critical to model development (initial models associated with development of idle and load-based emission rates were problematic until this database error was identified and corrected by the author).

*Synchronization Errors:* Data were checked for synchronization errors. An example plot of such a check is presented in Figure 4-5 where part of the trip for Bus 360 is used. The selected area covers about 200 seconds. Their GPS data are shown as the green/red part in the left figure. The figure on the right illustrates the time series plot for the area checked. The speed for red points in both figures is 0 mph. Although NO<sub>x</sub> correlates well to engine load and engine speed, vehicle speed doesn't correlate well to engine data and NO<sub>x</sub> emissions data. Bus 360 was equipped with an earlier version ECM that did not provide vehicle speed. GPS velocity data were used in place of the ECM data. According to Sensor's report, data synchronization was only done between emissions data and engine data, not for vehicle speed for emissions data (Ensfled 2002).

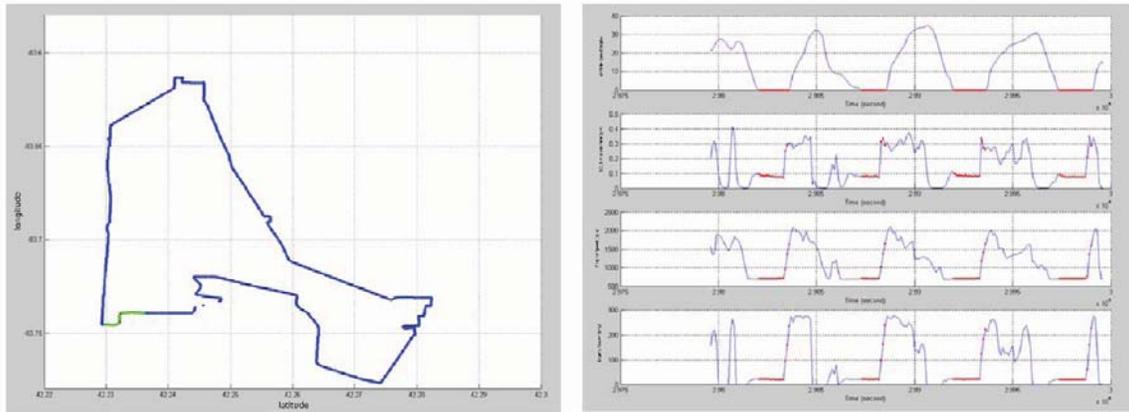


Figure 4-5 Example Check for Synchronization Errors for Bus 360

All bus data were checked for this type of error and such errors were identified in all of the test data for six buses (buses 360, 361, 363, 364, 375, 377). Coincidentally, these six buses had GPS dropout problems, too. From Frey's work (Frey and Zheng 2001), small errors in synchronization do not substantially impact estimate of total trip emissions. Such deviations will influence the estimate for micro-scale analysis. To choose the right delay time to remove the GPS data and vehicle speed data, the author compared the impacts of using a 2-second, 3-second, and 4-second delay. Figure 4-6 illustrates histograms of engine power for zero speed data based on three different proposed time delay options. A 3-second delay is chosen because engine power distribution for zero speed data based on a 3-second delay is more reasonable. Comparing to the 2-second delay results, zero speed data contain fewer data points with higher engine power ( $>150$  brake horsepower) for 3-second delay. Meanwhile, zero speed data contain more data points with lower engine power ( $<20$  brake horsepower) for a 3-second delay than 4-second delay time.

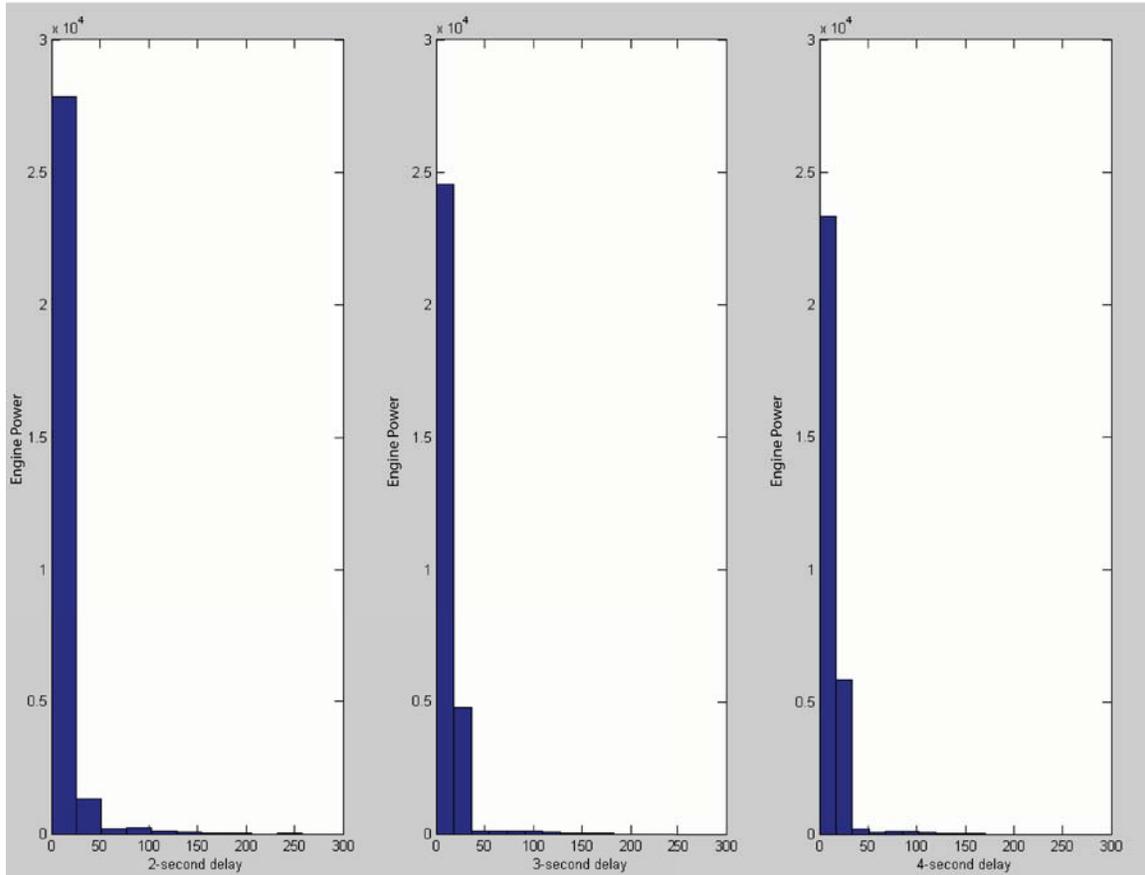


Figure 4-6 Histograms of Engine Power for Zero Speed Data Based on Three Different Time Delays

Road Grade Validation: According to Sensor's report, the GPS data were used for grade calculation. Combining the velocity at time  $t$  with the difference in altitude between time  $t$  and  $t-1$  second, the instantaneous grade is computed as shown in Equation 4-2 (Ensfield 2002).

$$\text{Grade}_t = \frac{\text{velocity}_t}{\text{altitude}_t - \text{altitude}_{t-1}} \quad \text{Equation 4-2}$$

where  $\text{grade}_t$  : Road grade at time  $t$   
 $t$  : time,  $t$  or  $t-1$  second  
 $\text{velocity}_t$  : vehicle speed in feet per second at time  $t$   
 $\text{altitude}$  : altitude in feet at time  $t$  or  $t-1$

The calculation formula can generate significant errors given the uncertainty in the GPS position, particularly at low speeds where there is less of a differential in distance over the one-second interval (Ensfield 2002). In the real world, the maximum recommended grade for use in road design depends upon the type of facility, the terrain on which it is built, and the design speed. Figure 4-7 is directly cited from Traffic Engineering (Roess et al. 2004) to present a

general overview of usual practice. Roess et al. (2004) indicated that these criteria represent a balance between the operating comfort of motorists and passengers and the practical constraints of design and construction in more severe terrains.

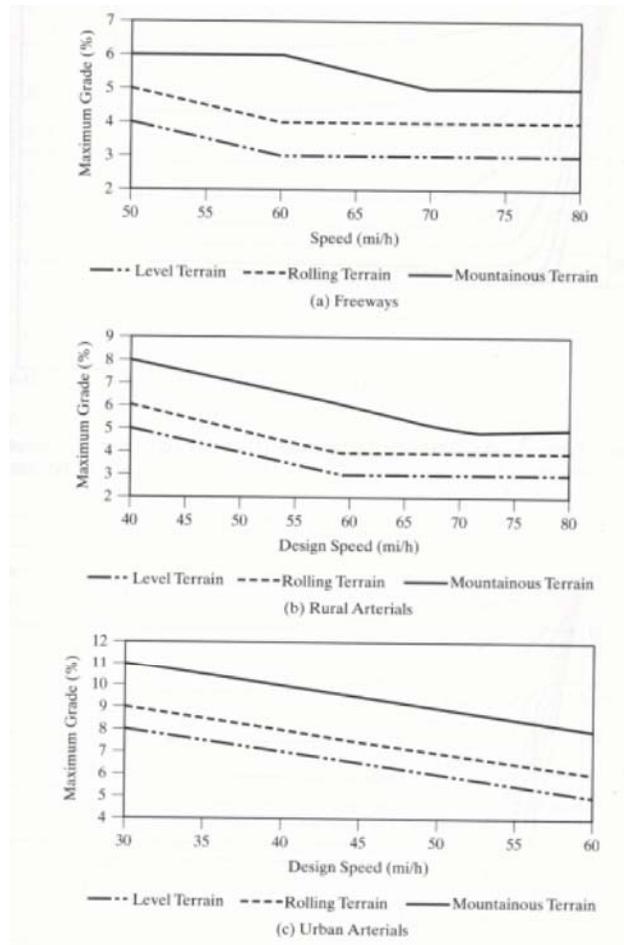


Figure 4-7 General Criteria for Maximum Grades (Roess et al. 2004)

The modeler screened the grade data in the database and found that 0.42% of the data have higher grade (> 10%). Meanwhile, 2% of the road grade data have higher rate of change (> 5%). This means some road grade data are dubious or erroneous. Considering Sensors, Inc. recommendations, road grade data would only be used as reference, and would not be used directly in model development.

#### 4.1.5 Database Formation

The data dictionaries of the source files were reviewed for parameter content. Not all variables reported will be included in explanatory analysis. A standard file structure was designed to accommodate the available format. Emissions rate data with units of grams/second were selected to develop the proposed emission rate model. Because volumetric fuel rate, fuel

specific gravity, and fuel mass flow rate are used to calculate mass emissions (g/s), these variables will be excluded in further analysis. Similarly, because percent engine load, engine torque, and engine speed are used to calculate engine power (brake horsepower), only engine power (bhp) is selected to represent power related variables. Exhaust flow rate is excluded because it is back-computed from the mass emissions generated with the fuel flow method. Fuel economy is excluded because it is 30 second moving average data and computed for a test period by summing the fuel consumed and dividing by the distance traveled. Because GPS data were used for grade calculation and road grade data would only be used as reference, a dummy variable was created to represent different road grade ranges.

At the same time, variables that might be helpful in explaining variability in vehicle emissions were included in the proposed file structure although they were not provided in the original dataset. These variables include model year, odometer reading, and acceleration. Acceleration data were derived from speed data using central difference method. Table 4-3 summarizes the parameter list for explanatory analysis.

Table 4-3 List of Parameters Used in Explanatory Analysis for Transit Bus

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	License number; Model year; Odometer reading; Engine size; Instrument configuration number
Roadway Characteristics	Dummy variable for road grade range
Onroad Load Parameters	Engine power (bhp); Vehicle speed (mph); Acceleration (mph/s)
Engine Operating Parameters	Throttle position (0 – 100%); Engine oil temperature (deg F); Engine oil pressure (kPa); Engine warning lamp (Binary); Engine coolant temperature (deg F); Barometric pressure reported from ECM (kPa)
Environmental Conditions	Ambient temperature (deg C); Ambient pressure (mbar); Ambient relative humidity (%); Ambient absolute humidity (grains/lb air)
Vehicle Emissions	HC, CO, NO <sub>x</sub> emission (in g/sec)

#### 4.1.6 Data Summary

After the post-processing procedure was completed, the summary of the emissions and activity data as well as environmental and roadway characteristics is given in Table 4-4.

Table 4-4 Summary of Transit Bus Database

Bus ID	360	361	363	364	372	375	377	379	380	381	382	383	384	385	386
Numbers of Seconds of Data	7606	5153	7623	5284	5275	7323	7809	7880	8006	7282	3136	7943	8453	8423	10339
Vehicle Operation															
Average Speed (mph)	11.116	25.804	14.626	19.046	21.45	16.814	12.518	15.118	13.035	16.335	19.947	18.253	18.262	16.559	17.319
Average Engine Power (bhp)	71.952	87.536	65.822	79.599	72.395	86.307	78.121	84.82	72.987	65.724	85.224	67.249	64.199	62.512	62.979
Emission Data															
Average CO (g/s)	0.029652	0.018965	0.022419	0.020627	0.016582	0.031844	0.028571	0.030731	0.052504	0.034294	0.052822	0.026207	0.036183	0.023527	0.047062
Average Nox (g/s)	0.11049	0.1484	0.066047	0.12341	0.087625	0.13697	0.074597	0.10658	0.10393	0.090166	0.14089	0.11873	0.10457	0.095998	0.10635
Average HC (g/s)	0.001838	0.001304	0.000239	0.003492	0.002371	0.001377	0.000557	0.001807	0.001073	0.000609	0.00132	0.001803	0.00137	0.001698	0.00147
Environmental Characteristics															
Average Ambient Temperature (deg C)	20.358	16.666	25.623	20.358	21.375	17.5	26.012	23.788	23.648	22.465	21.746	21.282	18.17	21.842	20.389
Average Ambient Pressure (mbar)	977.16	971.08	965.69	985.58	982.05	977.52	973.08	974.27	973.22	987.82	994.71	983.55	992.7	991.34	985.65
Average Humidity (grains/ (lb air))	24.512	26.745	88.396	33.227	32.494	24.394	70.653	70.818	67.525	46.016	27.868	44.646	22.494	29.766	37.239

## 4.2 Heavy-duty Vehicle Dataset

The heavy-duty vehicle emission dataset is prepared by the U.S. EPA National Risk Management Research Laboratory (NRMRL) (U.S. EPA 2001b). EPA's Onroad Diesel Emissions Characterization (ODEC) facility has been collecting real-world gaseous emissions data for many years (U.S. EPA 2001c). The on-road facility incorporated a 1990 Kenworth T800 tractor-trailer as its test vehicle to collect this database. When this truck was purchased, it had already logged over 900,000 miles and was due for an overhaul of its Detroit Diesel Series 60 engine. The vehicle was tested prior to having this work done and after the overhaul. NRMRL collected the test data for U.S. EPA from 1999 to 2000 and included all the results and findings in a report titled: "Heavy Duty Diesel Fine Particulate Matter Emissions: Development and Application of On-Road Measurement Capabilities" (U.S. EPA 2001c).

### 4.2.1 Data Collection Method

The general capabilities of the ODEC facility are shown in Figure 4-8. The facility is designed to collect data while traveling along the public roadways using a 1990 Kenworth T800 tractor-trailer. This truck was tested using two types of tests. During 'parametric' testing, the truck systematically follows a test matrix representing the full range of load, grade, speed and acceleration conditions. During 'highway' testing, the truck travels along an interstate highway with no specific agenda other than covering the distance safely and efficiently; speed and acceleration vary randomly with grade, speed limit, and traffic effects. Tables 4-5 and 4-6 summarize the tests finished by NRMRL for U.S. EPA.

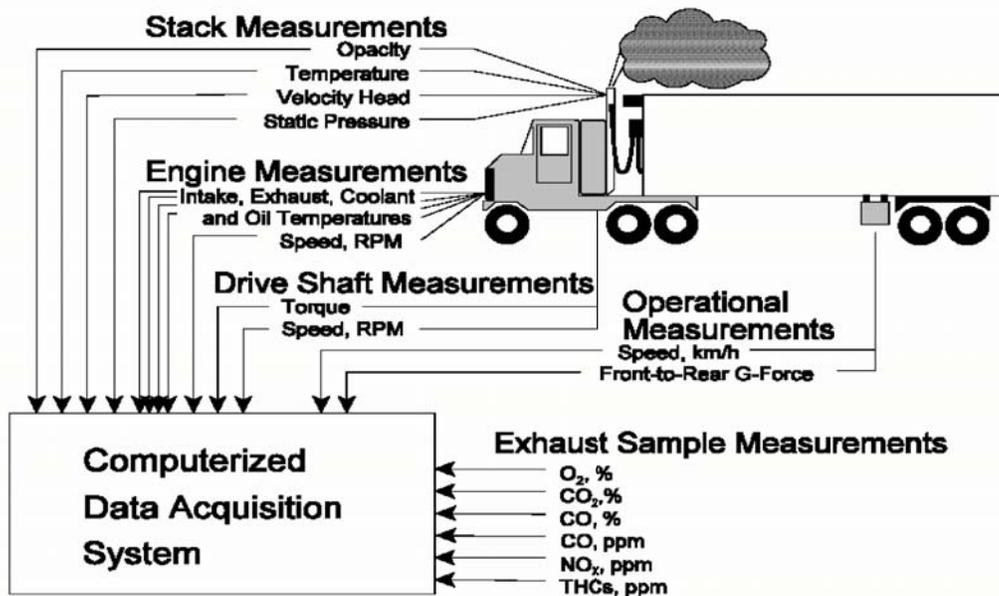


Figure 4-8 Onroad Diesel Emissions Characterization Facility (U.S. EPA 2001c)

Table 4-5 Onroad Tests Conducted with Pre-Rebuild Engine

Test ID	Load lb GCW	Grade(s) %	Comments
3F00V	79280	Zero	Constant Speed Testing
3F00C	79280	Zero	Cost Down & Acceleration
3F00A	79280	Zero	Governed Acceleration & Short-shift Acceleration
3H00V	61060	Zero	Constant Speed Testing
3H00C	61060	Zero	Cost Down & Acceleration
3H00A	61060	Zero	Governed Acceleration & Short-shift Acceleration
3E00V	42840	Zero	Constant Speed Testing
3E00C	42840	Zero	Cost Down & Acceleration
3E00A	42840	Zero	Governed Acceleration & Short-shift Acceleration
3F0GA	79280	Zero	Governed Acceleration
3F0SA	79280	Zero	Short-shift Acceleration
3F0V	79280	Zero	Constant Speed Testing
3H0GA	61060	Zero	Governed Acceleration
3H0SA	61060	Zero	Short-shift Acceleration
3H0V	61060	Zero	Constant Speed Testing
3E0GA	42840	Zero	Governed Acceleration
3E0SA	42840	Zero	Short-shift Acceleration
3E0V	42840	Zero	Constant Speed Testing
3F3&6	79280	3.1, 6.0	Uphill Grade Tests
3H3&6	61060	3.1, 6.0	Uphill Grade Tests
3E3&6	42840	3.1, 6.0	Uphill Grade Tests
3F-SEQ	79280	Zero	Dyno Sequence Simulations
3DRI	79280	Various	Open Highway Tests – Tunnel
3FIL	61060	Various	Open Highway Tests – Filters
3DIOX*	61060	Various	Open Highway Tests – Dioxin

\*Note: These tests are not available.

Table 4-6 Onroad Tests Conducted with Post-Rebuild Engine

Test ID	Load lb GCW	Grade(s) %	Comments
5F0V	74000	Zero	Constant Speed Testing
5F0C*	74000	Zero	Cost Down & Acceleration
5F0A*	74000	Zero	Governed Acceleration & Short-shift Acceleration
5H0V	61440	Zero	Constant Speed Testing
5H0C*	61440	Zero	Cost Down & Acceleration
5H0A*	61440	Zero	Governed Acceleration & Short-shift Acceleration
5E0V	42600	Zero	Constant Speed Testing
5E0C*	42600	Zero	Cost Down & Acceleration
5E0A*	42600	Zero	Governed Acceleration & Short-shift Acceleration
5F3&6	74000	3.1, 6.0	Uphill Grade Tests
5H3&6	61440	3.1, 6.0	Uphill Grade Tests
5E3&6	42600	3.1, 6.0	Uphill Grade Tests
5F-SEQ*	74000	Zero	Dyno Sequence Simulations
5Plume	61440	Various	Open Highway Tests – Plume
5NOxB*	61440	Various	Open Highway Tests – Burst
5DIOX*	61440	Various	Open Highway Tests – Dioxin

\*Note: These test results are not available.

#### 4.2.2 Heavy-duty Vehicle Data Parameters

A total of 42 files were collected for the pre-rebuild engine and a total of 38 file collected for the post-rebuild engine. Each file represents data collected for a different engine and test. Preliminary analysis of individual files indicated that the format of files was same for all available files. The data fields included in each file are summarized in Table 4-7 below.

Table 4-7 List of Parameters Given in Heavy-duty Vehicle Dataset Provided by U.S. EPA

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	Vehicle make/model; Model year; Engine type; Engine Rating; Vehicle maintenance history
Onroad Load Parameters	Truck load weight (lb); Vehicle speed (mph); Measured engine power (bhp)
Engine Operating Parameters	Engine speed (RPM); Shaft volts; Torque volts; Fuel H/C ratio; Fuel factor; Engine intake air temperature (deg F); Engine exhaust air temperature (deg °F); Engine coolant temperature (deg °F); Engine oil temperature (deg °F)
Environment Conditions	Barometric pressure (inches Hg); Ambient humidity (%)
Vehicle Emissions	CO, NO <sub>x</sub> , and HC emission (in ppm, g/hr, g/kg fuel and g/hp-hr units)

#### 4.2.3 Data Quality Assurance/Quality Control Check

Although a total of 80 tests were finished for that project, preliminary screening found that there were some test files missing from the data DVD provided by U.S. EPA to the researchers. The missing test files include: 3DIOX, 5E0C, 5H0C, 5F0C, 5F-SEQ, 5NOxB, and 5DIOX. For quality assurance purposes, the available data files were screened to check for errors or possible problems. Possible sources of errors for data collection should be considered before developing the model. The types of errors checked are listed below.

*Loss of Data:* Measured horsepower (engine power) and emission data were missing for some tests. Tests 3F-SEQ, 3FIL1, 3FIL2, and 3FIL3 had no measured horsepower data for the entire test. These test files couldn't be included in emission model development. In addition, tests 3E00A, 3E00C, 3E00V, 3F0GA, 3F0SA, 3F0V, 3H0SA, 3FIL4, 3FIL5, 3FIL7, 3FIL8, 3FIL9, 3FIL10, and 5H0V had no HC emission data. This problem will be fixed by removing these tests for HC emission model development. Test 3H0SA also had no CO emission data and this problem will be treated by removing this test for CO emission model development.

*Duplicated Records:* A notable issue was duplicate records with different emission values for same time in some test files. After communicating with Mr. Brown who prepared this dataset for EPA, the reason was identified: the data were recorded at rates as high as 10 Hz to improve the resolution of the data. To keep consistent with other test files, these data were post-processed as one data point for each second.

*Erroneous Load Data:* The “measured horsepower” field is engine power data calculated from measurement of the drive shaft torque and rotational speed. Results from the literature

review show that engine power is a major explanatory variable of possible erroneous load data. This variable was screened to check for errors or possible problems. An example of a check of measured horsepower is given in Figure 4-9. The observed relationship between measured horsepower and engine speed is to some extent a relationship between vehicle speed and engine speed which can be found in “Fundamentals of Vehicle Dynamics” (Gillespie 1992). At a given gear ratio, the relationship between engine speed and road speed is to some extent a linear relationship. The geometric progression in the left figure reflects the choices made in selection of transmission gear ratios. The right figure shows a problematic linear relationship between measured horsepower and vehicle speed. Essentially, the right figure appears to show no gear changes as vehicle speed increases, indicating that measured horsepower has been calculated incorrectly for this test. Such problems exist in the series of tests 3DRI and test 5Plume. These test files were removed from emission model development.

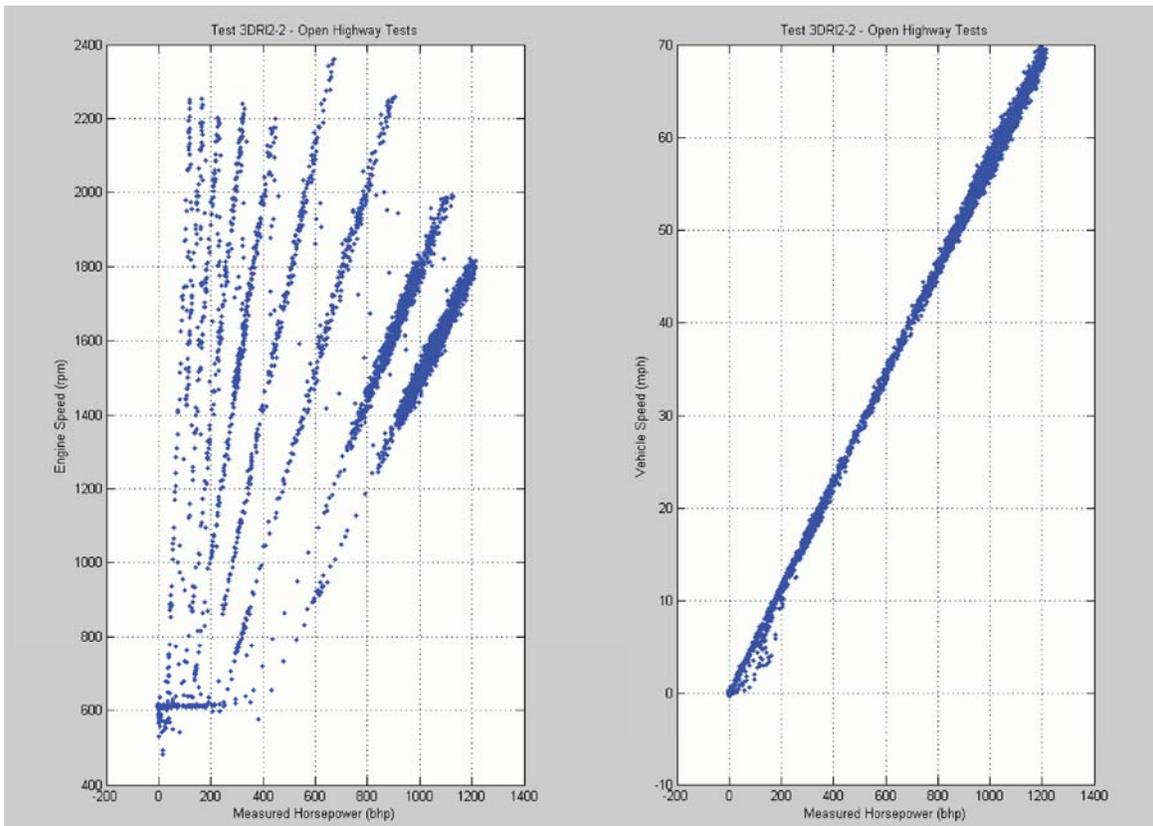


Figure 4-9 Example Check for Erroneous Measured Horsepower for Test 3DR12-2

*Vehicle Speed Validation:* The author reviewed NRMRL’s report (U.S. EPA 2001c) related to vehicle speed validation. Vehicle speed data were measured with a Datron LS1 optical speed sensor. The product literature specifies an accuracy of +/- 0.2% and a reproducibility of +/- 0.1% over the measurement range of 0.5 to 400 kph. Figure 4-10 from NRMRL’s report

correlates the speed measurement to a drive shaft speed sensor that was scaled using a National Institute of Standards and Technology (NIST)-traceable frequency source. The outliers at the low-speed indicated when the truck was turning (the tractor and the trailer-mounted speed sensor traveled less distance than the tractor does during turns). Notwithstanding these points, the correlation is a good indication of speed measurement precision.

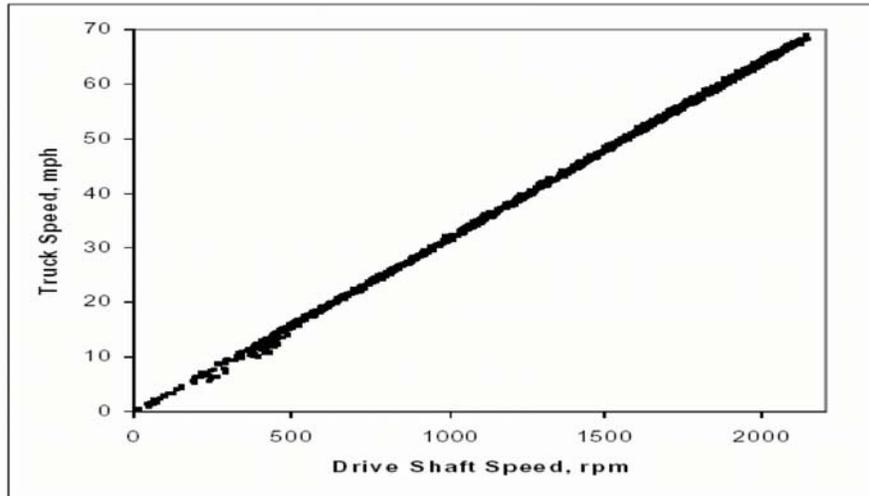


Figure 4-10 Vehicle Speed Correlation (U.S. EPA 2001c)

At the same time, NRMRL provided Figure 4-11 (U.S. EPA 2001c) to show the precision for four ranges of vehicle speed, along with similar estimates of accuracy. This figure will help researchers deal with speed measurement noise in the future.

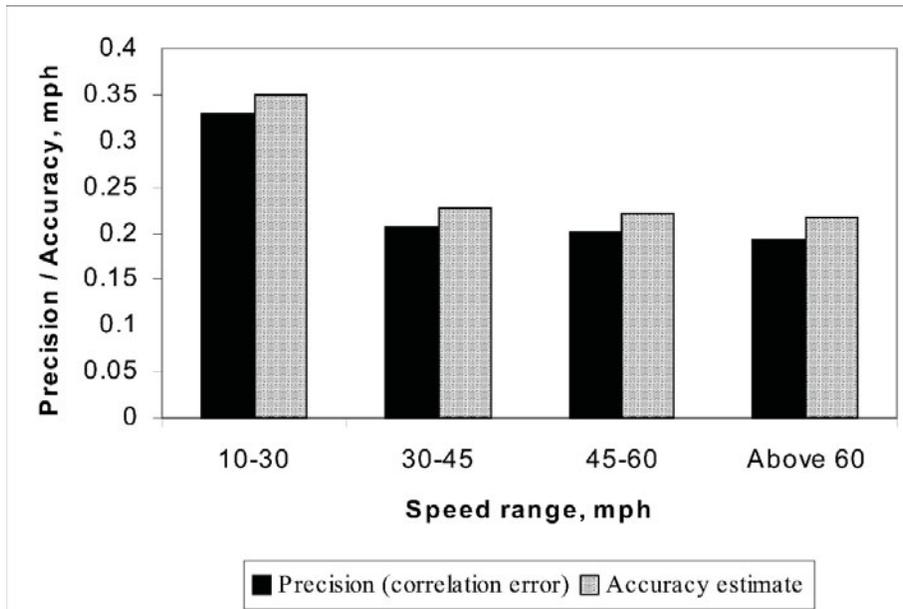


Figure 4-11 Vehicle Speed Error for Different Speed Ranges (U.S. EPA 2001c)

#### 4.2.4 Database Formation

The data dictionaries of the source files were reviewed for parameter content (Table 4-8). Not all variables reported are included in explanatory analysis. A standard file structure was designed to accommodate the available format. Emissions data with units of gram/second were selected to develop the proposed emission model. All variables used to calculate mass emissions were excluded in further analysis. Similarly, because the “measured horsepower” field is calculated from measurements of drive shaft torque and rotational speed, only “measured horsepower” is used to represent power related variables. At the same time, variables like acceleration that might be helpful in explaining variability in vehicle emissions were included in the proposed file structure although they were not provided in the original dataset. Acceleration data were derived from speed data using the central difference method.

Table 4-8 List of Parameters Used in Explanatory Analysis for HDDV

Category	Parameters
Test Information	Date; Time
Vehicle Characteristics	Vehicle make/model; Model year; Engine type; Engine rating; Vehicle maintenance history
Onroad Load Parameters	Truck load weight (lb); Vehicle speed (mph); Acceleration (mph/s); Measured engine power (bhp)
Engine Operating Parameters	Engine intake air temperature (deg F); Engine exhaust air temperature (deg F); Engine coolant temperature (deg F); Engine oil temperature (deg F)
Environment Conditions	Barometric pressure (Hg), Ambient moisture (%)
Vehicle Emissions	CO, NO <sub>x</sub> , and HC emission (in g/s units)

#### 4.2.5 Data Summary

After the post-processing procedure was completed, a summary of the emissions and activity data as well as environmental and roadway characteristics is given in Table 4-9.

Table 4-9 Summary of Heavy-Duty Vehicle Data U.S. EPA 2001c).

Test ID	Number of Seconds of Data	Vehicle Operation		Emission Data			Environment Characteristics	
		Average Speed (mph)	Average Engine Power (bhp)	Average CO (g/s)	Average NO <sub>x</sub> (g/s)	Average HC (g/s)	Barometric Pressure (Hg)	Ambient Moisture (%)
3F00V	4430	43.55	163.10	0.11633	0.27983	0.001442	28.273	1.6874
3F00C	7991	36.49	323.79	0.08200	0.19566	0.001166	28.272	1.6874
3F00A	1904	43.55	475.12	0.17476	0.34262	0.001471	28.272	1.6874
3H00V	3718	43.66	130.99	0.08386	0.22701	0.001429	28.273	1.6874
3H00C	7593	39.43	112.50	0.07456	0.17866	0.001414	28.272	1.6874
3H00A	1959	48.04	218.50	0.20521	0.32078	0.001751	30.423	1.3573
3E00V	3863	41.41	123.42	0.10896	0.21157	NA	28.273	1.6874
3E00C	7962	39.31	104.95	0.07489	0.14908	NA	28.272	1.6874
3E00A	1810	50.15	197.07	0.22324	0.26108	NA	30.137	1.9020
3F0GA	577	35.93	302.14	0.23114	0.41269	NA	29.995	0.4685
3F0SA	792	36.26	287.45	0.25140	0.37947	NA	29.995	0.4685
3F0V	3635	41.65	152.23	0.14879	0.28413	NA	29.995	0.4685
3H0GA	594	33.81	253.63	0.30036	0.48494	0.002159	29.690	1.6059
3H0SA	707	34.27	223.73	NA	0.32498	NA	29.690	1.6059
3H0V	3331	41.53	143.38	0.08892	0.27712	0.002436	28.020	0.4742
3E0GA	421	32.91	233.93	0.37978	0.30728	0.000589	29.976	0.5812
3E0SA	571	31.99	180.73	0.23652	0.33325	0.003042	29.976	0.5812
3E0V	3395	42.64	103.63	0.08879	0.25745	0.002805	29.976	0.5812
3F3&6	8629	36.59	131.00	0.14409	0.31374	0.001426	28.282	1.2520
3H3&6	10573	43.13	107.06	0.16769	0.27507	0.001753	28.273	1.6874
3E3&6	9825	44.74	121.69	0.16617	0.23913	0.001839	28.250	1.5716
3FIL4	12456	66.54	152.91	0.06994	0.29925	NA	29.238	0.3886
3FIL5	13738	58.76	129.99	0.06354	0.22315	NA	29.238	0.3886
3FIL6	6415	66.94	130.11	0.06273	0.20833	0.001409	29.238	0.3886
3FIL7	10678	62.76	164.82	0.07042	0.28353	NA	29.854	0.1480
3FIL8	12248	64.70	147.26	0.06688	0.26035	NA	29.773	0.1484
3FIL9	11956	65.62	153.44	0.06551	0.20905	NA	29.418	0.1502
3FIL10	12367	63.71	167.73	0.07481	0.35788	NA	30.132	0.1466
5F0V	4895	32.87	96.09	0.10716	0.23558	0.002828	30.101	0.5761
5H0V	4091	42.36	126.14	0.12564	0.30933	NA	30.179	0.6091
5E0V	4407	42.60	105.84	0.10681	0.29045	0.002894	30.278	0.8601
5F3&6a	6971	36.24	147.99	0.13716	0.31607	0.003111	28.004	0.9070
5F3&6b	5058	38.69	133.54	0.14044	0.30661	0.001924	28.009	0.8862
5H3&6a	6919	39.74	133.01	0.12723	0.28763	0.002397	28.024	0.8138
5H3&6b	6951	39.44	148.26	0.15400	0.32910	0.002807	28.014	1.2149
5E3&6	10807	46.01	124.07	0.13981	0.27674	0.002827	28.024	1.0131

## CHAPTER 5

### 5. METHODOLOGICAL APPROACH

The following chapter lays the theoretical foundation of the conceptual framework of model development. This chapter outlines the statistical methods, addresses issues that arise in statistical modeling, and presents the solutions that are employed to address these issues. This chapter will serve as a guide or “road map” for the underlying methodology of the model development process.

#### 5.1 Modeling Goal and Objectives

The goal of this research is to provide emission rate models that fill the gap between the existing models and ideal models for predicting emissions of  $\text{NO}_x$ , CO, and HC from heavy-duty diesel vehicles. Problems in existing models, like EPA’s MOBILE series and CARB’s EMFAC series of models, have been highlighted in previous chapters. U.S. EPA is currently developing a new set of modeling tools for the estimation of emissions produced by on-road and off-road mobile sources. MOVES, a new model under development by EPA’s OTAQ, is a modeling system designed to better predict emissions from on-road operations. The philosophy behind MOVES is the development of a model that is as directly data-driven as possible, meaning that emission rates are developed from second-by-second or binned data.

Using second-by-second data collected from on-road vehicles, this research effort will develop models that predict emissions as a function of on-road variables known to affect vehicle emissions. The model should be robust and ensure that assumptions about the underlying distribution of the data are verified and the properties of parameter estimates are not violated. With limited available data, this study focuses on development of an analytical methodology that is repeatable with a different data set from across space and across time. As more data become available, the proposed model will need to be re-estimated to ensure that the model is transferable across additional HDV engine types, operating conditions, environmental conditions, and even perhaps geographical regions.

## 5.2 Statistical Method

The purpose of statistical modeling was to determine which explanatory variables significantly influence vehicle emissions so that the data can be stratified by those variables and a corresponding regression relationship can be developed. For many statistical problems there are several possible solutions. In comparing the means of two small groups, for instance, we could use a  $t$  test, a  $t$  test with a transformation, a Mann-Whitney U test, or one of several others. The choice of method depends on the plausibility of normal assumptions, the importance of obtaining a confidence interval, the ease of calculation, etc.

Parametric or non-parametric approaches to evaluation can be applied. Parametric methods are used when the distribution is either known with certainty or can be guessed with a certain degree of certainty. These methods are meaningful only for continuous data which are sampled from a population with an underlying normal distribution or whose distribution can be rendered normal by mathematical transformation. Analysts must be careful to ensure that significant errors are not introduced when assumptions are not met. In contrast, nonparametric methods make no assumptions about the distribution of the data or about the functional form of the regression equation. Nonparametric methods are especially useful in situations where the assumptions required by parametric are in question. Brief overviews and underlying theories of statistical methods that might be used in this research are addressed in the following sections.

### 5.2.1 Parametric Methods

#### 5.2.1.1 The $t$ -Test

Student's  $t$ -test is one of the most commonly used techniques for testing whether the means of two groups are statistically different from each other. This test tries to determine whether the measured difference between two groups is large enough to reject the null hypothesis or whether such differences are just due to "chance". The formula for the  $t$ -test (Equation 5-1) is a ratio. The numerator of the ratio is just the difference between the two means or averages. The denominator is a measure of the variability or dispersion of the data.

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{(Equation 5-1)}$$

where  $\bar{x}_1$  and  $\bar{x}_2$  are the sample means,  $s_1^2$  and  $s_2^2$  are the sample variances,  $n_1$  and  $n_2$  are the sample sizes and  $t$  is a Student  $t$  quantile with  $n_1 + n_2 - 2$  degrees of freedom.

Usually a significance level of 0.05 (or equivalently, 5%) is employed in statistical analyses. The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis  $H_0$ , if it is in fact true. Another index is p-value which is the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis  $H_0$  is true. The p-value is compared with the actual significance level of the test and, if it is smaller, the result is significant. That is, if the null hypothesis were to be rejected at the 5% significance level, this would be reported as “ $p < 0.05$ ”.

The assumptions for  $t$ -test include: 1) the populations are normally distributed; 2) variances in the two populations are equal; and 3) the populations are independent. The results of the analysis may be incorrect or misleading when assumptions are violated. For example, if the assumption of independence for the sample values is violated, then the two-sample  $t$  test is simply not appropriate. If the assumption of normality is violated or outliers are present, the two-sample  $t$  test may not be the most powerful available test. This could mean the difference between detecting a true difference or not. A nonparametric test or employing a transformation may result in a more powerful test.

#### 5.2.1.2 Ordinary Least Squares Regression

Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that one variable can be predicted from the other, or others (Neter et al. 1996). There are many different kinds of regression models, like the linear regression model, exponential regression model, logistic regression model, and so on. Among them, linear regression is a commonly used and easily understood statistical method. Linear regression explores relationships that can be described by straight lines or their generalization to many dimensions. Regression allows a single response variable to be described by one or more predictor variables.

Ordinary least squares (OLS) regression is a common statistical technique for quantifying the relationship between a continuous dependent variable and one or more independent variables (Neter et al. 1996). The dependent variables may be either continuous or discrete. Neter et al. (1996) provides the basic OLS regression equation for a single variable regression model as shown in Equation 5-2:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_i X_i + \varepsilon_i \quad (\text{Equation 5-2})$$

where:

- $\hat{Y}_i$  = value of the response variable in the  $i^{\text{th}}$  trial
- $\hat{\beta}_0, \hat{\beta}_i$  = estimators of regression parameters
- $X_i$  = value of the predictor variable in the  $i^{\text{th}}$  trial
- $\varepsilon_i$  = random error term with mean  $E\{\varepsilon_i\} = 0$  and variance  $\sigma^2$   $\{\varepsilon_i\} = \sigma^2$ ;  
 $\varepsilon_i$  and  $\varepsilon_j$  are uncorrelated so that their covariance is zero.

The parameters of the OLS regression equation,  $\beta_0$  and  $\beta_i$ , are found by the least squares method, which requires that the sum of squares of errors be minimized. Gauss-Markov theorem (Neter et al. 1996) states that, among all unbiased estimators that are linear combinations of  $y$ s, the OLS estimators of regression coefficients have the smallest variance; i.e., they are the best linear unbiased estimators. The Gauss-Markov Theorem does not tell one to use least squares all the time, but it strongly suggests use of least squares (Neter et al. 1996).

In linear regression, there are key assumptions that must be met, including:

- $Y_i$  are independent normal random variables;
- The expected value of the error terms  $\varepsilon_i$  is zero;
- The error terms  $\varepsilon_i$  are assumed to have constant variance  $\sigma^2$ ;
- The error terms  $\varepsilon_i$  are assumed normally distributed;
- The error terms  $\varepsilon_i$  are assumed to be uncorrelated so that their covariance is zero; and
- The error terms  $\varepsilon_i$  are independent of the explanatory variable

If the above assumptions are violated the regression equation may yield biased results (Neter et al. 1996). For example, if the explanatory variable is not independent of the error term, larger sample sizes do not lead to lower standard errors for the parameters, and the parameter estimates (slope, etc.) are biased. If the error is not distributed normally, for example, there may

be fat tails. Consequently, use of the normal distribution may underestimate true 95% confidence intervals.

### 5.2.1.3 Robust Regression

OLS models generally rely on the normality assumption and are often fitted by means of the least squares estimators. However, the sensitivity of these estimation techniques is related to this underlying assumption which has been identified as a weakness that can lead to erroneous interpretations (Copt and Heritier 2006). Robust regression procedures dampen the influence of outlying cases, as compared to OLS estimation, in an effort to provide a better fit for the majority of cases. Robust regression procedures are useful when a known, smooth regression function is to be fitted to data that are “noisy”, with a number of outlying cases, so that the assumption of a normal distribution for the error terms is not appropriate (Neter et al. 1996). The method of moments (MM) estimators are designed to be both highly robust against outliers and highly efficient.

## **5.2.2 Nonparametric Methods**

Nonparametric methods have several advantages compared with parametric methods. Nonparametric methods require no or very limited assumptions to be made about the format of the data, and they may therefore be preferable when the assumptions required for parametric methods are not valid (Whitley and Ball 2002). Nonparametric methods can be useful for dealing with unexpected, outlying observations that might be problematic with a parametric approach. Nonparametric methods are intuitive and are simple to carry out by hand, for small samples at least.

However, nonparametric methods may lack power as compared with more traditional approaches (Siegel 1988). This lack of power is a particular concern if the sample size is small or if the assumptions for the corresponding parametric method hold true (e.g., normality of the data). Nonparametric methods are geared toward hypothesis testing rather than estimation of effects. It is often possible to obtain nonparametric estimates and associated confidence intervals, but this process is not generally straightforward. In addition, appropriate computer software for nonparametric methods can be limited, although the situation is improving.

### 5.2.2.1 Chi-Square Test

The Chi-square (Koehler and Larnz 1980), best known goodness-of-fit test, assumes that the observations are independent and that the sample size is reasonably large. This method can

be used to test whether a sample fits a known distribution, or whether two unknown distributions from different samples are the same. The test can detect major departures from a logistic response function, but is not sensitive to small departures from a logistic response function. The test assumptions are that the sample is random and that the measurement scale is at least ordinal (Conover 1980; Neter et al. 1996).

Pearson's chi-square goodness of fit test statistic is shown in Equation 5-3 (StatsDirect 2005):

$$T = \sum_{j=1}^i \frac{(O_j - E_j)^2}{E_j} \quad (\text{Equation 5-3})$$

where  $O_j$  are observed counts,  $E_j$  are corresponding expected count and  $c$  is the number of classes for which counts/frequencies are being analyzed.

The test statistic is distributed approximately as a chi-square random variable with  $c-1$  degrees of freedom. The test has relatively low power (chance of detecting a real effect) with all but large numbers or big deviations from the null hypothesis (all classes contain observations that could have been in those classes by chance).

The handling of small expected frequencies is controversial. Koehler and Larnz asserted that the chi-square approximation is adequate provided all of the following are true: total of observed counts ( $N \geq 10$ ; number of classes ( $c \geq 3$ ; all expected values  $\geq 0.25$  (Koehler and Larnz 1980).

#### 5.2.2.2 Kolmogorv-Smirnov Two-Sample Test

The Kolmogorov-Smirnov (K/S) two-sample test (Chakravart and Roy 1967) compares the empirical distribution functions of two samples,  $E_1$  and  $E_2$ . The Kolmogorov-Smirnov test is a nonparametric test, which can be used to test whether two or more samples are governed by the same distribution by comparing their empirical distribution functions.

The Kolmogorov-Smirnov two sample test statistic can be defined as shown in Equation 5-4 (Chakravart and Roy 1967):

$$D = \left| E_1(i) - E_2(i) \right| \quad (\text{Equation 5-4})$$

where  $E_1$  and  $E_2$  are the empirical distribution functions for the two samples.

The Kolmogorov-Smirnov (K/S) two-sample test provides an improved methodology over the chi-squared test since data do not have to be assigned arbitrarily to bins. Further, it is a non-parametric test so a distribution does not have to be assumed. However, the main disadvantage to the K/S is similar to the chi-square in that the orders of magnitude of separate tests that would have to be conducted to test all the possible combinations of variables in the datasets is logistically infeasible (Hallmark 1999).

### 5.2.2.3 Wilcoxon Mann-Whitney Test

The Wilcoxon Mann-Whitney Test (Easton and McColl 2005) is one of the most powerful of the nonparametric tests for comparing two populations. This test is used to test the null hypothesis that two populations have identical distribution functions against the alternative hypothesis that the two distribution functions differ only with respect to location (median), if at all.

The Wilcoxon Mann-Whitney test does not require the assumption that the differences between the two samples are normally distributed. In many applications, the Wilcoxon Mann-Whitney Test is used in place of the two sample *t*-test when the normality assumption is questionable. This test can also be applied when the observations in a sample of data are ranks, that is, ordinal data rather than direct measurements.

The Mann Whitney U statistic is defined as shown in Equation 5-5 (StatsDirect 2005):

$$U = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - \sum_{i=n_1+1}^{n_2} R_i \quad (\text{Equation 5-5})$$

where samples of size  $n_1$  and  $n_2$  are pooled and  $R_i$  are the ranks.

U can be resolved as the number of times observations in one sample precede observations in the other sample in the ranking. Wilcoxon rank sum, Kendall's S and the Mann-Whitney U test are exactly equivalent tests. In the presence of ties the Mann-Whitney test is also equivalent to a chi-square test for trend.

### 5.2.2.4 Analysis of Variance (ANOVA)

ANOVA (Analysis of Variance) (Neter et al. 1996), sometimes called an F test, is closely related to the *t* test. The major difference is that, where the *t* test measures the difference between the means of two groups, an ANOVA tests the difference between the means of two or more groups. ANOVA modeling does not require any assumptions about the nature of the statistical relation between the response and explanatory variables, nor do they require that the explanatory variables be quantitative.

The ANOVA, or single factor ANOVA, compares several groups of observations, all of which are independent, but each group of observations may have a different mean. A test of great importance is whether or not all the means are equal. The advantage of using ANOVA rather than multiple *t*-tests is that it reduces the probability of a type-I error (making multiple comparisons increases the likelihood of finding something by chance). One potential drawback to an ANOVA is that it can only tell that there is a significant difference between groups, not which groups are significantly different from each other. The breakdowns of the total sum of squares and degrees of freedom, together with the resulting mean squares, are presented in an ANOVA table such as Table 5-1.

Table 5-1 ANOVA Table for Single-Factor Study (Neter et al. 1996)

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	Expected Mean Square E(MS)
Between treatments	$SSTR = \sum n_i (\bar{Y}_i - \bar{Y}_{..})^2$	$r - 1$	$MSTR = \frac{SSTR}{r - 1}$	$\sigma^2 + \frac{\sum n_i (\mu_i - \mu)^2}{r - 1}$
Error (within treatments)	$SSE = \sum \sum (y_{ij} - \bar{Y}_i)^2$	$n_T - r$	$MSE = \frac{SSE}{N_T - r}$	$\sigma^2$
Total	$SSTO = \sum \sum (y_{ij} - \bar{Y}_{..})^2$	$n_T - 1$		

A factorial ANOVA can examine data that are classified on multiple independent variables. A factorial ANOVA can show whether there are significant main effects of the independent variables and whether there are significant interaction effects between independent variables in a set of data. Interaction effects occur when the impact of one independent variable depends on the level of the second independent variable (Neter et al. 1996). Computation can be performed with standard statistical software such as SAS<sup>®</sup>.

#### 5.2.2.5 HTBR

HTBR (Breiman et al. 1984) is a forward step-wise variable selection method, similar to forward stepwise regression. This method is also known as Classification and Regression Tree (CART) analysis. This technique generates a “tree” structure by dividing the sample data

recursively into a number of groups. The groups are selected to maximize some measure of difference in the response variable in the resulting groups. As Washington et al. summarized in 1997 (Washington et al. 1997a), this method is based upon iteratively asking and answering the following questions: (1) which variable of all of the variables ‘offered’ in the model should be selected to produce the maximum reduction in variability of the response? and (2) which value of the selected variable (discrete or continuous) results in the maximum reduction in variability of the response? The HTBR terminology is similar to that of a tree; there are branches, branch splits or internal nodes, and leaves or terminal nodes (Washington et al. 1997a).

To explain the method in mathematical terms, the definitions are presented by Washington et al. (Washington et al. 1997a). The first step is to define the deviance at a node. A node represents a data set containing L observations. The deviance,  $D_a$ , can be estimated as shown in equation 5-6:

$$D_a = \sum_{l=1}^L (y_{l,a} - \bar{x}_a)^2 \quad (\text{Equation 5-6})$$

where

- $D_a$  = total deviance at node a, or the sum of squared error (SSE) at the node
- $y_{l,a}$  =  $l^{\text{th}}$  observation of dependent variable y at node a
- $\bar{x}_a$  = estimated mean of L observations in node a

Next, the algorithm seeks to split the observation at node a on a value of an independent variable,  $X_1$ , into two branches and corresponding nodes  $b$  and  $c$ , each containing M and N of the original L observations ( $M+N=L$ ) of the variable  $X_1$ . The deviance reduction function evaluated over all possible  $X$ s then can be defined as shown in Equations 5-7 thru 5-9:

$$\Delta_{(allX)} = D_a - D_b - D_c \quad (\text{Equation 5-7})$$

$$D_b = \sum_{m=1}^M (y_{m,b} - \bar{x}_b)^2 \quad (\text{Equation 5-8})$$

$$D_c = \sum_{n=1}^N (y_{n,c} - \bar{x}_c)^2 \quad (\text{Equation 5-9})$$

where

$$\begin{aligned}
 \Delta_{(allX)} &= \text{the total deviance reduction function evaluated over the domain of all Xs} \\
 D_b &= \text{total deviance at node b} \\
 D_c &= \text{total deviance at node c} \\
 y_{m,b} &= m^{\text{th}} \text{ observation on dependent variable y in node b} \\
 y_{n,c} &= n^{\text{th}} \text{ observation on dependent variable y in node c} \\
 \bar{x}_b &= \text{estimated mean of M observations in node b} \\
 \bar{x}_c &= \text{estimated mean of N observations in node c}
 \end{aligned}$$

The variable  $X_k$  and its optimum split  $X_{k(i)}$  is sought so that the reduction in deviance is maximized, or more formally when (as shown in equation 5-10):

$$\Delta_{(allX)} = \sum_{l=1}^L (y_{l,a} - \bar{x}_a)^2 - \sum_{m=1}^M (y_{m,b} - \bar{x}_b)^2 - \sum_{n=1}^N (y_{n,c} - \bar{x}_c)^2 = \max \quad (\text{Equation 5-10})$$

where

$$\begin{aligned}
 \Delta_{(allX)} &= \text{the total deviance reduction function evaluated over the domain of all Xs} \\
 y_{l,a} &= l^{\text{th}} \text{ observation of dependent variable y at node a} \\
 \bar{x}_a &= \text{estimated mean of L observations in node a} \\
 y_{m,b} &= m^{\text{th}} \text{ observation on dependent variable y in node b} \\
 y_{n,c} &= n^{\text{th}} \text{ observation on dependent variable y in node c} \\
 \bar{x}_b &= \text{estimated mean of M observations in node b} \\
 \bar{x}_c &= \text{estimated mean of N observations in node c}
 \end{aligned}$$

The maximum reduction occurs at a specific value  $X_{k(i)}$ , of the independent variable  $X_k$ . When the data are split at this point, the remaining samples have a much smaller variance than the original data set. Thus, the reduction in node  $a$  deviance is greatest when the deviances at nodes  $b$  and  $c$  are smallest. Numerical search procedures are employed to maximize Equation 5-10 by varying the selection of variables used as a basis for a split and the value to use for each variable at a split.

In growing a regression tree, the binary partitioning algorithm recursively splits the data in each node until the node is homogenous or the node contains too few observations. If left

unconstrained, a regression tree model can “grow” until it results in a complex model with a single observation at each terminal node that explains all the deviance. However, for application purposes, it is desirable to create criteria to balance the model’s ability to explain the maximum amount of deviation with a simpler model that is easy to interpret and apply. Some software, such as S-Plus™, allows the user to select such criteria. The software allows the user to interact with the data in the following manner to select variables and help simplify the final model:

- Response variable: the response variable is selected by the user from a list of fields from the data set;
- Predictor variables: one or more independent variables can be selected by the user from a list of fields associated with the dataset;
- Minimum number of observations allowed at a single split: sets the minimum number of observations that must be present before a split is allowed (default is 5);
- Minimum node size: sets the allowed sample size at each node (default is 10); and
- Minimum node deviance: the deviance allowed at each node (default is 0.01).

However, unlike OLS regression models, a shortcoming of HTBR is the absence of formal measures of model fit, such as *t*-statistics, F-ratio, and r-square, to name a few. Thus, the HTBR model is used to guide the development of an OLS regression model, rather than as a model in its own right. Similar uses of HTBR techniques have been developed and applied in previous research papers (Washington et al. 1997a; Washington et al. 1997b; Fomunung et al. 1999; Frey et al. 2002).

### **5.3 Modeling Approach**

The model development process will start by using HTBR both as a data reduction tool and for identifying potential interactions among the variables. Then OLS Regression or Robust Regression is used with the identified variables to estimate a preliminary “final” model. After that, we need to check the model for compliance with normality assumptions and goodness of fit.

Several diagnostic tools are available to perform these checks. Once a preliminary “final” model is obtained, regression coefficients are examined using their *t*-statistics and correlation coefficients to determine which variables should be removed or retained in the model for further analysis. However this procedure can lead to the removal of potentially important inter-correlated explanatory variables. In fact, variable agreement with underlying scientific principles

of combustion, pollutant formation and emission controls (cause-effect relationships) should be the basis for the ultimate decisions regarding variable selection. Thus, a  $t$ -statistic may indicate that a parameter is insignificant (at level of significance = 0.05), while theory indicates that such a parameter should be retained in the model for further analysis. This type of error is usually referred to as a type II error (Fomunung 2000).

F-statistics and adjusted coefficient of multiple determination,  $R^2$  are used to determine the effect-size of the parameters. Usually, adding more explanatory variables to the regression model can only increase  $R^2$  and never reduce it, because SSE can never become larger with more X variables and total sum of squares (SSTO) is always the same for a given set of responses. The adjusted coefficient of multiple determination can adjust  $R^2$  by dividing each sum of squares by its associated degrees of freedom. The F-test is used to test whether the parameter can be dropped even if the  $t$ -statistic is appropriate.

In multiple regression analysis, the predictor or explanatory variables tend to be correlated among themselves and with other variables related to the response variable but not included in the model. The effects of multicollinearity are many and can be severe. Neter et al. (Neter et al. 1996) have documented a few of these: when multicollinearity exists the interpretation of partial slope coefficients becomes meaningless; multicollinearity can lead to estimated regression coefficients that vary widely from one sample to another; and there may be several regression functions that provide equally good fits to the data, making the effects of individual predictor variables difficult to assess.

There are some informal diagnostic tools suggested to detect this problem. A frequently used technique is to calculate a simple correlation coefficient between the predictor variables to detect the presence of inter-correlation among independent variables. Large values of correlation is an indication that multicollinearity may exist. Large changes in the estimated regression coefficients when a predictor variable is added or deleted are also an indication. Finally, multicollinearity may be a problem if estimated regression coefficients are calculated with an algebraic sign that is the opposite of that expected from theoretical considerations or prior experience (i.e., the beta coefficient is compensating for the beta coefficient of a correlated explanatory variable).

A formal method of detecting this problem is the variance inflation factor (VIF), which is a measure of how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related (Neter et al. 1996). This method is widely used because it can provide quantitative measurements of the impact of multicollinearity. The largest VIF value among all Xs is used to assess the severity of multicollinearity. As a

rule of thumb, a VIF in excess of 10 is frequently used as an indication that multicollinearity is severe.

Diagnostic plots are examined to verify normality and homoscedasticity (i.e., homogeneity of variance) assumptions as well as the goodness of fit. Because of the difficulty in assessing normality, it is usually recommended that non-constancy of error variance should be investigated first (Neter et al. 1996). The plots used to identify any patterns in the residuals are considered as informal diagnostic tools and include plots of the residuals versus the fitted values and plot of square root of absolute residuals versus the fitted values. The normality of the residuals can be studied from histograms, box plots, and normal probability plots of the residuals. In addition, comparisons can be made of observed frequencies with expected frequencies if normality exists. Usually, heteroscedasticity and/or inappropriate regression functions may induce a departure from normality. When OLS is applied to heteroskedastic models the estimated variance is a biased estimator of the true variance. OLS either overestimates or underestimates the true variance, and, in general it is not possible to determine the nature of the bias. The variances, and the standard errors, may therefore be either understated or overstated.

#### **5.4 Model Validation**

Model validity refers to the stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function, and the ability to generalize inferences drawn from the regression function. Validation is a useful and necessary part of the model-building process (Neter et al. 1996).

Two basic ways of validating a regression model are internal and external. Internal validation consists of model checking for plausibility of signs and magnitudes of estimated coefficients, agreement with earlier empirical results and theory, and model diagnostic checks such as distribution of error terms, normality of error terms, etc. Internal validation will be performed as part of the model estimation procedure.

External validation is the process to check the model and its predictive ability with the collection of new data, such as data from another location or time, or using a holdout sample. Considering there are only 15 buses/engines in the data set, it is not practical to split the data set and hold a sample for validation purposes. Splitting the data set will definitely influence the regression estimators. However suggestions and procedures for external validation will be provided.

## CHAPTER 6

### 6. DATA SET SELECTION AND ANALYSIS OF EXPLANATORY VARIABLES

#### 6.1 Data Set Used for Model Development

Development of a modal model designed to predict emissions on a second-by-second basis as a function of engine load requires the availability of appropriate emission test data. Modal modeling required the availability of second-by-second vehicle emissions data, collected in parallel with corresponding revealed engine load data. In 2004, only two data sets could be identified for use in this modeling effort. U.S. EPA provided two major HDV activity and emission databases to develop the emission rate model (Ensfield 2002) (U.S. EPA 2001b). One database is a transit bus database, which included emissions data collected on diesel transit buses operated by the AATA in 2001, and another database is heavy HDV (HDV8B) database prepared by NRMRL in 2001. The transit database consisted of data collected from 15 buses with the same type of engines while the HDV8B database consisted of only one truck engine tested extensively on-road under pre-rebuild and post rebuild engine conditions. To decide whether it is suitable to combine these two data sets or treat them individually, two dummy variables were added to the databases to describe vehicle types. For the first dummy variable named “bus”, 1 was assigned for transit bus, and 0 for others. For the second dummy variable, 1 was assigned for HDDV with pre-rebuild engine, and 0 for others. HTBR was applied to all data sets to examine whether transit buses behave differently from HDDVs or not. The regression trees and results for  $\text{NO}_x$ , CO, and HC emission rates are given in Figures 6-1 to Figure 6-3.

