STATISTICAL APPROACH TO THE DEVELOPMENT OF A MICRO SCALE MODEL FOR ESTIMATING EXHAUST EMISSIONS FROM LIGHT DUTY GASOLINE VEHICLES

by

BENJAMIN AFOTEY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2008

DEDICATED TO MY PARENTS AND GOD ALMIGHTY

"pillars of my strength"

ACKNOWLEDGEMENTS

First, I would like to express my deepest and profound gratitude to the Almighty God for the strength, wisdom and knowledge he provided me throughout this research.

Second, I would like to express my heartfelt thanks to my research advisor, Dr. Melanie Sattler. Her words of encouragement and motivation provided a source of inspiration that propelled me to successfully complete this demanding and challenging research. She was always available to give me a patient ear and to answer all my questions throughout the entire program. In fact, her role in my doctorial program was priceless and would always be remembered.

Third, I would like to thank Dr. Stephen Mattingly for his guidance during the research study. The ideas and knowledge he contributed during the data analysis part played an important role in my research.

Fourth, I would like to thank Dr. Victoria Chen for the invaluable knowledge and tips she provided during the statistical analysis. Her timely responses to my questions during the analysis were vital to my successful completion of this research.

Fifth, I thank Dr. Sia Ardekani for his advice, contribution and guidance he provided during this research and throughout my doctorial program.

I also thank my dissertation committee member Dr. Andrew Kruzic for his support and assistance during this research.

I thank Paul Shover for his assistance during the instrument installation.

I also thank Jaya Nagireddi for helping in the data collection.

I would like to thank Barbara Howser, at the science and engineering library, for her assistance with databases during the literature review write up.

I also thank Nancy Rowe, at the statistical department for her assistance in writing the SAS (Statistical Analysis System) codes.

I would like to thank the Civil Engineering department at UTA for providing me this great opportunity to pursue my doctorate under the guidance of some great minds.

Finally, I am grateful to my parents, brothers, sisters and church for their support in any small way. Without their prayers and love I would not have had this opportunity to pursue and complete my graduate study.

November 11, 2008

ABSTRACT

STATISTICAL APPROACH TO THE DEVELOPMENT OF A MICRO SCALE MODEL FOR ESTIMATING EXHAUST EMISSIONS OF LIGHT DUTY GASOLINE VEHICLES

Benjamin Afotey, PhD.

The University of Texas at Arlington, 2008

Supervising Professor: Dr. Melanie Sattler

Today, air pollution is taking a growing toll on human health, the environment and the economy, despite decades of efforts to combat it. Although it was once a primary urban phenomenon in industrialized countries, today air pollution has spread worldwide. Mobile, industrial and natural sources constitute the major sources of air pollution.

Around the world, major cities in industrialized countries have in recent times been battling air pollution from mobile sources. Beijing, New Delhi, Dallas/ Fort Worth and Los Angeles are no exceptions.

Pronounced interest has focused on mobile source (vehicle exhaust) emissions in recent decades. A rapid increase in the number of vehicles in use and a corresponding increase in vehicle miles traveled (VMT), especially in urban areas, have made vehicle

emissions suspected culprits for some major health and environmental problems observed among urban populations. At the state and regional levels, transportation and air quality engineers are developing various transportation models to help estimate vehicle exhaust emission.

This research involves development of a statistical model for vehicle tailpipe emissions estimation. Second-by-second data collection was carried out using an On Board emissions measurement System (OBS-1300 system) which was installed in the 2007 Dodge Charger Car acquired by the Civil Engineering Department of the University of Texas at Arlington (UTA).

The test procedure involved 40 hours of second-by-second emissions data collection. Two roadway types, arterial and highway, were considered and data was collected for two different time periods, off peak and peak.

The model built contains predictor variables such as velocity and acceleration, and thus is able to address driving dynamics and with excellent potential of estimating second-by-second vehicle emissions.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

1.1 Introduction

Despite decades of efforts to combat it, air pollution today is taking a growing toll on human health, the environment and the economy. Although it was once a primary urban phenomenon in industrialized countries, today air pollution has spread worldwide. Mobile, industrial and natural sources constitute the major sources of air pollution.

Around the world, major cities in industrialized countries have in recent times been battling air pollution from mobile sources. These cities include Beijing and Shanghai in China, New Delhi in India, Tokyo in Japan and Athens in Greece. Dallas/ Fort Worth, Texas and Los Angeles, California in the United States are no exceptions. According to the U.S Environmental Protection Agency, all mobile sources in the U.S contributed about 79% of all carbon monoxide (CO) emissions, 53% of all nitrogen oxides (NO_x) emissions, 43% of all volatile organic compounds (VOC) emissions, 21% of all particulate matter ($PM_{2.5}$) emissions and 30% of all carbon dioxide (CO_2) emissions (USEPA, 2003).

1.2 Why Developing a Microscale Emissions Model for Predicting Vehicle Emissions is Important

During recent decades, pronounced interest has focused on mobile source (vehicle exhaust) emissions and their health and environmental impacts. A rapid increase in the number of vehicles in use and a corresponding increase in vehicle miles traveled (VMT), especially in urban areas, have made vehicle emissions suspected culprits for some major health problems observed among urban populations.

For instance, long-term exposure to nitrogen dioxide and nitric oxide (NO_x) concentrations as low as 10-30 ppm can contribute to pulmonary fibrosis and emphysema, according to Varon and Fromm (1998). Carbon monoxide (CO) can lead to carbon monoxide poisoning (i.e. CO can combine with hemoglobin of the blood and reduce the blood's ability to transport oxygen to cell tissues), and impair visual perception, exercise capacity and manual dexterity. Ozone (O_3) formed from a photochemical reaction between hydrocarbons (HC) and nitrogen oxides (NO_x) can also cause lung tissue damage and respiratory illness.

Particulate matter (especially PM2.5, with a diameter less than 2.5 microns) can cause elevated health problems and even death to asthmatic and cardiovascular patients. Increased carbon dioxide (CO_2) in the stratosphere can cause global warming and lead to climate change. A World watch Institute study of 1990 also showed that air pollution, with most pollutants from vehicle exhaust, causes as many as 50, 000 deaths per year and costs as much as $40 billion a year in health care and lost productivity in the United States. (World watch Institute Study, 1990)

In order to curb the alarming rise of air pollutants, the Clean Air Act, which was last amended in 1990, required the Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS) to regulate air pollutants that are viewed as harmful to the public and the environment. For the Clean Air Act (CAA) to achieve its objectives, organizations at the state level must develop strategies to reduce vehicle emissions as part of State Implementation Plans. This task faced by environmental professionals required the development of emission models to accurately quantify emissions from roadways, and based on that, development of efficient emission reduction programs, strategies and technologies capable of addressing vehicle exhaust emissions.

This led to the development of the currently used emission model, MOBILE 6.2, developed by the Environmental Protection Agency (EPA). The most important input parameters are vehicle miles traveled (VMT) for various vehicle categories and average speed. Unfortunately, the model is not suitable for evaluating emissions impacts for small scale projects, such as changes in ramp and intersection geometric design, traffic-signal control, and intelligent transportation systems since it does not address driving dynamics.

The problem with the model is that it uses average speed as the only variable to represent driving dynamics when calculating vehicle emissions. As a matter of fact, driving dynamics cannot be appropriately characterized by average speed since a large number of different driving patterns can result in the same average speed. In essence it

is the change in speed, that is acceleration or deceleration, which represents driving dynamics.

The Mobile Emission Assessment System for Urban and Regional Evaluation (MEASURE) is another model developed by the Georgia Institute of Technology to estimate carbon monoxide, oxides of nitrogen and volatile organic compounds. MEASURE differs from the U.S Environmental Protection Agency's current model (MOBILE model) in that it estimates exhaust emissions as a function of vehicle operating modes, such as cruise, acceleration, deceleration and idling, rather than average vehicle speed. Researchers believe MEASURE model reflects on-road emissions better than MOBILE because it is a modal model. However, one big disadvantage with this model is that it contains over thirty variables and this makes the model very difficult to handle.

Due to the difficulty in representing acceleration or deceleration in a macroscale emission model (i.e. the MOBILE model) and the complexity of using the MEASURE model, this research focuses on developing a micro scale emission model that can estimate second-by-second CO_2 emissions. The model is developed from data collected using an On-Board measurement System (OBS).

The interest in CO_2 estimation is because of the rate at which CO_2 is increasing in the atmosphere, leading to global warming and a corresponding change in global climate. Therefore, by developing CO_2 emission models, strategies can be developed to help reduce CO_2 and hence reduce global warming.

1.3 Overview of Dissertation Report

The remaining chapters of the dissertation are as follows:

Chapter 2

The literature review discusses the background and impacts of the major vehicle exhaust pollutants regulated by the Clean Air Act. It also reviews similar emission models already developed and which are in use.

Chapter 3

Chapter 3 discusses the methodology used in on-road data collection. In addition, it spells out the statistical tools and software considered in the model building approach.

Chapter 4

This chapter presents results and discussion of the analysis conducted to arrive at the desired statistical (least-squares regression) model for vehicle exhaust emissions estimation.

Chapter 5

The final chapter concludes and recommends future work related to the project.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

Motor transportation systems, like every other fossil-fuel-combustion-based system, produce unwanted by-products in the form of air emissions. As discussed earlier, transportation related activities account for the highest percentage of total emissions from all sources in recent decades.

The vehicle transportation system, comprised of the motor vehicle fleet, different types of roadways, intersections, and operational control strategies (e.g. transportation control measures, signalization schemes), differs from stationary and natural sources of air pollution in several important ways. These include:

- The composition of the motor vehicle fleet changes over time.
- Emission rates of the individual vehicles are highly dependent upon the vehicle operational mode (acceleration, deceleration, cruise, idle, cold start).
- The vehicles emit pollutants at different emission rates for each pollutant depending on the emission-producing activity in which the vehicle is engaged (see Table 2.1).

The passage of the Clean Air Act Amendments of 1990 and the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) changed the way air quality impacts are considered in the transportation planning process significantly. Together,

the two acts established the concept of "conformity", which requires that all federally-funded or significant regional transportation projects show that they will not violate the National Ambient Air Quality Standards, increase the severity or magnitude of existing NAAQS violations, or delay attainment of NAAQS (FHWA, 2000).

The conformity rules issued by the Federal Highway Authority (FHWA) require that micro-scale or project level analyses be performed to determine vehicle emissions using the latest versions of EPA approved analytical computer models (FHWA, 2000). The models currently in use for determining vehicle emissions such as MOBILE6 were developed for macro scale analyses and therefore have several deficiencies when applied to micro scale level analyses.

This chapter provides background information on important vehicle emissions, their standards, measurement procedures and the types of instruments and devices used in their measurement. It also discusses important microscopic (micro scale) and macroscopic emission models, and their relevance to air emissions estimations.

2.2 Vehicle Emission Sources

Air emissions come from many different sources:

i. Stationary (point & area) sources such as factories and power plants.

ii. Smaller area sources like dry cleaners, gas stations and painting operations.

iii. On-road mobile sources including cars, buses, and trucks.

iv. Non-road mobile sources such as construction equipment, airplanes, boats and trains.

v. Natural sources such as windblown dust and volcanic eruptions.

Motor vehicles are the dominant emission sources of volatile organic compounds (VOCs), nitrogen oxides (NO_x), carbon monoxide (CO), particulate matter (PM) and carbon dioxides (CO_2) in urban areas. These pollutants result from either the combustion or evaporation of motor vehicle fuel. In the United States, apart from the very small fraction of alternative-fuel vehicles, most motor vehicles use either gasoline or diesel fuel. Light-duty vehicles use gasoline fuel, while heavy-duty trucks are diesel-powered. Various motor vehicle activities lead to varying degrees of emissions production, as shown in Table 2.1 below.

Table 2.1 Emission Producing Vehicle Activities

EMISSION PRODUCING VEHICLE ACTIVITY	TYPES OF EMISSIONS PRODUCED
Vehicle Miles Traveled	Running Exhaust Emissions ($CO,VOC,NO_x,PM_{10}, SO_x$) Running Evaporative Emissions (VOC)
Cold Engine Starts	Elevated Running Exhaust Emissions ($CO,VOC,NO_x,PM_{10}, SO_x$)
Hot Engine Starts	Elevated Running Exhaust Emissions ($CO,VOC,NO_x,PM_{10}, SO_x$)
Engine “Hot Soaks” (shut downs)	Evaporative Emissions (VOC)
Engine Idling	Running Exhaust Emissions ($CO,VOC,NO_x,PM_{10}, SO_x$) Running Evaporative Emissions (VOC)
Exposure to Diurnal and Multi-Day Diurnal Temperature Fluctuation	Evaporative Emissions (VOC)
Vehicle Refueling	Evaporative Emissions (VOC)
Modal Behavior (e.g., High Power Demand, Heavy Engine Loads, or Engine Motoring)	Elevated Running Exhaust Emissions ($CO,VOC,NO_x,PM_{10}, SO_x$)

CO = Carbon Monoxide; VOC = Volatile Organic Compound; NO_x = Oxides of Nitrogen; PM_{10} = Particulate Matter (less than 10 microns in diameter); SO_x = Oxides of Sulfur

Source: Adapted from Guensler, 1993

Motor vehicle emission rates are also affected by vehicle parameters, fuel parameters, vehicle operating conditions and environmental conditions, as listed in Table 2.2 below.

Table 2.2 Vehicle Parameters, Fuel Parameters, Vehicle Operating Conditions, and Environmental Conditions Known to Affect Motor Vehicle Emission Rates

Vehicle Parameters • Vehicle class (weight, engine size, HP, etc.)* • Model year • Accrued vehicle mileage • Fuel delivery system (carbureted or fuel injected) • Emission control system • On-board computer control system • Control system tampering • Inspection and maintenance history	Fuel Parameters • Fuel type • Oxygen content • Fuel volatility • Sulfur content (SO_x precursor) • Benzene content • Olefin and aromatic content • Lead and metals content • Trace sulfur (catalyst effects)*
Vehicle Operating Conditions • Cold or hot start mode (unless treated separately) • Average vehicle speed • Modal activities that cause enrichment* • Load (e.g. A/C, heavy loads, or towing) • Trip lengths and trips/day* • Influence of driver behavior*	Vehicle Operating Environment • Altitude • Humidity • Ambient temperature • Diurnal temperature range • Road grade*

Note:

These components are not explicitly included in the USEPA Mobile emission rate model

Source: Guensler, 1993

2.2.1 Hydrocarbon Emissions

Hydrocarbon emissions result from incomplete combustion in the engine, - or when unburned fuel is emitted from the engine as exhaust, and also through direct evaporation of fuel from vehicle engines into the atmosphere. Absorption of the fuel into the oil layer and crevices in the combustion chamber, and leakages around the valves, contribute to evaporative hydrocarbon or VOC emissions. Even higher levels of hydrocarbons are emitted from the engine during cold starts. During cold starts, the engine's emission control system (catalytic converter) is cold and so cannot help in any oxidation process, leading to elevated levels of exhaust emissions. Other sources of hydrocarbon emissions occur at gas stations when refueling.

2.2.2 Nitrogen Oxide Emissions

Nitrogen oxide emissions occur when molecular nitrogen in air combines with molecular oxygen during combustion. Nitrogen oxides formation is favored by elevated temperatures and excess oxygen. Increased amounts result as temperature increases due to increasing vehicle speed.

2.2.3 Carbon Monoxide Emissions

Carbon monoxide emissions from motor vehicles result from incomplete combustion of fuel. When conditions are poor for combustion, carbon monoxide emissions increase. Elevated levels are experienced during cold starts and "jack rabbit" accelerations. High carbon monoxide levels are also produced when the weather is very cold or at high elevations where there is less oxygen in the air for fuel combustion.

Automobile exhaust produces as much as 95 percent of all CO emissions (U.S EPA 2000).

2.2.4 Particulate Matter Emissions

Gasoline engines emit less particulate than diesel engines. A light-duty diesel vehicle, for instance, will emit from 10-30 times as much particulate as a gasoline engine without a catalytic converter (Lederer, 2001).

2.3 Pollutant Impacts

Vehicle emissions continue to be a major source of air pollution in the United States and other industrialized countries such as China, India and Japan. Vehicle emissions are of great concern in recent times because of their increased production due to increased vehicle miles traveled (VMT), especially in industrialized countries. These pollutants have the potential of causing various health problems to humans and damage to agriculture through acid rain. Hence, elevated amounts can be detrimental to the environment (USEPA, July 2003).

The major pollutants from gasoline vehicles are hydrocarbons (HC), also referred to as volatile organic compounds (VOC); carbon monoxide (CO); and oxides of nitrogen (NO_x). Another gas emitted from gasoline vehicles which has aroused attention due to global warming is carbon dioxide (CO_2). Ozone (O_3) is produced indirectly from gasoline vehicles. In addition to these pollutants, regulated pollutants of concern to EPA are particulate matter (PM), which is mostly produced from diesel vehicles, and sulfur dioxide (SO_2). Lead is also regulated by EPA, but various emission reduction programs have virtually eliminated lead from gasoline.

2.3.1 Carbon Monoxide

Carbon monoxide (CO) is a colorless, practically odorless and tasteless gas. The U.S National Air Quality Standard (NAAQS) for CO is 35 ppm for the one-hour average and 9 ppm for the 8-hour average.

Carbon monoxide is a poisonous gas, and when it bonds with hemoglobin in the blood stream, it reduces the ability of the blood to deliver oxygen to organs and tissues. At low concentrations, it causes fatigue in healthy people and chest pains in people with heart disease. At elevated concentrations, impaired vision, dizziness, headaches, nausea and poor learning ability may result.

2.3.2 Nitrogen Oxides

Nitrogen oxides are highly reactive gases that contain nitrogen and oxygen in varying amounts. Nitrogen dioxide (NO_2), which is a common pollutant, is most often seen to combine with particles in the air to form as a reddish-brown layer over many urban areas. The other component of nitrogen oxides is nitric oxide (NO). Motor vehicles, electric utilities, and other industrial, commercial and residential sources that burn fuel are the primary sources of nitrogen oxides. Motor vehicles emit a greater percentage of the NO_x, about 55 % (EPA, 2006).

About 95 percent of the NO_x is emitted as NO, which is readily converted to NO_2 in the atmosphere. NO_2 is a primary precursor in ozone formation. The NAAQS for NO_2 is 0.053 ppm (USEPA, 2000).

NO_x reacts with water vapor in the atmosphere to form acid rain, which may cause deterioration of cars, buildings, historic monuments and also put the lives of

fishes in lakes and streams at risk by making them acidic. Figure 2.1 shows the percentage contributions of nitrogen oxides from the major sources.

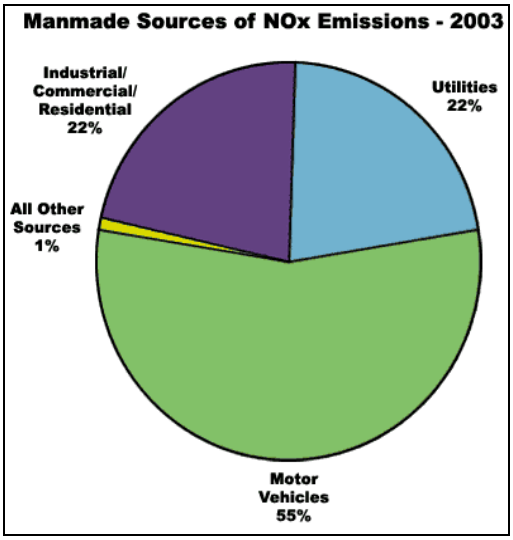


Figure 2.1 Human Sources of NO_x
Source: USEPA, March 2006

2.3.3 Hydrocarbons or Volatile Organic Compounds

Hydrocarbons are chemical compounds that contain hydrogen and carbon atoms. Hydrocarbon-based fuels such as gasoline and diesel are mostly used to power motor vehicles. In the presence of sunlight, hydrocarbons react with nitrogen oxides to form ozone. Hydrocarbons include many toxic compounds that cause cancer, lung irritation and other health effects. In addition to motor vehicle sources, hydrocarbons are also produced from industrial and natural processes.

2.3.4 Ozone

Ozone is a gas that occurs both in the earth's upper atmosphere and at ground level. Ozone which occurs in the upper atmosphere (stratosphere) is "good ozone" that shields us from the sun's ultraviolet rays. However, "bad ozone" is found at ground

level (troposphere) and is formed as a result of a reaction between volatile organic compounds (VOC) and nitrogen oxides in the presence of sunlight. Ground level ozone is colorless and very reactive. Figure 2.2 and Equation (1) shows a simplified version of photochemical reaction process that leads to ozone formation.

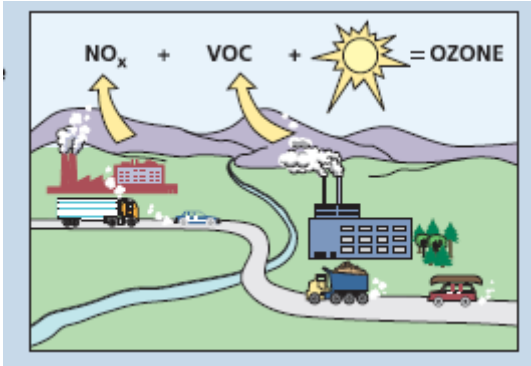


Figure 2.2 Photochemical Reaction Process
(Source: http://www.epa.gov/airtrends/2005/ozonenbp.pdf)

$$\text{VOC} + \text{NOx} \xrightarrow{\textit{sunlight}} \text{Ozone} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2.1$$

The current National Ambient Air Quality Standard for ozone is the 8-hour standard of 0.08 ppm. Ground level ozone is the primary ingredient in photochemical smog, which is a concern during the summer months due to strong sunlight and hot weather that triggers its formation.

Elevated levels of ground level ozone can cause a number of health problems, including chest pains, coughing and throat irritation. It can worsen bronchitis, emphysema and asthma. It can also reduce lung function, and repeated exposure may permanently scar lung tissues. Also at risk are healthy adults and children who may experience difficulty in breathing when exposed to high levels of ozone pollution, which occur most frequently during summer months. Ground level ozone also damages

vegetation and ecosystems. According to USEPA, it is responsible for an estimated $500 million in reduced crop production each year in the United States (USEPA, June 2006).

2.3.5 Particulate Matter

Particulate matter, also called particle pollution or PM, is a complex mixture of extremely tiny particles and liquid droplets. They come from a variety of sources such as construction sites, tilled fields, unpaved roads, stone crushing, wood burning and fuel combustion processes such as in the production of electricity and on-road vehicles. Other particles are formed in the atmosphere due to chemical reactions. Larger particles are visible as dust or smoke and settle out rapidly. The tiniest particles can be suspended in the air for long periods of time and are the most harmful to human health.

Particulate matter 2.5 (PM2.5) includes particles with an aerodynamic diameter less than 2.5 microns. They are so small that they are not typically visible to the naked eye. In the atmosphere, they are significant contributors to haze. They can cause significant health problems because the particles are too small to be filtered out and so they penetrate deep into the lungs and gets deposited in bronchioles or alveolar sacs. They also increase the risk of death from respiratory or cardiac disease for the elderly. Particulate matter with diameter less than 2.5 microns in diameter (PM2.5) is believed to elevate these health problems because they are even finer particulates. Virtually all particulate matter from mobile sources is PM2.5 (USEPA, 2000).

Particulate matter 10 (PM10) is mainly coarse particles that get deposited in the nose and throat but do not penetrate into lungs. They may be responsible for throat irritation.

PM10 has primary standards for short-term and long-term exposures. The standard for short-term exposure (24-hours) is 150 $\mu g/m^3$ and cannot be exceeded more than once per year averaged over a three-year period. The long-term standard is an annual arithmetic mean of 50 $\mu g/m^3$ averaged over three-years (USEPA, 2000). PM 2.5 also has primary standards for short-term and long-term exposures. The standard for short-term exposure (24-hours) is 65 $\mu g/m^3$ and cannot be exceeded more than once per year averaged over a three-year period. The long-term standard is an annual arithmetic mean of 15 $\mu g/m^3$ averaged over three years. The secondary standards are the same as the primary standards (USEPA, 2000).

2.3.6 Sulfur Dioxide

Sulfur dioxide is the main component of sulfur oxides (SO_x). These gases are formed when fuel containing sulfur (mainly coal and oil) is burned at power plants and during metal smelting and other industrial processes. Mobile emissions are a minor source of sulfur dioxide emissions. According to USEPA (2001), over 65% of the SO_2 emitted into the atmosphere comes from combustion of fuels used in electrical power plants. Short- term exposure to high concentrations of SO_2 results in breathing impairment, especially for asthmatic children and adults. Long-term exposure can result in respiratory illness or aggravated cardiovascular disease (USEPA, 2001).

In the presence of water, sulfur dioxide (SO_2) is converted into sulfuric acid in the atmosphere to form acid deposition (acid rain). It contributes to acidification of soils, lakes, streams and has adverse impacts on ecosystems. Exposure of vegetation to sulfuric acid can increase foliar injury and decrease plant growth and yield. Acid deposition can also accelerate the corrosion of building materials such as steel, iron, zinc, limestone, concrete and other protective materials (USEPA, 2001).

There are both short-term and long-term NAAQS for SO_2. The short-term (24-hour) standard of 0.14 ppm is not to be exceeded more than once per year. The long-term standard specifies an annual arithmetic mean not to exceed 0.03 ppm. The secondary standard (3-hour) of 0.50 ppm is not to be exceeded more than once per year (USEPA, 2006).

2.3.7 Lead

The primary source of lead in the United States is metal processing operations, with lead smelting plants producing the highest levels. Mobile sources were the greatest source of lead about thirty years ago, but the complete elimination of lead from gasoline has eliminated motor vehicles as a significant source.

Lead is a poisonous substance and can adversely affect kidneys, liver, nervous system and blood cells. High levels of lead (above 80 μg per deciliter of blood) can cause convulsions, coma and even death. Lower levels of lead (about 10 μg per deciliter of blood) can impair mental and physical development (Lederer, 2001).

2.3.8 Carbon Dioxide

Carbon dioxide has aroused concern worldwide because it is one of the gases that cause the greenhouse effect. The sources of carbon dioxide are mainly industrial and mobile, with the latter producing greater amounts in recent decades. Figure 2.3 shows carbon dioxide emissions by end-use sector from 1990-2006, while Figure 2.4 shows amounts of carbon dioxide from various transportation sources.

Figure 2.3 Emissions of Carbon Dioxide, 1990-2006
Source: U.S Department of Energy, 2006

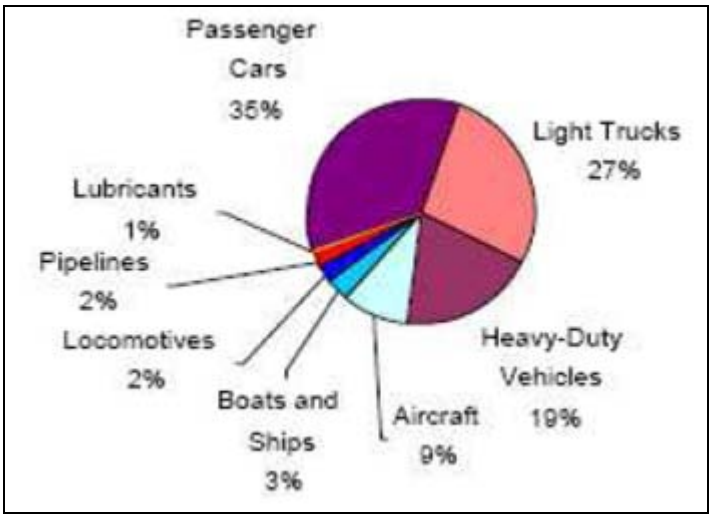


Figure 2.4 2003 Transportation Greenhouse Gas Emissions, by USEPA, 2006

2.4 The Clean Air Act

The Clean Air Act is a federal pollution law covering the entire United States. The 1990 Clean Air Act is the current version of the law. Under the law, the Environmental Protection Agency (EPA) sets ambient air quality standards to protect human health and the environment. The law allows individual states to have stronger pollution controls, but states are not allowed to have weaker pollution controls than those set for the whole country.

States have to develop state implementation plans (SIPs) that explain how each state will do its job and stay within emission standards under the Clean Air Act. A state implementation plan (SIP) is a collection of regulations a state uses to clean up its polluted area(s). EPA is the governing body that approves the SIP. If a SIP is unacceptable, EPA can decide to take over the enforcement of the Clean Air Act in the state.

2.4.1 Air Quality Standards

EPA's National Ambient Air Quality Standards (NAAQS) are federal standards, established through extensive scientific review that set allowable concentration limits for certain pollutants in order to protect public health and welfare. Two types of National Ambient Air Quality Standards were established by the Clean Air Act. These are:

The primary standard**:** sets limits to protect public health, including health of sensitive populations such as asthmatics, children and the elderly.

The secondary standard: sets limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation and buildings.

Carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), lead (Pb), ozone (O_3) and particulate matter (PM) are the six criteria pollutants for which standards have been set under the authority of the Clean Air Act. Table 2.3 lists the current NAAQS. An ambient standard is yet to be set for CO_2 due to its contribution to greenhouse effect and global warming.

Table 2.3 National Ambient Air Quality Standards

Pollutant	Primary Standard	Averaging Times	Secondary Standards
Carbon Monoxide	9ppm(10 mg/m^3)	8-hour	None
	35ppm(40 mg/m^3)	1-hour	None
Lead	1.5 μg/m^3	Quarterly average	Same as Primary
Nitrogen Dioxide	0.053 ppm (100 μg/m^3)	Annual (Arithmetic Mean)	Same as Primary
Particulate Matter (PM10)	50 μg/m^3	Annual (Arithmetic Mean)	Same as Primary
	150 μg/m^3	24-hour	
Particulate Matter (PM2.5)	15 μg/m^3	Annual (Arithmetic Mean)	Same as Primary
	65 μg/m^3	24-hour	
Ozone	0.08 ppm	8-hour	Same as Primary
	0.12 ppm	1-hour (Applies only in limited areas)	Same as Primary
Sulfur Oxides	0.03 ppm	Annual (Arithmetic Mean)	-
	0.14 ppm	24-hour	-
	-	3-hour	0.5ppm (1300 μg/m^3)

Source: http://www.epa.gov/air/criteria.html

2.4.2 Nonattainment Areas

To determine which areas have air pollution problems, monitoring networks were established by EPA, States and local agencies to measure outdoor air pollutants. Monitoring data is then analyzed to determine if standards are met. If levels of any pollutant exceed the standards, then EPA, in cooperation with the state, designates the area as nonattainment. Once the area in question meets the standards for healthy air, EPA redesignates that area back to attainment. Two cities currently on EPA's list as nonattainment areas for ozone are Dallas/Fort Worth (DFW) and Los Angeles, with DFW classified as "moderate nonattainment" and Los Angeles classified as "extreme nonattainment" for ozone. Figure 2.5 shows the counties within DFW that are nonattainment for ozone.

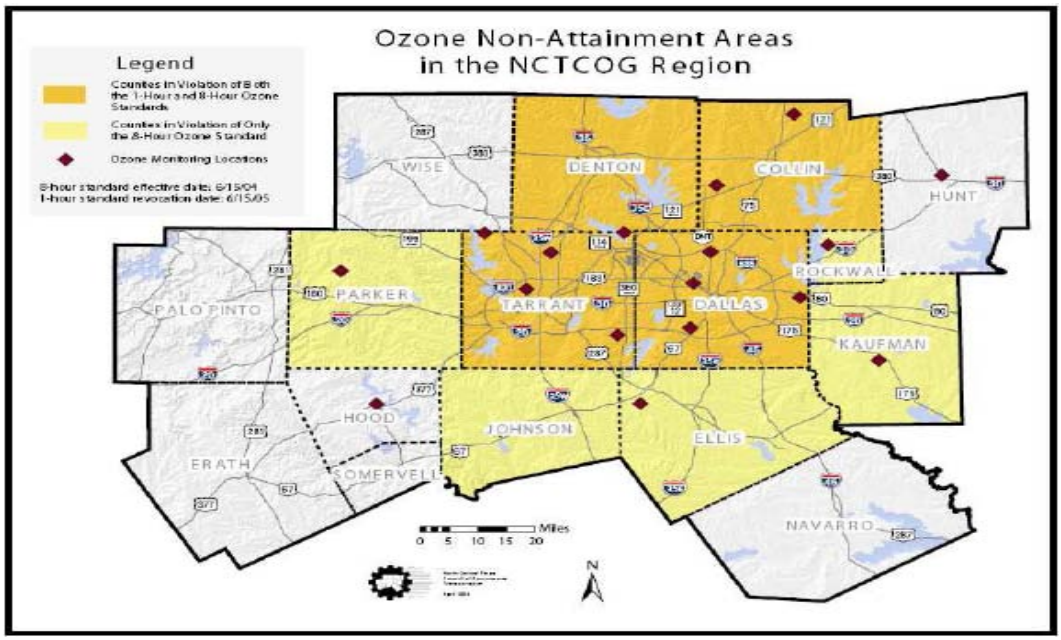


Figure 2.5 Ozone Nonattainment Areas in DFW Region
Source: NCTCOG, 2007

Table 2.3 below shows that the number of areas designated as nonattainment for air quality has decreased between 1992 and 2002, which demonstrates that air quality is improving in the United States. The information concerning ozone in Table 2.4 is based on the old 1-hour standard.

Table 2.4 Number of Areas Designated as Nonattainment

Pollutant	1992	2002
Carbon monoxide	78	24
Lead	13	3
Nitrogen dioxide	1	0
Ozone	134	74
Particulate Matter (PM10)	84	68
Sulfur Dioxide	53	26
All Pollutants	363	195

Source: U.S. EPA, July/August 2003

2.5 Why Measure / Estimate Emissions?

State Implementation Plans (SIPs) require quantitative estimates of emission reductions. Measuring and estimating emissions is necessary to predict ambient pollutant concentrations. NO_x and VOC emissions, for instance, are input into a regional (macro scale) photochemical model to predict ozone concentrations. Ozone concentrations are predicted before and after controls, to ensure that controls are sufficient to bring a region into compliance with CAA standards. If emissions are overestimated, the SIP will be overly stringent and costly. If emissions are

underestimated, the SIP will fail to achieve compliance with NAAQS. Therefore accurate emission measurement and estimation is critical to achieving air quality compliance cost effectively. In addition, it permits the application of efficient control strategies to reduce emissions (Sattler, 2005).

2.6 Emission Measurement Methods

The day-to-day rise in vehicle mile traveled (VMT) caused by the increasing population, coupled with increasing annual miles travelled per vehicle and a corresponding rise in vehicle tailpipe emissions, has warranted emission measurements. In order to curb pollutant concentrations exceeding national standards, emission measurement is necessary. This permits us to accurately quantify pollutants from all sources for a particular region and therefore be able to develop the appropriate control strategies to reduce their concentration to meet national standards.

Three major types of vehicle measurement methods are currently in use. These are: (i) Dynamometer tests (Federal Test procedure); (ii) Remote sensing; and (iii) On-board measurement methods.

2.6.1 Dynamometer Testing Method

Dynamometer testing is a method where vehicle emissions are measured under laboratory conditions during a driving cycle that simulates vehicle road operation. The driving cycle is intended to represent typical driving patterns in an urban area. A driving cycle is composed of a unique profile of stops, starts, constant speed cruises, accelerations and decelerations and is typically characterized by an overall average speed. Different driving cycles are used to represent driving under different conditions.

The vehicle can be driven over the cycle with both a cold start and hot start, and the exhaust (tailpipe) emissions are measured during the trip. A shortcoming of the dynamometer test is that it may not be representative of actual driving conditions. The dynamometer tests are often used in regulatory procedures (Federal Testing Procedure) to check compliance of light-duty vehicles and light-duty trucks with federal emission standards or to inspect in-use vehicles.

Data obtained from driving cycles are also used to develop emission estimation models such as EMFAC, MOBILE6, Georgia Tech's MEASURE and UC Riverside's modal emission model. Figure 2.6 shows a dynamometer used in vehicle emission measurement. Figure 2.7 and Figure 2.8 show light-duty vehicles on a dynamometer device undergoing an emissions test.

Figure 2.6 Chassis Dynamometer

Figure 2.7 Vehicle Undergoing Dynamometer Testing

Figure 2.8 Vehicle Undergoing Dynamometer Testing
Source (Figs. 2.6-2.8): Mustangdyne, March 2005

Figure 2.9 shows a monograph of EPA's Federal Test Procedure (FTP). The FTP uses dynamometer testing to determine compliance of light-duty vehicles and light duty trucks with federal emission standards. Pre-production and production line vehicles are tested using the FTP as part of the vehicle certification process. (Sattler, 2005)

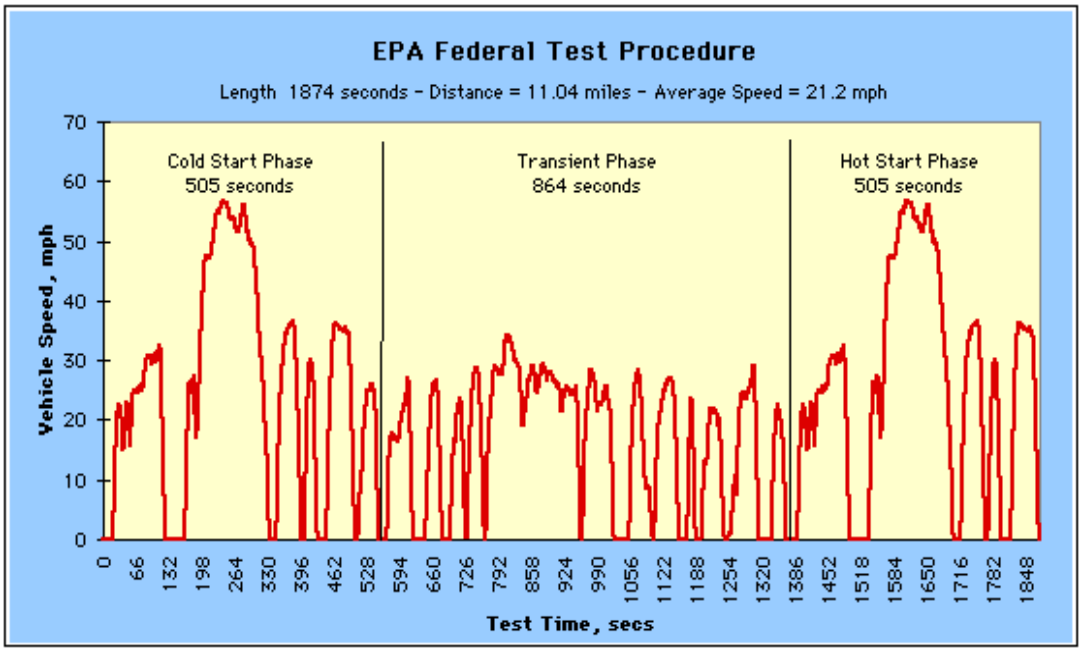


Figure 2.9 Federal Test Procedure
Source: Sattler, Transportation and Air Quality Notes, Fall 2005

2.6.2 Remote Sensing Method

Remote sensing is one of the methods used in measuring pollutant levels in a vehicle's exhaust while it is traveling down the road. This emission measurement is different from other methods since the equipment is not physically attached to the vehicle as it is with other equipment used to measure vehicle emissions today. In addition to being a fast, mobile and unobtrusive emission measuring method, remote sensing also helps in identifying excessive polluters as well as vehicles that are registered outside the boundaries of the test area. Pollutants measured by the remote sensing device (RSD) system are tailpipe hydrocarbons (HC), carbon monoxide (CO), carbon dioxide (CO_2) and nitrogen oxides (NO_x). Remote sensing, however, is not able to measure "evaporative" emissions and gasoline vapors that vent directly into the air

from hot engines and the fuel system, which are significant sources of hydrocarbon pollution capable of exceeding tailpipe emissions on some hot days (City of Albuquerque, 2005).

2.6.2.1 How Remote Sensing Works

Remote sensing device works by employing an infrared absorption principle to measure emissions. The RSD system consists of an IR source, a detector, video ID camera and computer. It operates by continuously projecting a beam of IR radiation across the roadway. The IR light signals are strong at the RSD's detector when there is no vehicle emission in the path. If there is some amount of CO, CO_2, HC present in the path, it will absorb a portion of the light for the pollutant's unique wavelength.

The device stores the readings of the ambient conditions present at the time a vehicle passes through the beam produced by the RSD device and performs electronic calibrations. The RSD system then measures the CO to CO_2 and CO to exhaust HC ratios in front of the vehicle and in the exhaust plume behind the vehicle. This measurement takes place at the rate of 100 per second and is adjusted for ambient conditions (City of Albuquerque, 2005).

The RSD calculates the vehicle's CO emission rate by comparing the "behind" measurement to the expected ratio for ideal combustion, using the "before" measurement of the ambient conditions as base. To ensure validity of the data, the system conducts several quality audit checks. All these calculations take place within 0.7 seconds (City of Albuquerque, 2005). In a similar manner, exhaust HC is calculated.

Thus the RSD compares the total carbon content of exhaust HC, CO and CO_2 to the total carbon content of the gasoline the car burns.

Measuring NO_x emissions with the RSD system is slightly different. It employs an ultraviolet (UV) light source in addition to, and collinear with, infrared due to the fact that NO_x absorption characteristics are stronger and more selective in the ultraviolet light spectrum. The RSD system uses the video camera to digitize a color image of the rear of the tested vehicle, including the license plate, and stores emissions information for each monitored vehicle based on its license plate. This makes it suitable to be used for inspection and enforcement purposes in identifying high emitting vehicles (City of Albuquerque, 2005).

The major advantage of remote sensing is that it is possible to measure a large number of on-road vehicles per day. The major disadvantage of remote sensing is that it only gives an instantaneous (snap shot) estimate of emissions at a specific location. There are also difficulties in siting the device that makes it impractical to use remote sensing as a means of vehicle emission measurement at many locations of interest, such as that close to intersections or across multiple lanes of heavy traffic (City of Albuquerque, March 2005).

Figures 2.10 and 2.11 show the measurement procedure of vehicle exhaust emissions by remote sensing method.