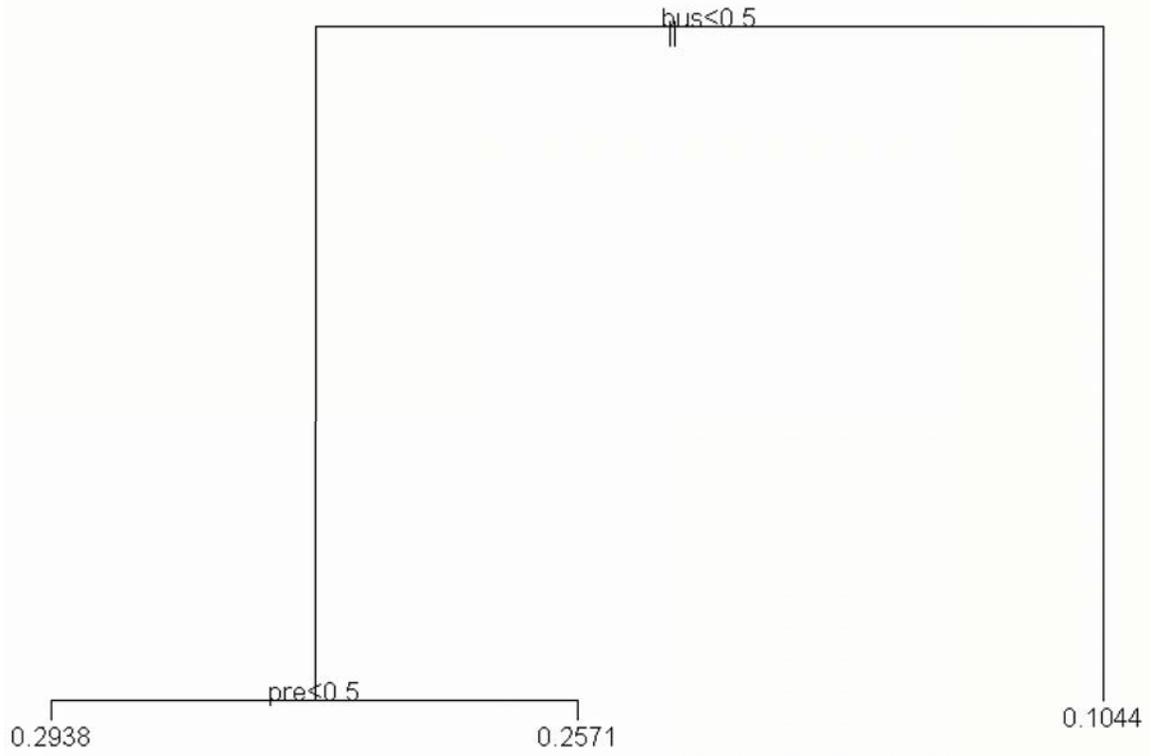


Figure 6-1 HTBR Regression Tree Result for NO<sub>x</sub> Emission Rate for All Data Sets

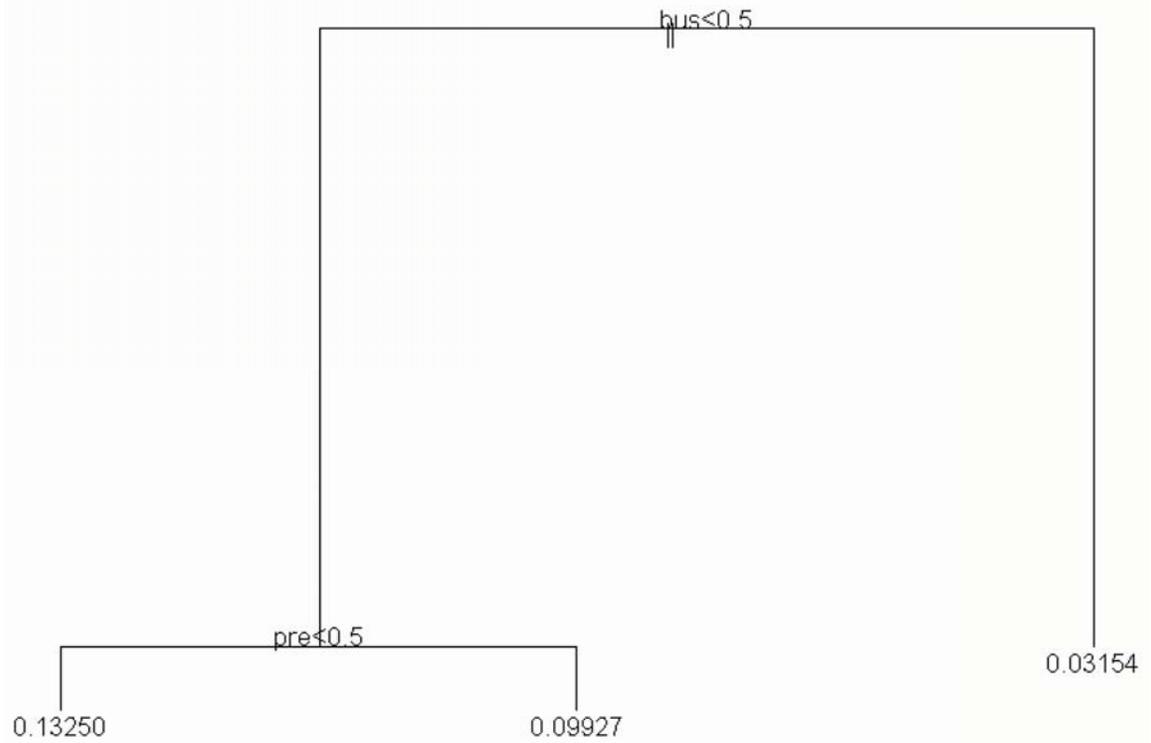


Figure 6-2 HTBR Regression Tree Result for CO Emission Rate for All Data Sets

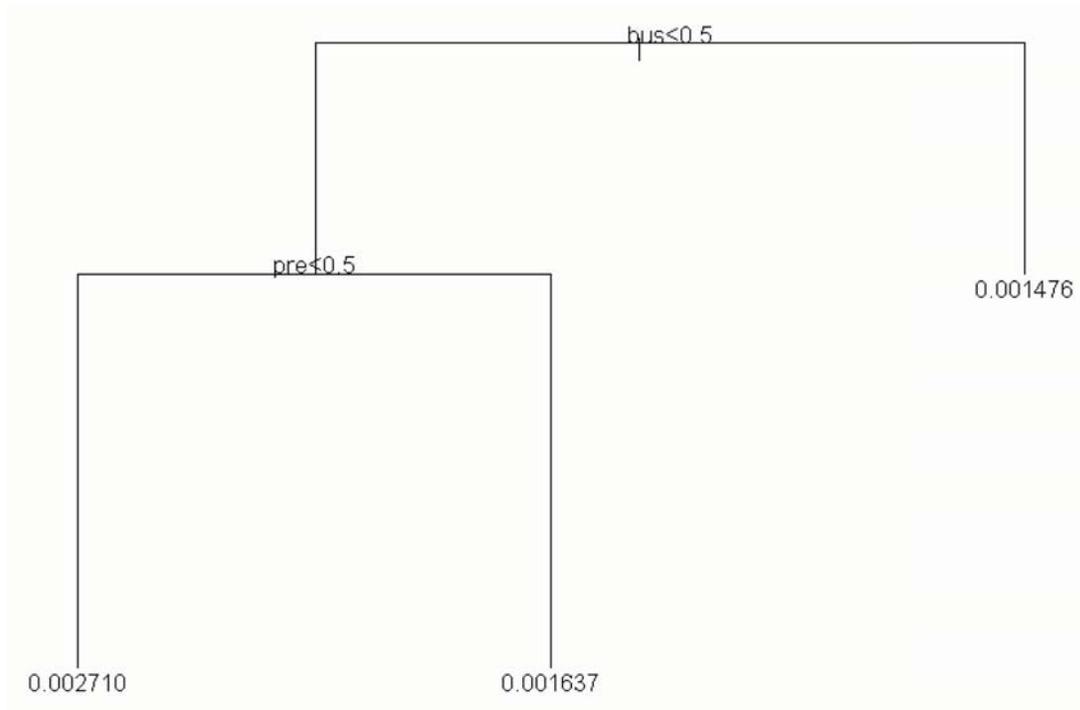


Figure 6-3 HTBR Regression Tree Result for HC Emission Rate for All Data Sets

Dummy variable for bus is selected as the first split for all three trees above. Therefore transit bus and HDDV should be treated separately. Since there are 15 engines in the transit bus data set and one engine (pre-rebuild and post-rebuild for the same engine) in the HDDV data set, the transit bus data set should be used for the final version of the conceptual model development.

## 6.2 Representative Ability of the Transit Bus Data Set

The transit bus data set was collected by Sensors, Inc. in Oct. 2001 (Ensfield 2002). The buses tested came from the AATA and included 15 New Flyer models with Detroit Diesel Series 50 engines. All of the buses were of model years 1995 and 1996. All of the bus tested periods lasted approximately 2 hours. The buses operated during standard AATA bus routes and stopped at all regular stops although the buses did not board or discharge any passengers (Ensfield 2002). The routes were mostly different for each test, and were selected for a wide variety of driving conditions (see Figure 4-1).

Figure 6-4 shows the speed-acceleration matrix developed with second-by-second data. There are two high speed/acceleration frequency peaks here. One is the bin of speed  $\leq 2.5$  mph and acceleration  $(-0.25 \text{ mph/s}, 0.25 \text{ mph/s})$  and contains 26.11% of the observations, while the other is the combination of several adjacent bins which covers speed  $(22.5 \text{ mph}, 47.5 \text{ mph})$  and acceleration  $(-0.75 \text{ mph/s}, 0.75 \text{ mph/s})$ .

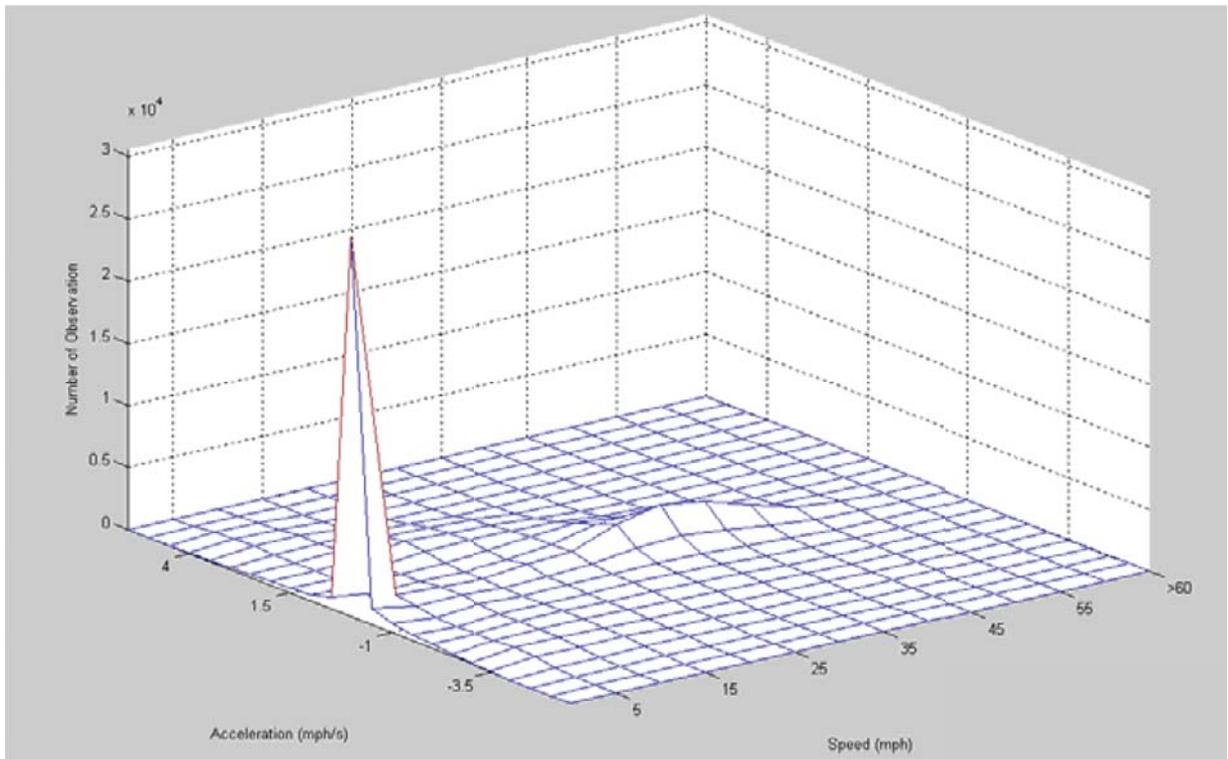


Figure 6-4 Transit Bus Speed-Acceleration Matrix

Georgia Institute of Technology researchers collected more than 6.5 million seconds of transit bus speed and position data using Georgia Tech Trip Data Collectors ( an onboard computer with GPS receiver, data storage, and wireless communication device) installed on two Metropolitan Atlanta Rapid Transit Authority (MARTA) buses in 2004 (Yoon et al. 2005b). With second-by-second data, the research team developed transit bus speed/acceleration matrices for the combinations between roadway facility type (arterial or local road) and time range (morning, midday, afternoon, night). For each matrix, two high acceleration/deceleration frequency peaks were also found. This finding is consistent with the AATA data set, indicating at least that the on-road operations of the buses in Ann Arbor are similar to operations in the Atlanta region.

This data set was collected under a wide variety of environmental conditions, too. The temperature ranged from 10 °C to 30 °C, the relative humidity ranged from 15% to 65%, while the barometric pressure ranged from 960 mbar to 1000 mbar (Figure 6-5). So we can use this data set to examine the impact of environmental conditions on emissions.

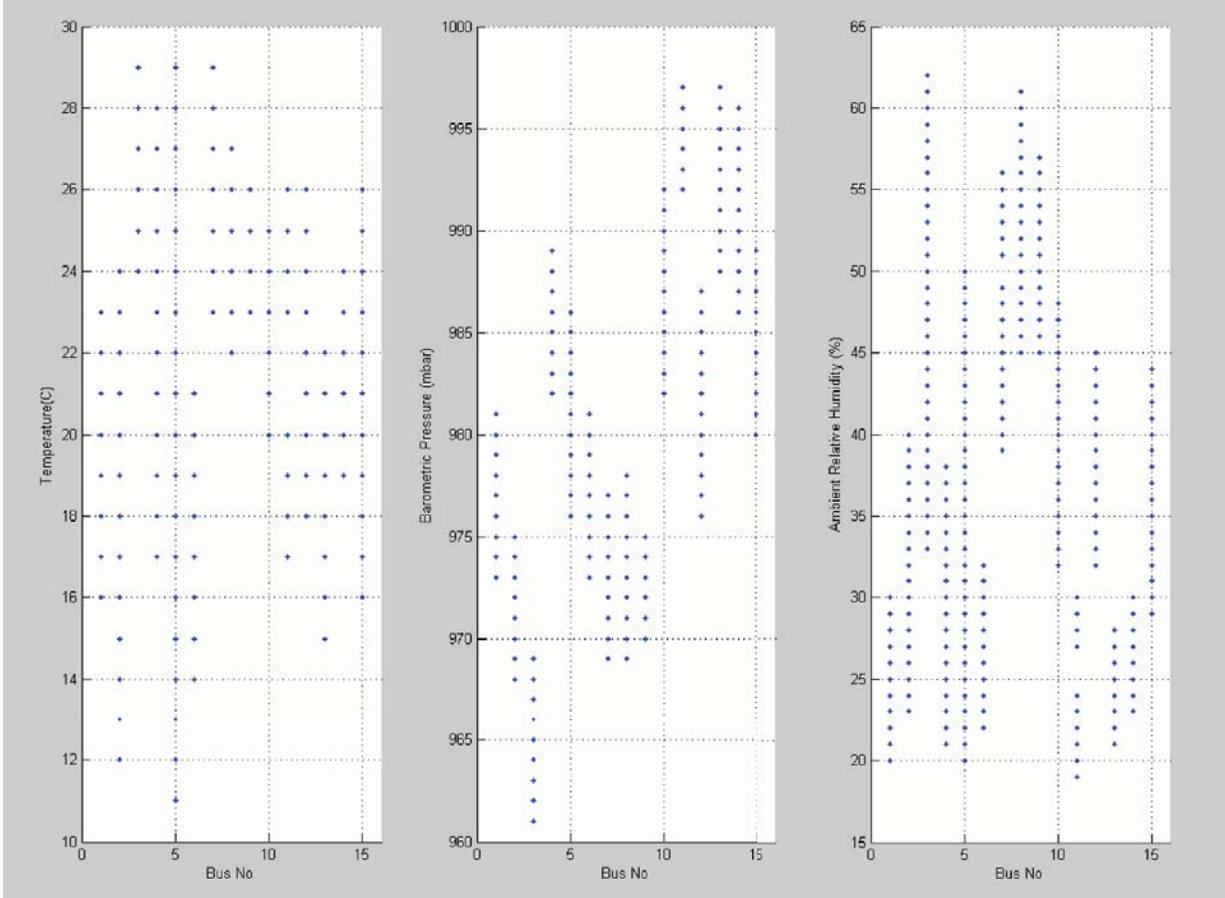


Figure 6-5 Test Environmental Conditions

Transit buses tested were provided by the AATA and all of them are New Flyer models with Detroit Diesel Series 50 engines. Since these buses utilized consistent engine technologies (i.e., fuel injection type, catalytic converter type, transmission type, and so on), the ability of estimated emission models to incorporate the effect of other types of vehicle technologies is limited. Another limitation is the consideration of the effects of emission control technology deterioration on emission levels since these buses were only 5 or 6 years old during the test.

### 6.3 Variability in Emissions Data

#### 6.3.1 Inter-bus Variability

Data are presented to illustrate the variability in observed data. Inter-bus variabilities are illustrated using median and mean of  $\text{NO}_x$ , CO, and HC emission rates for each bus from Figures 6-6 to 6-8. The difference between median and mean is an indicator of skewness for the distribution of emission rates.

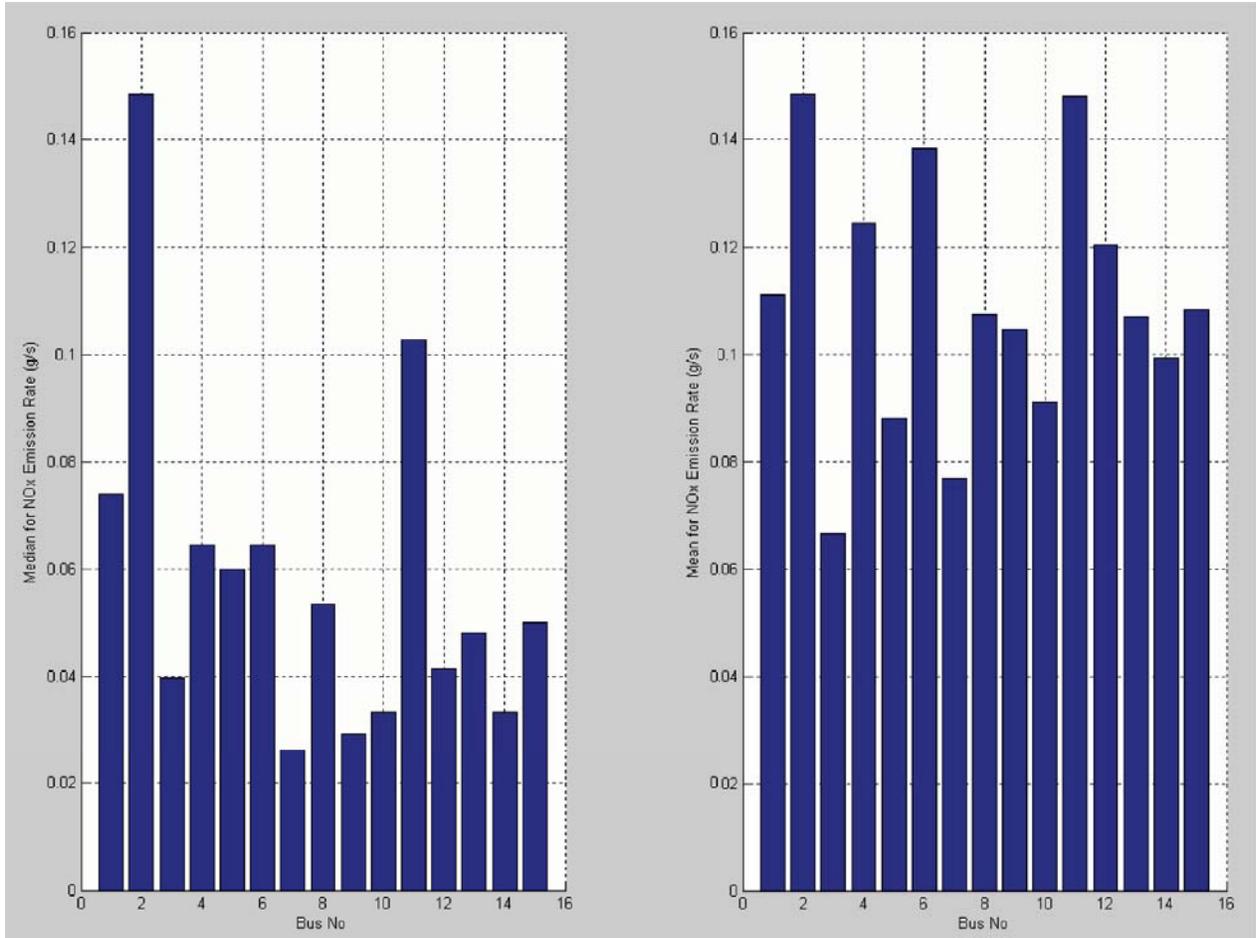


Figure 6-6 Median and Mean of NO<sub>x</sub> Emission Rates by Bus

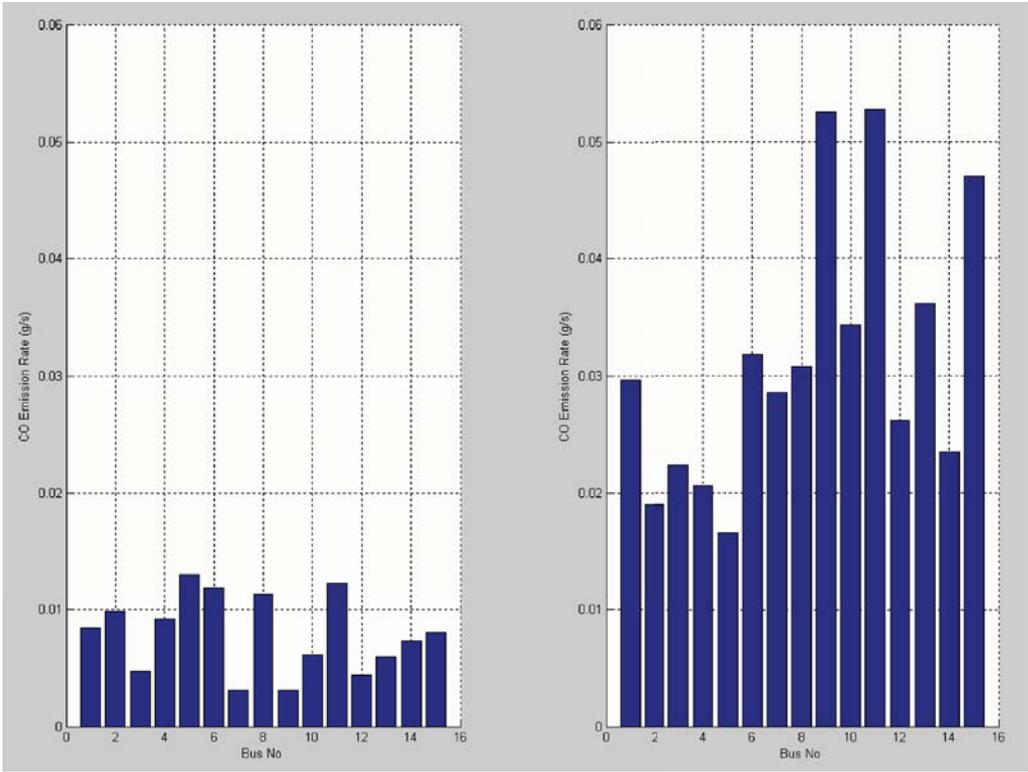


Figure 6-7 Median and Mean of CO Emission Rates by Bus

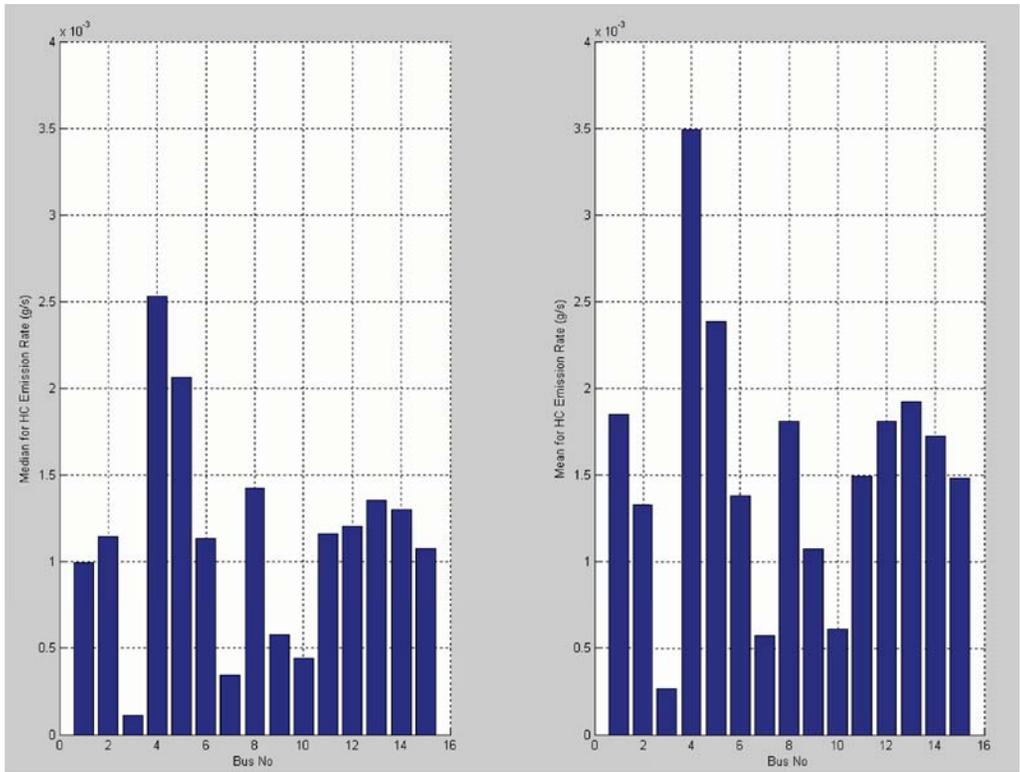


Figure 6-8 Median and Mean of HC Emission Rates by Bus

The purpose of inter-bus variability analysis was to characterize the range of variability in vehicle average emissions among all of the buses, to determine whether the data set is relatively homogeneous. Although there are some clusters among the buses as suggested from Figures 6-6 to 6-8 and some skewness in the distribution as suggested by upper tails in Figure 6-9, it is not obvious that this data set lacks homogeneity and should be separated into different groups. Thus, this data set is treated as a single group for purposes of analysis and model development.

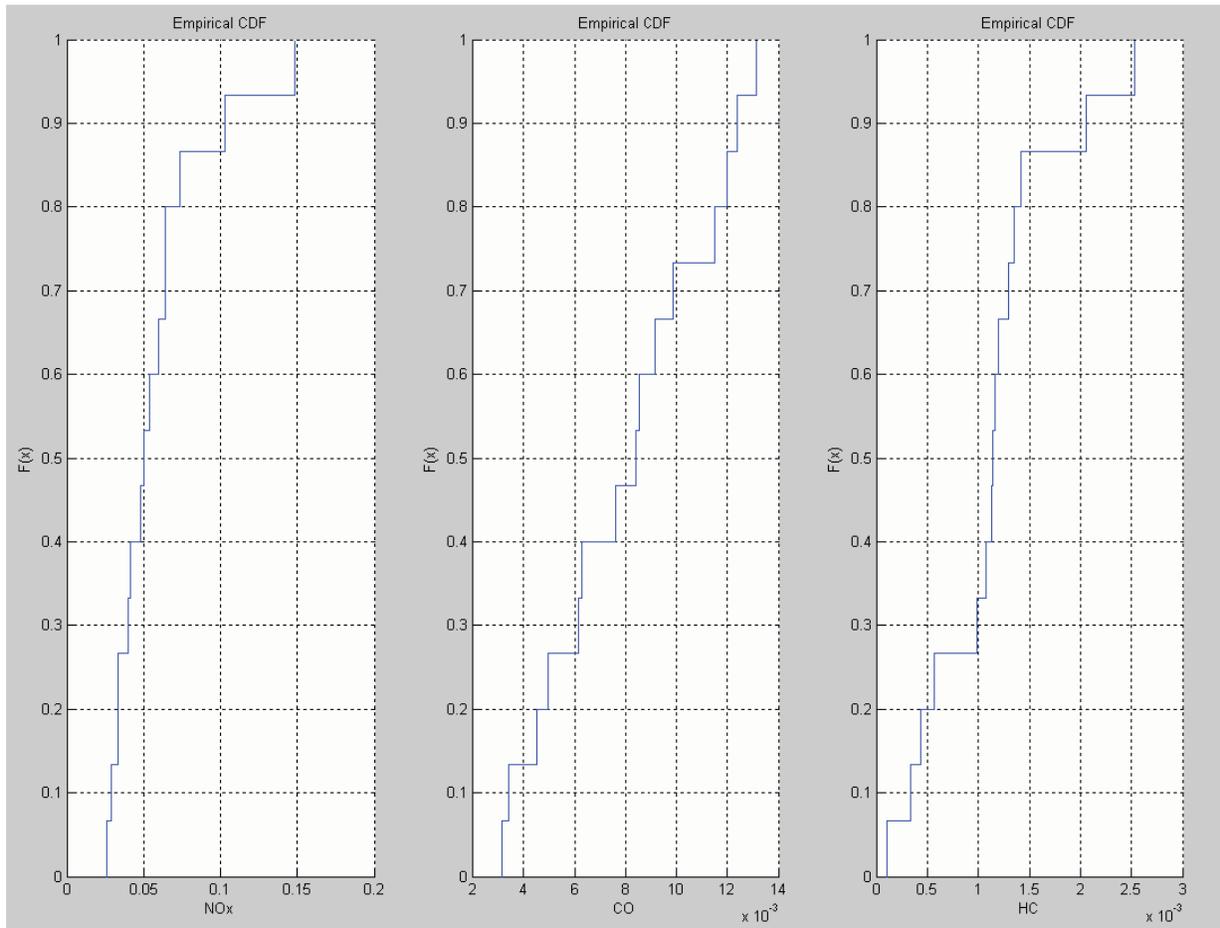


Figure 6-9 Empirical Cumulative Distribution Function Based on Bus Based Median Emission Rates for Transit Buses

### 6.3.2 Descriptive Statistics for Emissions Data

Applicable numerical summary statistics, such as variable means and standard deviations, are presented in Table 6-1. Relatively simple graphics such as histograms and boxplots describing variable distributions are presented in Figures 6-10 to 6-12. It may also be necessary to assess whether the individual variables are normally distributed prior to any further analysis using parametric methods that are based upon this assumption.

Table 6-1 Basic Summary Statistics for Emissions Rate Data for Transit Bus

*** Summary Statistics for data in: transitbus.data ***			
	NO <sub>x</sub>	HC	CO
Min:	0.000000e+000	0.000000e+000	0.000000e+000
1st Qu.:	3.030000e-003	2.195000e-002	4.200000e-004
Mean:	3.183675e-002	1.052101e-001	1.438709e-003
Median:	7.540000e-003	5.058000e-002	9.300000e-004
3rd Qu.:	2.197000e-002	1.731100e-001	1.840000e-003
Max:	3.057700e+000	2.427900e+000	6.679000e-002
Total N:	1.075350e+005	1.075350e+005	1.075350e+005
NA's :	0.000000e+000	0.000000e+000	0.000000e+000
Std Dev.:	8.479305e-002	1.162344e-001	1.956353e-003

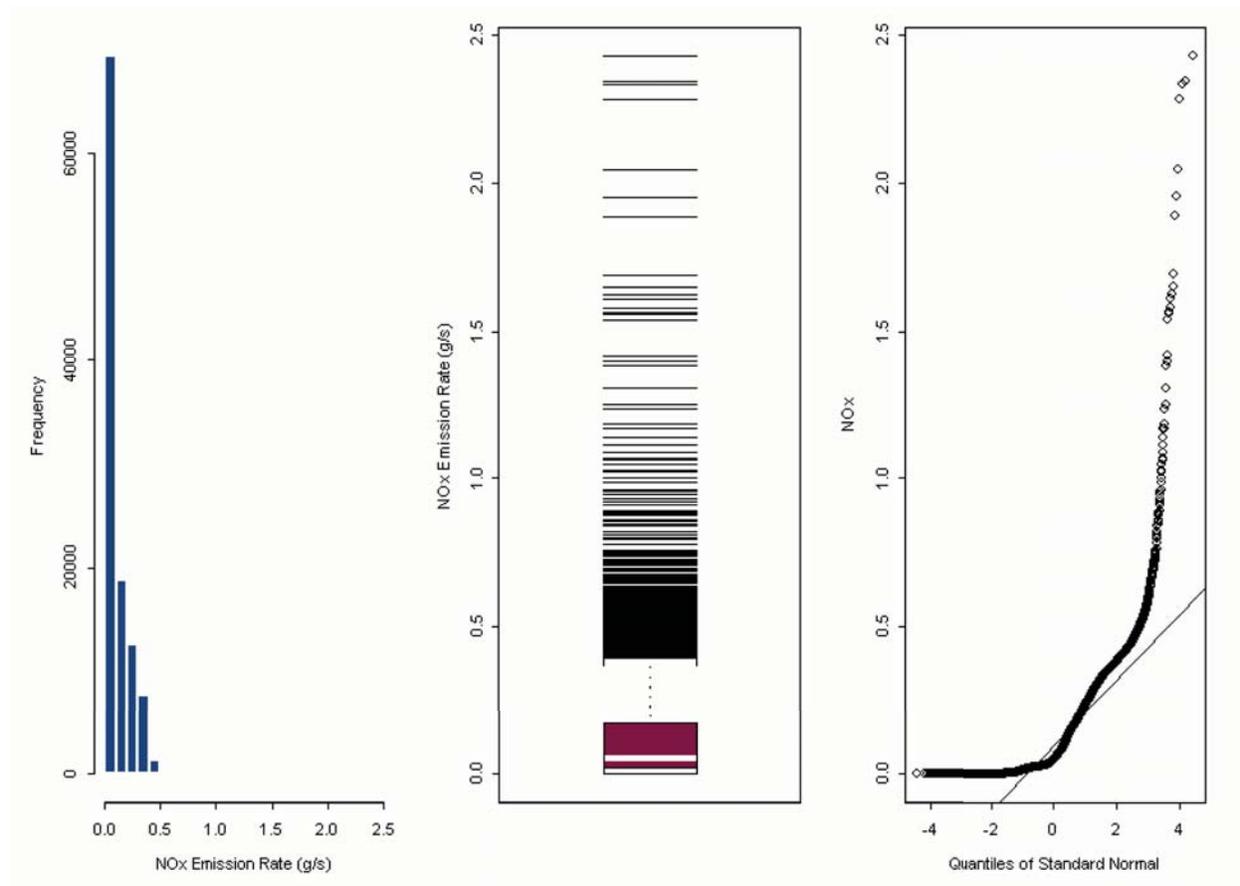


Figure 6-10 Histogram, Boxplot, and Probability Plot of NO<sub>x</sub> Emission Rate

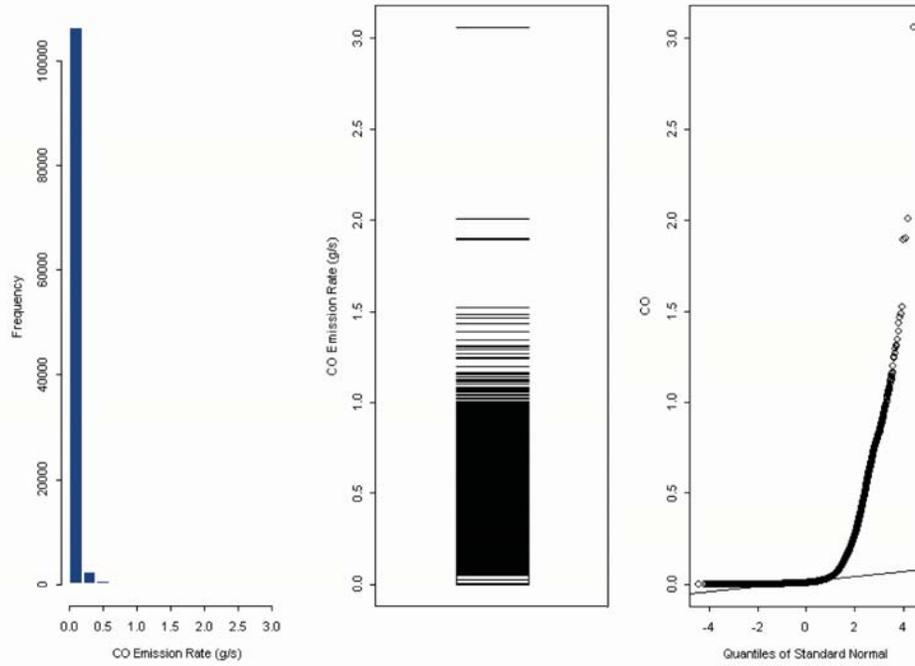


Figure 6-11 Histogram, Boxplot, and Probability Plot of CO Emission Rate

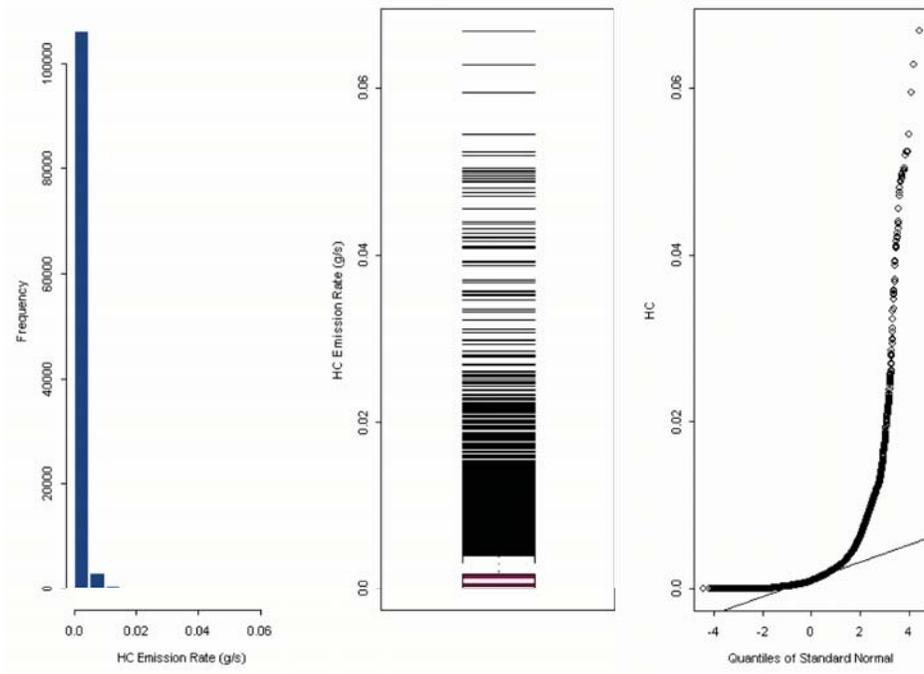


Figure 6-12 Histogram, Boxplot, and Probability Plot of HC Emission Rate

Further analysis indicated that there are some zero values in the emission data. There might be several reasons for zero values. Missing data caused by loss of communication between instruments or failure of a particular vehicle were recorded as zero in the data set. Those zero values were already identified in the data post-processing procedure in Chapter 4. Zero values might also have occurred when the reference air contained significant amounts of a pollutant so the instrument systematically reported negative emission values. Sensors, Inc. suggested that negative data should be set to zero. Thus these negative values were artificially recorded as zero, not observed by test equipment as zero. These zero values would create truncation issues in the model, since the Sensors, Inc. transit bus data set contained only valid positive emission data. Usually, truncation is found when a random variable is not observable over its entire range. Truncation could not be treated as a missing data problem as the missing observations are random. In statistics consideration or analysis can be limited to data that meet certain criteria or to a data distribution where values above or below a certain point have been eliminated (or cannot occur). A program was written in MATLAB<sup>®</sup> to check for the presence of zero emissions estimates in the data set. There were 1.45% zero values for NO<sub>x</sub> emissions, 1.65% zero values for CO emissions and 3.84% zero values for HC emissions. Since negative emission values were not observable for the transit bus data set, further analysis will focus on truncated data sets with valid positive emission data only.

The numerical summary statistics such as variable means and standard deviations for truncated emission data are presented in Table 6-2, and relatively simple graphics such as histograms and boxplots describing variable distributions are presented from Figures 6-13 to 6-15. The mean of truncated NO<sub>x</sub> emission data increases 1.26%, while the mean of truncated CO emission data increases 1.23% and the mean of truncated HC emission data increases 0.99%, compared with the means of the original data set.

Table 6-2 Basic Summary Statistics for Truncated Emissions Rate Data

	NO <sub>x</sub>	CO	HC
Min:	1.000000e-005	1.000000e-005	1.000000e-005
1st Qu.:	2.256000e-002	3.190000e-003	4.700000e-004
Mean:	1.067578e-001	3.236955e-002	1.496171e-003
Median:	5.243500e-002	7.770000e-003	9.900000e-004
3rd Qu.:	1.749625e-001	2.246000e-002	1.880000e-003
Max:	2.427900e+000	3.057700e+000	6.679000e-002
Total N:	1.059760e+005	1.057650e+005	1.034050e+005
NA's :	0.000000e+000	0.000000e+000	0.000000e+000
Std Dev.:	1.163785e-001	8.539871e-002	1.973375e-003

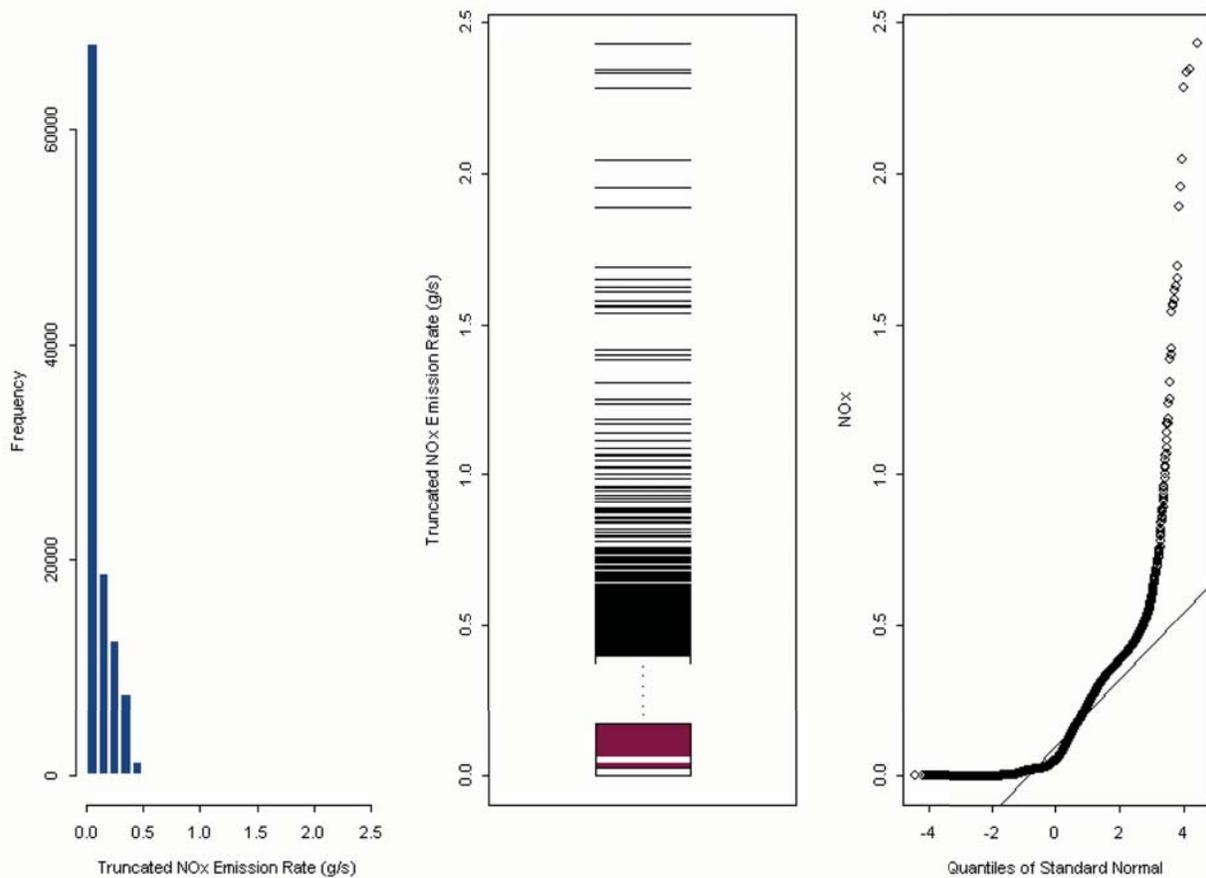


Figure 6-13 Histogram, Boxplot, and Probability Plot of Truncated NO<sub>x</sub> Emission Rate

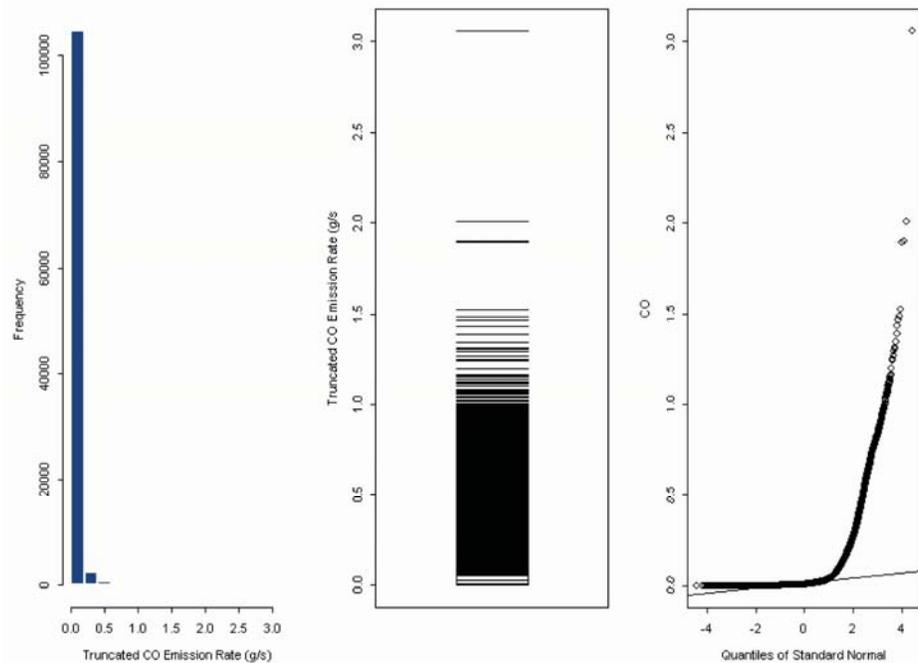


Figure 6-14 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate

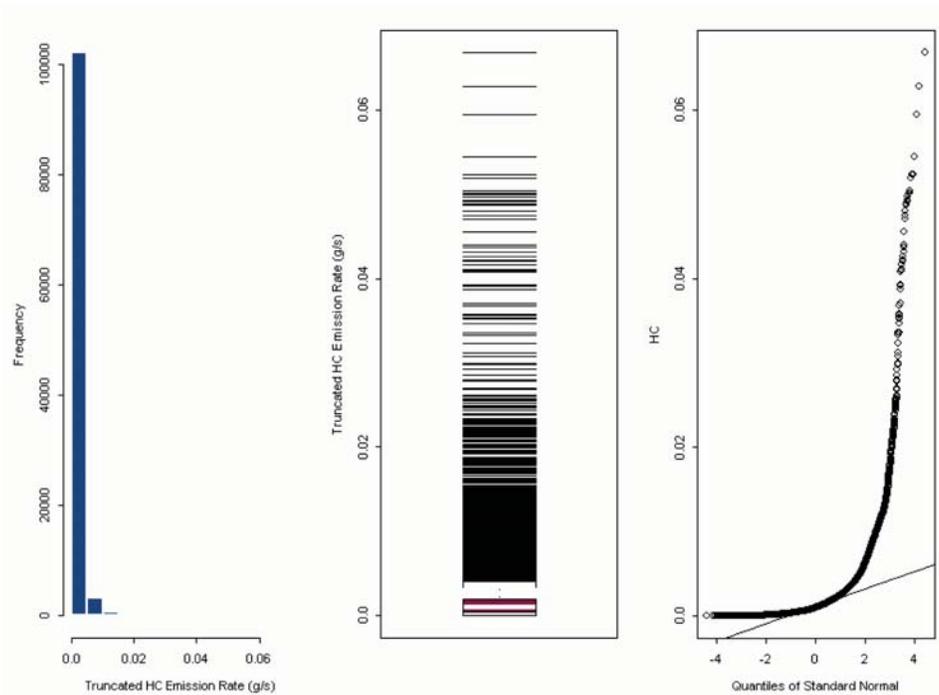


Figure 6-15 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate

These boxplots for truncated emission data show that there are some obvious outliers in the measured emissions of all three pollutants, and the histograms suggest a high degree of non-normality, also indicated in the probability plots. There is thus a need to transform the response

variable to correct for this condition. Transformations are used to present data on a different scale. In modeling and statistical applications, transformations are often used to improve the compatibility of the data with assumptions underlying a modeling process, to linearize the relation between two variables whose relationship is non-linear, or to modify the range of values of a variable (Washington et al. 2003).

### 6.3.3 Transformation for Emissions Data

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<sup>®</sup> 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 are 0.22875 for truncated NO<sub>x</sub>, -0.0648 for truncated CO, 0.14631 for truncated 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 with the power transformation. The lambda values used for transformations are 1/4 for truncated NO<sub>x</sub>, 0 for truncated CO, 0 for truncated HC. Histograms, boxplots and normal-normal plots describing transformed variable distributions are presented in Figures 6-16 to 6-18, where a great improvement is noted.

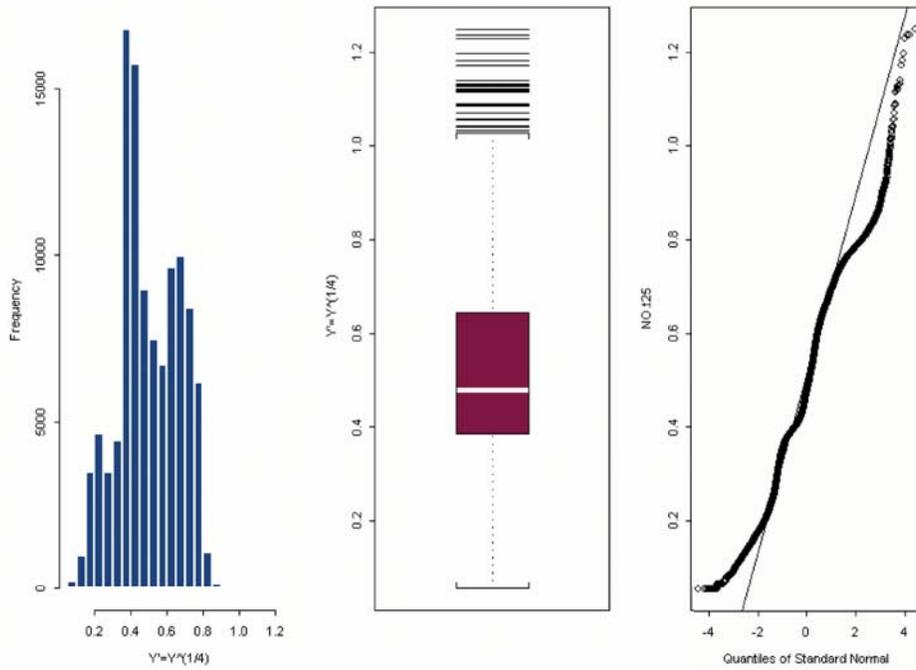


Figure 6-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO<sub>x</sub> Emission Rate

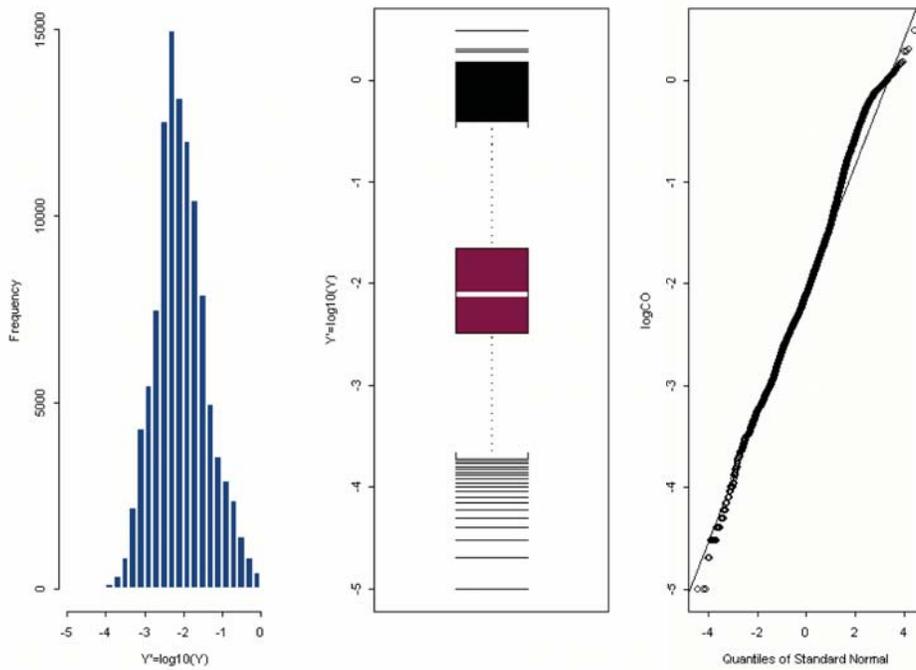


Figure 6-17 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate

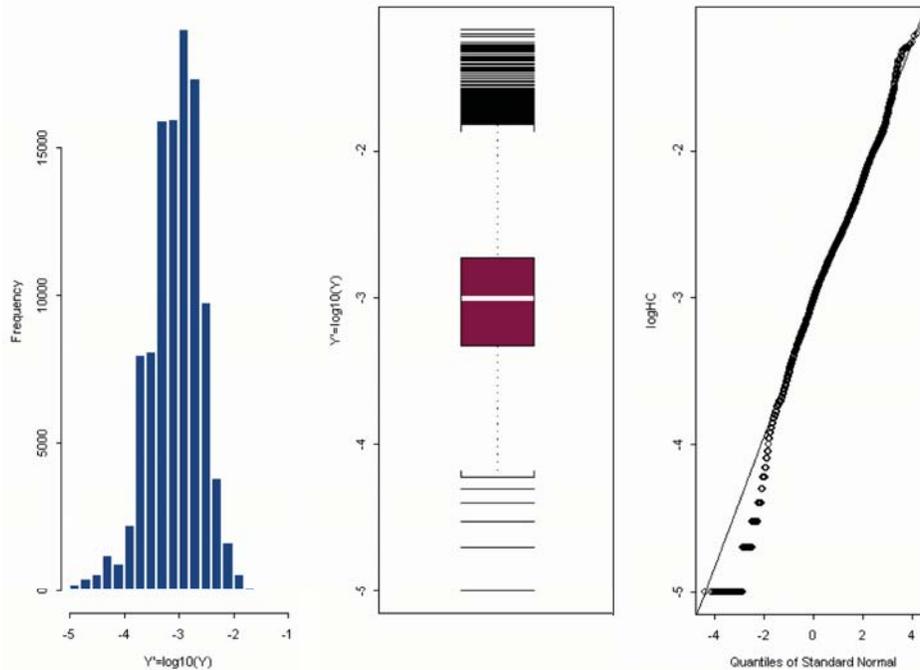


Figure 6-18 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate

Although transformations can result in improvement of a specific modeling assumption such as linearity or normality, they can often result in the violation of others. Thus, transformations must be used in an iterative fashion, with continued checking of other modeling assumptions as transformations are made. Dr. Washington suggested the comparisons should always be made on the original untransformed scale of Y when comparing statistical models and these comparisons extend to goodness of fit statistics and model validation exercises (Washington et al. 2003).

#### 6.3.4 Identification of High Emitter

From a modeling viewpoint, it is important to accurately predict the number of ‘high emitter’ vehicles in the fleet (older technology, poorly maintained, or tampered vehicles that emit significantly elevated emissions relative to the fleet average under all operating conditions) and the fraction of activities that yield high emissions for normal emitting vehicles. Historic practices to identify ‘high emitters’ in a data set have relied on judgment to set cut points that are often indefensible from a statistical, and sometimes even practical, perspective. U.S. EPA uses five times the prevailing emission standards as the cut point across all pollutants (U.S. EPA 1993), while CARB has defined different emission regimes ranging from normal to super emitters and used different criteria for each regime (CARB 1991; Carlock 1994) (see Table 6-3) .

Table 6-3 CARB Emission Regime Definition (Carlock 1994)

Emitter Status	NO <sub>x</sub>	CO	HC
Normal	≤ 1 standard	< 1 standard	< 1 standard
Moderate	1 to 2 standard	1 to 2 standard	1 to 2 standard
High	2 to 3 standard	2 to 6 standard	2 to 4 standard
Very High	3 to 4 standard	6 to 10 standard	5 to 9 standard
Super	> 4 standard	> 10 standard	> 9 standard

In contrast, the methodology employed in MEASURE database development at Georgia Tech is statistically based. Wolf et al. used regression tree techniques to classify vehicles into classes that behave similarly, exhibit similar technology characteristics, and exhibit similar mean emission rates under standardized testing conditions (Wolf et al. 1998). The cut points within each technology class are then defined on the basis of pre-selected percentiles of a normal distribution of the emission rates for each pollutant. The analysis by Wolf et al. specified a cut point of 97.73 percent (that is, mean + 2 standard deviations), which implies that approximately 2.27 percent of the vehicles in each technology class are high emitters.

For this research, although inter-bus variability exists in the data set, these 15 buses should be treated as one technology class because they shared the same fuel injection type, catalytic converter type, transmission type, and their model year and odometer reading were similar. Just as in Wolf's approach, the emissions value located at two standard deviations above the mean of the normalized emissions distribution is used as a cutpoint to distinguish between normal and high emission points. Theoretically, this method will consistently identify approximately 2.27 percent of the data as high emission points. That means 97.73 percent of the population should fall into the normal status. Analysis results showed that 0.33 percent of NO<sub>x</sub> emission, 3.76 percent of CO emission, and 1.37 percent of HC emissions were identified as high emission points. After assigning those high emissions points to different buses, the distribution is shown in Table 6-4.

Table 6-4 Percent of High Emission Points by Bus

	NO <sub>x</sub>	CO	HC
bus 360	0.02%	2.80%	5.06%
bus 361	0.32%	1.08%	0.25%
bus 363	0.06%	3.10%	0.00%
bus 364	0.04%	0.87%	7.38%
bus 372	0.00%	0.13%	1.96%
bus 375	0.69%	3.16%	0.27%
bus 377	0.00%	4.44%	0.00%
bus 379	0.67%	2.85%	1.17%
bus 380	0.52%	7.67%	0.69%
bus 381	0.10%	4.76%	0.14%
bus 382	1.14%	8.12%	0.36%
bus 383	0.88%	3.44%	1.82%
bus 384	0.50%	5.10%	1.33%
bus 385	0.55%	2.10%	0.60%
bus 386	0.20%	6.63%	0.57%
Total	0.36%	3.81%	1.38%

For each individual bus, the highest proportion is 1.14 percent for bus 382 for NO<sub>x</sub> emissions, 8.12 percent for bus 380 for CO emissions, and 7.38 percent for bus 364 for HC emissions. No evidence from Table 6-4 suggests that there are some “high emitters” (older technology, poorly maintained, or tampered vehicles) in the data set. This conclusion makes sense since all buses were only 5 or 6 years old during the test. Another finding indicated that a small fraction of a bus’s observed activity exhibited disproportionately high emissions. Activities found in the literature include hard accelerations at low speeds, moderate acceleration at high speeds, or equivalent accelerations against gravity (Fomunung 2000). Given that high emissions points make up only 0.33 percent of the data set for NO<sub>x</sub>, 3.76 percent for CO, and 1.37 percent for HC, it is not necessary to develop two different models for normal emissions and high emissions. Based on this analysis, these 15 buses should be treated as one technology class since no high emitters were identified.

#### 6.4 Potential Explanatory Variables

There are four main groups of parameters that affect vehicle emissions as indicated in the literature (Guensler 1993; Clark et al. 2002). These groups are: 1) vehicle characteristics, including vehicle type, make, model year, engine type, transmission type, frontal area, drag coefficient, rolling resistance, vehicle maintenance history, etc.; 2) roadway characteristics, including road grade and possibly pavement surface roughness, etc.; 3) on-road load parameters, like

on-road driving trace (sec-cy-sec) or speed/acceleration profile, vehicle payload, on-road operating modes, driver behavior, etc.; and 4) environmental conditions, including humidity, ambient temperature, and ambient pressure (Feng et al. 2005; Guensler et al. 2005).

In general, emissions from HDDVs are more likely to be a function of brake-horsepower load on the engine (especially for  $\text{NO}_x$ ) than emissions from light-duty gasoline vehicles, because instantaneous emissions levels of diesel engines are highly correlated with the instantaneous work output of the engine (Ramamurthy et al. 1999; Feng et al. 2005). That is, in particular, the higher the engine load, the higher emissions for  $\text{NO}_x$ . The emissions modeling framework (from which most of the items below are derived) is outlined in the Regional Applied Research Effort (RARE) report (Guensler et al. 2006). The goal of that modeling regime was to predict on-road load and then apply appropriate emission rates to the load. Most of the items outlined below are related to the amount of engine load that a vehicle will experience. Although each of the variables below is important, the values are not always available in on-road testing data (although in the future we need to make sure that these data are all collected). But, engine load in the AATA database could be used in emission rate model development for this research. Also, there are some factors, such as temperature and humidity, that may affect emission rates independent of load, or perhaps interacting with load. The model should incorporate such variables.

#### **6.4.1 Vehicle Characteristics**

Factors related to vehicle characteristics influencing heavy-duty diesel vehicle emissions which are summarized in the literature include vehicle class (i.e., weight, engine size, horsepower rating), model year, vehicle mileage, emission control system (i.e., engine exhaust aftertreatment system), transmission type, inspection and maintenance history, etc. (Guensler 1993; Clark et al. 2002).