Figure 2.10 Measurement of Vehicle Emissions by Remote Sensing

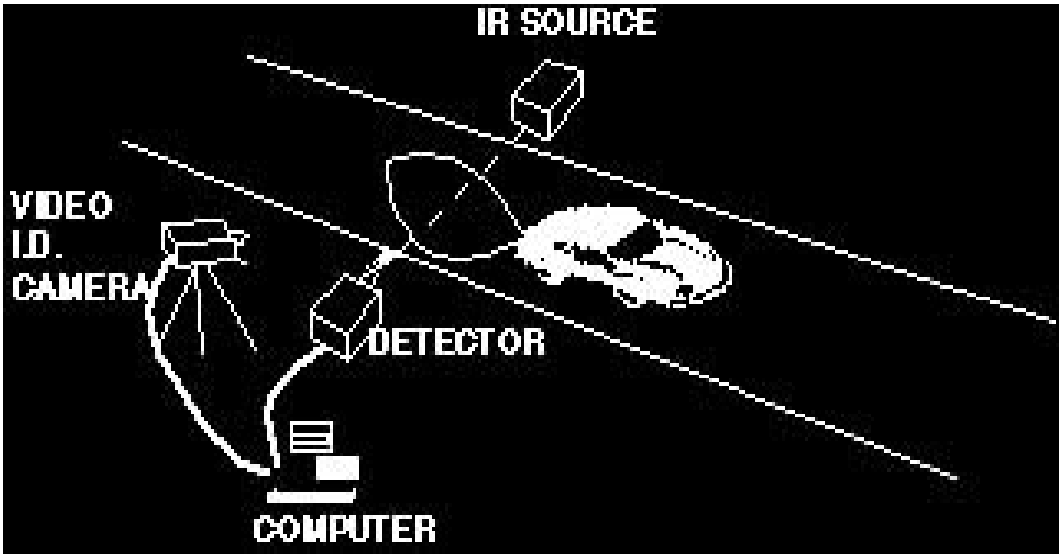


Figure 2.11 Measurement of Vehicle Emission by Remote Sensing Method
Source: City of Albuquerque, March 2005

2.6.3 On-Board Measurement Method

On-board emission measurement is widely recognized as a desirable emission measurement approach since data is collected under real- world conditions at any location and time traveled by the vehicle. Analysis and representation of variability in vehicle emissions as a result of variations in facility (roadway) characteristics, vehicle location, vehicle operation, driver behavior or other factors can be done more reliably with on-board emission measurement method than with other methods. This is because measurements are obtained during real world driving, thus eliminating concerns about non-representativeness that is often an issue with dynamometer testing, and at any location, eliminating the siting restrictions associated with remote sensing.

The Horiba OBS-1300 system, shown in Figure 2.12, consists mainly of a MEXA-1170 HNDIR analyzer, data integration unit which houses a MEXA-720 NO_x analyzer, a power supply unit, a data logger PC and other accessories. The NO_x sensor is attached to the tailpipe of the van, as shown in Figure 2.13. The OBS and its accessories are installed in the test vehicle of interest.

Figure 2.12 OBS-1300 Setup

Figure 2.13 NO_x Sensor Attached to Tailpipe

While the test vehicle is being driven down the road, OBS systems logs concentrations of HC, CO, CO_2, and NO_x; vehicle velocity; exhaust temperature and pressure; and ambient temperature, pressure and humidity on a second-by second basis.

2.6.4 Advantages and Disadvantages of Vehicle Emissions Measurement Techniques

Table 2.5 below lists advantages and disadvantages associated with the various vehicle emissions measurement techniques.

Table 2.5 Advantages and Disadvantages of Vehicle Emissions Measurement Techniques

Measurement Technique	Advantages	Disadvantages
Dynamometer	1. Can test specific velocities 2. Accurate and reliable since pattern of acceleration, decelerations and velocities is the same for all vehicles	1. Pattern may not be representative of actual driving conditions.
Remote-Sensing	1. Can sample many vehicles. 2. Real world data.	1. Provides just a snapshot of emissions. 2. Just meaningfully measures one vehicle at a time. Problem if there are several lanes of traffic. 3. Problem of not being able to identify hot spots.
On-board System	1. Can evaluate microscale impact. 2. Measures emissions for actual driving conditions rather than simulated conditions. 3. Can measure emissions from various driving patterns. 4. Can measure emissions at various times and places.	1. Can only measure one vehicle at a time. 2. Time intensive. 3. Involves significant technical difficulties.

2.7 Formation of Pollutants in Engines and Prediction of NO_x Emissions as a Function of Velocity and Acceleration

2.7.1 Formation of Pollutants in Gasoline Engines

The main emissions generated in gasoline vehicle engines are nitrogen oxides, carbon monoxide, and hydrocarbons, or volatile organic compounds. In the engine, the elemental nitrogen in the combustion air reacts with oxygen in the presence of heat to form nitrogen oxides. The amount of NO_x produced increases with combustion temperature, as shown in Figure 2.14.

Air-to-fuel ratio strongly influences emissions, as shown in Figure 2.15. Carbon monoxide and hydrocarbons are produced at low air-to-fuel ratios due to insufficient air supplied to the combustion chamber. At high air-to-fuel ratios, carbon monoxide and hydrocarbon emissions are generally reduced and carbon dioxide emissions increased, but the engine operates rougher. At stoichiometric conditions, the engine temperature is the greatest, which maximizes NO_x emissions. In recent times, the current practice for modern vehicles equipped with catalytic converters is to burn at just less than the stoichiometric air-to-fuel ratio (ER=0.98-0.99) to allow favorable operation of the catalytic converter, which must simultaneously oxidize CO and HC and reduce NO_x. With this overall target in mind, the methods of controlling the flow rate of fuel to provide the right macromixture (overall AFR) and the methods of mixing the air and fuel to provide a homogeneous micromixture in every part of the cylinder are very important.

Today, achieving the desired AFR is accomplished by fuel injectors, as compared with carburetors in the past.

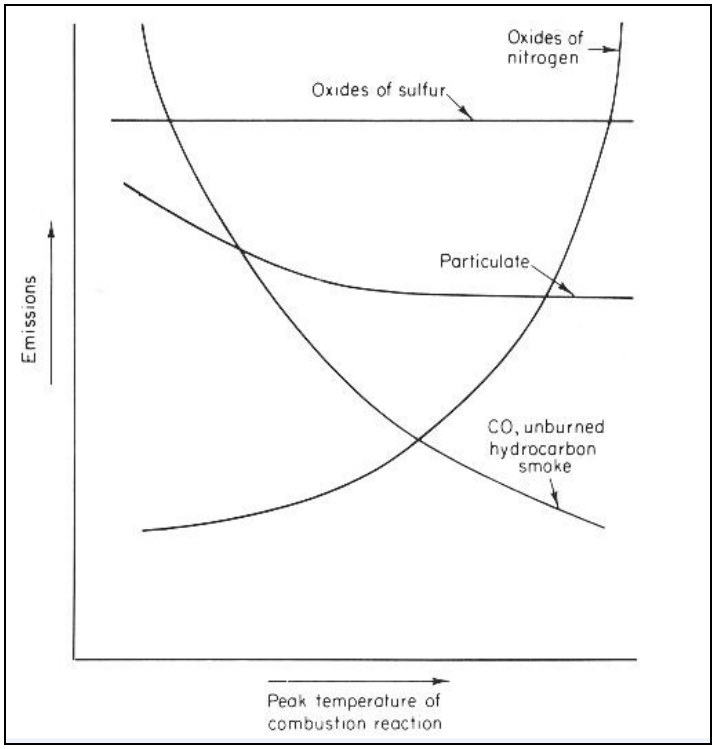


Figure 2.14 Plots of Emissions versus Temperature
Source: Boubel, 1994

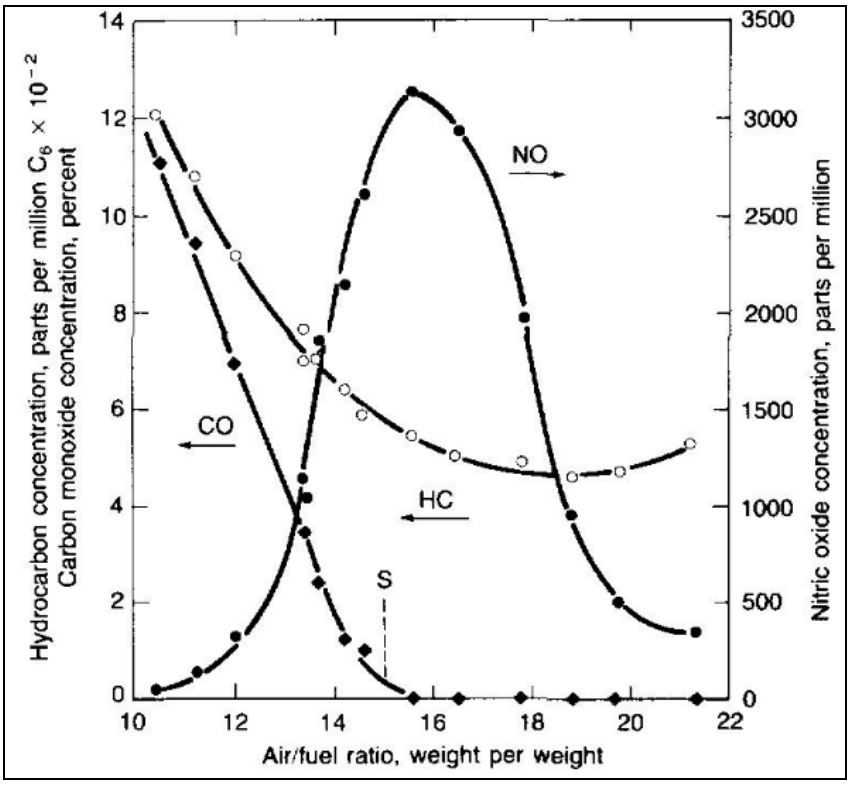


Figure 2.15 Effects of Air/Fuel Ratio on Nitrogen Oxide, Hydrocarbon and Carbon Monoxide Exhaust Emissions
Source: Cooper and Alley, 2002

2.7.2 Variation of NOx with Velocity

Figure 2.16 shows a plot of how nitrogen oxides (NO_x) in g/mile changes with velocity or speed in mph, at zero acceleration, according to repeated runs of MOBILE5a. The figure shows light-duty fleet-averaged emission factors.

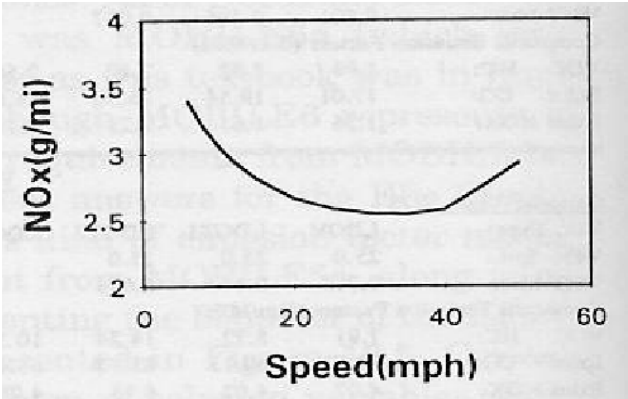


Figure 2.16 Light-Duty Vehicle Emissions vs. Speed from MOBILE
Source: Cooper and Alley, 2002

A review of Figure 2.16 above shows that NO_x decreases with speed until 30-40 mph, and then begins to increase as speed increases.

2.7.3 Variation of Emissions with Driving Mode

Frey et al. (2003) measured emissions as a function of driving mode for 10 light-duty gasoline vehicles tested on 3 sites. The average emission rate for the acceleration mode was the highest for all pollutants for all vehicles. The cruise mode had the 2nd-highest emission rate for all vehicles. In almost all cases, deceleration was the 3rd-highest emission rate, with idle emission rate being lowest.

2.8 Emission Models

Emissions estimates are important for developing emissions control strategies, determining applicability of permitting and control programs, ascertaining the effects of sources and appropriate mitigation strategies, and a number of other related applications by users including federal, state and local agencies, consultants and industry. Emission models are hence important for regulation purposes.

2.8.1 Macroscopic (Macroscale) Emission Models

The most widely used macroscale emission model is MOBILE, developed by the U.S Environmental Protection Agency, (USEPA). It was developed based on laboratory dynamometer driving tests. The model predicts carbon monoxide, hydrocarbon and nitrogen oxides emissions as a function of a number of factors, with vehicle speed constituting the only traffic measure. The MOBILE model has undergone significant revision since its initial development. MOBILE6 is hence the most recent updated version. Other models such as NONROAD, PART5, and MOBTOX are included in the MOBILE6 package to estimate off-road emissions, PM emissions and air toxics emissions, respectively.

According to Heirings et al. (2001), the primary revisions included in MOBILE6 are light duty vehicle basic exhaust emission rate, start emissions, speed/cycle corrections, inspection and maintenance, gasoline fuel composition, air conditioning, diurnal and resting losses, “gross liquid leakers,” heavy-duty vehicles, fleet characterization, trip characteristics, new emission standards and idle emissions. The model is currently used to demonstrate conformity of transportation and air -quality plans, develop emissions inventories and provide emission estimates for dispersion and photochemical air -quality modeling.

MOBILE is designed for estimating regional emissions for input into a regional photochemical model. MOBILE6 reports area-wide emission rates for each vehicle type in grams of pollutant per vehicle mile travelled. MOBILE6 enables the modeler to input the average speed by roadway type and then calculates area-wide emission factors as a

weighted sum of the four facility-specific emission factors (local/ collectors, arterial, freeways, and ramps). Emission rates can be combined with estimates of travel activity (total vehicle miles traveled or VMT) to develop highway vehicle emission inventories expressed in tons per year, day, month, or year. While MOBILE6 is useful for developing regional emission inventories, it lacks the temporal and special resolution needed to evaluate the "micro" scale impact of MSERS, such as signal timing optimization and intersection improvements (Frey et al, 2003). To evaluate the air quality benefits of MSERs, it is necessary to evaluate localized changes in emissions at a fixed location, which macroscopic models are not designed to do (Frey et al, 2001). MOBILE6 basis its emission factors on standardized driving cycles and cannot be adjusted to account for increased or decreased numbers of accelerations due to MSERS. The latest version of MOBILE, the regional mobile emissions model, estimates CO_2 emissions, but its estimates according to the User's Guide are "very simplistic". They are based on fuel economy, but are not adjusted for speed, temperature, or fuel content. SYNCHRO is a much more simplified macroscopic emission model that predicts vehicle emissions by first predicting fuel consumption, calculated as a function of vehicles miles, total delay in vehicle-hr/hr, and total stops in stops per hour. Vehicle emissions are then estimated by multiplying the fuel consumption by an adjustment factor, which differs depending on the type of emissions (CO, NO_x or HC). SYNCHRO is known to have shortcomings in terms of accuracy, and also does not include CO_2.

Transyt-7F, PasserII-90, HCS and SIGNAL97 are other macroscopic traffic models but unfortunately, they do not include emission predictors (Unal et al., 2003).

2.8.2 Microscopic (Microscale) Emission Models

2.8.2.1 Look-Up Tables

The most basic form of the microscale emissions model is a multidimensional lookup table that stores emissions values corresponding to a combination of speed and acceleration or deceleration, and/or speed in the last several seconds of vehicle operation - (Barth et al., 2000). The values in look-up tables have historically come from dynamometer testing, but may also come from on-board testing.

CORSIM is an example of software that uses look-up tables to estimate emissions. CORSIM determines the total emissions on each link by applying default emission rates, based on speed and acceleration, to each vehicle for each second the vehicle travels on the given link (Lederer, 2001).

Look-up tables are straightforward to implement and the computational cost is very low. However, look-up tables cannot explicitly account for the time dependence of emissions in response to vehicle operation. It is also inconvenient to introduce other influencing factors such as road grade into such a model. Also, the emission factors in CORSIM have not been kept up-to-date.

2.8.2.2 MEASURE Model

The Mobile Emission Assessment System for Rural and Urban Evaluation (MEASURE) model is a second microscale mobile emission model recently developed by researchers at Georgia Institute of Technology. The model development was conducted under a cooperative agreement with the U.S Environmental Protection Agency (EPA) and Federal Highway Administration. MEASURE predicts on-road emissions as a function of vehicle operating mode (cruise, acceleration, deceleration, idle), specific vehicle characteristics (including model year, engine size), operating conditions (road grade, traffic flow, etc.) and speed/acceleration profiles. The model was developed within a GIS framework, which allows for facility-level aggregations of microscopic traffic simulations or disaggregation of traditional macroscopic four-step travel demand forecasting models to develop emission-specific vehicle activity data (Guensler, 1998).

The vehicle technology module, vehicle activity module, vehicle emission module and reporting module are the four major modules of MEASURE. While the vehicle technology module takes regional vehicle registration data and outputs location and time specific emission technology group distribution, the vehicle activity module takes regional planning model results and joins them with appropriate speed and acceleration lookup tables to produce location and time specific estimates of emission specific modes of vehicle activity. Weighted, least-square, regression models developed from large databases of vehicle emission tests (iterative, linear approach, etc.) constitute the vehicle emissions module. Finally, the reporting module combine estimates into a

gridded, hourly, format that is used as input into a regional photochemical models (Guensler, 1998).

Although a great deal of unexplained variability exists even within the enhanced analyses of the MEASURE model, the model framework is capable of providing significant improvements over the current average speed models. It uses over 30 variables as inputs, however, and thus is complicated to implement.

2.8.3 Integration

INTEGRATION, another microscopic traffic model, computes fuel consumption for each vehicle on a second-by-second basis as a function of speed and acceleration. The vehicle's tailpipe emissions are then estimated on a second-by-second basis as a function of the fuel consumption, ambient temperature, and the extent to which the vehicle's catalytic converter has already been warmed up during an earlier portion of the trip (Lederer, 2001). Dynamometer testing data was used in developing the model and a light-duty gasoline vehicle was used in the data collection.

2.8.4 Ahn's Non-Linear Regression Model and Neural Network Model

Ahn (1998) developed models for estimating emission rates using two main predictor variables: vehicle velocity and acceleration. Second-by second measurements of light-duty gasoline vehicle emissions made by Oak Ridge National Laboratory were used in the development of the models. Dynamometer testing data from the laboratory was used in developing the models and light-duty gasoline vehicles were used in data collection. As part of this research, two types of mathematical models, nonlinear regression models and neural network models were studied.

2.8.4.1 Non-Linear Regression Model

$$Log(MOE_e) = \sum_{i=0}^{3}\sum_{j=0}^{3}\left(k^e{}_{i,j} \times s^i \times a^j\right) \qquad \text{Eqn. 2.1}$$

Where MOE_e = Fuel consumption and emission rates (liters/hr or mg/s)

k = model regression coefficient

s = speed (m/s)

a = acceleration (m/s^2)

2.8.4.2 Neural Network Models

$$MOE_e = F^3\left(W^3F^2\left(W^2F^1\left(W^1p + b^1\right) + b^2\right) + b^3\right) \qquad \text{Eqn.2.2}$$

Where MOE_e = Fuel consumption and emission rates (liters/hr or mg/s)

W^1, W^2, W^3 = model coefficients

b^1, b^2, b^3 = bias coefficients

p = an input vector containing pairs of speed and acceleration used as predictor variables

F^1 = nonlinear transfer function (hyperbolic tangent sigmoid, $F= 1/(1+e^{-n})$

F^2 and F^3 = nonlinear transfer function (logarithmic sigmoid, $F= (e^n-e^{-n})/(e^n+e^{-n})$

The R^2 values are 0.95 for the Non-linear regression model and 0.99 for the Neural network models. These show very good model fit.

2.8.5 Fomunung's Statistical Model for Nitrogen Oxides from Light-Duty Gasoline Vehicles

Fomunung (1999) developed a statistical model for estimating nitrogen oxide emissions from light duty gasoline vehicles. Data used was obtained from in-use vehicle

emissions testing database compiled by the USEPA, Office of Mobile Sources. Light-duty gasoline vehicles were tested for this model.

$$E_p(g/s) = 0.0259 \times FTPbag2 \times anti\log\left\{\begin{array}{l} 0.0225(AVGSPD) + 0.3434(IPS.120) + \\ 0.6329(ACC.6) + 0.0247(DEC.2) + 0.0083(finj1) + \\ 0.0028(finj2) - 0.0021(cat1) + 0.0026(cat2) + \\ 0.0003(cat3) - 0.0085(flag1) - 0.0068(flag2) \end{array}\right\} \quad \text{Eqn.2.3}$$

where:

E_p= predicted emission rate under (g/s) under tested driving conditions.

AVGSPD = average speed of cycle in miles per hour.

IPS.120 = percent of cycle time spent with inertia power surrogate greater than 120 mph^2/s.

ACC.6 = percent of cycle time spent accelerating at rate greater than 6mph/s.

DEC.2 = percent of cycle time spent decelerating at rate greater than 6mph/s.

finj1 = an interaction variable between fuel injection type carburetor and odometer reading less than 25,000 miles.

finj2 = variable representing that have carburetors with odometer readings between 25,000 miles and 50,000 miles.

cat1 = variable for vehicle that have only oxidation type catalyst and odometer reading between 50,000miles and 100,000miles.

cat2 = variable for vehicle that have 3-way catalyst type convertor and mileage between 25,000 miles and 50,000 miles.

cat3 = variable for vehicle that have 3-ay catalyst type convertor and mileage between 50,000 miles and 100,000 miles.

flag1 = variable with fuel injection type port with odometer reading and odometer reading between 50,000 miles and 100,000 miles and is also a high emitter.

flag2 = variable with throttle body fuel injector type and odometer reading between 50,000 miles and 100,000 miles and is also a high emitter.

The R^2 value for this model is 0.62. This demonstrates a reasonable fit.

2.8.6 Hung, Tong, and Cheung's Modal Emissions Model for HC, CO, and NO_x

Hung, Tong, and Cheung (2005) developed a vehicle emissions model based on mode.

Their model is useful in evaluating microscale impacts of local traffic management schemes. It is based on on-road data collected from 4 vehicles in Hong Kong. HC, CO, and NOx emissions are predicted as functions of vehicle speed for acceleration, deceleration, and cruising modes.

2.8.7 Toth-Nagy's Artificial Neural Network Model for CO and NO_x

Toth-Nagy et al. (2006) developed an artificial neural network-based model for predicting emissions of CO and NO_x from heavy-duty diesel conventional and hybrid vehicles.

ANN models were trained on engine and exhaust emissions data collected from dynamometer tests and then used to predict emissions based on engine speed and torque data. Predictions were then compared with actual emissions data collected from chassis dynamometer tests of similar vehicles. The methodology sounds promising, but was not applied to gasoline vehicles.

2.8.8 Jazcilevich's Emissions Model Based on On-Board System Data

Jazcilevich et al. (2007) developed a methodology for estimating vehicular emissions using a car simulator, a basic traffic model, and a geographical information system.

The experimental data for the car simulator was obtained using on-board measurement system (Semtech) data from 17 vehicles and laboratory Fourier transform IR (FTIR) measurements with a dynamometer following typical driving cycles. The car simulator uses this data to generate emission factors every second for HC, CO, CO_2, and NO_x. These emission factors, together with information on car activity and velocity profiles in conjunction with a basic traffic model, provide emissions per second of a sample fleet. The geographical information system was used to localize these road emissions, with Mexico City used as a case study.

2.8.9 A Statistical Model of Vehicle Emissions and Fuel Consumption based on Second-by Second Dynamometer Testing Data

In the past, vehicle emissions models were mainly developed for CO, NO_x, and HC. CO_2 models were scarcely developed, until recently when global warming and climate change has become a global issue due to the alarming rate at which greenhouse gases increase in the troposphere.

Most CO_2 emissions models developed from vehicular sources are macro scale and are based mainly on average speed. Data used in developing such models were obtained from fixed driving cycles, which do not adequately capture the effects of driving dynamic and vehicle dynamics on emissions. CO_2 emissions are poorly predicted from these models.

A second by second model is better able to predict vehicle emissions, not only because it addresses the dynamics of driving but also takes into account traffic variables that could not be measured by a macro scale model, when predicting emissions.

This article (A Statistical Model of Vehicle Emissions and Fuel Consumption) used regression to develop an equation to predict CO_2. Regression based models typically employ functions of instantaneous vehicle speed and acceleration as explanatory variables. This helps in overcoming limitations such as sparseness and non flexibility.

"This paper presents EMIT (EMIssions from Traffic), which is a simple statistical model for instantaneous tailpipe emissions (CO_2, CO, HC and NO_x) and fuel consumption. The paper derived its explanatory variables from a load-based approach, in order to realistically reproduce the emissions behavior".

Data used for developing the EMIT model was obtained from NCHRP (National Corporative Highway Research Program) vehicle emissions database. The data was collected using a dynamometer testing procedure, where second-by- second speed, engine out and tailpipe emissions rates of CO_2, CO, HC and NO_x were measured for 344 light-duty vehicles. Only speed and emissions data were obtained from the database, and acceleration was computed from the change in second-by-second speed per unit time.

Using the data, the tailpipe CO_2 emissions model was developed as:

$$\boxed{TP_{CO_2} = \begin{cases} \alpha_{CO_2} + \beta_{CO_2} v + \delta_{CO_2} v^3 + \zeta_{CO_2} av, if P_{tract} > 0 \\ \alpha'_{CO_2}, if P_{tract} = 0 \end{cases}} \qquad \text{Eqn.2. 4}$$

where:

$$P_{tract} = A.v + B.v^2 + C.v^3 + M.a.v + M.g.\sin\vartheta.v$$

v: vehicle speed (m/s)

a: vehicle acceleration(m/s^2)

A: rolling resistance term (kW/m/s)

B: speed-correction to rolling resistance term (kW/(m/s)^2)

C: air drag resistance term (kW/(m/s)^3)

M: vehicle mass (kg)

g: gravitational constant (9.81m/s^2)

ϑ: road grade (degrees)

2.9 Fuel Consumption Models

Many factors that influence fuel consumption also influence emission rates. Models developed to estimate fuel consumption can thus indicate the anticipated form of equations for estimating emissions. Four such fuel consumption models are listed below.

2.9.1 Acceleration Term Models (S.E Moore, 1982)

Watson, Milkins and Marshall (1979) developed a set of relationships to estimate fuel consumption and emissions as a function of average speed, v and positive accelerations, PKE. These relationships are of the form:

$$E_i = k_{i,1} + \frac{k_{i,2}}{\bar{v}} + k_{i,3}\bar{v} + k_{i,4}PKE \qquad \text{Eqn2.5}$$

where PKE is defined as:

$$\boxed{\sum \frac{\left(v_f^{\,2} - v_s^{\,2}\right)}{D}}$$ for dv/dt> 0

$k_{i,1,2,3,4}$ are constants, v_f and v_s are final and start velocities for positive accelerations, i relates to different expressions for fuel consumption or HC, CO, NOx, CO_2 emissions.

The constants in the above equation were estimated, using regression analysis, for a single vehicle.

2.9.2 The Vehicle Mix Model (Lecture Notes, Ardekani, Fall 2006)

$$\boxed{F = K_1 + K_2 T} \qquad \text{Eqn.2.6}$$

Where F = Fuel Consumption (Lit/Km)

T = temperature in what units (Degree Fahrenheit)

K_1 = a + bM M = Engine Mass (Kg)

K_2 =C. EC, where EC = Engine Capacity (Liters)

2.9.3 Transyt – 7F Model (Lecture Notes, Ardekani, Fall 2006)

$$\boxed{F = K_1 TT + K_2 D + K_3 S} \qquad \text{Eqn. 2.7}$$

Where F = Fuel Consumption (units?)

TT = Vehicle-Miles Traveled (miles)

D = Vehicle Hours of Delay (hours)

S = Number of Stops

2.9.4 Steady Speed Model **(**Vincent et al: Lecture Notes, Ardekani, fall 2006)

$$\boxed{F = a + bV_x + cV_x^{\,2}} \qquad \text{Eqn. 2.8}$$

Where, F = Fuel consumption (liters/km); V_x = Cruise speed (km/hr)

2.9.5 Steady Speed Model (Post et al: Lecture Notes, Ardekani, Fall 2006)

$$F = a + \frac{b}{V_x} + cV_x^2 \qquad \text{Eqn. 2.9}$$

Where, F = Fuel consumption (liters/km); V_x = Cruise speed (km/hr); b<0

The models developed are linear regression models containing vehicle speed, vehicle acceleration and acceleration x velocity term. Vehicle acceleration in my models explains the dynamics of driving. However, in the fuel consumption models, vehicle velocity is the only variable.

The positive signs of vehicle velocity and acceleration are plausible and consistent with our expectation. The findings suggest that CO_2 emissions increase with velocity and acceleration, which is consistent with theory and current models.

2.10 Objective of the Research

The objective of the research is to develop a microscale model for estimating vehicle emissions of carbon dioxide. A statistical (least square regression) approach using statistical analysis system (SAS) software will be employed during the model development process.

The model developed in this research will overcome shortcomings associated with some of the existing models:

- MOBILE: The MOBILE model currently in use has average velocity and vehicle mile traveled as the major variables that impact vehicle emissions. It, however, does not address driving dynamics (acceleration, deceleration), which contributes significantly to vehicle emissions. It cannot, thus, be used to evaluate emission impacts

of transportation system measures that affect accelerations and decelerations, such as traffic signal coordination.

- SYNCHRO: SYNCHRO predicts emissions by first predicting fuel consumption. Vehicle miles traveled and numbers of stops are the other parameters used by this model. It, however, does not address driving dynamics, and hence cannot be used to evaluate emissions impacts of transportation system measures that affect accelerations and decelerations. SYNCHRO is known to have shortcomings in terms of accuracy, and also does not include CO_2.
- Look-Up Tables (CORSIM): Look-up tables cannot explicitly account for the time dependence of emissions in response to vehicle operation. It is also inconvenient to introduce other influencing factors such as road grade into such a model. Also, the emission factors in CORSIM have not been kept up-to-date.
- MEASURE: In the case of the Georgia Tech MEASURE model, many variables were considered, some of which negligibly impact vehicle emissions, but which make the model complicated and difficult to use.
- INTEGRATION: INTEGRATION predicts emissions from computed fuel consumption as a function of velocity and acceleration obtained from a dynamometer test.
- Ahn's Models: The models are complicated to use in emission estimation.
- Fomunung's Model: This model only applies to NO_x. Like the MEASURE model, this model estimates emissions by considering several variables, some of which negligibly impacts emissions.

- Hung, Tong, and Cheung's Modal Emissions Model**:** Since it was developed for Hong Kong, it is not necessarily representative of vehicles with US emission controls. In addition, the model does not distinguish among various accelerations within the acceleration mode category, which can produce substantially different emissions. Also, it does not include CO_2.
- Toth-Nagy's Artificial Neural Network Model. The methodology sounds promising, but the work was conducted for heavy-duty diesel vehicles and not for gasoline vehicles. Also, it does not include CO_2.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Previous chapters have discussed the benefits of addressing driving dynamics, which some emission models did not consider. Research work has shown that vehicle emissions are significant in the acceleration mode; hence, analysis of vehicle activity is important in emission estimation. The focus of this research is hence on developing a model that incorporates acceleration and deceleration. Unlike other models, such as the MEASURE model that used GIS based software in its analysis, this research uses Statistical Analysis System software in most of its analysis and the model building process.

The model will be developed using data collected under real world traffic conditions using an On-Board System (OBS) emissions measurement unit, in contrast with other models that use dynamometer data in their model building process. This chapter discusses the study equipment and its accessories, the data collection procedure and the statistical tools employed in the analysis and model development.

3.2 Study and Data Collection Equipment

3.2.1 Study Vehicle

The study vehicle used was a rented vehicle from Enterprise Incorporated, located on East Abrams, Arlington, Texas. The Civil Engineering Department of the University of Texas at Arlington (UTA) acquired the car for this study. It was a 2007

gasoline-powered Dodge Charger, shown in Figure 3.1 below. Information about the car is given in Table 3.1.

Figure 3.1 2007 Dodge Charger

Table 3.1 Specifications of 2007 Dodge Charger Car

Parameter	Value
Standard Engine	2.7 L V6
Power	149KW, 200 HP @ 5,800 rpm
Power	190 ft lb, 258 Nm@ 4,850 rpm
Fuel Tank capacity	18 gallons
Fuel type/ system	Gas engine / Sequential electronic fuel injected
Standard transmission	4 speed automatic
Cylinders	6
Compression	9.7
Weight, lb	3820

3.2.2 OBS -1300

The equipment used in the on-road data collection was acquired from Horiba Instruments. It is an On-Board emission measurement System (OBS 1300) installed in the on-road study vehicle and which measures the exhaust gases of the vehicle. The equipment is mainly composed of two on-board gas analyzers, a laptop computer equipped with data logger software, a power supply unit, a tailpipe attachment and other accessories. The OBS measures nitrogen oxides (NOx), carbon monoxide (CO), carbon dioxide (CO_2) and hydrocarbons (HC).

The various components of the OBS1300 set up are described in details as follows:

3.2.2.1 Batteries

The batteries, shown in Fig. 3.2, are the source of electrical energy on which the whole OBS runs.

Figure 3.2 Two 12V Deep Cycle Batteries

The OBS setup runs on two deep cycle batteries of 12 V each, acquired from Trojan batteries. These two batteries connected to the Power Supply Unit (PSU) should last for around four hours after fully charged. A battery monitor is also provided to monitor the voltage of the batteries.

3.2.2.2 Power Supply Unit (PSU)

Figure 3.3 shows the PSU unit.

Figure 3.3 Power Supply Unit (PSU)

The PSU performs a dual function. It converts the battery output power (24 V in this case) to AC current and supplies it to all the units of the OBS-1300 setup while collecting data in the field. It also serves as a battery charger by converting AC input power to DC current. The PSU is connected to the battery and data integration unit.

3.2.2.3 Data Integration Unit (DIU) and MEXA-720 HNDIR

The data integration unit (DIU) houses the Mexa-720 NO_x analyzer, which measures the nitrogen oxide emissions. It also acts as an interface unit for each sensor,

and analyzer of the entire OBS-1300 setup as well as the data logger PC (a Dell laptop in this case). Figure 3.4 shows the DIU unit and the MEXA-1170 HNDIR unit.

Figure 3.4 DIU and MEXA-1170 HNDIR Unit

The individual MEXA-720 analyzer unit and the NO_x sensor are shown in Figure 3.5.

Figure 3.5 MEXA-720 Unit and NO_x Sensor (Source: Horiba Instruments Inc.)

The MEXA-720 NO_x analyzer is a non-sampling type zirconium sensor that measures NO_x concentrations. It comes pre-installed in the DIU unit, as discussed earlier.

Figure 3.6 shows the NO_x sensor probe being attached to the tail pipe attachment.

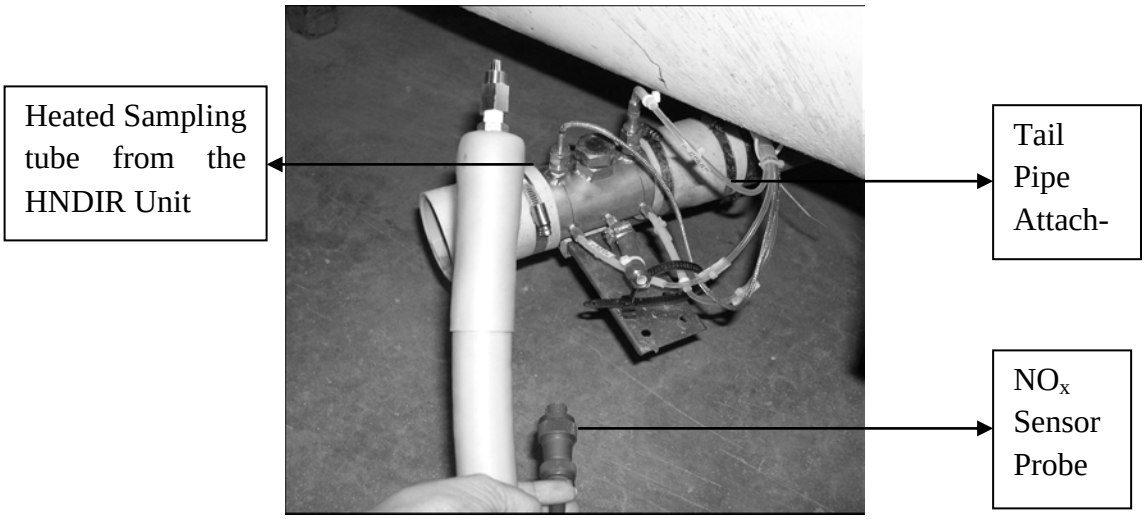


Figure 3.6 NO_x Sensor Probe Being Attached to the Tail Pipe Attachment

3.2.2.4 MEXA-1170 HNDIR Unit

This unit uses a Heated Non-Dispersive Infrared (HNDIR) detection technique to measure CO, HC and CO_2 emissions exiting the tailpipe of the car. The HNDIR technique is based on the principle of selective absorption. Here a gas species absorbs infrared radiation in an amount directly proportional to its molecular concentration.

The HNDIR unit has a heated tube attached to the tail pipe attachment, as shown in Figure 3.6, which takes in the sample for analysis. This unit also connects the analog output to DIU.

3.2.2.5 Data Logger PC

A DELL laptop loaded with an OBS-1000 series data logging software is provided. A PCMCIA card is included with it for data import. This unit logs the earlier indicated pollutant concentrations, A/F ratio, exhaust pipe temperature, ambient temperature and ambient humidity data.

3.2.2.6 Remote Controller

A remote controller is provided for the rider (or data logger), who takes care of the data logging procedures on the laptop during the run, to operate the MEXA-1170 HNDIR unit from the back seat of the study vehicle. Figure 3.7 shows the remote controller.

Figure 3.7 Remote Controller

It holds calibration buttons like “CAL”, “ZERO”, and “SPAN” and other buttons like “PURGE”, “RESET” and “MEASURE”. The remote facilitates the tasks of HNDIR unit calibration and initiation of data sampling functions for the rider at the back with the earlier mentioned buttons.

3.2.2.7 Geographic Positioning System (GPS)

A GPS unit is provided in this setup to log the velocity, latitude and altitude of the vehicle when on a data collection run. It is placed at a convenient location on the side of the van just outside of the window to obtain accurate GPS points throughout the run.

3.2.2.8 Humidity/Temperature Sensor

A humidity/temperature sensor is also provided in the OBS-1300 setup that measures the humidity and ambient temperature of the outside air on a second by second basis. It acts as a mini weather station for the whole system.

3.3 Data Collection Procedures

3.3.1 Introduction

The data collection and preparation procedures were the most intensive, complicated and challenging areas of the study. Large amounts of data were collected on a second –by-second basis. As mentioned earlier, the data collection for this research project was carried out using an OBS-1300 system, which was installed in the 2007 Dodge Charger car.

The test procedure was 40 hours of second-by-second emissions data collection. 20 hours of the 40 hours was from a highway facility and the remaining 20 hours was collected from an arterial facility. For each of the facilities, 10 hours of the data was collected during the peak hour time period, defined as times from 6:30 am to 9:00 am (morning peak) and 4:00 pm to 6:30 pm (evening peak). The other 10 hours of data was collected during the off-peak time periods, defined as times from 9:00 am to 4:00 pm. Data collection starts Monday afternoon at1:00 pm and ends at 9:00 am or 10:30 am on Friday, on a weekly basis. Monday morning and Friday afternoons are exempted due to extreme traffic density which might not be representative of the rest of the data. Data collection was carried out from January through May of 2007.

3.3.2 Data Collection Routes

3.3.2.1 Arterial Test Track

Figure 3.8 shows the arterial route followed during this research. From UTA Blvd., travel North on Cooper to Division; East on Division to Collins; South on Collins to Pioneer; West on Pioneer to Cooper; North on Cooper to UTA Blvd.

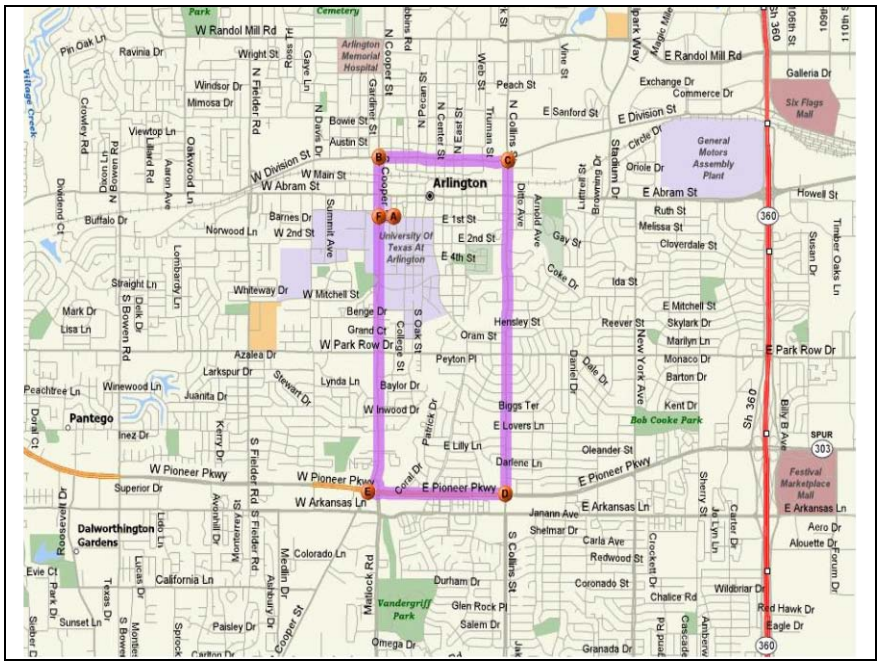


Figure 3.8 Arterial Test Track

3.3.2.2 Highway Test Track

Figure 3.9 shows the highway route followed during this research. From Cooper, travel west on I-30 to 820; South on 820 to I-20; East on I-20 to Spur-408; North on Spur-408 to Loop-12; North on Loop-12 to I-30; West on I-30 to Cooper.

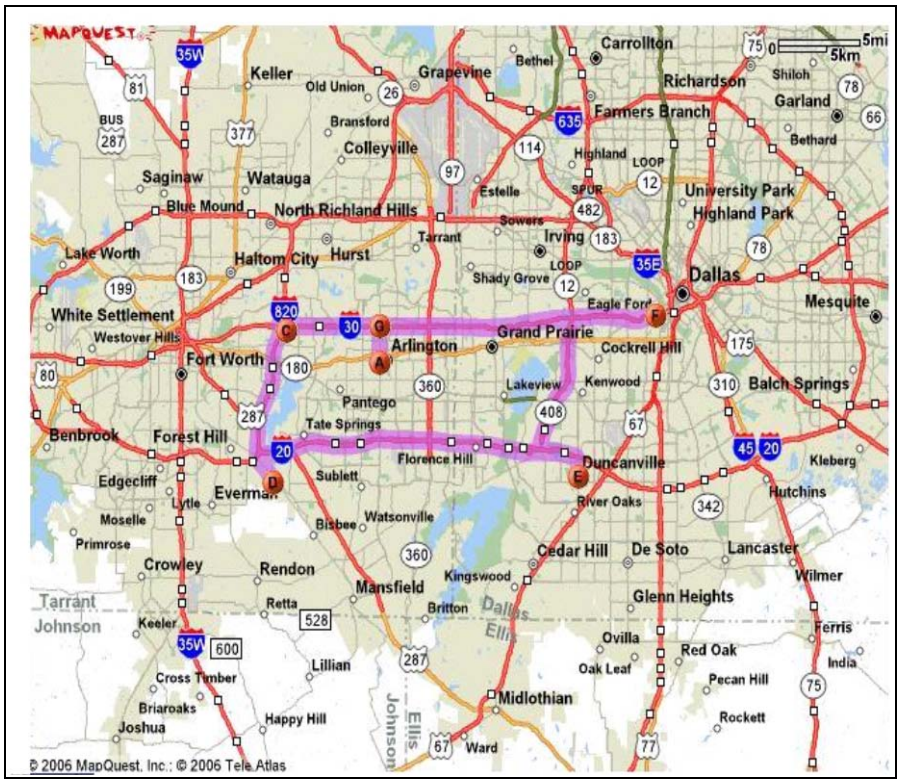


Figure 3.9 Highway Test Track

3.3.3 Data Quality Control

The procedures listed below were conducted to ensure quality data:

1. Warming up of the OBS-1300 system was carried out for 45 minutes to 1 hour before starting data collection. This was done to tune the system to measure and log accurate data.

2. The required calibrations for the analyzers as well as the NOx sensor were duly carried out to maintain the accuracy of the whole setup.

3. The data collection was carried out during the days when the university was in session, as traffic patterns in and around the city of Arlington was greatly impacted by this.

4. Testing was conducted from Monday afternoon through Friday morning.

5. No collection of data was carried out during rainy days, since water can damage the NO_x sensor.

6. The speed of the vehicle during data collection was approximately maintained with the flow speed of the traffic. This was done to imitate real traffic conditions.

7. The two 12V batteries used for the system were fully charged before the start of data collection. Data collection was immediately suspended whenever the voltage reading showed less than 21V.

8. Regular inspections were carried out for the OBS to see to that the emissions coming from the exhaust and those measured by the device were not varying with time and usage.

3.3.4 Configuration of Data Logging Software

For the data logging software to log the correct values of the measured emissions and other required parameters, the software was configured to a set of values provided by the Horiba Instruments, Inc. Table 3.2 lists the configured set of values.

Table 3.2 Configured Value Ranges in the Analog Digital Conversion Setup

Parameters	Range of Values	Units
NO_x	0-4989	ppm
Air to Fuel Ratio (AFR)	0-100	
Exhaust Temperature	0-1000	Deg. C
Exhaust Pressure	0-200	kPa
Ambient Temperature	0-150	Deg. C
Ambient Pressure	0-100	kPa
Ambient Humidity	0-100	%
Velocity	0-500	kmph
Revolutions	0-5000	rpm

In addition, it was estimated by that there would be a 1 second delay in the logging of the NO_x data and 3 seconds for HC, CO and CO_2 data. Horiba Instruments Inc. attributed this delay to the time it took to convert the measured concentration from analog to digital output on the laptop. The data analysis spreadsheets were adjusted to account for this delay.

3.3.5 Warm-Up and Calibration Procedures before Data Collection Session

Before the start of data collection every day, it is necessary to have two 12V batteries (DC) completely charged. The following are the steps for the warm-up and

calibration procedure that were followed before a morning or afternoon data collection session (typically lasting 4 hours):

- Using AC Power, turn on the DIU followed by the HNDIR unit.
- Let the system warm up for around 45 minutes to 1 hr., then switch off both units and change the power source from AC to DC source by connecting the units to the two 12 V batteries.
- Turn on the DIU then the HNDIR unit.
- Open the valve of the zero gas cylinder making sure it is set at 100kPa.
- Purge for 5 minutes by pressing the “PURGE” button on the HNDIR unit.
- Press “RESET”.
- Press “ZERO” button on the HNDIR unit. Allow 30 to 60 seconds.
- Connect the span gas cylinder to the HNDIR unit and open its valve.
- Press “RESET”.
- Press “SPAN” button and make sure enough span gas is flowing through the units. This can be known when the ball within the MEXA 1170HNDIR compartment in front of the unit rises up between two blue bands. Wait for around 2 minutes.
- Press “RESET”.
- Press “CAL” and the OBS-1300 will do both the zero and span calibrations. It will reset by itself at the end of the process.

3.3.6 Calibration and Data Collection Procedures during a Data Collection Session

The ‘runs’ for data collection were started every day only after the earlier mentioned warm-up and calibration procedures were completed. The OBS-1300 setup

runs on the DC power of the two batteries for a maximum of 4 hours during data collection. The calibration and data collection procedures during a 4 – hour data collection session are as follows:

a) Turn on the data logger PC and start the OBS-1300 data logging software.

b) Put the system in PURGE by pressing the PURGE button on the remote controller.

c) Perform a pitot tube calibration on the laptop to stabilize the exhaust flow rate value. It is OK to allow a tolerance or accuracy level of between -0.05 to .05.

d) Now start the engine by turning on the ignition.

e) Activate the GPS if not activated by selecting the appropriate port in which a satellite's signal is available.

f) Enter a filename and get ready to start logging data by activating the "START" button on the Laptop.

g) Now, with the help of the remote controller, put the system in 'MEASURE" mode.

h) Now, start data logging by clicking on the START button when the desired starting point is reached and saving of data begins in the file.

i) Stop data logging at the end of the run and turn off the engine of the car.

j) REPEAT steps 'b' to 'h' for the second and subsequent runs until the assigned hours of data collection are completed.

3.3.7 Weekly Calibration of NO_x Sensor

NO_x sensor calibration was carried out every week throughout the data collection period. This is required to help maintain accuracy of NO_x emissions measurement. The setup for the NO_x sensor calibration is shown in Figure 3.10. The following are the steps involved here:

- A calibration unit is employed in the NO_x sensor calibration. It consists of a sensor adaptor, a flow meter, bubbler, and water inlet.
- The NO_x sensor has to be fixed in the adaptor of the calibration unit.
- The calibration gas used for this process is oxygen free N_2, and the flow of this gas from the cylinder is regulated through a regulatory valve.
- The exhaust outlet of the calibration unit is connected to a long Teflon tube, through which the calibration gas is safely discharged into the outside air.

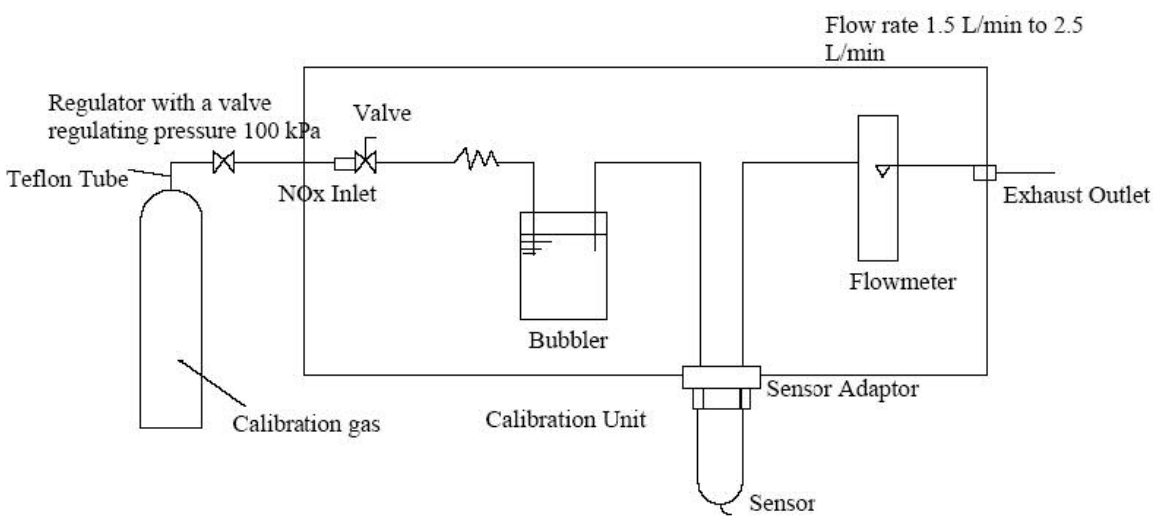


Figure 3.10 NO_x Sensor Calibration Setup (Adapted from Horiba's OBS Instruction Manual)

- Before calibration, the NO_x sensor is connected to the calibrator, as shown in Figure 3.10 above.
- Teflon tubing is also connected from the calibration gas to the calibrator, as shown in Figure 3.10.
- The valve of the calibration gas is turned on, making sure the regulator valve is set to 100kPa.
- The yellow valve on the calibrator is slowly turned on till the ball in the flow meter rises up between two blue bands.
- Allow about 2 minutes for the ball to equilibrate.
- Now, press CAL/SET key on the Mexa-720, the zero LEV will blink. This sets the sensor to zero.
- Press CAL/SET key again, the span LEV will blink. This sets the NO_x sensor to the concentration of the calibration gas on the gas bottle. This completes the calibration steps.
- Finally press CAL/SET and press ENTER.

3.3.8 Maintenance and Diagnostics Procedures

To ensure consistently smooth and good data collection without frequent interruptions due to OBS 1300 unit malfunction, inability of batteries to stay charged and calibration issues throughout the period, proper maintenance and diagnostic procedures were followed.

3.3.8.1 Maintenance Procedures

(i) Batteries

1. Set the charger at 12V, 2Amp when charging the batteries overnight. One can set charger at 12V, 10Amp when charging batteries for a shorter period of time (say 2-3 hours).

Problem: Leaving the batteries to charge overnight at 12V, 10Amp or for longer hours causes the acid to boil up and destroys the lead.

2. Check the acid level of the battery at least once every 3 weeks and top it off only with *distilled water* in case the acid level has gone down.

(ii) OBS 1300 Unit

1. Need regular calibration of unit. The calibration procedures are described in Sections 3.3.5 and 3.3.6.

2. Loosened CO, HC, CO_2 cable connections: Need to check and tighten all connecting cables once every two weeks.

3. Check for hydrocarbon hang up in the heated line by performing the following every two weeks:

- Disconnect the heated line from the tailpipe and connect it to a tee.
- Put a short piece of tubing on one end of the tee before opening to ambient pressure.
- Connect the other leg of the tee to the nitrogen zero gas: use a needle valve or regulator to limit the flow out of the gas cylinder.

- Flow gas and adjust until you just get enough flow out of the open leg of the tee that will make a piece of paper flutter; this should be checked while in measure.
- Do this right after zero calibrating the analyzers.
- When in measure and pulling zero gas through the heated line, look for high hydrocarbon signals.
- If you are seeing high numbers, then try to clean the heated line with hot water.

3.3.8.2 Diagnostic Procedures

After driving around and collecting data, it is possible that expansion, contraction and bumping back and forth on the roadway could cause the connections to loosen up. Drifting of OBS 1300 unit could also result from longer periods of driving. This can affect the emission values and most importantly the flow rate value of the pitot tube.

(i) Negative Flow Rate Values

Possibilities of getting negative flow rate values and how to solve them:

1. Negative flow rate values can indicate loosened pressure lines connected to the tailpipe attachment and at the back of the DIU. Hence those connections must be tightened.

2. Another possibility is a **"**kink" in the pressure line can result in negative flow rates. The line needs to be straightened.

3. Drift can occur due to constant usage of units, resulting in negative flow rate values; hence, the operator needs to purge the system and flow zero gas as well (see checklist for purging & zeroing). Also calibrate the pitot tube.

Pitot tube calibration on the lap top**:** Each time we calibrate, we have to allow about 1 to 2 minutes for the flow rate values to stabilize; a good flow rate range to stop calibration and start data collection is -0.05 - 0.05m^3/min. That is when the flow rate value bounces between these numbers, then we are OK to start logging data.

Also, it is very important that we make sure that the engine's ignition is turned off before we do the pitot tube (flow) calibration.

(ii) Inconsistent CO, HC, CO_2 , NO_x, Values

Possibilities of inconsistent CO, HC, CO_2 or NO_x, emission values and how to solve them:

1. Use of unit for a long time without calibration. Need to calibrate Unit and NO_x Sensor connected to the tailpipe attachment.

2. Condensed H_2O in exhaust line connected at the back of the Mexa 1170 HNDIR. Need to drain condensed water in the exhaust line.

3. Hand held remote is not set to MEASURE just before clicking LOG START to begin logging data. We must hence make sure we set the hand held remote to MEASURE before we start logging data.

4. It is also important we check for any leak of the sample line monthly.

3.3.9 Installation Procedure

The OBS 1300 is installed in any test vehicle by using the following procedure:

1. Connect Analog cable at the back of MEXA 1170 HNDIR **to** XMS2S- 37 at the back of the battery monitor.

2. Connect GPS DIU line at the back of the DIU **to** RS-232C (GPS) at the back of the PC.

3. Connect Temperature & Humidity sensor at the back of DIU **to** NO 05393 and attach to vehicle using a tape.

4. Connect NO_x sensor line, SFC-8 to NO_x sensor at the back of DIU.

5. Connect Power in -100- 240VAC line **to** power in at the back of DIU.

6. Connect Power out2 (PC) line **to** Power out2 PC (031029-1) at the back of DIU.

7. Connect Power OUT1 at the back of DIU **to** MEXA model THP-B4T-30.

8. Connect D. Press HIG**,** D. Press low from the back of DIU **to** D. Press HIG**,** D. Press low at the tailpipe attachment.

9. Connect Remote control line **to** OMRONXM2S-og at the back of MEXA 1170HNDIR.

10. Connect Main power **to** 10A 250V ~ SNDF1.

11. Connect H-line (heated line) from the tail pipe attachment **to** ZCAT 3035-13 at the back of MEXA 1170HNDIR.

12. Connect NO_x sensor and flow meter to the tailpipe attachment.

3.4 Statistical Modeling

3.4.1 Introduction

One of the challenges of this study is to select an appropriate statistical modeling method for analyzing the data. The method should be able to identify the most important predictor variables, as well as to identify a relationship between the predictor

variables and the response variable. The model variables should be representative of the true variable responsible for accurate on-road emission estimation.

Based upon the research needs, multiple linear regression (MLR) was found to be the most appropriate statistical method to use for this study. Multiple linear regression is a procedure that determines the best fitting regression function (a regression surface or a response surface) among several variables. In general, the multiple linear regression model can be represented as follows:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_i$$

Where:

Y_i= Response variable

X_k= Predictor variables (*k=0,1,2,….p-1,)* ; *p* is number of variables

β_0 = *Y* intercept of the regression plane

β_k = Parameters (*k=0,1,2,….p-1,)* ; *p* is number of parameters

ε_i = Random error in *Y* for observation *i*

3.4.2 Model Building Process

The completion of the data collection phase resulted in the accumulation of a large amount of CO_2 emissions data, which is used as the response variable with average velocity, acceleration, deceleration and power demand (velocity x acceleration) and time of the day (peak/off-peak) representing the predictor (independent) variables.

The data is then divided into a training dataset and testing dataset. The training dataset is used in model estimation, while the testing dataset is used for model

validation. A Statistical analysis system (SAS) software is employed for the data analysis. The model estimation process involves:

(1) Fitting a preliminary model – Plotting the response variable against each predictor variable.

(2) Checking model assumptions:

a. Use the plots of residuals vs. predicted values to verify that the error variance is constant

b. Use the plots of residuals vs. predicted values to verify that the current model form is appropriate.

c. Use a normal probability plot of the residuals to verify that the errors are normally distributed.

d. Use a plot of the residuals over time to verify that there is no serial correlation.

(3) Performing diagnostics:

a. Checking x-y-outliers (Bonferroni outlier test): this test is to identify data points which are suspicious of being outliers.

b. Leverage values: these are diagonal elements of the “Hat matrix”. If an observation of the hat matrix is high, then observation i is x-outlying and has high leverage.

c. Influence: this test is to detect if any data point considered as an outlier and hence removed from the dataset has the potential of affecting the fitted regression function to any substantial extent.

d. Variance inflation (VIF); Variance inflation is used to check the multicollinearity problem. Higher variance inflation represents the increasing variation in coefficients due to high correlation between the predictors.

(4) Where there is enough reason, such as interruption in data collection by a passing train, outliers are removed and preliminary analysis repeated.

(5) Performing any necessary transformations, e.g. squaring of response variable.

At this point model assumptions are rechecked by:

(6) Explaining the regression output

a. Explained variability

b. Significance of predictors

(7) Exploring interactions: Using partial regression plots and adding possibly useful interactions to the set of predictors.

(8) Obtaining a set of at least two potentially "best" models (backward deletion, best subsets, and stepwise regression).

The next stage in the model estimation process is to ensure all predictors are significant at $\alpha = 0.10$ level and multicolinearity is not a problem. This is accomplished by:

a. Verifying model assumptions

a. Checking diagnostics (outliers, leverage, influence, VIF)

b. Discussing and justifying choice of best overall model

(10) Presenting and interpreting meaning of final model.

(11) Discussing fit of the model and interpreting inferences.

Details of the model estimation process will be presented in Chapter 4.

CHAPTER 4

MODEL DEVELOPMENT AND RESULTS

As mentioned earlier, a multiple linear regression approach is used to fit the emissions data. The emissions data from both arterial and highway facilities are used to develop statistical models in this research: carbon dioxide emissions models. Six variables including the response variable were considered during the model building process for both models, with the most significant variables left in the model at the end of the process.

4.1 Building the Preliminary Regression Model for Carbon Dioxide (CO_2)

4.1.1 Variables

Predictor variables used were:

- Vehicle Velocity (Vel) in miles per hour,
- Vehicle Acceleration (Acc) in miles per hour per second,
- Vehicle Deceleration (Dec) in miles per hour per second,
- Power Demand (PD) in mile^{2}per hour^2 per second, (acceleration * velocity)
- Time of Day (TD), (Traffic Period), unitless; 1 or 0 (peak or off-peak).

The response variable used was:

- Carbon dioxides (CO_2) emission rate in grams/sec

Although the OBS will measure ambient temperature and relative humidity, the temperature/relative humidity sensor was not functioning when the data for this

research was collected. We need to include the preceding sentence because you mention in Ch. 3 that the OBS measures temperature and humidity, so someone may legitimately as the question of why they weren't included in the model-building process.

4.1.2 Fitting the Preliminary Model for Arterial Data

Multiple Linear Regression (MLR) is a procedure that determines the best fitting regression function (a regression surface or a response surface) among the several variables. In general, the multiple linear regression models can be represented as follows:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon_i , \qquad \text{Eqn. 4.1}$$

Where:

β_0 = *Y* intercept of the regression plane,

β_k = Parameters (*k=0, 1, 2….p-1,)*; *p* is number of parameters,

ε_i = Random error in *Y* for observation *i.*

To fit a model for the variables presented, SAS software is used. To find the unbiased estimators (Regression Coefficients, b_k) for unknown parameters β_k, the Least Squares Method is used to minimize the residuals error (residual, $e_i = y_i - \hat{y}$) and the assumption $E[\varepsilon_i] = 0$ is considered. After regressing "y_i- Carbon Dioxide" on "x_1- Vehicle Velocity", "x_2- Vehicle Acceleration", "x_3- Vehicle Deceleration", "x_4- Power Demand" and "x_5- Time of Day", the result (SAS output) obtained is as follows:

Table 4.1 Parameter Estimates

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.90912	0.00861	105.55	<.0001	0
Vel	Velocity	1	0.00960	0.00021723	44.19	<.0001	1.05587
Acc	Acceleration	1	0.95488	0.00628	152.11	<.0001	2.51470
Dec	Deceleration	1	-0.05074	0.00571	-8.88	<.0001	2.87665
PD	Power demand	1	0.00759	0.00022749	33.35	<.0001	5.00135
TD	Time of Day	1	0.15151	0.00661	22.93	<.0001	1.00151

From the SAS table in Table 4.1 above:

$b_0 = 0.90912$

$b_1 = 0.0096$

$b_2 = 0.95488$

$b_3 = -0.05074$

$b_4 = 0.00759$

$b_5 = 0.15151$

So, the estimated regression function is:

$$\boxed{\hat{y} = 0.90912 + 0.0096x_1 + 0.95488x_2 - 0.05074x_3 + 0.00759x_4 + 0.15151x_5}$$

Eqn. 4.2

At this point, the regression coefficients b_1, for example, can be interpreted as the change in response variable per unit increase in the x_1 variable when all other predictor variables remain constant. However, due to correlation between predictor variables, this might not be the exact case.

4.2 Evaluating the Existing Model

In the multiple linear regression method, evaluating the efficiency of the model can be quite challenging and time consuming since several combinations of predictor variables can be used to build a model. To evaluate the influence of the same predictor variable on the response variable, we need to come up with the best combination of predictor variables and, hence, the best MLR model. To assess whether or not the existing MLR model is reasonable, several tests and procedures will be used. The preliminary plots are the easiest and fastest to start with during this evaluation stage.

4.2.1 Preliminary Plots

To conduct a snap evaluation of the preliminary model, plots are very useful. Instead of dealing with many complicated numbers, eyeballing was used to assess existing essential factors visually and identify problems within the plots in satisfying our first assumptions such as constant variance, normality, correlation and outliers.

4.2.1.1 Response Variables vs. Predictor Variables and Predictor Variables vs. Predictor Variables

A scatter plot between CO_2 mass emission rate (response variable) and 5 different predictor variables can easily convey the relationship between the response and each of the predictor variables. In Figure 4.1 below, which is the SAS output for the scatter plots of CO_2, “Vehicle Velocity”, “Vehicle Acceleration”, “Vehicle Deceleration”, “Power Demand” and “Time of Day”, the objective is to identify any curvilinearity, funnel shape and/or outliers and fix them. In addition, any linear trend within the scatter plots is assessed. An assessment is conducted for predictor variables

vs. predictor variable (Velocity, Acceleration, Deceleration, Power Demand, Time of Day) plots to identify potential multicollinearity.

In Figure 4.1 below, each linear trend or coverage within the scatter plots has been shown. Any noticeable trend detected between the predictor vs. predictor plots show a potential correlation between the variables and that is not good. This is treated later on in the model development process.

Referring to Figure 4.1:

- A correlation between "CO_2" and "Vehicle Acceleration" is noticeable. A slight linear trend shows a relationship between these two variables. This means that "Vehicle Acceleration" can noticeably influence "CO_2" emission rate and the positive slope of the trend conveys that, as "Vehicle Acceleration" increases, the mass emission rate of "CO_2" increases as well.
- Plots do not show a high correlation between "CO_2" with "Vehicle Velocity" and "Time of Day". This means that the two variables may not highly influence "CO_2", although this is just a basic evaluation. Further tests and assessments will be conducted in order to evaluate influences caused by the interaction terms as well.
- "CO_2" and "Power Demand" seem to be slightly correlated, which means this predictor variable is influencing the response variable.
- "Vehicle Acceleration" and "Power Demand" and "Vehicle Deceleration" and "Power Demand", on the other hand, look to be highly correlated and might lead to a multicollinearity problem. This is the inter-correlation of independent variables and

occurs when two or more predictors are highly correlated. In the remaining plots, the points seem to be fairly scattered and variables seem to be uncorrelated.

- In some predictor vs. predictor plots as well as response vs. predictor plots, some points appear to be potential x-outliers. Later some procedures will be conducted to evaluate whether they should be thrown out.

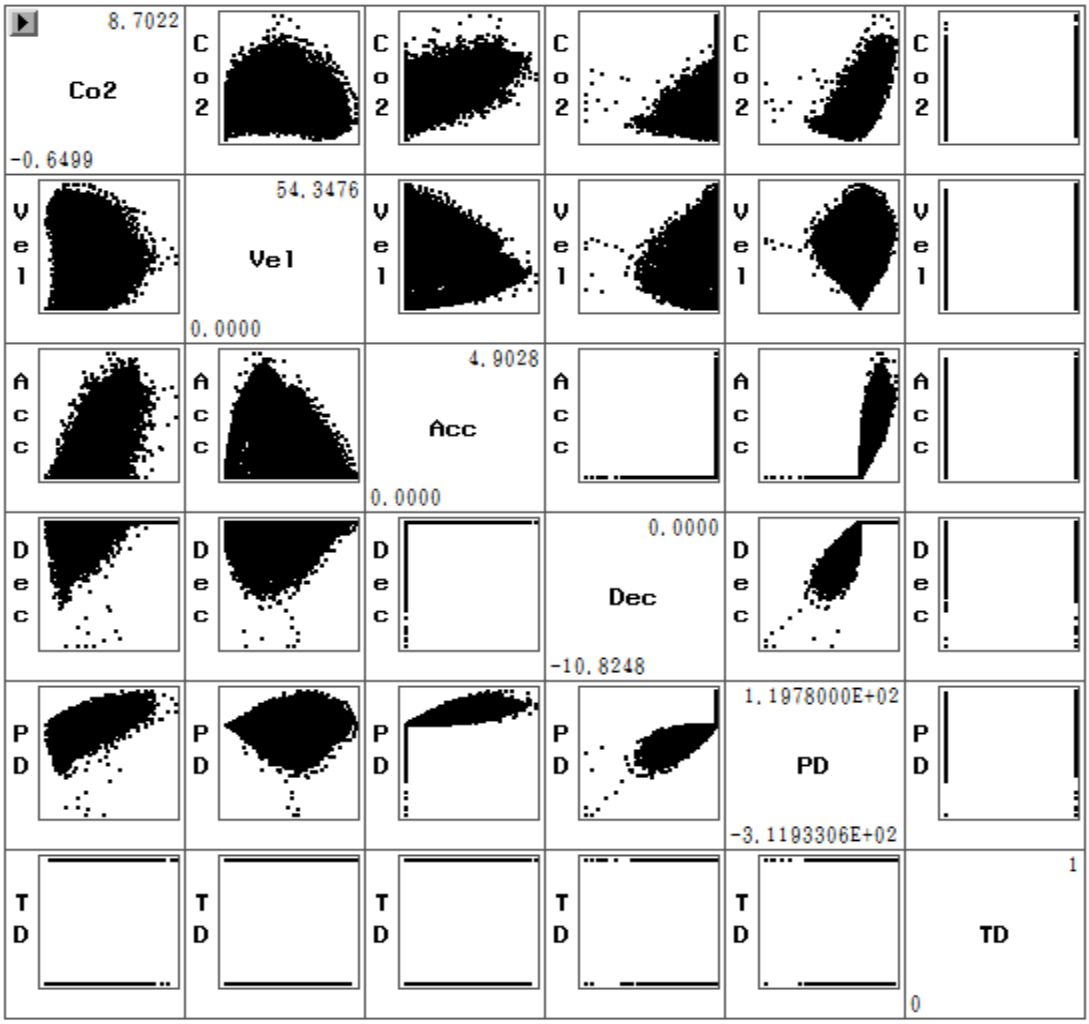


Figure 4.1 Scatter Plots of Response and Predictor Variables

In addition to the plots, another SAS output is used to evaluate the correlations between the variables. This output is a table that summarizes the Correlation Coefficient

(Correlation Matrix) between the variables (correlation between the response - predictor and also between the predictor- predictor variables). Correlation Coefficient (r) is a measure that indicates a linear relationship and varies between “-1” and “1”. The higher the value of the “Correlation Coefficient”, the stronger the linear relation between the pair of variables. Table 4.2 shows the correlation between the response variable and predictor variables.

Table 4.2 Correlation between Variables

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	Dec	PD	TD
Co2 CO2	1.00000	0.05348	0.76028	0.29817	0.61979	0.06577
Vel Velocity	0.05348	1.00000	-0.10829	0.16249	0.03015	0.03796
Acc Acceleration	0.76028	-0.10829	1.00000	0.26661	0.68986	-0.00019
Dec Deceleration	0.29817	0.16249	0.26661	1.00000	0.73260	0.00041
PD Power demand	0.61979	0.03015	0.68986	0.73260	1.00000	-0.00094
TD Time of Day	0.06577	0.03796	-0.00019	0.00041	-0.00094	1.00000

From Table 4.2, we have the following preliminary conclusions:

$\mathbf{r_{y1}}$ = 0.05 indicates that CO_2 and “Vehicle Velocity” are somewhat positively correlated.

$\mathbf{r_{y2}}$ = 0.76 indicates that CO_2 and “Vehicle Acceleration” are positively highly correlated.

$\mathbf{r_{y3}}$ = 0.30 indicates that CO_2 and “Vehicle Deceleration” are somewhat positively correlated.

$\mathbf{r_{y4}}$ = 0.62 indicates that CO_2 and “Power Demand” are positively highly correlated.

$\mathbf{r_{y5}}$ = 0.07 indicates that CO_2 and "Time of Day" are somewhat positively correlated.

$\mathbf{r_{12}}$ = -0.10 indicates that "Vehicle Velocity" and "Vehicle Acceleration" are somewhat negatively correlated.

$\mathbf{r_{13}}$ = 0.16 indicates that "Vehicle Velocity" and "Vehicle Deceleration" are somewhat positively correlated.

$\mathbf{r_{14}}$ = 0.03 indicates that there is little correlation between "Vehicle Velocity" and "Power Demand".

$\mathbf{r_{15}}$ = 0.04 indicates that there is little correlation between "Vehicle Velocity" and "Time of Day".

$\mathbf{r_{23}}$ = 0.26 indicates that "Vehicle Acceleration" and "Vehicle Deceleration" are somewhat positively correlated.

$\mathbf{r_{24}}$ = 0.69 indicates that "Vehicle Acceleration" and "Power Demand" are positively highly correlated.

$\mathbf{r_{25}}$ = -0.002 indicates that there is little correlation between "Vehicle Acceleration" and "Time of Day".

$\mathbf{r_{34}}$ = 0.73 indicates that there is little correlation between "Vehicle Deceleration" and "Power Demand" are positively highly correlated.

$\mathbf{r_{35}}$ = 0.0004 indicates that there is little correlation between "Vehicle Deceleration" and "Time of Day".

$\mathbf{r_{45}}$ = -0.0009 indicates that there is little correlation between "Power Demand" and "Time of Day".

From the numbers presented in Table 4.2, it can be deduced that a correlation exists between predictor variables in at least two cases, “Vehicle Acceleration” and “Power Demand” and “Vehicle Deceleration” and “Power Demand”. This may be a potential multicollinearity problem. Further tests will help in making a final decision on the multicolinearity issue.

4.2.2 Residuals Plots

Residual plots can reveal useful information about the performance of the existing model. We are looking to detect problems such as “curvilinearity”, as shown in Figure 4.3, and “funnel shape”, as shown in Figure 4.2, to assess “constant variance” and “outliers”.

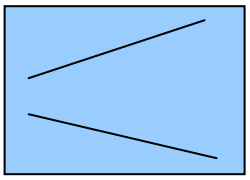


Figure 4.2 Funnel Shape

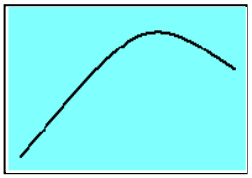


Figure 4.3 Curvilinearity

4.2.2.1 Residuals vs. $\hat{y}_i$

From Figure 4.4 below, we cannot detect any funnel shape. So far what can be concluded is that the “constant variance” assumption is probably OK. Further tests will be conducted to verify the conclusion from this plot.

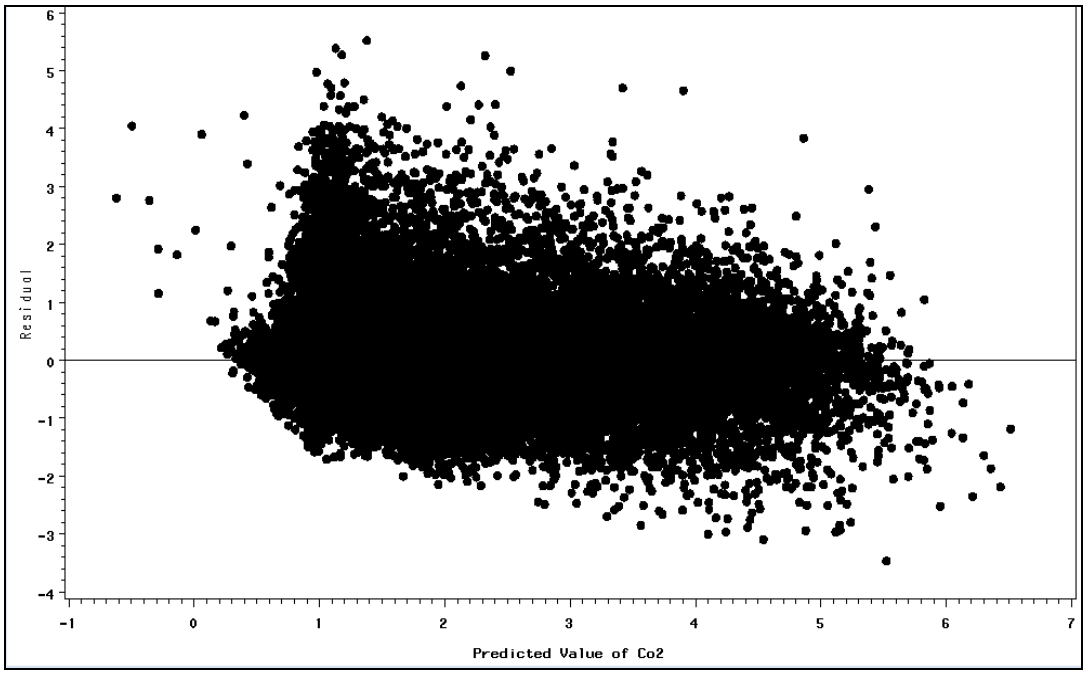


Figure 4.4 Residual vs. Predicted Value of CO_2

4.2.2.2 Residuals (e) vs. Predictor Variables

In this series of plots shown in Figure 4.5, curvature cannot be detected. Curvature would indicate violation of the assumed model form, which is currently a linear equation. There are also some possible outliers.

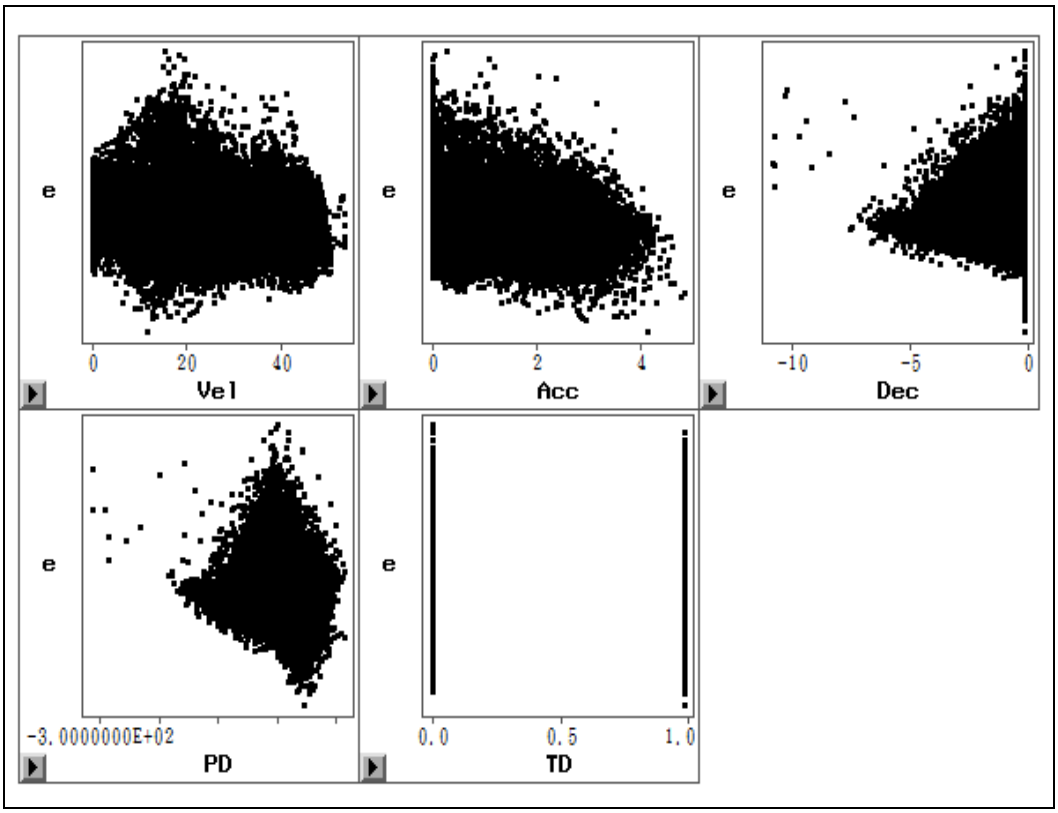


Figure 4.5 Residual vs. Predictor Variables

4.2.2.3 Residuals vs. Interaction Terms

Basically, what we are looking for in Figure 4.6 plots below is a trend between the residuals and any interaction terms. If a trend is seen, we can suggest adding the associated interaction terms to the model.

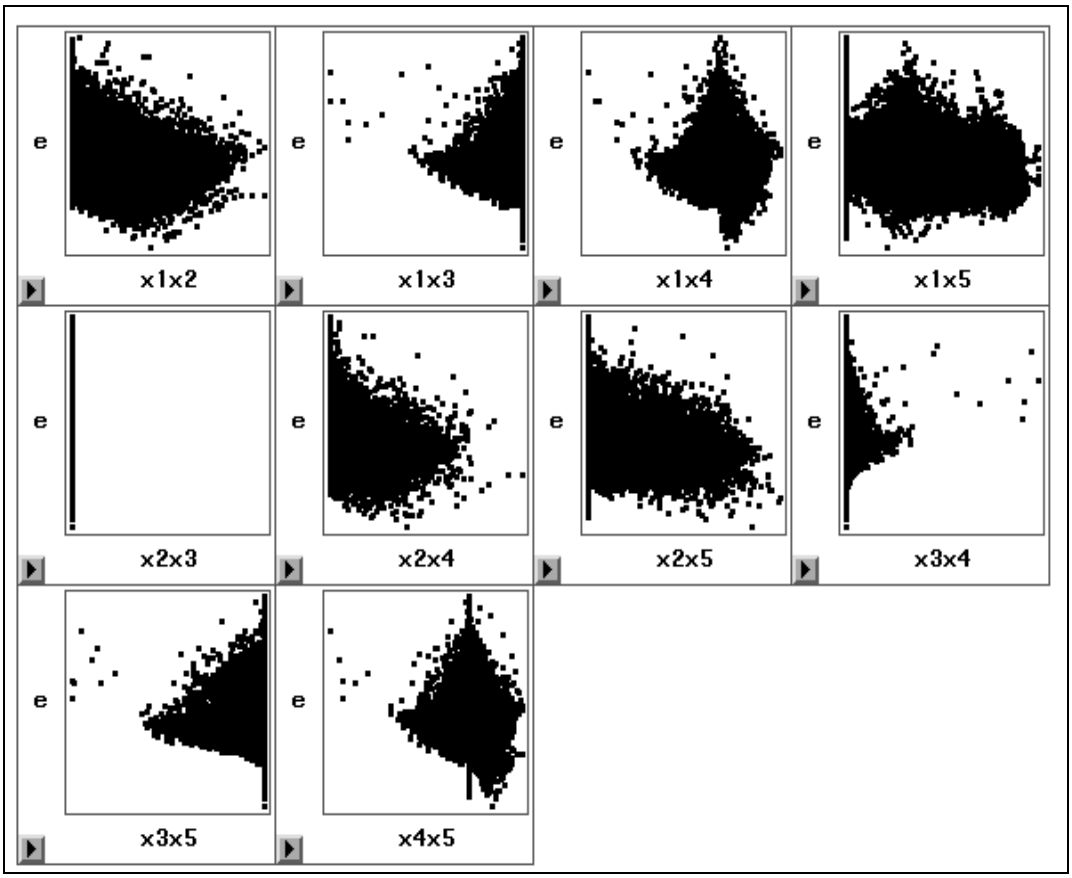


Figure 4.6 Residuals vs. Interaction Terms

As shown in Figure 4.6 above, there is not a strong trend among the plots. There might be a slight trend between residuals and x1x3, x1x2, x1x4 and x2x4, but at this point we are not suggesting that any interaction terms should be added.

4.2.3 Normal Probability Plot of e_i's

A normality plot delivers information about the distribution of residuals and allows one to study whether the residuals are normally distributed. The normality plot shown in Figure 4.7 looks to deviate slightly from a straight-line. There is slight indication of a shorter right tail and a longer left tail than the normal distribution.

However, on the strength of the straightness of the plot, normality may be OK. Further tests will be conducted and a final decision on normality will be made later.

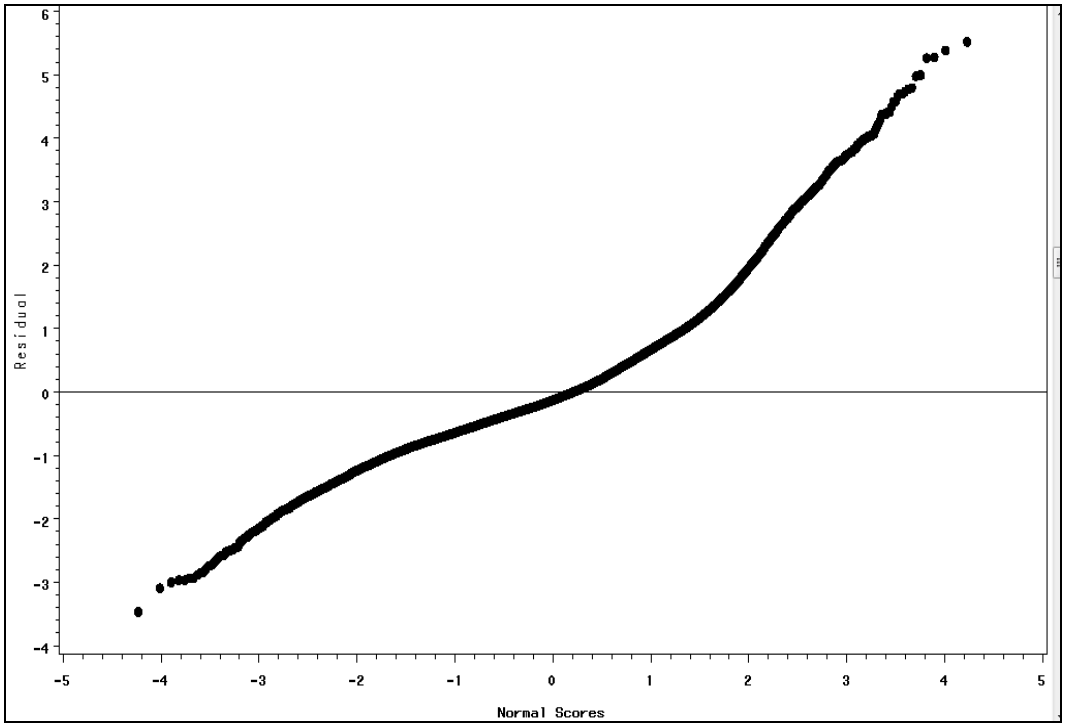


Figure 4.7 Normal Probability Plot

4.2.4 Time Series Plot

The time series plot in Figure 4.8 below shows a slight upward trend, which might indicate serial correlation between the observations.

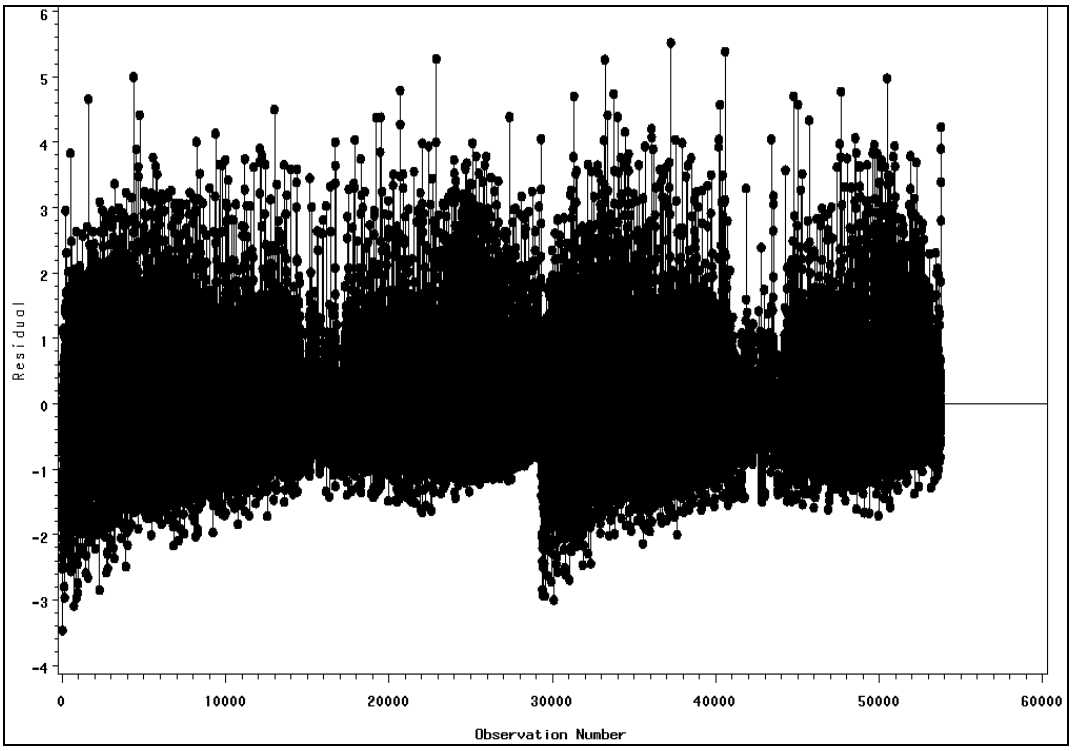


Figure 4.8 Time Series Plot

4.3 Verifying the Model Assumptions

In this section, tests and diagnostics are presented to further verify model assumptions that have been assessed visually from the plots.