The effect of vehicle class on emissions is significant. Five main factors that cause a vehicle to demand engine power are vehicle speed, vehicle acceleration, drive train inertial acceleration, vehicle weight, and road grade. As the required power and work performed by the vehicle increase, the amount of fuel burned to produce that power also increases, and the applicable emission rates also generally increase. Thus, emissions vary as a function of vehicle class and vehicle configuration. The higher truck classes with larger engines are heavier and, thus, typically produce more emissions. Vehicle configurations with large frontal areas and high drag coefficients will yield higher emissions when operated at higher speeds and/or accelerated at higher rates.

The concept of vehicle technology groups is to identify and track subsets of vehicles that have similar on-road load responses and similar laboratory emission rate performance. The basic premise is that vehicles in the same heavy-duty vehicle class, employing similar drive train systems, and of the same size and shape have similar load relationships. There is also an important practical consideration in establishing vehicle technology groups. Researchers need to be able to identify these vehicles in the field during traffic counting exercises.

The starting point for technology group criteria is a visual classification scheme. Yoon et al. (Yoon et al. 2004a) developed a new HDV visual classification scheme called the X-scheme based on the number of axles and gross vehicle weight ratings (GVWR) as a hybrid scheme between the FHWA truck and U.S. EPA HDV classification schemes. With field-observed HDV volumes, emissions rates estimated using the X-scheme were 34.4% and 32.5% higher for NO<sub>x</sub> and PM, compared to using the standard U.S. EPA guidance (U.S. EPA 2004c). The X-scheme reflects vehicle composition in the field more realistically than does the standard U.S. EPA guidance (U.S. EPA 2004c), which shifted heavy-HDV volumes into light- or medium-HDV volumes 21% more frequently than the X-scheme. Figure 6-19 shows X-scheme classes and their typical figures (Yoon et al. 2004a).

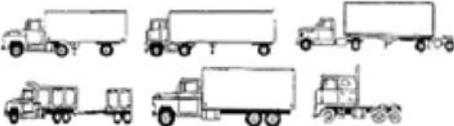
X Class	EPA Class	Typical Figures
X1	HDV2b, HDV3,HDV4, HDV5,HDV6, HDV7	
X2	HDV8a	
X3	HDV8b	

Figure 6-19 The X Classes and Typical Vehicle Configurations

Vehicle age and model year effects are accounted for because some vehicle models have much lower average emissions. Researchers from West Virginia University reported that most regulated emissions from engines produced by Detroit Diesel Corporation have declined over the years and the expected trend of decreasing emission levels with the model year of the engine

is clear and consistent for PM, HC, CO and NO<sub>x</sub>, starting with the 1990 models (Prucz et al. 2001). Information on vehicle age can be obtained from a registration database using vehicle identification numbers and truck manufacturer records. The registration database can be sorted by calendar year and show vehicles registered in the given year by model year. However, given the differences noted between field-observation fleet composition and registration data in the light-duty fleet (Granell et al. 2002), significant additional research efforts designed to model the on-road subfleet composition (classifications and model year distributions) are even more warranted for HDVs. It is also important to keep in mind that heavy-duty engines accumulate miles of travel very rapidly and that engine rebuilding is a common practice. Hence, the age of the vehicle does not necessarily equal the age of the engine. Previous field work in Atlanta indicates that on-road surveys provide better information on fleet composition (Ahanotu 1999). To refine the model, appropriate data sets that include detailed information on engine type, transmission type, etc. will be needed to appropriately subdivide the observed on-road groups and continue to develop respective emission rates. The data collection challenge in this area is daunting, but it is worthwhile to perform once to provide a library of information that can be used in a large number of modeling applications.

Vehicle weight is critical to the demand engine power that must be supplied to produce the tractive force needed to overcome inertial and drag forces and then influence vehicle emissions. NO<sub>x</sub> emissions increase as the vehicle weight increases and this relationship does not vary much from vehicle to vehicle (Gajendran and Clark 2003). The effects of vehicle age, engine horsepower ratings, transmission type, and engine exhaust aftertreatment were also investigated in other literature (Clark et al. 2002; Feng et al. 2005).

The vast majority of heavy-duty vehicles are normal emitters, but a small percentage of vehicles are high-emitters under every operating condition, typically because they have been tampered with or they are malfunctioning (i.e., defective or mal-maintained engine sensors or actuators). As the vehicle ages, general engine wear and tear will increase emission rates moderately due to normal degradation of emission controls of properly functioning vehicles. On the other hand, as vehicles age, the probability increases that some of the vehicles will malfunction and produce significantly higher emissions (i.e., become high-emitters). Probability functions that classify vehicles within specific model years (and later, within specific statistically-derived vehicle technology groups) are currently being developed through the assessment of certification testing and various roadside emissions tests. Obtaining additional detailed sources of data for developing failure models appears to be warranted.

After engine horsepower at the output shaft has been reduced by power losses associated with fluid pressures, operation of air conditioning, and other accessory loads, there is still an

additional and significant drop in available power from the engine before reaching the wheels. Power is required to overcome mechanical friction within the transmission and differential, internal working resistance in hydraulic couplings and friction of the vehicle weight on axle bearings. The combined effect of these components is parameterized as drive train efficiency. However, the more difficult and more significant component of power loss in the drive train is associated with the inertial resistance of drive train components rotational acceleration (Gillespie 1992).

A heavy-duty truck drive train is significantly more massive than its light-duty counterpart. The net effect of drive train inertial losses when operating in higher gears on the freeway may not be significant enough to be included in the model (relative to the other load-related components in the model for these heavy vehicles). However, recent studies appear to indicate very high truck emission rates (gram/second) in “creep mode” stop and start driving activities noted in ports and rail yards. Thus, high inertial loads for low gear, low speed, and acceleration operations may contribute significantly to emissions from mobile sources in freight transfer yards and therefore should not be ignored (Guensler et al. 2006).

The inertial losses are a function of a wide variety of physical drive train characteristics (transmission and differential types, component mass, etc.) and on-road operating conditions. To refine the use of inertial losses in the modal model, new drive train testing data will be designed to evaluate the inertial losses for various engine, drive shaft, differential, axle, and wheel combinations and to establish generalized drive train technology classes. Then, gear selection probability matrices for each drive train technology class and gear and final drive ratio data can be provided in lookup tables for model implementation, in place of the inertial assumptions currently employed. However, data are currently significantly lacking for development of such lookup tables.

#### **6.4.2 Roadway Characteristics**

The three basic geometric elements of a roadway are the horizontal alignment, the cross-slope or amount of super-elevation and the longitudinal profile or grade. Among them, road grade has been shown to have significant impact on engine load and vehicle emissions (Guensler 1993). Other roadway characteristics, such as lane width, are also noted to have a significant impact on the speed-acceleration profiles of heavy-duty vehicles and can therefore affect engine load (Grant et al. 1996).

### **6.4.3 Onroad Load Parameters**

Onroad load parameters include on-road driving trace (second-by-second) or speed/acceleration profile, engine load, on-road operating modes (i.e., idling, motoring, acceleration, deceleration, and cruise), driver behavior, and so on. Vehicle speed and acceleration are integral components for the estimation of vehicle road load, and therefore engine load. Previous studies indicated that increased engine power requirements could result in the increase in NO<sub>x</sub> emissions (Ramamurthy and Clark 1999; Feng et al. 2005). Clark et al. reported that the vehicle applications and duty cycles can have an effect on the emission produced (Clark et al. 2002). This study found that over a typical day of use for any vehicle, one that stops and then accelerates more often might produce higher distance-specific emissions, providing all else is held constant.

Passenger and freight payloads together with the vehicle tare weight contribute to the demand for power that must be supplied to produce the tractive force needed to overcome inertial and drag forces. Passenger loading functions for transit operations can be obtained through analysis of fare data or on-board passenger count programs. On the heavy-duty truck side, on-road freight weight distributions by vehicle class can be derived from roadside weigh station studies. Ahanotu conducted detailed weigh-in-motion studies in Atlanta and found that reasonable load distributions by truck class and time of day could be applied in such a modal modeling approach (Ahanotu 1999). Although additional field studies are warranted to examine the validity of the Atlanta results over time and the transferability of findings in Atlanta to other metropolitan areas (especially considering the potential variability in commodity transport, such as agricultural goods, that may occur in other areas), the modeling methodology seems appropriate.

### **6.4.4 Environmental Conditions**

Environmental conditions under which the vehicle is operated include humidity, ambient temperature, and ambient pressure. U.S. EPA is currently conducting studies to find the effect of ambient conditions on HDDV emissions (NRC 2000). The current MOBILE6.2 model includes correction factors to account for the impact of environmental conditions on vehicle emission rates. Given the lack of compelling additional data available for analysis, it may be necessary to ignore the effects of these environmental parameters (altitude, temperature, and humidity) or simply incorporate the existing MOBILE6.2 correction factors. Preliminary analyses of the data and methods used to derive the MOBILE6.2 environmental correction factors indicate that the embedded equations in MOBIL6.2 probably need to be revisited.

### 6.4.5 Summary

It is impossible for modeler to include all explanatory variables identified in the literature review for model development because the explanatory variables available for model development and model validation are only a subset of potential explanatory variables identified above. Therefore, the conceptual model will only include available variables and derived variables in the data set provided.

### 6.5 Selection of Explanatory Variables

As mentioned earlier, available explanatory variables for transit buses are only a subset of potential explanatory variables identified. In brief, available explanatory variables can be summarized as:

- *Test information*: date, time;
- *Vehicle characteristics*: license number; model year, odometer reading, engine size, instrument configuration number;
- *Roadway characteristics*: road grade (%);
- *Onroad load parameters*: engine power (bhp), vehicle speed (mph), acceleration (mph/s);
- *Engine operating parameters*: throttle position (0 – 100%), engine oil temperature (deg F), engine oil pressure (kPa), engine warning lamp (Binary), engine coolant temperature (deg F), barometric pressure reported from ECM (kPa);
- *Environmental conditions*: ambient temperature (deg C), ambient pressure (mbar), ambient relative humidity (%), ambient absolute humidity (grains/lb air).

The most important question related to engine power is how to simulate engine power in the real world for application purposes. Georgia Institute of Technology researchers developed a transit bus engine power demand simulator (TB-EPDS), which estimates transit bus power demand for given speed, acceleration, and road grade conditions (Yoon et al. 2005a; Yoon et al. 2005b). Speed-acceleration-road grade matrices were developed from speed and location data obtained using a Georgia Tech Trip Data Collector. The researchers conclude that speed-acceleration-road grade matrices at the link level or the route level are both acceptable for regional inventory development. However, for micro-scale air quality impact analysis, link-based ma-

trices should be employed (Yoon et al. 2005a). Although significant uncertainties still exist for inertial loss which is significant at low speeds and motoring mode with negative engine power, this research showed that using engine power as load data is possible for application purposes. Thus we concluded that engine power could be used as load data in estimated emission models.

The relationships between explanatory variables were investigated using S-Plus®. Three variables were excluded because they have only a single value for all records, and they are engine size, instrument configuration number and engine warning lamp. There are 14 explanatory variables included in correlation analysis. The correlation matrix is shown in Table 6-5.

Table 6-5 Correlation Matrix for Transit Bus Data Set

\*\*\* Correlations for data in: transitbus.data \*\*\*

	model.year	odometer	temperature	baro
model.year	1.000000000	-0.655273106	0.047048515	0.394378106
odometer	-0.655273106	1.000000000	0.186771499	-0.704310642
temperature	0.047048515	0.186771499	1.000000000	-0.326938545
baro	0.394378106	-0.704310642	-0.326938545	1.000000000
SCB.RH	0.068411842	0.343814465	0.488214011	-0.632480147
humid	0.030997734	0.39026148	0.751260451	-0.649522446
grade	-0.004241021	0.00052737	-0.005590441	0.002384338
vehicle.speed	-0.014916204	-0.062908098	-0.225478003	0.054918347
throttle.position	-0.00186824	0.009346571	-0.09113266	-0.014470281
oil.temperature	0.051759069	-0.011881827	0.042676227	-0.026744091
oil.pressure	0.050521339	-0.098442472	-0.073256993	0.034212231
coolant.temperature	0.206727241	-0.117710067	0.077114798	0.045844706
eng.bar.press	0.137781076	-0.248876183	-0.260525088	0.371021489
engine.power	-0.006066455	0.021283229	-0.059512654	-0.035718725
	SCB.RH	humid	grade	vehicle.speed
model.year	0.0684118427	0.030997734	-0.004241021	-0.014916204
odometer	0.3438144652	0.390261480	0.00052737	-0.062908098
temperature	0.4882140119	0.751260451	-0.005590441	-0.225478003
baro	-0.6324801472	-0.649522446	0.002384338	0.054918347
SCB.RH	1.0000000000	0.931879078	-0.006075112	-0.034502697
humid	0.9318790788	1.000000000	-0.006411009	-0.117870984
grade	-0.0060751123	-0.006411009	1.000000000	0.000896568
vehicle.speed	-0.0345026977	-0.117870984	0.000896568	1.000000000
throttle.position	0.0134235743	-0.024720165	0.020186507	0.387705398
oil.temperature	0.096018579	0.087317807	-0.007116669	0.018641433
oil.pressure	-0.0498528376	-0.077649741	0.009836954	0.567493814

coolant.temperature	0.2005559889	0.171558840	-0.014531524	0.072998199
eng.bar.press	-0.3663829274	-0.373540032	0.002132063	0.143270319
engine.power	0.0257436423	-0.003279122	0.021662091	0.303209657
acc	0.0000403711	0.003340728	0.012930076	0.000224126
	throttle.position	oil.temperature	oil.pressure	
model.year	-0.001868240	0.051759069	0.050521339	
odometer	0.009346571	-0.011881827	-0.098442472	
temperature	-0.091132660	0.042676227	-0.073256993	
baro	-0.014470281	-0.026744091	0.034212231	
SCB.RH	0.013423574	0.096018570	-0.049852837	
humid	-0.024720165	0.087317807	-0.077649741	
grade	0.020186507	-0.007116669	0.009836954	
vehicle.speed	0.387705398	0.018641433	<b>0.567493814</b>	
throttle.position	1.000000000	0.012077329	<b>0.681336402</b>	
oil.temperature	0.012077329	1.000000000	-0.117896787	
oil.pressure	<b>0.681336402</b>	-0.117896787	1.000000000	
coolant.temperature	0.059605193	0.335667341	-0.298083257	
eng.bar.press	0.102861968	0.059886972	0.022549030	
engine.power	<b>0.959310116</b>	0.007171781	<b>0.656609695</b>	
acc	<b>0.660747116</b>	-0.004185245	0.465493435	
	coolant.temperature	eng.bar.press	engine.power	
model.year	0.206727200	41 0.137781076	-0.006066455	
odometer	-0.117710000	67 -0.248876183	0.021283229	
temperature	0.077114700	98 -0.260525088	-0.059512654	
baro	0.045844700	06 0.371021489	-0.035718725	
SCB.RH	0.200555900	88 -0.366382927	0.025743642	
humid	0.171558800	40 -0.373540032	-0.003279122	
grade	-0.014531500	24 0.002132063	0.021662091	
vehicle.speed	0.072998100	99 0.143270319	0.303209657	
throttle.position	0.059605100	93 0.102861968	<b>0.959310116</b>	
oil.temperature	0.335667300	41 0.059886972	0.007171781	
oil.pressure	-0.298083200	57 0.022549030	<b>0.656609695</b>	
coolant.temperature	1.000000000	00 0.284506753	0.050584845	
eng.bar.press	0.284506700	53 1.000000000	0.089702976	
engine.power	0.050584800	45 0.089702976	1.000000000	

All variable pairs with correlation coefficients greater than 0.5 were scrutinized and subjected to further analysis, which invariably helped in paring down the number of variables. The values in the correlation matrix show that throttle position and engine power, ambient relative humidity and ambient absolute humidity are highly correlated (higher than 0.90). Model

year and odometer, odometer and barometric pressure, barometric pressure and ambient relative humidity, barometric pressure and ambient absolute humidity, ambient absolute humidity and temperature, oil pressure and throttle position, oil pressure and vehicle speed, oil pressure and engine power, throttle position and acceleration, engine power and acceleration are moderately correlated (higher than 0.50). Other pairs of variables, however, have only slight correlations.

The relationship between throttle position and engine power is shown in Figure 6-20. Since engine power is derived from percent engine load, engine torque, and engine speed, and previous studies indicated that increased engine power requirements could result in the increase in NO<sub>x</sub> emissions (Ramamurthy and Clark 1999; Feng et al. 2005), the author retained engine power in the database.

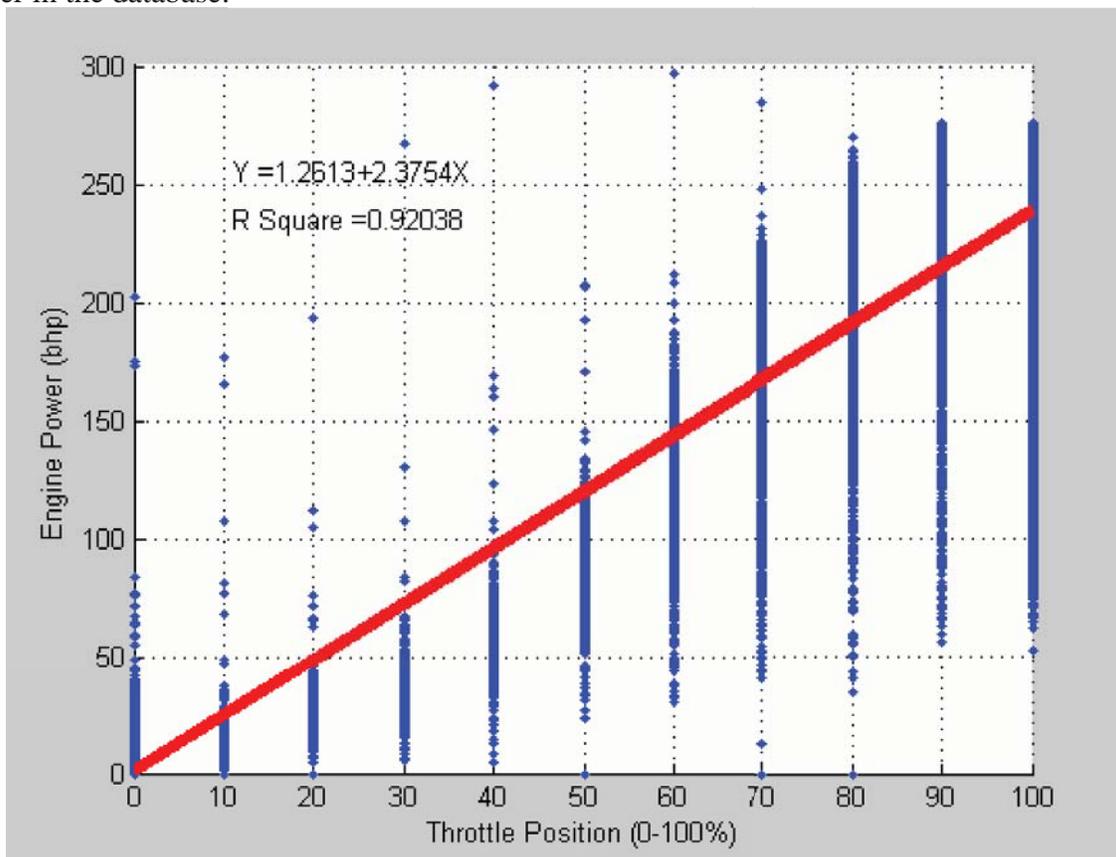


Figure 6-20 Throttle Position vs. Engine Power for Transit Bus Data Set

Ambient relative humidity and ambient absolute humidity provide the same information in two different ways, and either is enough to consider the influence of ambient humidity on emissions. The author retained ambient relative humidity in the database.

Three other findings related to the correlation matrix are:

1. All environmental characteristics, like temperature, humidity, and barometric pressure, are moderately correlated with each other (Figure 6-21), which indicates modelers should consider such relationships when developing environmental factors.
2. Engine power is correlated with not only on-road load parameters such as vehicle speed, acceleration, and road grade, but also engine operating parameters such as throttle position and engine oil pressure. Engine power in this data set is derived from measured engine speed, engine torque and percent engine load. On the other hand, engine power could be derived theoretically from vehicle speed, acceleration and road grade using an engine power demand equation. So, engine power can connect on-road modal activity with engine operating conditions at this level. This fact strengthens the importance of introducing engine power into a conceptual emissions model and to improve the ability to simulate engine power for regional inventory development.
3. Engine operating parameters, like throttle position (0 – 100%), engine oil pressure (kPa), engine oil temperature (deg F), engine coolant temperature (deg F), and barometric pressure reported from ECM (kPa), are highly or moderately related to on-road operating parameters. For example, engine power and throttle position are highly correlated, while oil pressure and vehicle speed, oil pressure and engine power, throttle position and acceleration are moderately correlated. Although engine operating parameters may have power to explain the variability of emission data, it is difficult to obtain such data in the real world for modeling purposes. These four variables are retained for further analysis of their relationships with emissions. Although these four variables will be excluded from the emission model at this time, analysis of these potential relationships may indicate a need for further research in this area.

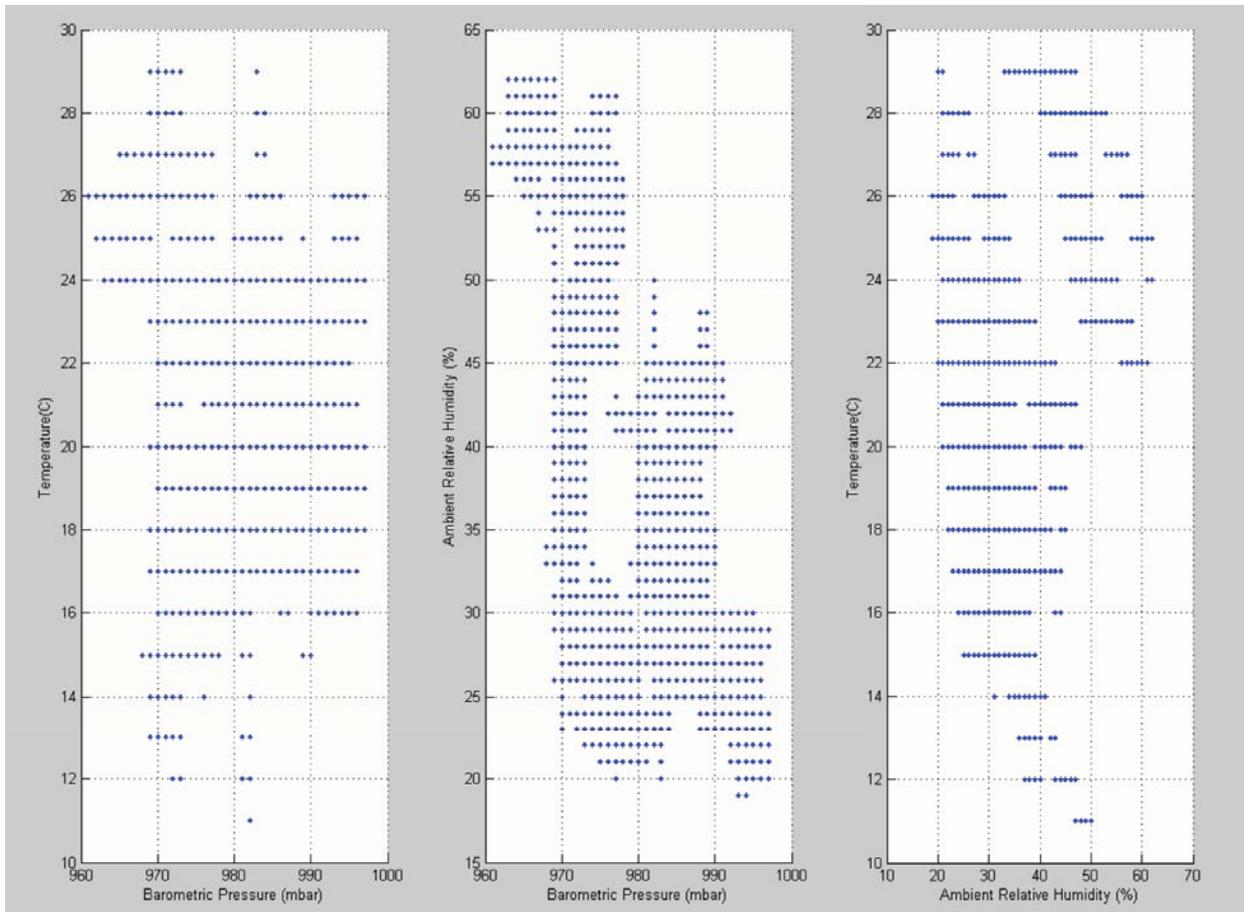


Figure 6-21 Scatter plots for environmental parameters

## CHAPTER 7

### 7. MODAL ACTIVITY DEFINITIONS DEVELOPMENT

#### 7.1 Overview of Current Modal Activity Definitions

Current research suggests that vehicle emission rates are highly correlated with modal vehicle activity. Modal activity is a vehicle activity characterized by cruise, idle, acceleration or deceleration operation. Consequently, a modal approach to transportation-related air quality modeling is becoming widely accepted as more accurate in making realistic estimates of mobile source contribution 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 done in several recent researches (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 significant portion of variability of emission data. Usually, driving can be divided into four modes: acceleration, deceleration, cruise, and idle. But driving mode definitions in literature were somewhat arbitrary. To define the driving modes or choose more reasonable definitions for the proposed modal emissions model, current driving mode definitions used in different modal emission models need to be investigated first.

##### *MEASURE's Definitions*

Researchers at Georgia Tech developed the MEASURE model in 1998 (Guensler et al. 1998). This model was developed from more than 13,000 laboratory tests conducted by the EPA and CARB using standardized test cycle conditions and alternative cycles (Bachman 1998). Modal activities variables were introduced into the MEASURE model as follows: *acceleration* (mph/sec), *deceleration* (mph/sec), *cruise* (mph) and percent in idle time. In addition, two surrogate variables were also developed, *inertial power surrogate* (IPS) (mph<sup>2</sup>/s), which was defined as acceleration times velocity and *drag power surrogate* (DPS) (mph<sup>3</sup>/s), which was defined as acceleration times velocity squared. Within each mode, several 'cut points', or threshold values,

were specified and used to create several categories. In total, six threshold values were defined for acceleration, three for deceleration, five for cruise modes, seven for IPS, and seven for DPS. Modal activity surrogate variables were added as percent of cycle time spend in specified operating conditions (Fomunung et al. 1999).

#### *NCSU's Definitions*

Dr. Frey at NCSU defined four modes of operation (idle, acceleration, deceleration, and cruise), for U.S. EPA's MOVES' model in 2001 (Frey and Zheng 2001; Frey et al. 2002). The following description is directly cited from his report (Frey et al. 2002).

Idle is defined as based upon zero speed and zero acceleration. The acceleration mode includes several considerations. First, the vehicle must be moving and increasing in speed. Therefore, speed must be greater than zero and the acceleration must be greater than zero. However, vehicle speed can vary slightly during events that would typically be judged as cruising. Therefore, in most instances, the acceleration mode is based upon a minimum acceleration of 2 mph/sec. However, in some cases, a vehicle may accelerate slowly. Therefore, if the vehicle has had a sustained acceleration rate averaging at least 1 mph/sec for at least three seconds or more, that is also considered acceleration. Deceleration is defined in a similar manner as acceleration, except that the criteria for deceleration are based upon negative acceleration rates. All other events not classified as idle, acceleration, or deceleration, are classified as cruising. Thus, cruising is approximately steady speed driving but some drifting of speed is allowed.

#### *Physical Emission Rate Estimator's (PERE's) Definitions*

Dr. Nam developed his definitions when he introduced his Physical Emission Rate Estimator (PERE) model in 2003 (Nam 2003). Idle is defined as speed less than 2 mph. Acceleration mode is based on acceleration rate greater than 1 mph/sec. However, deceleration is based on deceleration rate less than -0.2 mph/sec. Other events are classified as cruise mode and the acceleration range is between -0.2 mph/sec and 1 mph/sec. Nam also mentioned in his report that the definition of cruise (based only on acceleration) will change depending on the speed in future studies.

*Summary*

Current driving mode definitions related to modal emission models are all significantly different from each other. NCSU used one absolute critical value, 2 mph/sec, for acceleration and deceleration mode. However, PERE chose two different critical values, 1 mph/sec and -0.2 mph/sec, for acceleration and deceleration mode individually. The critical values, 2 mph/sec, 1 mph/sec, or 0.2 mph/sec, were chosen somewhat arbitrarily. MEASURE used several threshold values to add modal activity surrogate variables. Table 7-1 summarizes these modal activity definitions.

Table 7-1 Comparison of Modal Activity Definition

	<b>MEASURE</b>	<b>NCSU</b>	<b>PERE</b>
Idle	Speed=0, Acc=0	Speed=0, Acc=0	Speed<2
Acceleration	Acc>6, Acc>5, Acc>4, Acc>3, Acc>2, Acc>1	Acc>2 or Acc>1 for three seconds	Acc>1
Deceleration	Acc<-3, Acc<-2, Acc<-1	Acc<-2 or Acc<-1 for three seconds	Acc<-0.2
Cruise	Speed>70, Speed>60, Speed>50, Speed>40, Speed>30	Other events	-0.2<Acc<1

Note: Unit for speed is mph, unit for acceleration is mph/sec.

**7.2 Proposed Modal Activity Definitions and Validation**

Although the current mode definitions could all explain some variability in different emission data sets (Barth et al. 1996; Bachman 1998; Fomunung et al. 1999; Frey et al. 2002; Nam 2003; Barth et al. 2004), they differ significantly from each other. Determining whether to accept current definitions or develop new definitions is therefore a challenge.

MEASURE’s definitions were developed based on cycle test data and modal activity surrogate variables were added as percent of cycle time spent in specified operating conditions. Obviously, this definition is not suitable for second-by-second data. PERE’s definition could not assign all data into appropriate modes. Idle mode was defined as zero speed and zero acceleration in NCSU’s definitions. Although idle mode is defined theoretically as zero speed and zero acceleration, idle mode could not be defined in this manner without considering unavoidable measurement error and measurement noise. Based on this analysis, it seems more reasonable to develop new definitions for this proposed modal emission model, where such definitions can be derived through empirical analysis of the data. In fact, the definition of modal activity is depen-

dant on the available speed/acceleration data and data quality. For example, a lack of zero speed records does not mean that there is no idle activity in the data set.

The initial proposed modal activity definitions were defined as follows:

- Idle is defined as based on speeds less than 2.5 mph and absolute acceleration less than 0.5 mph/sec.
- Acceleration mode is based upon a minimum acceleration of 0.5 mph/sec.
- Deceleration is defined in a manner similar to acceleration, except that the criteria for deceleration are based upon negative acceleration rates.
- All other events not classified as idle, acceleration, or deceleration, are classified as cruise.

At the same time, several different critical values were chosen to examine the reasonableness of the proposed criteria. Four different mode definitions using different critical values are shown in Table 7-2.

Table 7-2 Four Different Mode Definitions and Modal Variables

	Idle	Acceleration	Deceleration	Cruise
Definition 1	Speed $\leq 2.5$ & abs(acc) $\leq 0.5$	Acc $> 0.5$	Acc $< -0.5$	Other
Definition 2	Speed $\leq 2.5$ & abs(acc) $\leq 1$	Acc $> 1$	Acc $< -1$	Other
Definition 3	Speed $\leq 2.5$ & abs(acc) $\leq 1.5$	Acc $> 1.5$	Acc $< -1.5$	Other
Definition 4	Speed $\leq 2.5$ & abs(acc) $\leq 2$	Acc $> 2$	Acc $< -2$	Other

Note: Unit for speed is mph, unit for acceleration is mph/sec.

A program was written in MATLAB™ to determine the driving mode for second-by-second data and estimate the average value of emissions for each of the driving modes. At the same time, average modal emission rates were estimated for each mode based on different modal activity definitions in Table 7-2. Figures 7-1 to 7-3 present a comparison of average modal emission rates for different pollutants (NO<sub>x</sub>, CO, and HC).

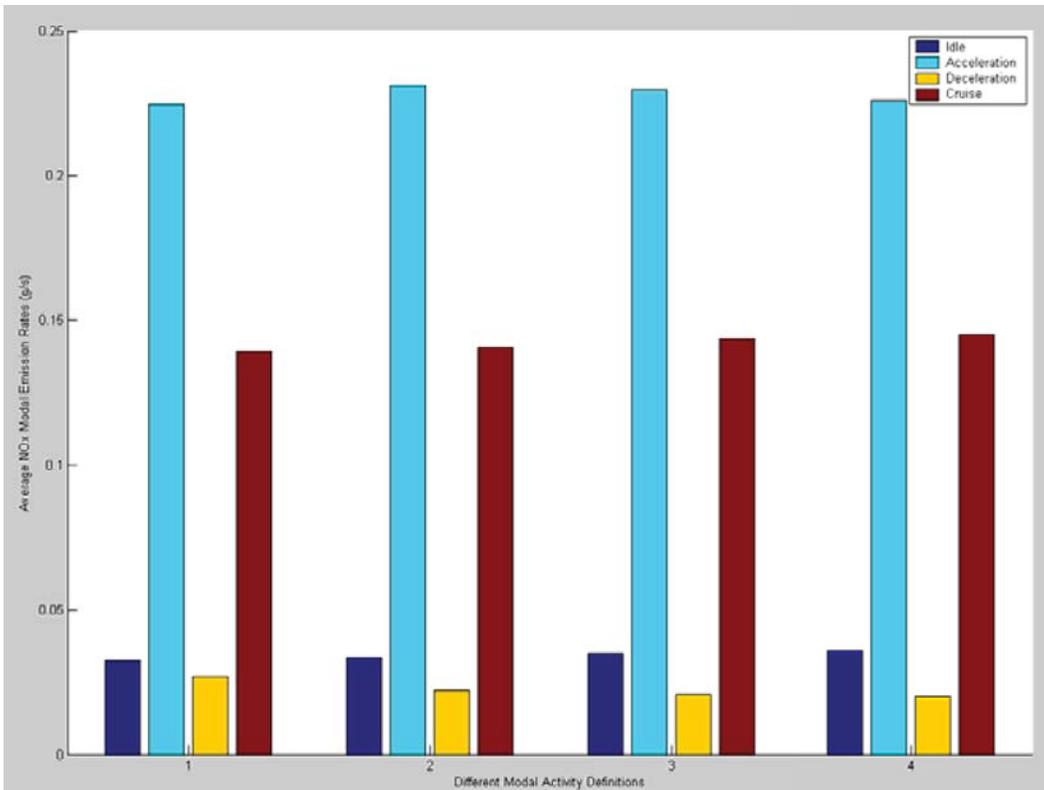


Figure 7-1 Average NO<sub>x</sub> Modal Emission Rates for Different Activity Definitions

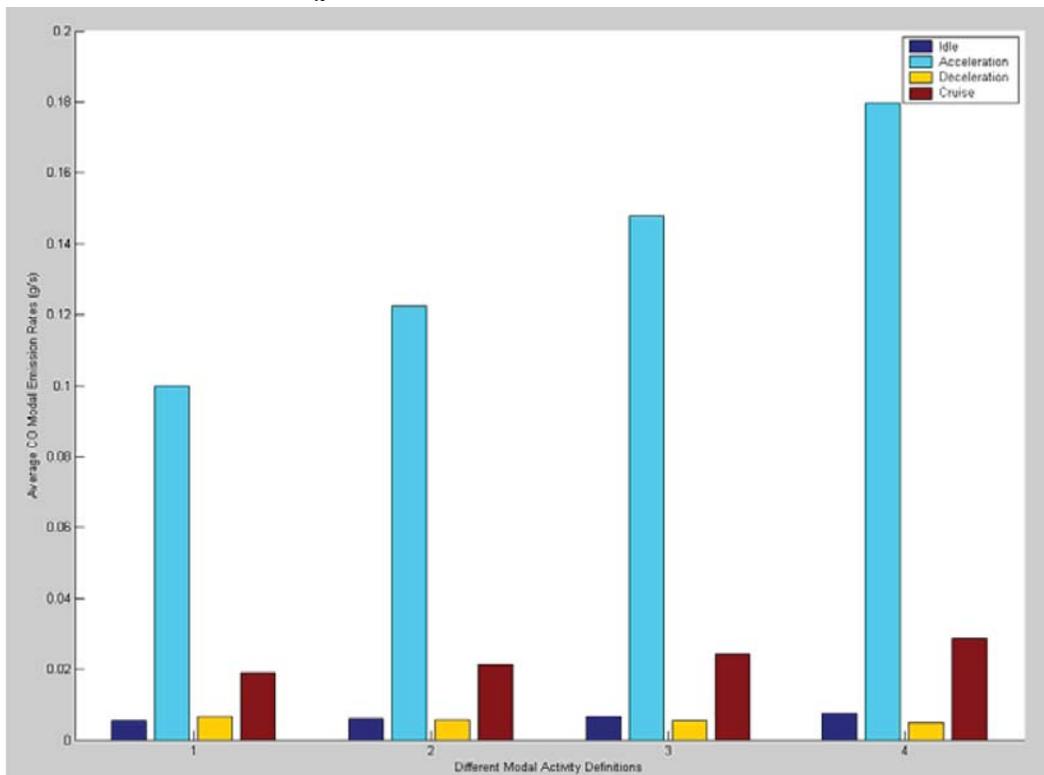


Figure 7-2 Average CO Modal Emission Rates for Different Activity Definitions

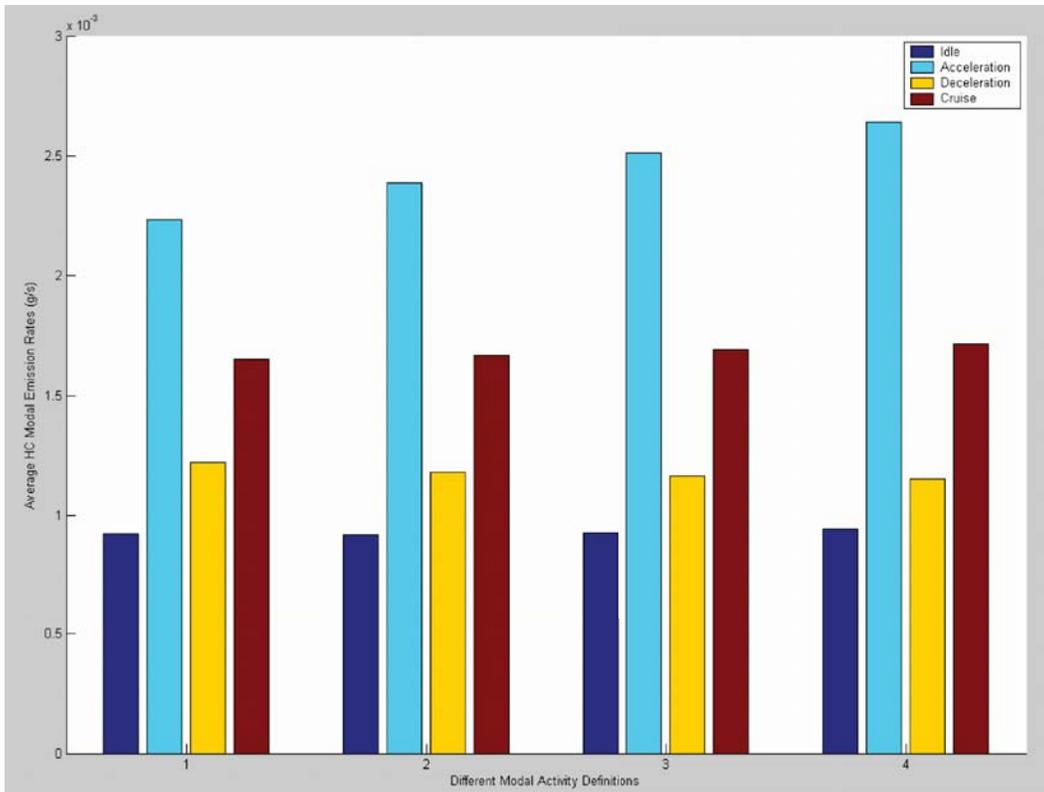


Figure 7-3 Average HC Modal Emission Rates for Different Activity Definitions

These four different modal activity definitions show a kind of consistent pattern. The average emissions during the acceleration mode are significantly higher than any other driving mode for all of the pollutants. The average emission rate during deceleration mode is the lowest of the four modes for  $\text{NO}_x$  and CO emissions while the average emission rate during idle mode is the lowest of the four modes for HC emissions. The average cruising emission rate is typically higher than the average idling and decelerating emission rate, except for CO emission in definitions 3 and 4.

To assess whether the average modal emission rates are statistically significantly different from each other, two-sample tests were estimated for each pair. Lilliefors tests for goodness of fit to a normal distribution were first used for each mode based on different modal activity definitions. The results show that all of them reject the null hypothesis of normal distribution at 5% level. A Kolmogorov-Smirnov two-sample test was chosen to take place of the *t*-test because the assumption of normal distribution 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 are presented in Table 7-3 in terms of p-values where

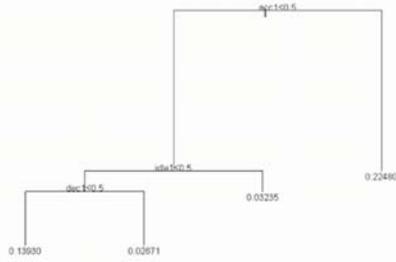
“Acc” represents acceleration mode while “Dec” represents deceleration mode. The cases where the p-value is less than 0.05 indicate that the distributions are different at the 5% level. All p-values for 72 possible pairwise comparisons are lower than 0.05, indicating that the distributions for these pairs are statistically different from each other.

Table 7-3 Results for Pairwise Comparison for Modal Average Estimates In Terms of P-value

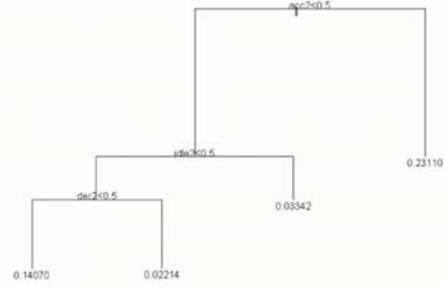
		Idle-Acc	Idle-Dec	Idle-Cruise	Acc-Dec	Acc-Cruise	Dec-Cruise
Definiton1	NO <sub>x</sub>	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton2	NO <sub>x</sub>	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton3	NO <sub>x</sub>	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0
Definiton4	NO <sub>x</sub>	0	0	0	0	0	0
	CO	0	0	0	0	0	0
	HC	0	0	0	0	0	0

The modal emission analysis results suggest that all four mode definitions proposed in Table 7-2 appear reasonable. These modal definitions allow some explanation of differences in emissions based upon driving mode, as revealed by the fact that the modal emission distributions differ from each other. A further step is taken here to see which mode definition would be identified as the most appropriate definition by utilizing HTBR technique. For each definition, three dummy variables are added to represent idle, acceleration, and deceleration mode. The regression trees are developed between emission data and these three dummy variables for each definition are shown in Figures 7-4 to 7-6. The sensitivity test results based on these regression trees for NO<sub>x</sub>, CO, and HC are summarized in Table 7-4.

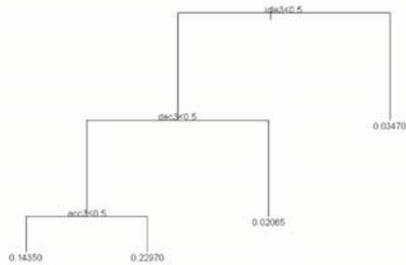
(a) Definition 1



(b) Definition 2



(c) Definition 3



(d) Definition 4

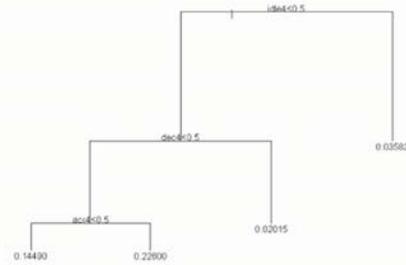
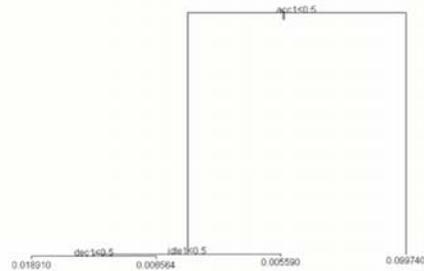
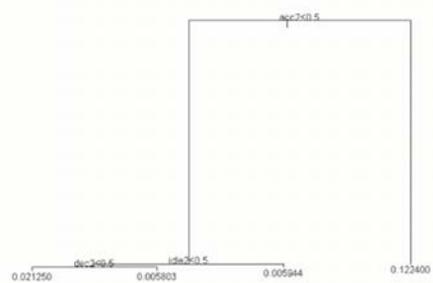


Figure 7-4 HTBR Regression Tree Result for NO<sub>x</sub> Emission Rate

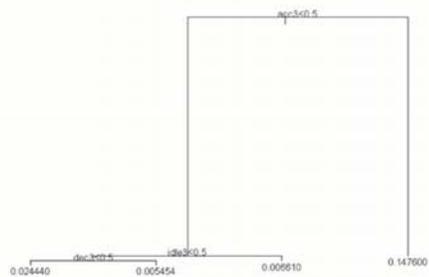
(a) Definition 1



(b) Definition 2



(c) Definition 3



(d) Definition 4

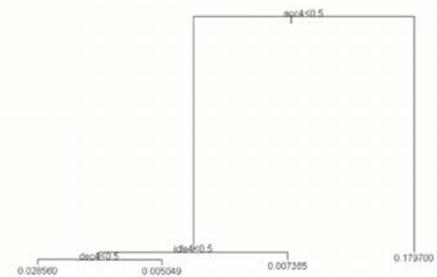


Figure 7-5 HTBR Regression Tree Result for CO Emission Rate

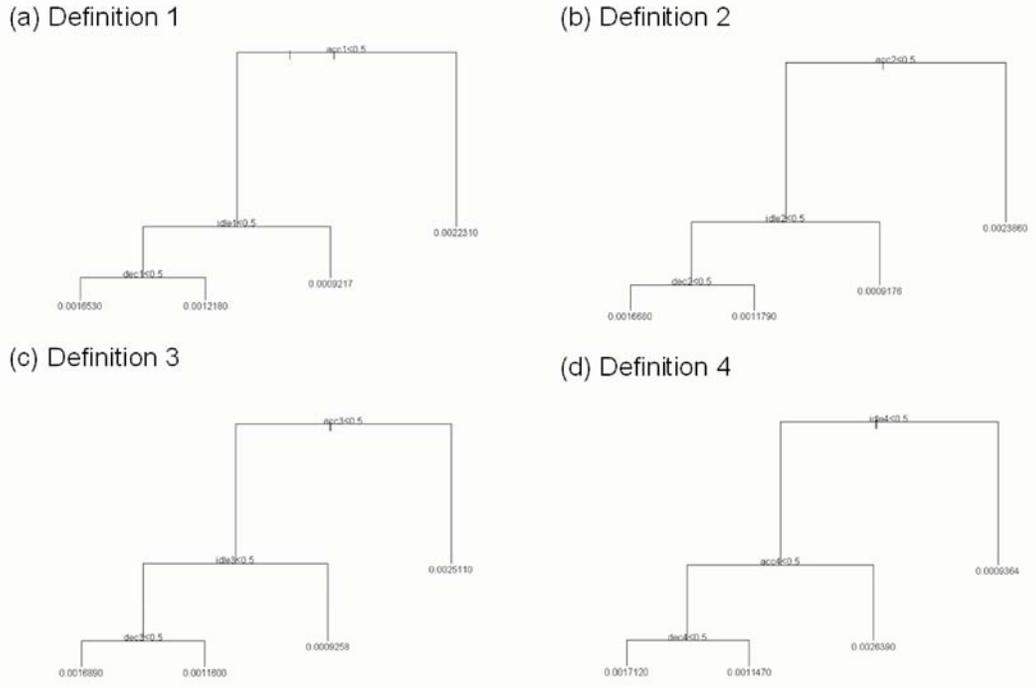


Figure 7-6 HTBR Regression Tree Result for HC Emission Rate

Table 7-4 Sensitivity Test Results for Four Mode Definition

NO <sub>x</sub>	Mode	Number	Deviance	Mean ER	Residual Mean Deviance
		105976	1435.00	0.10680	
Definition 1					0.006967 = 738.3 / 106000
	Idle	29541	11.04	0.03235	
	Acceleration	25931	320.90	0.22480	
	Deceleration	22242	41.32	0.02671	
	Cruise	28262	365.10	0.13930	
Definition 2					0.007658 = 811.5/106000
	Idle	31064	16.05	0.03342	
	Acceleration	18894	206.50	0.23110	
	Deceleration	16644	21.14	0.02214	
	Cruise	39374	567.80	0.14070	
Definition 3					0.00856 = 907.1 / 106000
	Idle	32010	23.07	0.03470	
	Acceleration	13417	130.50	0.2297	
	Deceleration	12768	14.27	0.02065	
	Cruise	47781	739.30	0.14350	

<b>NO<sub>x</sub></b>	<b>Mode</b>	<b>Number</b>	<b>Deviance</b>	<b>Mean ER</b>	<b>Residual Mean Deviance</b>
Definition 4					0.009397 = 995.8 / 106000
	Idle	32717	30.240	0.03583	
	Acceleration	8719	77.150	0.22600	
	Deceleration	9452	9.191	0.02015	
	Cruise	55088	879.200	0.14490	
CO					
		105765	771.300	0.032370	
Definition 1					0.005795 = 612.9 / 105800
	Idle	29287	2.166	0.005590	
	Acceleration	25866	559.400	0.099740	
	Deceleration	22456	3.903	0.006564	
	Cruise	28156	47.380	0.018910	
Definition 2					0.005486283 = 580.2 / 105800
	Idle	30764	4.185	0.005944	
	Acceleration	18864	484.900	0.122400	
	Deceleration	16919	2.410	0.005803	
	Cruise	39218	88.710	0.021250	
Definition 3					0.005293 = 559.8 / 105800
	Idle	31691	9.131	0.006610	
	Acceleration	13402	410.100	0.147600	
	Deceleration	13035	1.861	0.005454	
	Cruise	47637	138.700	0.024440	
Definition 4					0.005239 = 554 / 105800
	Idle	32375	15.5200	0.007365	
	Acceleration	8712	339.1000	0.179700	
	Deceleration	9681	0.7047	0.005049	
	Cruise	54997	198.7000	0.028560	
HC					
		103405	0.40270	0.0014960	
Definition 1					3.648e-006 = 0.3772 / 103400
	Idle	28780	0.09337	0.0009217	
	Acceleration	25122	0.09143	0.0022310	
	Deceleration	22287	0.07644	0.0012180	
	Cruise	27216	0.11600	0.0016530	
Definition 2					3.629e-006 = 0.3752 / 103400
	Idle	30250	0.09492	0.0009176	
	Acceleration	18330	0.06668	0.0023860	
	Deceleration	16805	0.05355	0.0011790	
	Cruise	38020	0.16010	0.0016680	

NO <sub>x</sub>	Mode	Number	Deviance	Mean ER	Residual Mean Deviance
Definition 3					3.636e-006 = 0.376 / 103400
	Idle	31157	0.09651	0.0009258	
	Acceleration	12999	0.04355	0.0025110	
	Deceleration	12970	0.04256	0.0011600	
	Cruise	46279	0.19330	0.0016890	
Definition 4					3.656e-006 = 0.378 / 103400
	Idle	31849	0.09835	0.0009364	
	Acceleration	8443	0.02944	0.0026390	
	Deceleration	9613	0.03257	0.0011470	
	Cruise	53500	0.21760	0.0017120	

### 7.3 Conclusions

Comparison of modal average estimates shows that the average modal emission rates are statistically different from each other for three different pollutants. HTBR regression tree results demonstrate that all four definitions can work well to divide the database. Comparisons of residual mean deviance indicate that definition 1 has the smallest residual mean deviance for NO<sub>x</sub> (definition 4 for CO and definition 2 for HC). However, differences were small. At this time, it is difficult to choose one definition for three pollutants based just on sensitivity analysis results in this chapter. The analysis results in this section indicate that driving mode definition could not be transferred directly from one research study to another research study. A better approach would be to test several different critical values and obtain the most suitable definition instead of testing only one definition developed from other research. For this research, more analysis will be performed in the chapters that follow to develop the most suitable driving mode definitions.

## CHAPTER 8

### 8. IDLE MODE DEVELOPMENT

In Chapter 7, the concept of driving modes was introduced and several sensitivity tests (comparison of modal average estimates, comparison of HTBR regression tree results, and comparison of residual mean deviance) were performed for four different mode definitions. Based on sensitivity analysis results, it is difficult to choose one definition for three pollutants at this moment. More analysis will be performed next to develop the most suitable driving mode definition. This chapter will focus on developing the suitable definition for idle mode.

Theoretically, idle mode is usually defined as zero speed and zero acceleration. In real world data collection efforts, this definition must be refined due to the presence of speed measurement error. In this research, idle mode will be defined by speed and acceleration. The critical value could not be deduced directly from previous research. It is better to test several critical values statistically and identify the most suitable idle definition.

#### 8.1 Critical Value for Speed in Idle Mode

Three critical values were tested to get the appropriate critical value for speed in defining idle activity. Figures 8-1 to 8-3 illustrate engine power vs. emission rates for three pollutants for three critical speed values: 1 mph, 2.5 mph, and 5 mph. Figure 8-4 compares engine power distributions for these three critical values. Because engine power distributions for three pollutants exhibit similar patterns, only NO<sub>x</sub> emissions are shown in Figure 8-4. Tables 8-1 and 8-2 provide the engine power distribution for these three critical values in two ways: by number and percentage.

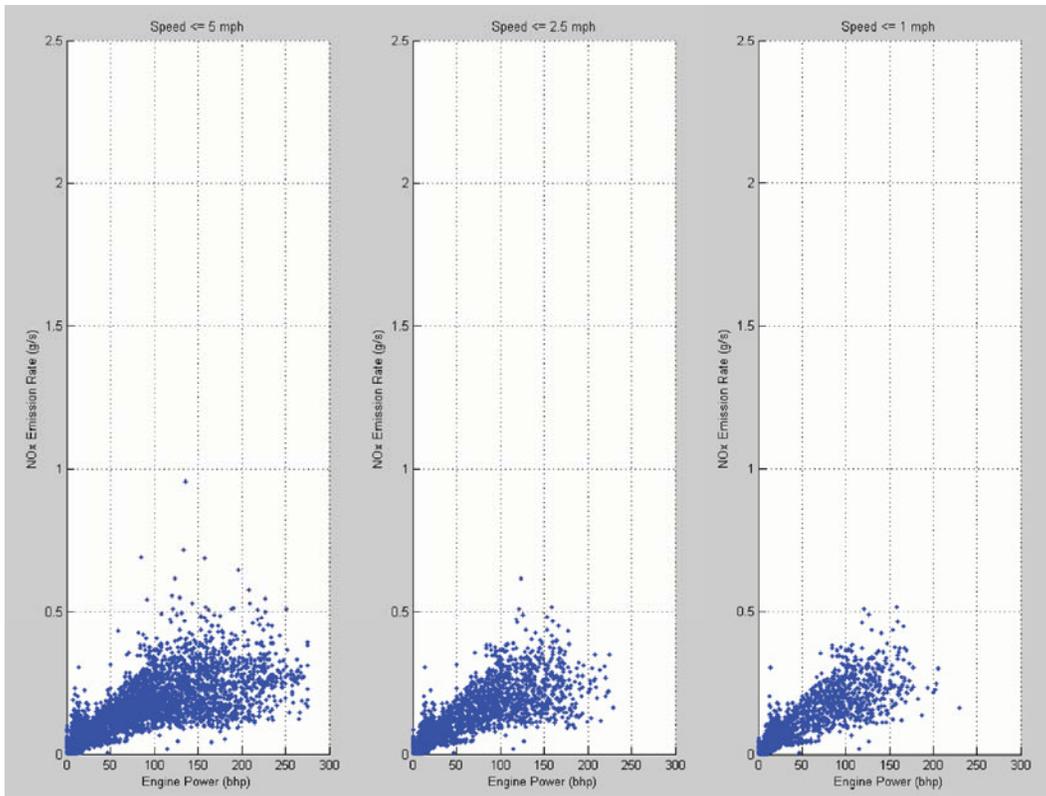


Figure 8-1 Engine Power vs. NO<sub>x</sub> Emission Rate for Three Critical Values

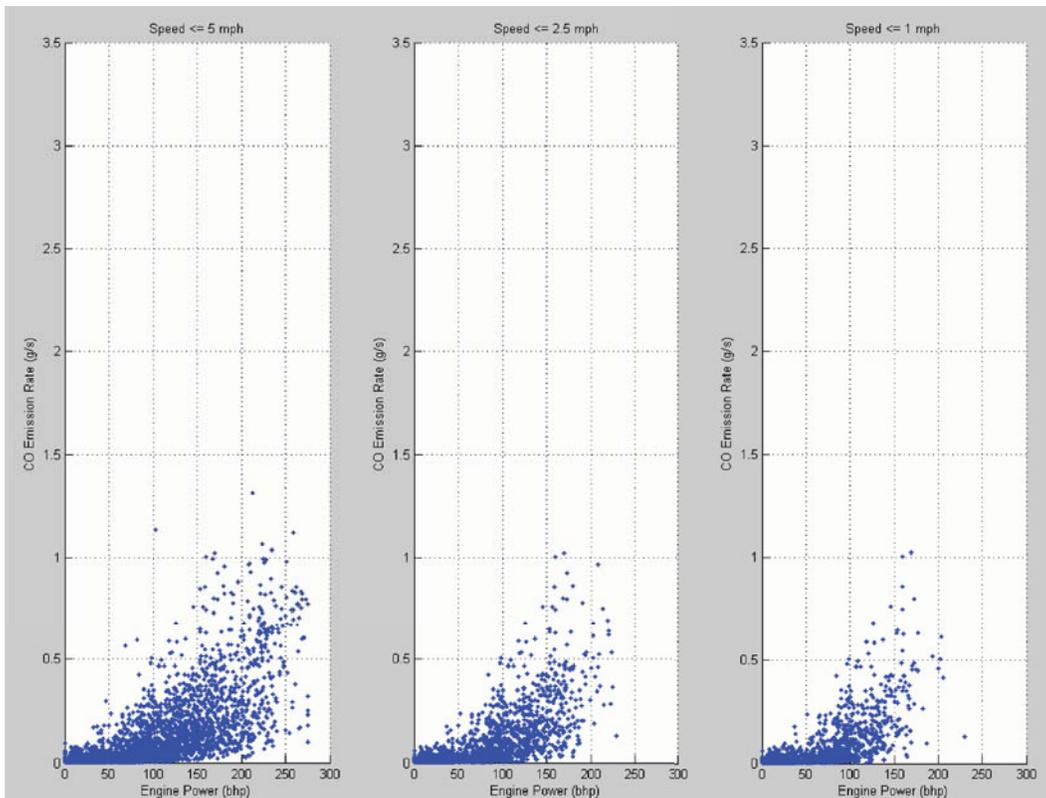


Figure 8-2 Engine Power vs. CO Emission Rate for Three Critical Values

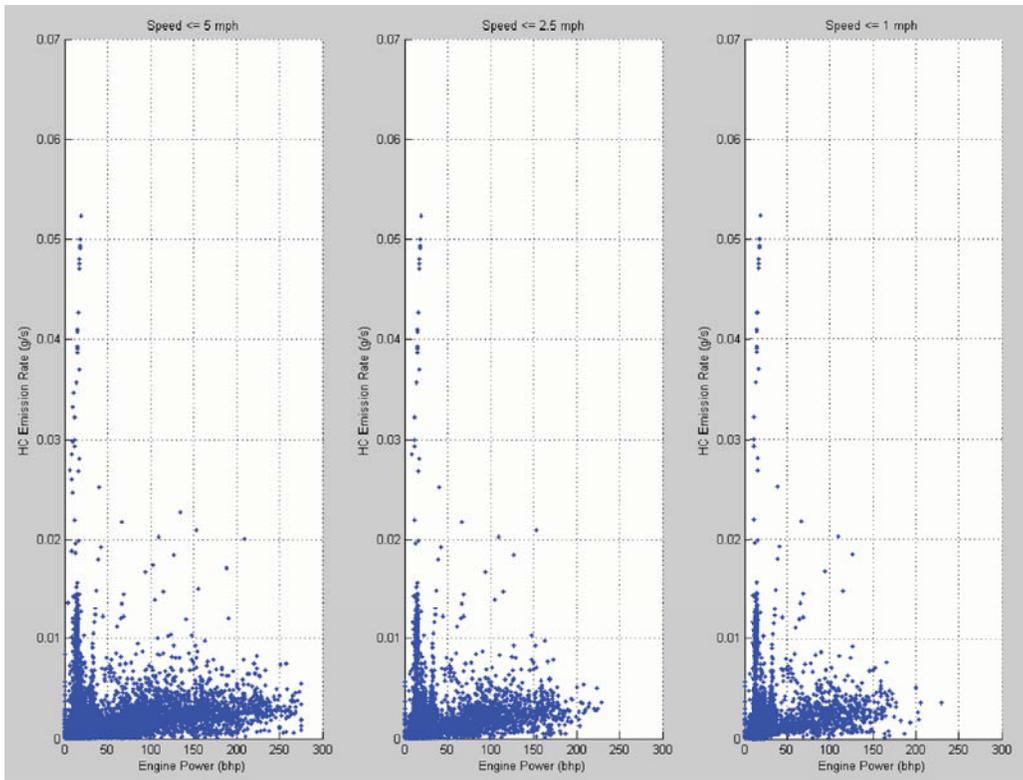


Figure 8-3 Engine Power vs. HC Emission Rate for Three Critical Values

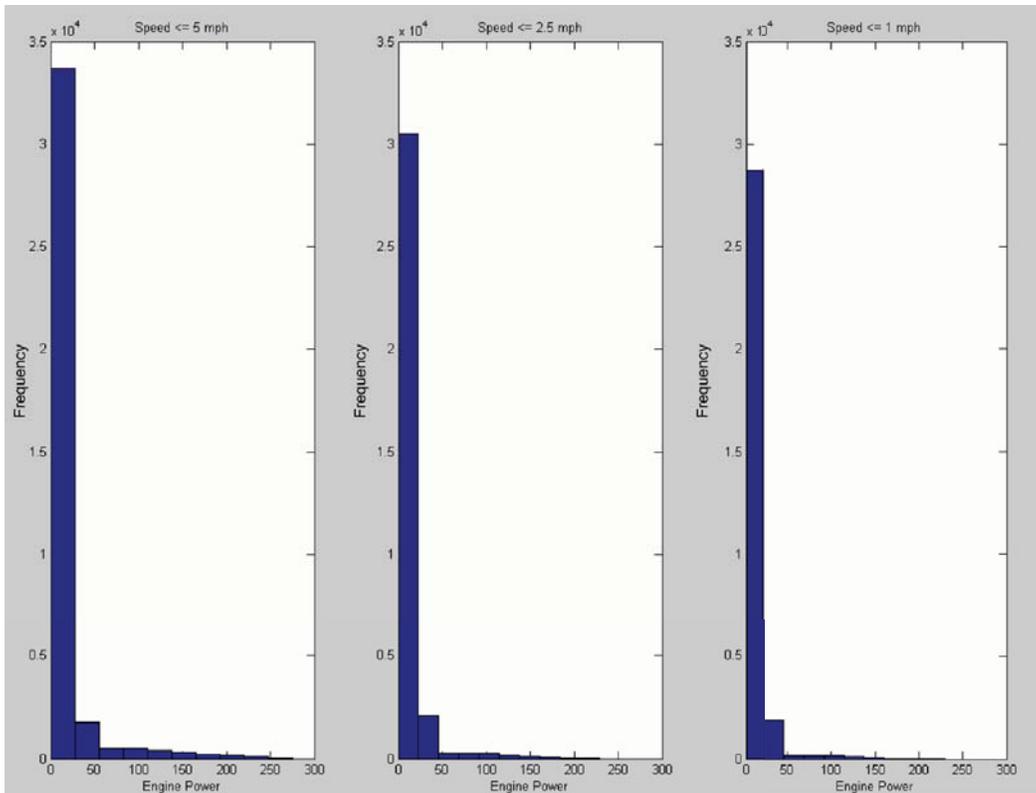


Figure 8-4 Engine Power Distribution for Three Critical Values based on NO<sub>x</sub> Emissions

Table 8-1 Engine Power Distribution for Three Critical Values for Three Pollutants

Speed	Pollutant	Engine Power (brake horsepower (bhp))					Total
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	
≤ 5 mph	NO <sub>x</sub>	31631	2272	1323	152	2348	37726
	CO	31258	2269	1316	149	2342	37334
	HC	30737	2264	1321	147	2284	36753
≤ 2.5mph	NO <sub>x</sub>	29222	2098	1196	83	1143	33742
	CO	28880	2096	1189	81	1139	33385
	HC	28373	2093	1194	80	1106	32846
≤ 1 mph	NO <sub>x</sub>	27516	2011	1100	51	700	31378
	CO	27217	2010	1093	51	699	31070
	HC	26713	2007	1099	48	680	30547

Table 8-2 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants

Speed	Pollutant	Engine Power (brake horsepower (bhp))					Total
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	
≤ 5 mph	NO <sub>x</sub>	83.84%	6.02%	3.51%	0.40%	6.22%	100%
	CO	83.73%	6.08%	3.52%	0.40%	6.27%	100%
	HC	83.63%	6.16%	3.59%	0.40%	6.21%	100%
≤ 2.5mph	NO <sub>x</sub>	86.60%	6.22%	3.54%	0.25%	3.39%	100%
	CO	86.51%	6.28%	3.56%	0.24%	3.41%	100%
	HC	86.38%	6.37%	3.64%	0.24%	3.37%	100%
≤ 1 mph	NO <sub>x</sub>	87.69%	6.41%	3.51%	0.16%	2.23%	100%
	CO	87.60%	6.47%	3.52%	0.16%	2.25%	100%
	HC	87.45%	6.57%	3.60%	0.16%	2.23%	100%

Based on the analysis above, a critical value of 5 mph includes more data points with higher engine power (>50 bhp) than 2.5 mph and 1 mph. However, there is no large difference for engine power distributions between 2.5 mph and 1 mph. These two critical values for speed will be tested further with different acceleration values in the next section. The results will be used to make a final decision with regards to deceleration mode.