4.3.1 Test for Constant Variance (Modified Levene Test)

A further test is conducted to confirm the initial assumption from the plots that the “constant variance” assumption is OK. Figure 4.9 and Table 4.3 are considered in this test.

Ho: Variances are equal

Test: vs.

H1: Variances are not equal

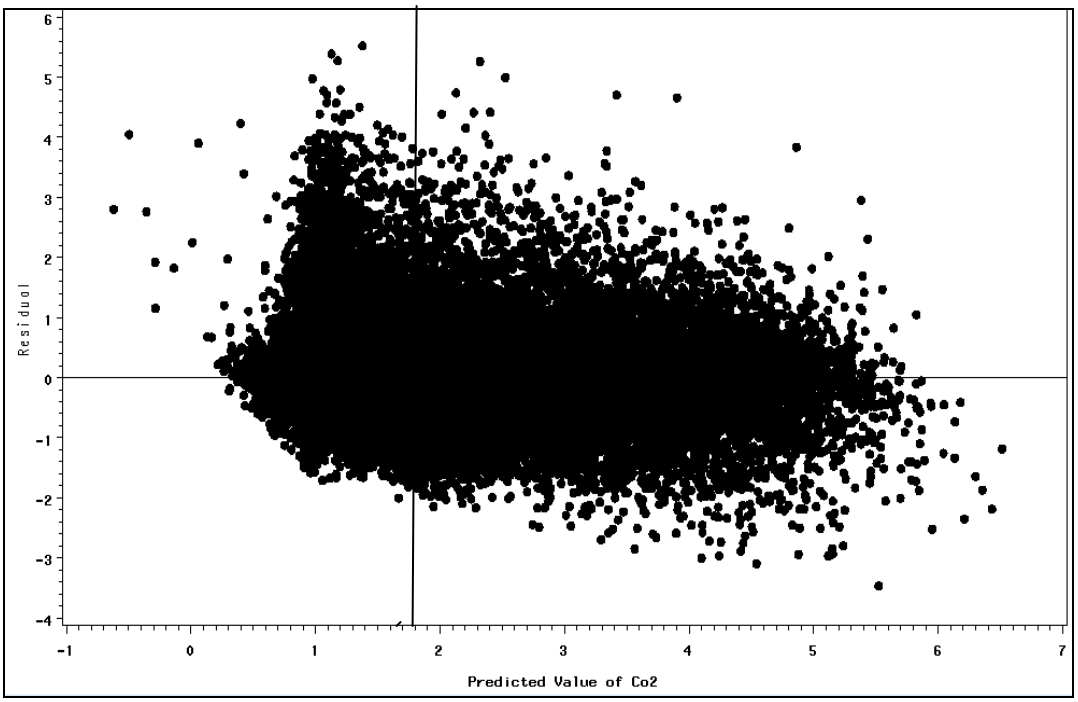


Figure 4.9 Residual vs. Predicted Value of CO_2

As discussed previously, we concluded from the residual plots that the “constant variance” assumption is OK. The Modified Levene Test will be used to confirm this conclusion. This test does not depend on the normality of the error terms. To conduct this test, we divide the data into two groups of about the same number of observations.

The cutoff line on $\hat{y}_i$ will fall around value “1.8”. This line is illustrated on the residual vs. $\hat{y}_i$ plot in Figure 4.9.

Table 4.3 SAS Output for Testing Constant Variance

The TTEST Procedure

Statistics

Variable	group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
d	1	26876	0.4633	0.47	0.4766	0.5512	0.5559	0.5606	0.0034
d	2	26919	0.6104	0.6167	0.623	0.5224	0.5269	0.5313	0.0032
d	Diff (1-2)		-0.156	-0.147	-0.138	0.5383	0.5416	0.5448	0.0047

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	54E3	1.43	<.5321
d	Satterthwaite	Unequal	54E3	1.42	<.5321

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
d	Folded F	26875	26918	1.11	<.3257

Table 4.3 shows the result of conducting the test using SAS and using 2 groups with about the same number of observations in each of them. Assuming α=0.1 for the confidence level, the p-value for the test turns out to be greater than α=0.1. Thus:

P-value=0.5321>0.1=$\alpha \Rightarrow$ we fail to reject H_0.

This leads us to conclude that the assumption of “constant variance” is OK.

4.3.2 Test for Normality

Further testing is conducted to support the initial assumption that normality generally may be OK, as follows:

H_0: Normality is OK

Test: Vs.

H_1: Normality is violated

The following table is obtained from the SAS output.

Table 4.4 Normality Test from SAS Output

Pearson Correlation Coefficients, N = 53795
Prob > |r| under H0: Rho=0

	e	enrm
e	1.00000	0.96674
Residual		<.0001
enrm	0.96674	1.00000
Normal Scores	<.0001	

We need to find $c(\alpha, n)$ in order to compare it with the correlation value $\hat{\rho}$ (from the SAS output) in Table 4.4 with $\alpha = 0.10$, we have $c(\alpha = 0.10, n = 53795) = 0.989$; this implies $\hat{\rho} = 0.967 < 0.989 \Rightarrow$ reject Ho. Thus, the Normality is not OK from the test. However, the correlation value $\hat{\rho}$ is close to the $c(\alpha, n)$ value. Further, normality is a desired assumption and not a required one. Hence, we consider normality from the plot to be somewhat OK.

4.3.3 Outlier Diagnostic

Y-Outlier: Y-Outliers can be detected by using the Bonferroni test.

If the studentized deleted residuals (Rstudent-APPENDIX A) are less than the Bonferroni test cutoff, then there are no y-outliers.

Bonferroni test cutoff: t(1-α/2n, n-p-1) = t(1-0.10/(2*53795)); 53795-6-1) = t(0.99999, 53788) = 3.291. When compared with the Rstudent values in APPENDIX A, the values are less than 3.291; we then conclude that there are no y – outliers.

X-Outlier: Leverage values (h_{ii}) are utilized to check whether x-outliers exist or not. The h_{ii} values (APPENDIX A) are compared with the value $\frac{2p}{n}$ = 2(6)/53795 = 0.0002. If any h_{ii} value from APPENDIX A>0.0002, then it can be concluded that x-outliers exist. We can detect a few observations greater than 0.0002, but we do not have any reason for deleting them; hence, we leave them in the dataset and continue with the analysis.

Influence on the Fitted Values: DFFITS

The DFFITS values for any outliers (APPENDIX A) are compared with the value $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{6}{53795}}$ = 0.021. If the DFFITS value for an outlier > 0.021, then that outlier has high influence on the estimated regression equation and fitted values. From APPENDIX A, there are a couple of DFFITS values higher than $2\sqrt{\frac{p}{n}}$ = 0.021.

Cut-off DFBETA (influence on the individual LSEs):

The DFBETA values (APPENDIX A) are compared with the value $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{53795}}$ = 0.009. If the DFBETA value for an outlier >0.009, then that outlier has high influence on the least square estimates. From the value of DFBETA in APPENDIX A, a couple of observations have higher values than 0.009.

4.3.4 Multicollinearity - Variance Inflation

From the assessment of the predictor variables using the Correlation Matrix, a potential multicollinearity problem was discovered. To further assess the

multicollinearity problem, tolerance $(1-R^2)^{-1}$ or the variance inflation factor (VIF), which is simply the reciprocal of tolerance, is used. If the VIF is high, there is a high multicollinearity problem and therefore the correlation between predictor variables is high. One guideline is to avoid models with any VIF greater than 5.

Table 4.1, repeated here for convenience, shows only one number slightly greater than "5" under the "Variance Inflation" column. Therefore, up to this point, multicollinearity may not be a problem.

Table 4.1 SAS Output for Assessing Variance Inflation

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.90912	0.00861	105.55	<.0001	0
Vel	Velocity	1	0.00960	0.00021723	44.19	<.0001	1.05587
Acc	Acceleration	1	0.95488	0.00628	152.11	<.0001	2.51470
Dec	Deceleration	1	-0.05074	0.00571	-8.88	<.0001	2.87665
PD	Power demand	1	0.00759	0.00022749	33.35	<.0001	5.00135
TD	Time of Day	1	0.15151	0.00661	22.93	<.0001	1.00151

4.3.5 Partial Regression Plots

Partial regression plots can evaluate the effect of adding an additional variable to the model. Partial regression plots are formed by:

(1) Calculating the residuals of regressing the response variable against the independent variables but omitting Xi,

(2) Calculating the residuals from regressing Xi against the remaining independent variables,

(3) Plotting the residuals from (1) against the residuals from (2).

The resulting plots are shown in Figure 4.10-4.19 Before considering the addition of any interaction terms to our model, the correlation between the interaction

terms vs. predictor variable in Tables 4.5 and 4.6 are examined. The results show many high correlation values (>0.7). To reduce correlation between interaction terms and current predictor variables, standardized variables will be utilized. The correlations for the standardized variables are shown in Table 4.7 and 4.8. From these tables, the correlations between standardized interaction terms and current predictors are less or equal to 0.6 except stdx2x4.

Figure 4.19 shows scatter plots of residual vs. standardized interaction terms. Trends are difficult to see on both Figures 4.10-4.19 and 4.20. However, stdx1x2, stdx1x3, stdx1x4, stdx2x3 and stdx2x4 seem to show some trends and would be added to the model.

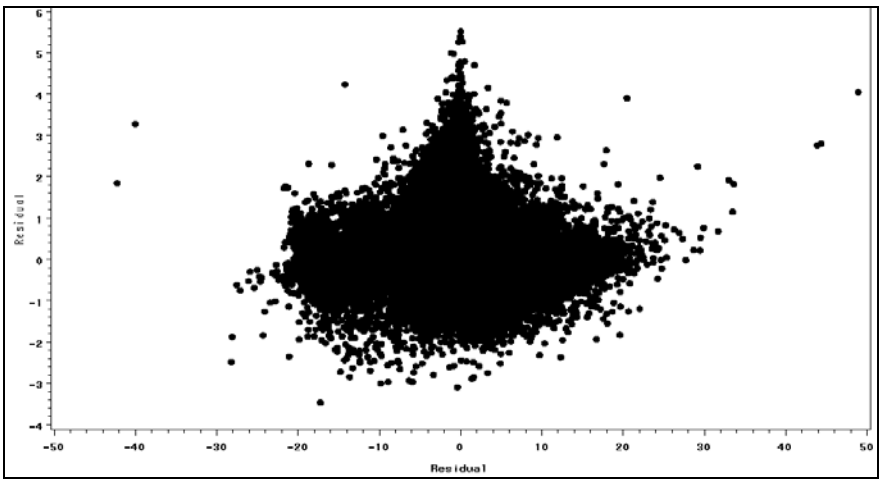


Figure 4.10 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x2/x1x2x3x4x5)

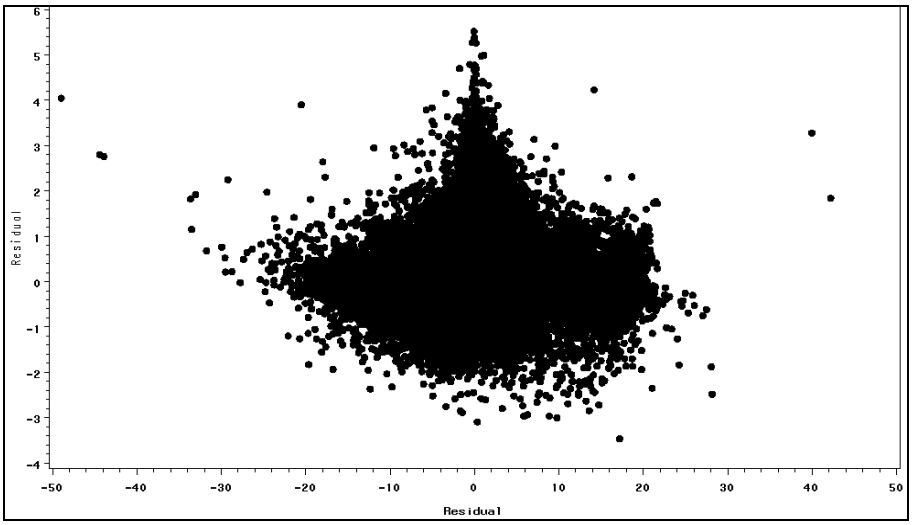


Figure 4.11 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x3/x1x2x3x4x5)

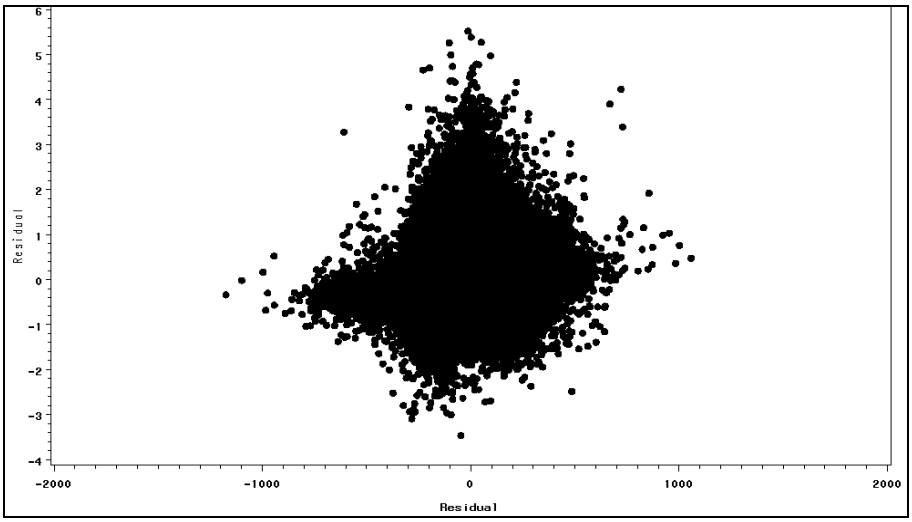


Figure 4.12 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x4/x1x2x3x4x5)

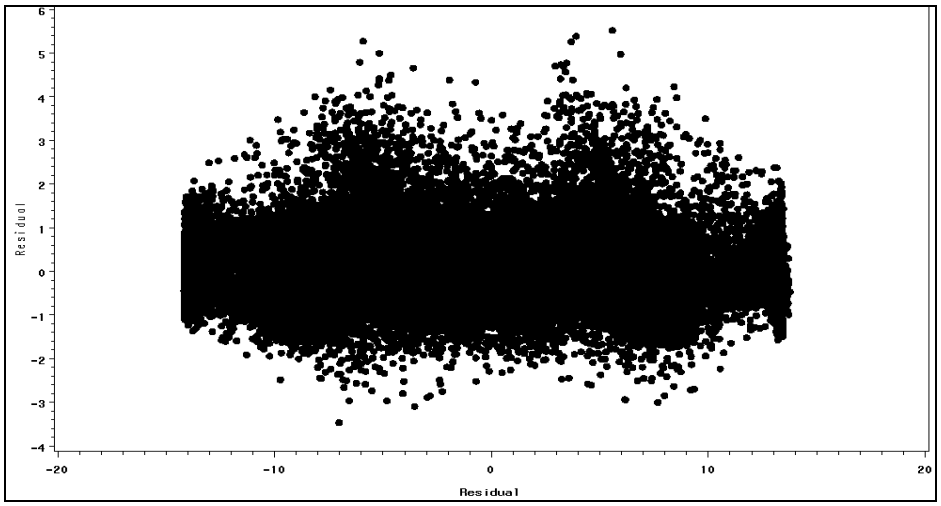


Figure 4.13 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x5/x1x2x3x4x5)

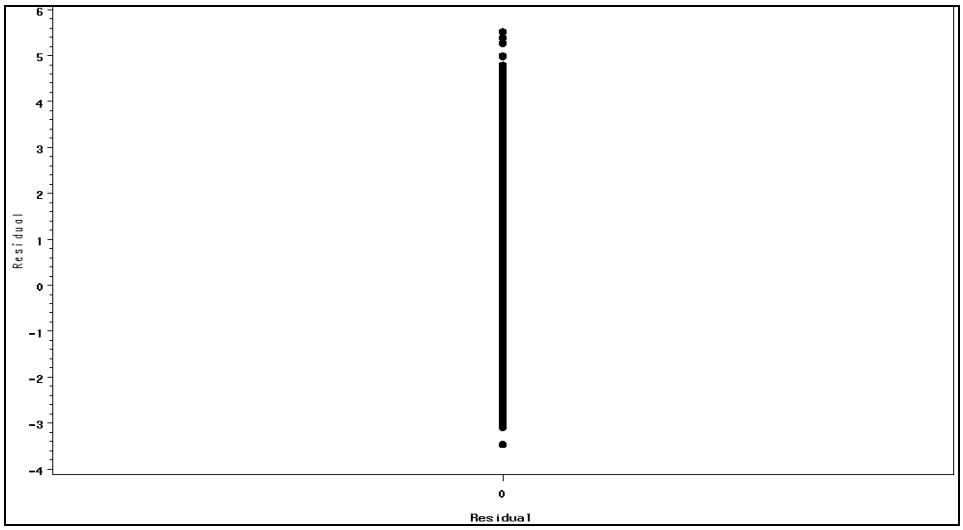


Figure 4.14 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x3/x1x2x3x4x5)

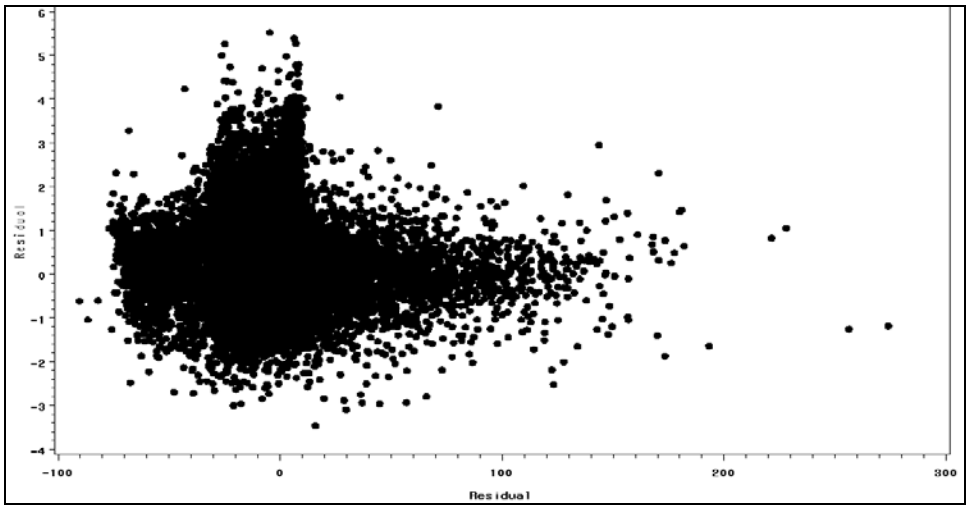


Figure 4.15 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x4/x1x2x3x4x5)

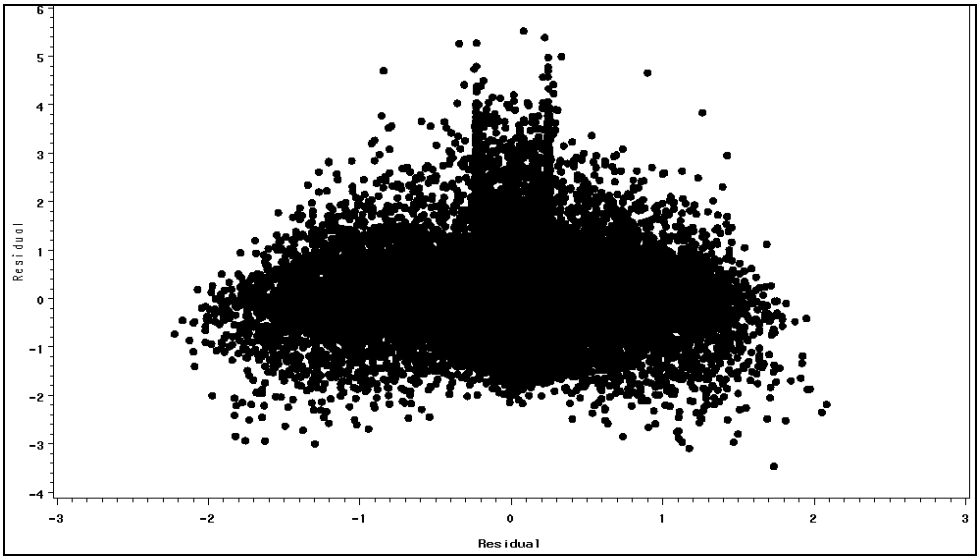


Figure 4.16 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x5/x1x2x3x4x5)

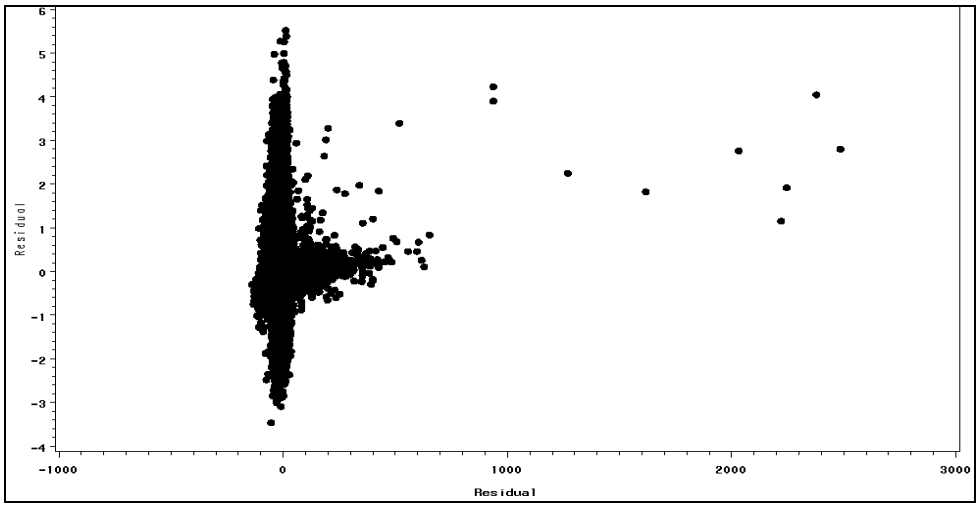


Figure 4.17 Residual e(Y/x1x2x3x4x5) vs. Residual e(x3x4/x1x2x3x4x5)

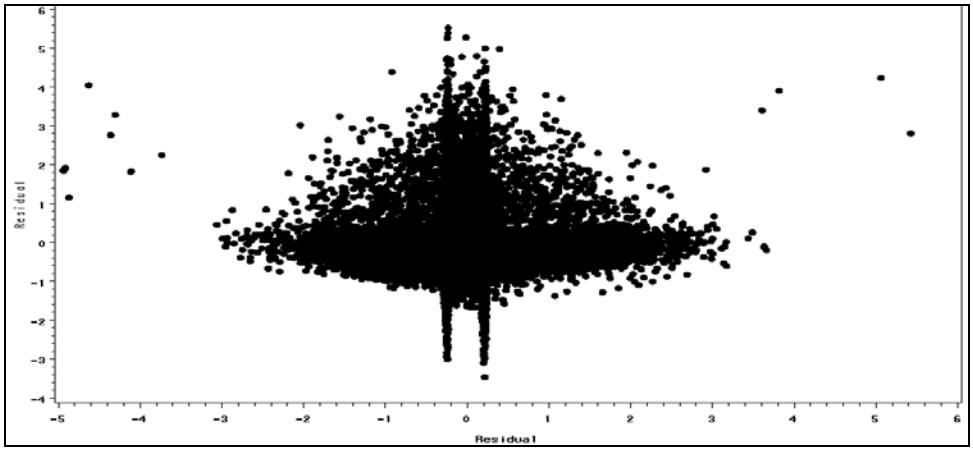


Figure 4.18 Residual e(Y/x1x2x3x4x5) vs. Residual e(x3x5/x1x2x3x4x5)

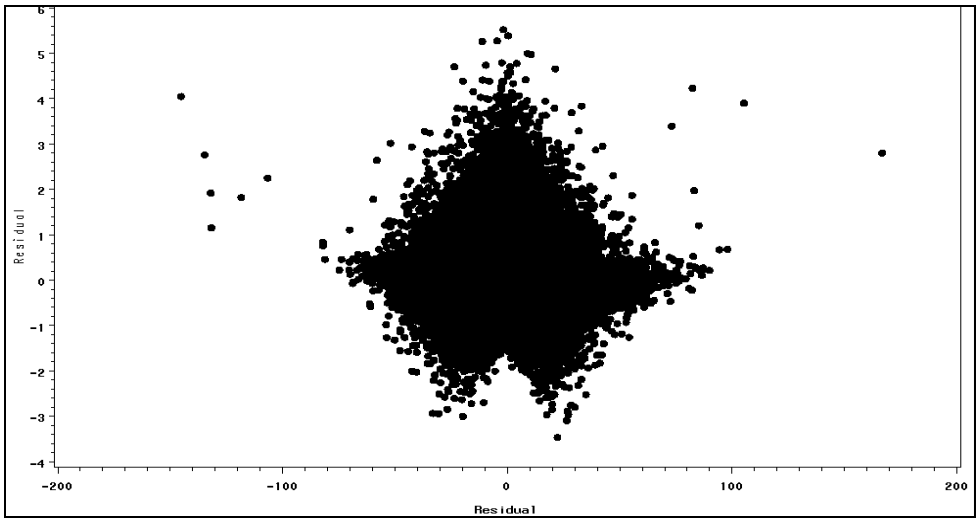


Figure 4.19 Residual e(Y/x1x2x3x4x5) vs. Residual e(x4x5/x1x2x3x4x5)

Table 4.5 Correlation between Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	Dec	PD	TD	x1x2	x1x3
Co2 CO2	1.00000	0.05348	0.76028	0.29817	0.61979	0.06577	0.70911	0.31265
Vel Velocity	0.05348	1.00000	-0.10829	0.16249	0.03015	0.03796	0.09749	-0.04098
Acc Acceleration	0.76028	-0.10829	1.00000	0.26661	0.68986	-0.00019	0.87243	0.27382
Dec Deceleration	0.29817	0.16249	0.26661	1.00000	0.73260	0.00041	0.28922	0.85921
PD Power demand	0.61979	0.03015	0.68986	0.73260	1.00000	-0.00094	0.77970	0.82949
TD Time of Day	0.06577	0.03796	-0.00019	0.00041	-0.00094	1.00000	0.00284	-0.00396
x1x2	0.70911	0.09749	0.87243	0.28922	0.77970	0.00284	1.00000	0.29704

Table 4.6 Correlation between Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	Dec	PD	TD	x1x2	x1x3
x1x3	0.31265	-0.04098	0.27382	0.85921	0.82949	-0.00396	0.29704	1.00000
x1x4	0.55916	0.03407	0.58512	0.61617	0.95825	-0.00104	0.75122	0.79123
x1x5	0.06416	0.47366	-0.05875	0.08223	0.01271	0.78396	0.03857	-0.01501
x2x3	.	.	.	.	.	.	.	.
x2x4	0.68554	-0.05198	0.90181	0.19280	0.65737	-0.00636	0.90189	0.19801
x2x5	0.53311	-0.07261	0.68026	0.18577	0.47377	0.36066	0.59608	0.19079
x3x4	-0.19722	-0.05588	-0.16292	-0.82397	-0.66929	-0.00649	-0.17674	-0.86300
x3x5	0.19222	0.11216	0.19058	0.68988	0.51357	-0.30728	0.20674	0.59877
x4x5	0.45183	0.02220	0.50569	0.53341	0.73260	0.02835	0.57261	0.60643

Table 4.7 Correlation between Standardized Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	Dec	PD	TD	stdx1x2	stdx1x3
Co2 CO2	1.00000	0.05348	0.76028	0.29817	0.61979	0.06577	-0.41082	-0.07996
Vel Velocity	0.05348	1.00000	-0.10829	0.16249	0.03015	0.03796	-0.27560	0.14444
Acc Acceleration	0.76028	-0.10829	1.00000	0.26661	0.68986	-0.00019	-0.52483	-0.16612
Dec Deceleration	0.29817	0.16249	0.26661	1.00000	0.73260	0.00041	-0.17040	-0.58431
PD Power demand	0.61979	0.03015	0.68986	0.73260	1.00000	-0.00094	-0.12277	-0.15241
TD Time of Day	0.06577	0.03796	-0.00019	0.00041	-0.00094	1.00000	-0.02026	0.01348
stdx1x2	-0.41082	-0.27560	-0.52483	-0.17040	-0.12277	-0.02026	1.00000	0.14516

Table 4.8 Correlation between Standardized Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	Dec	PD	TD	stdx1x2	stdx1x3
stdx1x3	-0.07996	0.14444	-0.16612	-0.58431	-0.15241	0.01348	0.14516	1.00000
stdx1x4	-0.06355	-0.05641	-0.18723	-0.23839	0.09667	-0.00351	0.58093	0.57944
stdx1x5	-0.02697	-0.06625	-0.01692	0.01155	-0.00192	-0.00684	0.04143	-0.03236
stdx2x3	0.29842	-0.22606	0.50497	-0.69726	-0.14300	-0.00051	-0.23774	0.39973
stdx2x4	0.57922	-0.07359	0.81271	-0.06461	0.39246	-0.00753	-0.31196	-0.06487
stdx2x5	-0.03217	-0.01697	-0.02611	-0.00032	-0.01053	0.00003	0.03121	-0.00513
stdx3x4	-0.09431	-0.05899	-0.04068	-0.78213	-0.54811	-0.00743	0.06065	0.28286
stdx3x5	-0.00131	0.01157	-0.00032	-0.03871	-0.01648	-0.00007	-0.00524	0.04399
stdx4x5	-0.01594	-0.00193	-0.01053	-0.01651	-0.01588	0.00017	0.00538	0.00893

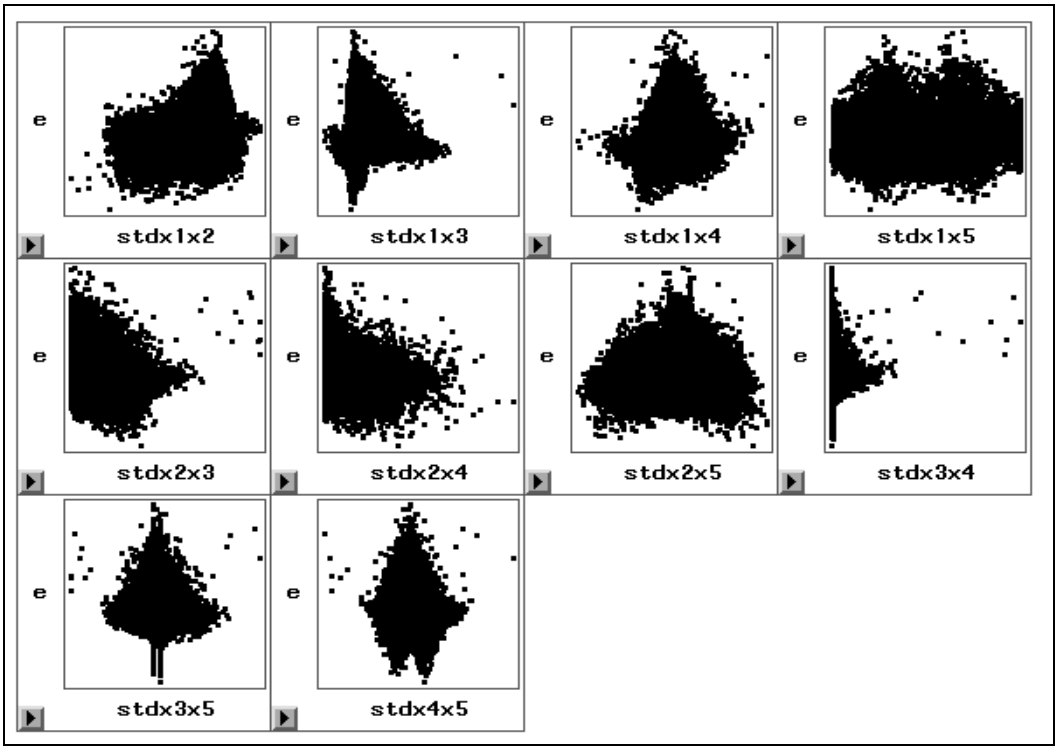


Figure 4.20 Scatter Plots of Residual vs. Standardized Interaction Terms

4.4 Model Selection (Best Potential Model)

In multiple linear regression analysis, the most important part is selecting the best model. This is due to the need to explain the data in the simplest way and, hence, remove unnecessary predictor variables. Time will be saved in measuring unnecessary predictors. Three methods to consider in "best model" selection are best subset selection, stepwise regression, and backward deletion. Each of these methods is discussed in turn.

4.4.1 Best Subset Selection

The best subset selection is a procedure that uses the branch and bound algorithm (Chen, Spring 2006) to find a specified number of best models containing one, two, three variables and so on, up to the single model containing all of the explanatory variables. The criteria used for the best model determination are: a) High R^2, the coefficient of multiple determination, the percent of the variance in the dependent variable that can be explained by all of the independent variables taken together; b) High R_a^2 , the adjusted coefficient of multiple determination, which is a version of R^2 that has been adjusted for the number of predictors in the model (R-Squared tends to over-estimate the strength of the association, especially if the model has more than one independent variable); c) low C_p or close to p (number or parameters), C_p criterion, estimates the normalized expected value of the squared difference between a fitted regression model and a true model. The SAS program was used to run the best subset procedure to obtain the output shown in Table 4.9 below.

Table 4.9 SAS Output for the Best Subset Method

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
1	0.5780	0.5780	5809.800	Acc
1	0.3841	0.3841	33194.30	PD

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
2	0.5967	0.5967	3175.638	Vel Acc
2	0.5953	0.5954	3363.954	Acc PD

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
3	0.6094	0.6094	1375.794	Vel Acc PD
3	0.6050	0.6050	2002.254	Vel Acc stdx1x2

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
4	0.6134	0.6135	808.6403	Vel Dec stdx1x4 stdx2x3
4	0.6134	0.6135	808.6403	Vel Acc Dec stdx1x4

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
5	0.6172	0.6173	276.4572	Vel Dec TD stdx1x4 stdx2x3
5	0.6172	0.6173	276.4572	Vel Acc Dec TD stdx1x4

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
6	0.6190	0.6190	15.1144	Vel Acc PD TD stdx1x2 stdx1x4
6	0.6186	0.6186	89.0613	Vel Dec TD stdx1x3 stdx1x4 stdx2x3

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
7	0.6191	0.6191	14.8893	Vel PD TD stdx1x2 stdx1x3 stdx1x4 stdx2x3
7	0.6191	0.6191	14.8893	Vel Dec PD TD stdx1x2 stdx1x3 stdx1x4

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
8	0.6191	0.6192	9.0000	Vel PD TD stdx1x2 stdx1x3 stdx1x4 stdx2x3 stdx2x4
8	0.6191	0.6192	9.0000	Vel Dec PD TD stdx1x2 stdx1x3 stdx1x4 stdx2x4

In conclusion, using the idea of R^2 leveling off and C_p close to P (the number of parameters), the following two best models were selected.

- Vel , Acc, PD, TD, stdx1x2 and stdx1x4 (1st choice)
- Vel , Dec, PD, TD, stdx1x2, stdx1x3, stdx1x4 and stdx2x4 (2nd choice)

4.4.2 Stepwise Regression

Stepwise regression is a procedure that combines the backward elimination and forward selection methods. This method allows the addition and removal of variables at anytime in the process and finally selects the “best” model. The result from the SAS program is shown in Table 4.10 below.

Table 4.10 SAS Output for Stepwise Regression Method

Variable Acc Entered: R-Square = 0.5780 and C(p) = 5809.800

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	46827	46827	73685.1	<.0001
Error	53793	34185	0.63550		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.21183	0.00395	59802	94101.8	<.0001
Acc	1.12319	0.00414	46827	73685.1	<.0001

Variable Vel Entered: R-Square = 0.5967 and C(p) = 3175.638

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	48339	24169	39791.5	<.0001
Error	53792	32673	0.60740		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.90154	0.00732	9211.45395	15165.3	<.0001
Vel	0.01084	0.00021731	1512.03033	2489.34	<.0001
Acc	1.14518	0.00407	48107	79201.4	<.0001

Table 4.10 - continued

Variable PD Entered: R-Square = 0.6094 and C(p) = 1375.794

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	49372	16457	27979.3	<.0001
Error	53791	31640	0.58820		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.00672	0.00763	10243	17415.0	<.0001
Vel	0.00952	0.00021615	1141.50120	1940.67	<.0001
Acc	0.98178	0.00559	18154	30864.2	<.0001
PD	0.00599	0.00014282	1033.48885	1757.04	<.0001

Variable TD Entered: R-Square = 0.6132 and C(p) = 843.4201

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	49679	12420	21320.9	<.0001
Error	53790	31333	0.58251		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.92974	0.00830	7308.90263	12547.2	<.0001
Vel	0.00933	0.00021527	1094.45197	1878.85	<.0001
Acc	0.98082	0.00556	18118	31102.8	<.0001
PD	0.00601	0.00014213	1041.05171	1787.17	<.0001
TD	0.15170	0.00661	306.50213	526.17	<.0001

Variable stdx1x4 Entered: R-Square = 0.6164 and C(p) = 391.4505

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	49939	9987.83595	17289.5	<.0001
Error	53789	31073	0.57768		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.88830	0.00849	6319.49195	10939.4	<.0001
Vel	0.00994	0.00021629	1220.43168	2112.63	<.0001
Acc	1.02759	0.00596	17170	29722.5	<.0001
PD	0.00495	0.00015014	626.71436	1084.88	<.0001
TD	0.15145	0.00659	305.52192	528.87	<.0001
stdx1x4	0.13692	0.00645	260.38453	450.74	<.0001

Table 4.10 - continued

Variable stdx1x2 Entered: R-Square = 0.6190 and C(p) = 15.1144

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	50148	8358.02550	14565.9	<.0001
Error	53788	30864	0.57381		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.98168	0.00978	5784.92131	10081.6	<.0001
Vel	0.00757	0.00024891	530.24009	924.07	<.0001
Acc	0.93242	0.00776	8292.82016	14452.2	<.0001
PD	0.00613	0.00016208	821.79794	1432.18	<.0001
TD	0.15024	0.00656	300.59955	523.87	<.0001
stdx1x2	-0.12797	0.00671	208.97328	364.19	<.0001
stdx1x4	0.21286	0.00756	454.94451	792.85	<.0001

Variable stdx1x3 Entered: R-Square = 0.6191 and C(p) = 14.8893

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	50157	7165.35133	12490.9	<.0001
Error	53787	30855	0.57365		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.97920	0.00979	5733.12935	9994.19	<.0001
Vel	0.00771	0.00025144	539.47690	940.44	<.0001
Acc	0.93361	0.00776	8301.95062	14472.3	<.0001
PD	0.00596	0.00016759	725.91941	1265.45	<.0001
TD	0.15043	0.00656	301.36849	525.36	<.0001
stdx1x2	-0.13429	0.00689	218.17739	380.33	<.0001
stdx1x3	-0.02107	0.00523	9.30627	16.22	<.0001
stdx1x4	0.23910	0.00998	329.38780	574.20	<.0001

Table 4.10 - continued

Variable stdx2x4 Entered: R-Square = 0.6192 and C(p) = 9.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	50162	6270.24804	10931.9	<.0001
Error	53786	30850	0.57357		
Corrected Total	53794	81012			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.99495	0.01129	4458.37748	7773.00	<.0001
Vel	0.00736	0.00028143	391.85250	683.18	<.0001
Acc	0.89168	0.01683	1610.61366	2808.04	<.0001
PD	0.00633	0.00021291	507.11864	884.14	<.0001
TD	0.15077	0.00656	302.62132	527.61	<.0001
stdx1x2	-0.15185	0.00930	152.88660	266.55	<.0001
stdx1x3	-0.02544	0.00546	12.45970	21.72	<.0001
stdx1x4	0.25239	0.01104	299.62612	522.39	<.0001
stdx2x4	0.01671	0.00595	4.52507	7.89	0.0050

All variables left in the model are significant at the 0.05 level. No other variable met the 0.05 significance level for entry into the model.

- Vel , Acc, PD, TD and stdx1x4, stdx2x4 (1st choice)
- Vel, Acc, PD, TD, stdx1x2, stdx1x3 and stdx1x4 (2nd choice)

4.4.3 Backward Deletion

This is the simplest method among the three methods of variable selection. This method starts with all the predictor variables in the model and removes those with highest p-value greater than alpha critical (selected as 0.0500), one after the other. This is because they are statistically insignificant. We ended up with the SAS program selecting the model in Table 4.11 below.

Table 4.11 SAS Output for Backward Deletion Method

```
            All Variables Entered: R-Square = 0.6192 and C(p) = 9.0000
The model is not of full rank. A subset of the model which is of full rank is chosen.


                                Analysis of Variance

                                       Sum of           Mean
    Source                   DF       Squares         Square    F Value    Pr > F

    Model                     8         50162     6270.24804    10931.9    <.0001
    Error                 53786         30850        0.57357
    Corrected Total       53794         81012


                      Parameter     Standard
          Variable     Estimate        Error   Type II SS  F Value  Pr > F

          Intercept     0.99443      0.01128   4461.12820  7777.79  <.0001
          Vel           0.00737   0.00028143    393.82599   686.62  <.0001
          Acc           0.93800      0.01741   1664.94671  2902.77  <.0001
          Dec           0.04632      0.00994     12.45970    21.72  <.0001
          PD            0.00466   0.00042068     70.26690   122.51  <.0001
          TD            0.15077      0.00656    302.62132   527.61  <.0001
          stdx1x2      -0.13021      0.00881    125.34780   218.54  <.0001
          stdx1x4       0.25239      0.01104    299.62612   522.39  <.0001
          stdx2x4       0.01671      0.00595      4.52507     7.89  0.0050
```

All variables left in the model are significant at the 0.0500 level.

In conclusion, the model with p-value<0.05 was selected.

- Vel, Acc, Dec, PD, TD, stdx1x2 and stdx1x4, stdx2x4 (1st choice), if we consider significance level at 0.1.

4.5 Potential "Best" Models

Summarizing the results obtained from the three methods of model search, we conclude with the following best models.

Model A: Vel, Acc, PD, TD, stdx1x2 and stdx1x4 (1st choice)

Model B: Vel, Acc, PD, TD, stdx1x2, stdx1x3 and stdx1x4 (2nd choice)

4.5.1 Building the Regression Function

MODEL A

The regression function for Model A can be built from Table 4.12 below.

Table 4.12 SAS Output for Parameter Estimate for Model A

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.98168	0.00978	100.41	<.0001	0
Vel	Velocity	1	0.00757	0.00024891	30.40	<.0001	1.40531
Acc	Acceleration	1	0.93242	0.00776	120.22	<.0001	3.89143
PD	Power demand	1	0.00613	0.00016208	37.84	<.0001	2.57359
TD	Time of Day	1	0.15024	0.00656	22.89	<.0001	1.00160
stdx1x2		1	-0.12797	0.00671	-19.08	<.0001	2.97008
stdx1x4		1	0.21286	0.00756	28.16	<.0001	1.62310

The regression function is:

$$\hat{y} = 0.982 + 0.008x_1 + 0.923x_2 + 0.006x_4 + 0.150x_5 - 0.128x_1x_2 + 0.213x_1x_4 \qquad \text{Eqn. 4.3}$$

Where

$\hat{y} = CO_2$, (gram/sec)

x_1 = Vehicle Velocity, V,(mi/hr)

x_2 = Vehicle Acceleration, V,(mi/hr/sec)

x_4 = Power Demand, PD, ($mi^2/hr^2/sec$)

x_5 = Time of Day, TD,

From the correlation matrix in Table 4.13 below and the scatter plot matrix in Figure 4.21 below, a strong correlation does not exist between the predictor variables.

Table 4.13 Correlation between Predictor Variables for Model A

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	PD	TD	stdx1x2	stdx1x4
Co2 CO2	1.00000	0.05348	0.76028	0.61979	0.06577	-0.41082	-0.06355
Vel Velocity	0.05348	1.00000	-0.10829	0.03015	0.03796	-0.27560	-0.05641
Acc Acceleration	0.76028	-0.10829	1.00000	0.68986	-0.00019	-0.52483	-0.18723
PD Power demand	0.61979	0.03015	0.68986	1.00000	-0.00094	-0.12277	0.09667
TD Time of Day	0.06577	0.03796	-0.00019	-0.00094	1.00000	-0.02026	-0.00351
stdx1x2	-0.41082	-0.27560	-0.52483	-0.12277	-0.02026	1.00000	0.58093
stdx1x4	-0.06355	-0.05641	-0.18723	0.09667	-0.00351	0.58093	1.00000

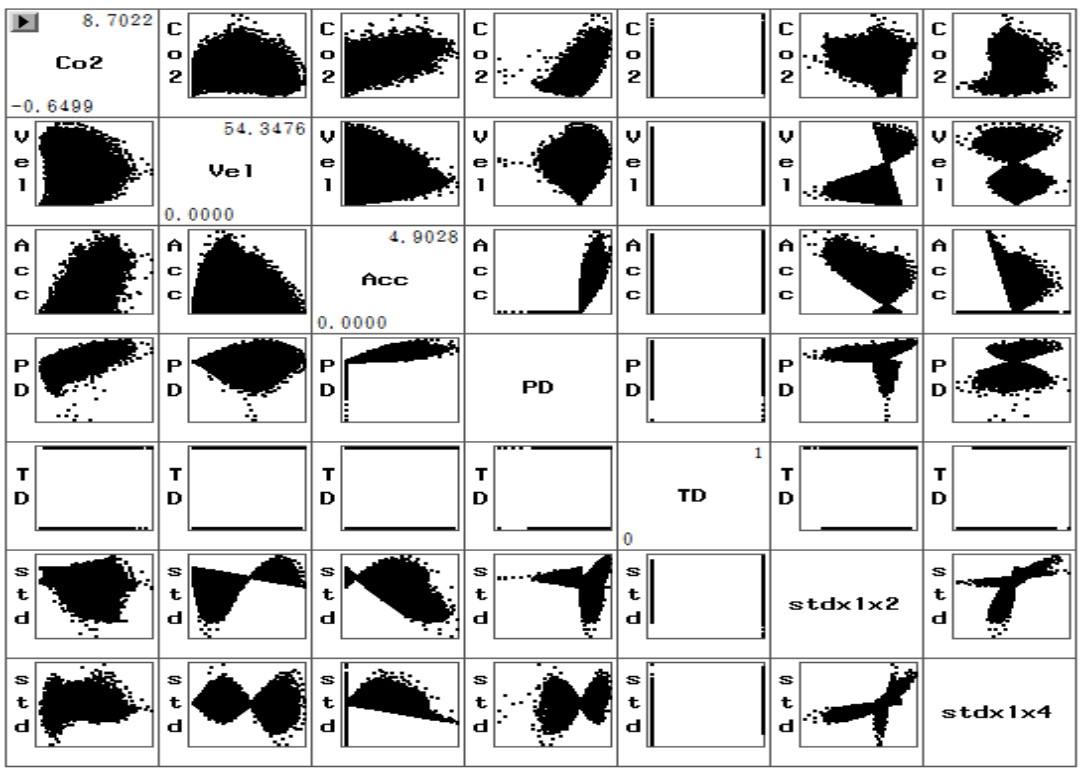


Figure 4.21 Scatter Plot Matrix of Response and Predictor Variables

This model compared to the preliminary model shows an improvement in terms of linear relationship between response variable and predictor variables and less multicolinearity between predictor variables.

MODEL B

The regression function for Model B can be built from Table 4.14 below.

Table 4.14 SAS Output for Parameter Estimate for Model B

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97920	0.00979	99.97	<.0001	0
Vel	Velocity	1	0.00771	0.00025144	30.67	<.0001	1.43447
Acc	Acceleration	1	0.93361	0.00776	120.30	<.0001	3.89708
PD	Power demand	1	0.00596	0.00016759	35.57	<.0001	2.75237
TD	Time of Day	1	0.15043	0.00656	22.92	<.0001	1.00166
stdx1x2		1	-0.13429	0.00689	-19.50	<.0001	3.13234
stdx1x3		1	-0.02107	0.00523	-4.03	<.0001	1.90268
stdx1x4		1	0.23910	0.00998	23.96	<.0001	2.82846

The regression function is:

$$\hat{y} = 0.979 + 0.008x_1 + 0.934x_2 + 0.006x_4 + 0.150x_5 - 0.134x_1x_2 - 0.021x_1x_3 + 0.239x_1x_4$$

Eqn. 4.4

Where

$\hat{y} = CO_2$ (grams/sec)

x_1 = Vehicle Velocity, Vel,(mi/hr)

x_2 = Vehicle Acceleration, Acc,(mi/hr/sec)

x_3 = Vehicle Deceleration, Acc,(mi/hr/sec)

x_4 = Power Demand, PD, ($mi^2/hr^2/sec$)

x_5 = Time of Day, TD,

From the correlation matrix in Table 4.15 above and the scatter plot matrix in Figure 4.22 below, there is no high correlation or multicollinearity between predictor variables.

Table 4.15 Correlation between Predictor Variables for Model B

Pearson Correlation Coefficients, N = 53795

	Co2	Vel	Acc	PD	TD	stdx1x2	stdx1x3	stdx1x4
Co2 CO2	1.00000	0.05348	0.76028	0.61979	0.06577	-0.41082	-0.07996	-0.06355
Vel Velocity	0.05348	1.00000	-0.10829	0.03015	0.03796	-0.27560	0.14444	-0.05641
Acc Acceleration	0.76028	-0.10829	1.00000	0.68986	-0.00019	-0.52483	-0.16612	-0.18723
PD Power demand	0.61979	0.03015	0.68986	1.00000	-0.00094	-0.12277	-0.15241	0.09667
TD Time of Day	0.06577	0.03796	-0.00019	-0.00094	1.00000	-0.02026	0.01348	-0.00351
stdx1x2	-0.41082	-0.27560	-0.52483	-0.12277	-0.02026	1.00000	0.14516	0.58093
stdx1x3	-0.07996	0.14444	-0.16612	-0.15241	0.01348	0.14516	1.00000	0.57944
stdx1x4	-0.06355	-0.05641	-0.18723	0.09667	-0.00351	0.58093	0.57944	1.00000

.

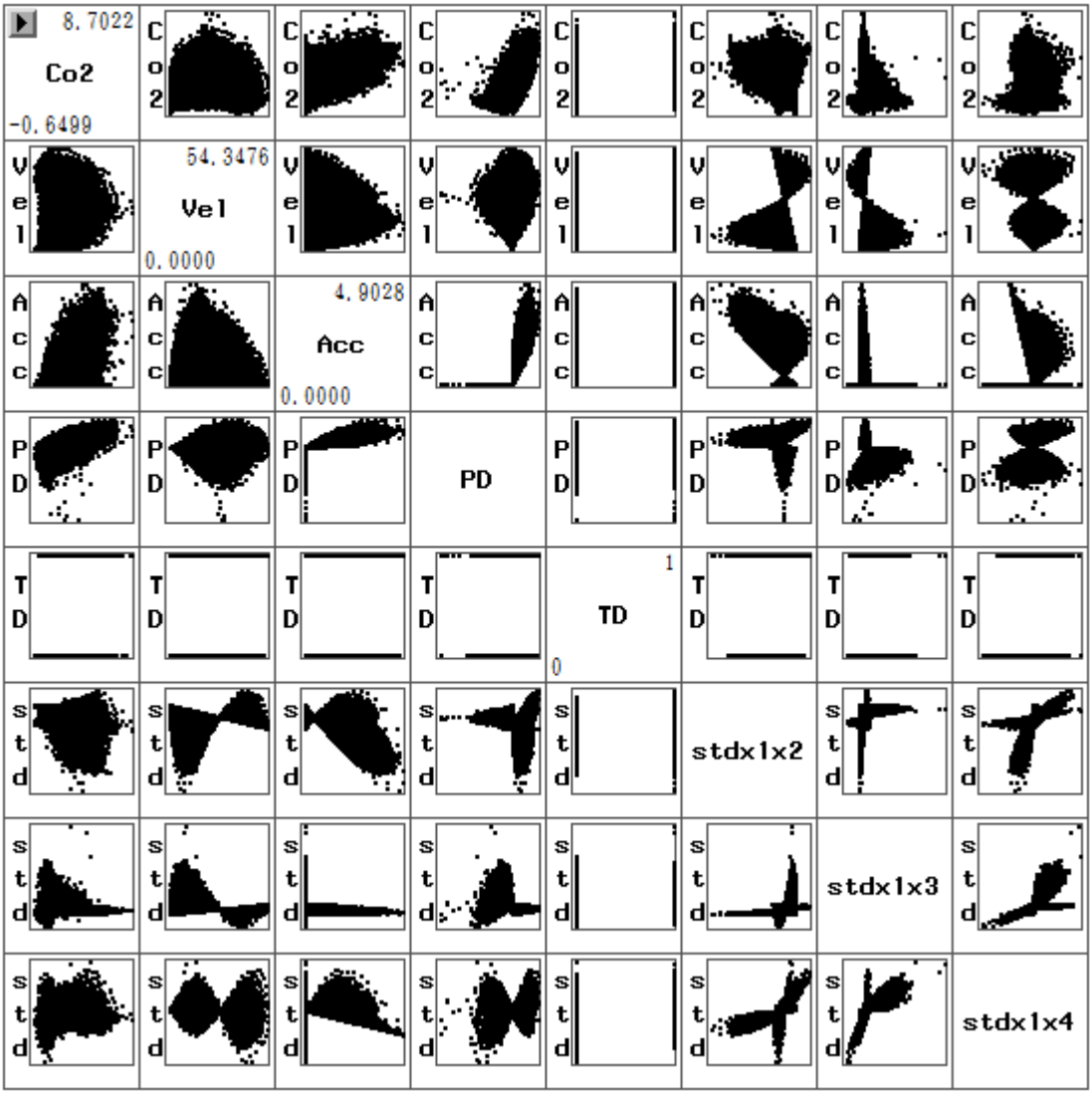


Figure 4.22 Scatter Plot Matrix of Response and Predictor Variables

The best model is selected from the two models above. To do this, various tests including diagnostic, normality and constant variance tests for the two models are conducted and make sure they satisfy them. The model that satisfies all the tests is selected as the "best" overall model.

4.6 Verifying Model Assumptions

Table 4.16 below compares Model A with Model B

Table 4.16 Model Assumption Verification

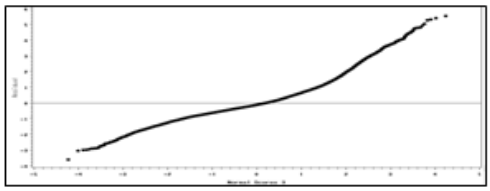
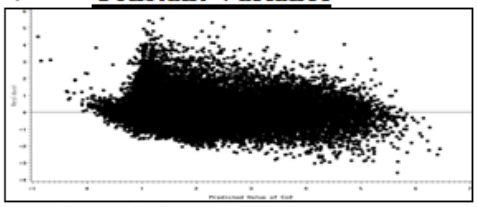
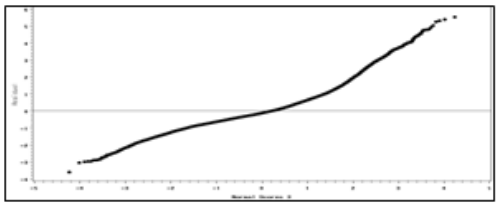
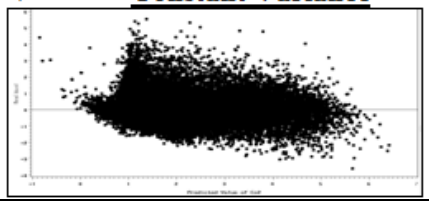
1. VERIFY ASSUMPTIONS: ❖ **Normality (MODEL A)**  Not straight enough ⇒ Normality may be **OK** $\hat{\rho}$ **=0.97 (SAS output)** $c(\alpha = 0.10, n = 53795) = 0.989$ $\hat{\rho}$ **<0.989 ⇒ Normality not OK** ❖ **Constant Variance** 	**1.VERIFY ASSUMPTIONS:** ❖ **Normality (MODEL B)**  Not straight enough ⇒ Normality may be **OK** $\hat{\rho}$ **=0.97(SASoutput)** $c(\alpha = 0.10, n = 53795) = 0.989$ $\hat{\rho}$ **slightly < 0.989 ⇒ Normality not OK** ❖ **Constant Variance** 

Table 4.16- continued

<table>
<tr>
<td>Not funnel shape, variance is constant.
Using the modified Levene test with two groups n_1=26389 and n_2=27406, we have the following SAS output:

Variances	DF	t Value	Pr > |t|
Equal	54E3	-33.95	<.2355

For $\alpha = 0.10 \Rightarrow p-value > 0.10 \Rightarrow$ fail to reject Ho, meaning constant variance (same as plot conclusion).</td>
<td>Not funnel shape, variance is constant.
Using the modified Levene test with two groups n_1=26470 and n_2=27325, we have the following SAS output:

Variances	DF	t Value	Pr > |t|
Equal	54E3	-34.35	<.2350

For $\alpha = 0.10 \Rightarrow p-value > 0.10 \Rightarrow$ fail to reject Ho, meaning constant variance (same as plot conclusion).</td>
</tr>
<tr>
<td><u>Variance Inflation</u>
We need to verify serious multicollinearity problem with model.

variable	Variance inflation
Vel	1.40531
Acc	3.89143
PD	2.57359
TD	1.00610
Stdx1x2	2.79008
Stdx1x4	1.62310

Since Max (VIF_k)= 3.89143<5 and AVG(VIF_k)=2.21494<5 $\Rightarrow$ Serious multicollinearity is not a problem.</td>
<td><u>Variance Inflation</u>
We need to verify serious multicollinearity problem with model.

variable	Variance inflation
Vel	1.43447
Acc	3.89708
PD	2.75237
TD	1.00166
Stdx1x2	1.13234
Stdx1x3	1.90268
Stdx1x4	2.82846

Since Max (VIF_k)= 3.89708<5 and AVG(VIF_k)= 2.13558<5 $\Rightarrow$ Serious multicollinearity is not a problem.</td>
</tr>
</table>

4.7 Testing Goodness of Fit for Each Model

Analysis of Variance (ANOVA) and t-test tables from the SAS output will be used to test the goodness of fit for each model:

Model A

Tables 4.17 and 4.18 show the SAS output of the ANOVA and T-test respectively, for testing the goodness of fit for Model A.

Table 4.17 SAS ANOVA Output for Model A

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	50148	8358.02550	14565.9	<.0001
Error	53788	30864	0.57381		
Corrected Total	53794	81012			

Root MSE	0.75750	R-Square	0.6190
Dependent Mean	1.74046	Adj R-Sq	0.6190
Coeff Var	43.52302		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.98168	0.00978	100.41	<.0001	0
Vel	Velocity	1	0.00757	0.00024891	30.40	<.0001	1.40531
Acc	Acceleration	1	0.93242	0.00776	120.22	<.0001	3.89143
PD	Power demand	1	0.00613	0.00016208	37.84	<.0001	2.57359
TD	Time of Day	1	0.15024	0.00656	22.89	<.0001	1.00160
stdx1x2		1	-0.12797	0.00671	-19.08	<.0001	2.97008
stdx1x4		1	0.21286	0.00756	28.16	<.0001	1.62310

Table 4.18 SAS T-test Output for Model A

The TTEST Procedure

Statistics

Variable	group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
d	1	26389	0.4516	0.4582	0.4648	0.5427	0.5474	0.5521	0.0034
d	2	27406	0.6096	0.6158	0.6221	0.5253	0.5297	0.5341	0.0032
d	Diff (1-2)		-0.167	-0.158	-0.149	0.5352	0.5384	0.5416	0.0046

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	54E3	-33.95	<.2355
d	Satterthwaite	Unequal	54E3	-33.93	<.2355

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
d	Folded F	26388	27405	1.07	<.9015

Model B

Tables 4.19 and 4.20 show the SAS output of the ANOVA and T-test respectively, for testing the goodness of fit for Model B.

Table 4.19 SAS ANOVA Output for Model B

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	50157	7165.35133	12490.9	<.0001
Error	53787	30855	0.57365		
Corrected Total	53794	81012			

Root MSE	0.75739	R-Square	0.6191
Dependent Mean	1.74046	Adj R-Sq	0.6191
Coeff Var	43.51687		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97920	0.00979	99.97	<.0001	0
Vel	Velocity	1	0.00771	0.00025144	30.67	<.0001	1.43447
Acc	Acceleration	1	0.93361	0.00776	120.30	<.0001	3.89708
PD	Power demand	1	0.00596	0.00016759	35.57	<.0001	2.75237
TD	Time of Day	1	0.15043	0.00656	22.92	<.0001	1.00166
stdx1x2		1	-0.13429	0.00689	-19.50	<.0001	3.13234
stdx1x3		1	-0.02107	0.00523	-4.03	<.0001	1.90268
stdx1x4		1	0.23910	0.00998	23.96	<.0001	2.82846

Table 4.20 SAS T-test Output for Model B

The TTEST Procedure

Statistics

Variable	group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
d	1	26470	0.4507	0.4573	0.4639	0.5416	0.5462	0.5509	0.0034
d	2	27325	0.6105	0.6168	0.6231	0.5263	0.5307	0.5352	0.0032
d	Diff (1-2)		-0.169	-0.159	-0.15	0.5352	0.5384	0.5416	0.0046

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	54E3	-34.35	<.2350
d	Satterthwaite	Unequal	54E3	-34.33	<.2350

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
d	Folded F	26469	27324	1.06	<.9011

From the ANOVA and t-test tables above**:**

❖ The regression is significant for both models. This is because the p value (P_r>F) is less than $\alpha = 0.10$.

❖ The coefficient of multiple determination (R^2) measures how well the regression model fits the data set of CO_2 emitted. $R^2 = (SSR/SSTO) = 0.6190 \Rightarrow$ It can be concluded that 62% of total variation in the mass emission rate of CO_2 is explained by the introduction of velocity, acceleration, power demand, time of day, velocity*acceleration, and velocity*power demand in the model. The model appears to fit well since the value is close to 1 ($0 \le R^2 \le 1$). The remaining 38% is not accounted for, or unexplained.

❖ Similarly for model B, $R^2 = (SSR/SSTO) = 0.6191 \Rightarrow$ It can be concluded that 62% of total variation in the mass emission rate of CO_2 is explained by the introduction of velocity, acceleration, power demand, time of day, (velocity*acceleration), (velocity*deceleration) and (velocity*power demand) in the model. The model appears to fit well since the value is close to 1 ($0 \le R^2 \le 1$). The remaining 38% is not accounted for, or unexplained. Several factors may have contributed to the unaccounted portion of the dataset. These may include changes in weather conditions, road grade, and air conditioning usage.

4.8 Summary of the Best Selected Model

Model A**:** Passed all the tests except normality test. Normality test, however, is a desired test, not a required one. In addition to that it is a simpler model to use compared

to Model B. Moreover, Model A was selected as one of the 2 best models in two model selection methods.

Model B: Passed all tests except the normality test, though the normality plot showed that normality may be OK. Model B was selected as one of the 2 best models in one of the three model selection methods.

Best model from a statistical point is model Model A**.**

$$\boxed{\hat{y} = 0.982 + 0.008x_1 + 0.923x_2 + 0.006x_4 + 0.150x_5 - 0.128x_1x_2 + 0.213x_1x_4} \quad \text{Eqn. 4.3}$$

4.9 Fitting the Preliminary Model for Highway Data

Table 4.21 Parameter Estimates

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97915	0.02215	44.21	<.0001	0
Vel	Velocity	1	0.02459	0.00029903	82.25	<.0001	1.23382
Acc	Acceleration	1	0.57065	0.01962	29.08	<.0001	2.83860
Dec	Deceleration	1	-0.21946	0.01784	-12.30	<.0001	4.37172
PD	Power demand	1	0.01178	0.00030283	38.89	<.0001	6.95577
TD	Time of Day	1	0.03357	0.00743	4.52	<.0001	1.04951

The result (SAS output) obtained is as follows:

From the SAS table in Table 4.21 above:

$b_0 = 0.97915$

$b_1 = 0.02459$

$b_2 = 0.57065$

$b_3 = -0.21946$

$b_4 = 0.01178$

$b_5 = 0.03357$

So, the estimated regression function is:

$$\boxed{\hat{y} = 0.979 + 0.025x_1 + 0.571x_2 - 0.219x_3 + 0.012x_4 + 0.034x_5} \qquad \text{Eqn. 4.5}$$

4.10 Evaluating the Existing Model

The same evaluation procedure used for the arterial data is employed.

4.10.1 Preliminary Plots

The same procedure used in evaluating the arterial data is employed.

4.10.1.1 Response Variables vs. Predictor Variables and Predictor Variables vs. Predictor Variables

A similar scatter plot analysis between the mass emission rate of CO_2 (response variable) and the 5 different predictor variables are made with regards to the highway data. The plots are shown in Figure 4.23 below.

Referring to Figure 4.23:

A correlation between "CO_2" and "Vehicle Acceleration" is noticeable. A slight linear trend shows a relationship between these two variables. This means that "Vehicle Acceleration" can noticeably influence "CO_2" emission rate and the positive slope of the trend conveys that as "Vehicle Acceleration" increases, the mass emission rate of "CO_2" increases as well.

Plots do not show a high correlation between "CO_2" with "Vehicle Velocity" and "Time of Day". This means that the two variables may slightly influence "CO_2", although this is just a basic evaluation. Further tests and assessments will be conducted in order to evaluate influences caused by the interaction terms as well.

"CO_2" and "Power Demand" seem to be slightly correlated, which means this predictor variable is influencing the response variable.

"Vehicle Acceleration" and "Power Demand" and "Vehicle Deceleration" and "Power Demand", on the other hand, look to be highly correlated and might lead to a multicollinearity problem. This is the inter-correlation of independent variables and occurs when two or more predictors are highly correlated. In the remaining plots, the points seem to be fairly scattered and variables seem to be uncorrelated.

In some predictor vs. predictor plots as well as response vs. predictor plots, some points appear to be potential x-outliers. Later some procedures will be conducted to evaluate whether they should be thrown out.

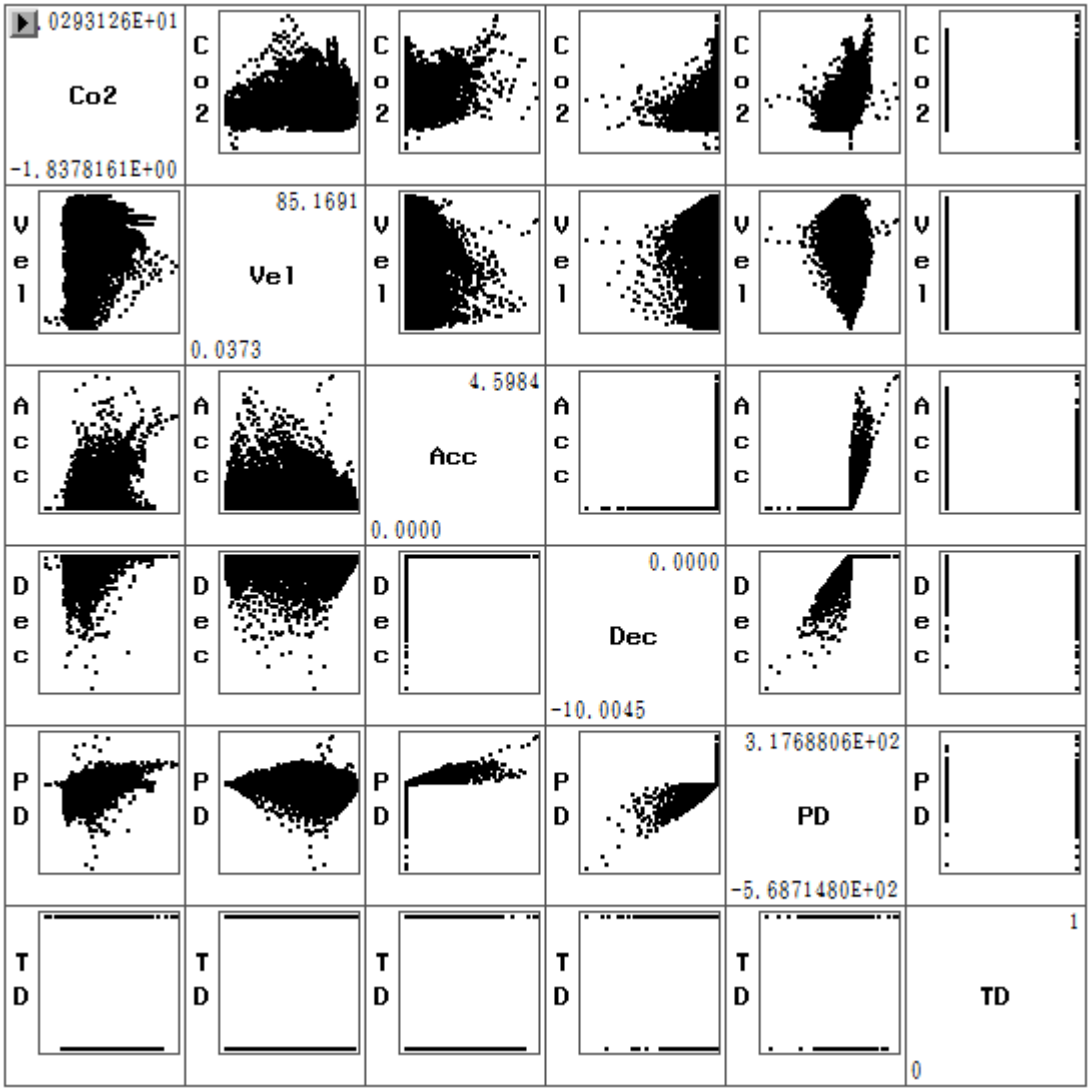


Figure 4.23 Scatter Plots of Response and Predictor Variables

Similarly, Table 4.20 shows the correlation between the response variable and predictor variables for the highway data.

Table 4.22 Correlation between Variables

Pearson Correlation Coefficients, N = 55342						
	Co2	Vel	Acc	Dec	PD	TD
Co2 CO2	1.00000	0.26077	0.31009	0.31733	0.41598	-0.04149
Vel Velocity	0.26077	1.00000	-0.27262	0.21424	0.00146	-0.21521
Acc Acceleration	0.31009	-0.27262	1.00000	0.24699	0.65491	0.06141
Dec Deceleration	0.31733	0.21424	0.24699	1.00000	0.79511	-0.07042
PD Power demand	0.41598	0.00146	0.65491	0.79511	1.00000	-0.01157
TD Time of Day	-0.04149	-0.21521	0.06141	-0.07042	-0.01157	1.00000

From Table 4.22, we have the following preliminary conclusions:

$\mathbf{r_{y1}}$ = 0.26 indicates that CO_2 and "Vehicle Velocity" are somewhat positively correlated.

$\mathbf{r_{y2}}$ = 0.31 indicates that CO_2 and "Vehicle Acceleration" are somewhat positively correlated.

$\mathbf{r_{y3}}$ = 0.32 indicates that CO_2 and "Vehicle Deceleration" are somewhat positively correlated.

$\mathbf{r_{y4}}$ = 0.42 indicates that CO_2 and "Power Demand" are somewhat positively correlated.

$\mathbf{r_{y5}}$ = -0.04 indicates that there is little correlation between CO_2 and "Time of Day".

$\mathbf{r_{12}}$ = -0.27 indicates that "Vehicle Velocity" and "Vehicle Acceleration" are somewhat negatively correlated.

$\mathbf{r_{13}}$ = 0.21 indicates that “Vehicle Velocity” and “Vehicle Deceleration” are somewhat positively correlated.

$\mathbf{r_{14}}$ = 0.001 indicates that there is little correlation between “Vehicle Velocity” and “Power Demand”.

$\mathbf{r_{15}}$ = -0.22 indicates that “Vehicle Velocity” and “Time of Day” are somewhat negatively correlated.

$\mathbf{r_{23}}$ = 0.25 indicates that “Vehicle Acceleration” and “Vehicle Deceleration” are somewhat positively correlated.

$\mathbf{r_{24}}$ = 0.65 indicates that “Vehicle Acceleration” and “Power Demand” are positively highly correlated.

$\mathbf{r_{25}}$ = -0.06 indicates that there is little correlation between “Vehicle Acceleration” and “Time of Day”.

$\mathbf{r_{34}}$ = 0.80 indicates that there is little correlation between “Vehicle Deceleration” and “Power Demand” are positively highly correlated.

$\mathbf{r_{35}}$ = -0.07 indicates that there is little correlation between “Vehicle Deceleration” and “Time of Day”.

$\mathbf{r_{45}}$ = -0.01 indicates that there is little correlation between “Power Demand” and “Time of Day”.

From the numbers presented in Table 4.22, it can be deduced that a correlation exists between predictor variables in at least two cases, “Vehicle Acceleration” and “Power Demand” and “Vehicle Deceleration” and “Power Demand”. This may be a

potential multicollinearity problem. Further tests will help in making a final decision on the multicolinearity issue.

4.10.2 Residual Plots for Highway Data

4.10.2.1 Residuals vs. $\hat{y}_i$

From Figure 4.24 below, we cannot detect any funnel. The plot does not show any curvilinear shape but a couple of potential outliers can be observed. So far what can be concluded is that the "constant variance" assumption is probably OK. Further tests will be conducted to verify the conclusion from this plot.

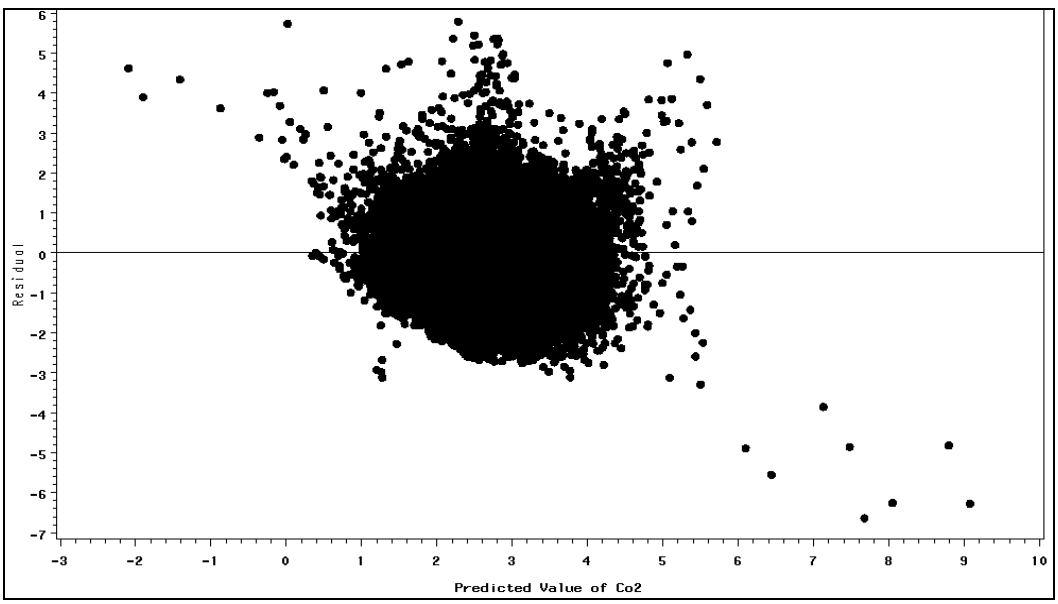


Figure 4.24 Residual vs. Predicted Value of CO_2

4.10.2.2 Residuals (e) vs. Predictor Variables

In this series of plots shown in Figure 4.25, curvature cannot be detected. Curvature would indicate violation of the assumed model form, which is currently a linear equation. There are also some possible outliers.

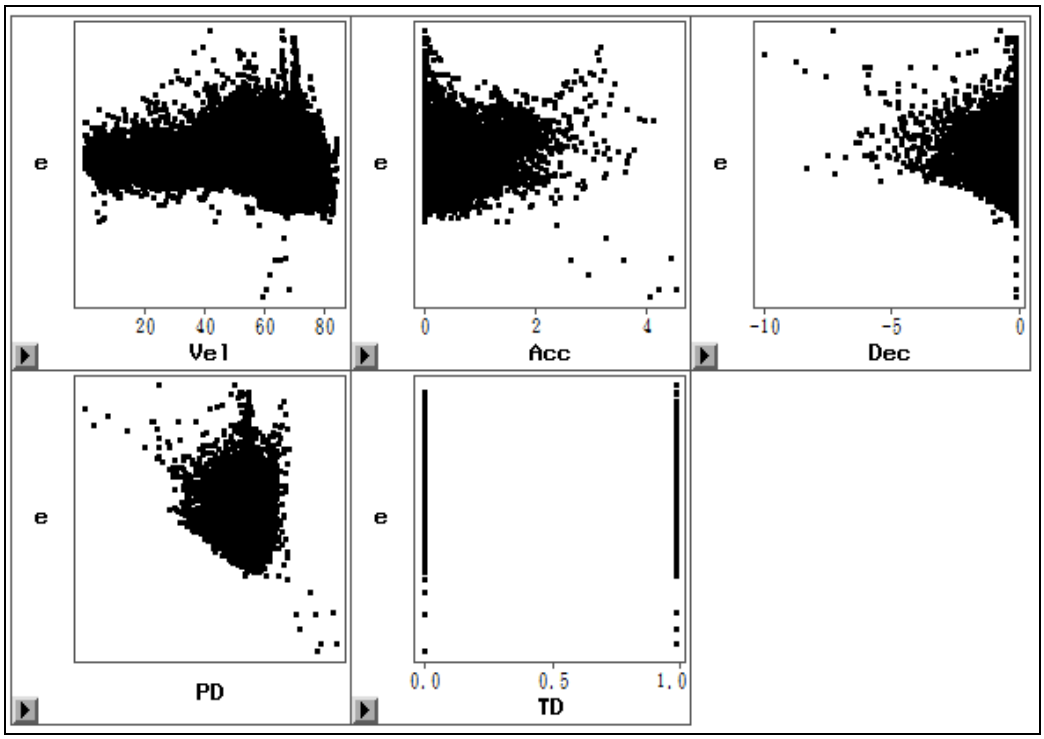


Figure 4.25 Residual vs. Predictor Variables

4.10.2.3 Residuals vs. Interaction Terms

Basically, what we are looking for in Figure 4.26 plots below is a trend between the residuals and any interaction terms. If a trend is seen, we can suggest adding the associated interaction terms to the model.

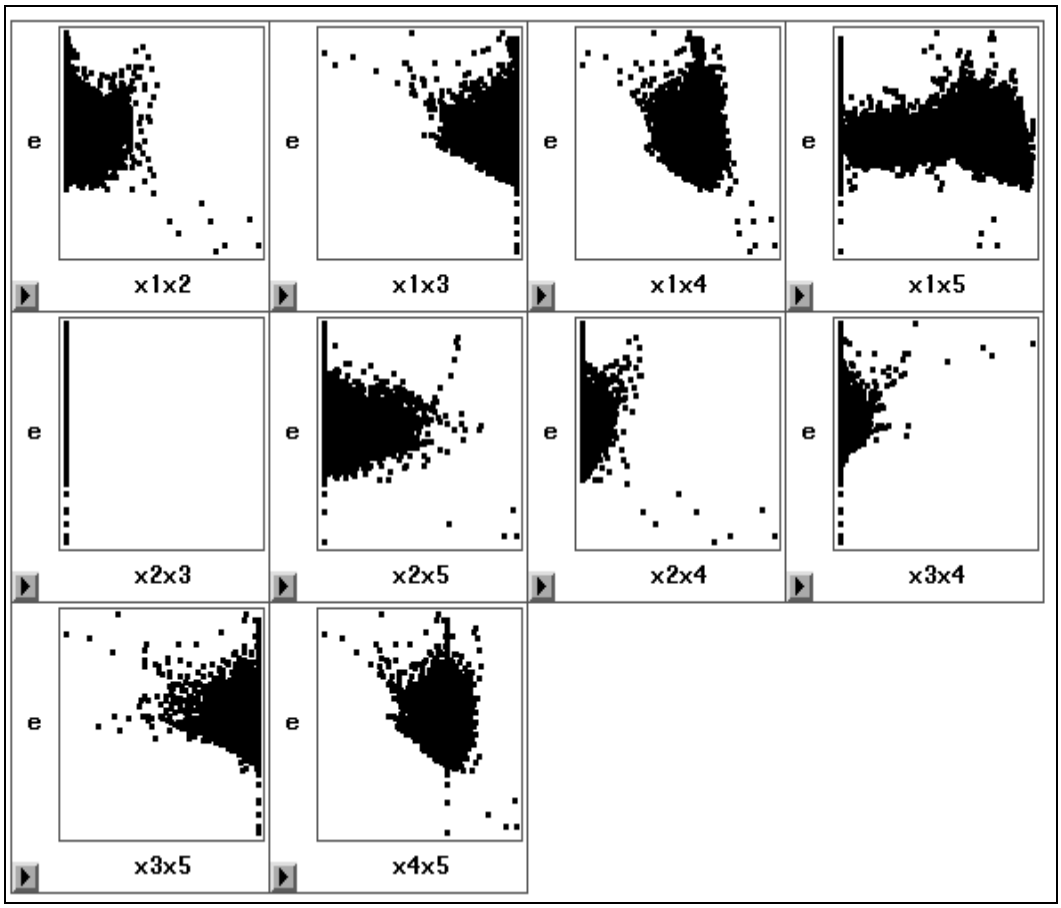


Figure 4.26 Residuals vs. Interaction Terms

As shown in Figure 4.26 above, there is not a strong trend among the plots. There might be a slight trend between residuals and x2x5, x2x4 and x3x4, but at this point we are not suggesting that any interaction terms should be added.

4.10.3 Normal Probability Plot of e_i's

The normality plot shown in Figure 4.27 does not deviate much from a straight-line. There is slight indication of a longer right tail and a shorter left tail than the normal distribution. However, on the strength of the straightness of the plot, normality seems

generally OK. Further tests will be conducted and a final decision on normality will be made later.

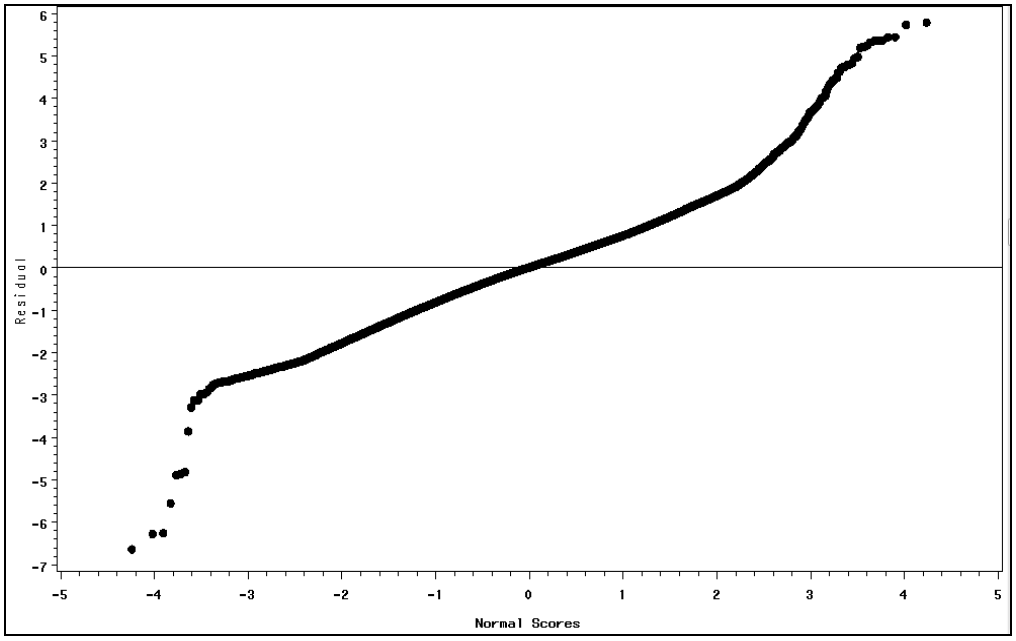


Figure 4.27 Normal Probability Plot

4.10.4 Time Series Plot

The time series plot in Figure 4.28 below shows a slight upward trend, which might indicate serial correlation between the observations.