## 8.2 Critical Value for Acceleration in Idle Mode

After setting the critical value for speed, the next step is to determine a critical value for acceleration. In total, four options were tested.

- Option 1: speed ≤ 2.5 mph and absolute acceleration ≤ 2 mph/s

- Option 2: speed  $\leq 2.5$  mph and absolute acceleration  $\leq 1$  mph/s
- Option 3: speed  $\leq 1$  mph and absolute acceleration  $\leq 2$  mph/s
- Option 4: speed  $\leq 1$  mph and absolute acceleration  $\leq 1$  mph/s

Using the same method as outlined in the previous section, Figures 8-5 to 8-7 illustrate engine power vs. emission rates for three pollutants for four options above. Figure 8-8 compares engine power distribution for data falling into these four options. Because engine power distributions for three pollutants exhibit a similar pattern, only NO<sub>x</sub> emissions are shown in Figure 8-8. Tables 8-3 and 8-4 provide the engine power distribution for four options in two ways: by number and percentage.

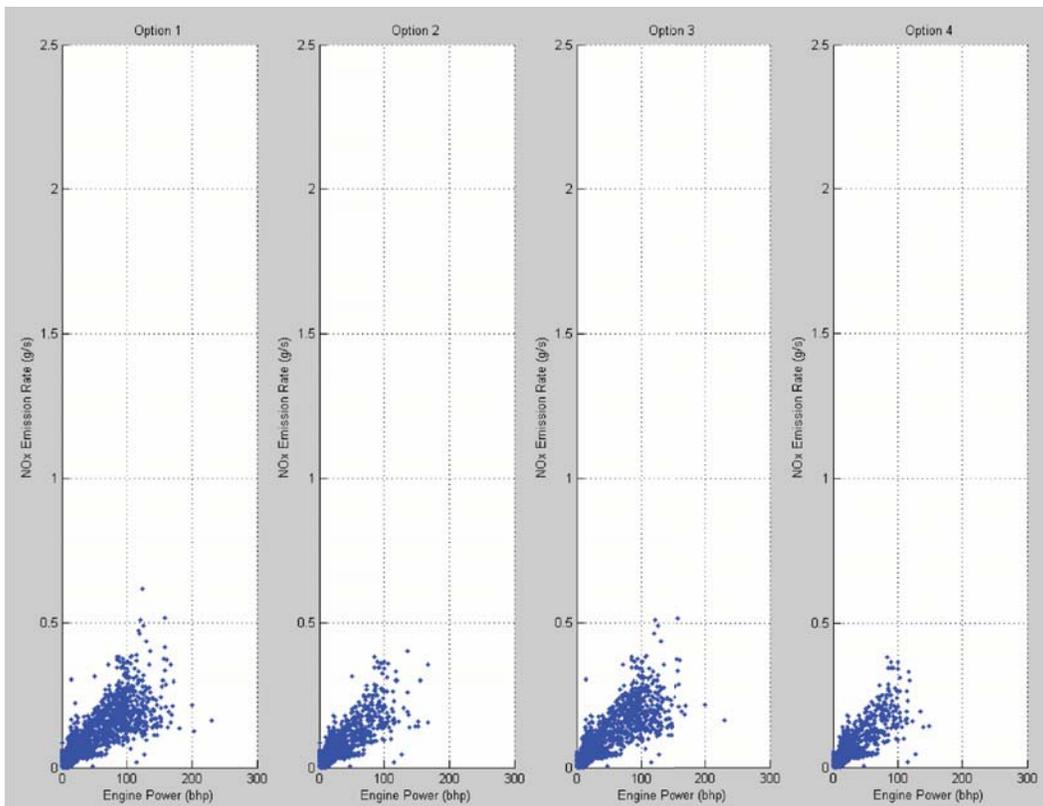


Figure 8-5 Engine Power vs. NO<sub>x</sub> Emission Rate for Four Options

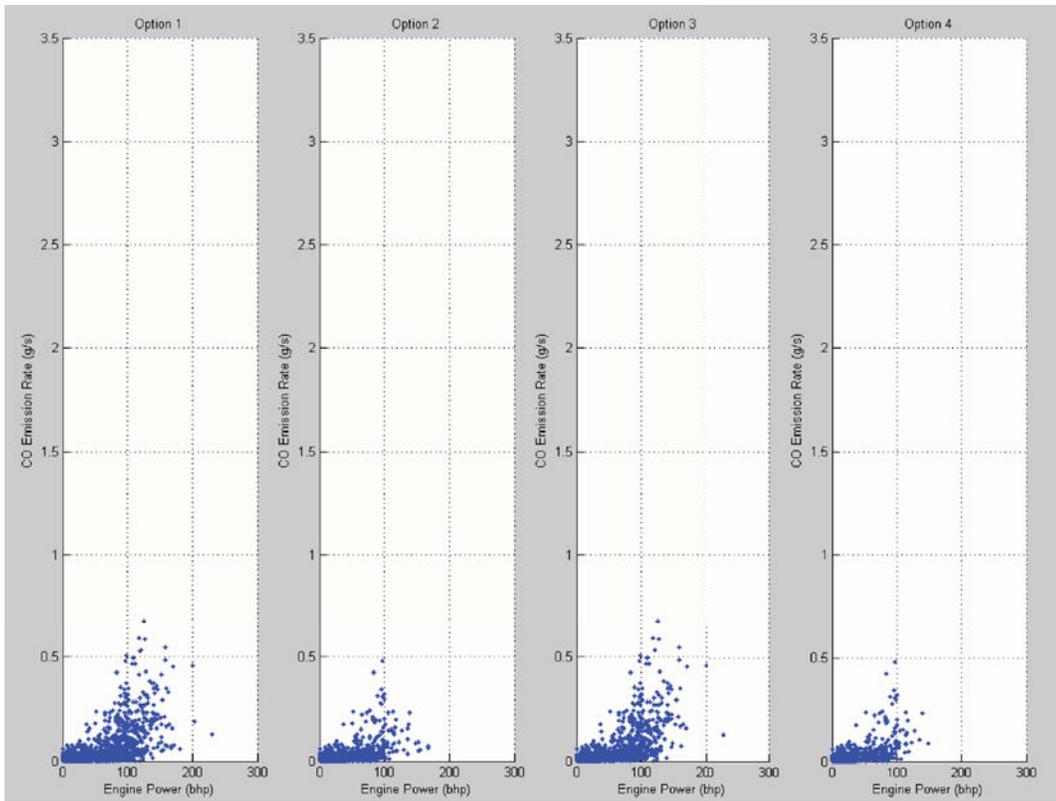


Figure 8-6 Engine Power vs. CO Emission Rate for Four Options

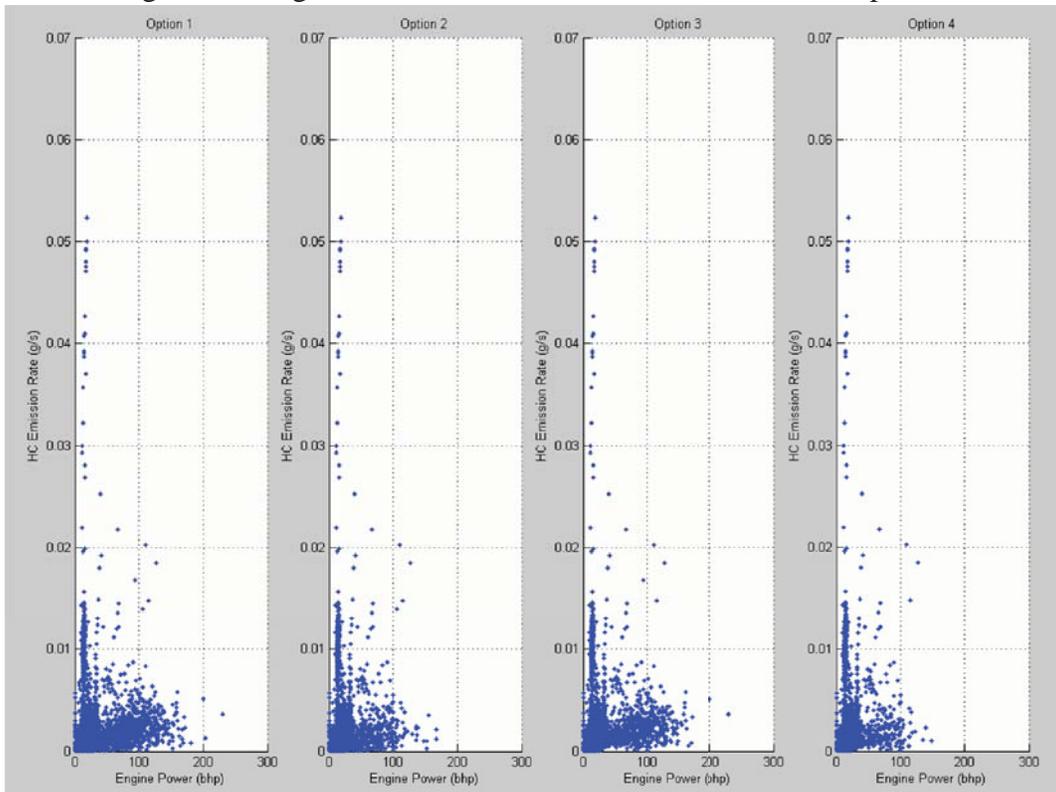


Figure 8-7 Engine Power vs. HC Emission Rate for Four Options

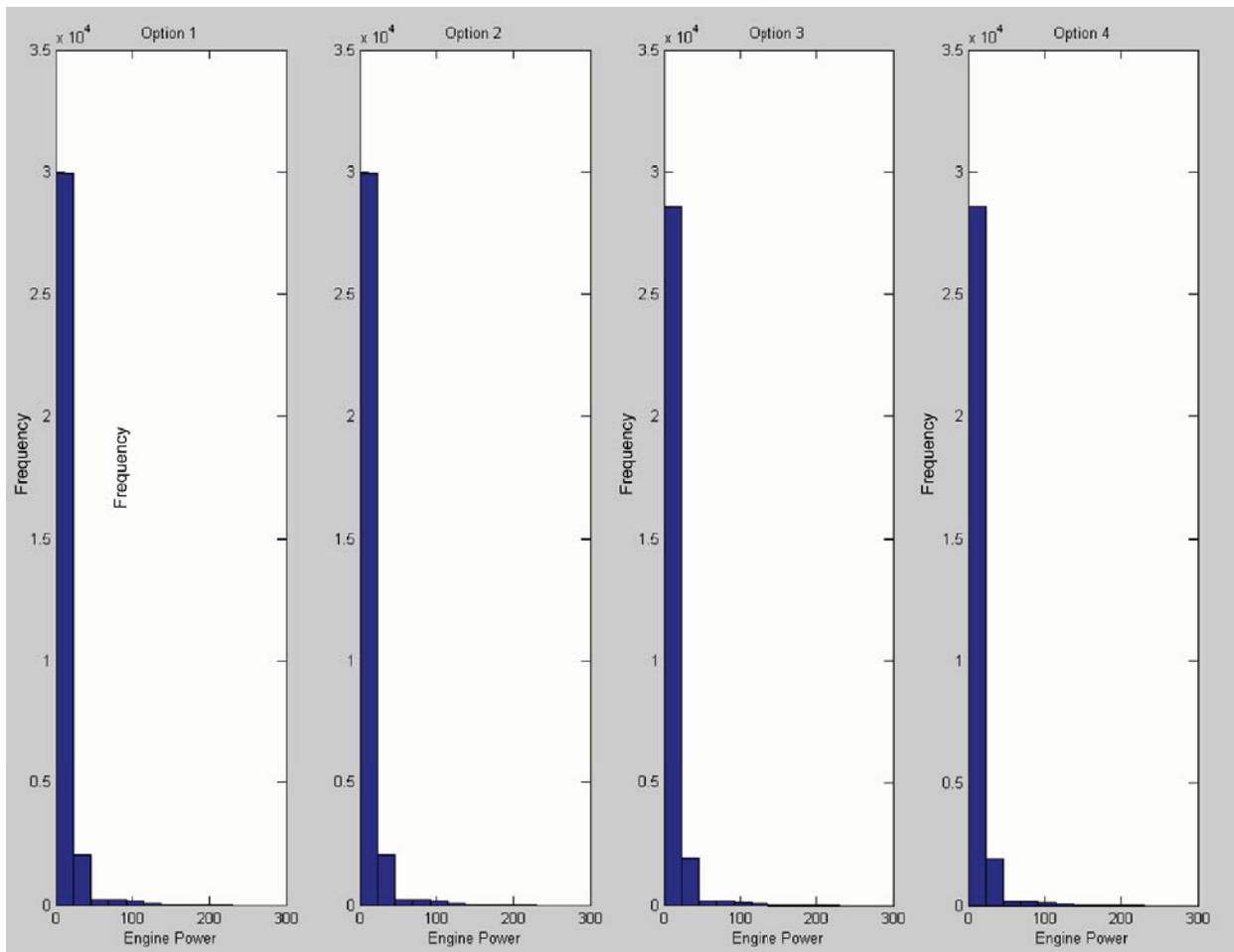


Figure 8-8 Engine Power Distribution for Four Options based on NO<sub>x</sub> Emission Rates

Table 8-3 Engine Power Distribution for Four Options for Three Pollutants

Pollutants		Engine Power (brake horsepower (bhp))					Total
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	
Option 1	NO <sub>x</sub>	28694	2075	1177	78	693	32717
	CO	28366	2073	1170	76	690	32375
	HC	27855	2070	1175	75	674	31849
Option 2	NO <sub>x</sub>	27571	2030	1120	53	290	31064
	CO	27284	2028	1114	51	287	30764
	HC	26771	2026	1119	51	283	30250
Option 3	NO <sub>x</sub>	27367	1999	1091	50	527	31034
	CO	27071	1998	1084	50	526	30729
	HC	26569	1995	1090	47	512	30213
Option 4	NO <sub>x</sub>	26719	1969	1057	34	205	29984
	CO	26446	1968	1051	34	204	29703
	HC	25944	1966	1056	32	198	29196

Table 8-4 Percentage of Engine Power Distribution for Three Critical Values for Three Pollutants

	Pollutants	Engine Power (brake horsepower (bhp))					Total
		[0 20)	[20 30)	[30 40)	[40 50)	≥ 50	
Option 1	NO <sub>x</sub>	87.70%	6.34%	3.60%	0.24%	2.12%	100.00%
	CO	87.62%	6.40%	3.61%	0.23%	2.13%	100.00%
	HC	87.46%	6.50%	3.69%	0.24%	2.12%	100.00%
Option 2	NO <sub>x</sub>	88.76%	6.53%	3.61%	0.17%	0.93%	100.00%
	CO	88.69%	6.59%	3.62%	0.17%	0.93%	100.00%
	HC	88.50%	6.70%	3.70%	0.17%	0.94%	100.00%
Option 3	NO <sub>x</sub>	88.18%	6.44%	3.52%	0.16%	1.70%	100.00%
	CO	88.10%	6.50%	3.53%	0.16%	1.71%	100.00%
	HC	87.94%	6.60%	3.61%	0.16%	1.69%	100.00%
Option 4	NO <sub>x</sub>	89.11%	6.57%	3.53%	0.11%	0.68%	100.00%
	CO	89.03%	6.63%	3.54%	0.11%	0.69%	100.00%
	HC	88.86%	6.73%	3.62%	0.11%	0.68%	100.00%

Based on the above analysis, data falling into option 2 and option 4 contain fewer data points with higher engine power (>50 bhp) than data falling into option 1 and option 3. But a large difference is not observed in the engine power distribution for data falling into option 2 and option 4. Based upon these results, the idle mode is defined as speed  $\leq 2.5$  mph and absolute acceleration  $\leq 1$  mph/s.

### 8.3 Emission Rate Distribution by Bus in Idle Mode

After defining “speed  $\leq 2.5$  mph and absolute acceleration  $\leq 1$  mph/s” as idle mode, emission rate histograms for each of the three pollutants for idle operations are presented in Figure 8-9. Figure 8-9 shows significant skewness for all three pollutants for idle mode. Inter-bus response variability for idle mode operations is illustrated in Figures 8-10 to 8-12 using median and mean of NO<sub>x</sub>, CO, and HC emission rates. Table 8-5 presents the same information in tabular form. The difference between median and mean is also an indicator of skewness.

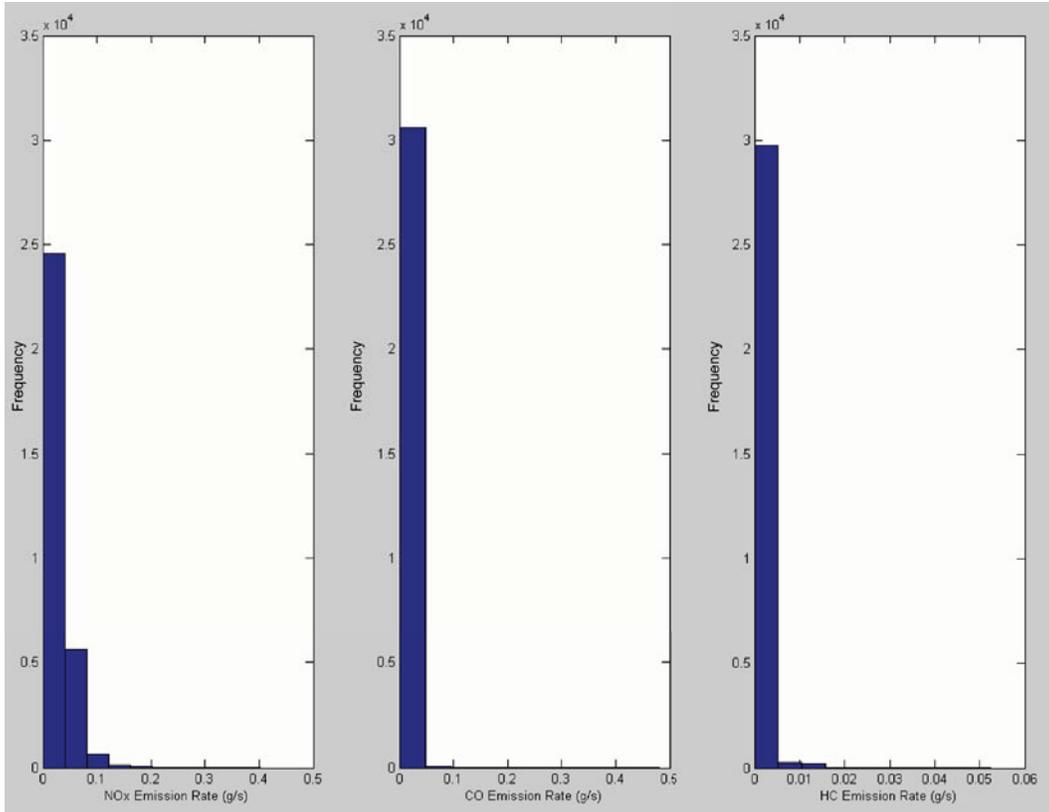


Figure 8-9 Histograms of Three Pollutants for Idle Mode

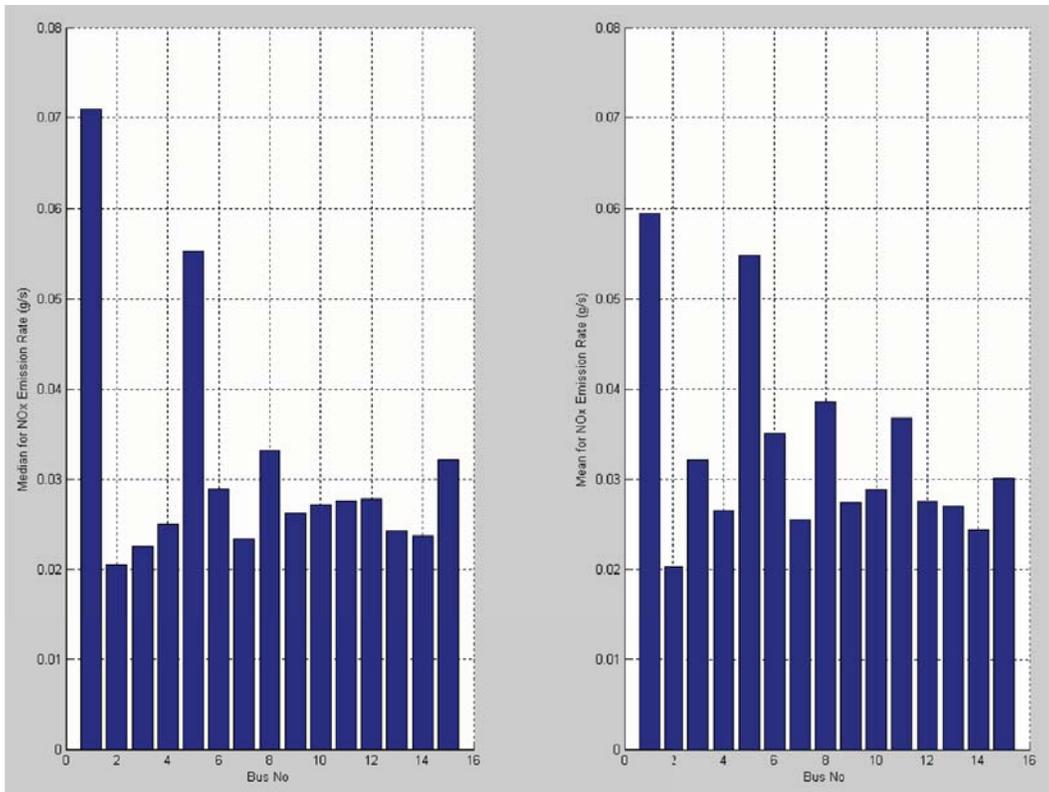


Figure 8-10 Median and Mean of  $\text{NO}_x$  Emission Rates in Idle Mode by Bus

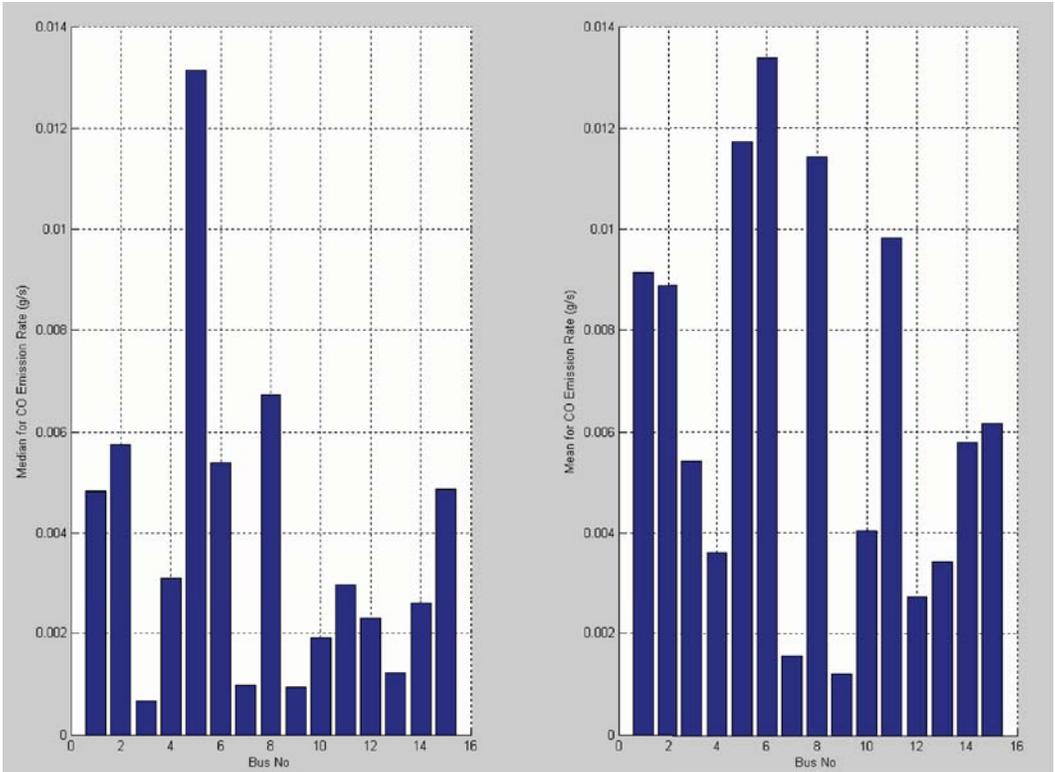


Figure 8-11 Median and Mean of CO Emission Rates in Idle Mode by Bus

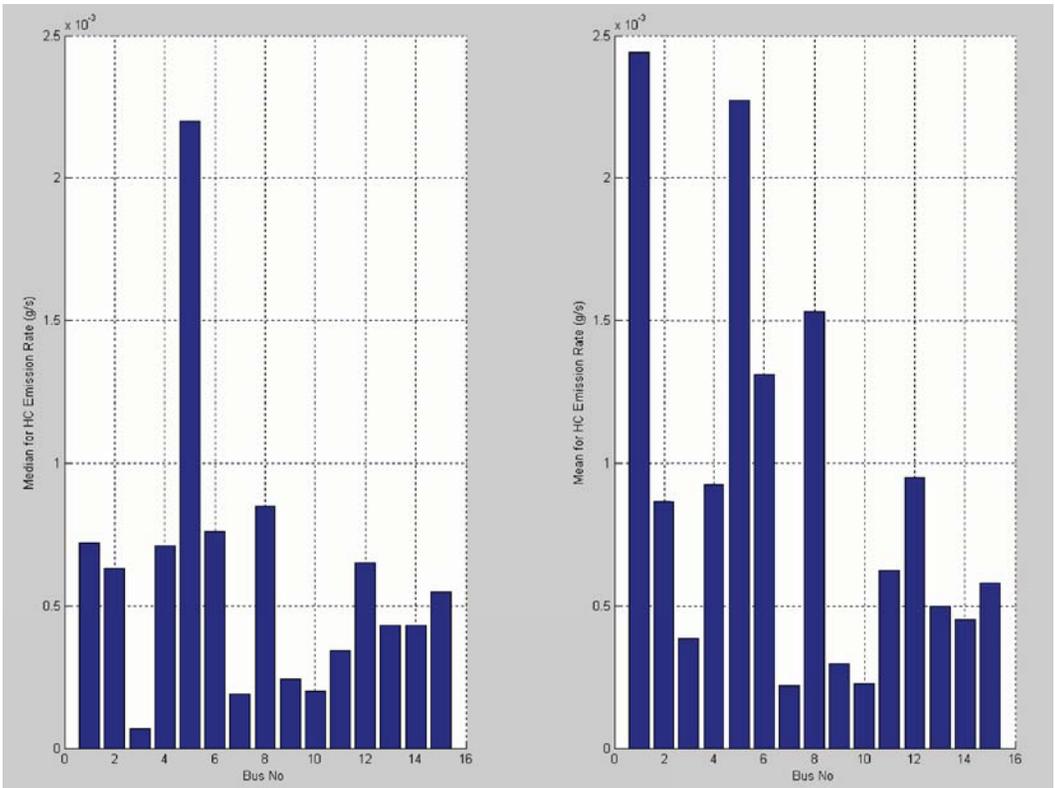


Figure 8-12 Median and Mean of HC Emission Rates in Idle Mode by Bus

Table 8-5 Median, and Mean of Three Pollutants in Idle Mode by Bus

Bus ID	NO <sub>x</sub>		CO		HC	
	Median	Mean	Median	Mean	Median	Mean
Bus 360	0.071020	0.059444	0.004830	0.009145	0.00072	0.002441
Bus 361	0.020455	0.020216	0.005740	0.008895	0.00063	0.000865
Bus 363	0.022555	0.032140	0.000670	0.005408	0.00007	0.000385
Bus 364	0.025050	0.026480	0.003110	0.003601	0.00071	0.000927
Bus 372	0.055210	0.054766	0.013150	0.011739	0.00220	0.002272
Bus 375	0.028880	0.035050	0.005390	0.013385	0.00076	0.001311
Bus 377	0.023370	0.025393	0.000960	0.001572	0.00019	0.000219
Bus 379	0.033210	0.038500	0.006730	0.011425	0.00085	0.001531
Bus 380	0.026200	0.027371	0.000930	0.001218	0.00024	0.000298
Bus 381	0.027115	0.028768	0.001915	0.004044	0.00020	0.000228
Bus 382	0.027605	0.036734	0.002980	0.009836	0.00034	0.000624
Bus 383	0.027790	0.027520	0.002290	0.002736	0.00065	0.000950
Bus 384	0.024210	0.026982	0.001205	0.003428	0.00043	0.000498
Bus 385	0.023750	0.024339	0.002590	0.005782	0.00043	0.000453
Bus 386	0.032140	0.030031	0.004860	0.006155	0.00055	0.000579

Figures 8-10 to 8-12 and Table 8-5 illustrate that bus 372 has the largest median and the second largest mean for CO and HC emissions, and the second largest median and the second largest mean for NO<sub>x</sub> emissions. The activity of bus 372 in terms of distribution of engine power by bus was compared to that of other buses in an effort to identify why the emission rates were significantly higher than for other buses. Table 8-6 and Figure 8-13 show that bus 372 has higher min (2<sup>nd</sup>), 1<sup>st</sup> quartile (2<sup>nd</sup>), median (1<sup>st</sup>), and 3<sup>rd</sup> quartile (2<sup>nd</sup>) engine power compared to the other 14 buses. Engine power in idle mode may include cooling fan, air compressor, air conditioner, and alternator loads (Clark et al. 2005). Considering test buses and engines are similar in many ways, this difference might be caused by variability across the engines, or may be associated with unrecorded air conditioner use. In analyzing the database, the modeler could not identify a contribution of air conditioner to engine power in idle mode. So, model development will include these data but readers should be cautioned that the noted variability is an indication that significant numbers of vehicles may need to be tested in the future if such inter-engine differences are significant in the fleet. In addition, the role of air conditioning usage on engine load in transit buses warrants additional future research.

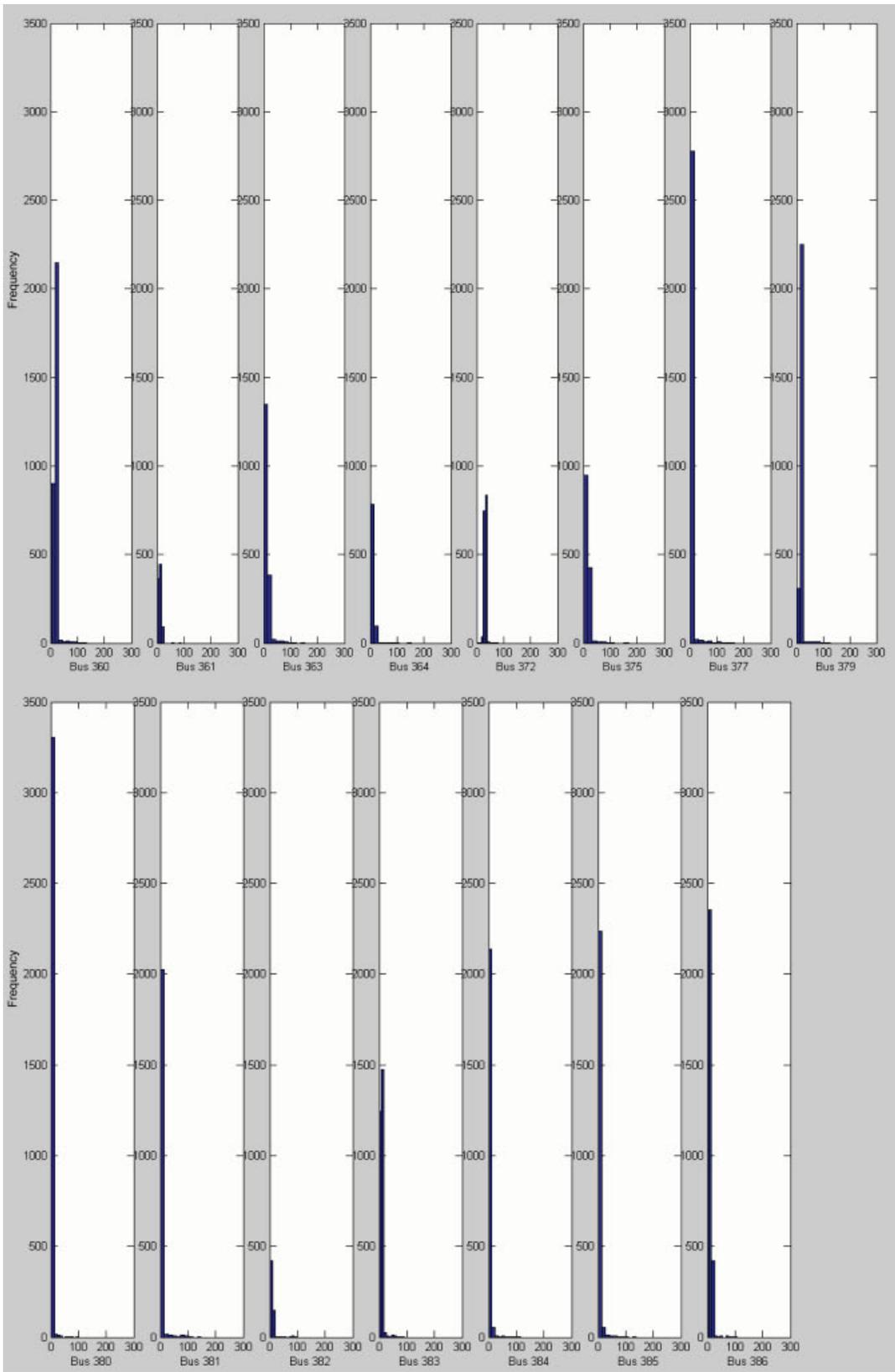


Figure 8-13 Histograms of Engine Power in Idle Mode by Bus

Table 8-6 Engine Power Distribution in Idle Mode by Bus

Bus ID	Min	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile	Max
Bus 360	3.92	15.36	18.7	19.83	135.43
Bus 361	0	5.35	12.52	13.83	89.47
Bus 363	0	13.1	13.34	15.16	152.94
Bus 364	0	13.18	13.85	14.99	154.51
Bus 372	0	26.44	31.84	33.10	79.08
Bus 375	0	12.52	13.81	18.08	167.72
Bus 377	0	8.5	9.17	9.85	166.86
Bus 379	0	15.86	17.15	19.42	126.64
Bus 380	2.67	7.85	8.49	9.17	100.99
Bus 381	0	8.7	10.49	11.17	148.28
Bus 382	0	7.35	8.52	13.89	99.04
Bus 383	0	7.16	10.03	12.5	91.86
Bus 384	0	6.01	7.34	8.51	117.39
Bus 385	0	4.53	7.19	8.51	139.05
Bus 386	4.68	9.18	13.33	14.46	105.44

## 8.4 Discussions

### 8.4.1 High HC Emissions

Figure 8-7 shows that there are some high HC emissions in idle mode. Based on definitions of “speed  $\leq 2.5$  mph and absolute acceleration  $\leq 1$  mph/s”, 388/30250=1.28% of data points in idle mode for HC are high emissions. These high emissions were noted in the HC emissions data, not in NO<sub>x</sub> and CO. All high HC emissions have been coded as high-idle 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 specific operating parameters related to these high-idle emissions were identified.

On the other hand, regression tree analysis results by bus and trip are presented in Figure 8-14. The left figure shows that these high HC emissions occurred in bus 360 and 372 while the right figure shows that these high HC emissions happened in bus 360 trip 4 and bus 372 trip 1. Even for HC emissions, Figure 8-14 shows that these high emissions are not a common situation in idle mode. There are 1529 idle segments in total for 15 buses, but most of these high HC emissions came just from three idle segments. These three idle segments are: bus 360 trip 4 idle

segment 1 (130 seconds), bus 360 trip 4 idle segment 38 (516 seconds) and bus 372 trip 1 idle segment 1 (500 seconds). More specifically, bus 360 trip 4 idle segment 1 contains 102 high HC emissions, bus 360 trip 4 idle segment 38 contains 264 high HC emissions, while bus 372 trip 1 idle slots contain 13 high HC emissions. Figures 8-15 to 8-17 illustrate time series plots for HC for these three idle segments while vehicle speed, engine speed, engine power, engine oil temperature, engine oil pressure, engine coolant temperature and ECM pressure are presented, too. These figures do not include  $\text{NO}_x$  and CO because  $\text{NO}_x$  and CO do not show such patterns as these three idle segments for HC. These three idle segments contain 379 high HC emissions in total. Thus about 98% of high emissions came from three idle segments only. Exclusion of these three idle segments based on all current information is difficult. The modeler prefers to keep these data since these outliers might reflect variability in the real world. However, future data collection efforts should seek to identify the causes of such events.

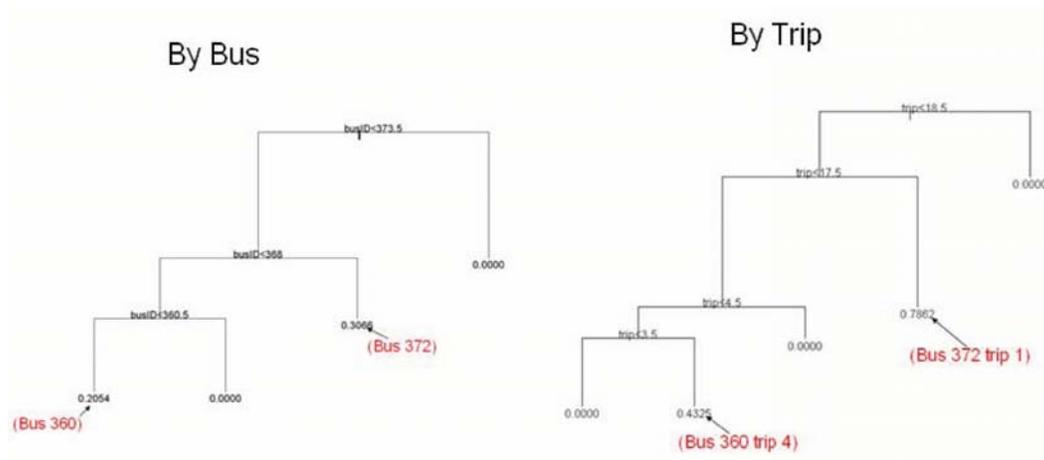


Figure 8-14 Tree Analysis Results for High HC Emission Rates by Bus and Trip

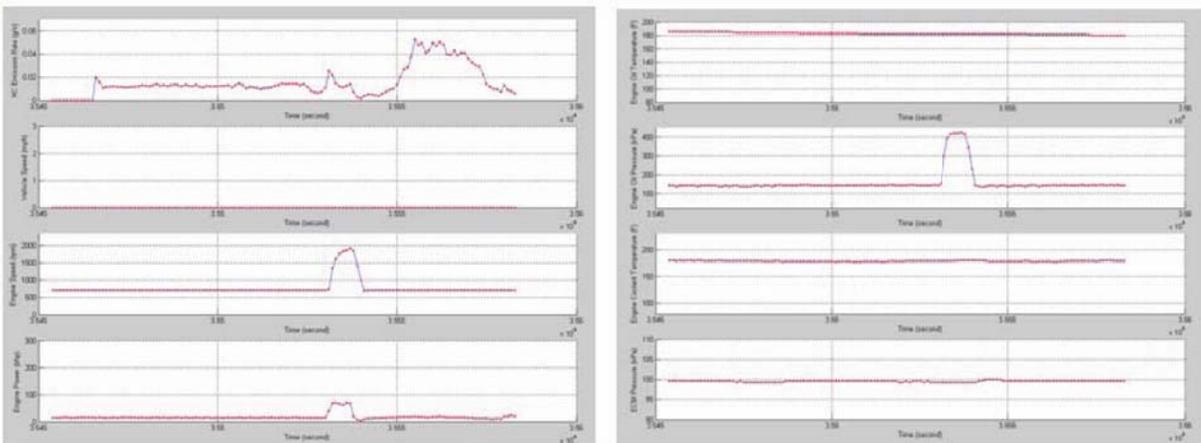


Figure 8-15 Time Series Plot for Bus 360 Trip 4 Idle Segment 1 (130 Seconds)

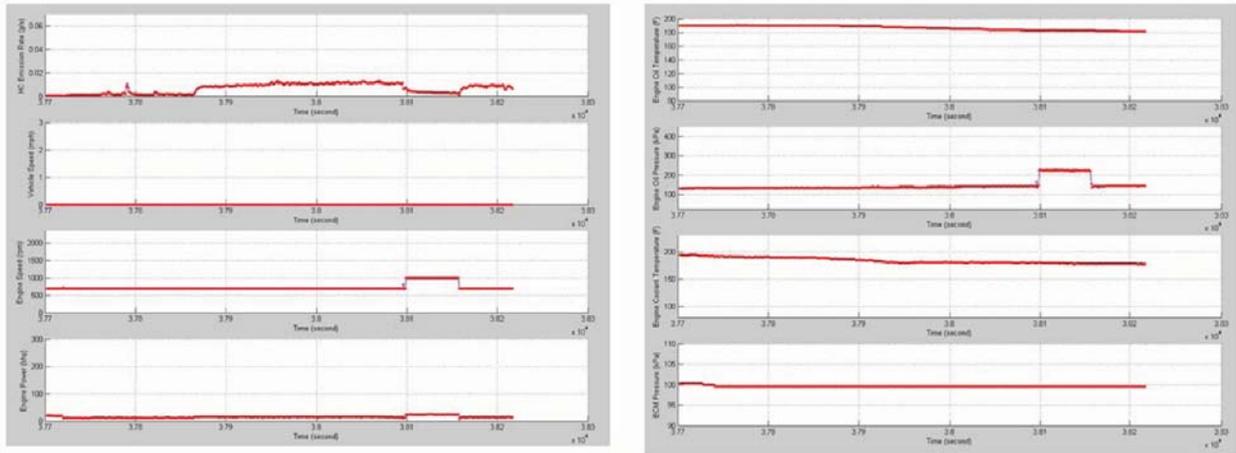


Figure 8-16 Time Series Plot for Bus 360 Trip 4 Idle Segment 38 (516 Seconds)

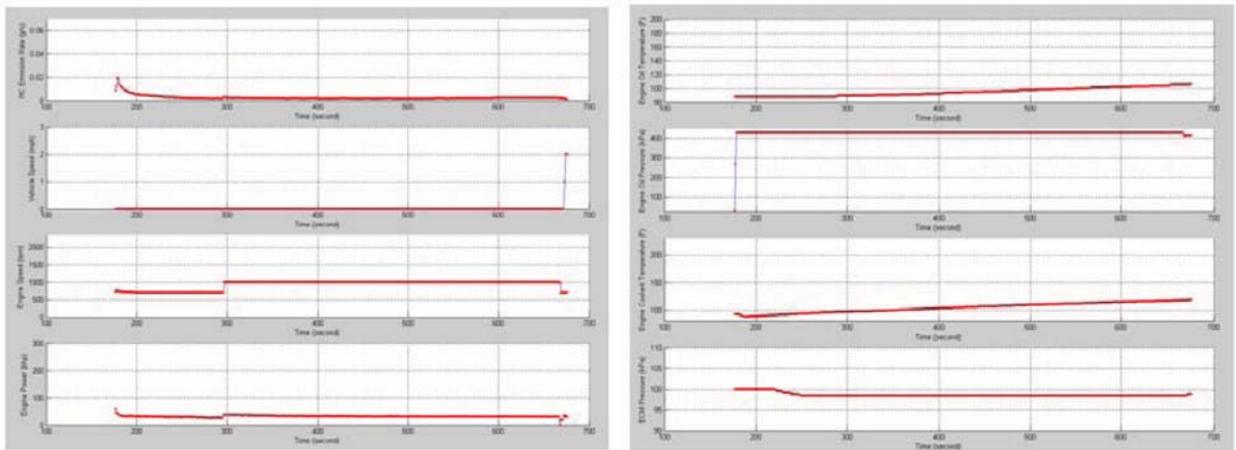


Figure 8-17 Time Series Plot for Bus 372 Trip 1 Idle Segment 1 (500 Seconds)

### 8.4.2 High Engine Operating Parameters

Figure 8-15 shows that engine speed once jumped to about 2000 rpm during bus 360 trip 4 idle segment 1, while corresponding engine power and engine oil pressure jumped, too. This jump lasted only 9 seconds. There are several reasons which might be responsible for this jump. Possibly bus 360 moved slowly from one location to another location while the GPS failed to detect the movement. Other explanations might be that the engine experienced a computer or sensor problem. This kind of jump, higher engine speeds (about 2000 rpm) accompanied by higher engine power and engine oil pressure in idle mode, did occur in the real world. The jump shown in Figure 8-16 was not such an occurrence since engine speed was only about 1000 rpm during that jump. After screening the whole dataset, another example of a jump is shown in Figure 8-18. The jump in bus 383 trip 1 idle segment 12 lasts 28 seconds. Since there are only two observations of such jumps in the whole database, there are not enough data to assess whether

they constitute a new mode. These observations might indicate that one should pay attention to slow movement during an idle segment. Since these two idle segments show some unusual activities, the modeler will retain them to avoid any bias in the results.

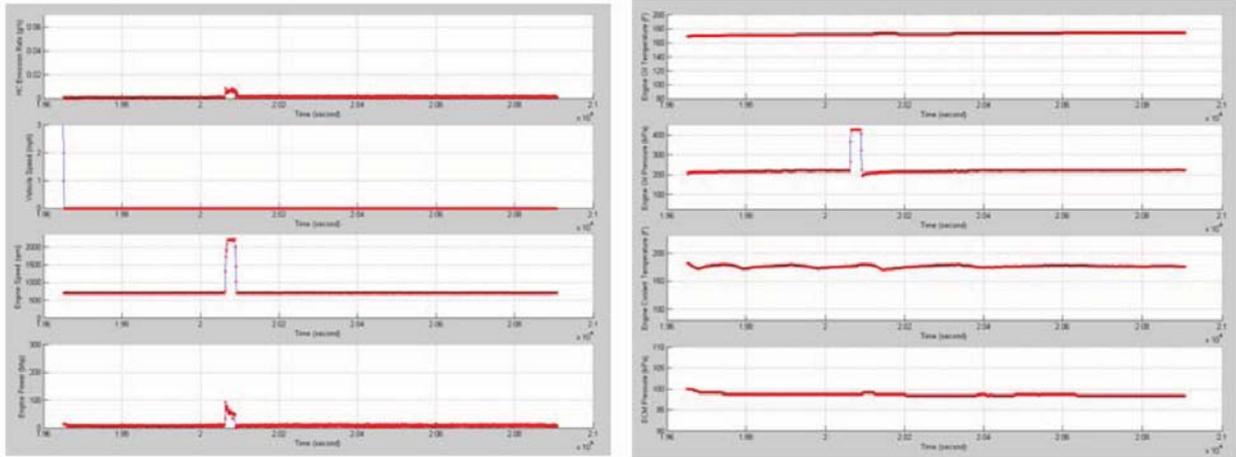


Figure 8-18 Time Series Plot for Bus 383 Trip 1 Idle Segment 12 (1258 Seconds)

### 8.5 Idle Emission Rates Estimation

Based on definition of “speed  $\leq 2.5$  mph and absolute acceleration  $\leq 1$  mph/s”, about 30% of available data are classified as idle mode. Usually, modelers estimate the idle emission rate by averaging all emission rates in idle mode. Although there are some data points with higher engine power ( $> 50$  bhp) in idle mode, about 90% of data in idle mode exhibit engine power between 0 and 20 bhp. After detailed analysis of all idle segments using time series plots, although some data may be incorrectly classified as the idle mode, no anomalies were noted. To avoid introducing any significant bias, a single idle emission rate is developed for each pollutant. When we treat all data as a whole and put them in the pool, the mean and confidence interval can reflect the distribution of emission rates in real world. Table 8-7 provides idle mode statistical analysis results for  $\text{NO}_x$ , CO, and HC.

Table 8-7 Idle Mode Statistical Analysis Results for NO<sub>x</sub>, CO, and HC

	NO <sub>x</sub>	CO	HC
minimum	0.00121	0.00002	0.00001
1 <sup>st</sup> Quartile	0.02201	0.00120	0.00026
mean	0.03342	0.00594	0.00092
median	0.02670	0.00293	0.00051
3 <sup>rd</sup> Quartile	0.03549	0.00554	0.00079
maximum	0.40259	0.48118	0.05232
skewness	4.45050	13.1840	11.6100
Total Number	31064	30764	30250

Due to the non-normality of emission rates, the median value (the value that divides observations into an upper and lower half) and the inter-quartile range (the range of values that includes the middle 50% of the observations) are the most appropriate for describing the distribution. The mean and skewness for the original data are presented in Table 8-8 as well. Although transformation for three pollutants already discussed based on the whole data set in Chapter 6, lambdas chosen by Box-Cox procedure for the whole data set and idle mode are different. Lambdas chosen by Box-Cox procedure for the whole data set are 0.22875 for NO<sub>x</sub>, -0.0648 for CO, 0.14631 for HC, while lambdas for idle mode are -0.19619 for NO<sub>x</sub>, -0.0625 for CO, 0.002875 for HC. At the same time, using transformation to estimate the mean and construct confidence intervals will create other problems. Therefore the modeler considers bootstrap, another class of general method, to obtain the estimation and construct confidence intervals.

The bootstrap is a procedure that involves choosing random samples with replacement from a data set and analyzing each sample the same way (Li 2004). To obtain the 95% confidence interval, the simple method is to take 2.5% and 97.5% percentile of the  $\beta$  replications  $T_1, T_2, \dots, T_\beta$  as the lower and upper bounds, respectively. The bootstrap function in this study will resample the emission data 1000 times and compute the mean, 2.5% and 97.5% percentile on each sample. Results are presented in Figure 8-20 and Table 8-8.

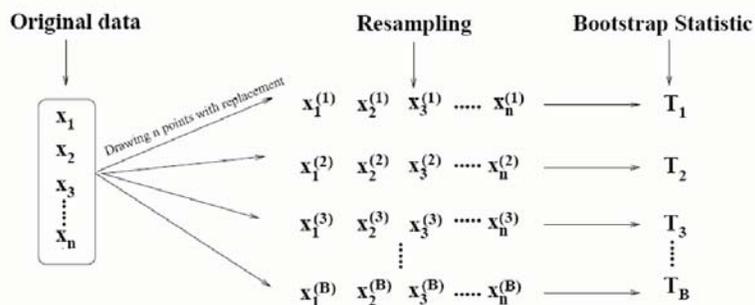


Figure 8-19 Graphical Illustration of Bootstrap (Adopted from Li 2004))

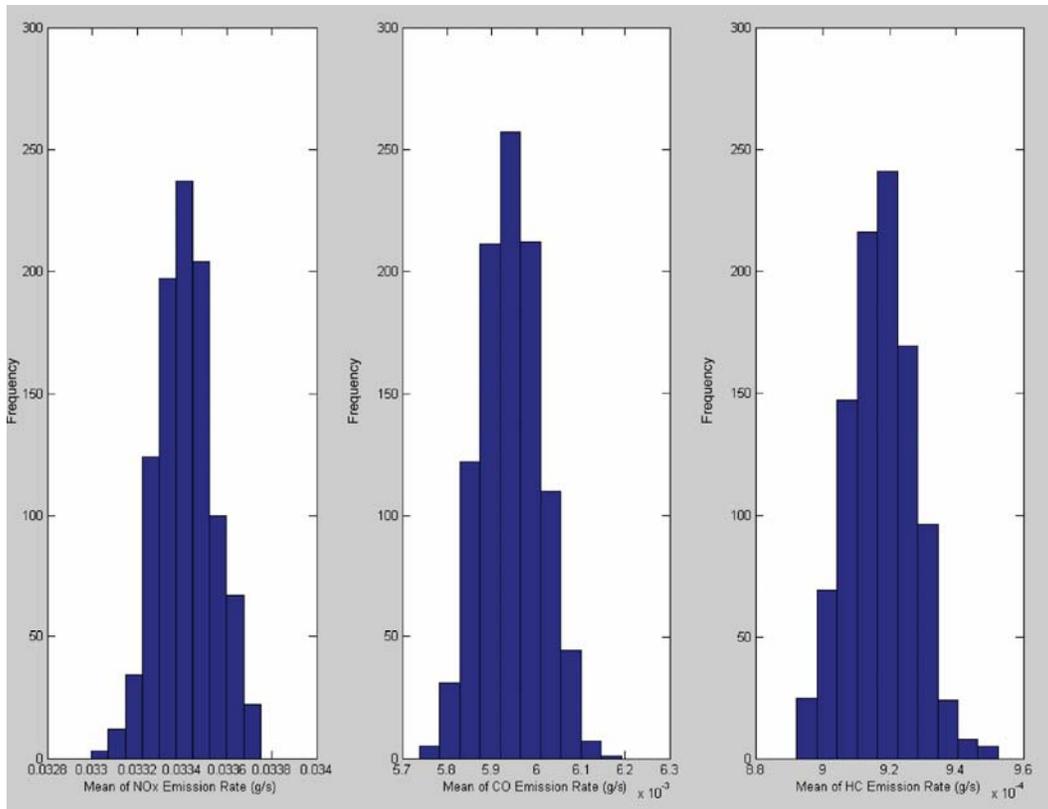


Figure 8-20 Bootstrap Results for Idle Emission Rate Estimation

Table 8-8 Idle Emission Rates Estimation and 95% Confidence Intervals Based on Bootstrap

		Average	2.5% Percentile	97.5% Percentile
NO <sub>x</sub>	Estimation	0.033415	0.010754	0.083266
	Confidence Interval	0.033162	0.010509	0.082279
		0.033669	0.010998	0.084252
CO	Estimation	0.0059439	0.00036116	0.028429
	Confidence Interval	0.0058184	0.00034446	0.028083
		0.0060693	0.00037775	0.028775
HC	Estimation	0.00091777	0.000059167	0.0037260
	Confidence Interval	0.00089742	0.000047572	0.0036412
		0.00093811	0.000070763	0.0038108

Based on table 8-9, the modeler recommends idle emission rates for NO<sub>x</sub> as 0.033415 g/s with 95% confidence interval (0.010754, 0.083266), CO as 0.0059439 g/s with 95% confi-

dence interval (0.00036116, 0.028429), HC as 0.00091777 g/s with 95% confidence interval (0.000059167, 0.0037260).

## 8.6 Conclusions and Further Considerations

In this research, idle mode is defined as “speed  $\leq$  2.5 mph and absolute acceleration  $\leq$  1 mph/s”. However the critical value could not be introduced from other research to this research directly. It is more appropriate to test several critical values and obtain the most suitable one instead of testing only one developed from other research.

Inter-bus variability analysis results indicate that bus 372 has the largest mean for NO<sub>x</sub>, CO, and HC emissions. Meanwhile, bus 372 has higher minimum (2<sup>nd</sup>), 1st Quartile (2<sup>nd</sup>), median (1<sup>st</sup>), and 3rd Quartile (2<sup>nd</sup>) engine power by comparison to the other 14 buses. Since test buses and engines are similar in most ways, this difference might be caused by variability of the engines or air conditioner usage. However, the contribution of the air conditioner to engine power in idle mode could not be identified in the database. Future research regarding the role of the air conditioner on engine power and emission rates in idle mode may be able to detect a difference.

Although some trips or some buses have higher mean and standard deviation than others, this kind of variability will decrease when all data in idle mode are treated as a whole. On the other hand, some elevated emissions events may simply reflect real world variability. Without additional evidence, modelers should treat all data as a whole instead of removing outliers and potentially biasing results.

There are two observations of an emissions jump that appears to be unrelated to engine speed, engine power, and engine oil temperature, in a single idle segment. The modeler first assumed that the bus moved too slowly from one location to another location for the GPS/ECM to detect the movement. Other explanations might be an engine computer problem or sensor problem. These two jumps might be evidence to support further research on slow movements during idle segments.

In summary, the modeler recommends idle emission rates for NO<sub>x</sub> as 0.033415 g/s with 95% confidence interval (0.010754, 0.083266), CO as 0.0059439 g/s with 95% confidence interval (0.00036116, 0.028429), HC as 0.00091777 g/s with 95% confidence interval (0.000059167, 0.0037260).