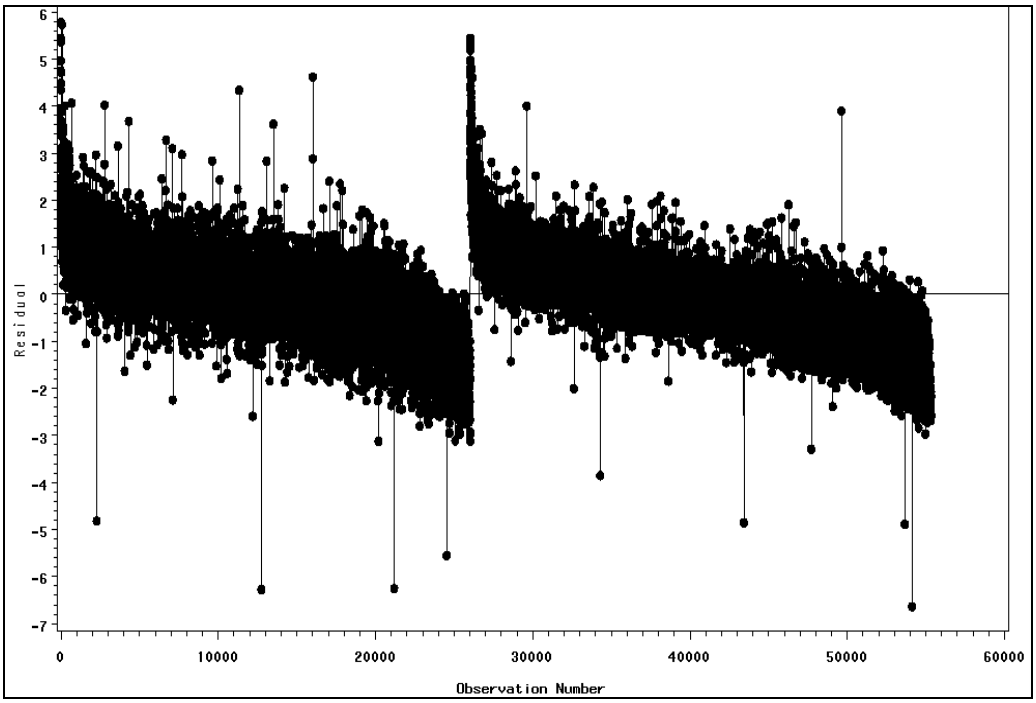


Figure 4.28 Time Series Plot

4.11 Verifying the Model Assumptions

In this section, tests and diagnostics are presented to further verify model assumptions that have been assessed visually from the plots.

4.11.1 Test for Constant Variance (Modified Levene Test)

A further test is conducted to confirm the initial assumption from the plots that the "constant variance" assumption is OK. Figure 4.29 and Table 4.21 are considered in this test.

Ho: Variances are equal

Test: vs.

H1: Variances are not equal

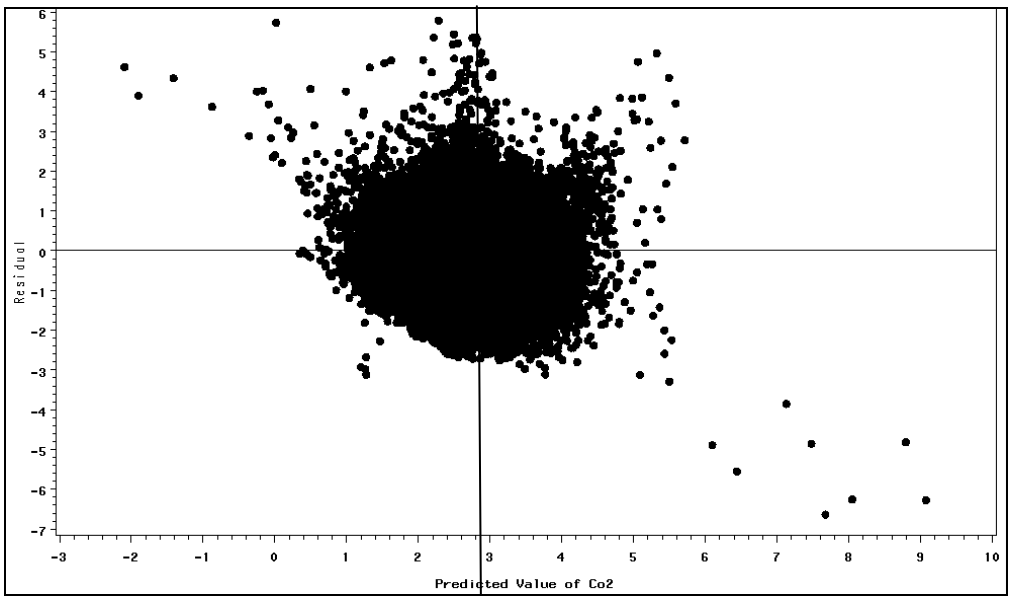


Figure 4.29 Residual vs. Predicted Value of CO_2

As discussed previously, we concluded from the residual plots that the "constant variance" assumption is OK. The Modified Levene Test will be used to confirm this conclusion. This test does not depend on the normality of the error terms. To conduct this test, we divide the data into two groups of about the same number of observations. The cutoff line on $\hat{y}_i$ will fall around value "2.8". This line is illustrated on the residual vs. $\hat{y}_i$ plot in Figure 4.29.

Table 4.23 SAS Output for Testing Constant Variance

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	55E3	12.83	<.3579
d	Satterthwaite	Unequal	55E3	12.84	<.3579

Table 4.23 shows the result of conducting the test using SAS and using 2 groups with about the same number of observations in each of them. Assuming α=0.1 for the confidence level, the p-value for the test turns out to be greater than α=0.1. Thus:

P-value=0.3579>0.1=$\alpha \Rightarrow$ we fail to reject H_0.

This leads us to conclude that the assumption of "constant variance" is OK.

4.11.2 Test for Normality

Further testing is conducted to support the initial assumption that normality is generally OK, as follows:

H_0: Normality is OK

Test: Vs.

H_1: Normality is violated

The following table is obtained from the SAS output.

Table 4.24 Normality Test from SAS Output

Pearson Correlation Coefficients, N = 55342
Prob > |r| under H0: Rho=0

	e	enrm
e	1.00000	0.99396
Residual		<.0001
enrm	0.99396	1.00000
Normal Scores	<.0001	

We need to find $c(\alpha, n)$ in order to compare it with the correlation value $\hat{\rho}$ (from the SAS output) in Table 4.22. With $\alpha = 0.10$, we have $c(\alpha = 0.10, n = 55342) = 0.989$; this implies $\hat{\rho} = 0.99396 > 0.989 \Rightarrow$ fail to reject Ho. Hence normality is OK.

4.11.3 Outlier Diagnostic

Y-Outlier: Bonferroni test cutoff: t(1-α/2n, n-p-1) = t(1-0.10/(2*55342)); 55342-6-1) = t(0.999999, 55335) = 3.291. When compared with the Rstudent values in APPENDIX A, the values are less than 3.291; we then conclude that there are no y – outliers.

X-Outlier: Leverage values (h_{ii}) are utilized to check whether x-outliers exist or not. The h_{ii} values (APPENDIX A) are compared with the value $\frac{2p}{n}$ = 2(6)/55342 = 0.0002. If any h_{ii} value from APPENDIX A>0.0002, then it can be concluded that x-outliers exist. We can detect a few observations greater than 0.0002, but we do not have any reason for deleting them; hence, we leave them in the dataset and continue with the analysis.

Influence on the Fitted Values: DFFITS

The DFFITS values for any outliers (APPENDIX A) are compared with the value $2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{6}{55342}}$ = 0.021. If the DFFITS value for an outlier > 0.021, then that outlier has high influence on the estimated regression equation and fitted values. From APPENDIX A, there are a few DFFITS values higher than $2\sqrt{\frac{p}{n}}$ = 0.021.

Cut-off DFBETA (influence on the individual LSEs)

The DFBETA values (APPENDIX A) are compared with the value $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{55342}}$ = 0.009. If the DFBETA value for an outlier >0.009, then that outlier

has high influence on the least square estimates. From the value of DFBETA in APPENDIX A, a few observations have higher values than 0.009.

4.11.4 Multicollinearity - Variance Inflation

From the assessment of the predictor variables using the Correlation Matrix, a potential multicollinearity problem was discovered. Table 4.21, repeated here for convenience, shows only one number slightly greater than “5” under the “Variance Inflation” column. Therefore, up to this point, multicollinearity may not be a problem.

Table 4.21 SAS Output for Assessing Variance Inflation

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97915	0.02215	44.21	<.0001	0
Vel	Velocity	1	0.02459	0.00029903	82.25	<.0001	1.23382
Acc	Acceleration	1	0.57065	0.01962	29.08	<.0001	2.83860
Dec	Deceleration	1	-0.21946	0.01784	-12.30	<.0001	4.37172
PD	Power demand	1	0.01178	0.00030283	38.89	<.0001	6.95577
TD	Time of Day	1	0.03357	0.00743	4.52	<.0001	1.04951

4.11.5 Assessing Partial Regression Plots for Highway Data

Following the same procedure as in the arterial data plots, partial regression graphs for highway data are plotted and analyzed. The resulting plots are shown in Figure 4.20-4.39. Trends are difficult to see on both Figures 4.30-4.39 and 4.340, hence, all terms will be added and the model selection procedures will be used to try to identify any helpful interaction terms.

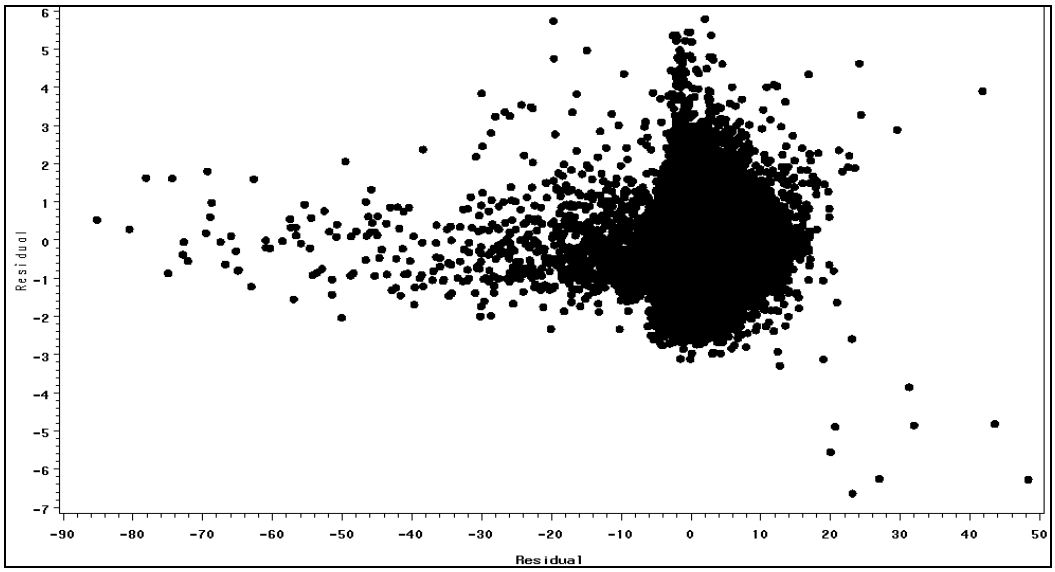


Figure 4.30 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x2/x1x2x3x4x5)

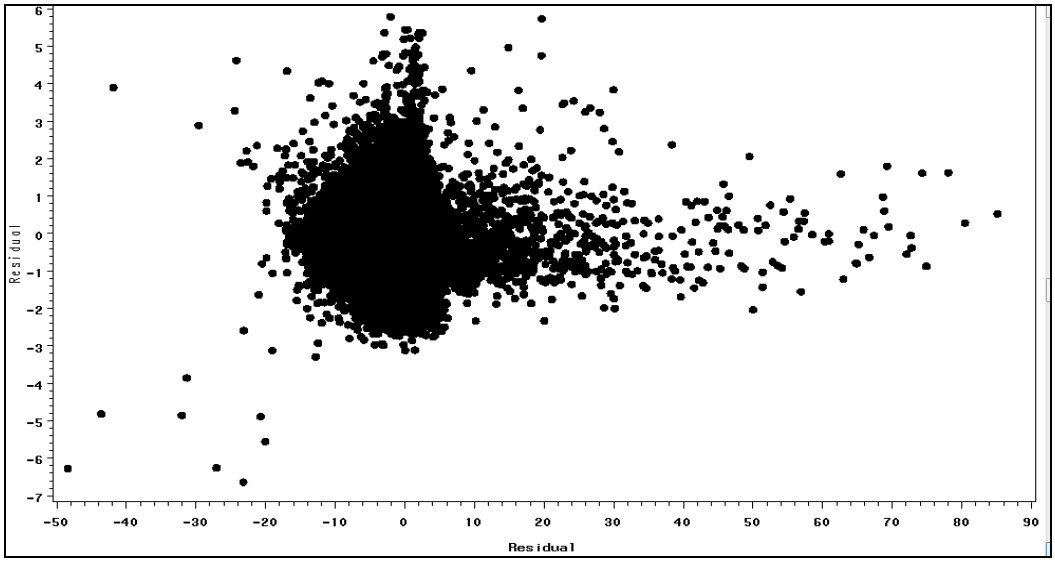


Figure 4.31 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x3/x1x2x3x4x5)

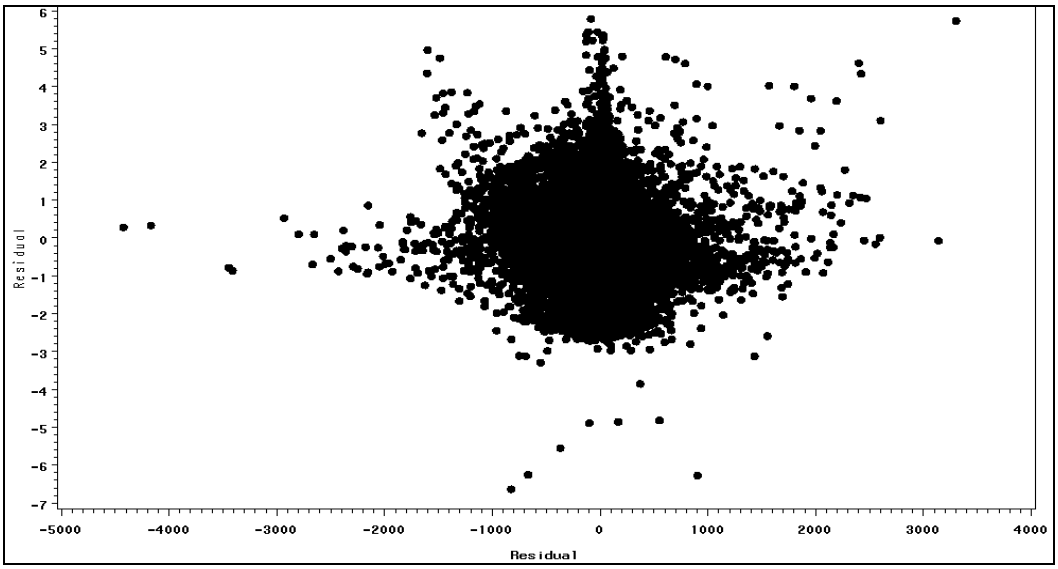


Figure 4.32 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x4/x1x2x3x4x5)

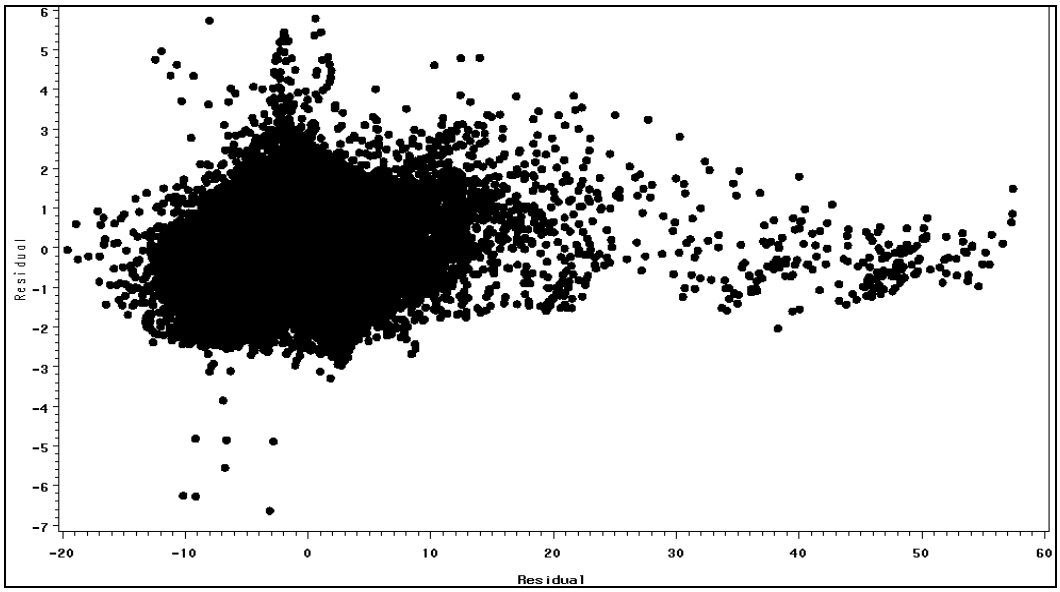


Figure 4.33 Residual e(Y/x1x2x3x4x5) vs. Residual e(x1x5/x1x2x3x4x5)

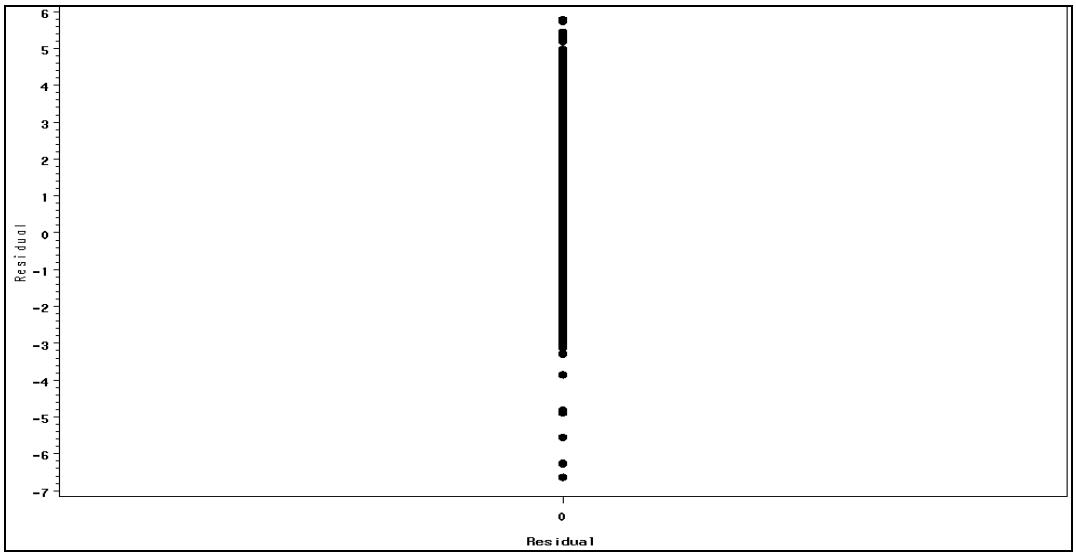


Figure 4.34 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x3/x1x2x3x4x5)

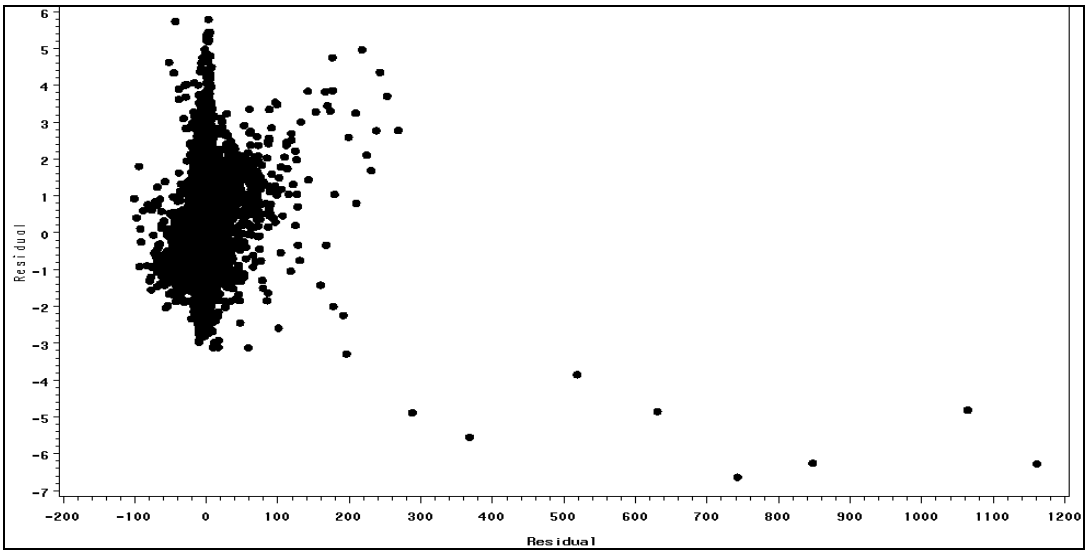


Figure 4.35 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x4/x1x2x3x4x5)

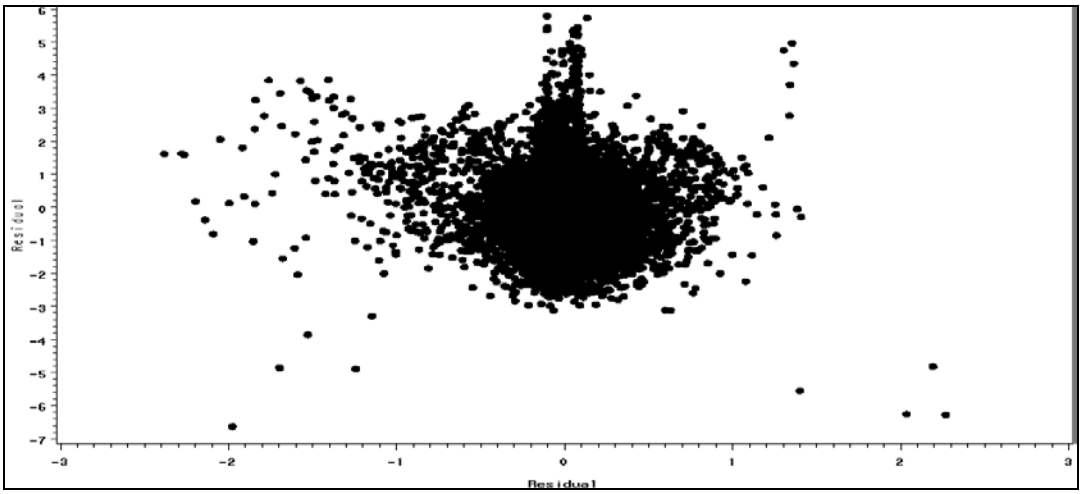


Figure 4.36 Residual e(Y/x1x2x3x4x5) vs. Residual e(x2x5/x1x2x3x4x5)

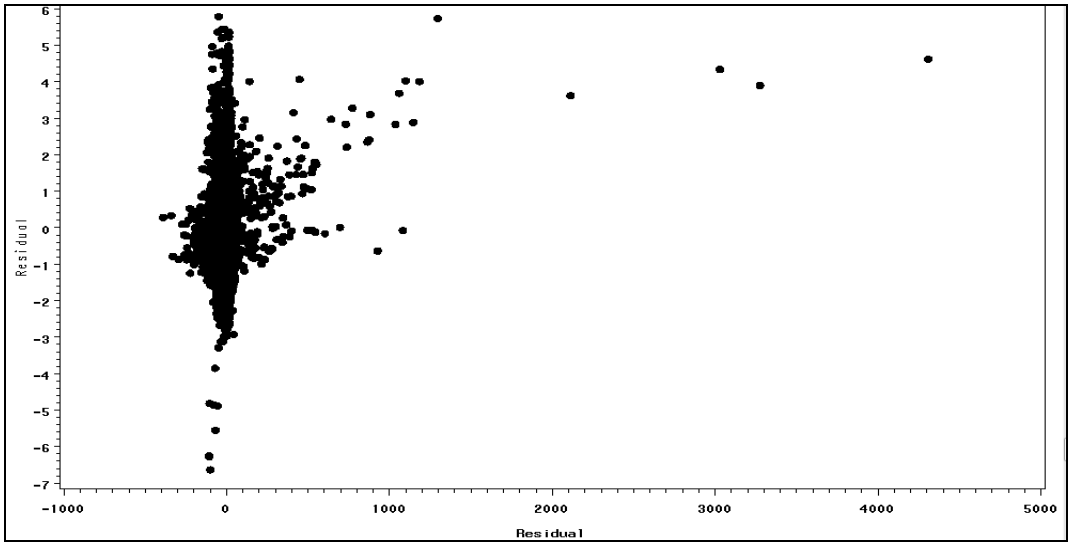


Figure 4.37 Residual e(Y/x1x2x3x4x5) vs. Residual e(x3x4/x1x2x3x4x5)

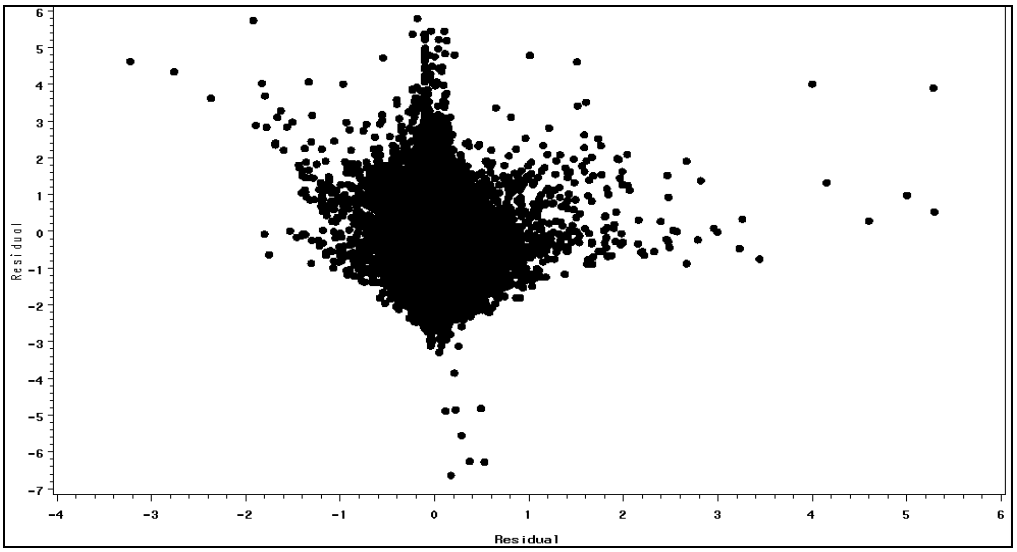


Figure 4.38 Residual e(Y/x1x2x3x4x5) vs. Residual e(x3x5/x1x2x3x4x5)

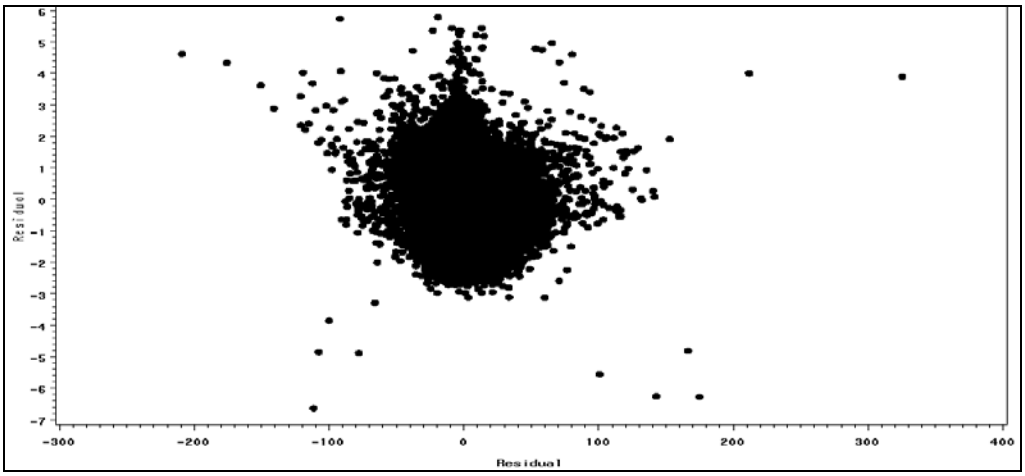


Figure 4.39 Residual e(Y/x1x2x3x4x5) vs. Residual e(x4x5/x1x2x3x4x5)

Table 4.25 Correlation between Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 55342								
	Co2	Vel	Acc	Dec	PD	TD	x1x2	x1x3
Co2 CO2	1.00000	0.26077	0.31009	0.31733	0.41598	-0.04149	0.39074	0.30086
Vel Velocity	0.26077	1.00000	-0.27262	0.21424	0.00146	-0.21521	-0.03739	0.02903
Acc Acceleration	0.31009	-0.27262	1.00000	0.24699	0.65491	0.06141	0.89523	0.27152
Dec Deceleration	0.31733	0.21424	0.24699	1.00000	0.79511	-0.07042	0.27877	0.91287
PD Power demand	0.41598	0.00146	0.65491	0.79511	1.00000	-0.01157	0.73381	0.87155
TD Time of Day	-0.04149	-0.21521	0.06141	-0.07042	-0.01157	1.00000	0.02177	-0.03192
x1x2	0.39074	-0.03739	0.89523	0.27877	0.73381	0.02177	1.00000	0.30647

Table 4.26 Correlation between Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 55342								
	Co2	Vel	Acc	Dec	PD	TD	x1x2	x1x3
x1x3	0.30086	0.02903	0.27152	0.91287	0.87155	-0.03192	0.30647	1.00000
x1x4	0.40726	0.00476	0.60613	0.73288	0.98303	-0.00766	0.72583	0.85353
x1x5	0.06512	0.11743	-0.00917	-0.00289	-0.00638	0.93339	0.02702	-0.02844
x2x3	.	.	.	.	.	.	.	.
x2x4	0.27794	-0.13902	0.80279	0.11791	0.49036	0.02286	0.77240	0.12962
x2x5	0.14745	-0.29268	0.69358	0.16091	0.41719	0.40754	0.56484	0.17689
x3x4	-0.13268	-0.07956	-0.08950	-0.77106	-0.59382	0.04934	-0.10102	-0.75913
x3x5	0.21918	0.24036	0.15790	0.81678	0.57961	-0.28430	0.17822	0.68349
x4x5	0.26907	0.01432	0.45111	0.66609	0.74990	0.00109	0.49431	0.69398

Table 4.27 Correlation between Standardized Interaction Terms and Predictors

Pearson Correlation Coefficients, N = 55342

	Co2	Vel	Acc	Dec	PD	TD	stdx1x2	stdx1x3
Co2 CO2	1.00000	0.26077	0.31009	0.31733	0.41598	-0.04149	-0.07081	-0.12862
Vel Velocity	0.26077	1.00000	-0.27262	0.21424	0.00146	-0.21521	0.29990	-0.26821
Acc Acceleration	0.31009	-0.27262	1.00000	0.24699	0.65491	0.06141	-0.61734	-0.12759
Dec Deceleration	0.31733	0.21424	0.24699	1.00000	0.79511	-0.07042	-0.12276	-0.61407
PD Power demand	0.41598	0.00146	0.65491	0.79511	1.00000	-0.01157	-0.16573	-0.22065
TD Time of Day	-0.04149	-0.21521	0.06141	-0.07042	-0.01157	1.00000	-0.04410	0.06652
stdx1x2	-0.07081	0.29990	-0.61734	-0.12276	-0.16573	-0.04410	1.00000	0.23005

Table 4.28 Correlation between Standardized Interaction Terms and Predictors

The CORR Procedure

Pearson Correlation Coefficients, N = 55342

	Co2	Vel	Acc	Dec	PD	TD	stdx1x2	stdx1x3
stdx1x3	-0.12862	-0.26821	-0.12759	-0.61407	-0.22065	0.06652	0.23005	1.00000
stdx1x4	-0.12065	0.00151	-0.36381	-0.46604	-0.25921	0.02592	0.57077	0.69000
stdx1x5	0.27329	0.67154	-0.08492	0.12325	0.02273	-0.02518	0.13195	-0.19541
stdx2x3	-0.09019	-0.37980	0.43181	-0.76737	-0.30674	0.10618	-0.29423	0.48715
stdx2x4	0.19121	-0.15311	0.70353	-0.09160	0.26268	0.02822	-0.40888	-0.00771
stdx2x5	-0.03069	-0.08587	0.13477	0.06683	0.05947	0.00727	-0.20743	-0.09302
stdx3x4	-0.10457	-0.08367	-0.03908	-0.74206	-0.53887	0.05081	0.03350	0.37717
stdx3x5	0.06277	0.12247	0.06567	0.43637	0.24670	-0.00819	-0.09096	-0.45447
stdx4x5	0.01276	0.02288	0.05920	0.24986	0.18264	-0.00136	-0.05343	-0.17817

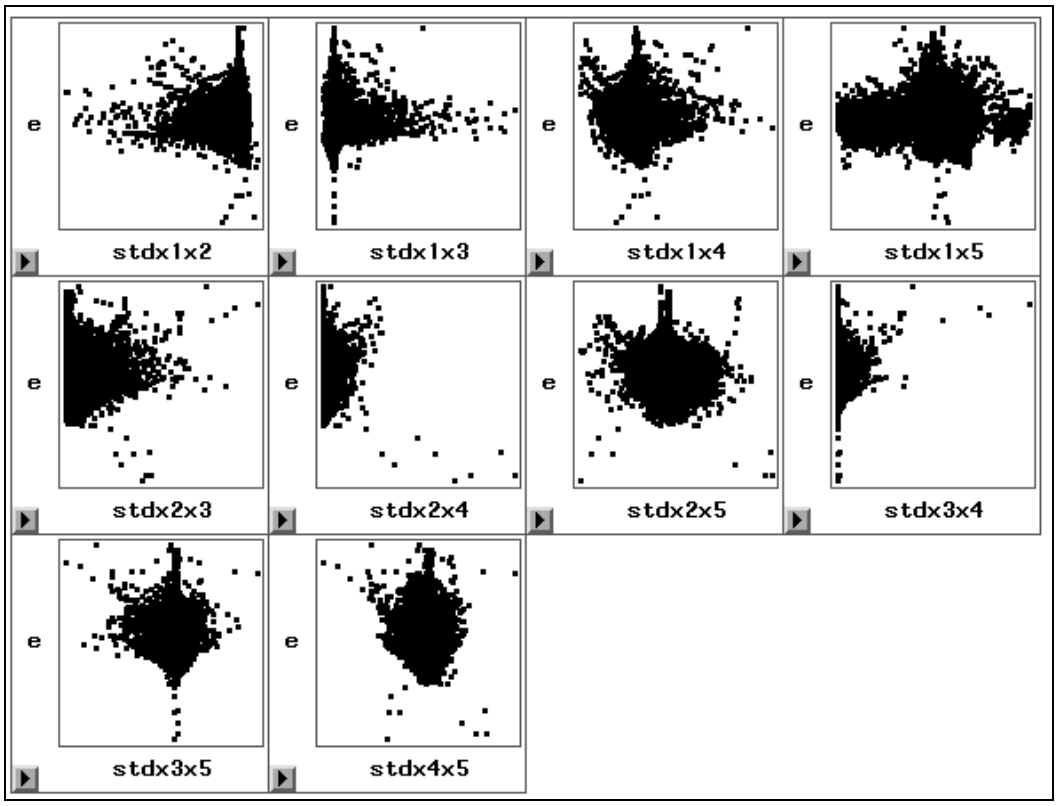


Figure 4.40 Scatter Plots of Residual vs. Standardized Interaction Terms

4.12 Model Selection (Best Potential Model) for Highway Data

The same selection process used for the arterial data is applied to the highway data.

4.12.1 Best Subset Selection

The result from the SAS program is shown in Table 4.29 below.

Table 4.29 SAS Output for the Best Subset Method

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
1	0.1730	0.1730	7134.010	PD
1	0.1007	0.1007	12598.97	Dec

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
2	0.2407	0.2407	2022.655	Vel PD
2	0.2249	0.2250	3213.572	Vel Acc

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
3	0.2650	0.2650	188.2306	Vel Acc stdx1x2
3	0.2646	0.2646	221.6498	Vel Acc PD

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
4	0.2669	0.2669	46.6880	Vel Acc PD stdx1x3
4	0.2668	0.2669	51.2218	Vel PD stdx1x2 stdx2x3

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
5	0.2672	0.2672	26.2143	Vel Acc PD stdx1x3 stdx2x4
5	0.2672	0.2672	26.5645	Vel Acc PD TD stdx1x3

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
6	0.2675	0.2675	5.2061	Vel Acc PD TD stdx1x3 stdx2x4
6	0.2675	0.2675	5.3663	Vel PD TD stdx1x2 stdx2x3 stdx2x4

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
7	0.2675	0.2675	7.0191	Vel Acc PD TD stdx1x3 stdx1x4 stdx2x4
7	0.2674	0.2675	7.0774	Vel PD TD stdx1x2 stdx1x4 stdx2x3 stdx2x4

Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
8	0.2674	0.2675	8.9998	Acc PD TD stdx1x2 stdx1x3 stdx1x4 stdx2x3 stdx2x4
8	0.2674	0.2675	9.0000	Acc Dec PD TD stdx1x2 stdx1x3 stdx1x4 stdx2x4

In conclusion, using the idea of R^2 leveling off and C_p close to P (the number of parameters), the following two best models were selected.

- Vel , Acc, PD, TD, stdx1x3 and stdx2x4 (1st choice)

- Vel , PD, TD, stdx1x2, stdx2x3 and stdx2x4 (2nd choice)

4.12.2 Stepwise Regression

The result from the SAS program is shown in Table 4.30 below.

Table 4.30 SAS Output for Stepwise Regression Method

Variable PD Entered: R-Square = 0.1730 and C(p) = 7134.010

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	9461.91599	9461.91599	11580.0	<.0001
Error	55340	45218	0.81709		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2.74110	0.00384	415738	508801	<.0001
PD	0.01312	0.00012195	9461.91599	11580.0	<.0001

Variable Vel Entered: R-Square = 0.2407 and C(p) = 2022.655

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	13163	6581.51044	8772.68	<.0001
Error	55339	41517	0.75023		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.48668	0.01824	4986.57329	6646.75	<.0001
Vel	0.01924	0.00027397	3701.10488	4933.31	<.0001
PD	0.01311	0.00011685	9444.61848	12589.0	<.0001

Table 4.30-continued

Variable stdx2x3 Entered: R-Square = 0.2650 and C(p) = 188.2306

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	14492	4830.74849	6651.90	<.0001
Error	55338	40188	0.72622		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.08998	0.02020	2115.38909	2912.87	<.0001
Vel	0.02425	0.00029390	4945.87446	6810.43	<.0001
PD	0.01484	0.00012183	10769	14828.6	<.0001
stdx2x3	0.28037	0.00655	1329.22459	1830.33	<.0001

Variable stdx1x2 Entered: R-Square = 0.2669 and C(p) = 51.2218

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	14593	3648.21542	5036.08	<.0001
Error	55337	40087	0.72442		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.04670	0.02050	1888.15170	2606.45	<.0001
Vel	0.02491	0.00029869	5036.52742	6952.53	<.0001
PD	0.01445	0.00012608	9511.76277	13130.2	<.0001
stdx1x2	-0.02415	0.00205	100.61621	138.89	<.0001
stdx2x3	0.25770	0.00682	1033.70091	1426.94	<.0001

Variable stdx2x4 Entered: R-Square = 0.2672 and C(p) = 26.6210

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	14612	2922.42312	4036.04	<.0001
Error	55336	40068	0.72408		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.06879	0.02094	1886.31441	2605.11	<.0001
Vel	0.02456	0.00030586	4670.27488	6449.93	<.0001
PD	0.01405	0.00014741	6580.20942	9087.67	<.0001
stdx1x2	-0.02161	0.00211	76.16803	105.19	<.0001
stdx2x3	0.22707	0.00904	456.40468	630.32	<.0001
stdx2x4	0.01311	0.00254	19.25395	26.59	<.0001

Table 4.30-continued

Variable TD Entered: R-Square = 0.2675 and C(p) = 5.3663

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14629	2438.15794	3368.60	<.0001
Error	55335	40051	0.72379		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.03429	0.02212	1581.86950	2185.54	<.0001
Vel	0.02484	0.00031102	4616.07190	6377.64	<.0001
PD	0.01404	0.00014740	6569.20211	9076.12	<.0001
TD	0.03581	0.00743	16.83202	23.26	<.0001
stdx1x2	-0.02184	0.00211	77.77648	107.46	<.0001
stdx2x3	0.22545	0.00905	449.32462	620.79	<.0001
stdx2x4	0.01335	0.00254	19.97313	27.60	<.0001

All variables left in the model are significant at the 0.05 level. No other variable met the 0.05 significance level for entry into the model.

- Vel, PD, TD, stdx1x2, stdx2x3 and stdx2x4 (1st choice)
- Vel , PD, stdx1x2, stdx2x3 and stdx2x4 (2nd choice)

4.12.3 Backward Deletion

We ended up with the SAS program selecting the model in Table 4.31 below.

Table 4.31 SAS Output for Backward Deletion Method

All Variables Entered: R-Square = 0.2675 and C(p) = 9.0000
The model is not of full rank. A subset of the model which is of full rank is chosen.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	14629	1828.65160	2526.42	<.0001
Error	55333	40051	0.72381		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.97849	0.02332	1274.53058	1760.86	<.0001
Vel	0.02482	0.00031661	4448.90034	6146.49	<.0001
Acc	0.31987	0.04394	38.36398	53.00	<.0001
Dec	-0.30927	0.03418	59.25204	81.86	<.0001
PD	0.01391	0.00055845	448.96515	620.28	<.0001
TD	0.03574	0.00743	16.74582	23.14	<.0001
stdx1x2	-0.02025	0.00363	22.49091	31.07	<.0001
stdx1x4	-0.00221	0.00727	0.06651	0.09	0.7618
stdx2x4	0.01265	0.00281	14.72101	20.34	<.0001

Variable stdx1x4 Removed: R-Square = 0.2675 and C(p) = 7.0919

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	14629	2089.87804	2887.37	<.0001
Error	55334	40051	0.72380		
Corrected Total	55341	54680			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.97940	0.02312	1298.78328	1794.40	<.0001
Vel	0.02481	0.00031487	4494.70855	6209.88	<.0001
Acc	0.32464	0.04102	45.33271	62.63	<.0001
Dec	-0.30219	0.02496	106.08784	146.57	<.0001
PD	0.01381	0.00046190	647.31911	894.33	<.0001
TD	0.03571	0.00743	16.72378	23.11	<.0001
stdx1x2	-0.02029	0.00363	22.61696	31.25	<.0001
stdx2x4	0.01290	0.00268	16.72089	23.10	<.0001

All variables left in the model are significant at the 0.0500 level.

In conclusion, the model with p-value<0.05 was selected.

- Vel, Acc, Dec, PD, TD, stdx1x2 and stdx2x4 (best choice), if we consider significance level at 0.1.

4.13 Potential "Best" Models

Summarizing the results obtained from the three methods of model search, we conclude with the following best models.

Model A: Vel, PD, TD, stdx1x2, stdx2x3 and stdx2x4 (1st choice)

Model B: Vel, Acc, PD, TD, stdx1x3 and stdx2x4 (2nd choice)

4.13.1 Building the Regression Function for the Highway Data

MODEL A

The regression function for Model A can be built from Table 4.32 below.

Table 4.32 SAS Output for Parameter Estimate for Model A

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	1.03429	0.02212	46.75	<.0001	0
Vel	Velocity	1	0.02484	0.00031102	79.86	<.0001	1.33586
PD	Power demand	1	0.01404	0.00014740	95.27	<.0001	1.64933
TD	Time of Day	1	0.03581	0.00743	4.82	<.0001	1.05061
stdx1x2		1	-0.02184	0.00211	-10.37	<.0001	1.30289
stdx2x3		1	0.22545	0.00905	24.92	<.0001	2.51043
stdx2x4		1	0.01335	0.00254	5.25	<.0001	2.22141

The regression function is:

$$\hat{y} = 1.034 + 0.025x_1 + 0.014x_4 + 0.036x_5 - 0.022x_1x_2 + 0.225x_2x_3 + 0.013x_2x_4$$

Eqn.4.6

Where

$\hat{y} = CO_2$, (gram/sec)

x_1 = Vehicle Velocity, V,(mi/hr)

x_2 = Vehicle Acceleration, Acc,(mi/hr/sec)

x_3 = Vehicle Deceleration, Dec,(mi/hr/sec)

x_4 = Power Demand, PD, ($mi^2/hr^2/sec$)

x_5 = Time of Day, TD, (1 for peak, 0 for offpeak)

From the correlation matrix in Table 4.33 below and the scatter plot matrix in Figure 4.41 below, a strong correlation does not exist between the predictor variables.

Table 4.33 Correlation between Predictor Variables for Model A

Pearson Correlation Coefficients, N = 55342

	Co2	Vel	PD	TD	stdx1x2	stdx2x3	stdx2x4
Co2 CO2	1.00000	0.26077	0.41598	-0.04149	-0.07081	-0.09019	0.19121
Vel Velocity	0.26077	1.00000	0.00146	-0.21521	0.29990	-0.37980	-0.15311
PD Power demand	0.41598	0.00146	1.00000	-0.01157	-0.16573	-0.30674	0.26268
TD Time of Day	-0.04149	-0.21521	-0.01157	1.00000	-0.04410	0.10618	0.02822
stdx1x2	-0.07081	0.29990	-0.16573	-0.04410	1.00000	-0.29423	-0.40888
stdx2x3	-0.09019	-0.37980	-0.30674	0.10618	-0.29423	1.00000	0.55079
stdx2x4	0.19121	-0.15311	0.26268	0.02822	-0.40888	0.55079	1.00000

.

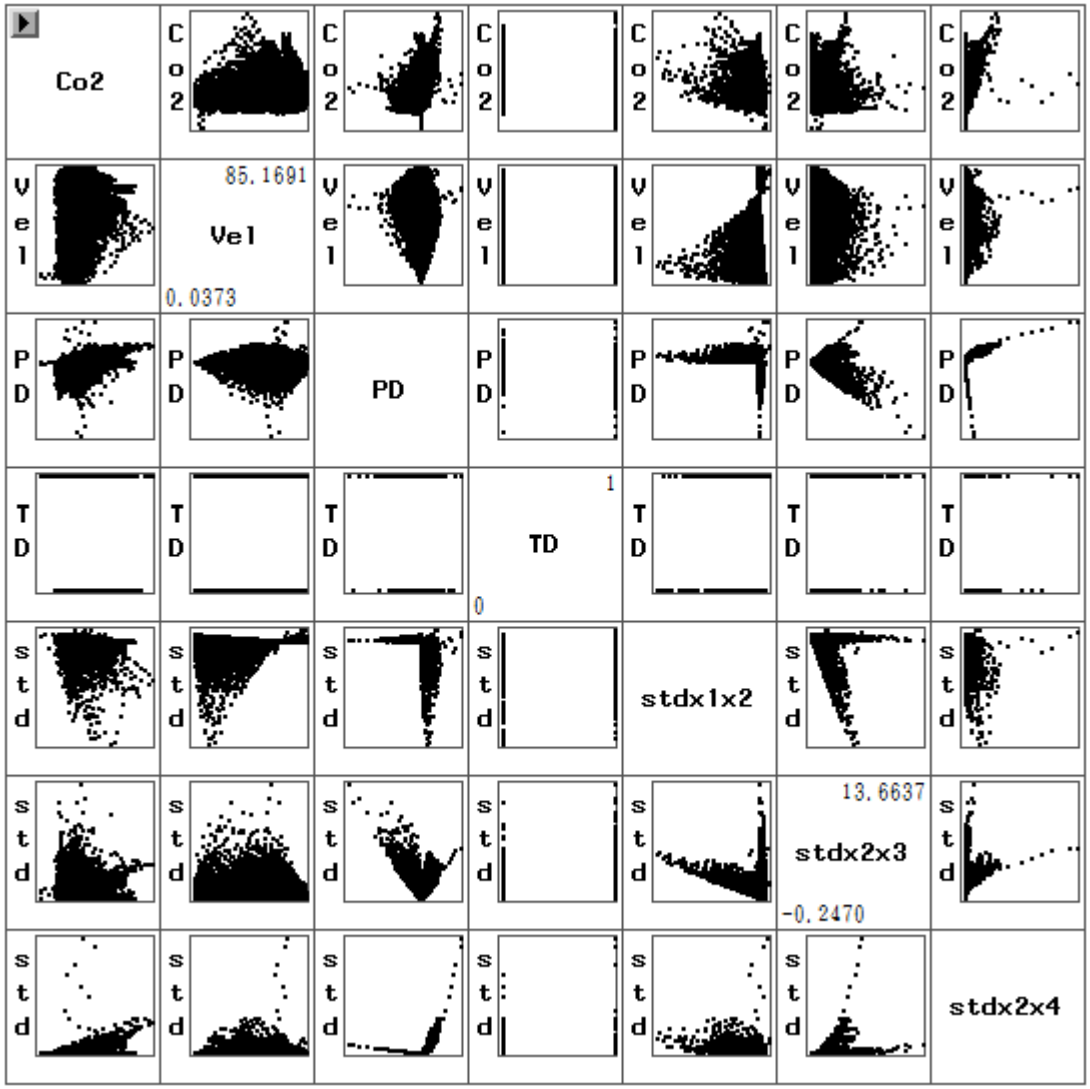


Figure 4.41 Scatter Plot Matrix of Response and Predictor Variables

This model compared to the preliminary model shows an improvement in terms of linear relationship between response variable and predictor variables and less multicolinearity between predictor variables.

MODEL B

The regression function for Model B can be built from Table 4.34 below.

Table 4.34 SAS Output for Parameter Estimate for Model B

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97819	0.02312	42.30	<.0001	0
Vel	Velocity	1	0.02482	0.00031175	79.62	<.0001	1.34214
Acc	Acceleration	1	0.63296	0.02439	25.95	<.0001	4.39021
PD	Power demand	1	0.00915	0.00017108	53.47	<.0001	2.22180
TD	Time of Day	1	0.03560	0.00742	4.80	<.0001	1.04940
stdx1x3		1	0.02617	0.00207	12.67	<.0001	1.15850
stdx2x4		1	0.01272	0.00263	4.83	<.0001	2.38311

The regression function is:

$$\boxed{\hat{y} = 0.978 + 0.025x_1 + 0.633x_2 + 0.009x_4 + 0.036x_5 + 0.026x_1x_3 + 0.013x_2x_4} \qquad \text{Eqn.4.7}$$

Where

$\hat{y} = CO_2$ (gram/sec)

x_1 = Vehicle Velocity, V,(mi/hr)

x_2 = Vehicle Acceleration, Acc.(mi/hr/sec)

x_3 = Vehicle Deceleration, Dec.(mi/hr/sec)

x_4 = Power Demand, PD, ($mi^2/hr^2/sec$)

x_5 = Time of Day, TD,

From the correlation matrix in Table 4.32 above and the scatter plot matrix in Figure 4.42 below, there is no high correlation or multicollinearity between predictor variables.

Table 4.35 Correlation between Predictor Variables for Model B

Pearson Correlation Coefficients, N = 55342

	Co2	Vel	Acc	PD	TD	stdx1x3	stdx2x4
Co2 CO2	1.00000	0.26077	0.31009	0.41598	-0.04149	-0.12862	0.19121
Vel Velocity	0.26077	1.00000	-0.27262	0.00146	-0.21521	-0.26821	-0.15311
Acc Acceleration	0.31009	-0.27262	1.00000	0.65491	0.06141	-0.12759	0.70353
PD Power demand	0.41598	0.00146	0.65491	1.00000	-0.01157	-0.22065	0.26268
TD Time of Day	-0.04149	-0.21521	0.06141	-0.01157	1.00000	0.06652	0.02822
stdx1x3	-0.12862	-0.26821	-0.12759	-0.22065	0.06652	1.00000	-0.00771
stdx2x4	0.19121	-0.15311	0.70353	0.26268	0.02822	-0.00771	1.00000

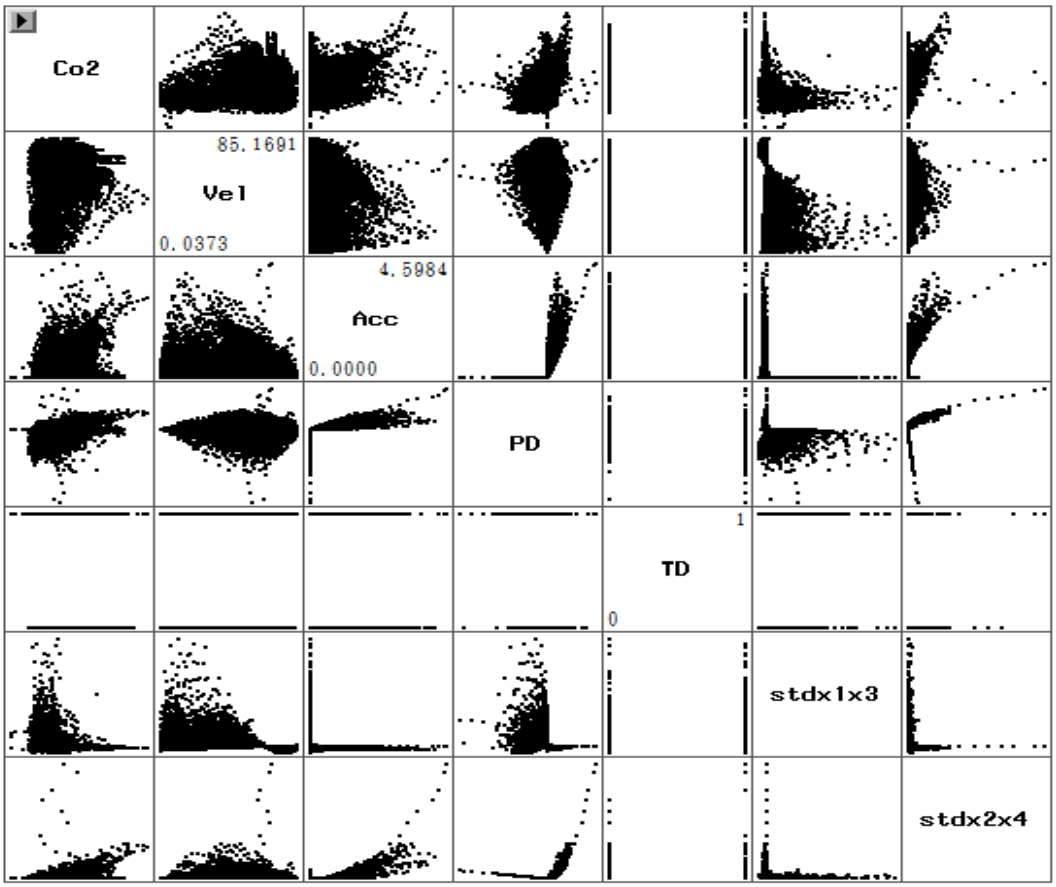


Figure 4.42 Scatter Plot Matrix of Response and Predictor Variables

The best model is selected from the two models above. To do this, various tests including diagnostic, normality and constant variance tests for the two models are conducted and make sure they satisfy them. The model that satisfies all the tests is selected as the **"**best" overall model.

4.14 Verifying Model Assumptions

Table 4.36 below compares Model A with model B

Table 4.36 Model Assumption Verification

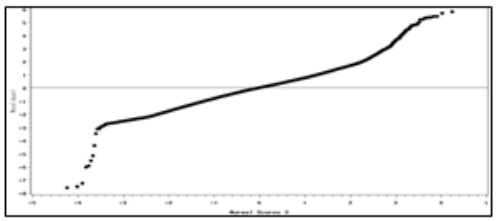
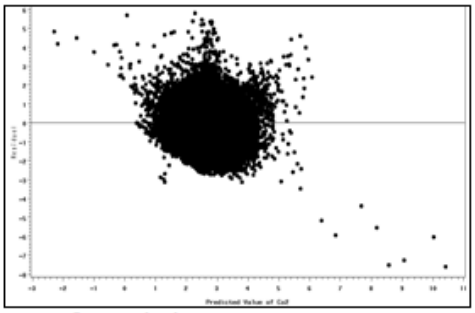
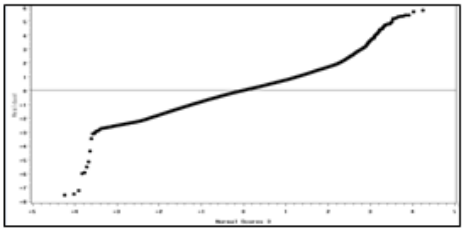
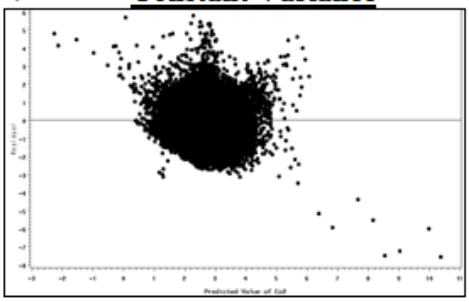
1. VERIFY ASSUMPTIONS: ❖ **Normality(MODEL A)**  Figure is straight enough $\Rightarrow$ Normality **OK** $\hat{\rho}$ **=0.99 (SAS output)** $c(\alpha = 0.10, n = 55342) = 0.989$ $\hat{\rho}$ **>0.989 $\Rightarrow$ Normality OK** ❖ **Constant Variance** 	**1.VERIFY ASSUMPTIONS:** ❖ **Normality(MODEL B)**  The figure is straight $\Rightarrow$ Normality **OK** $\hat{\rho}$ **=0.99 (SAS output)** $c(\alpha = 0.10, n = 55342) = 0.989$ $\hat{\rho}$ **>0.989 $\Rightarrow$ Normality OK** ❖ **Constant Variance** 

Table 4.36-continued

<table>
<tr>
<td>Not funnel shape, variance is constant. Using the modified Levene test with two groups n_1=27172 and n_2=28170, we have the following SAS output:

<table>
<tr><th>Variances</th><th>DF</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Equal</td><td>55E3</td><td>14.29</td><td><.2795</td></tr>
</table>
For $\alpha = 0.10 \Rightarrow p-value > 0.10 \Rightarrow$ fail to reject Ho, meaning constant variance (same as plot conclusion).</td>
<td>Not funnel shape, variance is constant. Using the modified Levene test with two groups n_1=27259 and n_2=28083, we have the following SAS output:

<table>
<tr><th>Variances</th><th>DF</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Equal</td><td>55E3</td><td>14.15</td><td><.2785</td></tr>
</table>
For $\alpha = 0.10 \Rightarrow p-value > 0.10 \Rightarrow$ fail to reject Ho, meaning constant variance (same as plot conclusion).</td>
</tr>
<tr>
<td><u>Variance Inflation</u>
We need to verify serious multicollinearity problem with model.
<table>
<tr><th>Variable</th><th>Variance inflation</th></tr>
<tr><td>Vel</td><td>1.33586</td></tr>
<tr><td>PD</td><td>1.64933</td></tr>
<tr><td>TD</td><td>1.05061</td></tr>
<tr><td>Stdx1x2</td><td>1.30289</td></tr>
<tr><td>Stdx2x3</td><td>2.51043</td></tr>
<tr><td>Stdx2x4</td><td>2.22141</td></tr>
</table>
Since Max (VIF_k) = 2.51043<5 and AVG (VIF_k) =1.67842<5 $\Rightarrow$ Serious multicollinearity is not a problem.</td>
<td><u>Variance Inflation</u>
We need to verify serious multicollinearity problem with model.
<table>
<tr><th>Variable</th><th>Variance inflation</th></tr>
<tr><td>Vel</td><td>1.34214</td></tr>
<tr><td>Acc</td><td>4.39021</td></tr>
<tr><td>PD</td><td>2.22180</td></tr>
<tr><td>TD</td><td>1.04940</td></tr>
<tr><td>Stdx1x3</td><td>1.15850</td></tr>
<tr><td>Stdx2x4</td><td>2.38311</td></tr>
</table>
Since Max (VIF_k) = 4.39021<5 and AVG (VIF_k) = 2.09086<5 $\Rightarrow$ Serious multicollinearity is not a problem.</td>
</tr>
</table>

4.15 Testing Goodness of Fit for Each Model

Analysis of Variance (ANOVA) and t-test tables from the SAS output will be used to test the goodness of fit for each model:

Model A

Tables 4.37 and 4.38 show the SAS output of the ANOVA and T-test, respectively, for testing the goodness of fit for Model A.

Table 4.37 SAS ANOVA Output for Model A

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14629	2438.15794	3368.60	<.0001
Error	55335	40051	0.72379		
Corrected Total	55341	54680			

Root MSE	0.85076	R-Square	0.2675
Dependent Mean	2.74685	Adj R-Sq	0.2675
Coeff Var	30.97211		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	1.03429	0.02212	46.75	<.0001	0
Vel	Velocity	1	0.02484	0.00031102	79.86	<.0001	1.33586
PD	Power demand	1	0.01404	0.00014740	95.27	<.0001	1.64933
TD	Time of Day	1	0.03581	0.00743	4.82	<.0001	1.05061
stdx1x2		1	-0.02184	0.00211	-10.37	<.0001	1.30289
stdx2x3		1	0.22545	0.00905	24.92	<.0001	2.51043
stdx2x4		1	0.01335	0.00254	5.25	<.0001	2.22141

Table 4.38 SAS T-test Output for Model A

The TTEST Procedure

Statistics

Variable	group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
d	1	27172	0.6756	0.6823	0.6889	0.5541	0.5588	0.5635	0.0034
d	2	28170	0.6091	0.6155	0.6218	0.5371	0.5415	0.546	0.0032
d	Diff (1-2)		0.0577	0.0668	0.076	0.5468	0.55	0.5533	0.0047

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	55E3	14.29	<.2795
d	Satterthwaite	Unequal	55E3	14.28	<.2795

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
d	Folded F	27171	28169	1.06	<.1050

Model B

Tables 4.39 and 4.40 show the SAS output of the ANOVA and T-test, respectively, for testing the goodness of fit for Model B.

Table 4.39 SAS ANOVA Output for Model B

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14629	2438.17726	3368.63	<.0001
Error	55335	40051	0.72379		
Corrected Total	55341	54680			

Root MSE	0.85076	R-Square	0.2675
Dependent Mean	2.74685	Adj R-Sq	0.2675
Coeff Var	30.97207		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	0.97819	0.02312	42.30	<.0001	0
Vel	Velocity	1	0.02482	0.00031175	79.62	<.0001	1.34214
Acc	Acceleration	1	0.63296	0.02439	25.95	<.0001	4.39021
PD	Power demand	1	0.00915	0.00017108	53.47	<.0001	2.22180
TD	Time of Day	1	0.03560	0.00742	4.80	<.0001	1.04940
stdx1x3		1	0.02617	0.00207	12.67	<.0001	1.15850
stdx2x4		1	0.01272	0.00263	4.83	<.0001	2.38311

Table 4.40 SAS T-test Output for Model B

The TTEST Procedure

Statistics

Variable	group	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
d	1	27259	0.6753	0.6819	0.6885	0.5537	0.5584	0.5631	0.0034
d	2	28083	0.6094	0.6157	0.6221	0.5374	0.5418	0.5463	0.0032
d	Diff (1-2)		0.057	0.0662	0.0753	0.5468	0.55	0.5533	0.0047

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
d	Pooled	Equal	55E3	14.15	<.2785
d	Satterthwaite	Unequal	55E3	14.14	<.2785

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
d	Folded F	27258	28082	1.06	<.1042

From the ANOVA and t-test tables above**:**

❖ The regression is significant for both models. This is because the p value (P_r>F) is less than $\alpha = 0.10$.

❖ $R^2 = (SSR / SSTO) = 0.2675 \Rightarrow$ It can be concluded that 27% of total variation in the mass emission rate of CO_2 is explained by the introduction of velocity, power demand, time of day, (velocity*acceleration), (acceleration*deceleration) and (acceleration*power demand) in the model. The model does not appear to fit that well since the value is so low ($0 \le R^2 \le 1$). The remaining 73% is not accounted for, or unexplained.

❖ Similarly for model B, $R^2 = (SSR / SSTO) = 0.2675 \Rightarrow$ It can be concluded that 27% of total variation in the mass emission rate of CO_2 is explained by the introduction of velocity, acceleration, power demand, time of day, (acceleration*deceleration) and (acceleration*power demand) in the model. The model does not appear to fit that well since the value is so low ($0 \le R^2 \le 1$). The remaining 73% is not accounted for, or unexplained.

❖ Similarly, factors such as changes in weather conditions, road grade, and air conditioning usage may have contributed to the unaccounted portion of the dataset.

4.16 Summary of the Best Selected Model

Model A: Passed all the tests and in addition to that it is a simpler model to use compared to Model B. Moreover, Model A was selected as one of the 2 best models in two model selection methods.

Model B: Passed all tests that normality is OK. Model B was selected as one of the 2 best models in two of the three model selection methods.

From a statistical point of view, the best model is Model A.

$$\boxed{\hat{y} = 1.034 + 0.025x_1 + 0.014x_4 + 0.036x_5 - 0.022x_1x_2 + 0.225x_2x_3 + 0.013x_2x_4}$$

Eqn.4.6

4.17 Selection of Best Model, Including Factors Other Than Statistical

Statistical factors, like R^2 values and C_p values, are only one set of factors to consider in choosing a model. Other factors include complexity of use, whether it is consistent with theory, and whether it gives realistic emission estimates. These additional factors are considered in this section

4.17.1 Other Models with High R^2 values

SAS output includes only the top two models in each category (number of terms),e.g. the top two models with two terms, the top two models with three terms, etc. However, in this research, the R^2 values for the top two models in each category were only slightly higher than models ranked lower. This section presents some of the lower ranked models, which had comparable R^2 values.

Eqns.4.8 - Eqn. 4.10 (arterial models) and Eqns.4.11 - Eqn. 4.13 (highway models) were selected from the other SAS output models for the following reasons:

(i) They are simple models.

(ii) The independent variables present are velocity, acceleration and velocity* acceleration. These variables are representative of variables found in other fuel consumption models such as Eqn.2.5,

and in practice, influence fuel consumption rates as well as exhaust emissions rate.

4.17.1.1 Arterial Models

$$\boxed{\hat{y} = 1.007 + 0.010Vel + 0.982Acc + 0.006\text{PD}}$$ R^2= 0.6094 Eqn. 4.8

$$\boxed{\hat{y} = 0.867 + 0.011Vel + 1.172Acc + 0.208\text{Vel} * \text{Acc}}$$ R^2=0.6050 Eqn.4.9

$$\boxed{\hat{y} = 0.902 + 0.011Vel + 1.145Acc}$$ R^2 = 0.5967 Eqn.4.10

$$\boxed{\hat{y} = 0.982 + 0.008Vel + 0.923Acc + 0.006PD + 0.150TD - 0.128Vel * Acc + 0.213Vel * PD}$$

R^2= 0.6190 Eqn.4.3

4.17.1.2 Highway Models

$$\boxed{\hat{y} = 1.071 + 0.024Vel + 0.702Acc + 0.009\text{PD}}$$ R^2 = 0.2646 Eqn. 4.11

$$\boxed{\hat{y} = 0.707 + 0.028Vel + 1.318Acc}$$ R^2 = 0.2250 Eqn. 4.12

$$\boxed{\hat{y} = 0.765 + 0.026Vel + 1.538Acc + 0.060\text{Vel} * \text{Acc}}$$ R^2= 0.2650 Eqn. 4.13

$$\boxed{\hat{y} = 1.034 + 0.025Vel + 0.014PD + 0.036TD - 0.022Vel * Acc + 0.225Ac * Dec + 0.013Acc * PD_4}$$

R^2= 0.2675 Eqn.4.6

4.17.2 Estimation of CO_2 Emissions from considered Arterial and Highway Models

This section presents case studies of CO_2 emissions calculated using different models.

Table 4.41 CO_2 Emissions Using Actual Values in Arterial Models, Acceleration=0 (Cruise Mode)

CO_2 emissions (gram/sec)								
Acc=0								
	5mi/hr	10mi/hr	15mi/hr	20mi/hr	25mi/hr	30mi/hr	35mi/hr	40mi/hr
Eqn.4.8	1.06	1.11	1.16	1.21	1.26	1.31	1.36	1.41
Eqn.4.9	0.92	0.98	1.03	1.09	1.14	1.20	1.25	1.31
Eqn.4.10	0.96	1.01	1.07	1.12	1.18	1.23	1.29	1.34
Eqn.4.3	1.10	1.14	1.18	1.22	1.26	1.30	1.34	1.38

Table 4.42 CO_2 Emissions Using Actual Values in Arterial Models, Acceleration=1mi/hr/sec

CO_2 emissions (gram/sec)								
Acc=1mi/hr/sec								
	5mi/hr	10mi/hr	15mi/hr	20mi/hr	25mi/hr	30mi/hr	35mi/hr	40mi/hr
Eqn.4.8	2.07	2.15	2.23	2.31	2.39	2.47	2.55	2.63
Eqn.4.9	3.13	4.23	5.32	6.42	7.51	8.61	9.70	10.80
Eqn.4.10	2.10	2.16	2.21	2.27	2.32	2.38	2.43	2.49
Eqn.4.3	6.74	22.15	48.20	84.91	132.26	190.27	258.92	338.23

Table 4.43 CO_2 Emissions Using Actual Values in Arterial Models, Acceleration=2mi/hr/sec

CO_2 emissions (gram/sec)								
Acc=2mi/hr/sec								
	5mi/hr	10mi/hr	15mi/hr	20mi/hr	25mi/hr	30mi/hr	35mi/hr	40mi/hr
Eqn.4.8	3.08	3.19	3.30	3.41	3.52	3.63	3.74	3.85
Eqn.4.9	5.35	7.48	9.62	11.75	13.89	16.02	18.16	20.29
Eqn.4.10	3.25	3.30	3.36	3.41	3.47	3.52	3.58	3.63
Eqn.4.3	12.39	43.16	95.23	168.60	263.27	379.24	516.51	675.08

Table 4.44 CO_2 Emissions Using Actual Values in Highway Models, Acceleration=0 (Cruise Mode)

CO_2 emissions (gram/sec)								
Acc=0								
	45mi/hr	50mi/hr	55mi/hr	60mi/hr	65mi/hr	70mi/hr	75mi/hr	80mi/hr
Eqn.4.11	2.15	2.27	2.39	2.51	2.63	2.75	2.87	2.99
Eqn.4.12	1.97	2.11	2.25	2.39	2.53	2.67	2.81	2.95
Eqn.4.6	2.18	2.30	2.43	2.55	2.68	2.80	2.93	3.05
Eqn.4.13	1.94	2.07	2.20	2.33	2.46	2.59	2.72	2.85

Table 4.45 CO_2 Emissions Using Actual Values in Highway Models, Acceleration=1mi/hr/sec

CO_2 emissions (gram/sec)								
Acc=1mi/hr/sec								
	45mi/hr	50mi/hr	55mi/hr	60mi/hr	65mi/hr	70mi/hr	75mi/hr	80mi/hr
Eqn.4.11	3.26	3.42	3.59	3.75	3.92	4.08	4.25	4.41
Eqn.4.12	3.29	3.43	3.57	3.71	3.85	3.99	4.13	4.27
Eqn.4.6	2.15	2.30	2.45	2.60	2.75	2.90	3.05	3.20
Eqn.4.13	6.17	6.60	7.03	7.46	7.89	8.32	8.75	9.18

Table 4.46 CO_2 Emissions Using Actual Values in Highway Models, Acceleration=2mi/hr/sec

CO_2 emissions (gram/sec)								
Acc=2mi/hr/sec								
	45mi/hr	50mi/hr	55mi/hr	60mi/hr	65mi/hr	70mi/hr	75mi/hr	80mi/hr
Eqn.4.11	4.37	4.58	4.79	5.00	5.21	5.42	5.63	5.84
Eqn.4.12	4.60	4.74	4.88	5.02	5.16	5.30	5.44	5.58
Eqn.4.6	3.29	3.59	3.90	4.20	4.51	4.81	5.12	5.42
Eqn.4.13	10.41	11.14	11.87	12.60	13.33	14.06	14.79	15.52

The trends in Tables 4.41-4.46 are what would be expected, in that emissions increase across each table as velocity increases. This is due to the fact that an increase in speed causes more fuel to be burned, correspondingly producing more tailpipe

emissions including carbon dioxide emissions. In addition, emissions predicted by a given equation for a given velocity increase. An increase in the change in velocity per unit time, acceleration, puts more load on the engine and requires more fuel to be burned producing more carbon dioxide emissions. For the cruise cases (acceleration = 0), emissions predicted by all 4 models at a given speed are very similar.

Emissions estimated for acceleration = 1 mi/hr/sec and 2 mi/hr/sec using Eq. 4.3 are very high compared with the estimates given by the other equations. Eq. 4.3 likely overestimates emissions at higher speeds, for acceleration = 1 mi/hr/sec and 2 mi/hr/sec, according to a check run using the carbon equivalent of gasoline published in the Code of Federal Regulations (40 CFR 600.113) and used by the Environmental Protection Agency in fuel economy calculations.

As speeds increase from 40 mi/hr arterial to 45 mi/hr highway, discontinuities exist between Tables 4.42 (arterial) and 4.45 (highway), and Tables 4.43 (arterial), and 4.46(highway). In most cases, emissions calculated by a given equation for 45 mph highway are greater than emissions estimated for 40 mph by the corresponding arterial equation. The discontinuity issue may be due to the fact that the highway model was built with velocity values less than 45mile/hr inclusive. The discontinuities do not occur for the cruise cases (acceleration = 0).

4.17.3 Final Choice of Models

From a statistical stand point, the models selected were Eqn. 4.3 (for arterial) and Eqn.4.6 (for highway), taking into account the model selection process. Factors such as R^2, Adjusted R^2, and C_p closer to P were considered. As explained earlier, the model with the highest R^2 value, highest adjusted R^2 value and C_p closest to P was selected as the best model from a statistical viewpoint, after passing all the models' assumption tests. Eqn. 4.3 & Eqn. 4.6 satisfied all these conditions and hence ended up as the selected models.

However, these models contain a lot of variables and interaction terms, which makes them complex. In addition, the negative coefficient in front of the interaction term, velocity* acceleration, which occurs in both equations, is not consistent with theory and practice. Theoretical models developed for estimating fuel consumption or emissions show positive coefficient values for velocity and acceleration variables, which indicates that emissions would increase with increase in these variables. This is consistent with what real world situations depict: as velocity and acceleration increase, fuel consumption and emissions also increase. In most emissions and fuel consumption models, such as Eqn. 4.13 below, velocity and acceleration are two important variables responsible for vehicle emissions increase.

Therefore, a model containing velocity acceleration and velocity*acceleration with all the terms positive will be plausible. Further the model should be simple.

Table 4.47 summarizes advantages and disadvantages associated with the various arterial models.

Table 4.47 Advantages and Disadvantages Associated with the Arterial Models

Equation	Number of Independent Variables/Variable Groups	R^2	Simple or Complex?	Fuel Economy basis?	Emission Estimates Realistic?
4.8	3	0.6094	Simple	No	Don't vary much with speed
4.9	3	0.6050	Simple	Yes	Seem reasonable
4.10	2	0.5968	Simple	Yes	Don't vary much with speed
4.3	6	0.6190	Complex	No, and interaction term has negative coefficient	Likely too high

For the arterial models, Eqn.4.3 (R^2=0.6190) has a higher R^2 value, but it likely over predicted emissions. In addition, due to the greater number of independent variables, it is more complex to use. Eqn.4.9 (R^2 = 0.6050) has about the same R^2 value as Eqn. 4.3, contains fewer variables, and has a basis in fuel economy models. In Tables 4.41 and 4.42, Eq. 4.8 and 4.10 do not predict much of an emissions increase with velocity, as would be expected. Eqn.4. 9 was hence selected as the final arterial model.

Table 4.48 summarizes advantages and disadvantages associated with the various highway models.

Table 4.48 Advantages and Disadvantages Associated with the Highway Models

Equation	Number of Independent Variables/Variable Groups	R^2	Simple or Complex?	Fuel Economy basis?	Emission Estimates Realistic?
4.11	3	0.2646	Simple	No	Yes
4.12	2	0.2250	Simple	Yes	Yes
4.6	6	0.2675	Complex	No	Yes
4.13	3	0.2650	Simple	Yes	Yes

Of the highway models, Eqn.4.6, due to the greater number of independent variables, is more complex to use. Eqn. 4.13 is a simpler model containing velocity, acceleration and velocity*acceleration, which are important independent variables present in most fuel consumption models such as the one in Eqn. 2.5. From Table 4.48, Eqn. 4.11 does not have a basis in fuel consumption models while the R^2 value of Eqn. 4.12 is lower. Therefore Eqn. 4.13 was selected as the final highway model.

Eqn.4.9 & Eqn. 4.13, which are much simpler and contains variables such as velocity and acceleration that are representative of vehicle emissions estimation, are selected as the final models for the arterial and highway datasets, respectively. These models have R^2 and adjusted R^2values almost as high as the models previously selected

purely on statistical rationale, in addition, the coefficients of all variables and interaction terms are positive.

Similarly, it should be noted that, the chosen arterial model can only be used over velocity ranges of 0 to 54 mile per hour, acceleration ranges of 0 to 4.9 mile per hour per second, and power demand ranges of 0 to 119 mi^2per hr^2 per sec. These ranges represent the range of values used in developing the model.

61% of total variation in the mass emission rate of CO_2 is explained by the chosen arterial model.

The remaining 39%, which was not accounted for in the model may be due to factors such as road grade, weather conditions, air conditioning usage, tire pressure (needed to be checked and made sure the amount of air in tire is consistent throughout the data collection period), road surface conditions and total vehicle weight.

Similarly, the chosen highway model can only be used over velocity ranges of 0 to 85 mile per hour, acceleration ranges of 0 to 4.5 mile per hour per second, and power demand ranges of 0 to 315 mi^2per hr^2 per sec. These ranges represent the range of dataset used in developing the model.

27% of total variation in the mass emission rate of CO_2 is explained by the highway model.

The remaining 73% is not accounted for, or unexplained. This unexplained part of the R^2 may be due to factors such as road grade, weather conditions, air conditioning usage, tire pressure (needed to be checked and made sure the amount of air in tire is consistent throughout the data collection period), road surface conditions and total vehicle weight

(which may have changed due to different drivers and passengers). Further, it may be due to the fact that the data used in developing the highway model contained speed data that is below 45miles per hour, which is representative of speeds found on an arterial facility, and may have represented anomalous conditions for the highway. Future research using freeway data should exclude velocity data less than 45 miles per hour.

The coefficients of both the velocity and the acceleration terms for the highway model are higher than for the arterial model. This is an indication that, on the highway facility, the changes in velocity and acceleration would produce higher CO_2 emissions compared with an arterial facility.

Unlike MOBILE 6.2 which contain vehicle velocity as the only traffic variable, the chosen arterial and highway models are functions of both velocity and acceleration and could better explain the dynamics associated with driving.

The models can be coupled with a micro scale traffic model to evaluate emissions impact of transportation system measures such as:

- traffic signal coordination
- speed limit reductions
- extended idling reductions
- bottleneck removal.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

A Multiple Linear Regression method which incorporates Statistical Analysis System (SAS) software was used to develop CO_2 emissions models for a light duty gasoline vehicle. This statistical method was used as a tool to identify useful explanatory variables.

5.1.1 Final Carbon Dioxide (CO_2) Models

The models to estimate carbon dioxide emissions on a second-by-second basis are:

1. Arterial model:

$$\boxed{\hat{y} = 0.867 + 0.011Vel + 1.172Acc + 0.208\text{Vel} * \text{Acc}}$$

2. Highway model:

$$\boxed{\hat{y} = 0.765 + 0.026Vel + 1.538Acc + 0.060\text{Vel} * \text{Acc}}$$

Where

$\hat{y} = CO_2$ (grams/sec),

Vel = Vehicle Velocity,(mi/hr),

Acc= Vehicle Acceleration,(mi/hr/sec),

Given the large set of second by second data collected, the number of variables examined and the surrogate nature of the variables, the overall R^2 value of 0.61 for CO_2 demonstrated a strong model fit to the arterial dataset, whereas the overall R^2 value of 0.27 for CO_2 demonstrated a weak model fit to the highway dataset. However, all final model variables are statistically significant as indicated by the P-values from the SAS output in the tables.

Further, unlike other models that predict vehicle emissions with velocity as the sole vehicle activity variable responsible for higher emissions, these models developed contain an additional variable, acceleration, which accounts for the dynamics involved in vehicle activity.

5.2 Recommendations

- On – board emissions measurement methods should be used in future studies to improve knowledge of real-world on-road tailpipe emissions.

- On-road measurement studies should be carefully designed, taking into account well-defined study objectives.

- Variability and uncertainty should be accounted for in the study design and when making inferences based upon measured data. That is, driving behavior during data collection should be as consistent as possible; in addition, variation in road grade, air conditioning usage and weather conditions under which data is collected should be

minimized. The impact, however, of accounting for these variability and uncertainty could lead to an improvement of the coefficient of determination (R^2).

- UTA and others should use on-road emissions data to aid in the design and evaluation of transportation control measures (TCMs) and transportation implementation programs (TIPs).

- Further research can be conducted by developing CO and HC models, as well as testing diesel vehicles.