## 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  $\geq$  -2 mph/s & acceleration < -1 mph/s
- Option 3: acceleration  $\geq$  -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))					Total
		(0 20)	(20 30)	(30 40)	(40 50)	≥ 50	
Option 1	NO <sub>x</sub>	9322	94	16	5	15	9452
	CO	9558	89	15	4	15	9681
	HC	9483	94	16	5	15	9613
Option 2	NO <sub>x</sub>	6748	127	101	42	174	7192
	CO	6800	126	99	42	171	7238
	HC	6754	125	99	42	172	7192
Option 3	NO <sub>x</sub>	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))					Total
		(0 20)	(20 30)	(30 40)	(40 50)	≥ 50	
Option 1	NO <sub>x</sub>	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 <sub>x</sub>	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 <sub>x</sub>	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%

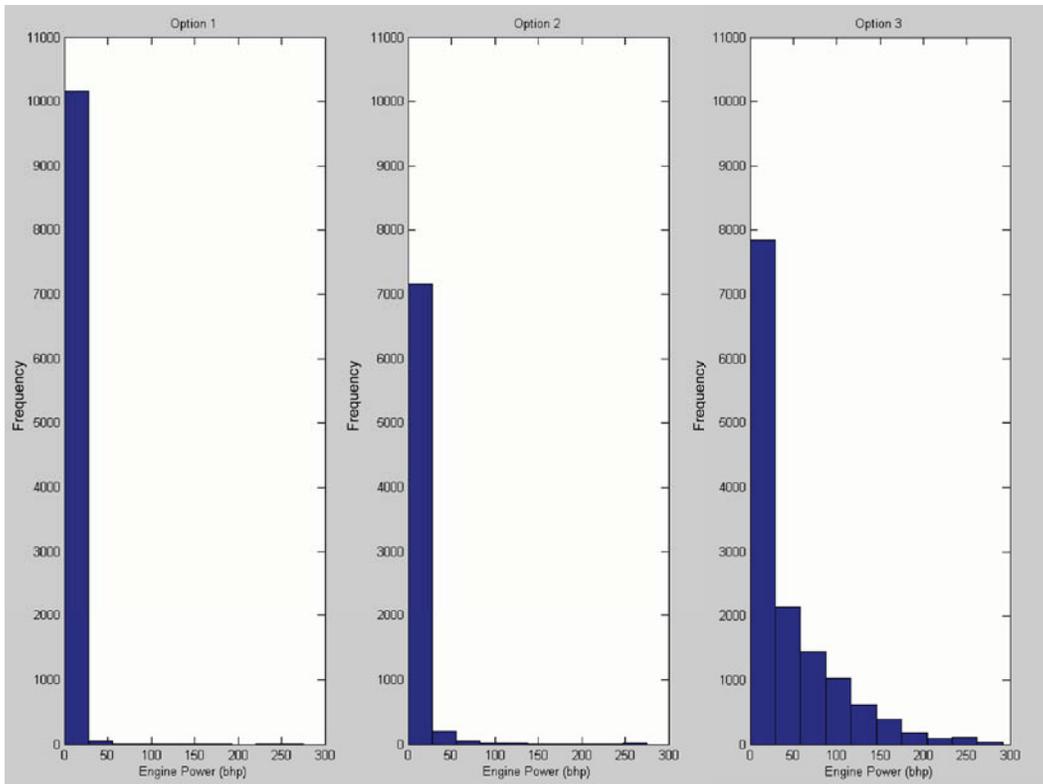


Figure 9-1 Engine Power Distribution for Three Options

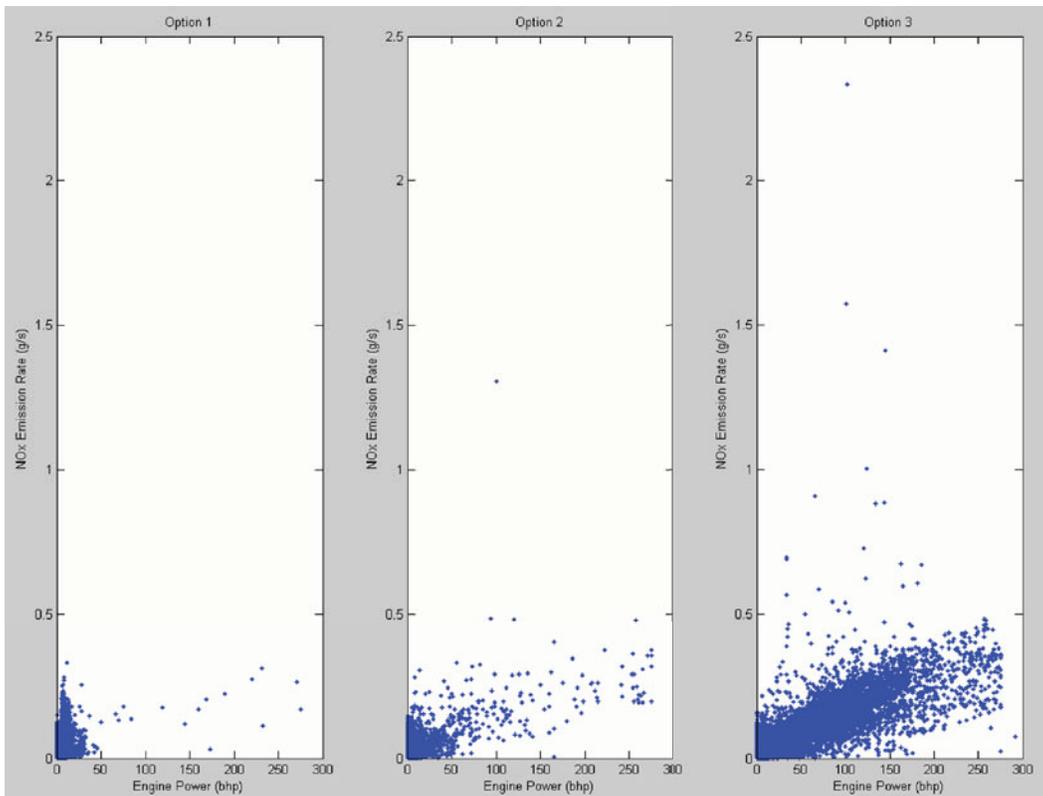


Figure 9-2 Engine Power vs.  $\text{NO}_x$  Emission Rate for Three Options

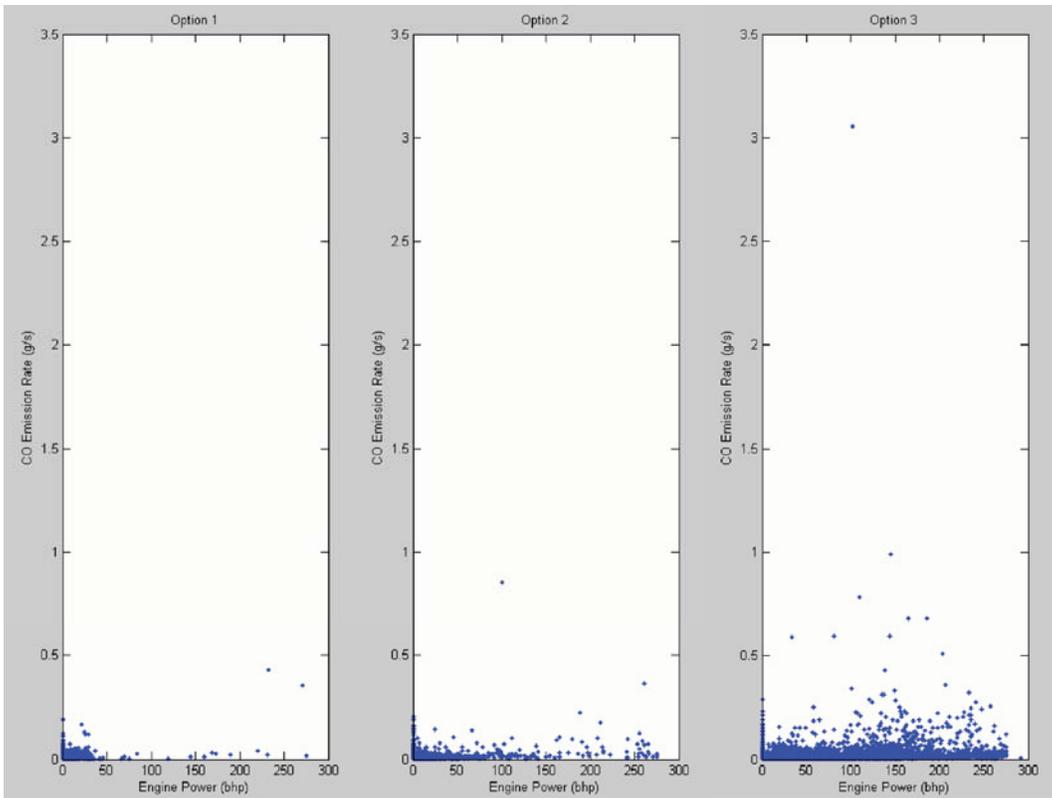


Figure 9-3 Engine Power vs. CO Emission Rate for Three Options

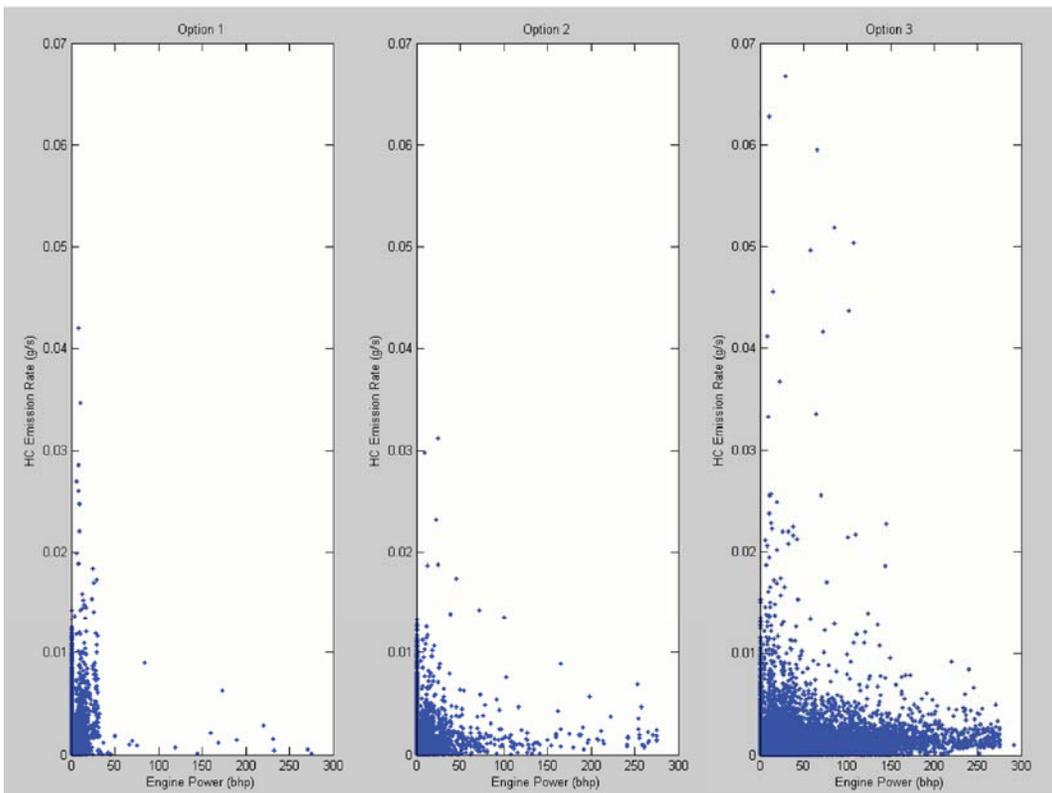


Figure 9-4 Engine Power vs. HC Emission Rate for Three Options

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<sub>x</sub>, 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.

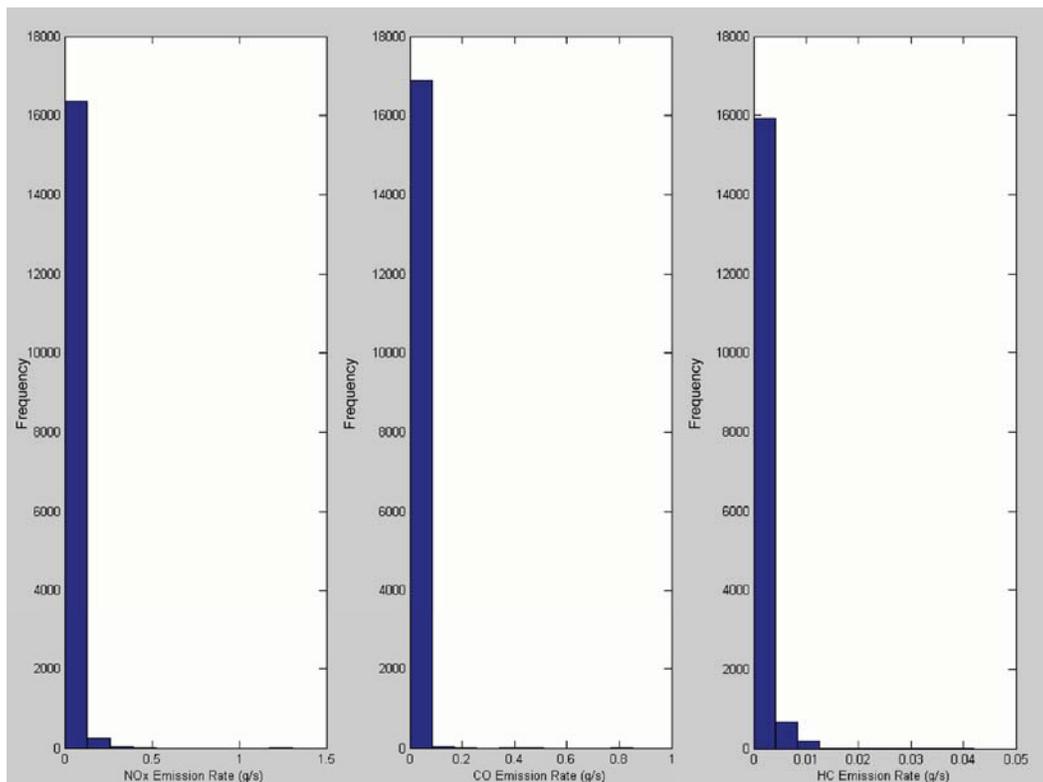


Figure 9-5 Histograms of Three Pollutants for Deceleration Mode

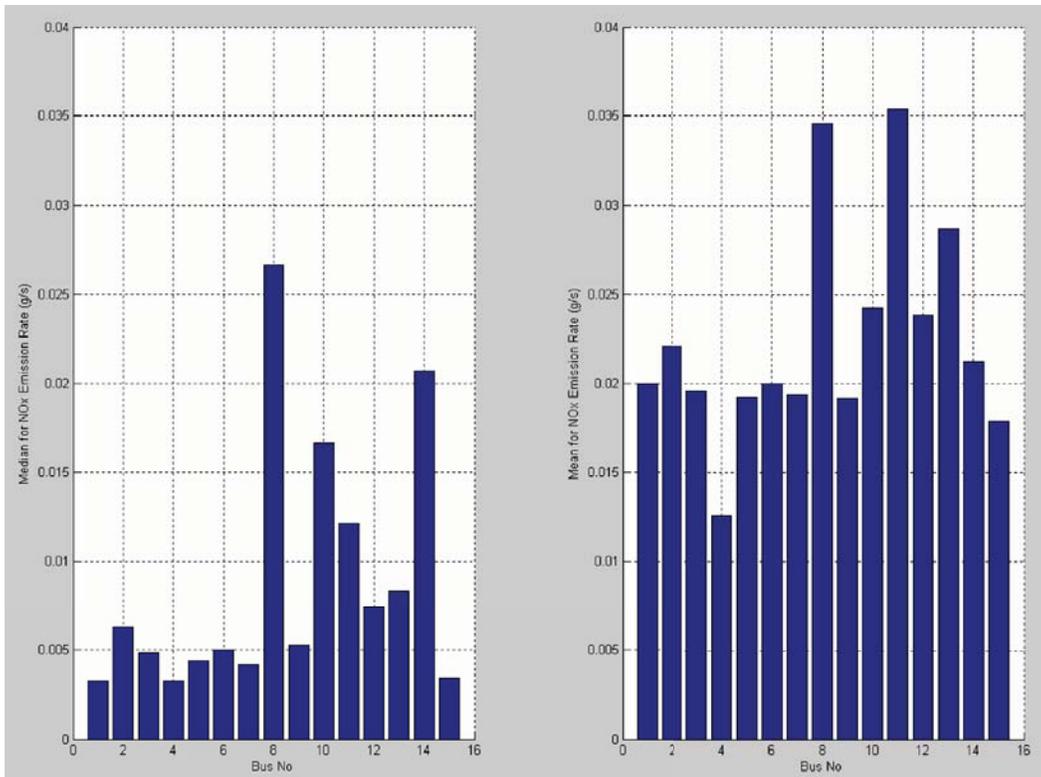


Figure 9-6 Median and Mean of NO<sub>x</sub> Emission Rates in Deceleration Mode by Bus

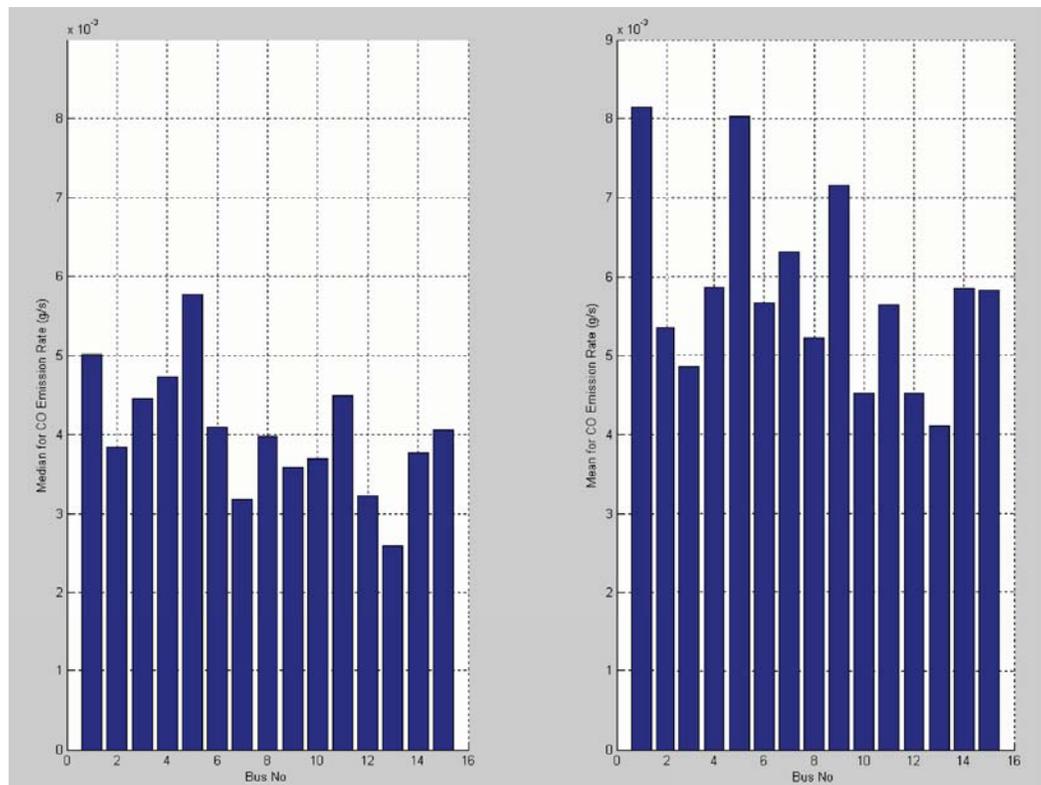


Figure 9-7 Median and Mean of CO Emission Rates in Deceleration Mode by Bus

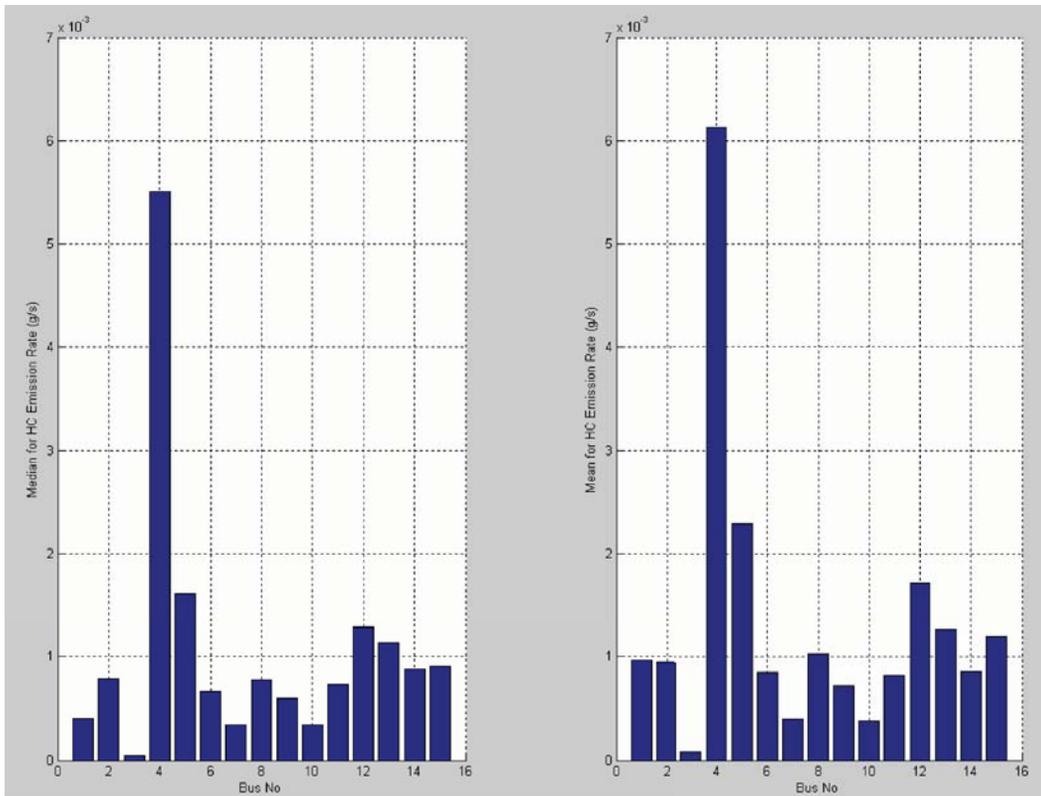


Figure 9-8 Median and Mean of HC Emission Rates in Deceleration Mode by Bus

Table 9-3 Median, and Mean for NO<sub>x</sub>, CO, and HC in Deceleration Mode by Bus

Bus ID	NO <sub>x</sub>		CO		HC	
	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<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub> 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 significantly different from those under positive engine power. Further analysis will examine this question. Moreover, bus 372 has the greatest 3<sup>rd</sup> Quartile engine power in deceleration mode, consistent with the finding in idle mode.

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

<b>Bus No</b>	<b>Minimum</b>	<b>1<sup>st</sup> Quartile</b>	<b>Median</b>	<b>3<sup>rd</sup> Quartile</b>	<b>Maximum</b>
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

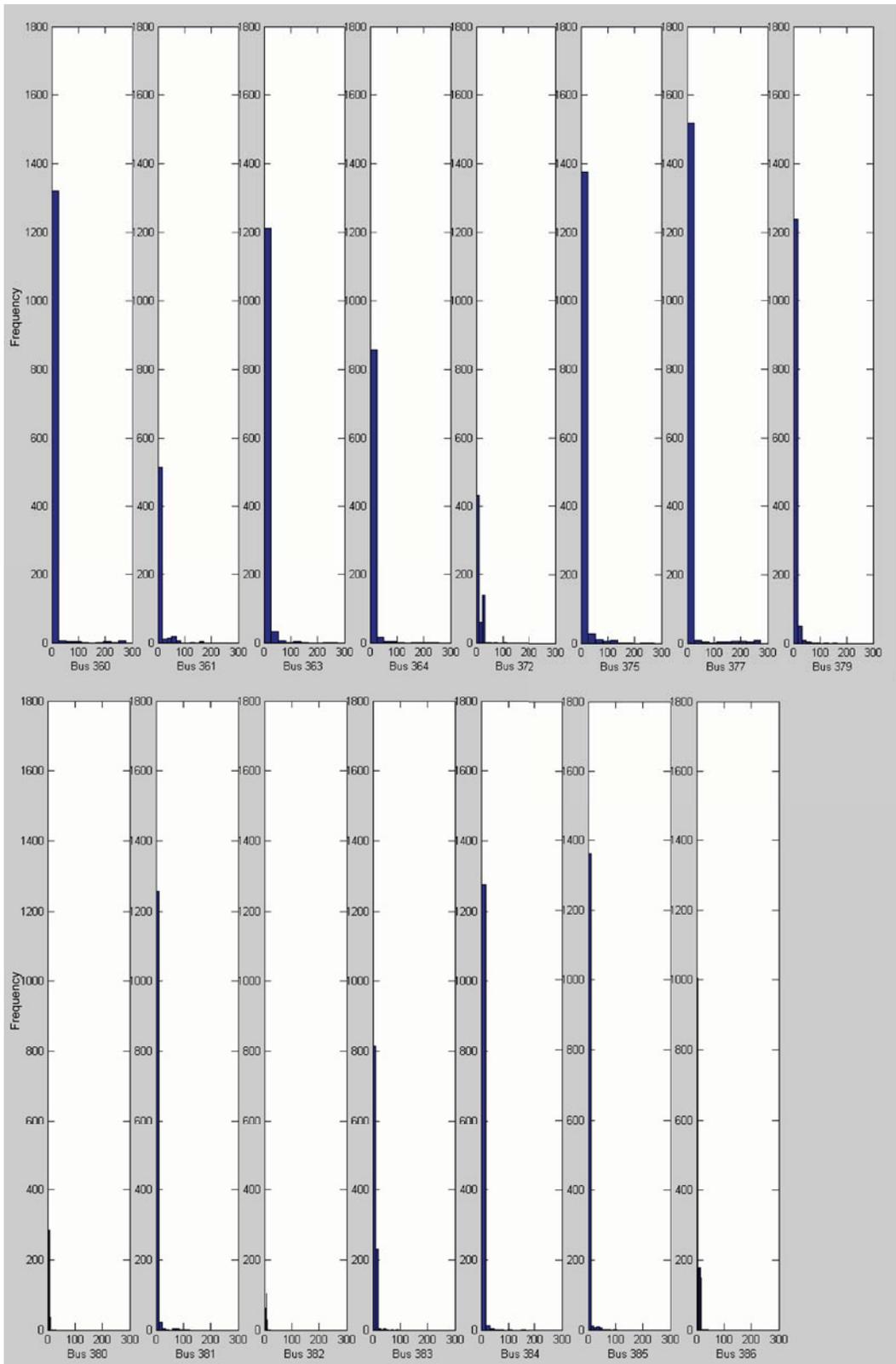


Figure 9-9 Histograms of Engine Power in Deceleration Mode by Bus

Based on definitions of “acceleration < -1 mph/s”, about 1% of data points with high engine power ( $\geq 50$  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.

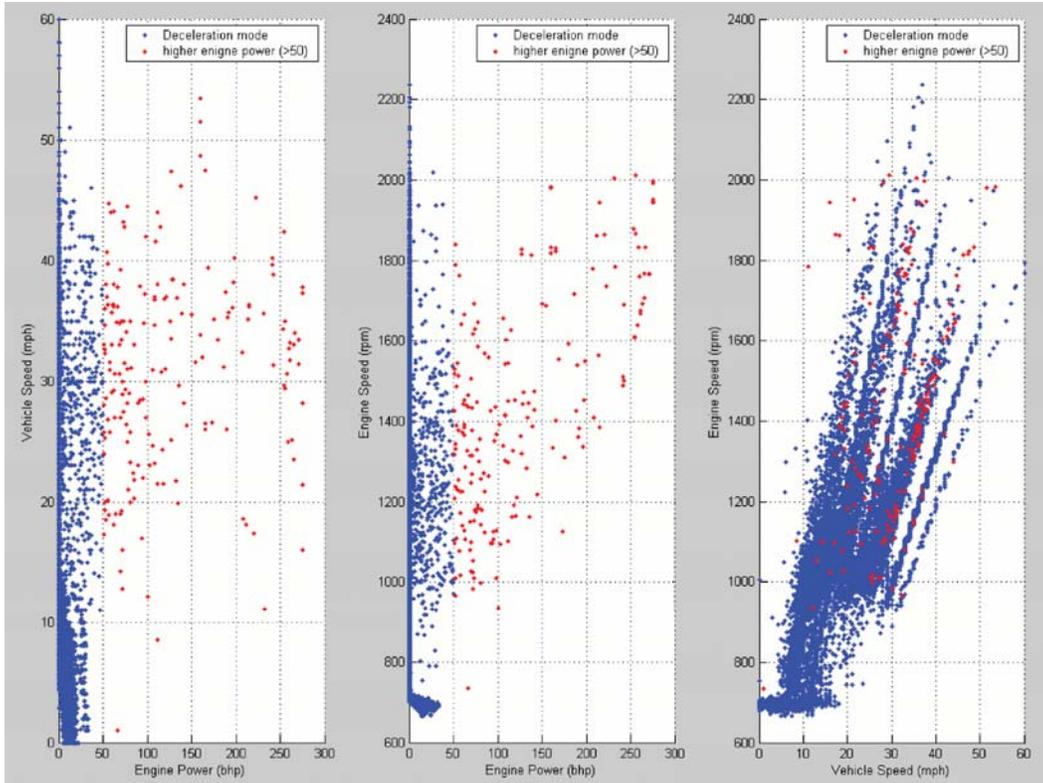


Figure 9-10 Engine Power vs. Vehicle Speed, Engine Power vs. Engine Speed, and Vehicle Speed vs. Engine Speed

### 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.

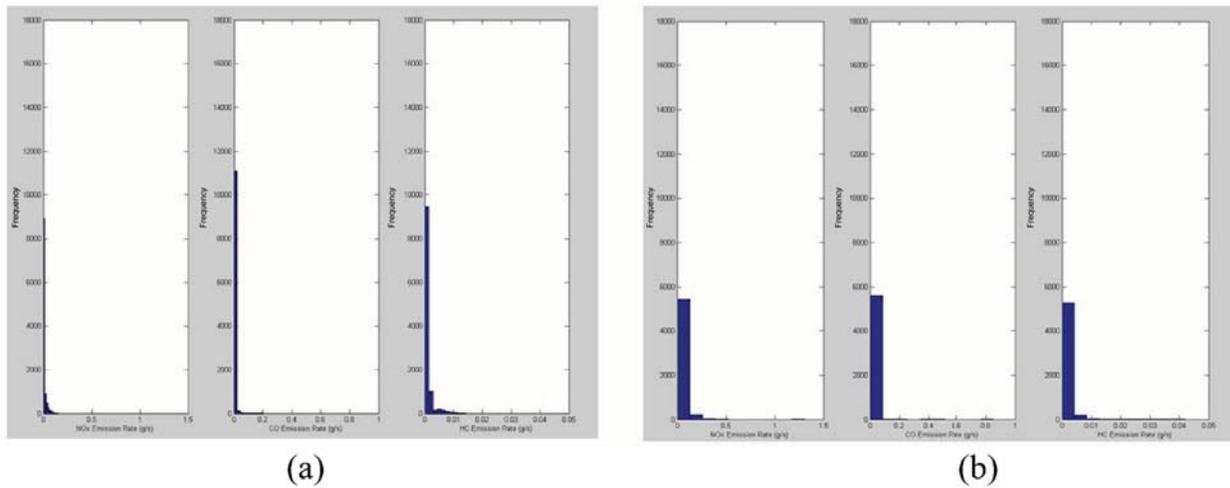


Figure 9-11 Histograms for Three Pollutants in Deceleration Motoring Mode (a) and Deceleration Non-Motoring Mode (b)

To test the differences between deceleration motoring mode and deceleration non-motoring mode, a Kolmogorov-Smirnov 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)

	<b>NO<sub>x</sub></b>	<b>CO</b>	<b>HC</b>
<b>Deceleration Mode</b>			
Number	16644	16919	16805
Minimum	0.00001	0.00001	0.00001
1 <sup>st</sup> Quartile	0.00182	0.00249	0.00039
Median	0.00611	0.00398	0.00068
3 <sup>rd</sup> 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
1 <sup>st</sup> Quartile	0.00124	0.00269	0.00041
Median	0.00272	0.00401	0.00067
3 <sup>rd</sup> 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
1 <sup>st</sup> Quartile	0.01973	0.00204	0.00034
Median	0.03431	0.00384	0.00069
3 <sup>rd</sup> 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  $\text{NO}_x$ .

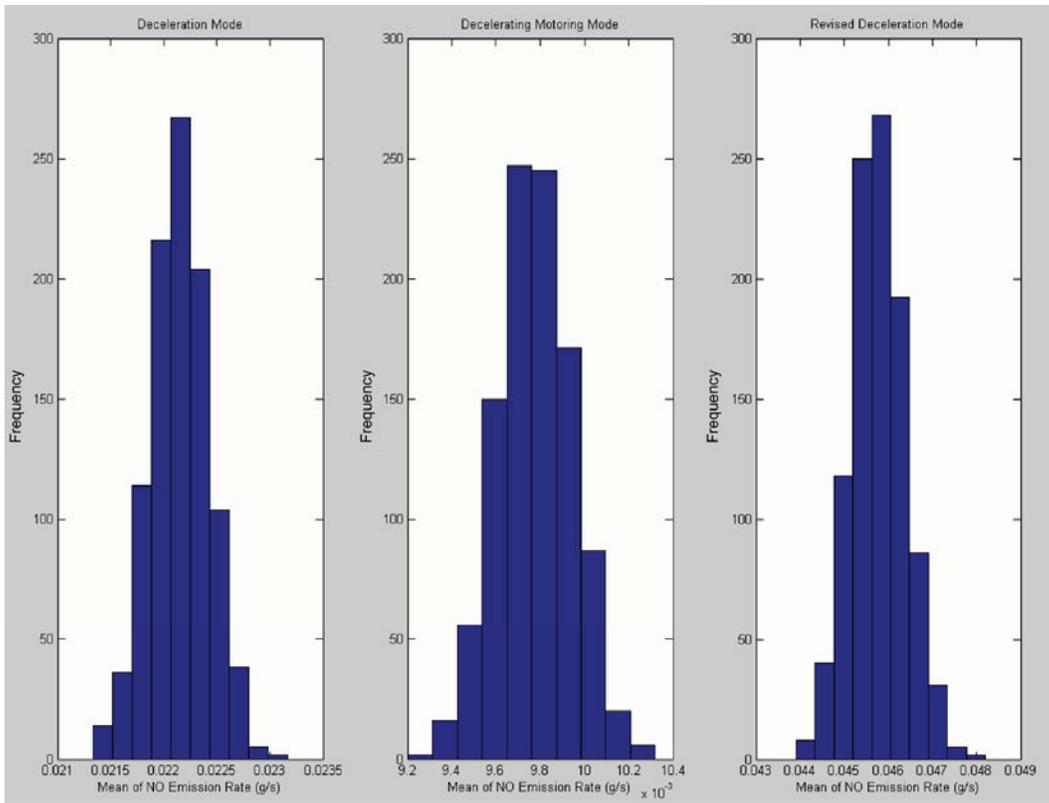


Figure 9-12 Bootstrap Results for NO<sub>x</sub> Emission Rate Estimation in Deceleration Mode

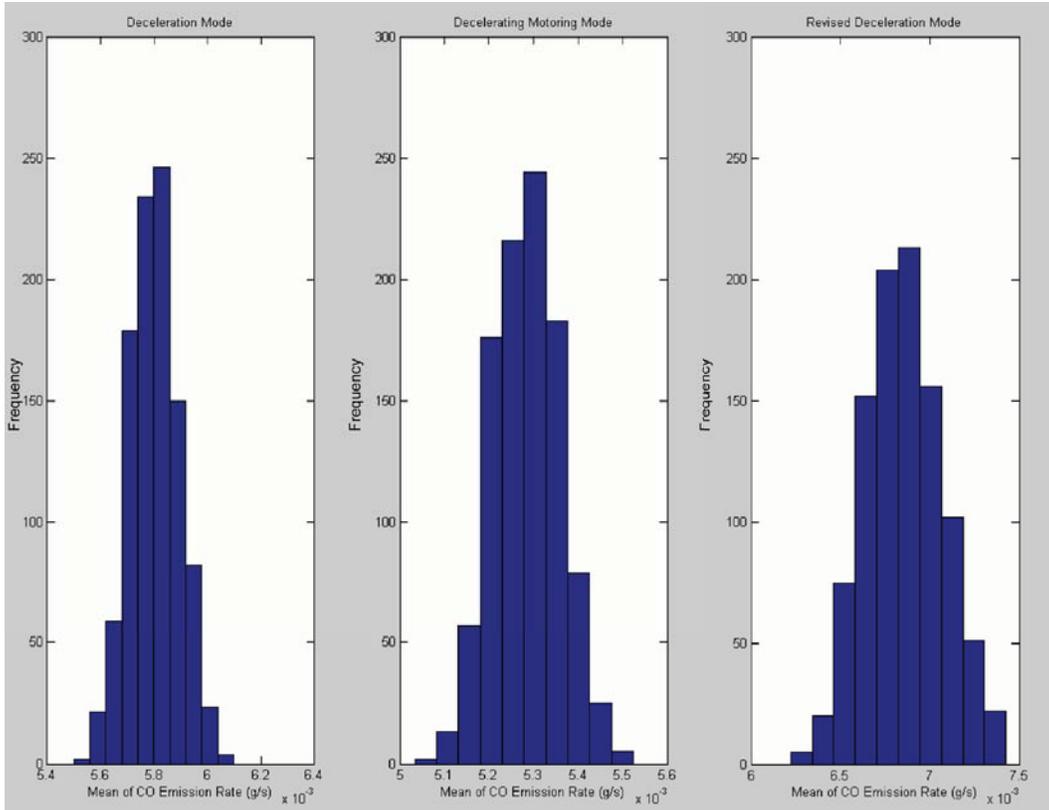


Figure 9-13 Bootstrap Results for CO Emission Rate Estimation in Deceleration Mode

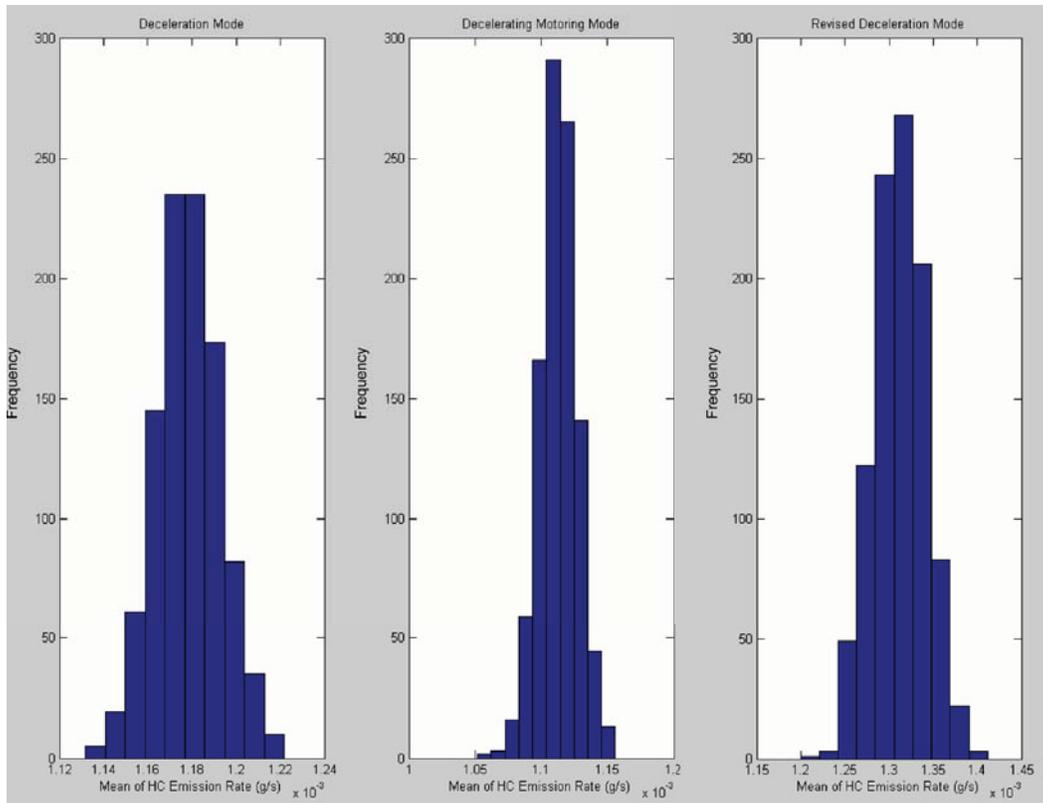


Figure 9-14 Bootstrap Results for HC Emission Rate Estimation in Deceleration Mode

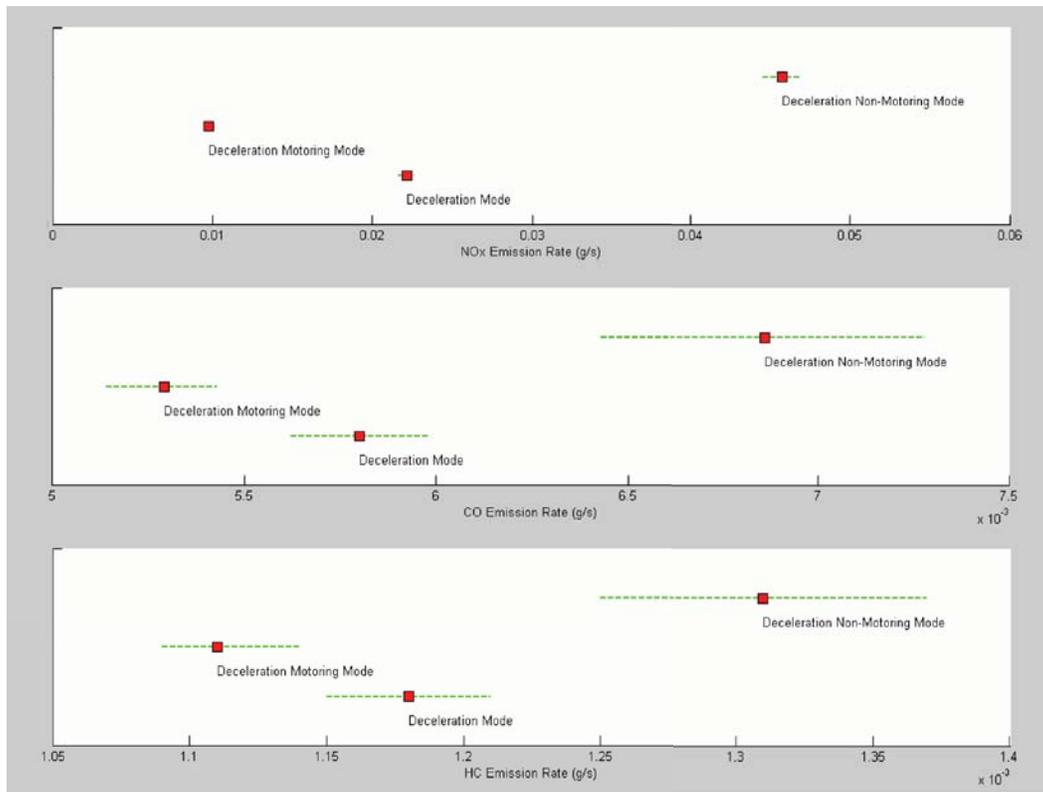


Figure 9-15 Emission Rate Estimation Based on Bootstrap for Deceleration Mode

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

		<b>Average</b>	<b>2.5% Percentile</b>	<b>97.5% Percentile</b>
<b>Deceleration Mode</b>				
NO <sub>x</sub>	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
<b>Deceleration Motoring Mode</b>				
NO <sub>x</sub>	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
<b>Deceleration Non-Motoring Mode</b>				
NO <sub>x</sub>	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<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub> 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 3<sup>rd</sup> 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<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub>.

## 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  $\leq 1$  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  $\leq 2$  mph/s
- Option 3: acceleration  $> 0$  mph/s and acceleration  $\leq 1$  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.

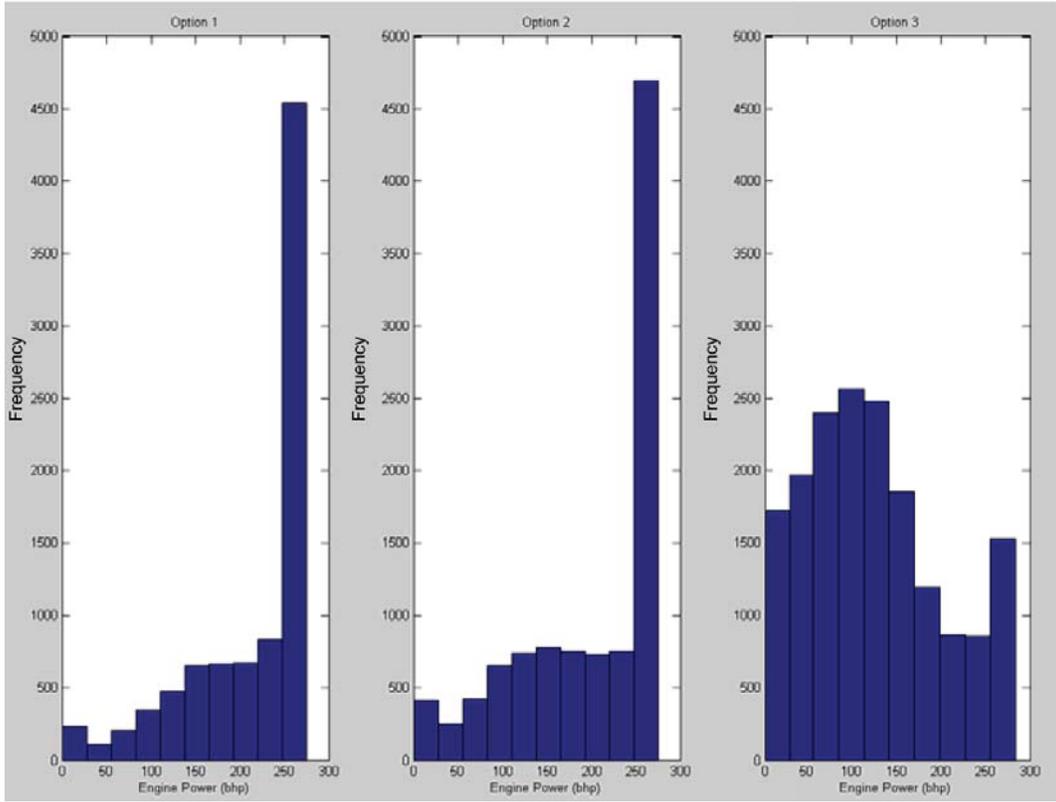


Figure 10-1 Engine Power Distribution for Three Options

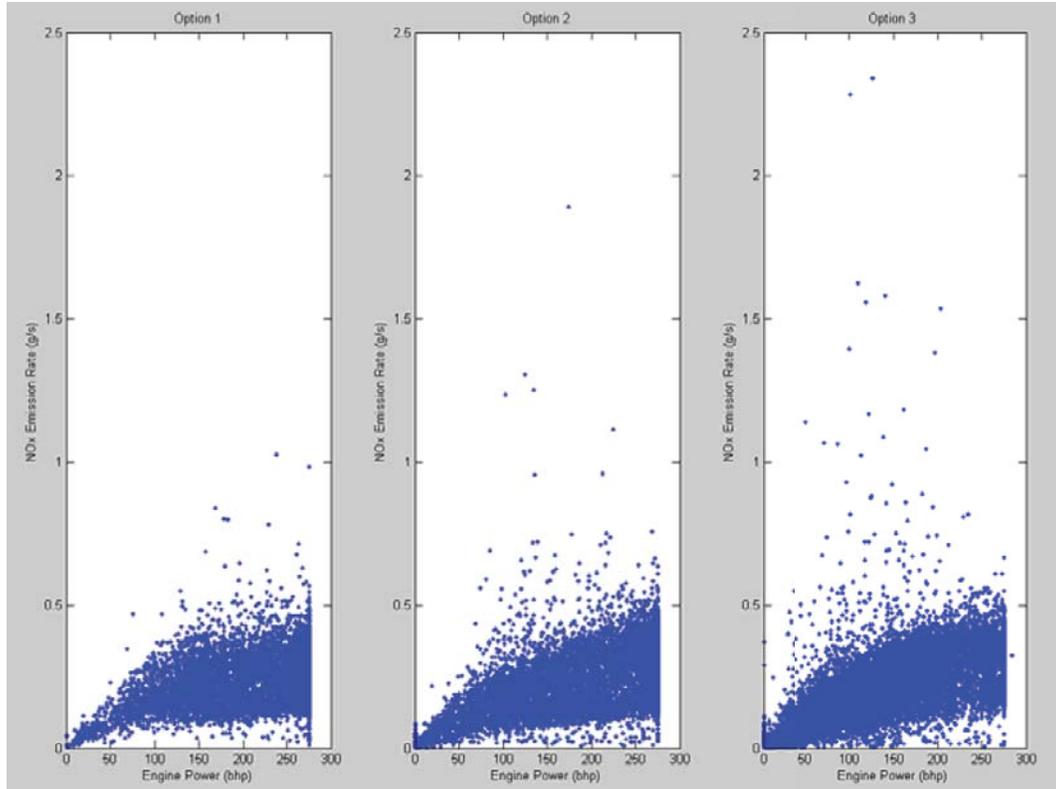


Figure 10-2 Engine Power vs. NO<sub>x</sub> Emission Rate (g/s) for Three Options

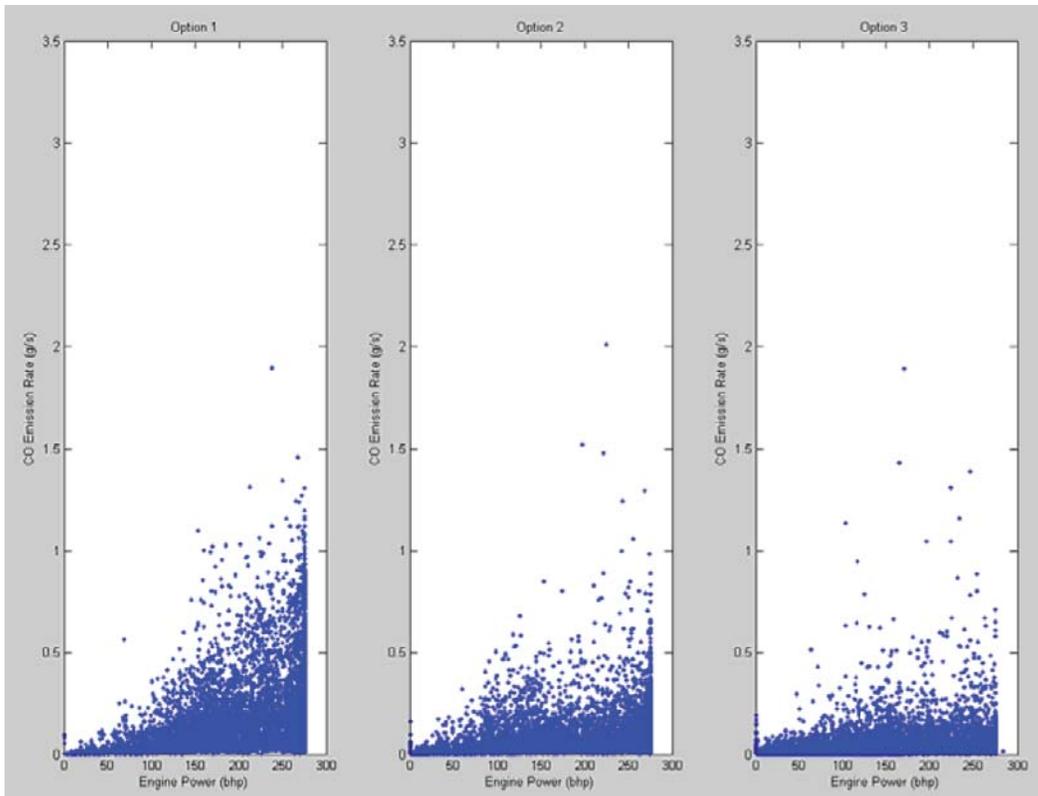


Figure 10-3 Engine Power vs. CO Emission Rate (g/s) for Three Options

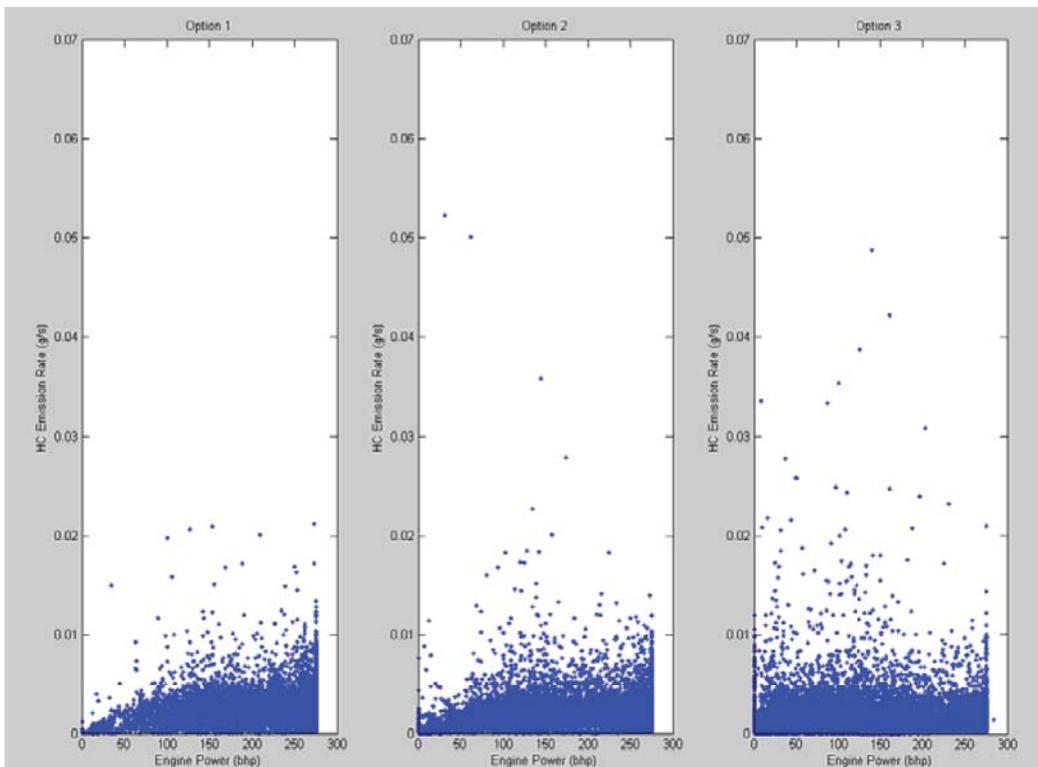


Figure 10-4 Engine Power vs. HC Emission Rate (g/s) for Three Options

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

Acceleration	Pollutants	Engine Power (brake horsepower (bhp))					Total
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	
Option 1	NO <sub>x</sub>	322	446	852	1229	5870	8719
	CO	319	444	851	1228	5870	8712
	HC	318	440	833	1203	5649	8443
Option 2	NO <sub>x</sub>	613	865	1358	1324	6015	10175
	CO	606	858	1355	1321	6012	10152
	HC	605	843	1328	1287	5824	9887
Option 3	NO <sub>x</sub>	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))					Total
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	
Option 1	NO <sub>x</sub>	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 <sub>x</sub>	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 <sub>x</sub>	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 ( $\geq 200$  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 <sub>x</sub>	935	1311	2210	2553	11885	18894
	CO	925	1302	2206	2549	11882	18864
	HC	923	1283	2161	2490	11473	18330
Percentage	NO <sub>x</sub>	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 <sub>x</sub>	15885	8988	7173	3536	3792	39374
	CO	15834	8940	7145	3529	3770	39218
	HC	15481	8600	6830	3394	3715	38020
Percentage	NO <sub>x</sub>	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%

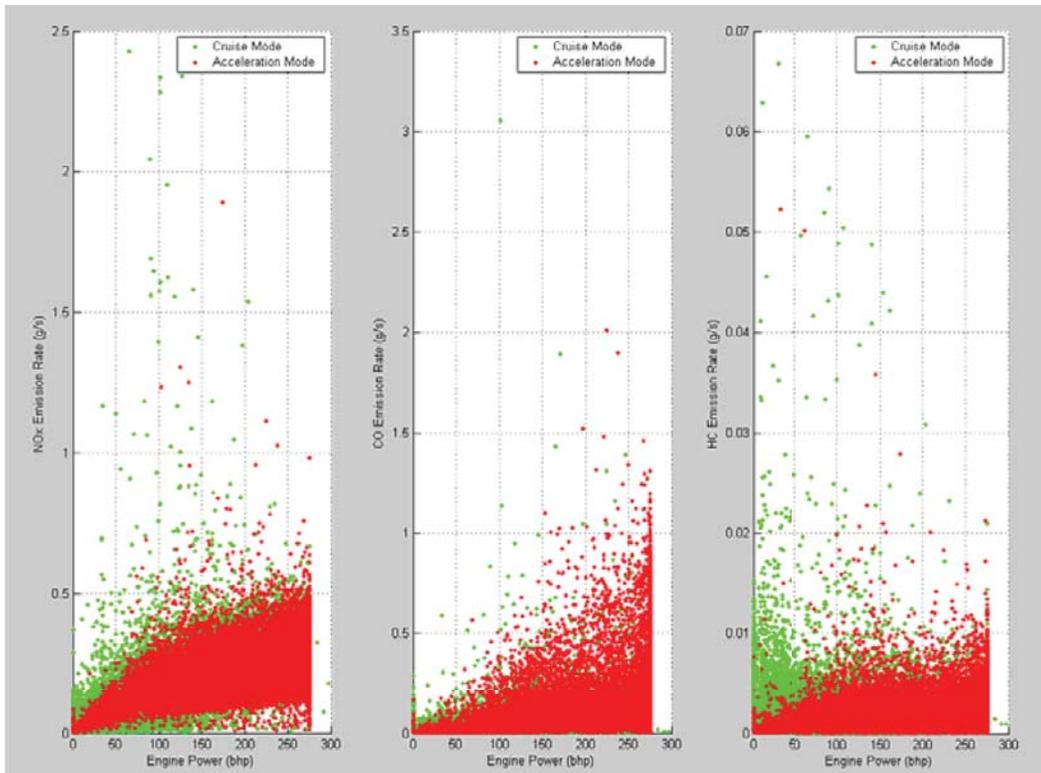


Figure 10-5 Engine Power vs. Emission Rate for Acceleration Mode and Cruise Mode

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.

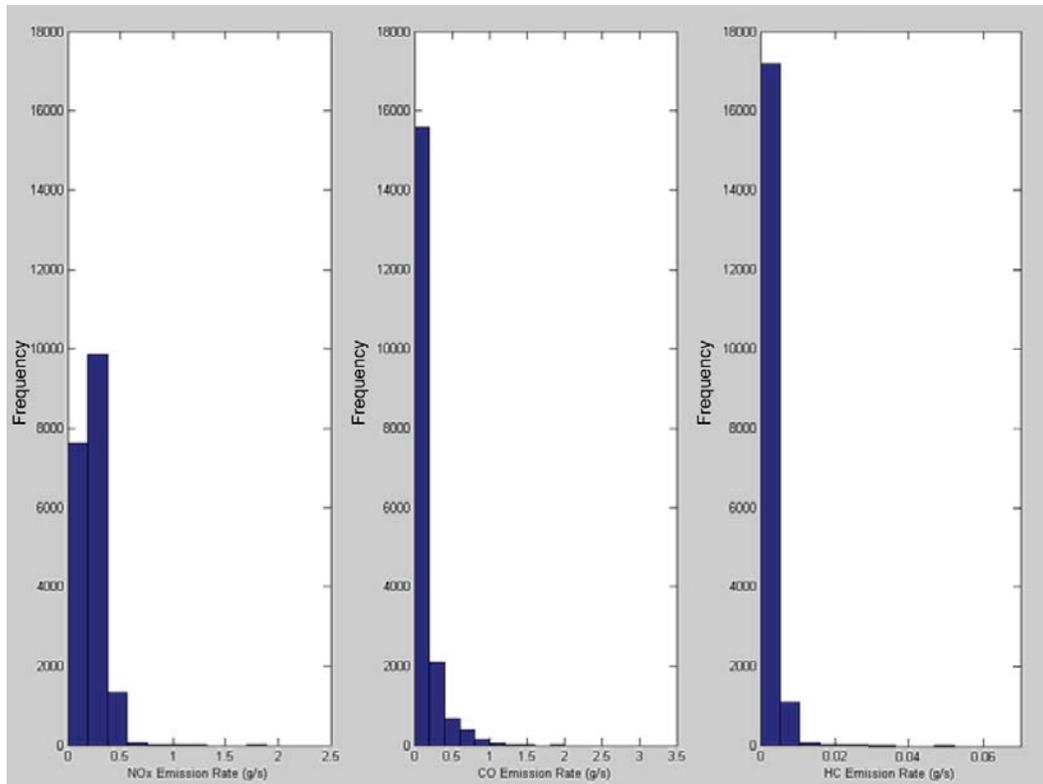


Figure 10-6 Histograms of Three Pollutants for Acceleration Mode

Inter-bus response variability for acceleration mode operations is illustrated in Figures 10-7 to 10-9 using median and mean of NO<sub>x</sub>, 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

Bus ID	NO <sub>x</sub>		CO		HC	
	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

Bus ID	NO <sub>x</sub>		CO		HC	
	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

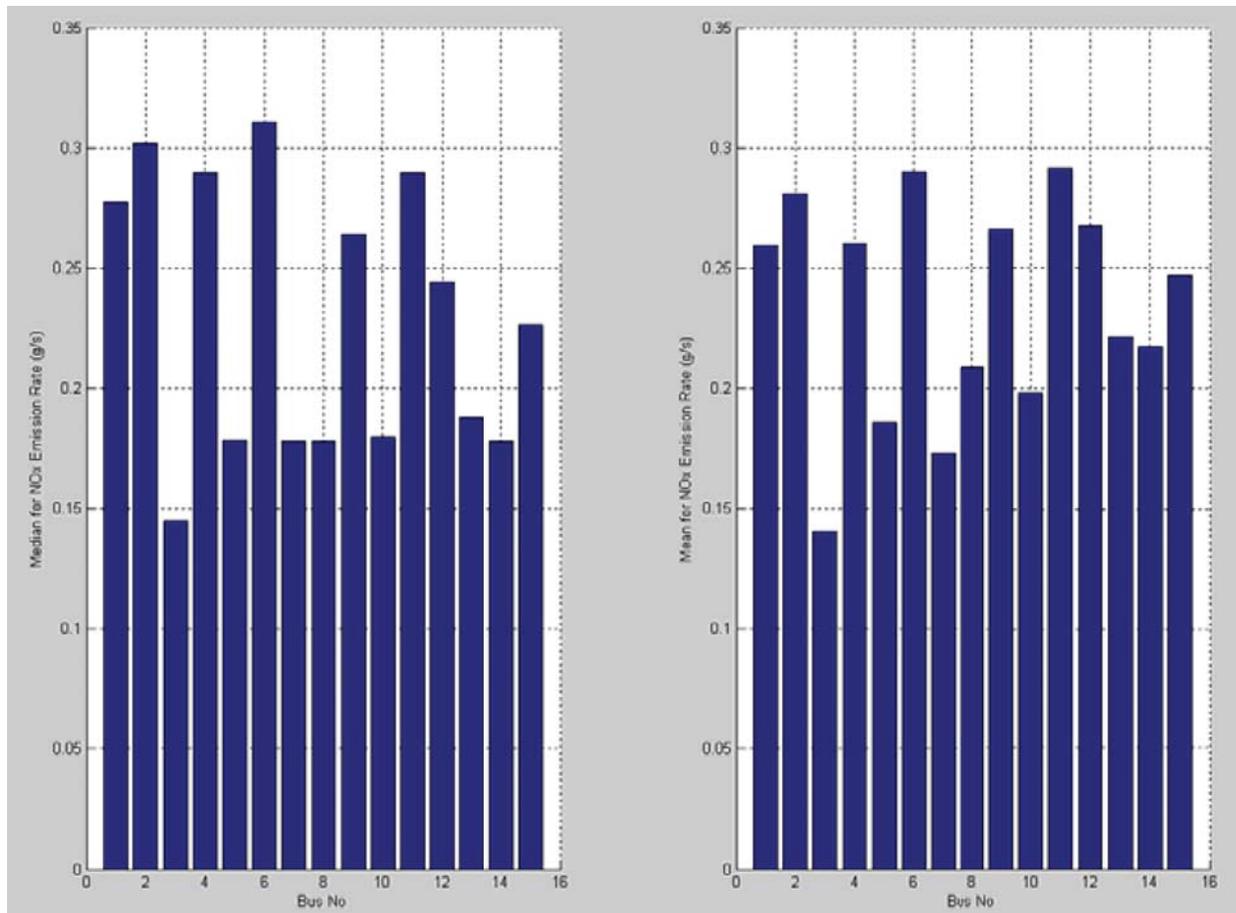


Figure 10-7 Median and Mean of NO<sub>x</sub> Emission Rates in Acceleration Mode by Bus

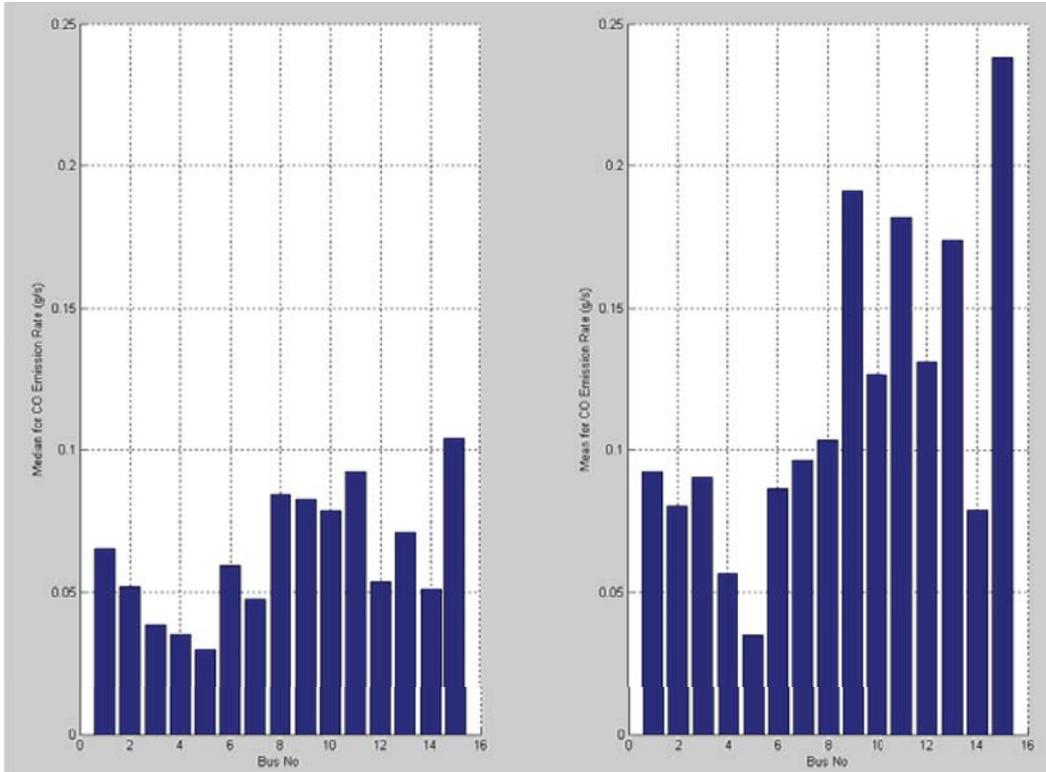


Figure 10-8 Median and Mean of CO Emission Rates in Acceleration Mode by Bus

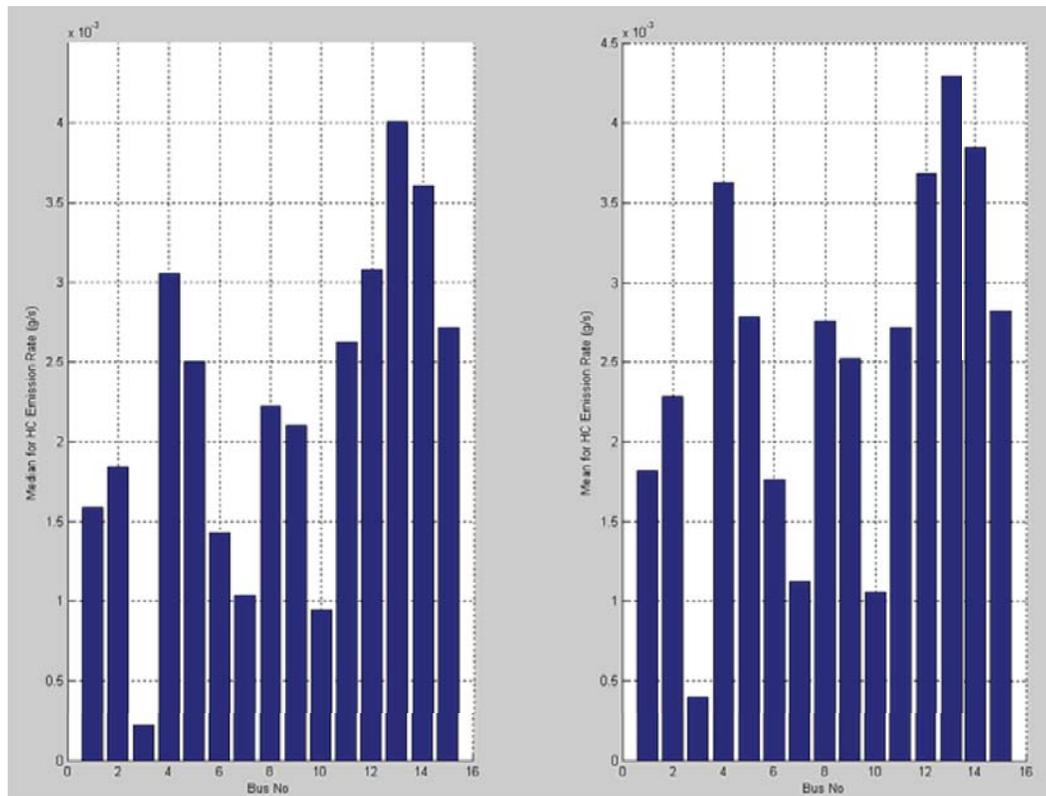


Figure 10-9 Median and Mean of HC Emission Rates in Acceleration Mode by Bus

Figures 10-7 to 10-9 and Table 10-4 illustrate that NO<sub>x</sub> 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	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> 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.

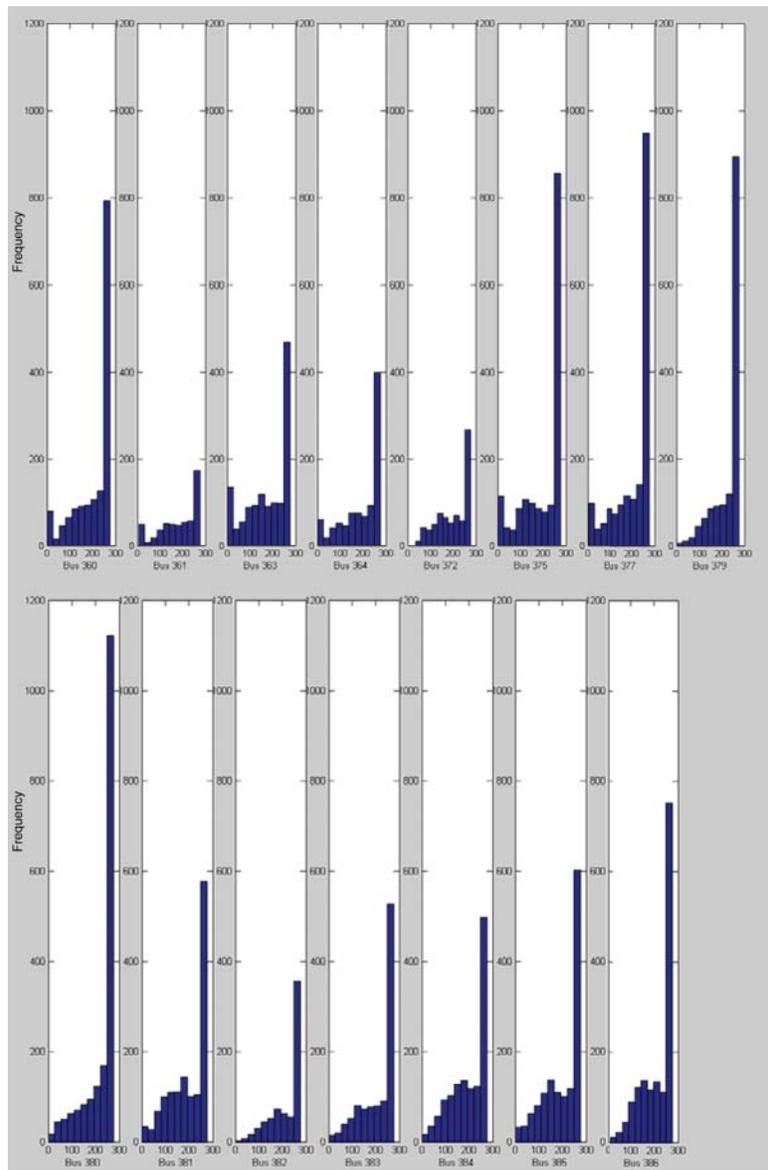


Figure 10-10 Histograms of Engine Power in Acceleration Mode by Bus

## 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<sub>x</sub>, 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<sub>x</sub>, 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<sub>x</sub>, CO, and HC, while Figures 10-14 to 10-16 present the same plots for truncated transformed emission rates for NO<sub>x</sub>, CO and HC, where a great improvement is noted.

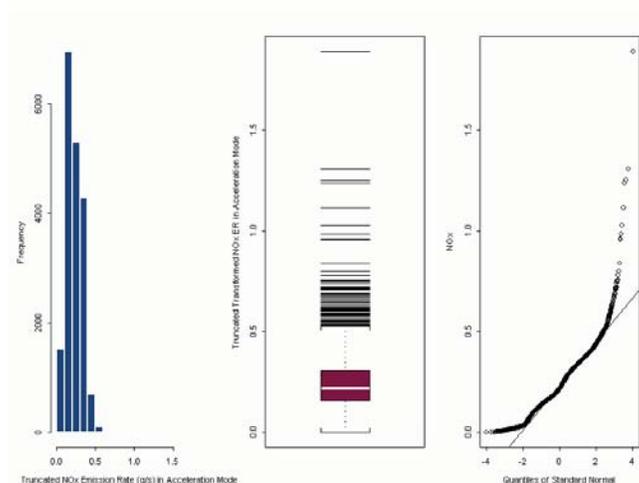


Figure 10-11 Histogram, Boxplot, and Probability Plot of Truncated NO<sub>x</sub> Emission Rate in Acceleration Mode

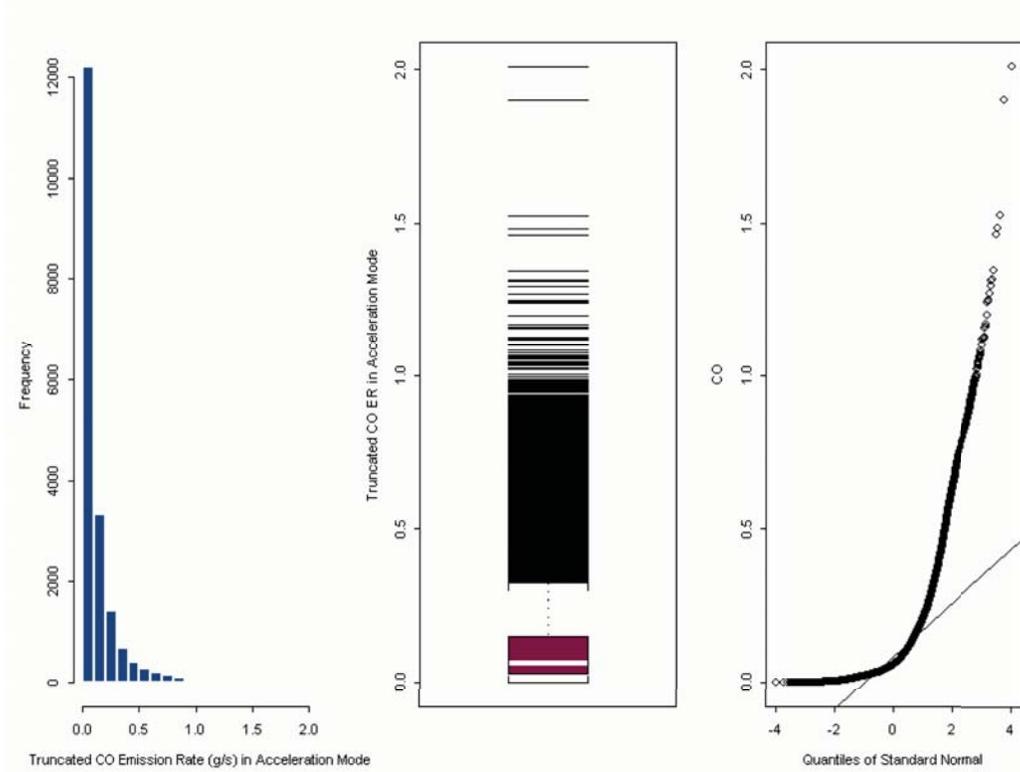


Figure 10-12 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Acceleration Mode

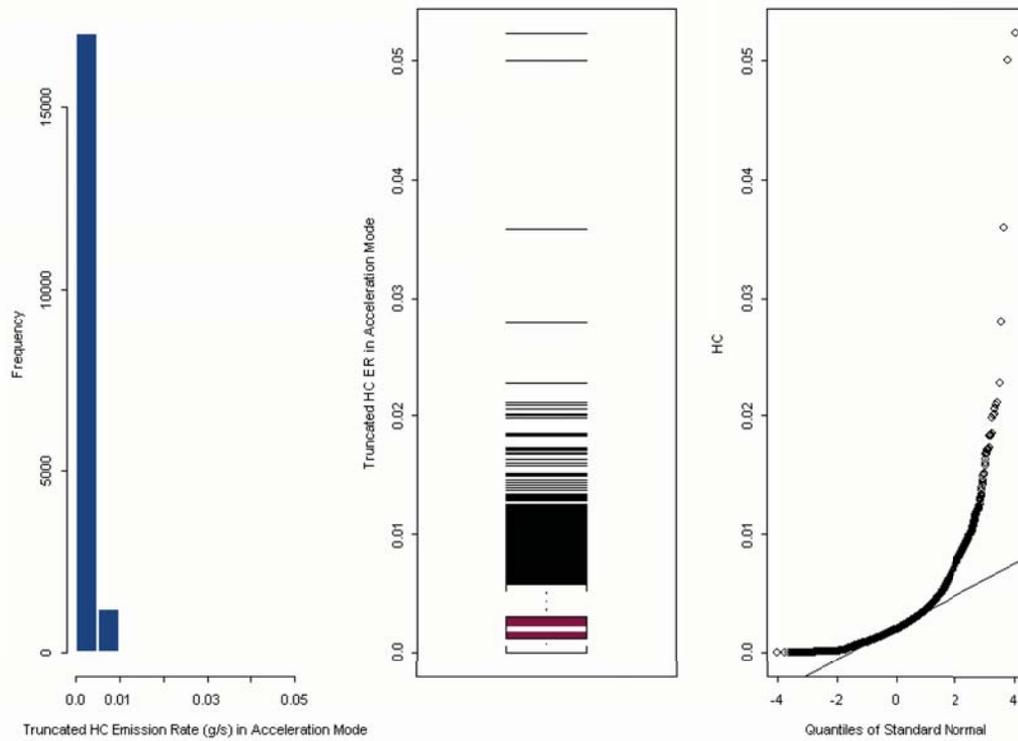


Figure 10-13 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Acceleration Mode

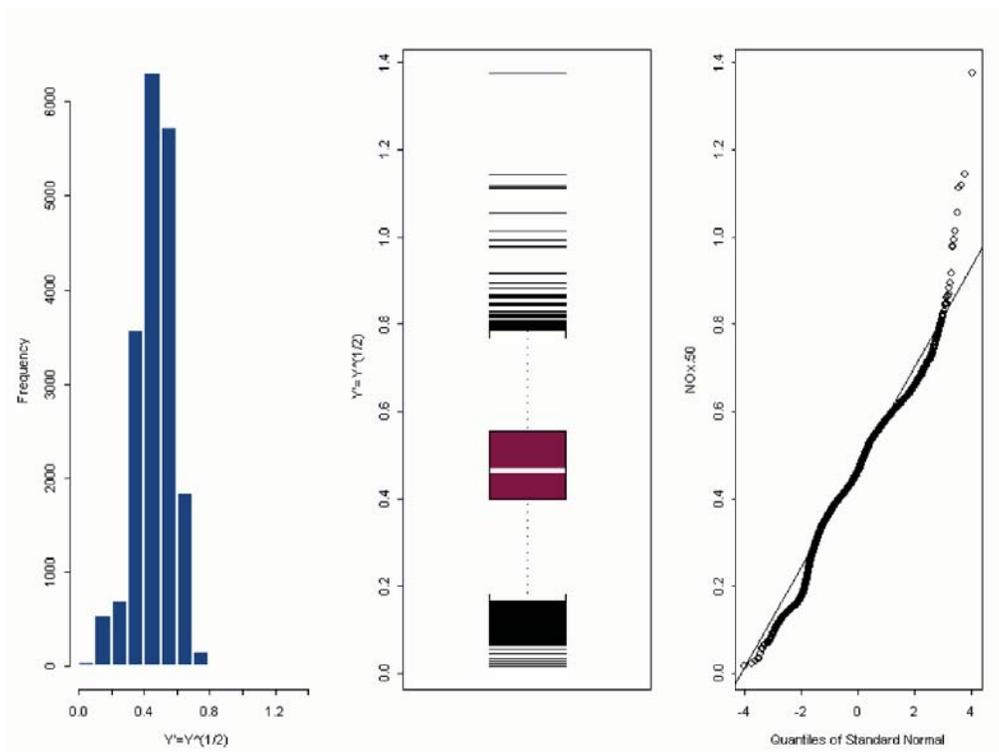


Figure 10-14 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO<sub>x</sub> Emission Rate in Acceleration Mode

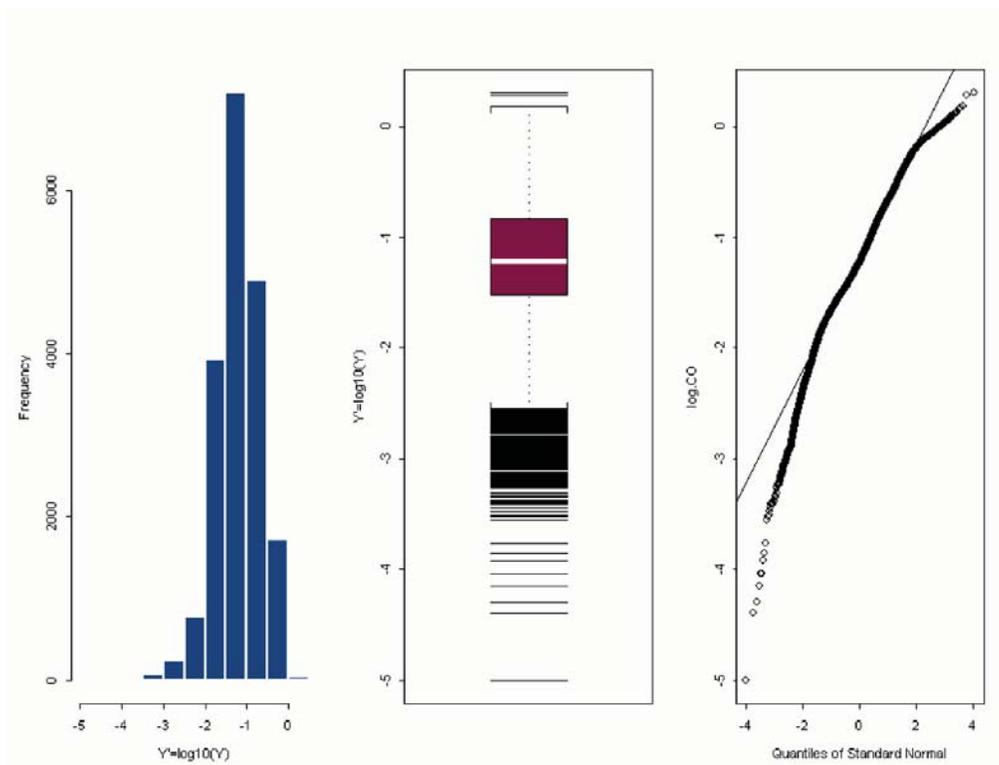


Figure 10-15 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate in Acceleration Mode

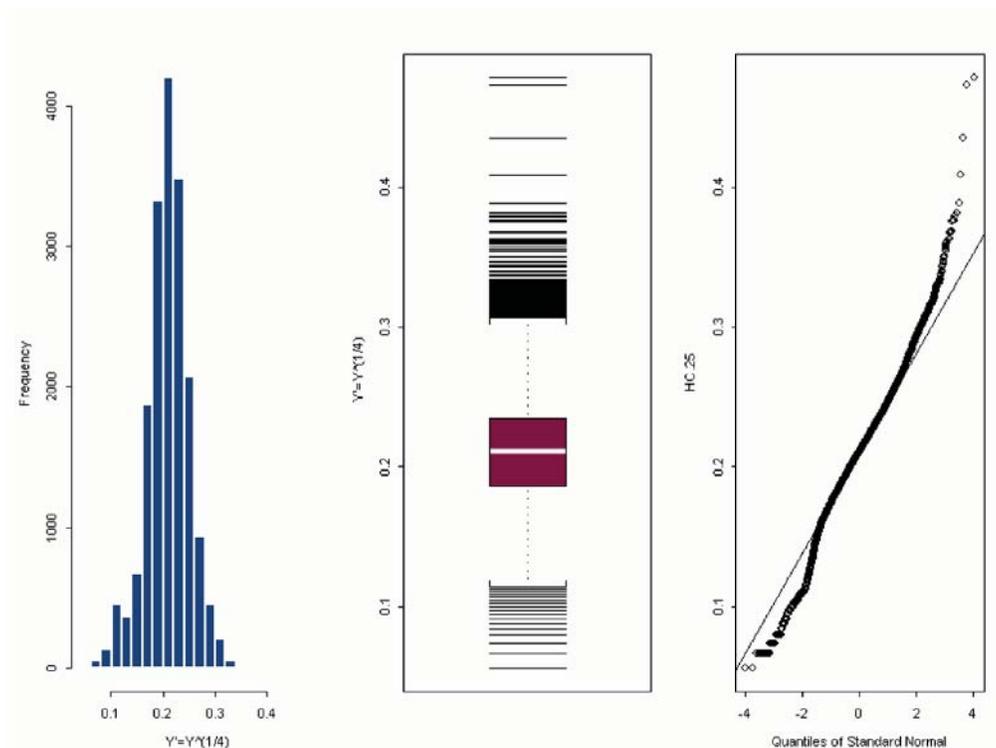


Figure 10-16 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate in Acceleration Mode

### 10.3.1.1 $\text{NO}_x$ HTBR Tree Model Development

Figure 10-17 illustrates the initial tree model used for truncated transformed  $\text{NO}_x$  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.



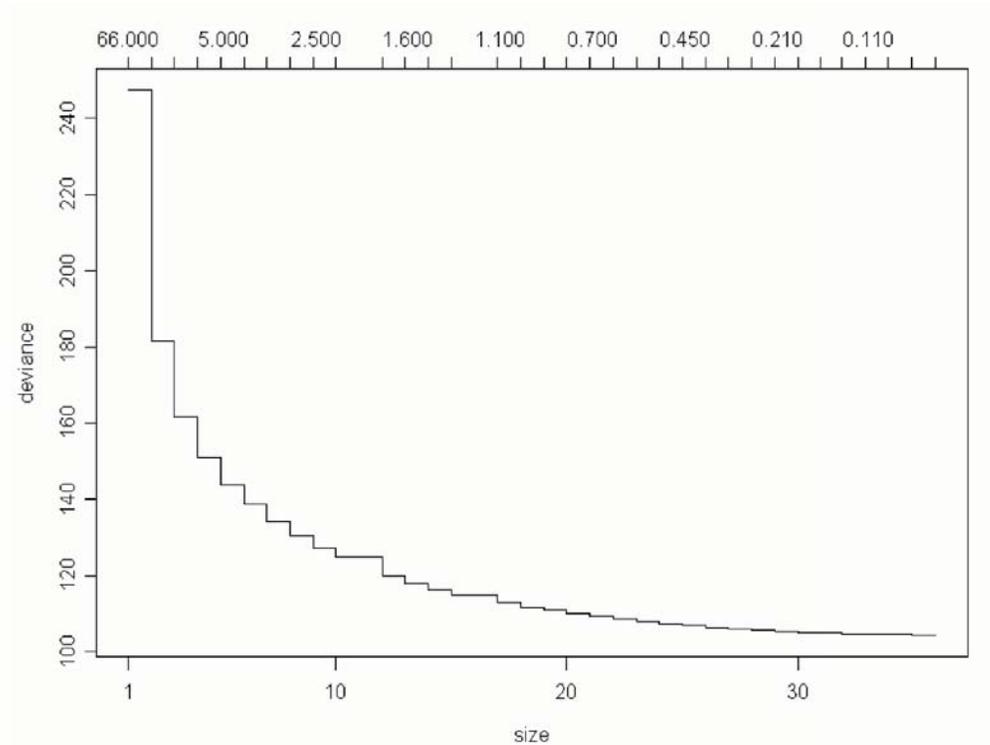


Figure 10-18 Reduction in Deviance with the Addition of Nodes of Regression Tree for Truncated Transformed NO<sub>x</sub> Emission Rate in Acceleration Mode

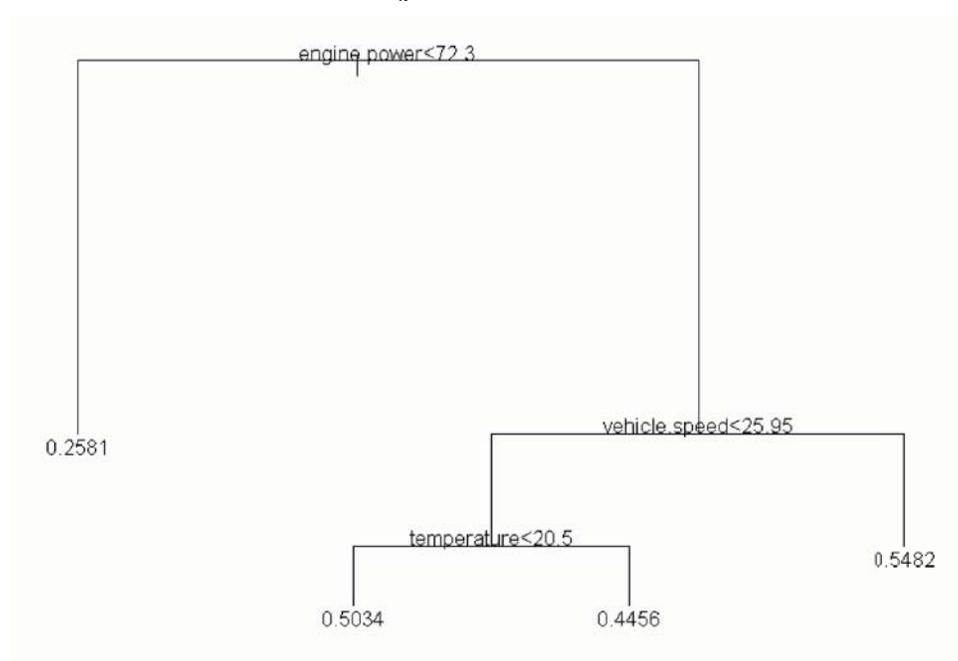


Figure 10-19 Trimmed Regression Tree Model for Truncated Transformed NO<sub>x</sub> Emission Rate in Acceleration Mode

Table 10-7 Trimmed Regression Tree Results for Truncated Transformed NO<sub>x</sub> Emission Rate in Acceleration Mode

```

Regression tree:
snip.tree(tree = tree(formula = NOx.50 ~ model.year + odometer + temperature +
  baro + humidity + vehicle.speed + oil.temperature + 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<sub>x</sub> emissions. This result is consistent with previous research results which verified the important effect of engine power on NO<sub>x</sub> 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<sub>x</sub> 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.

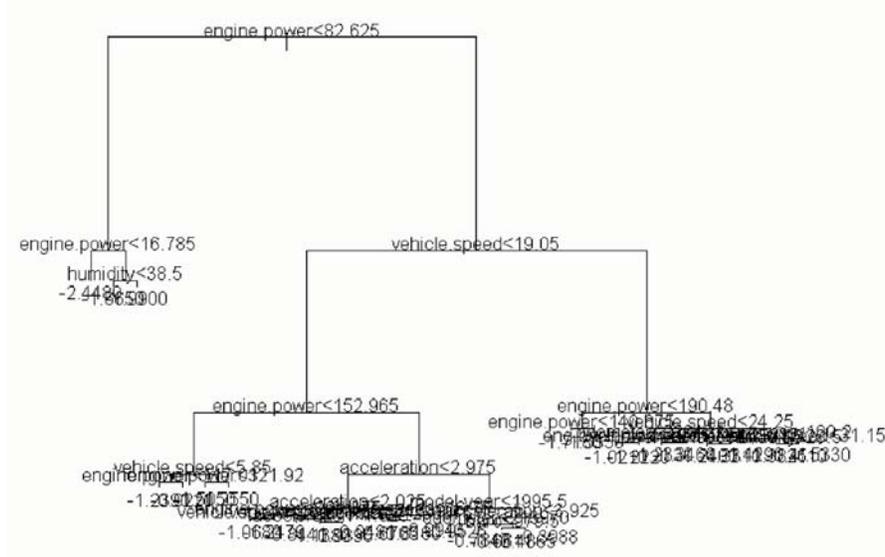


Figure 10-20 Original Untrimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode

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.temperure + 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.

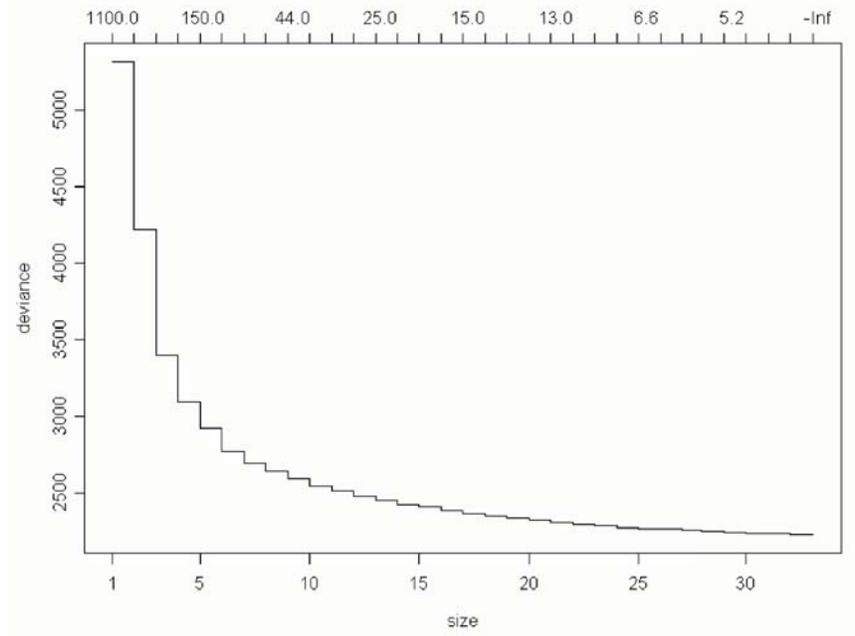


Figure 10-21 Reduction in Deviance with the Addition of Nodes of Regression Tree for Truncated Transformed CO Emission Rate in Acceleration Mode

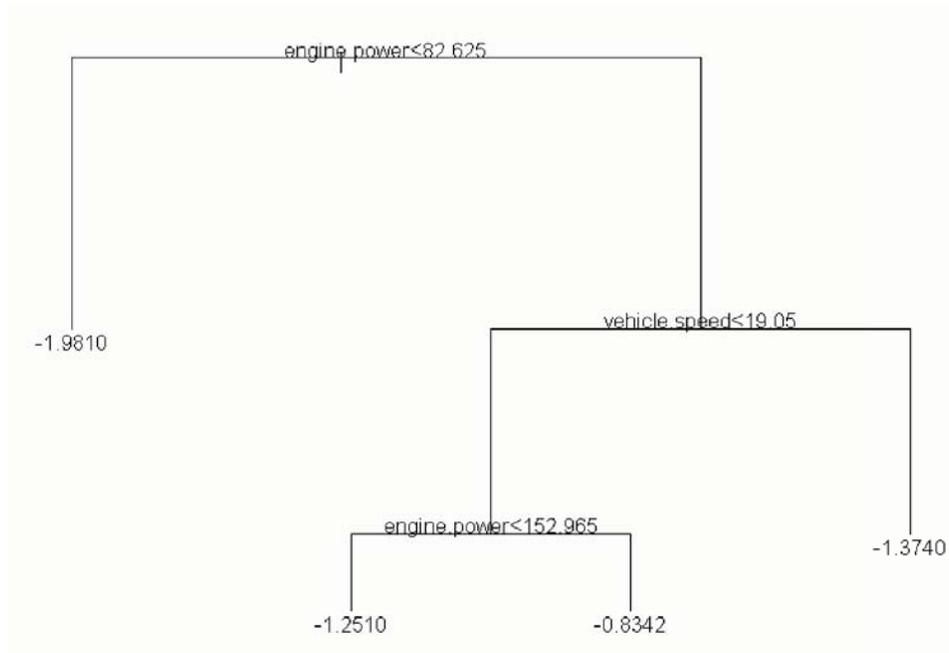


Figure 10-22 Trimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Acceleration Mode

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<sub>x</sub> 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.



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<sub>x</sub> and CO models.

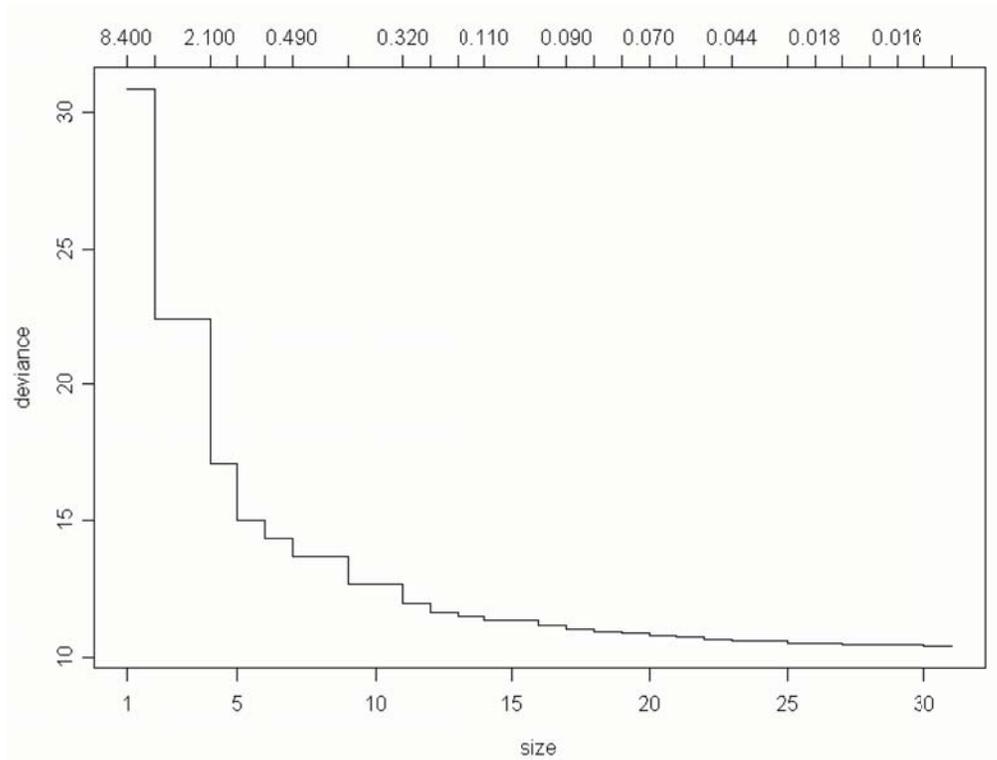


Figure 10-24 Reduction in Deviance with the Addition of Nodes of Regression Tree for Truncated Transformed HC Emission Rate in Acceleration Mode

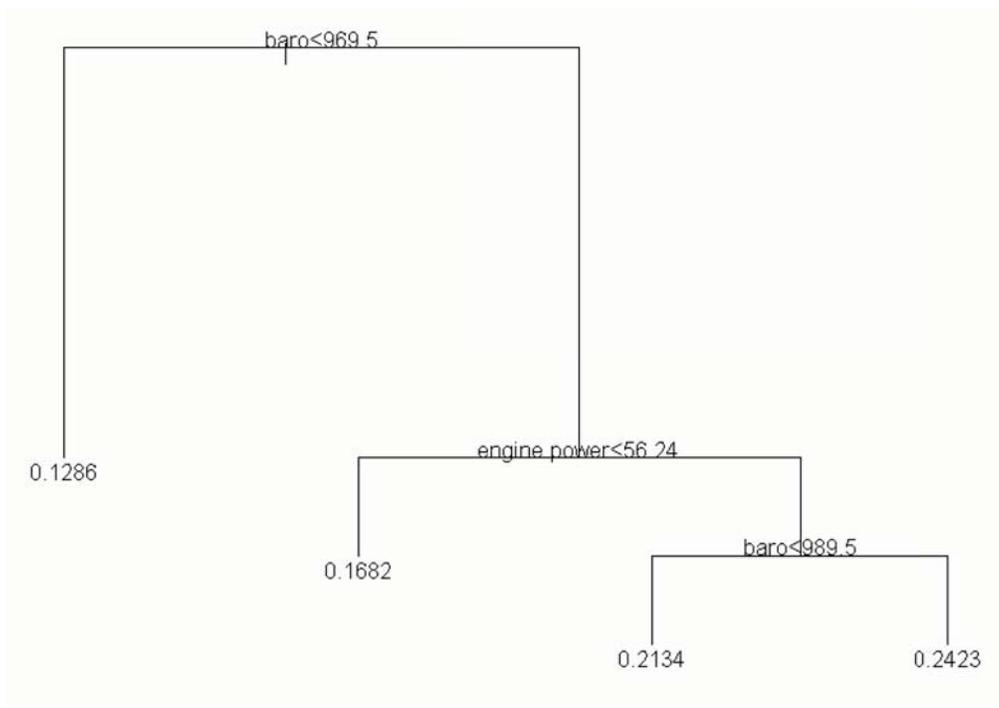


Figure 10-25 Trimmed Regression Tree Model for Truncated Transformed HC in Acceleration Mode

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.temperure + 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 ( $\text{baro} < 969.5$ ) 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 than 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.

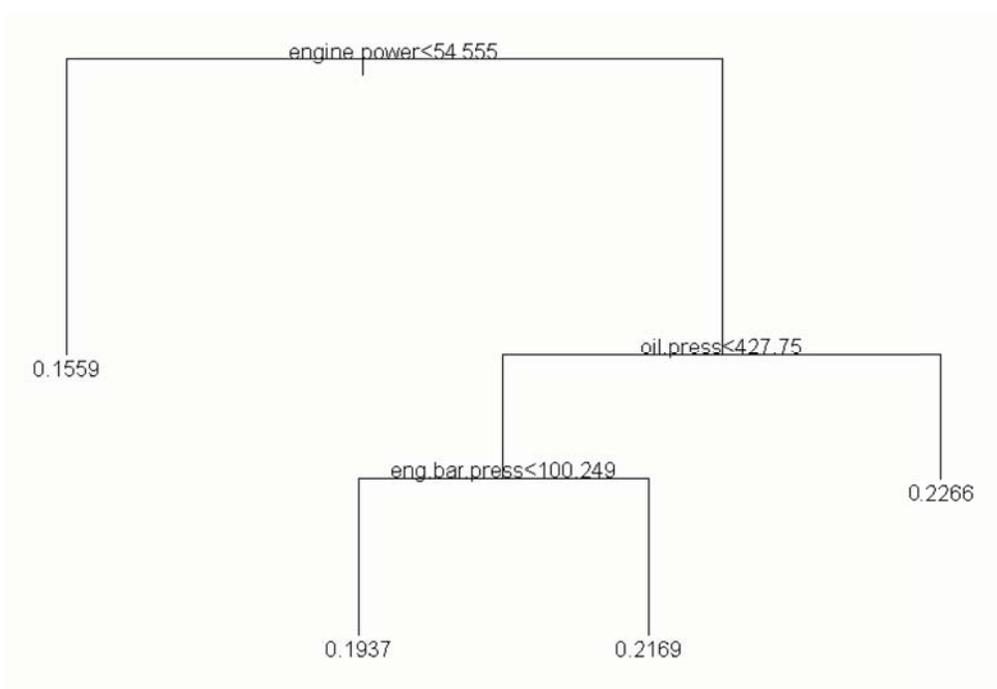


Figure 10-26 Secondary Trimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Acceleration Mode

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<sub>x</sub> 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.

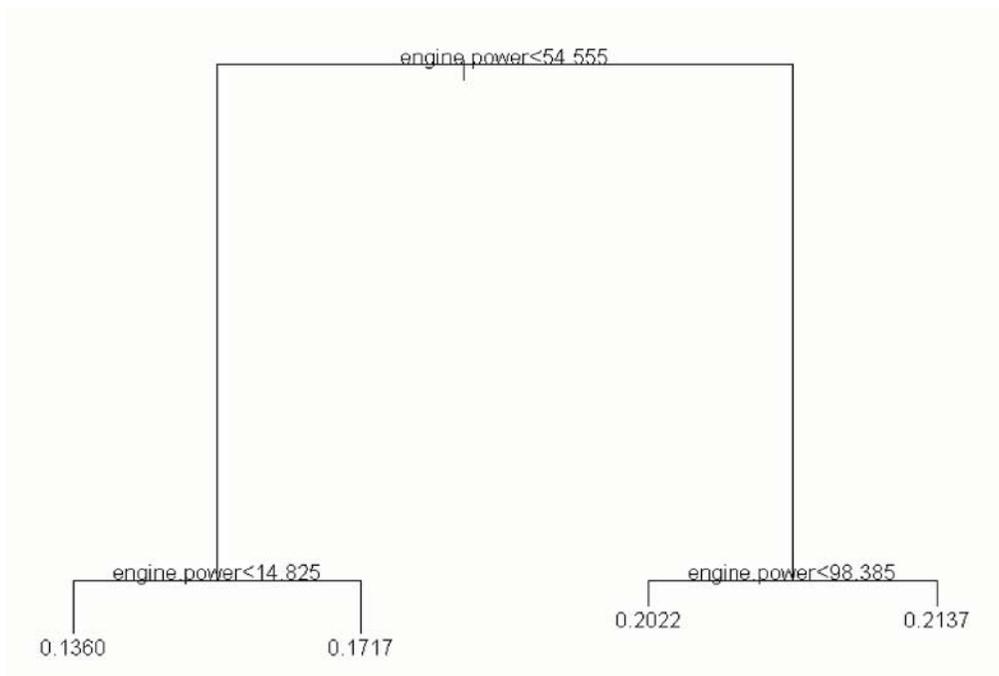


Figure 10-27 Final Regression Tree Model for Truncated Transformed HC and Engine Power in Acceleration Mode

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<sub>x</sub> Emission Rate Model Development for Acceleration Mode

Based on previous analysis, truncated transformed NO<sub>x</sub> 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:

$$Y = \beta_0 + \beta_1(\text{engine.power}) + \text{Error} \quad (1.1)$$

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

Table 10-14 Regression Result for NO<sub>x</sub> 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<sub>x</sub>. 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.

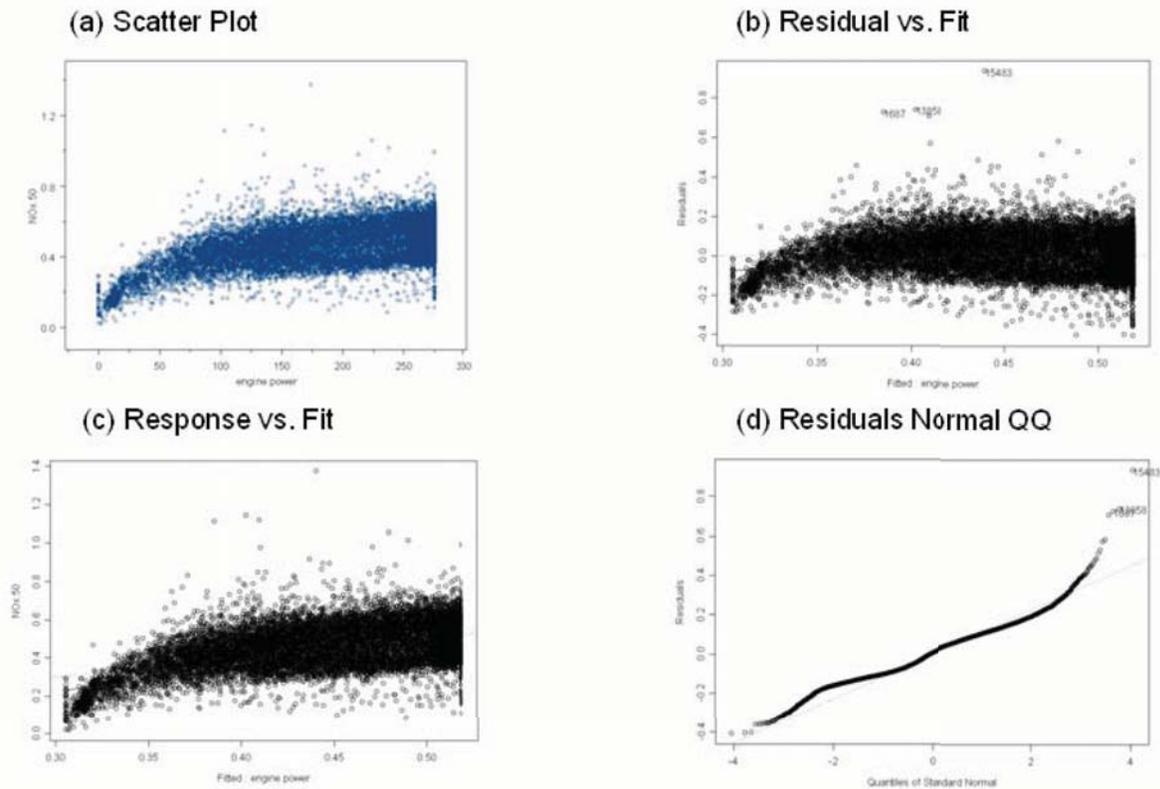


Figure 10-28 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.1

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.

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

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

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<sub>x</sub> 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
    
```

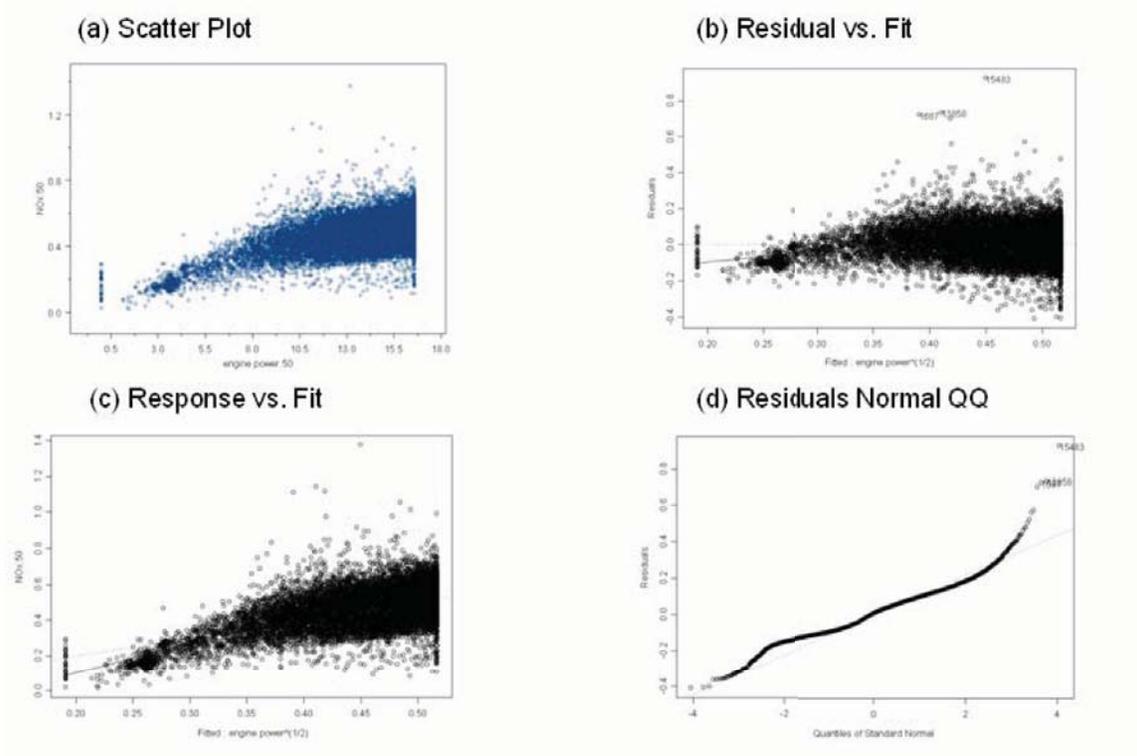


Figure 10-29 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.2

Table 10-16 Regression Result for NO<sub>x</sub> 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
    
```

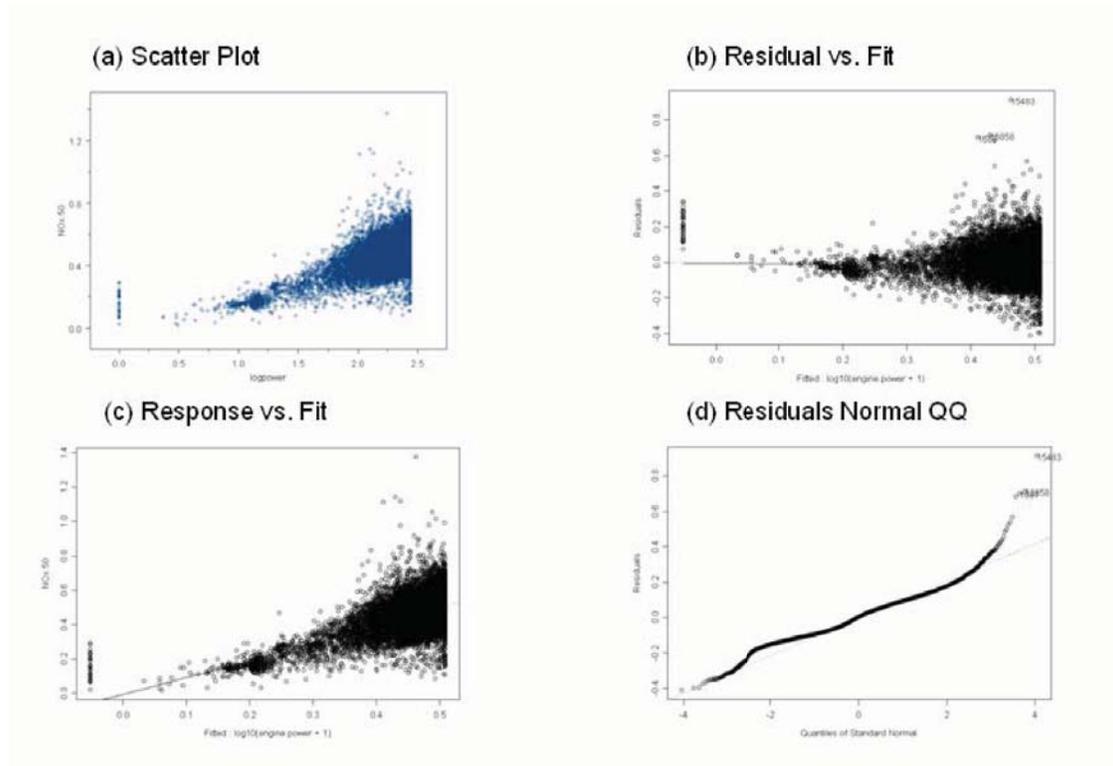


Figure 10-30 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.3

The results suggest that by using square root transformed engine power, the model increases the amount of variance explained in truncated transformed NO<sub>x</sub> 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<sup>2</sup> 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:

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

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

Table 10-17 Regression Result for NO<sub>x</sub> 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
    
```

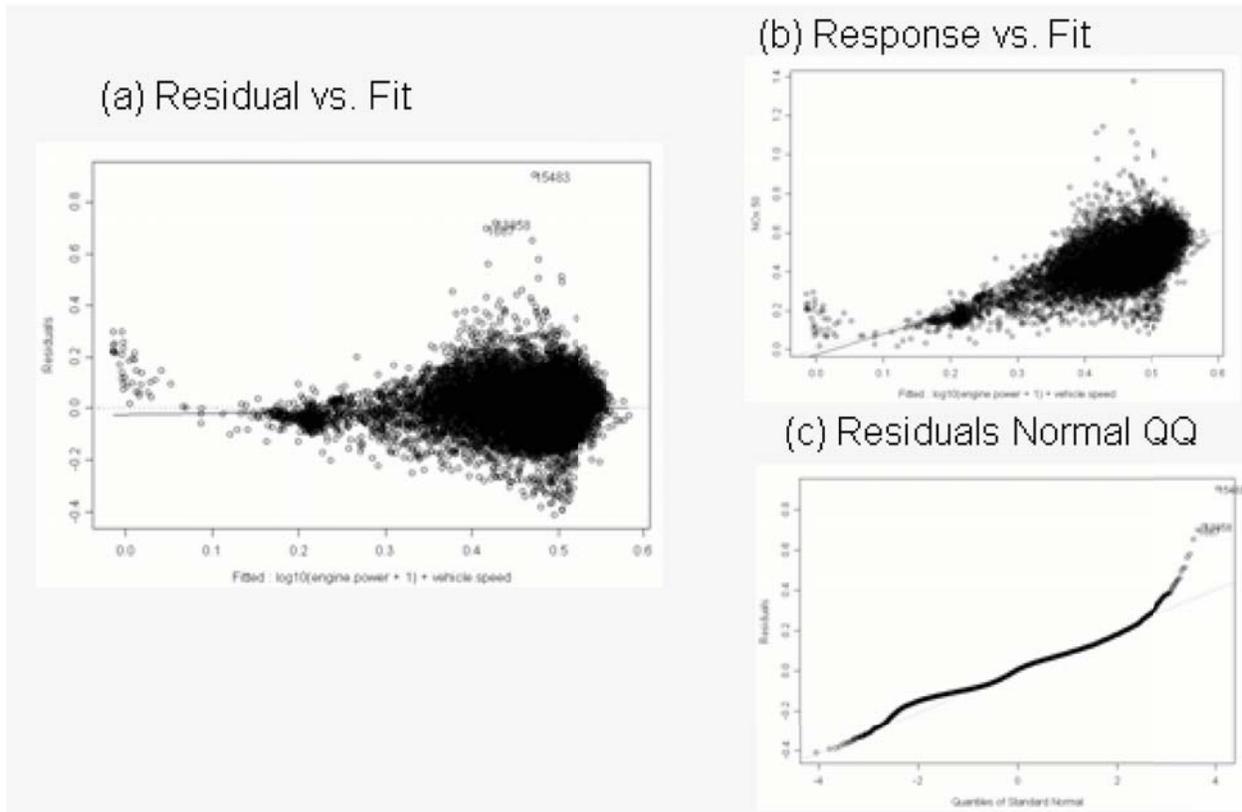


Figure 10-31 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.4

The results suggest that by using vehicle speed and transformed engine power, the model increases the amount of variance explained in truncated transformed NO<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub> 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:

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

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

Table 10-18 Regression Result for NO<sub>x</sub> 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

```

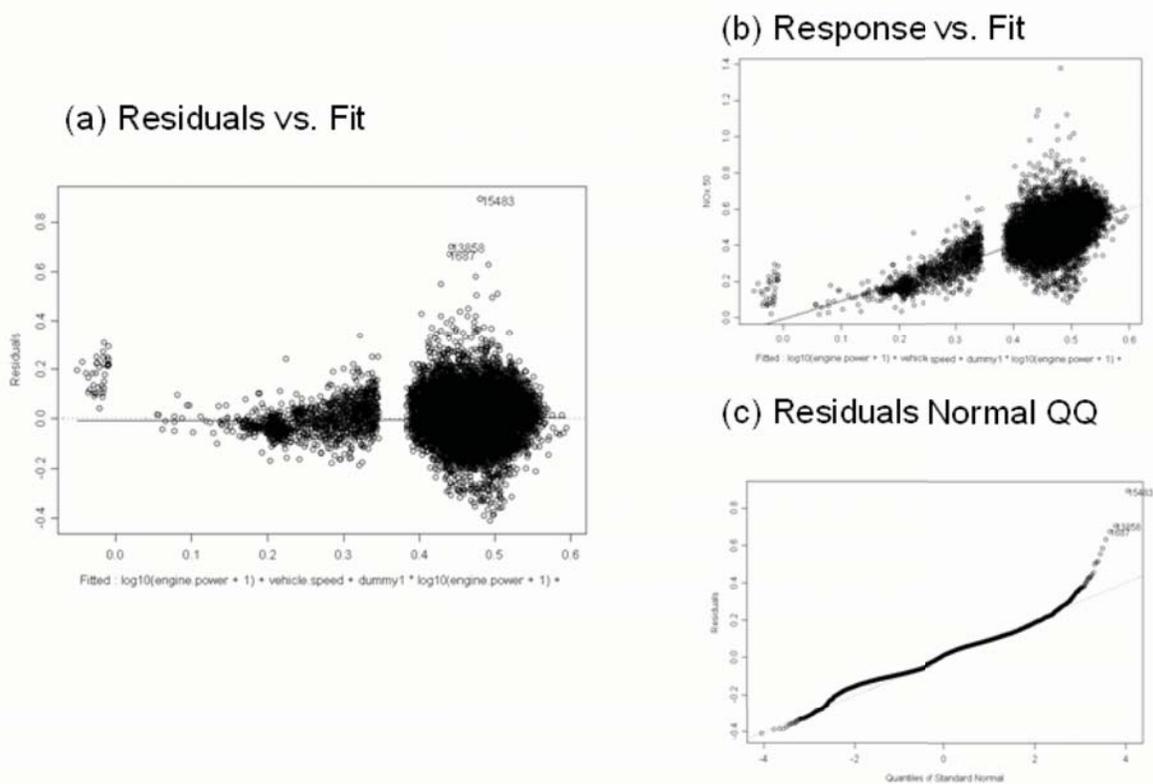


Figure 10-32 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.5

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<sub>x</sub> from about 37% (Model 1.4) to about 38% (Model 1.5).

Model 1.5 slightly improves the R<sup>2</sup> 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<sup>2</sup> value is an indication of precision. Usually, higher R<sup>2</sup> 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:

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

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

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<sub>x</sub> Emission Rate Models

	Coefficient of determination (R <sup>2</sup> )	Slope (β <sub>1</sub> )	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 <sup>5</sup> )	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<sub>x</sub> 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<sub>x</sub> 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  $\pm \sqrt{\text{MSE}}$  and about 90 percent fall between  $\pm 1.645 \sqrt{\text{MSE}}$ . 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<sub>x</sub> emission model for cruise mode is:

$$\text{NO}_x = [-0.0195 + 0.201 \log_{10}(\text{engine.power}+1) + 0.0019 \text{vehicle.speed}]^2$$

Analysis results support the observation that the final NO<sub>x</sub> 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:

$$Y = \beta_0 + \beta_1 \text{engine.power} + \text{Error} \quad (2.1)$$

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.

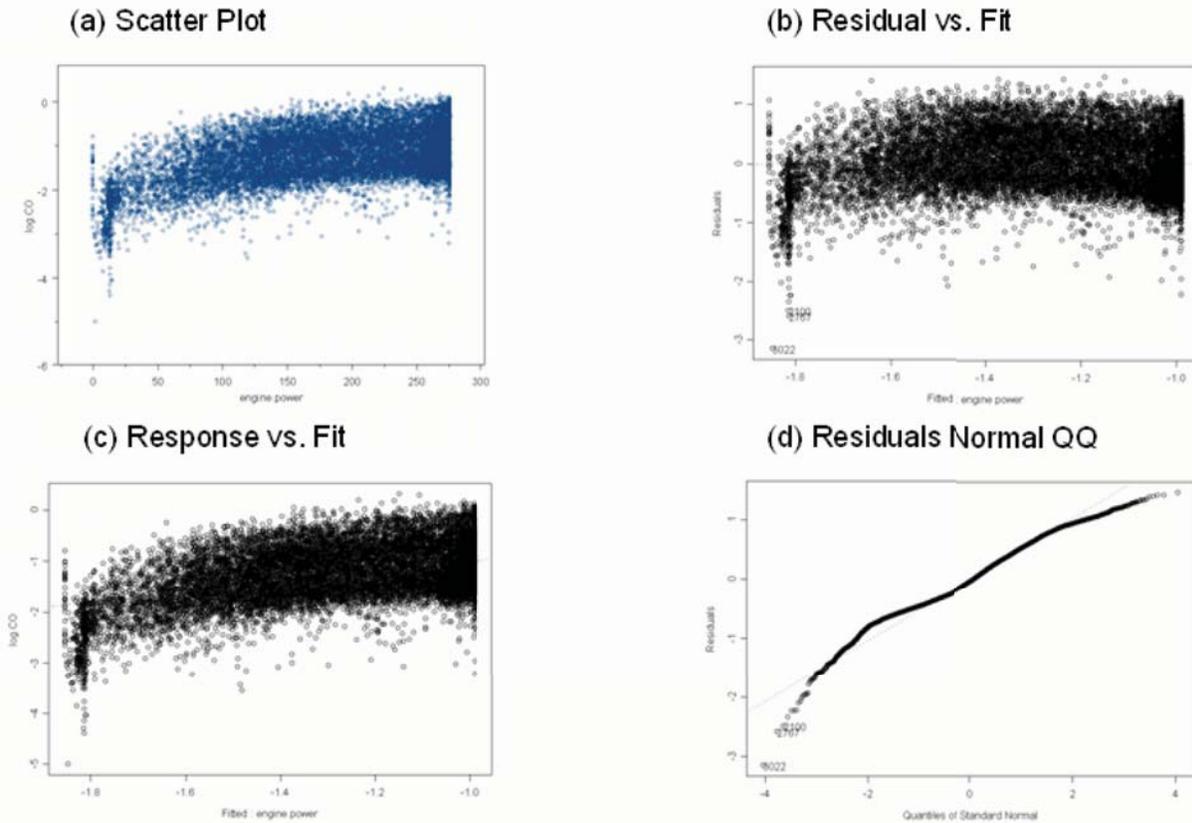


Figure 10-33 QQ and Residual vs. Fitted Plot for CO Model 2.1

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.

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

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

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
    
```

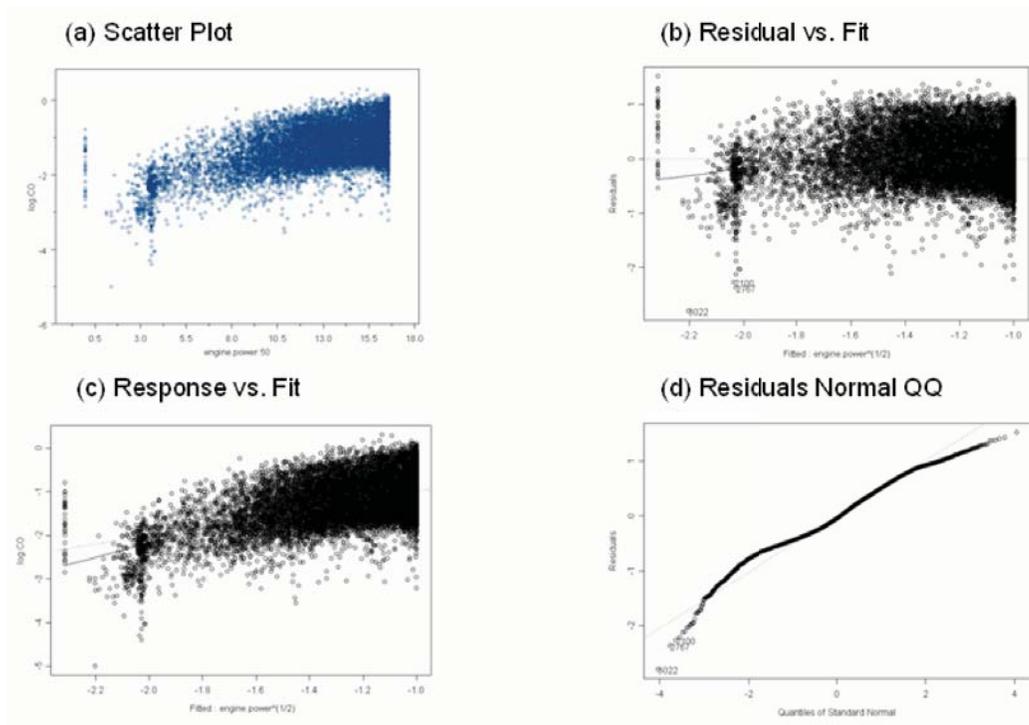


Figure 10-34 QQ and Residual vs. Fitted Plot for CO Model 2.2

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
    
```

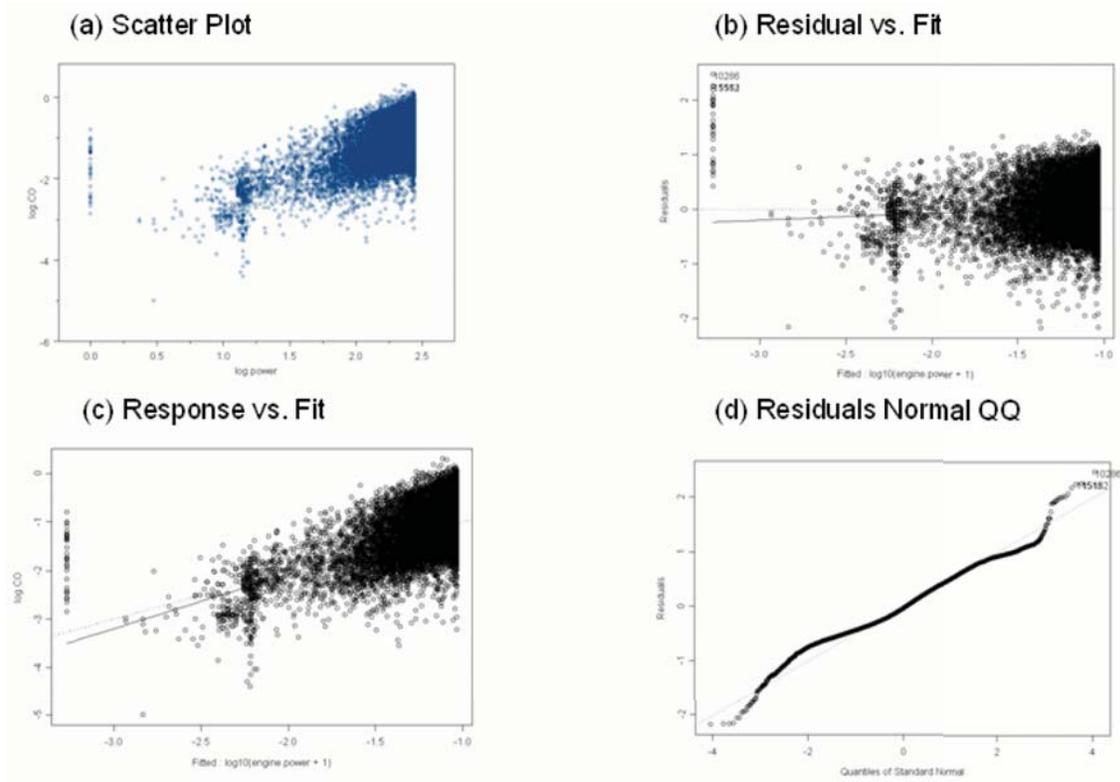


Figure 10-35 QQ and Residual vs. Fitted Plot for CO Model 2.3

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:

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

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

```

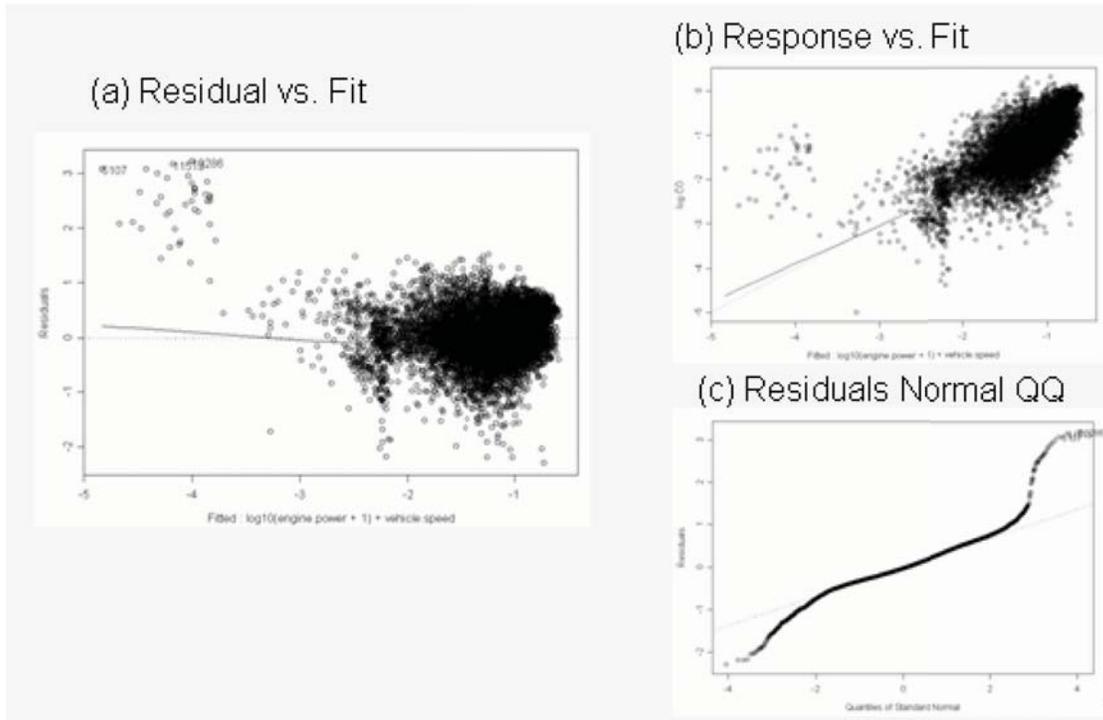


Figure 10-36 QQ and Residual vs. Fitted Plot for CO Model 2.4

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.

$$Y = \beta_0 + \beta_1 \log_{10}(\text{engine.power}+1) + \beta_2 \text{vehicle.speed} + \beta_3 \text{dummy1} + \beta_4 \text{dummy1} \log_{10}(\text{engine.power}+1) + \beta_5 \text{dummy1vehicle.speed} + \text{Error} \quad (2.5)$$

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

```

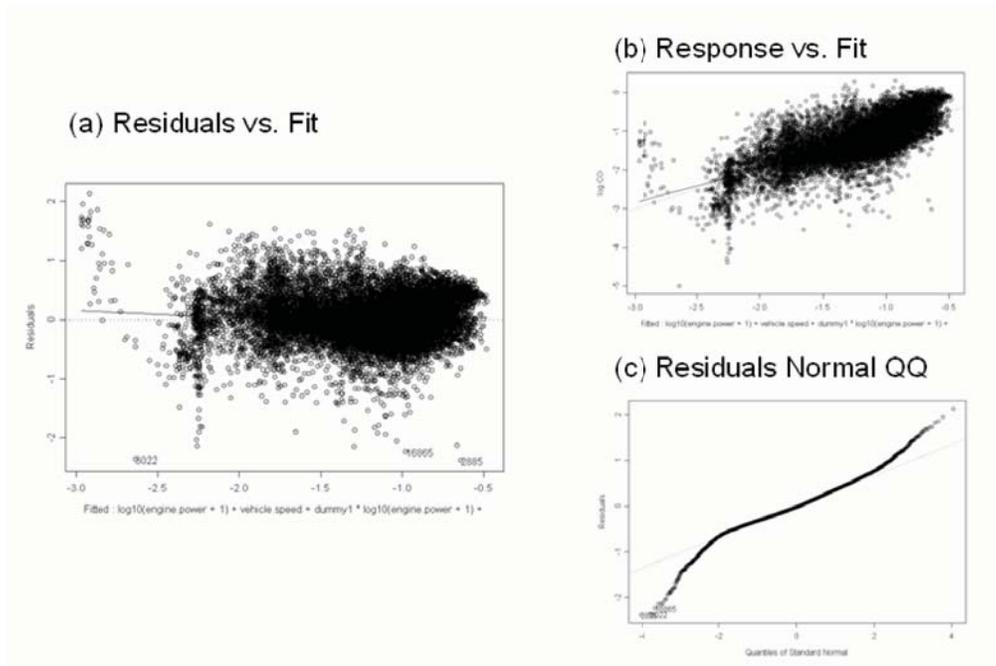


Figure 10-37 QQ and Residual vs. Fitted Plot for CO Model 2.5

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:

$$\begin{aligned}
 Y = & \beta_0 + \beta_1 \log_{10}(\text{engine.power}+1) + \beta_2 \text{vehicle.speed} + \beta_3 \text{dummy21} + \\
 & \beta_4 \text{dummy21} \log_{10}(\text{engine.power}+1) + \beta_5 \text{dummy21} \text{vehicle.speed} + \beta_6 \text{dummy22} + \\
 & \beta_7 \text{dummy22} \log_{10}(\text{engine.power}+1) + \beta_8 \text{dummy22} \text{vehicle.speed} + \beta_9 \text{dummy23} + \\
 & \beta_{10} \text{dummy23} \log_{10}(\text{engine.power}+1) + \beta_{11} \text{dummy23} \text{vehicle.speed} + \text{Error}
 \end{aligned} \tag{2.6}$$

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) + dummy22 * vehicle.speed + dummy23 * 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:
              (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

```

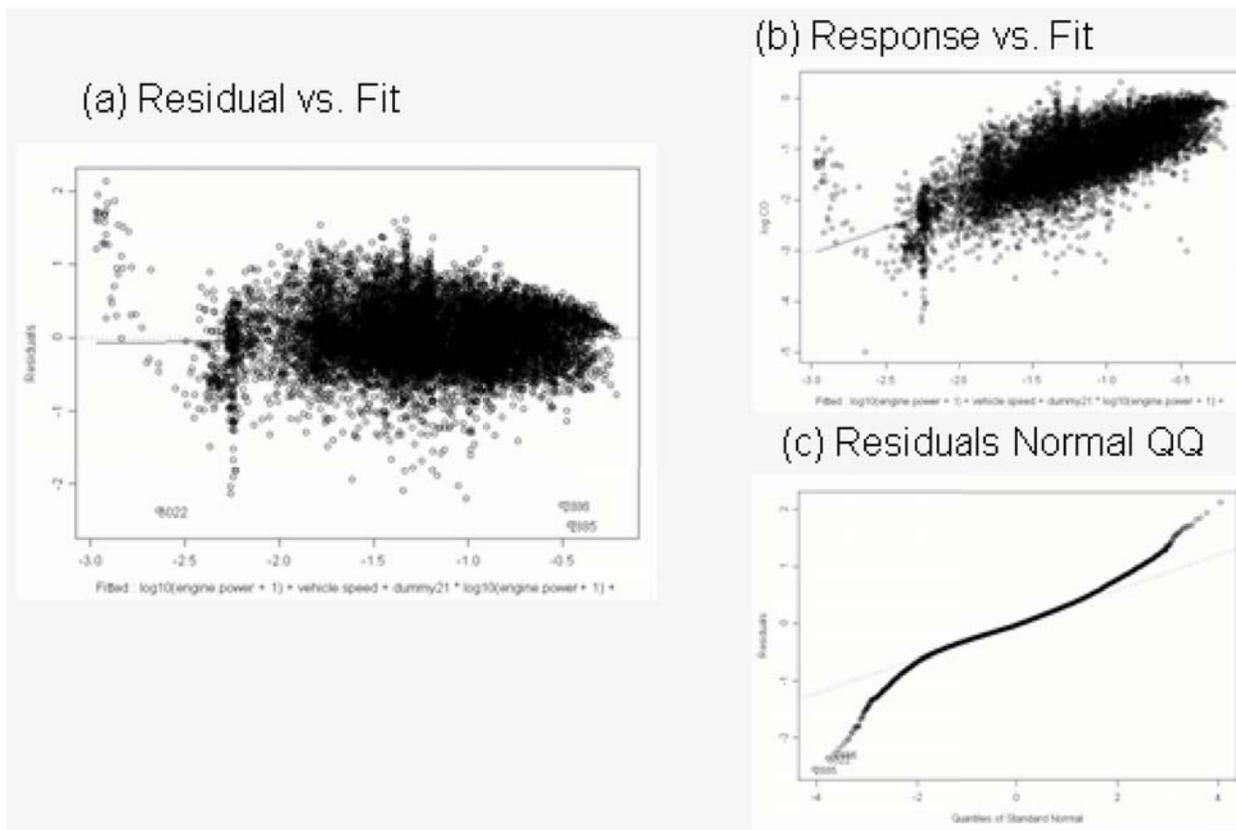


Figure 10-38 QQ and Residual vs. Fitted Plot for CO Model 2.6

Model 2.6 does improve the ability to explain variance by another 4% ( $R^2$  increases from 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 ( $\beta_1$ )	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 <sup>0.5</sup> )	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:

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

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:

$$Y = \beta_0 + \beta_1 \text{engine.power} + \text{Error} \quad (3.1)$$

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

```

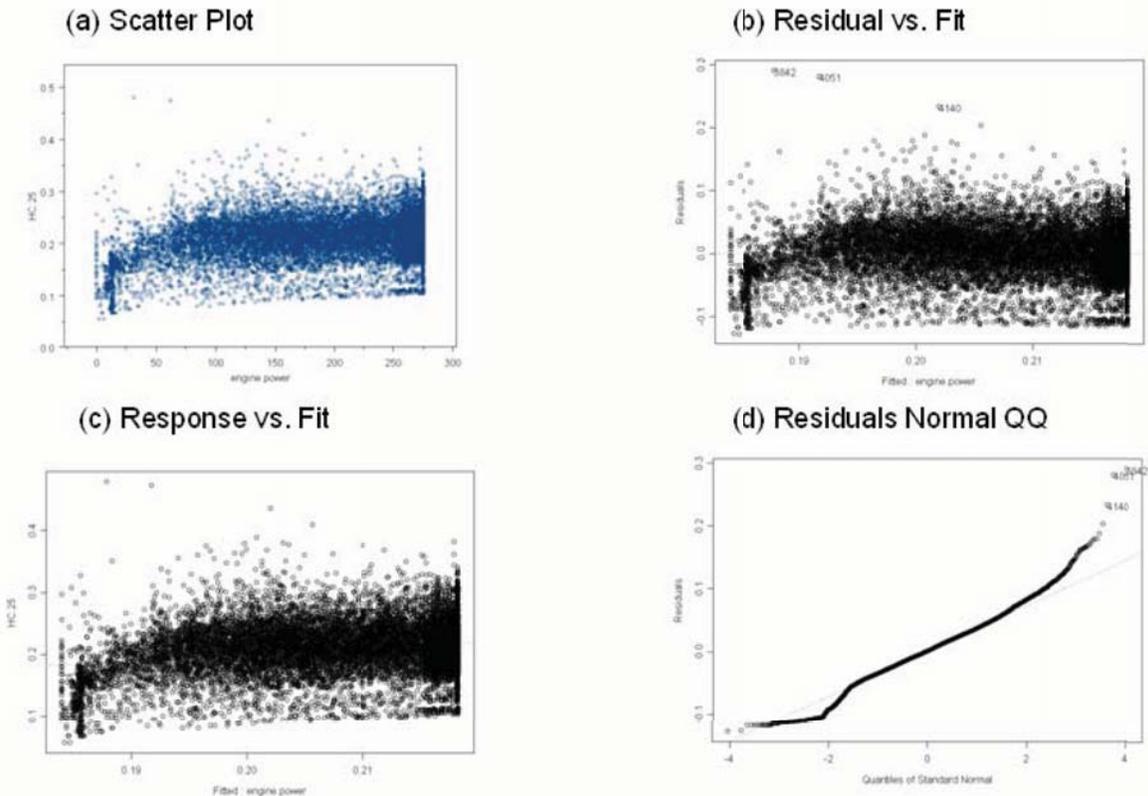


Figure 10-39 QQ and Residual vs. Fitted Plot for HC Model 3.1

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.

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

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

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
    
```

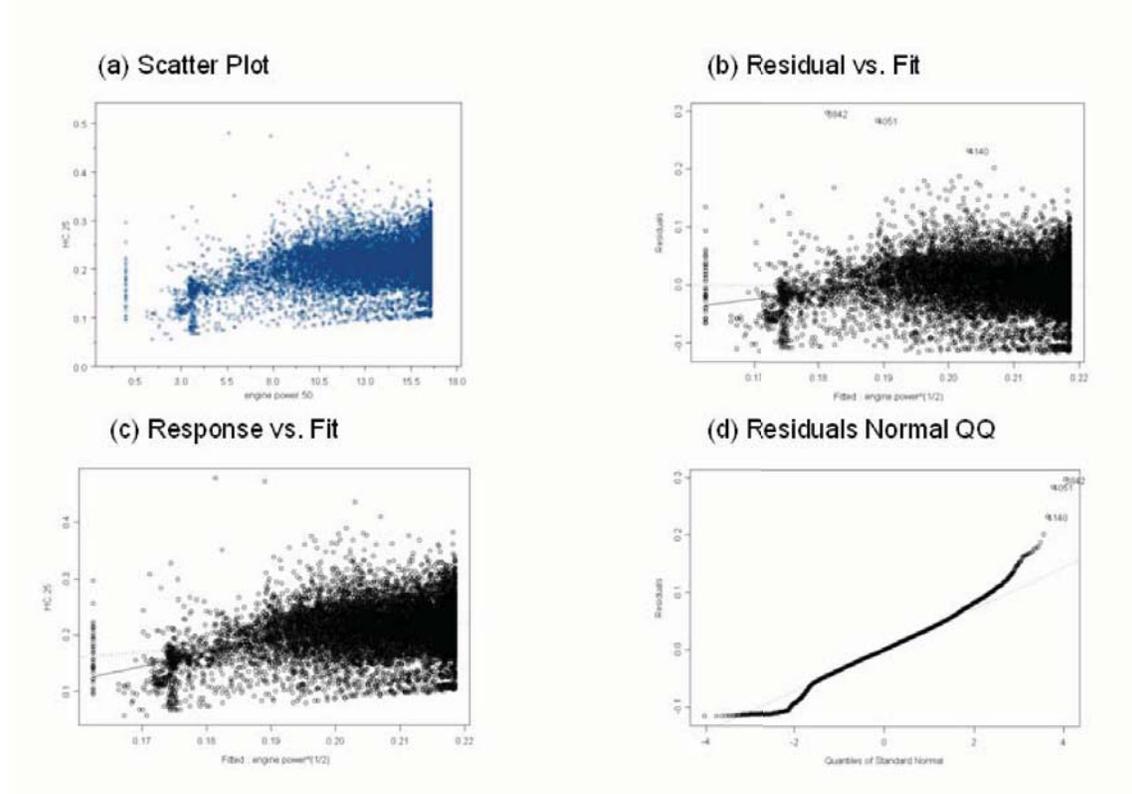


Figure 10-40 QQ and Residual vs. Fitted Plot for HC Model 3.2

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
    
```

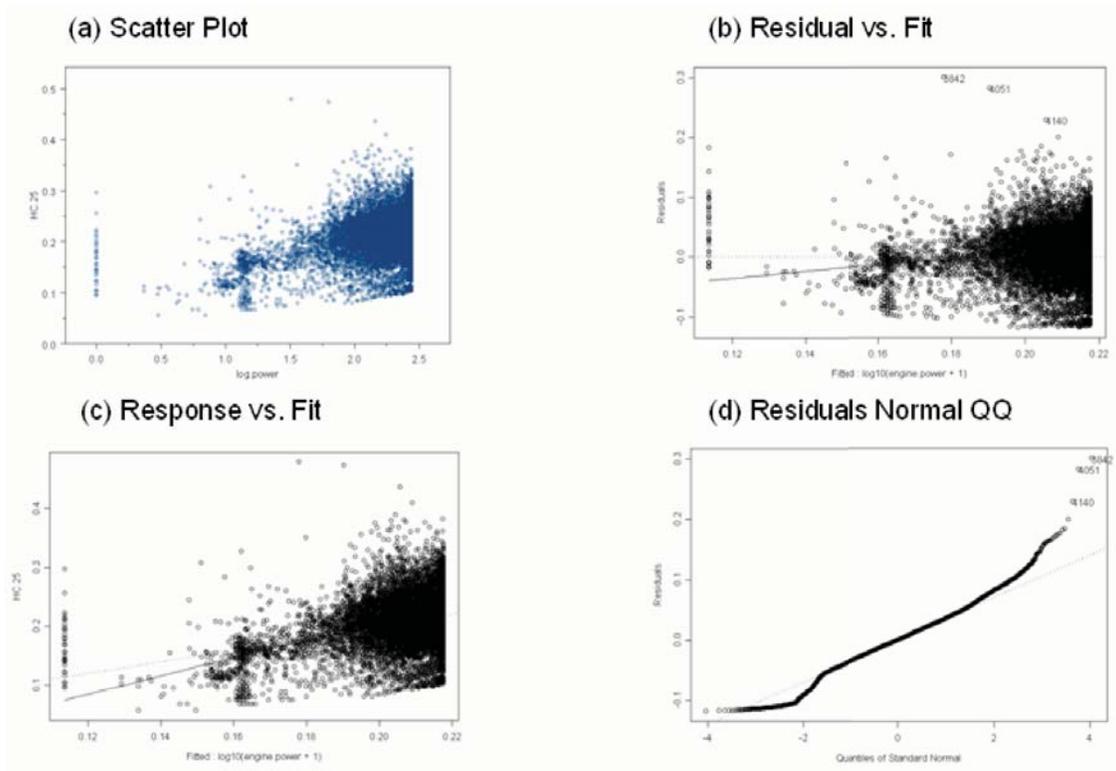


Figure 10-41 QQ and Residual vs. Fitted Plot for HC Model 3.3

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.

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

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

```

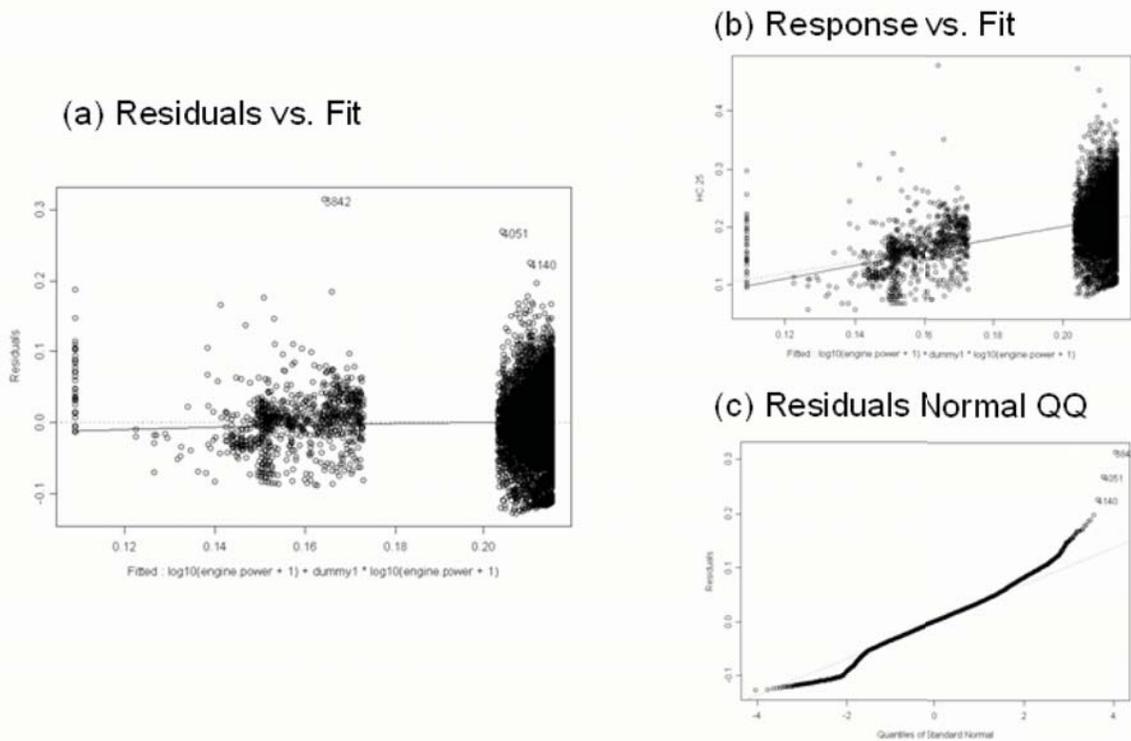


Figure 10-42 QQ and Residual vs. Fitted Plot for HC Model 3.4

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 ( $\beta_1$ )	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 <sup>0.5</sup> )	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:

$$\text{NO}_x = [-0.0195 + 0.201\log_{10}(\text{engine.power}+1) + 0.0019\text{vehicle.speed}]^2$$

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

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

## 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					All
		(0 50)	(50 100)	(100 150)	(150 200)	≥ 200	
Number	NO <sub>x</sub>	15885	8988	7173	3536	3792	39374
	CO	15834	8940	7145	3529	3770	39218
	HC	15481	8600	6830	3394	3715	38020
Percentage	NO <sub>x</sub>	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.

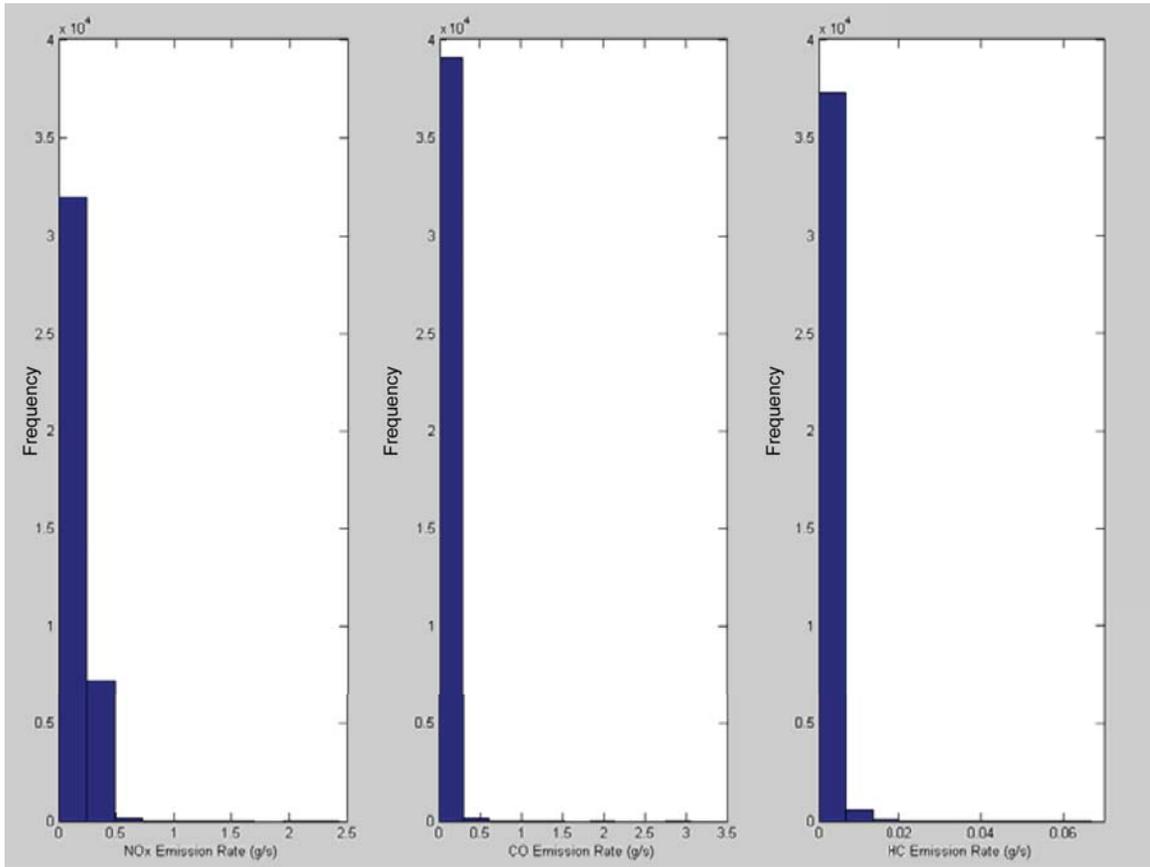


Figure 11-1 Histograms of Three Pollutants for Cruise Mode

### 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  $\text{NO}_x$ , 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.

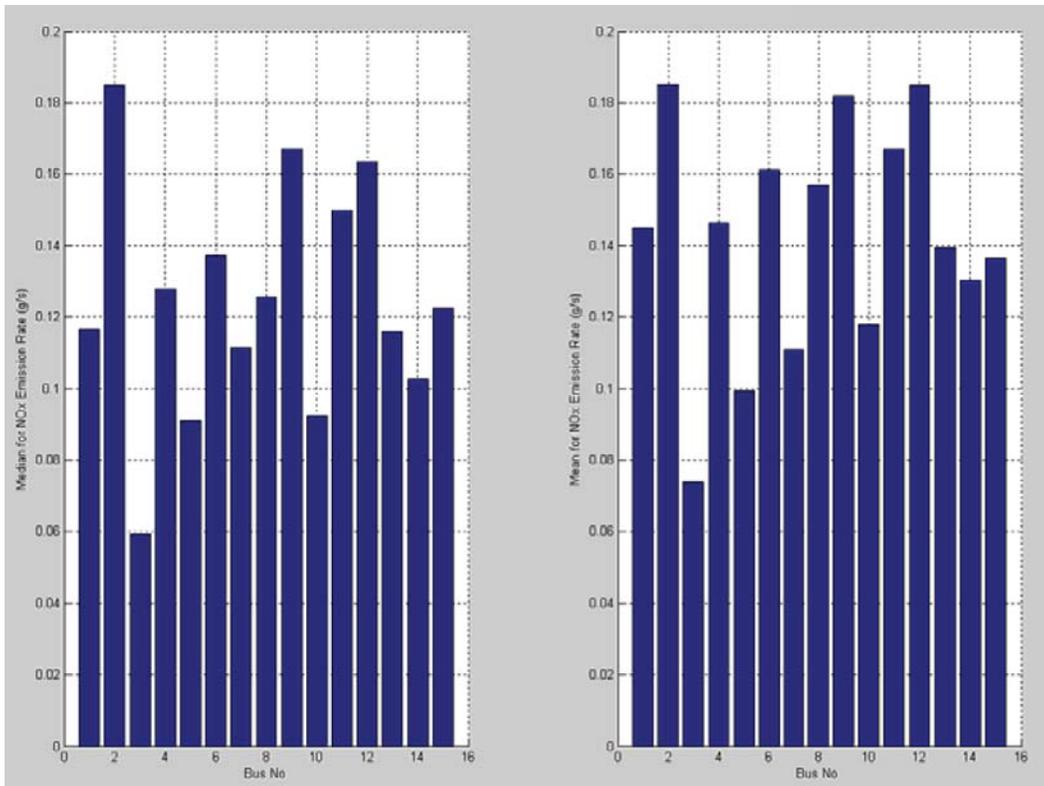


Figure 11-2 Median and Mean of NO<sub>x</sub> Emission Rates in Cruise Mode by Bus

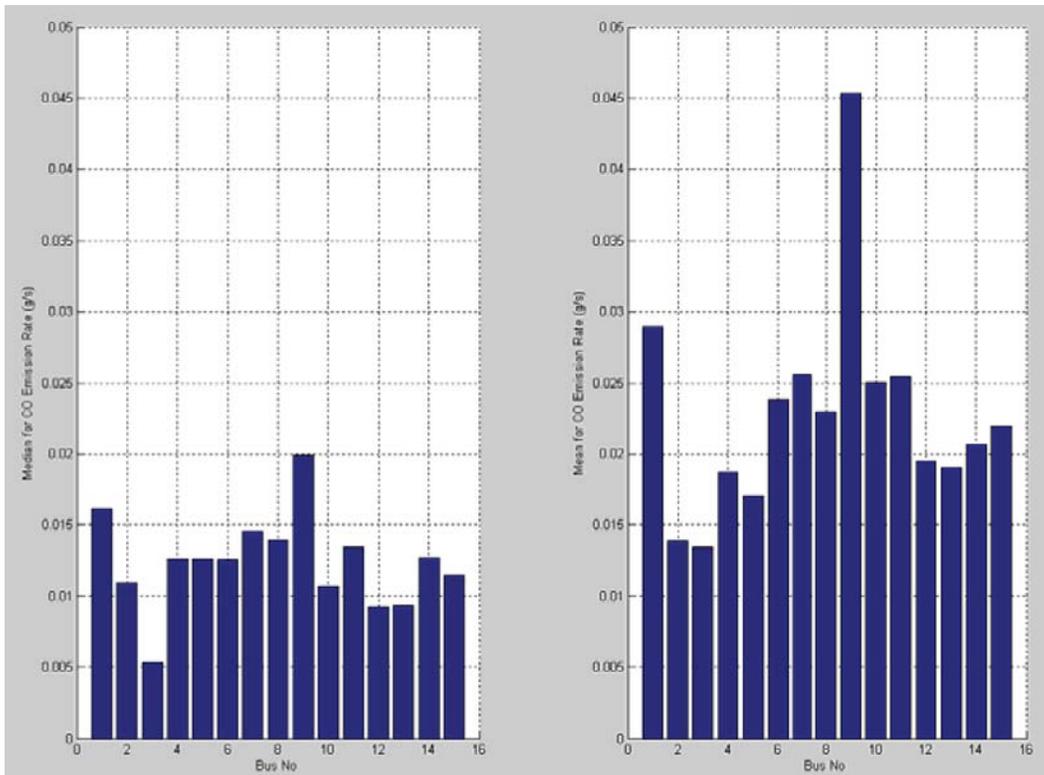


Figure 11-3 Median and Mean of CO Emission Rates in Cruise Mode by Bus

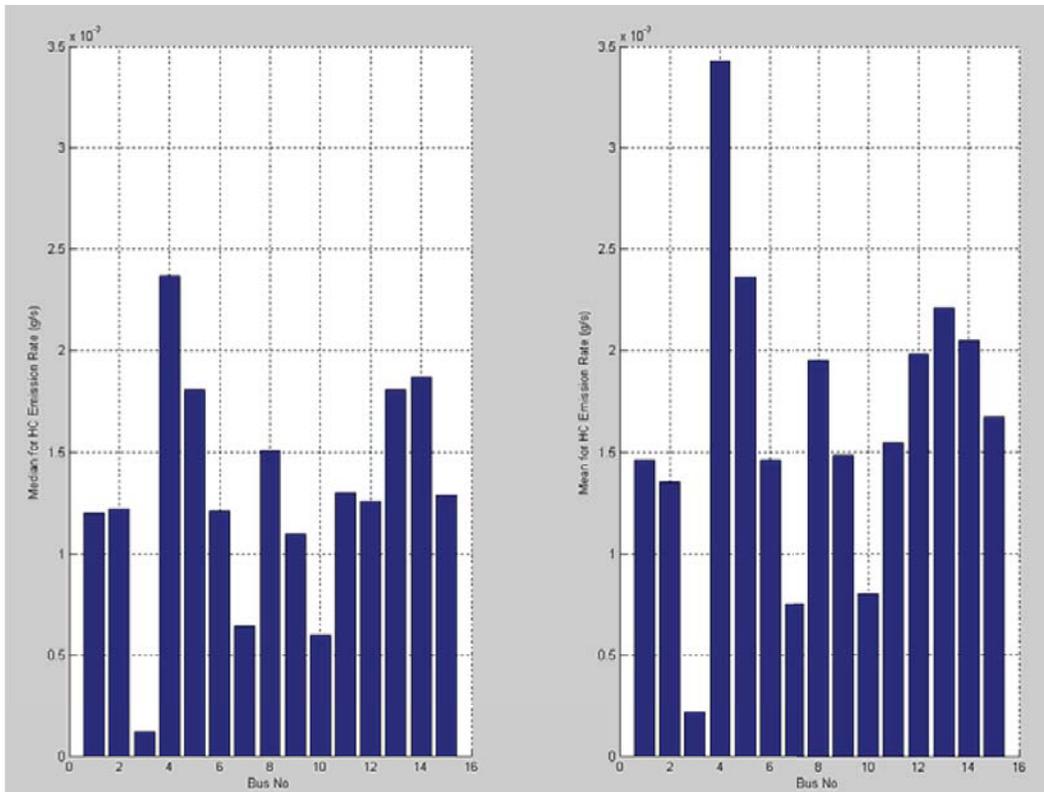


Figure 11-4 Median and Mean of HC Emission Rates in Cruise Mode by Bus

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

Bus ID	NO <sub>x</sub>		CO		HC	
	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<sub>x</sub> 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 1<sup>st</sup> quartile engine power in cruise mode while Bus 377 has the largest median and 3<sup>rd</sup> 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

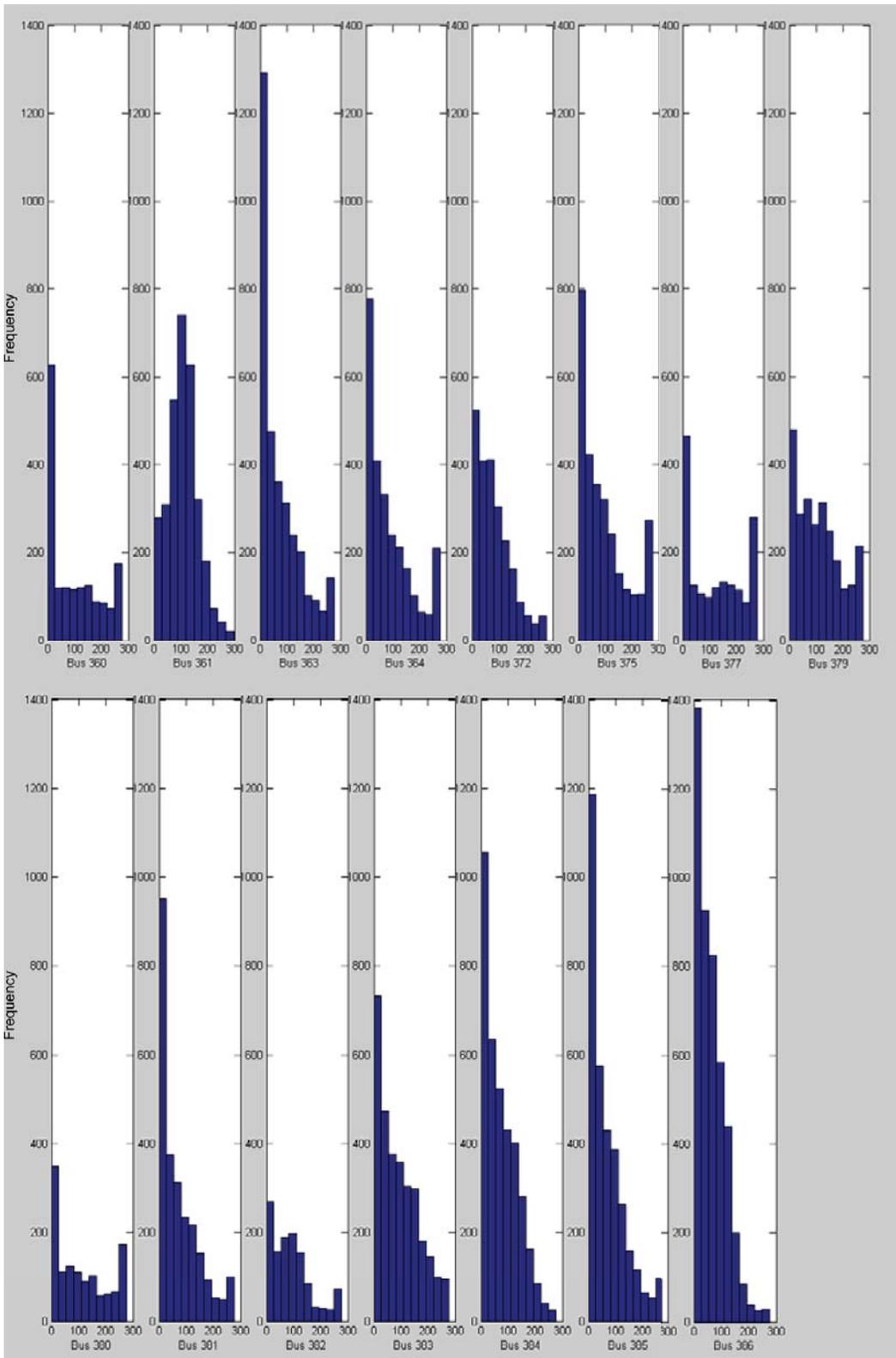


Figure 11-5 Histograms of Engine Power in Cruise Mode by Bus

## 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<sub>x</sub>, 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<sub>x</sub>, 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<sub>x</sub>, CO, and HC, while Figures 11-9 to 11-11 pres-

ent the same plots for truncated transformed emission rates for  $\text{NO}_x$ , CO and HC, where a great improvement is noted.

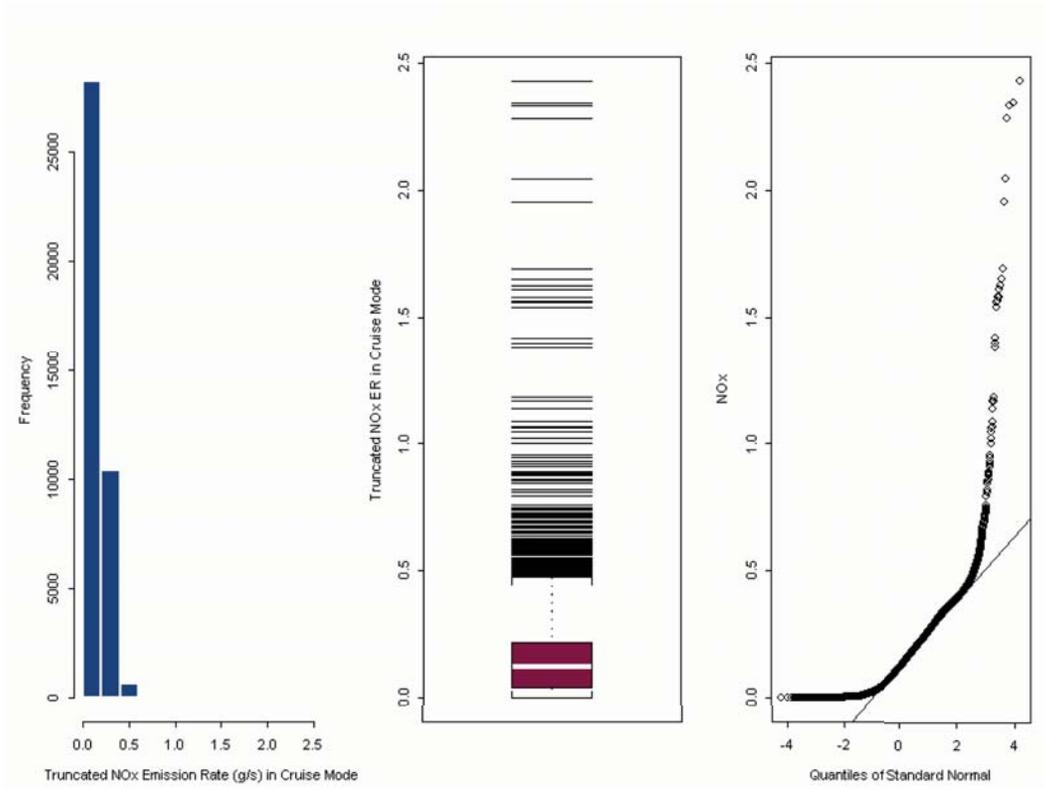


Figure 11-6 Histogram, Boxplot, and Probability Plot of Truncated  $\text{NO}_x$  Emission Rates in Cruise Mode

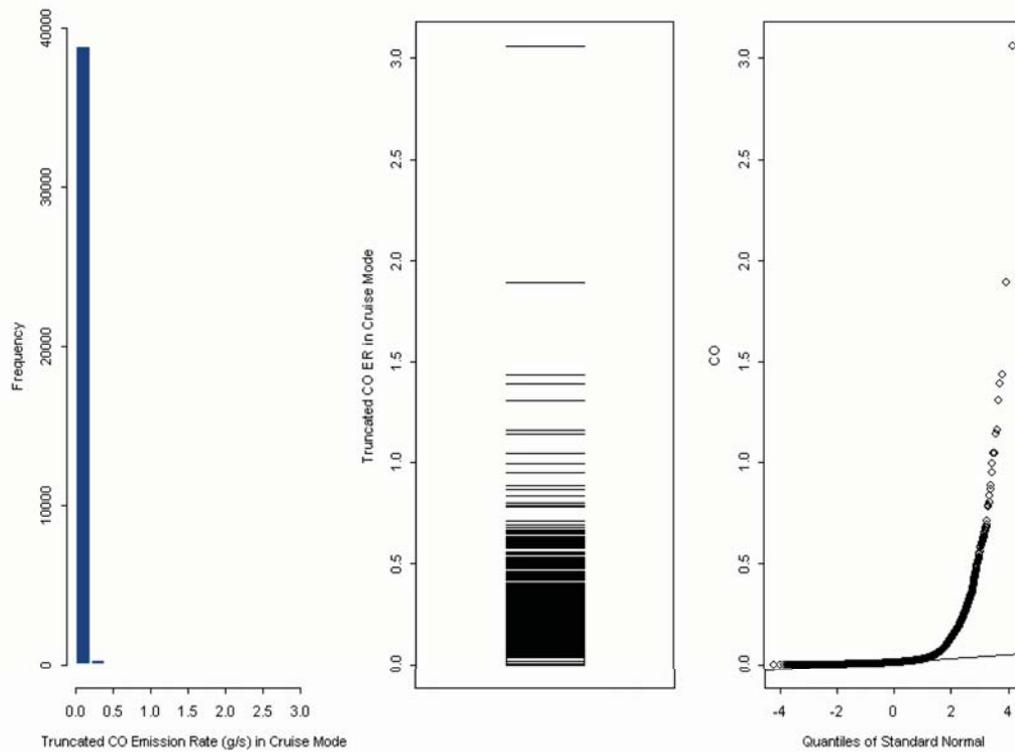


Figure 11-7 Histogram, Boxplot, and Probability Plot of Truncated CO Emission Rate in Cruise Mode

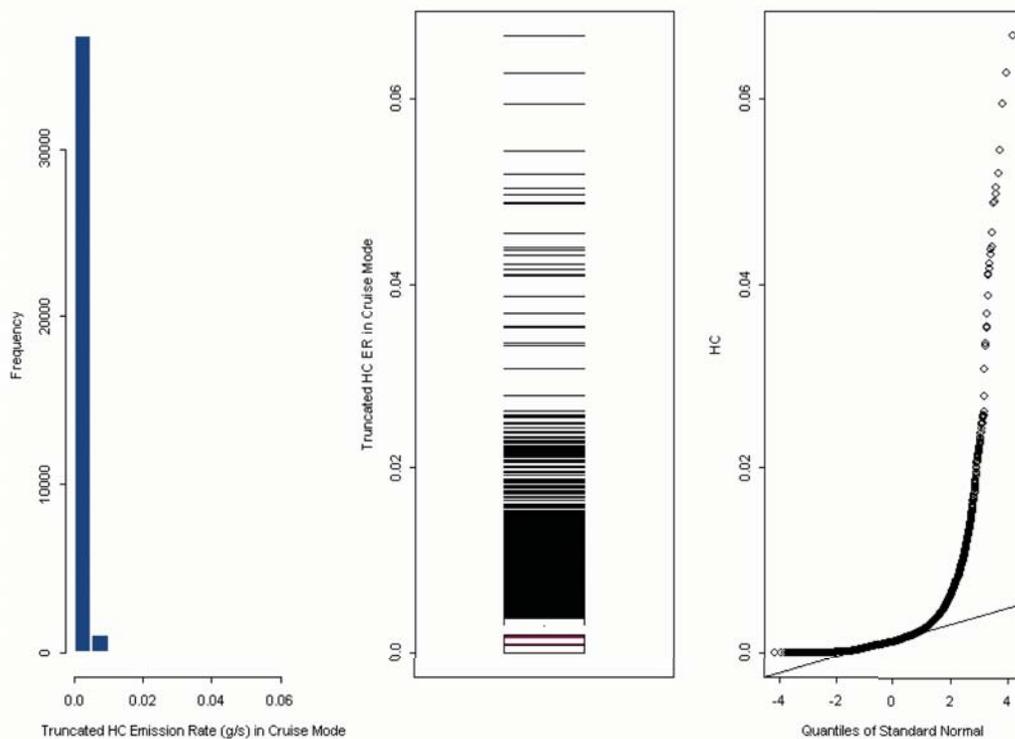


Figure 11-8 Histogram, Boxplot, and Probability Plot of Truncated HC Emission Rate in Cruise Mode

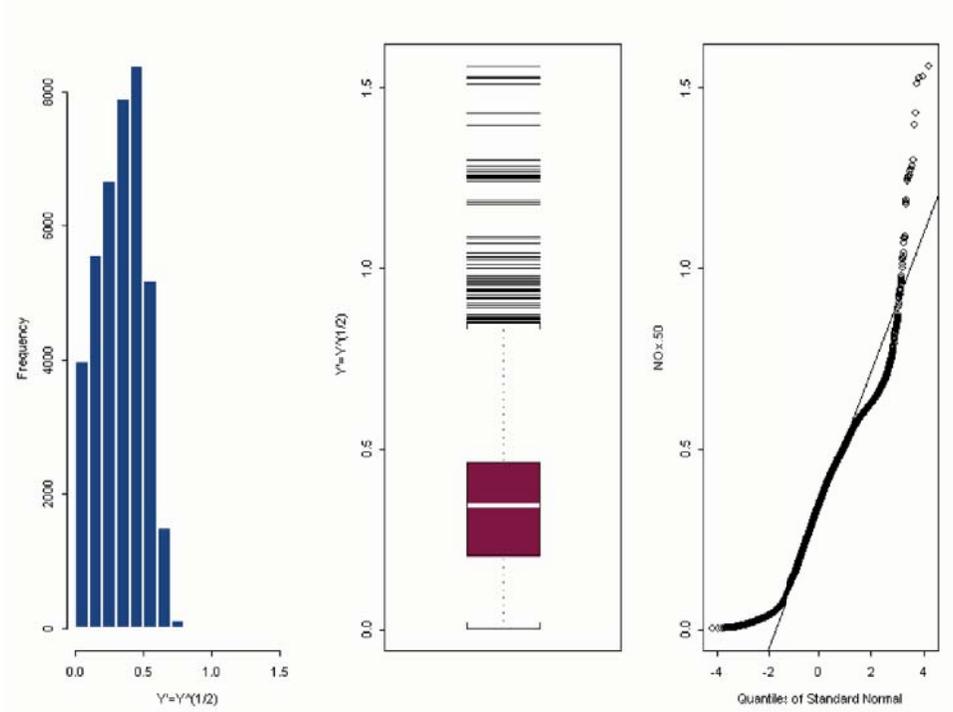


Figure 11-9 Histogram, Boxplot, and Probability Plot of Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

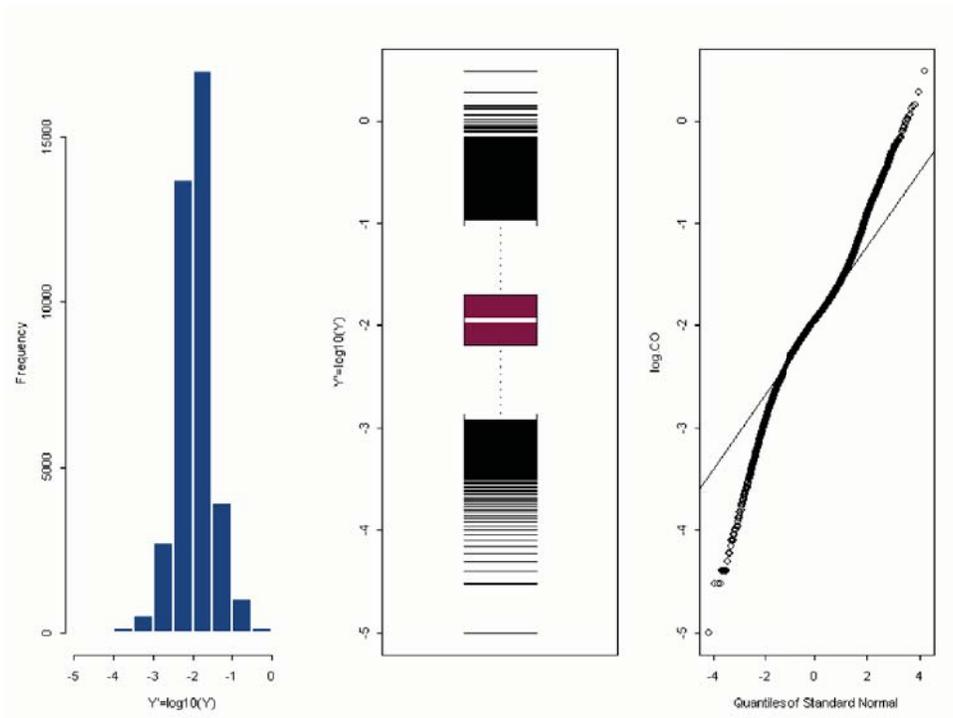


Figure 11-10 Histogram, Boxplot, and Probability Plot of Truncated Transformed CO Emission Rate in Cruise Mode

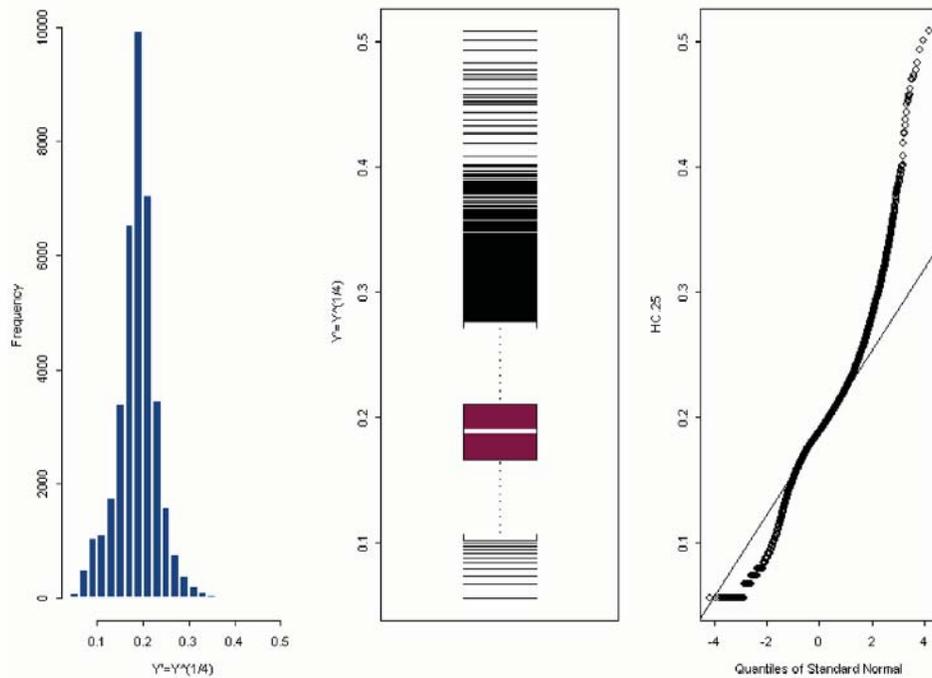


Figure 11-11 Histogram, Boxplot, and Probability Plot of Truncated Transformed HC Emission Rate in Cruise Mode

#### 11.2.1.1 $\Delta$ HTBR Tree Model Development

Figure 11-12 illustrates the initial tree model used for the truncated transformed  $\text{NO}_x$  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.

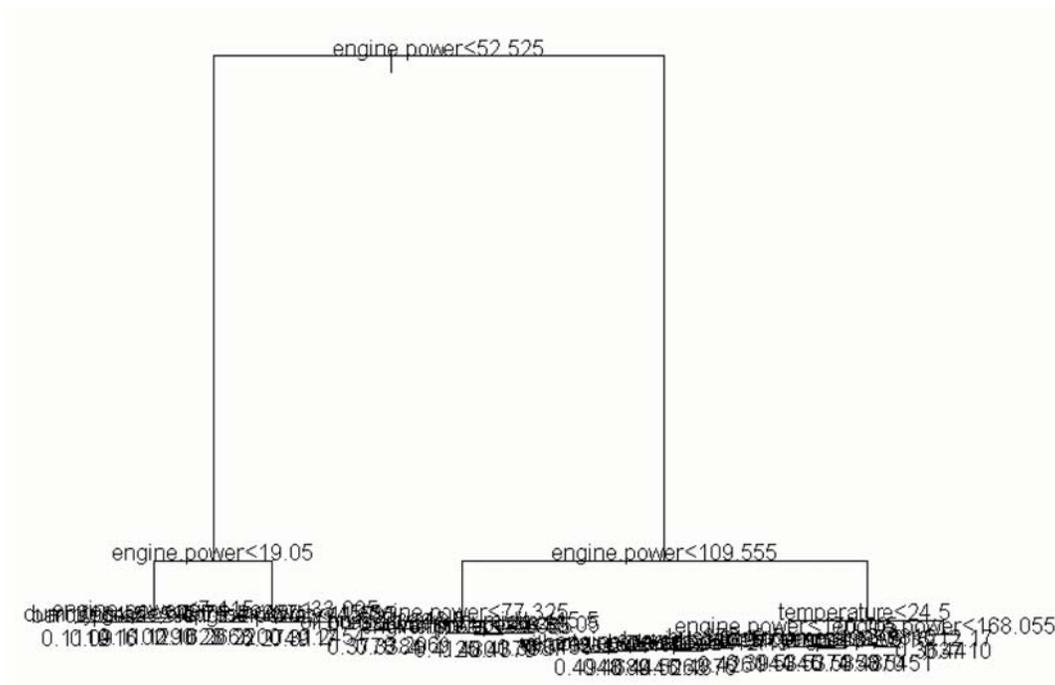


Figure 11-12 Original Untrimmed Regression Tree Model for Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

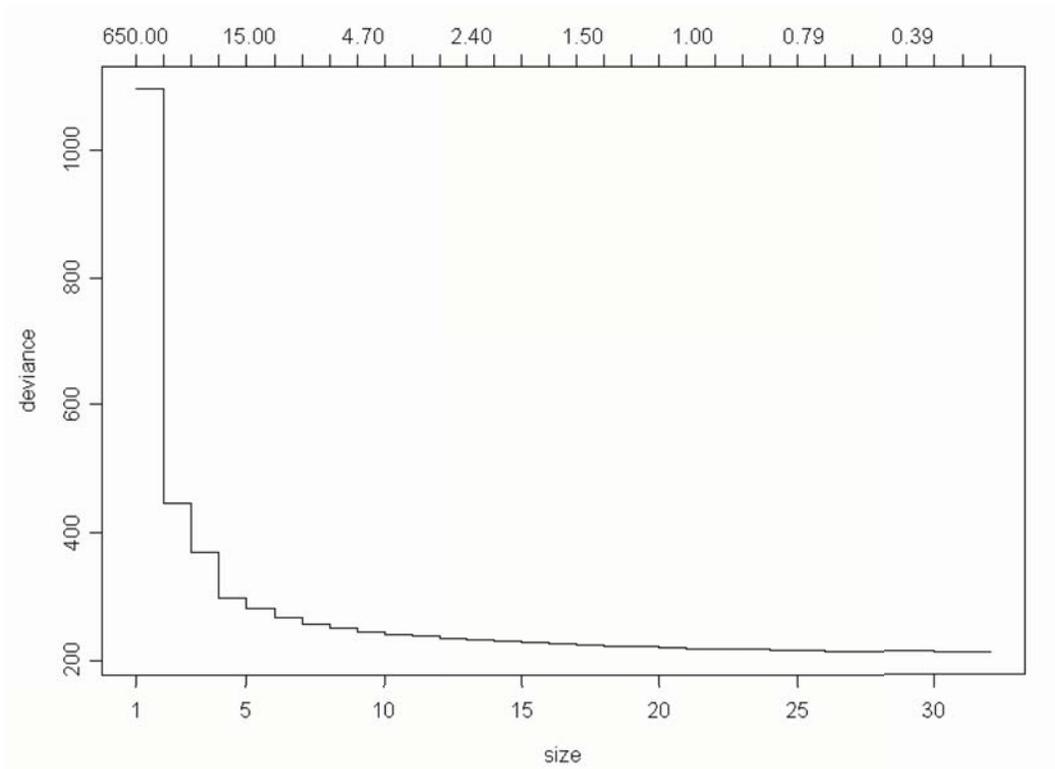


Figure 11-13 Reduction in Deviance with the Addition of Nodes of Regression Tree for Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

Table 11-4 Original Untrimmed Regression Tree Results for Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

```

Regression tree:
tree(formula = NOx.50 ~ model.year + odometer + temperature + baro + humidity +
vehicle.speed + oil.temperature + 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<sub>x</sub> cruise model will be developed based on this result.

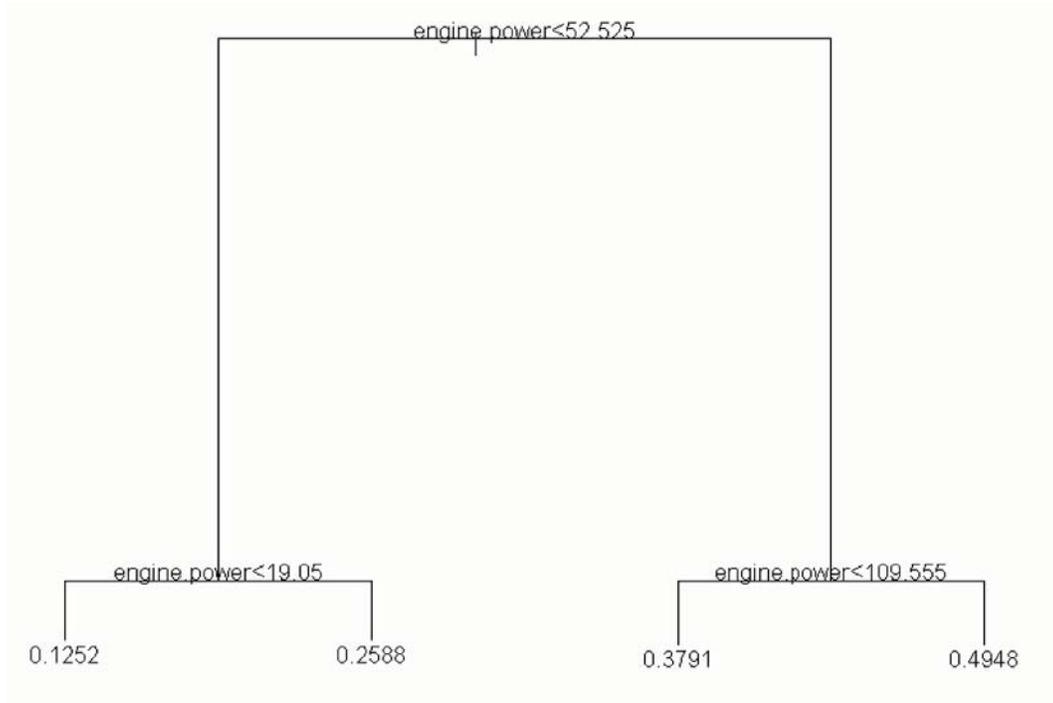


Figure 11-14 Trimmed Regression Tree Model for Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

Table 11-5 Trimmed Regression Tree Results for Truncated Transformed NO<sub>x</sub> Emission Rate in Cruise Mode

```

Regression tree:
snip.tree(tree = tree(formula = NOx.50 ~ model.year + odometer + temperature + baro
+ humidity + vehicle.speed + oil.temperature + 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<sub>x</sub> emissions. This finding is consistent with previous research results which verified the important effect of engine power on NO<sub>x</sub> 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.

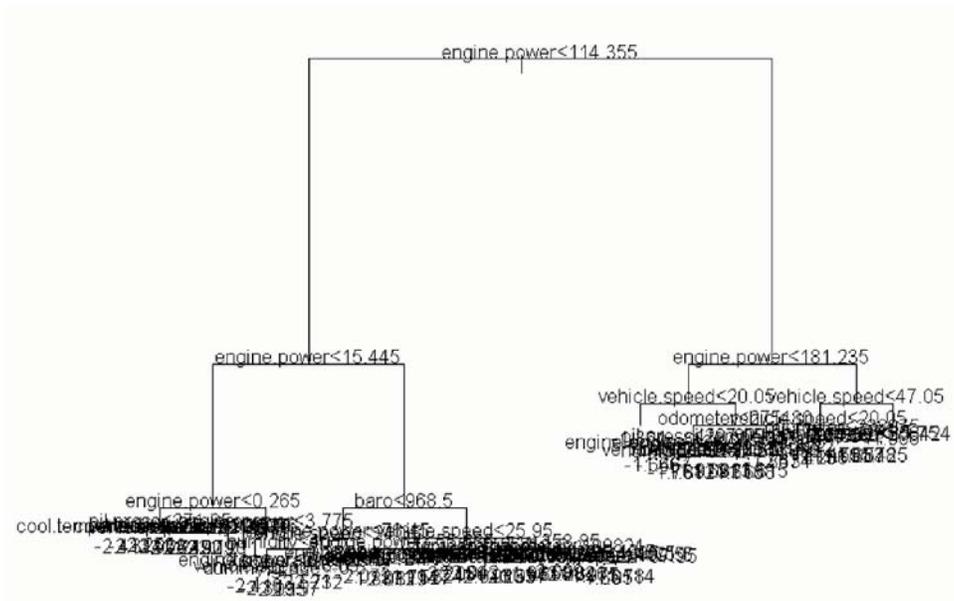


Figure 11-15 Original Untrimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Cruise Mode

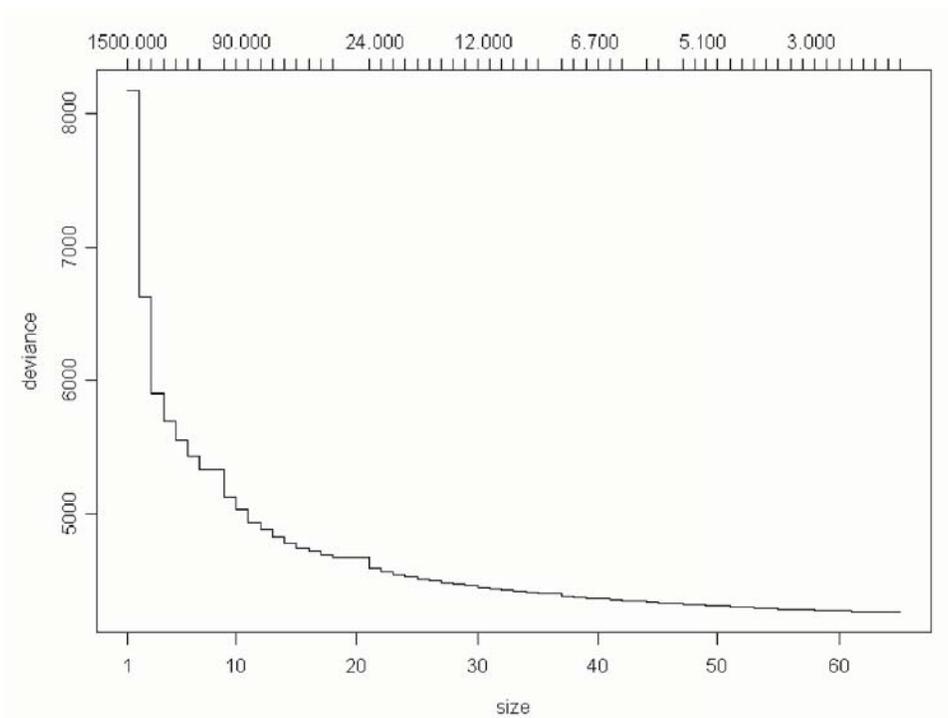


Figure 11-16 Reduction in Deviance with the Addition of Nodes of Regression Tree for Truncated Transformed CO Emission Rate in Cruise Mode

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.

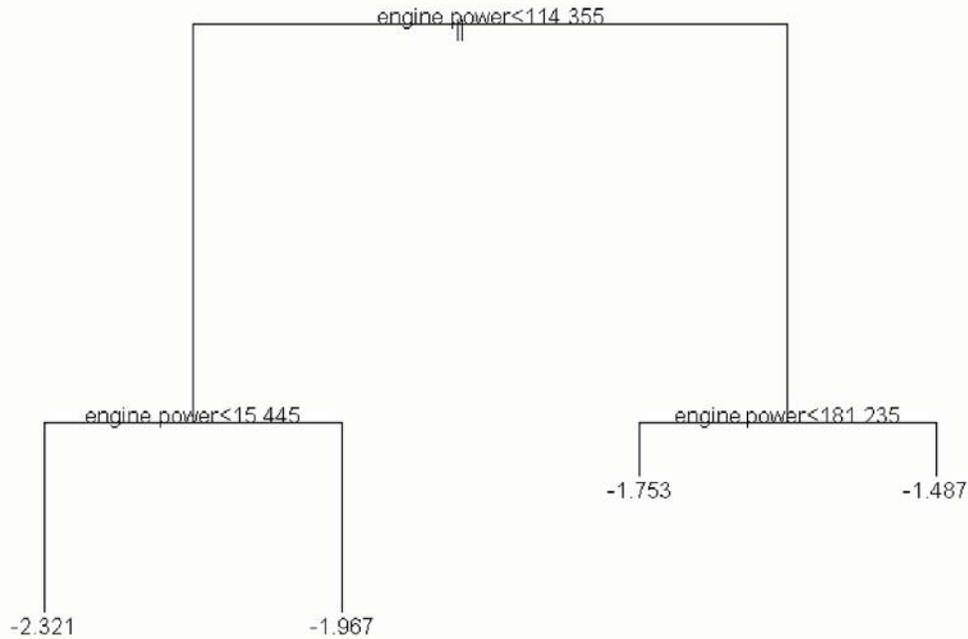


Figure 11-17 Trimmed Regression Tree Model for Truncated Transformed CO Emission Rate in Cruise Mode

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 *
  
```



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.temperure + 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.temperure" "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 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 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<sub>x</sub> and CO models.

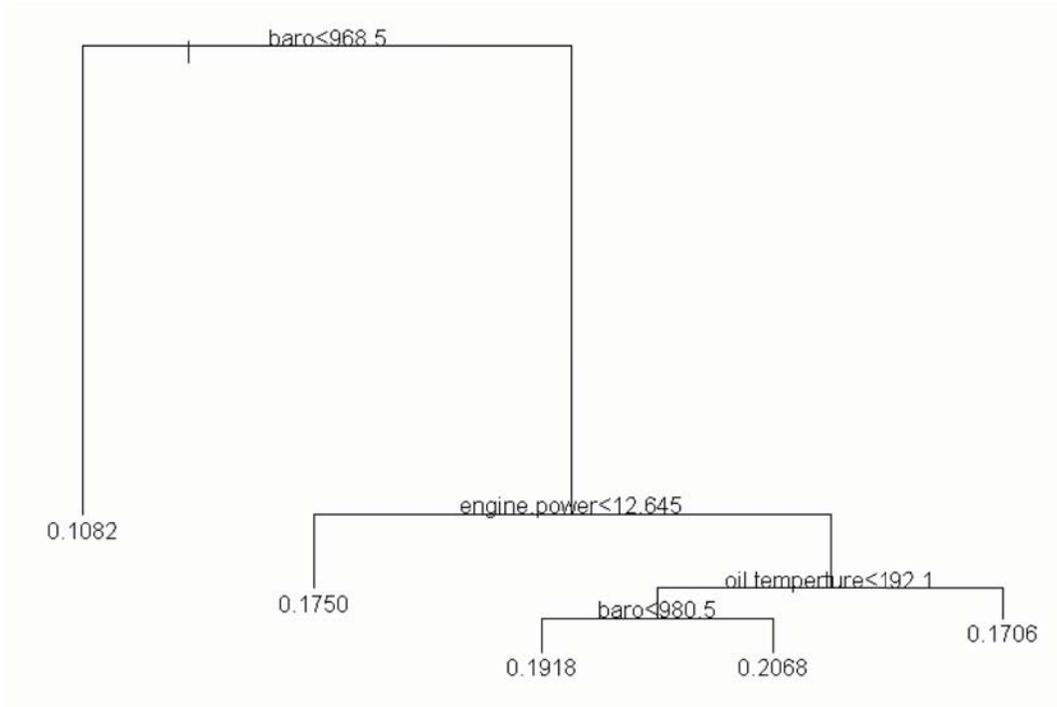


Figure 11-19 Trimmed Regression Tree Model for Truncated Transformed HC Emission Rate in Cruise Mode

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.temperature + 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.temperature"
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.temperature<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.temperature>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.

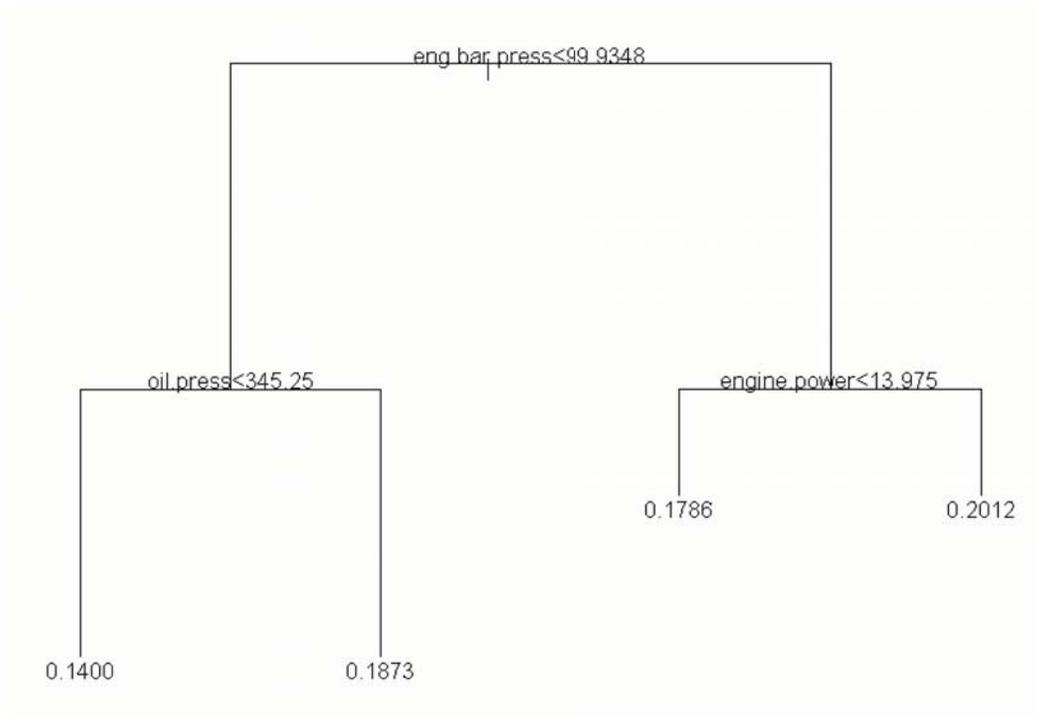


Figure 11-20 Secondary Trimmed Regression Tree Model for Truncated Transformed HC in Cruise Mode

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.temperature + 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<sub>x</sub> 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.

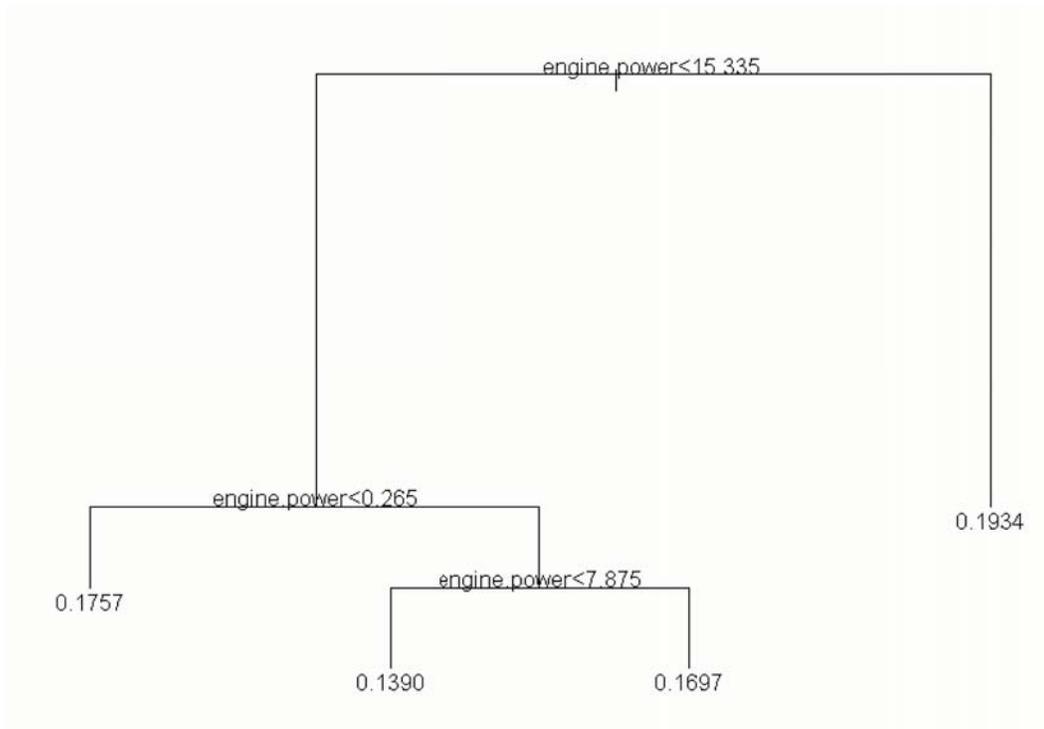


Figure 11-21 Final Regression Tree Model for Truncated Transformed HC and Engine Power in Cruise Mode

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<sub>x</sub> Emission Rate Model Development for Cruise Mode

Based on previous analysis, truncated transformed NO<sub>x</sub> 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:

$$Y = \beta_0 + \beta_1(\text{engine.power}) + \text{Error} \quad (1.1)$$

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

Table 11-12 Regression Result for NO<sub>x</sub> 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<sub>x</sub>. 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.

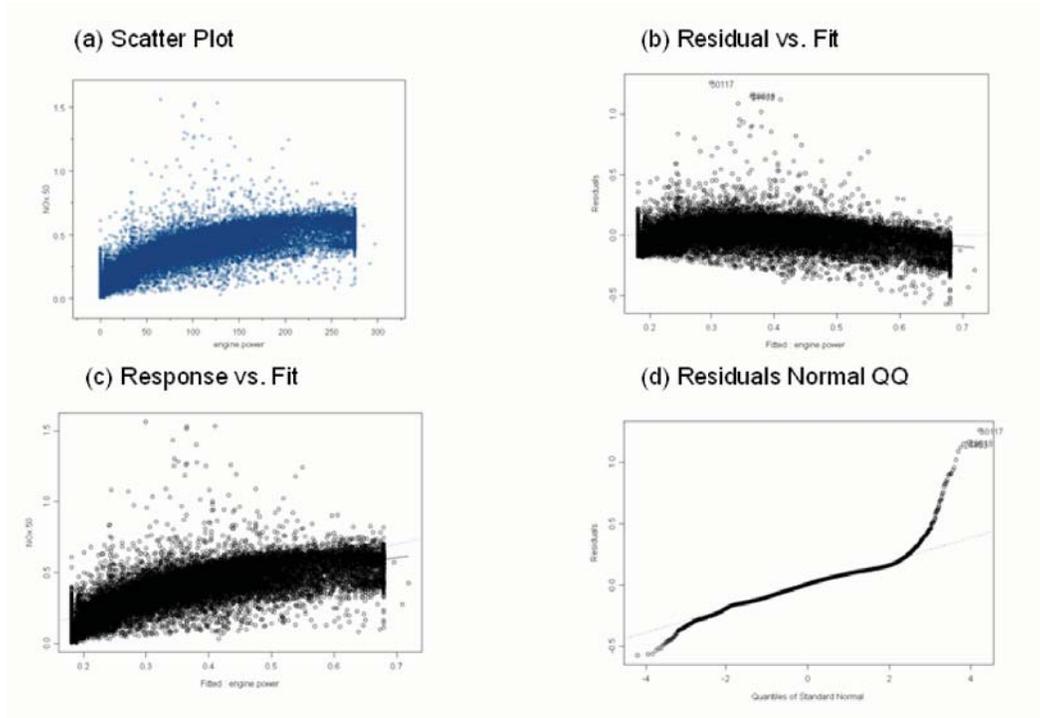


Figure 11-22 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.1

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.

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

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

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<sub>x</sub> 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
    
```

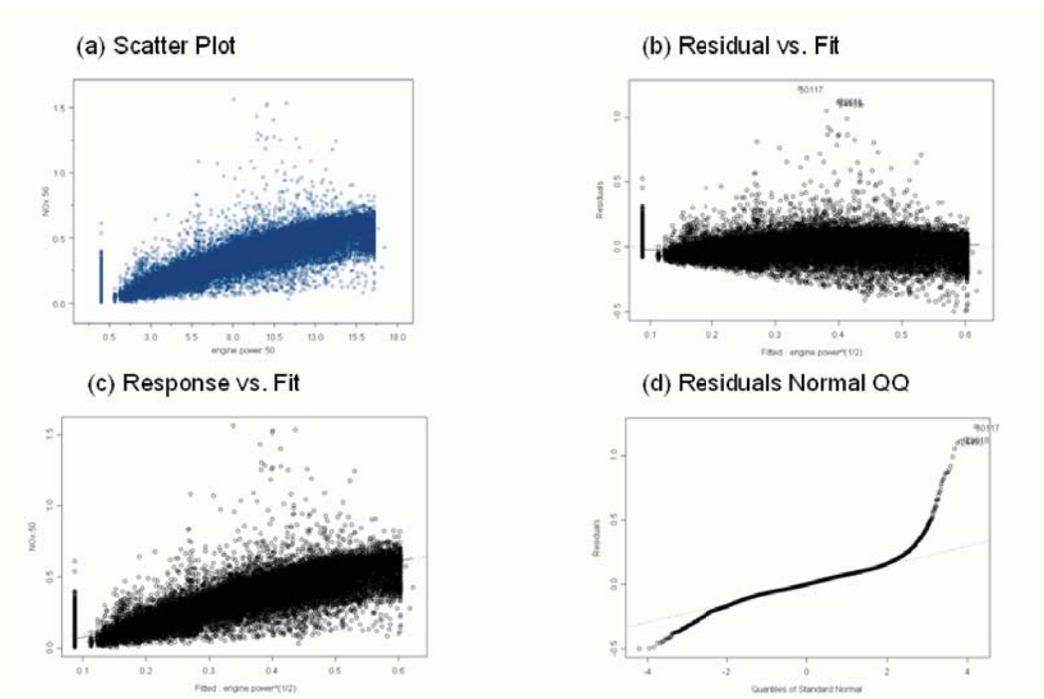


Figure 11-23 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.2

Table 11-14 Regression Result for NO<sub>x</sub> 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
    
```

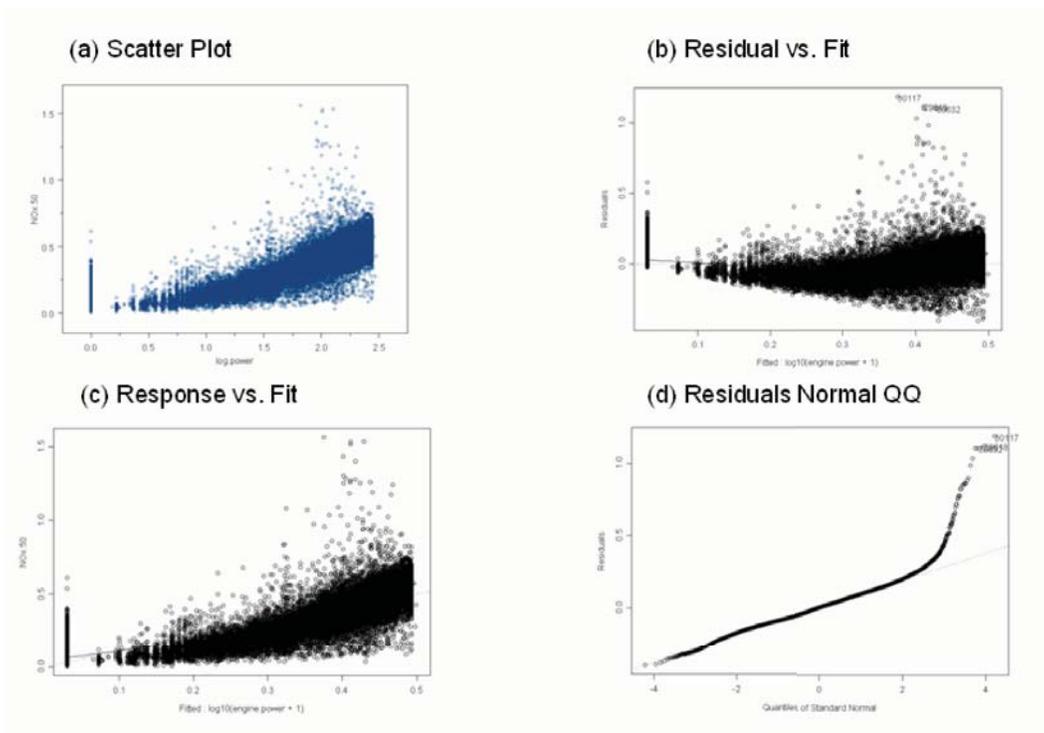


Figure 11-24 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.3

The results suggest that by using square root transformed engine power, the model increases the amount of variance explained in truncated transformed  $\text{NO}_x$  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  $\text{NO}_x$  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  $\text{NO}_x$  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
$\geq$ 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.

$$Y = \beta_0 + \beta_1 \text{engine.power}^{(1/2)} + \beta_2 \text{dummy1} + \beta_3 \text{dummy1engine.power}^{(1/2)} + \text{Error} \quad (1.4)$$

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

Table 11-15 Regression Result for NO<sub>x</sub> 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))
  dummy1      -0.0682    0.0026  -25.9438  0.0000
I(engine.power^(1/2)):dummy1
  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

```

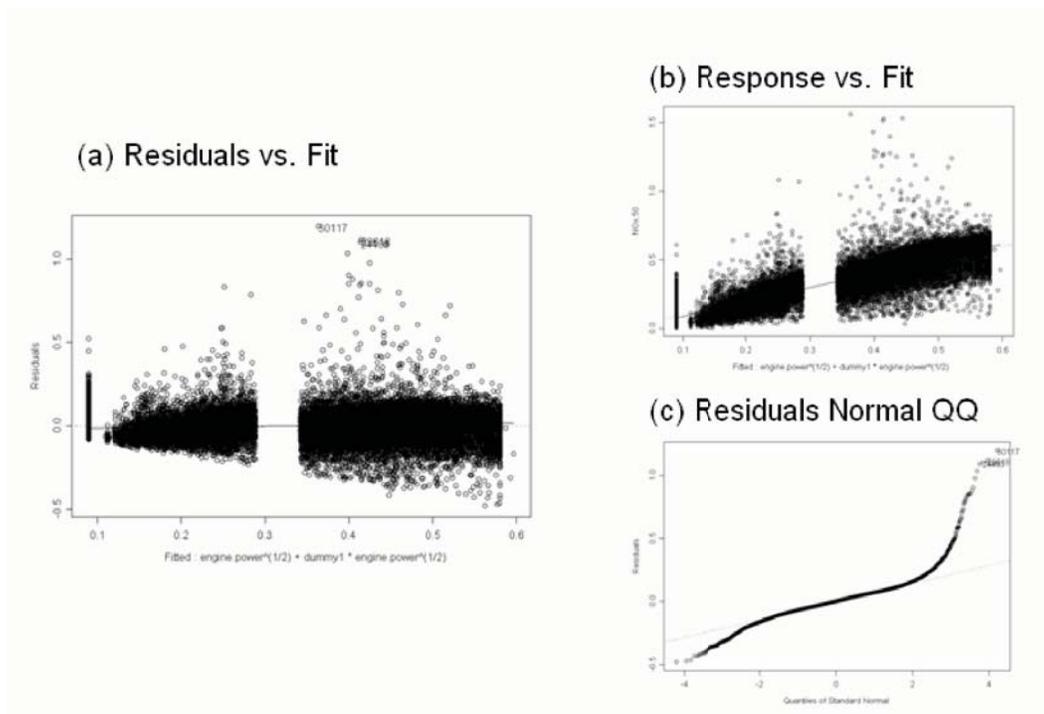


Figure 11-25 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1.4

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<sub>x</sub> 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 ( $\beta_1$ )	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 <sup>0.5</sup> )	0.614	0.975	0.07494	0.00707
Linear regression (log(power))	0.587	1.287	0.08043	0.00933
Linear regression (power <sup>0.5</sup> ) 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<sub>x</sub> 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<sub>x</sub> 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<sub>x</sub> emission rate model selected for cruise mode is:

$$\text{NO}_x = (0.087 + 0.0311(\text{engine.power})^{(1/2)})^2$$

Analysis results support the observation that the final NO<sub>x</sub> 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:

$$Y = \beta_0 + \beta_1 \text{engine.power} + \text{Error} \quad (2.1)$$

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.

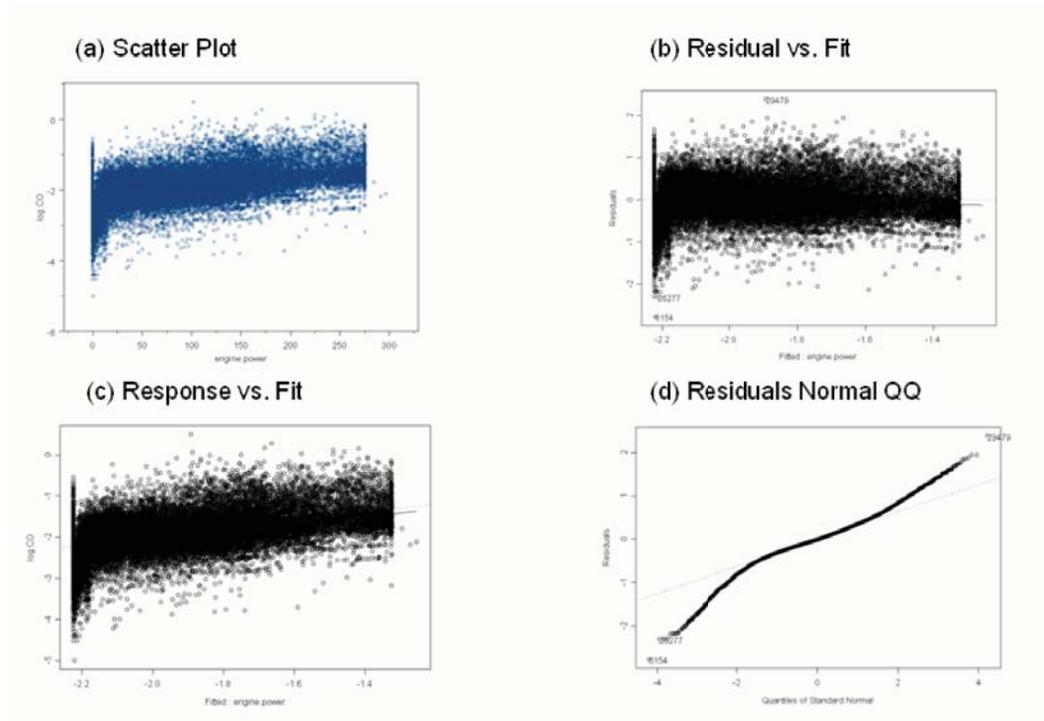


Figure 11-26 QQ and Residual vs. Fitted Plot for CO Model 2.1

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.

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

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

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
    
```

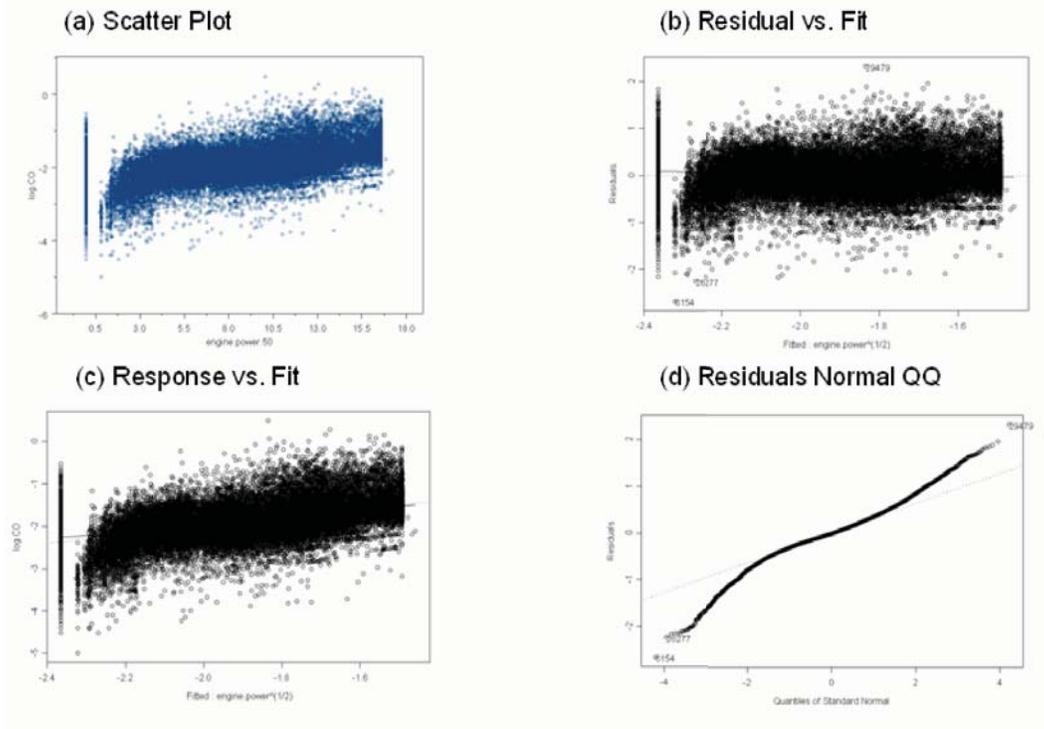


Figure 11-27 QQ and Residual vs. Fitted Plot for CO Model 2.2

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
    
```

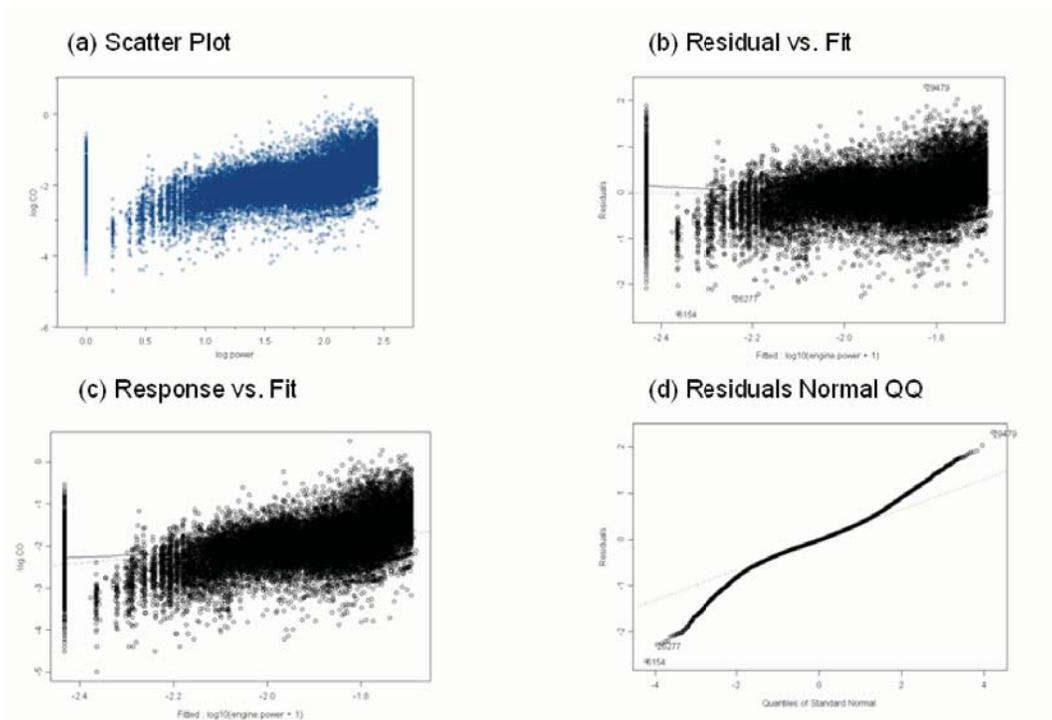


Figure 11-28 QQ and Residual vs. Fitted Plot for CO Model 2.3

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.

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

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
    
```

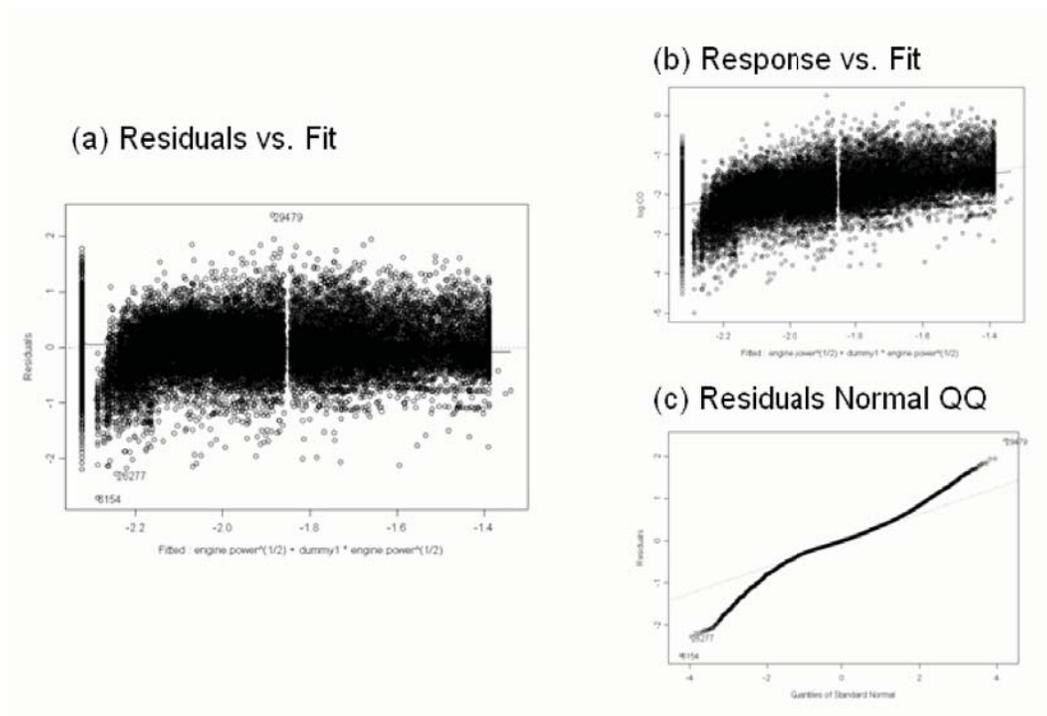


Figure 11-29 QQ and Residual vs. Fitted Plot for CO Model 2.4

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 ( $\beta_1$ )	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 <sup>0.5</sup> )	0.0899	1.984	0.04662	0.00804
Linear regression (log(power))	0.0659	2.560	0.04736	0.00866
Linear regression (power <sup>0.5</sup> ) 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:

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

### 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:

$$Y = \beta_0 + \beta_1 \text{engine.power} + \text{Error} \quad (3.1)$$

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.

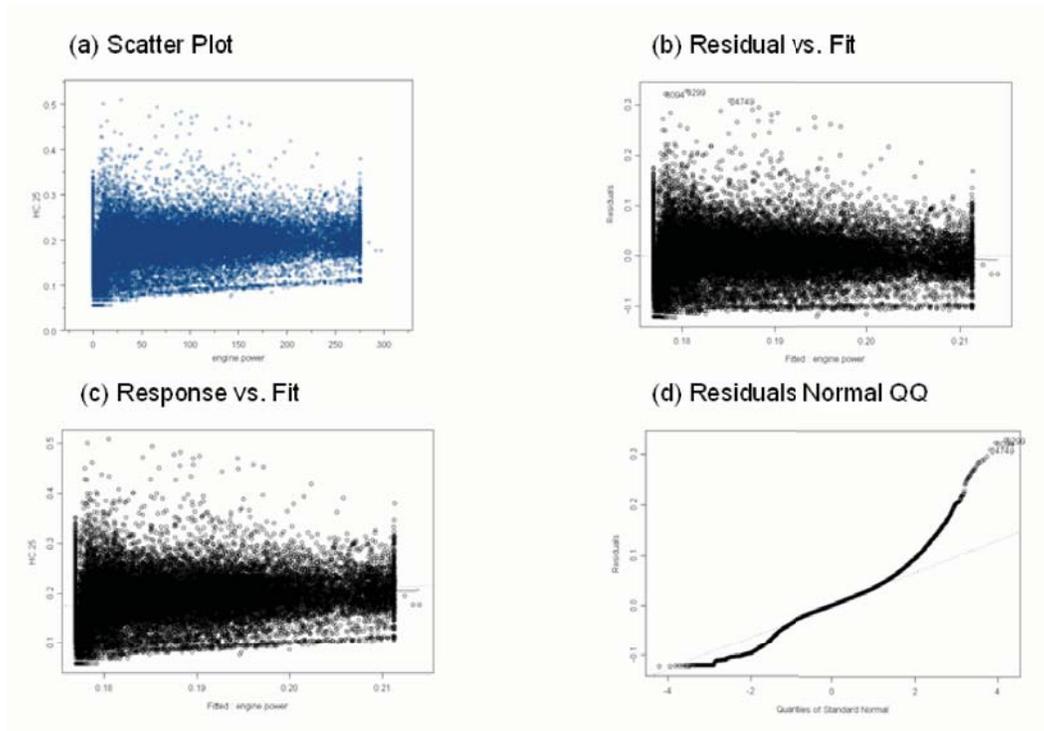


Figure 11-30 QQ and Residual vs. Fitted Plot for HC Model 3.1

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.

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

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

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
    
```

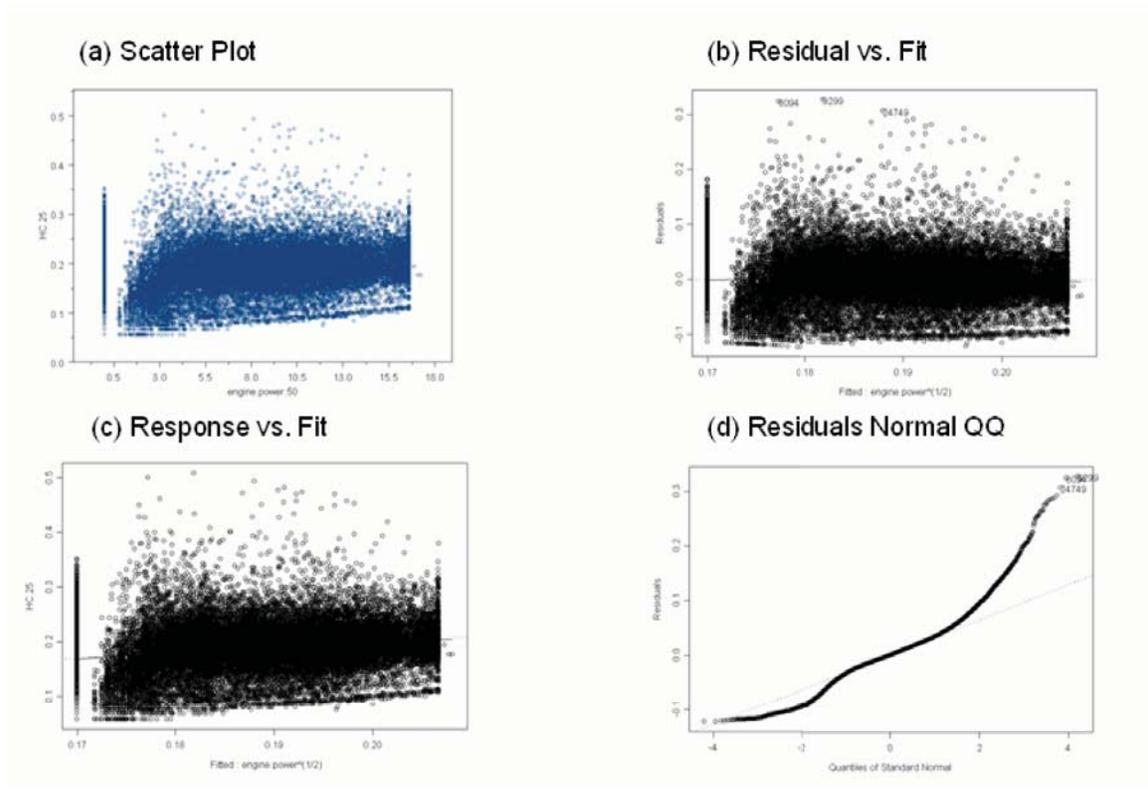


Figure 11-31 QQ and Residual vs. Fitted Plot for HC Model 3.2

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
    
```

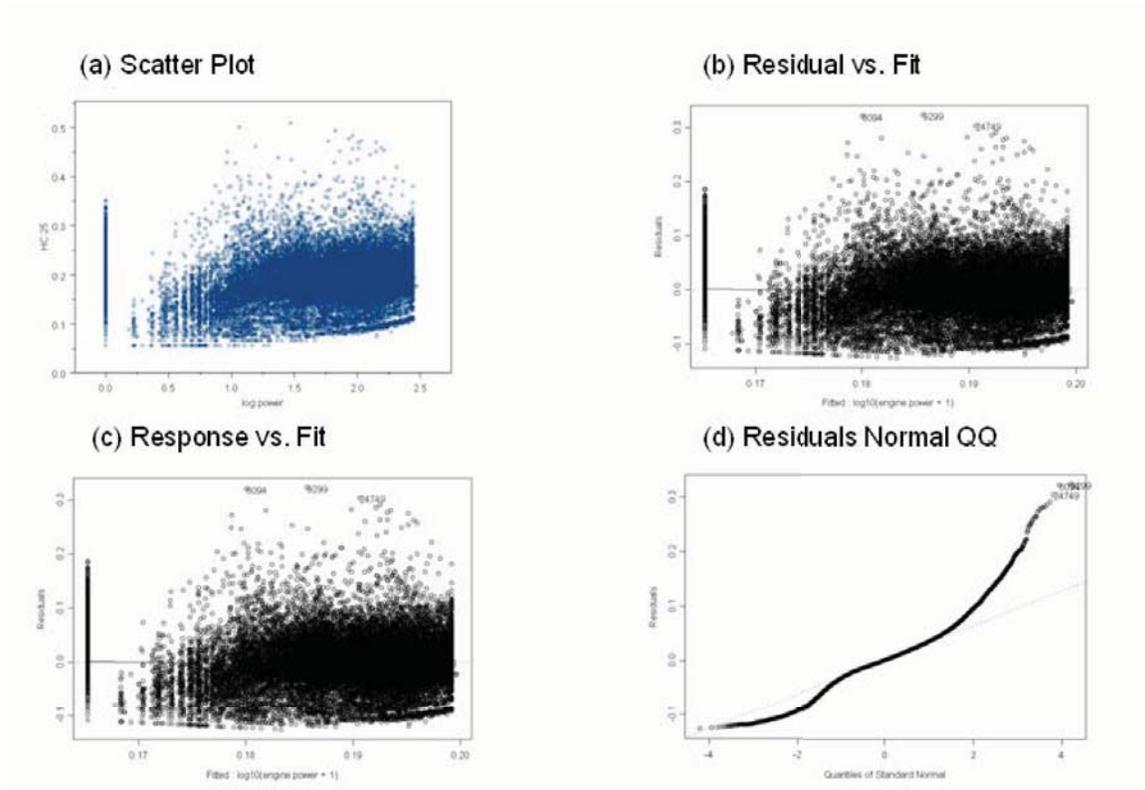


Figure 11-32 QQ and Residual vs. Fitted Plot for HC Model 3.3

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
$\geq$ 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.

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

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		

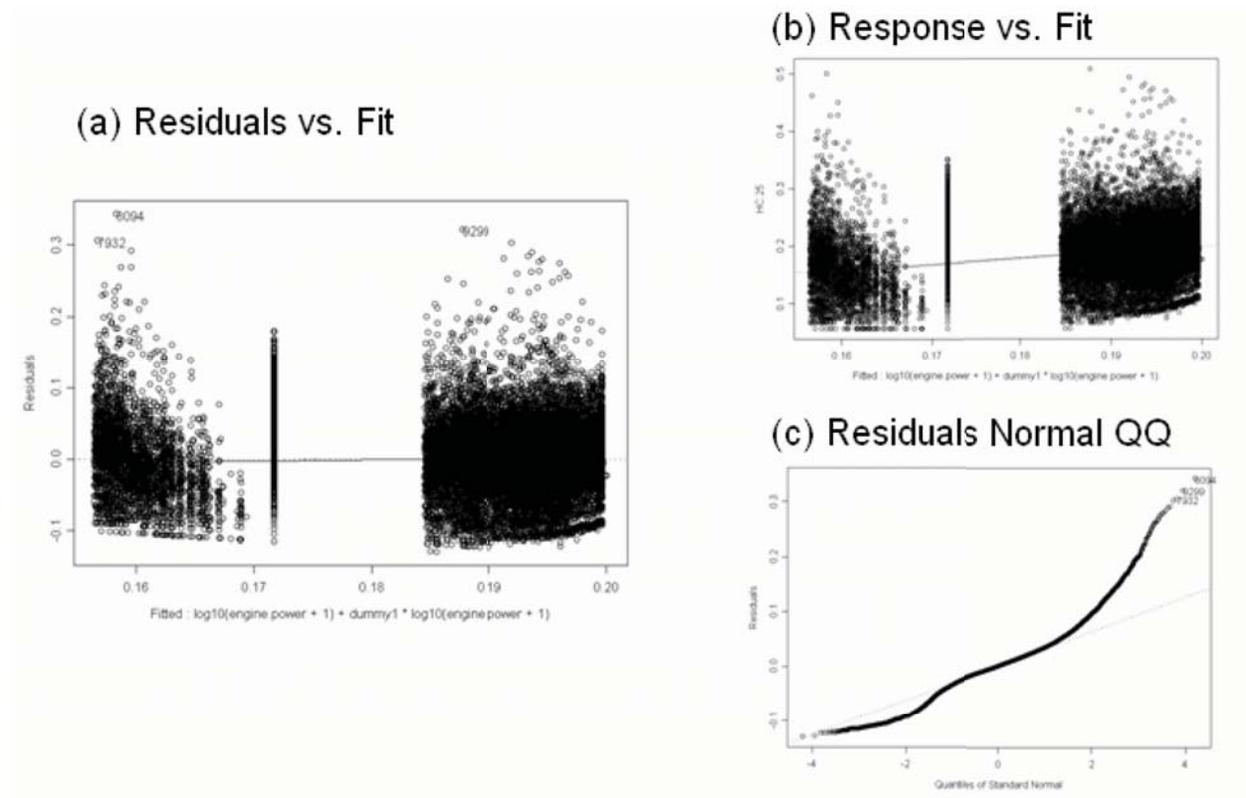


Figure 11-33 QQ and Residual vs. Fitted Plot for HC Model 3.4

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  $\beta$  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 ( $\beta_1$ )	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 <sup>0.5</sup> )	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 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 11-31 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 NO<sub>x</sub> 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 than 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:

$$\text{NO}_x = [0.0087 + 0.0311 (\text{engine.power})^{(1/2)}]^2$$

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

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

## 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<sub>x</sub>, 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/s<sup>2</sup>
- $\gamma$ : mass factor accounting for the rotational masses (~0.1)
- g: acceleration due to gravity
- grade: road grade
- C<sub>R</sub>: rolling resistance (~0.0135)
- $\rho$ : air density (1.2)
- C<sub>D</sub>: 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):

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

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<sub>x</sub> 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:
              (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

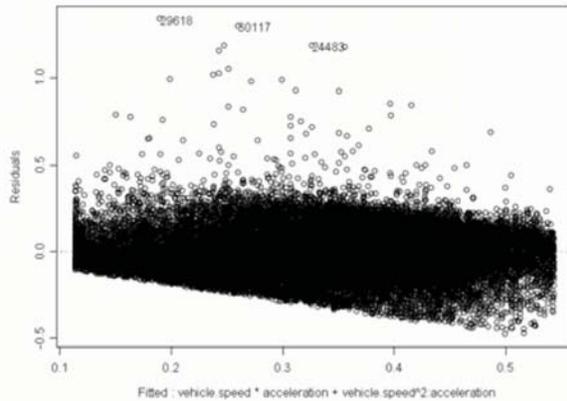
Analysis of Variance Table

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

```

(a) Residual vs. Fit



(b) Residuals Normal QQ

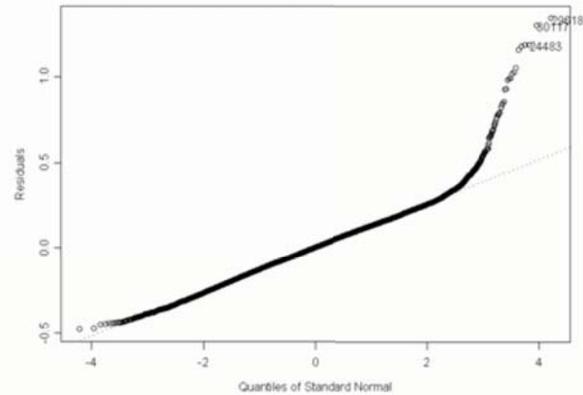


Figure 12-1 QQ and Residual vs. Fitted Plot for NO<sub>x</sub> Model 1

The results suggest that the surrogate variable model can explain about 37 % of the variance in truncated transformed NO<sub>x</sub>, 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:

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

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

where:

- RMSE: = root mean square error
- n: = number of observations
- $y_i$ : = observation 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 ( $\beta_1$ )	RMSE	MPE
NO <sub>x</sub>				
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<sub>x</sub> 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  $\text{NO}_x$  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  $\text{NO}_x$  emissions. When using the entire database at once (data are not broken into mode subsets for this derivation), the appropriate transformation for  $\text{NO}_x$  is  $1/4$  based on Box-Cox results, rather than the  $1/2$  value used in developing models for acceleration and cruise mode (see Chapters 10 and 11). The trimmed HTBR tree models for  $\text{NO}_x$  are illustrated in Figure 12-2 and Table 12-3.

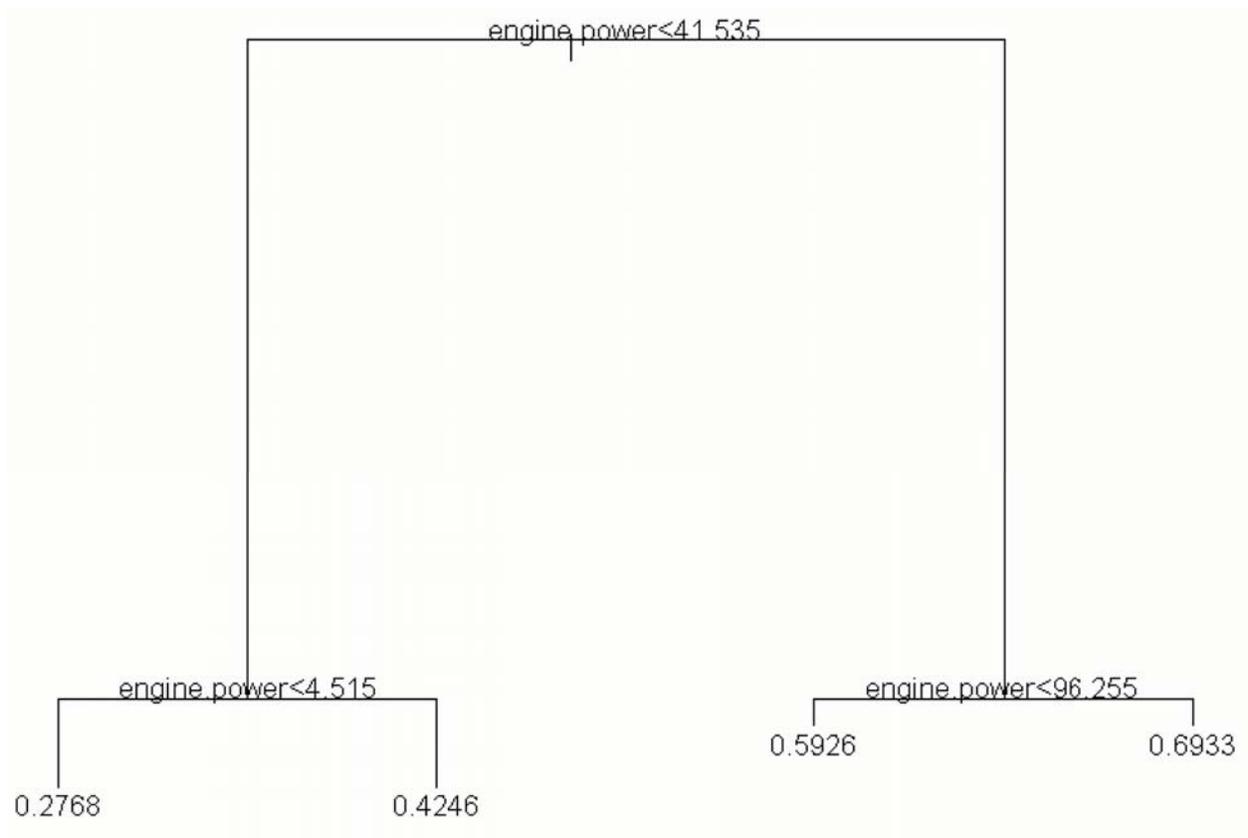


Figure 12-2 Trimmed Regression Tree Model for Truncated Transformed  $\text{NO}_x$

Table 12-3 Trimmed Regression Tree Results for Truncated Transformed NO<sub>x</sub>

```

Regression tree:
tree(formula = NOx.25 ~ engine.power + vehicle.speed + acceleration +
      oil.temperature + 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<sub>x</sub> 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<sub>x</sub> emissions is:

$$NO_x = [0.230 + 0.195\log_{10}(\text{engine.power}+1)]^4$$

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

Table 12-4 Regression Result for NO<sub>x</sub> 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
    
```

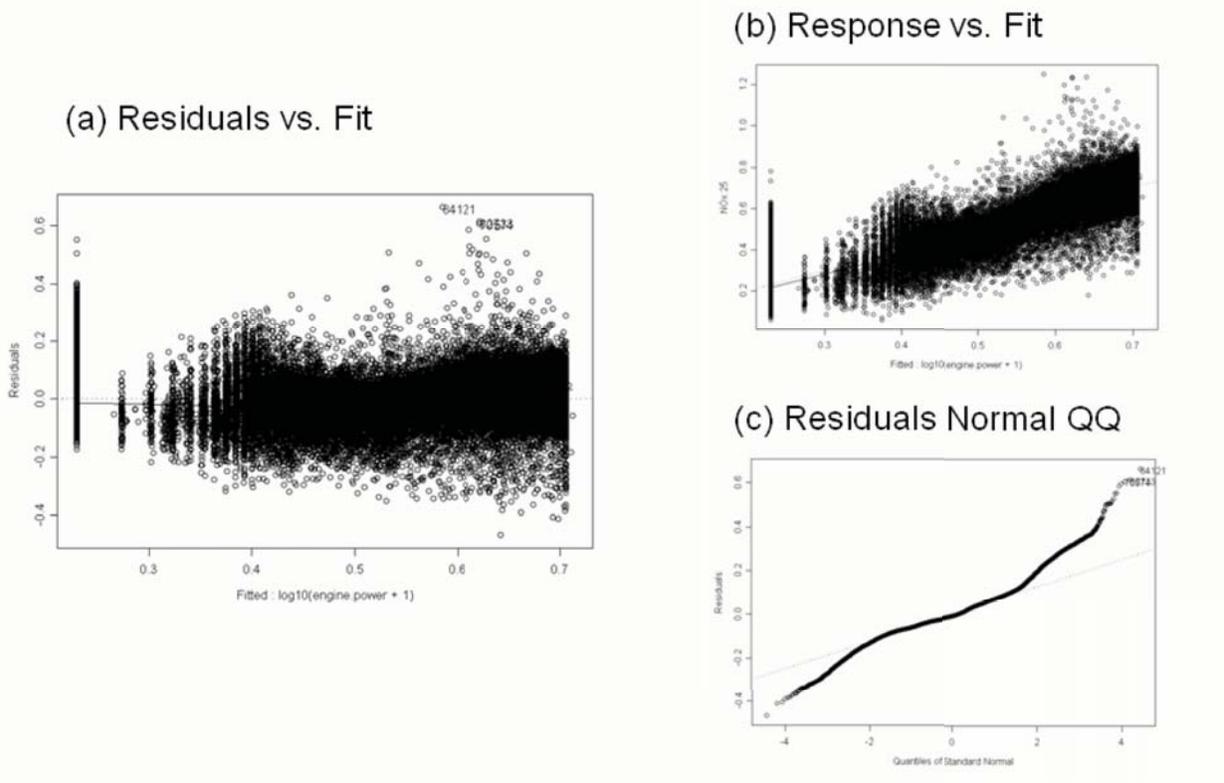


Figure 12-3 QQ and Residual vs. Fitted Plot for Load-Based Only NO<sub>x</sub> Emission Rate Model

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

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

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

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

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 <sup>2</sup> )	Slope (β <sub>1</sub> )	RMSE	MPE
NO <sub>x</sub>				
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<sub>x</sub>, 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<sub>x</sub> emission model for combined driving mode is:

$$\text{NO}_x = [0.113 + 0.0266(\text{engine.power}^{(1/2)})^2]$$

The final CO emission model for combined driving mode is:

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

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

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

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 <sup>2</sup> )	Slope ( $\beta_1$ )	RMSE	MPE
NO <sub>x</sub>				
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<sub>x</sub> 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<sub>x</sub>.

### 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 °F, 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

	<b>Coefficient of determination (R<sup>2</sup>)</b>	<b>Slope (<math>\beta_1</math>)</b>	<b>RMSE</b>	<b>MPE</b>
NO <sub>x</sub>				
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<sub>x</sub> 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<sub>x</sub> 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 significant for NO<sub>x</sub> 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  $\text{NO}_x$ , 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 NO<sub>x</sub>, 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<sub>x</sub>, 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<sub>x</sub>, 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 <sub>x</sub>	
Idle Mode	0.033415 g/s
Decelerating Motoring Mode	0.0097768 g/s
Deceleration Non-Motoring Mode	0.045777 g/s
Acceleration Mode	$NO_x = (-0.0195 + 0.201\log_{10}(\text{engine.power} + 1) + 0.0019\text{vehicle.speed})^2$
Cruise Mode	$NO_x = (0.0087 + 0.0311 (\text{engine.power})^{(1/2)})^2$
CO	
Idle Mode	0.0059439 g/s
Decelerating Motoring Mode	0.0052857 g/s
Deceleration Non-Motoring Mode	0.0068557 g/s
Acceleration Mode	$CO = 10^{(-3.747 + 1.341\log_{10}(\text{engine.power} + 1) - 0.0285\text{vehicle.speed})}$
Cruise Mode	$CO = 10^{(-2.223+0.0033\text{engine.power})}$
HC	
Idle Mode	0.00091777 g/s
Decelerating Motoring Mode	0.001113 g/s
Revised Deceleration Mode	0.001312 g/s
Acceleration Mode	$HC = (0.114 + 0.0426\log_{10}(\text{engine.power} + 1))^4$
Cruise Mode	$HC = (0.170 + 0.0022 (\text{engine.power})^{(1/2)})^4$

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<sup>2</sup> value is high for NO<sub>x</sub> 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<sub>x</sub>, 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/cmем/cmем\\_users\\_guide.pdf](http://pah.cert.ucr.edu/cmем/cmем_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 7<sup>th</sup> 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](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 NO<sub>x</sub> 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 Federal Regulations. National Archives and Records Administration.

- 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](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 2<sup>nd</sup> 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.