APPENDIX A

ARTERIAL SAS OUTPUT FOR X & Y OUTLIERS DETERMINATION

Output Statistics

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	-2.1845	-2.8654	0.0007	0.9999	-0.0737
2	-2.3474	-3.0792	0.0007	0.9998	-0.0841
3	-1.8662	-2.4477	0.0005	1.0000	-0.0567
4	-1.8755	-2.4602	0.0008	1.0003	-0.0707
5	-0.4088	-0.5362	0.0006	1.0006	-0.0127
6	-1.1803	-1.5481	0.0005	1.0004	-0.0357
7	-1.3338	-1.7493	0.0005	1.0003	-0.0406
8	-1.3338	-1.7493	0.0005	1.0003	-0.0406
9	-1.6415	-2.1529	0.0005	1.0001	-0.0480
10	-0.4731	-0.6205	0.0006	1.0006	-0.0146
11	-1.6941	-2.2220	0.0006	1.0002	-0.0544
12	-0.1005	-0.1319	0.0005	1.0006	-0.0030
13	-2.5165	-3.3007	0.0005	0.9994	-0.0711
14	-0.4241	-0.5562	0.0004	1.0005	-0.0117
15	-0.5667	-0.7432	0.0005	1.0005	-0.0161
16	-0.3658	-0.4798	0.0005	1.0006	-0.0105
17	-1.4331	-1.8795	0.0005	1.0002	-0.0400
18	-0.0514	-0.0675	0.0004	1.0005	-0.0014
19	-0.2973	-0.3899	0.0005	1.0006	-0.0086
20	-0.3532	-0.4633	0.0004	1.0005	-0.0098
21	-1.7178	-2.2530	0.0004	1.0000	-0.0468
22	-0.7036	-0.9228	0.0005	1.0005	-0.0203
23	-1.3722	-1.7996	0.0004	1.0002	-0.0367
24	-1.3722	-1.7996	0.0004	1.0002	-0.0367
25	-0.0522	-0.0684	0.0005	1.0006	-0.0015
26	-0.6311	-0.8277	0.0005	1.0006	-0.0188
27	-3.4641	-4.5441	0.0005	0.9984	-0.1057
28	-0.3584	-0.4700	0.0005	1.0005	-0.0101
29	-1.5140	-1.9856	0.0004	1.0001	-0.0419
30	-0.6736	-0.8834	0.0005	1.0006	-0.0206
31	-0.6736	-0.8834	0.0005	1.0006	-0.0206
32	-0.7496	-0.9831	0.0004	1.0004	-0.0201
33	0.2667	0.3498	0.0004	1.0005	0.0074
34	-1.8331	-2.4044	0.0006	1.0001	-0.0610
35	-0.4451	-0.5837	0.0005	1.0005	-0.0125
36	-2.0493	-2.6877	0.0005	0.9998	-0.0580
37	-0.4661	-0.6113	0.0005	1.0005	-0.0134
38	-1.1558	-1.5158	0.0005	1.0003	-0.0330

Output Statistics

Obs	DFBETAS Intercept	Vel	Acc	Dec	PD	TD
1	0.0156	0.0016	-0.0627	-0.0126	0.0255	-0.0113
2	0.0138	0.0050	-0.0767	-0.0249	0.0413	-0.0122
3	0.0145	-0.0010	-0.0416	-0.0003	0.0082	-0.0096
4	0.0079	0.0069	-0.0666	-0.0285	0.0429	-0.0099
5	0.0028	0.0003	-0.0104	-0.0017	0.0037	-0.0021
6	0.0113	-0.0034	-0.0164	0.0100	-0.0077	-0.0060
7	0.0090	0.0007	-0.0327	-0.0046	0.0109	-0.0069
8	0.0090	0.0007	-0.0327	-0.0046	0.0109	-0.0069
9	0.0132	-0.0019	-0.0314	0.0041	0.0014	-0.0084
10	0.0027	0.0007	-0.0125	-0.0029	0.0054	-0.0025
11	0.0083	0.0040	-0.0488	-0.0156	0.0256	-0.0089
12	0.0006	0.0001	-0.0025	-0.0006	0.0011	-0.0005
13	0.0166	0.0004	-0.0535	-0.0033	0.0131	-0.0130
14	0.0029	-0.0001	-0.0085	-0.0001	0.0016	-0.0022
15	0.0035	0.0004	-0.0127	-0.0016	0.0040	-0.0029
16	0.0021	0.0004	-0.0085	-0.0015	0.0031	-0.0019
17	0.0088	0.0008	-0.0311	-0.0034	0.0091	-0.0074
18	0.0003	-0.0000	-0.0010	0.0000	0.0002	-0.0003
19	0.0016	0.0005	-0.0072	-0.0016	0.0030	-0.0015
20	0.0021	0.0002	-0.0076	-0.0009	0.0022	-0.0018
21	0.0109	0.0003	-0.0346	-0.0017	0.0078	-0.0089
22	0.0035	0.0013	-0.0174	-0.0043	0.0077	-0.0037
23	0.0097	-0.0010	-0.0236	0.0029	0.0009	-0.0071
24	0.0097	-0.0010	-0.0236	0.0029	0.0009	-0.0071
25	0.0003	0.0001	-0.0012	-0.0002	0.0004	-0.0003
26	0.0028	0.0015	-0.0166	-0.0050	0.0083	-0.0033
27	0.0144	0.0094	-0.0947	-0.0319	0.0509	-0.0182
28	0.0019	0.0005	-0.0083	-0.0016	0.0032	-0.0019
29	0.0085	0.0015	-0.0336	-0.0053	0.0115	-0.0079
30	0.0027	0.0019	-0.0185	-0.0063	0.0101	-0.0035
31	0.0027	0.0019	-0.0185	-0.0063	0.0101	-0.0035
32	0.0047	0.0001	-0.0147	-0.0005	0.0031	-0.0039
33	-0.0014	-0.0003	0.0060	0.0011	-0.0023	0.0014
34	0.0053	0.0076	-0.0568	-0.0250	0.0368	-0.0097
35	0.0022	0.0007	-0.0105	-0.0024	0.0044	-0.0023
36	0.0100	0.0036	-0.0491	-0.0118	0.0212	-0.0107
37	0.0021	0.0010	-0.0116	-0.0032	0.0054	-0.0024

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
1531	1.0433	1.3681	0.0001	1.0000	0.0164
1532	0.2761	0.3620	0.0003	1.0004	0.0064
1533	0.4933	0.6469	0.0002	1.0002	0.0079
1534	1.2206	1.6007	0.0003	1.0001	0.0288
1535	-0.7807	-1.0237	0.0003	1.0003	-0.0180
1536	-0.7948	-1.0421	0.0001	1.0001	-0.0124
1537	-0.3532	-0.4632	0.0002	1.0003	-0.0067
1538	0.7213	0.9459	0.0004	1.0004	0.0194
1539	-0.7825	-1.0260	0.0001	1.0001	-0.0122
1540	0.7615	0.9985	0.0001	1.0001	0.0119
1541	0.9935	1.3028	0.0001	1.0001	0.0155
1542	0.0158	0.0207	0.0002	1.0003	0.0003
1543	-0.0597	-0.0782	0.0002	1.0003	-0.0011
1544	0.0425	0.0558	0.0003	1.0004	0.0010
1545	-0.1646	-0.2159	0.0003	1.0004	-0.0037
1546	0.9665	1.2673	0.0001	1.0001	0.0152
1547	0.4822	0.6323	0.0001	1.0002	0.0076
1548	1.2054	1.5807	0.0001	1.0000	0.0192
1549	1.6175	2.1210	0.0002	0.9998	0.0260
1550	-0.8314	-1.0902	0.0002	1.0001	-0.0142
1551	-1.7452	-2.2885	0.0002	0.9997	-0.0281
1552	-0.3478	-0.4560	0.0002	1.0003	-0.0058
1553	0.0546	0.0716	0.0002	1.0003	0.0009
1554	-0.1870	-0.2452	0.0001	1.0002	-0.0029
1555	-0.4385	-0.5750	0.0003	1.0004	-0.0101
1556	1.4771	1.9369	0.0002	0.9998	0.0240
1557	-0.5728	-0.7511	0.0001	1.0002	-0.0089
1558	-1.2158	-1.5943	0.0001	1.0000	-0.0189
1559	0.4543	0.5957	0.0001	1.0002	0.0071
1560	-0.0215	-0.0282	0.0002	1.0003	-0.0004
1561	-1.4754	-1.9347	0.0002	0.9999	-0.0299
1562	-0.2565	-0.3363	0.0002	1.0003	-0.0051
1563	-0.5245	-0.6878	0.0002	1.0002	-0.0090
1564	0.4336	0.5686	0.0002	1.0003	0.0088
1565	0.4575	0.5999	0.0002	1.0002	0.0074
1566	0.1613	0.2115	0.0002	1.0003	0.0029
1567	-0.8372	-1.0978	0.0002	1.0001	-0.0136
1568	1.4974	1.9636	0.0001	0.9998	0.0231
1569	-0.3481	-0.4564	0.0001	1.0002	-0.0055
1570	1.0719	1.4056	0.0001	1.0000	0.0169

	DFBETAS					
Obs	Intercept	Vel	Acc	Dec	PD	TD
1531	-0.0020	-0.0017	0.0102	0.0009	-0.0017	0.0055
1532	0.0007	-0.0020	0.0054	0.0034	-0.0043	0.0015
1533	-0.0007	-0.0011	0.0054	0.0011	-0.0016	0.0026
1534	0.0034	-0.0092	0.0245	0.0157	-0.0197	0.0066
1535	-0.0020	0.0057	-0.0153	-0.0097	0.0121	-0.0042
1536	0.0020	0.0007	-0.0066	0.0006	-0.0003	-0.0041
1537	0.0020	-0.0011	-0.0005	0.0032	-0.0036	-0.0018
1538	-0.0066	0.0054	-0.0044	-0.0128	0.0150	0.0036
1539	0.0019	0.0007	-0.0065	0.0006	-0.0002	-0.0041
1540	-0.0024	-0.0001	0.0053	-0.0018	0.0017	0.0039
1541	-0.0028	-0.0004	0.0075	-0.0017	0.0015	0.0052
1542	0.0000	-0.0001	0.0003	0.0001	-0.0002	0.0001
1543	-0.0000	0.0003	-0.0008	-0.0004	0.0005	-0.0003
1544	0.0001	-0.0003	0.0008	0.0005	-0.0006	0.0002
1545	-0.0004	0.0012	-0.0032	-0.0020	0.0025	-0.0009
1546	-0.0017	-0.0017	0.0095	0.0010	-0.0018	0.0051
1547	-0.0015	-0.0001	0.0034	-0.0011	0.0010	0.0025
1548	-0.0018	-0.0026	0.0128	0.0023	-0.0035	0.0063
1549	-0.0020	-0.0039	0.0178	0.0040	-0.0057	0.0085
1550	0.0039	-0.0015	-0.0030	0.0052	-0.0058	-0.0042
1551	0.0066	-0.0013	-0.0096	0.0070	-0.0075	-0.0090
1552	0.0015	-0.0006	-0.0014	0.0020	-0.0022	-0.0018
1553	-0.0002	0.0001	0.0002	-0.0003	0.0004	0.0003
1554	0.0004	0.0003	-0.0017	-0.0001	0.0002	-0.0010
1555	-0.0012	0.0032	-0.0085	-0.0055	0.0068	-0.0024
1556	-0.0014	-0.0041	0.0170	0.0045	-0.0063	0.0078
1557	0.0017	0.0001	-0.0041	0.0012	-0.0011	-0.0030
1558	0.0024	0.0018	-0.0113	-0.0006	0.0014	-0.0063
1559	-0.0009	-0.0007	0.0043	0.0003	-0.0006	0.0024
1560	0.0000	0.0001	-0.0003	-0.0001	0.0001	-0.0001
1561	-0.0022	0.0086	-0.0249	-0.0138	0.0175	-0.0079
1562	-0.0004	0.0015	-0.0043	-0.0023	0.0030	-0.0014
1563	0.0024	-0.0010	-0.0018	0.0033	-0.0037	-0.0027
1564	-0.0028	0.0018	-0.0002	-0.0048	0.0055	0.0022
1565	-0.0018	0.0005	0.0023	-0.0021	0.0023	0.0024
1566	-0.0008	0.0004	0.0004	-0.0013	0.0014	0.0008
1567	0.0033	-0.0008	-0.0043	0.0037	-0.0040	-0.0043
1568	-0.0032	-0.0019	0.0132	0.0000	-0.0008	0.0078

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
5536	0.0550	0.0721	0.0001	1.0002	0.0006
5537	0.3725	0.4884	0.0001	1.0002	0.0045
5538	-0.006786	-0.008898	0.0001	1.0002	-0.0001
5539	-0.000335	-0.000439	0.0001	1.0002	-0.0000
5540	-0.9771	-1.2812	0.0001	1.0000	-0.0118
5541	-0.4368	-0.5728	0.0000	1.0001	-0.0038
5542	-0.5194	-0.6811	0.0001	1.0001	-0.0057
5543	-0.1499	-0.1965	0.0000	1.0002	-0.0013
5544	-1.0556	-1.3841	0.0000	0.9999	-0.0091
5545	-0.1414	-0.1855	0.0001	1.0002	-0.0016
5546	-1.0949	-1.4356	0.0000	0.9999	-0.0099
5547	0.1442	0.1891	0.0001	1.0002	0.0014
5548	0.6965	0.9133	0.0001	1.0001	0.0066
5549	0.7719	1.0121	0.0001	1.0001	0.0081
5550	-0.6950	-0.9112	0.0001	1.0001	-0.0073
5551	-0.4494	-0.5892	0.0001	1.0002	-0.0052
5552	2.6126	3.4261	0.0001	0.9989	0.0348
5553	-0.1032	-0.1353	0.0001	1.0002	-0.0011
5554	-0.6116	-0.8020	0.0001	1.0001	-0.0066
5555	-0.3749	-0.4916	0.0000	1.0001	-0.0033
5556	0.8024	1.0522	0.0001	1.0001	0.0084
5557	-0.4121	-0.5404	0.0001	1.0002	-0.0047
5558	0.9902	1.2983	0.0001	1.0000	0.0098
5559	0.6531	0.8563	0.0001	1.0001	0.0081
5560	-0.2438	-0.3197	0.0001	1.0002	-0.0025
5561	-1.2338	-1.6178	0.0000	0.9999	-0.0106
5562	-1.2338	-1.6178	0.0000	0.9999	-0.0106
5563	-0.7745	-1.0156	0.0000	1.0000	-0.0070
5564	3.7681	4.9419	0.0001	0.9974	0.0378
5565	0.3526	0.4624	0.0000	1.0001	0.0031
5566	-0.8401	-1.1016	0.0001	1.0000	-0.0090
5567	-0.4916	-0.6446	0.0001	1.0001	-0.0053
5568	-0.0518	-0.0679	0.0001	1.0002	-0.0007
5569	0.6882	0.9024	0.0000	1.0001	0.0062
5570	-1.0631	-1.3940	0.0001	1.0000	-0.0139
5571	-0.4748	-0.6225	0.0000	1.0001	-0.0042
5572	-0.4319	-0.5663	0.0001	1.0001	-0.0043

	DFBETAS					
Obs	Intercept	Vel	Acc	Dec	PD	TD
5536	-0.0001	0.0001	-0.0001	-0.0002	0.0003	0.0003
5537	-0.0015	0.0019	-0.0013	-0.0019	0.0027	0.0019
5538	0.0000	-0.0000	0.0000	0.0000	-0.0000	-0.0000
5539	-0.0000	0.0000	-0.0000	-0.0000	0.0000	-0.0000
5540	0.0040	-0.0050	0.0033	0.0048	-0.0069	-0.0049
5541	-0.0008	0.0011	-0.0005	-0.0006	-0.0000	-0.0023
5542	0.0016	-0.0020	0.0013	0.0020	-0.0031	-0.0026
5543	-0.0002	0.0003	-0.0001	-0.0002	-0.0001	-0.0008
5544	-0.0007	0.0012	-0.0003	-0.0003	-0.0014	-0.0055
5545	0.0005	-0.0006	0.0004	0.0006	-0.0009	-0.0007
5546	0.0010	-0.0010	0.0010	0.0016	-0.0036	-0.0056
5547	-0.0002	0.0002	-0.0002	-0.0003	0.0006	0.0007
5548	-0.0010	0.0012	-0.0010	-0.0015	0.0028	0.0036
5549	-0.0021	0.0026	-0.0018	-0.0027	0.0042	0.0039
5550	0.0018	-0.0022	0.0016	0.0023	-0.0037	-0.0035
5551	0.0017	-0.0021	0.0014	0.0021	-0.0030	-0.0022
5552	0.0173	-0.0229	0.0123	0.0176	-0.0148	0.0143
5553	0.0003	-0.0003	0.0002	0.0004	-0.0006	-0.0005
5554	0.0018	-0.0023	0.0016	0.0023	-0.0035	-0.0031
5555	0.0002	-0.0002	0.0003	0.0004	-0.0011	-0.0019
5556	-0.0022	0.0027	-0.0019	-0.0027	0.0044	0.0040
5557	0.0015	-0.0018	0.0012	0.0018	-0.0027	-0.0021
5558	-0.0020	0.0025	-0.0018	-0.0027	0.0046	0.0050
5559	-0.0029	0.0036	-0.0024	-0.0035	0.0049	0.0032
5560	0.0006	-0.0007	0.0005	0.0007	-0.0012	-0.0012
5561	-0.0008	0.0012	-0.0002	-0.0002	-0.0018	-0.0064
5562	-0.0008	0.0012	-0.0002	-0.0002	-0.0018	-0.0064
5563	0.0007	-0.0007	0.0007	0.0011	-0.0025	-0.0040
5564	0.0142	-0.0191	0.0095	0.0137	-0.0085	0.0202
5565	-0.0002	0.0001	-0.0002	-0.0003	0.0010	0.0018
5566	0.0025	-0.0030	0.0021	0.0031	-0.0048	-0.0042
5567	0.0015	-0.0019	0.0013	0.0019	-0.0029	-0.0025
5568	0.0003	-0.0004	0.0002	0.0003	-0.0004	-0.0003
5569	-0.0006	0.0006	-0.0007	-0.0010	0.0022	0.0035
5570	0.0052	-0.0066	0.0043	0.0061	-0.0085	-0.0053
5571	0.0002	-0.0002	0.0003	0.0005	-0.0013	-0.0024
5572	0.0010	-0.0012	0.0009	0.0013	-0.0021	-0.0022
5573	0.0005	-0.0007	0.0004	0.0006	-0.0009	-0.0007
5574	0.0008	-0.0010	0.0007	0.0010	-0.0014	-0.0011
5575	0.0005	-0.0007	0.0002	0.0003	0.0005	0.0026
5576	0.0005	-0.0007	0.0002	0.0003	0.0005	0.0026
5577	-0.0000	0.0001	0.0000	0.0000	-0.0003	-0.0007
5578	0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0001

APPENDIX B

HIGHWAY SAS OUTPUT FOR X & Y OUTLIERS DETERMINATION

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	4.9675	5.8439	0.0020	0.9984	0.2585
2	4.3564	5.1244	0.0019	0.9991	0.2219
3	4.7540	5.5926	0.0019	0.9987	0.2470
4	3.7070	4.3599	0.0017	0.9998	0.1822
5	2.7757	3.2643	0.0017	1.0006	0.1343
6	5.7900	6.8059	0.0001	0.9952	0.0607
7	5.4485	6.4040	0.0000	0.9957	0.0449
8	2.1079	2.4784	0.0014	1.0008	0.0922
9	5.3642	6.3051	0.0001	0.9959	0.0642
10	4.4649	5.2473	0.0000	0.9972	0.0362
11	2.9119	3.4224	0.0005	0.9993	0.0748
12	2.4644	2.8966	0.0006	0.9998	0.0732
13	3.3820	3.9745	0.0002	0.9986	0.0555
14	3.5000	4.1130	0.0001	0.9984	0.0399
15	3.0762	3.6150	0.0002	0.9989	0.0460
16	2.6787	3.1480	0.0003	0.9993	0.0526
17	2.4447	2.8731	0.0004	0.9996	0.0545
18	1.7866	2.0999	0.0007	1.0004	0.0564
19	2.0856	2.4512	0.0006	1.0000	0.0584
20	4.4902	5.2771	0.0001	0.9972	0.0425
21	2.4283	2.8538	0.0004	0.9996	0.0555
22	3.5291	4.1472	0.0001	0.9983	0.0323
23	2.0541	2.4142	0.0006	1.0000	0.0572
24	2.0239	2.3787	0.0006	1.0000	0.0559
25	1.9745	2.3207	0.0006	1.0001	0.0569
26	2.2671	2.6642	0.0003	0.9997	0.0482
27	2.0941	2.4611	0.0005	0.9999	0.0538
28	2.1281	2.5009	0.0003	0.9998	0.0464
29	1.8911	2.2226	0.0006	1.0002	0.0557
30	1.0425	1.2256	0.0011	1.0011	0.0410
31	2.1866	2.5696	0.0004	0.9997	0.0483
32	1.8625	2.1889	0.0004	1.0000	0.0460
33	2.2449	2.6381	0.0003	0.9996	0.0451
34	3.9606	4.6545	0.0001	0.9978	0.0352
35	1.8108	2.1282	0.0006	1.0002	0.0502
36	4.7246	5.5535	0.0003	0.9971	0.1035
37	1.7315	2.0350	0.0006	1.0002	0.0478
38	1.5366	1.8059	0.0006	1.0003	0.0431
39	2.4987	2.9362	0.0001	0.9993	0.0340
40	3.6127	4.2454	0.0001	0.9982	0.0400
41	1.7990	2.1143	0.0006	1.0002	0.0500
42	1.0474	1.2312	0.0008	1.0008	0.0354
43	1.5725	1.8481	0.0006	1.0003	0.0448
44	2.8293	3.3246	0.0001	0.9990	0.0283

	DFBETAS					
Obs	Intercept	Vel	Acc	Dec	PD	TD
1	-0.0610	0.0374	0.2065	0.0232	-0.0628	0.0151
2	-0.0587	0.0383	0.1668	0.0036	-0.0362	0.0141
3	-0.0496	0.0272	0.2072	0.0402	-0.0791	0.0135
4	-0.0525	0.0358	0.1273	-0.0101	-0.0151	0.0128
5	-0.0419	0.0298	0.0862	-0.0170	-0.0004	0.0102
6	-0.0124	0.0124	0.0157	0.0078	-0.0265	0.0316
7	-0.0015	0.0034	-0.0024	0.0092	-0.0102	0.0303
8	-0.0298	0.0216	0.0539	-0.0172	0.0064	0.0082
9	-0.0161	0.0154	0.0214	0.0074	-0.0306	0.0294
10	-0.0067	0.0072	-0.0028	-0.0033	0.0084	0.0248
11	-0.0210	0.0147	0.0373	-0.0167	0.0109	0.0127
12	-0.0237	0.0175	0.0342	-0.0202	0.0142	0.0109
13	-0.0078	0.0033	0.0265	-0.0071	0.0054	0.0147
14	-0.0148	0.0134	0.0043	-0.0137	0.0154	0.0193
15	-0.0107	0.0074	0.0162	-0.0115	0.0109	0.0147
16	-0.0162	0.0123	0.0193	-0.0164	0.0142	0.0129
17	-0.0187	0.0147	0.0191	-0.0191	0.0165	0.0119
18	-0.0196	0.0150	0.0242	-0.0181	0.0136	0.0081
19	-0.0186	0.0137	0.0270	-0.0161	0.0115	0.0093
20	0.0234	-0.0228	-0.0118	0.0058	-0.0026	0.0199
21	-0.0149	0.0102	0.0265	-0.0125	0.0088	0.0107
22	0.0011	-0.0024	0.0054	-0.0022	0.0055	0.0170
23	-0.0161	0.0112	0.0300	-0.0117	0.0068	0.0087
24	-0.0165	0.0117	0.0279	-0.0131	0.0085	0.0088
25	-0.0152	0.0102	0.0320	-0.0095	0.0043	0.0081
26	-0.0153	0.0116	0.0183	-0.0152	0.0129	0.0108
27	-0.0131	0.0083	0.0300	-0.0083	0.0038	0.0087
28	-0.0158	0.0123	0.0164	-0.0161	0.0138	0.0103
29	-0.0138	0.0088	0.0335	-0.0068	0.0013	0.0075
30	-0.0136	0.0100	0.0217	-0.0097	0.0055	0.0043
31	-0.0137	0.0097	0.0214	-0.0124	0.0095	0.0099
32	-0.0170	0.0136	0.0151	-0.0178	0.0154	0.0092
33	-0.0131	0.0096	0.0182	-0.0127	0.0106	0.0105
34	0.0151	-0.0139	-0.0095	0.0077	-0.0029	0.0192
35	-0.0138	0.0094	0.0270	-0.0096	0.0052	0.0076
36	-0.0119	0.0056	0.0142	-0.0401	-0.0096	0.0183
37	-0.0133	0.0092	0.0252	-0.0096	0.0055	0.0073
38	-0.0145	0.0110	0.0183	-0.0135	0.0104	0.0071
39	-0.0139	0.0125	0.0048	-0.0136	0.0142	0.0138
40	-0.0183	0.0194	0.0120	0.0086	-0.0178	0.0227
41	-0.0118	0.0073	0.0299	-0.0059	0.0012	0.0072
42	-0.0131	0.0103	0.0137	-0.0129	0.0102	0.0049
43	-0.0129	0.0090	0.0234	-0.0095	0.0056	0.0067
44	-0.0064	0.0054	0.0037	-0.0068	0.0087	0.0148

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
496	1.6472	1.9354	0.0000	0.9998	0.0137
497	1.8974	2.2294	0.0001	0.9996	0.0161
498	1.0035	1.1794	0.0005	1.0005	0.0276
499	0.9303	1.0932	0.0002	1.0002	0.0160
500	1.6133	1.8956	0.0001	0.9998	0.0139
501	1.1310	1.3291	0.0004	1.0003	0.0256
502	1.3210	1.5522	0.0001	0.9999	0.0141
503	2.4409	2.8682	0.0001	0.9993	0.0241
504	1.3705	1.6104	0.0001	0.9999	0.0139
505	3.1752	3.7317	0.0003	0.9989	0.0695
506	1.1970	1.4066	0.0002	1.0001	0.0216
507	0.6346	0.7457	0.0002	1.0003	0.0113
508	2.0203	2.3738	0.0001	0.9995	0.0170
509	1.6300	1.9152	0.0001	0.9998	0.0153
510	1.6330	1.9187	0.0001	0.9998	0.0137
511	2.1969	2.5813	0.0000	0.9994	0.0181
512	1.3286	1.5611	0.0001	0.9999	0.0140
513	1.6672	1.9589	0.0001	0.9997	0.0139
514	1.6335	1.9194	0.0000	0.9998	0.0136
515	1.3679	1.6072	0.0001	0.9999	0.0138
516	1.2854	1.5103	0.0001	1.0000	0.0165
517	1.0919	1.2830	0.0001	1.0000	0.0140
518	0.5953	0.6995	0.0002	1.0003	0.0108
519	1.3305	1.5634	0.0001	0.9999	0.0137
520	0.8376	0.9842	0.0002	1.0002	0.0133
521	1.2875	1.5128	0.0001	0.9999	0.0143
522	2.0622	2.4231	0.0001	0.9995	0.0174
523	0.9049	1.0633	0.0002	1.0001	0.0132
524	0.9298	1.0925	0.0001	1.0001	0.0131
525	1.6384	1.9251	0.0001	0.9998	0.0137
526	1.3542	1.5912	0.0001	0.9999	0.0138
527	1.9254	2.2623	0.0000	0.9996	0.0152
528	2.1137	2.4836	0.0001	0.9995	0.0184
529	2.8661	3.3680	0.0002	0.9990	0.0422
530	1.9281	2.2655	0.0001	0.9996	0.0164
531	1.6209	1.9045	0.0001	0.9998	0.0136
532	1.1443	1.3446	0.0001	1.0000	0.0135
533	1.4401	1.6921	0.0001	0.9999	0.0138
534	0.5757	0.6765	0.0003	1.0004	0.0117
535	1.9048	2.2382	0.0001	0.9996	0.0160
536	1.1993	1.4092	0.0001	1.0000	0.0134
537	1.2304	1.4457	0.0001	1.0000	0.0135
538	2.2942	2.6958	0.0001	0.9994	0.0196
539	1.8423	2.1647	0.0001	0.9997	0.0154

	DFBETAS					
Obs	Intercept	Vel	Acc	Dec	PD	TD
496	-0.0027	0.0028	-0.0005	-0.0018	0.0036	0.0091
497	-0.0027	0.0037	-0.0033	0.0018	0.0008	0.0113
498	-0.0008	-0.0017	0.0215	0.0055	-0.0084	0.0032
499	-0.0036	0.0023	0.0067	-0.0035	0.0030	0.0043
500	-0.0048	0.0051	-0.0013	-0.0019	0.0035	0.0095
501	-0.0002	-0.0021	0.0183	0.0038	-0.0059	0.0039
502	-0.0044	0.0040	0.0017	-0.0042	0.0050	0.0071
503	0.0137	-0.0129	-0.0076	0.0057	-0.0011	0.0112
504	-0.0068	0.0069	-0.0005	-0.0041	0.0050	0.0083
505	-0.0108	0.0067	0.0151	-0.0209	-0.0132	0.0131
506	-0.0005	-0.0014	0.0131	0.0008	-0.0019	0.0046
507	-0.0055	0.0050	0.0012	-0.0056	0.0055	0.0037
508	0.0008	0.0002	-0.0043	0.0031	0.0001	0.0114
509	0.0007	-0.0013	0.0029	-0.0010	0.0024	0.0078
510	-0.0041	0.0045	-0.0013	-0.0017	0.0034	0.0095
511	0.0017	-0.0008	-0.0031	0.0040	-0.0025	0.0120
512	-0.0041	0.0036	0.0018	-0.0040	0.0048	0.0071
513	-0.0041	0.0045	-0.0016	-0.0012	0.0030	0.0098
514	-0.0035	0.0037	-0.0011	-0.0017	0.0034	0.0093
515	-0.0070	0.0072	-0.0007	-0.0039	0.0049	0.0084
516	-0.0019	0.0007	0.0060	-0.0025	0.0027	0.0059
517	-0.0052	0.0046	0.0022	-0.0052	0.0055	0.0059
518	-0.0054	0.0049	0.0010	-0.0055	0.0054	0.0035
519	-0.0067	0.0067	-0.0003	-0.0043	0.0052	0.0080
520	-0.0050	0.0043	0.0029	-0.0053	0.0052	0.0044
521	-0.0038	0.0032	0.0025	-0.0040	0.0047	0.0067
522	-0.0010	0.0020	-0.0030	0.0036	-0.0018	0.0119
523	-0.0058	0.0053	0.0017	-0.0057	0.0058	0.0051
524	-0.0075	0.0073	0.0001	-0.0060	0.0063	0.0058
525	-0.0041	0.0045	-0.0015	-0.0014	0.0031	0.0096
526	-0.0074	0.0077	-0.0008	-0.0038	0.0047	0.0085
527	0.0010	-0.0004	-0.0031	0.0017	0.0012	0.0105
528	0.0087	-0.0085	-0.0029	0.0038	0.0002	0.0098
529	-0.0021	0.0002	0.0068	-0.0090	-0.0093	0.0126
530	-0.0015	0.0025	-0.0037	0.0025	0.0003	0.0113
531	-0.0039	0.0043	-0.0013	-0.0016	0.0033	0.0094
532	-0.0060	0.0057	0.0009	-0.0052	0.0057	0.0066
533	-0.0068	0.0071	-0.0010	-0.0031	0.0042	0.0089
534	-0.0036	0.0027	0.0044	-0.0036	0.0031	0.0027
535	0.0051	-0.0053	0.0001	0.0018	0.0010	0.0089
536	-0.0067	0.0065	0.0001	-0.0049	0.0056	0.0072
537	-0.0056	0.0053	0.0007	-0.0047	0.0054	0.0070
538	0.0065	-0.0057	-0.0047	0.0043	-0.0021	0.0116
539	0.0039	-0.0042	0.0005	0.0011	0.0014	0.0087

Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS
8146	0.6933	0.8146	0.0001	1.0001	0.0075
8147	0.1391	0.1634	0.0001	1.0002	0.0012
8148	0.0262	0.0308	0.0003	1.0004	0.0006
8149	0.6141	0.7215	0.0001	1.0001	0.0057
8150	0.5043	0.5926	0.0001	1.0001	0.0044
8151	0.2389	0.2807	0.0000	1.0001	0.0020
8152	0.4906	0.5765	0.0001	1.0001	0.0044
8153	-0.3207	-0.3769	0.0001	1.0002	-0.0037
8154	0.0195	0.0230	0.0001	1.0002	0.0002
8155	0.2334	0.2743	0.0000	1.0001	0.0019
8156	0.4330	0.5088	0.0000	1.0001	0.0036
8157	-0.1531	-0.1798	0.0001	1.0002	-0.0016
8158	0.5359	0.6296	0.0001	1.0001	0.0046
8159	-0.0388	-0.0456	0.0001	1.0002	-0.0004
8160	0.2391	0.2809	0.0000	1.0001	0.0019
8161	0.6256	0.7351	0.0001	1.0001	0.0053
8162	0.5918	0.6953	0.0001	1.0001	0.0051
8163	0.5617	0.6600	0.0001	1.0001	0.0048
8164	0.3211	0.3773	0.0000	1.0001	0.0026
8165	0.4721	0.5547	0.0001	1.0001	0.0040
8166	-0.0494	-0.0581	0.0001	1.0002	-0.0005
8167	0.4895	0.5751	0.0001	1.0001	0.0041
8168	0.4104	0.4822	0.0001	1.0001	0.0037
8169	0.5815	0.6832	0.0001	1.0001	0.0056
8170	0.5052	0.5936	0.0001	1.0001	0.0046
8171	0.1583	0.1860	0.0001	1.0002	0.0014
8172	0.3649	0.4287	0.0001	1.0001	0.0031
8173	0.3218	0.3781	0.0000	1.0001	0.0025
8174	0.2620	0.3078	0.0000	1.0001	0.0021
8175	0.5337	0.6270	0.0001	1.0001	0.0048
8176	-0.3112	-0.3656	0.0001	1.0002	-0.0035
8177	0.2558	0.3006	0.0000	1.0001	0.0021
8178	-0.0807	-0.0948	0.0001	1.0002	-0.0008
8179	0.4592	0.5396	0.0001	1.0001	0.0040
8180	0.1315	0.1545	0.0001	1.0002	0.0011
8181	0.4252	0.4996	0.0001	1.0001	0.0038
8182	-0.0241	-0.0283	0.0001	1.0002	-0.0002
8183	0.0974	0.1145	0.0001	1.0002	0.0008
8184	0.2324	0.2731	0.0000	1.0002	0.0019
8185	0.1709	0.2008	0.0001	1.0002	0.0016
8186	0.4997	0.5871	0.0001	1.0001	0.0044
8187	0.6087	0.7152	0.0001	1.0001	0.0059
8188	0.4498	0.5285	0.0001	1.0001	0.0040
8189	0.1639	0.1925	0.0001	1.0002	0.0014

	DFBETAS					
Obs	Intercept	Vel	Acc	Dec	PD	TD
8146	-0.0030	0.0032	0.0023	0.0016	-0.0035	0.0042
8147	-0.0005	0.0005	-0.0002	-0.0001	0.0002	0.0009
8148	0.0001	-0.0001	0.0004	0.0002	-0.0002	0.0001
8149	-0.0018	0.0020	0.0007	0.0011	-0.0019	0.0037
8150	-0.0001	0.0004	-0.0010	0.0009	-0.0001	0.0029
8151	-0.0005	0.0005	-0.0003	-0.0000	0.0003	0.0014
8152	-0.0014	0.0016	-0.0000	0.0008	-0.0009	0.0030
8153	0.0015	-0.0014	-0.0003	0.0013	-0.0015	-0.0018
8154	-0.0001	0.0001	-0.0000	-0.0000	0.0000	0.0001
8155	-0.0004	0.0005	-0.0003	-0.0000	0.0003	0.0013
8156	-0.0000	0.0002	-0.0008	0.0005	0.0002	0.0025
8157	0.0008	-0.0008	0.0001	0.0004	-0.0005	-0.0009
8158	0.0002	0.0001	-0.0012	0.0010	-0.0002	0.0031
8159	0.0002	-0.0002	0.0000	0.0001	-0.0001	-0.0002
8160	-0.0003	0.0004	-0.0003	-0.0000	0.0003	0.0014
8161	-0.0008	0.0011	-0.0001	0.0011	-0.0013	0.0036
8162	-0.0010	0.0013	-0.0000	0.0010	-0.0012	0.0035
8163	0.0002	0.0001	-0.0011	0.0010	-0.0003	0.0032
8164	-0.0004	0.0006	-0.0005	0.0002	0.0002	0.0019
8165	0.0015	-0.0015	-0.0002	0.0006	0.0002	0.0022
8166	0.0002	-0.0002	0.0000	0.0001	-0.0001	-0.0003
8167	0.0002	0.0001	-0.0010	0.0007	0.0000	0.0028
8168	-0.0010	0.0012	-0.0005	0.0006	-0.0003	0.0025
8169	0.0030	-0.0031	0.0001	0.0012	-0.0002	0.0024
8170	0.0021	-0.0022	0.0002	0.0008	-0.0000	0.0022
8171	-0.0005	0.0006	-0.0002	-0.0001	0.0002	0.0010
8172	-0.0005	0.0007	-0.0006	0.0004	0.0001	0.0022
8173	-0.0001	0.0002	-0.0004	0.0001	0.0003	0.0018
8174	-0.0002	0.0002	-0.0003	-0.0000	0.0004	0.0014
8175	-0.0016	0.0018	0.0003	0.0010	-0.0013	0.0033
8176	0.0017	-0.0016	-0.0001	0.0013	-0.0015	-0.0018
8177	-0.0005	0.0006	-0.0004	0.0000	0.0003	0.0015
8178	0.0002	-0.0001	-0.0001	0.0002	-0.0002	-0.0004
8179	-0.0004	0.0007	-0.0009	0.0007	-0.0001	0.0027
8180	-0.0004	0.0005	-0.0001	-0.0001	0.0002	0.0008
8181	-0.0012	0.0015	-0.0002	0.0007	-0.0005	0.0026
8182	0.0000	-0.0000	-0.0000	0.0000	-0.0001	-0.0001
8183	-0.0003	0.0003	-0.0001	-0.0001	0.0002	0.0006
8184	-0.0005	0.0006	-0.0003	-0.0000	0.0003	0.0014
8185	-0.0006	0.0007	-0.0002	-0.0000	0.0002	0.0011
8186	-0.0013	0.0015	-0.0001	0.0008	-0.0009	0.0030
8187	-0.0022	0.0024	0.0011	0.0012	-0.0022	0.0037
8188	-0.0011	0.0014	-0.0003	0.0007	-0.0005	0.0028
8189	-0.0005	0.0006	-0.0002	-0.0001	0.0002	0.0010

REFERENCES

1. H. Christopher Frey; Nagui M. Rouphil; A. Unal; James D. Colyar; "Measurement of On- Road Tailpipe CO, NO, and Hydrocarbon Emissions using a portable instrument"; *Journal of Air and Waste Management Association*, June 24-28, 2001.

2. Nagui M. Rouphil; H. Christopher Frey; James D. Colyar; Alper Unal; "Vehicle Emissions and traffic Measures: Exploratory Analysis of Field Observations at Signalized Arterials"; *Annual meeting of the Transportation research Board*, January 7-11, 2001.

3. U.S Environmental Protection Agency home page. See http://www.epa.gov/otaq/invntory/overview/definitions.htm, (accessed December, 2006).

4. Yi Grace Qi; Hualiang Harry Teng; and Lei Yu; "Microscale Emissions Models Incorporating Acceleration and Deceleration"; *Journal of Transportation Engineering at ASCE*, May/June 2004.

5. Rakesh B. Singh; James J. Sloan; "Development and Application of a Microscale Emission Factor Model (MicroFac) for Mobile Source Emissions"; *Annual Emission Inventory Conference*, Las Vegas, April 11-15, 2005.

6. *Remote Sensing: A Supplemental tool for Vehicle Emission Control*; U.S Environmental Protection Agency Office of Mobile Sources, August 1993.

7. City of Albuquerque; *Remote Sensing: Vehicle Pollution Management*. http://www.cabq.gov/aircare/rst.html (accessed September 14, 2007).

8. Melanie Sattler; *Transportation and Air quality lecture notes* fall 2005.

9. Melanie Sattler, *Air pollution Dispersion modeling lecture notes*, fall 2005.

10. Randall Guensler; *Data Needs for Evolving Motor Vehicle Emission Modeling Approaches*; Georgia Tech., Atlanta, 1993.

11. A. Unal, H; Christopher Frey; Nagui M. Rouphail; James D. Colyer; *Hot Spot Analysis of Real World Vehicle Emissions based upon a Portable On-board Measurement System*; North Carolina, Raleigh, 2001.

12. John Neter; Michael H. Kutner; Christopher J. Nachtsheim; Williams Li; *Applied Linear Statistical Models*; Fifth edition, 1996.

13. Shauna L. Hallmark, Randall Guensler, and John D. Leonard II, *Stopline Distributions of Speed and Acceleration for Signalized Intersections*, School of Civil and Environmental Engineering, Georgia Tech., Atlanta, Georgia, 30332, 1998.

14. U.S Department of Transportation and Federal Highway Administration; *Air Quality and Transportation*; See http://www.tfhrc.gov/pubrds/03jul/10.htm (accessed January, 2007).

15. Robert Jourmard; Peter Jost; John Hickman; Dieter Hassel; *Hot Passenger Car Emissions Modeling as a Function of Instantaneous Speed and Acceleration*; 1994.

16. Environmental Protection Agency, *Clean Air Act*, http://www.epa.gov/oar/oaqps/pegcaa/pegcaa02.html (accessed January 2007).

17. Environmental Protection Agency; *Green Book-Criteria Pollutant;*, http://www.epa.gov/oar/oaqps/greenbk/o3co.htm (accessed September, 2007).

18. Alper Unal; Nagui M. Rouphail; H. Christopher Frey; *Effect of Arterial Signalization and Level of Service on Measured Vehicle Emissions*; Transportation Research record 1842 paper No. 03-2884.

19. Steven H. Cadle; Robert A. Gorse; Timothy C. Belian; Douglas R. Lawson;

"Real-World Vehicle Emission: A Summary of the Eighth Coordinating Research Council On-Road Vehicle Emissions Workshop"**;** *Journal of the Air & Waste Management Association*; March 1999.

21. Dr. Christopher Frey; Dr. Nagui Rouphail; Alper Unal; James Colyar; *Emissions reduction Through Better Traffic Management: An Empirical Evaluation Based Upon On-Road Measurements*; North Carolina State University: North Carolina, Raleigh, NC 27695, December 2001.

22. Ignatus Fomunung; Simon Washington; Randall Guensler; *Comparison of Measure and Mobile 5a predictions Using Laboratory Measurements of Vehicle Emission*; Transportation Planning and Air Quality; Georgia Institute of Technology: Atlanta, Georgia, 1999.

23. Horiba Instruments Inc.; *On-board Emission Measurement System, OBS-1000 Instruction Manual*; 2003.

24. Gary R. McGaughey; Nimish R. Desai; David T. Allen; Robert L. Seila; William A. Lonneman; Matthew P. Fraser; Robert A. Harle;, Alison K. Pollack; Jason M. Ivy; James H. Price; *Analysis of motor vehicle emissions in a Houston tunnel during the Texas Air Quality Study*; Atmospheric Environment 38, 2004.

25. Roger Westerholm; Karl-Erik Egeback; *Exhaust Emissions from Light-duty and Heavy-duty Vehicles: Chemical Composition, Impact of Exhaust after Treatment, and Fuel Parameters*; Environmental Health Perspectives, 102(suppl 4): 13-23 (1994).

26. U.S. Environmental Protection Agency; *Technology Transfer Network Support Center for Regulatory atmospheric Modeling*; http://www.epa.gov/scram001/aqmindex.htm, (accessed January, 2007).

27. Randall Guensler; Simon Washington; William Bachman; *Overview of the MEASURE Modeling Framework*; Transportation Planning and Air Quality III, 1998.

28. Peter J. Sturm; Raimund Almbauer; Christoph Sudy; Karl Pucher; *Application of Computational Methods for the Determination of Traffic Emissions*; Institute for Internal Combustion Engines and Thermodynamics: Technical University of Graz, Austria, 1997.

29. Wing-Tat Hung; Hing-Yan Tong; Chun-Shun Cheung; *A Modal Approach to Vehicular Emissions and Fuel Consumption Model Development*; Hong Kong Polytechnic University; *Hong Kong, SAR, 2005.*

30. H. Christopher Frey; Alper Unal; Jianjun Chen; and Song Li; *Modeling Mobile Emissions based upon In-use and Second-by-Second data: Development for Conceptual*

Approaches for EPA's new MOVES model; North Carolina State university, Raleigh, NC27695.

31. Michael H. Kutner; Christopher J. Nachtsheim; John Neter; Williams Li; *Applied Linear Statistical Models*; Fifth edition, 2004.

32. R. Hogg; J. Ledolter; A*pplied Statistics for Engineers and Physical Scientists;* 1992.

33. R. Mason; R. Gunst; J. Hess; *Statistical Design and Analysis of Experiments*; 1989.

34. Paul Richard Lederer; *Analysis and Prediction of Individual Vehicle Activity for Light-Duty Trucks on Freeway Rumps*; Dissertation, June 2001.

35. U.S Environmental Protection Agency, Office of Mobile Sources; *Remote Sensing: A Supplementary Tool for Vehicle Emission Control*, August 1993.

36.Joseph Veron; Paul E. Marik; Robert E. Fromm; Alfredo Gueler; *Carbon Monoxide Poisoning: A Review for Clinicians*,1999.

37. World Watch Institute; *Air Pollution Now Threatening Health Worldwide*, http://www.worldwatch.org/node/1586, (accessed January 2007).

38. Robert Jourmard; Peter Jost; John Hickman; Dieter Hassel; *Hot Passenger Car Emissions Modeling as a Function of Instantaneous Speed and Acceleration*; 1994.

39. Taosheng Jin; Lixin Fu; *Application of GIS to Modified Models of Vehicle Emission Dispersion*; Atmospheric Environment, July, 2005.

40. Environmental Protection Agency; *OAQPS-Taking Toxics Out of the Air*; http://www.epa.gov/oar/oaqps/takingtoxics/p1.html , (accessed October, 2007).

.

41. Bruno Ando; Salvatore Baglio; Salvatore Graziani; Nicola Pitrone; *Models for Air Quality Management and Assessment***;** IEEE Transactions on System, August 2000.

42. Randall Guensler; Simon Washington; William Bachman; *Overview of the Measure Modeling Framework*; Georgia Institute of Technology; Georgia, Atlanta, 30332.

43. Office of Integrated Analysis and Forecasting, EI – 81 (2007), U.S; *Carbon dioxide Emissions from Energy Sources 2006 flash estimates*; U.S department of energy, 2007.

44. U.S environmental Protection Agency and Office of Transportation and Air Quality; *Greenhouse Gas Emissions*; 2006.

45. S.A. Ardekani; *Fuel Consumption Model Class Notes, UTA*; 2007.

46. K. Ahn; *Microscopic Fuel Consumption and Emissions Modeling*; 1998.

47. C. David Cooper; F.C. Alley; Air Pollution Control: A design approach; 2002.

48. Victoria Chen; *Advanced Engineering Statistics Lecture Notes*, Spring 2006.

49. EPA Emission trend Report: Office of Air Quality Planning and Standards; Research Triangle Park, North Carolina, September 2003.

50. To´th-Nagy; Csaba; Conley, John J.; Jarrett, Ronald P.; and Nigel N. Clark. "Further Validation of Artificial Neural Network-Based Emissions Simulation Models for Conventional and Hybrid Electric Vehicles." Journal of the Air & Waste Management Association, Vol. 56, pp. 898-910, 2006

51. Jazcilevich, Aro´n D.; Garcı´a-Fragoso, Alejandro; Reynoso, Agustı´n Garcı´a; Grutter, Michel; Diego-Ayala, Ulises; Lents, Jim; and Nicole Davis. "A Vehicle Emissions System Using a Car Simulator and a Geographical Information System: Part 1--System Description and Testing." *Journal of the Air & Waste Management Association*, Vol. 57, pp. 1234-1240, 2007.

52. Hung, Wing-Tat; Tong, Hing-Yan; and Chun-Shun Cheung. "A Modal Approach to Vehicular Emissions and Fuel Consumption Model Development." Journal of the Air & Waste Management Association, Vol. 55, pp. 1431-1440, October 2005.

53. S.E. Moore, **An Analysis of Fuel Consumption Models,** Proceedings of the Joint SAE-A, AARB Second Conference on Traffic, Energy and Emissions, Melbourne, Australia, May 19-21, 1982.

54. Frey, H. Christopher; Unal, Alper; Rouphail, Nagui M.; and James D. Colyar. "On-Road Measurement of Vehicle Tailpipe Emissions Using a Portable Instrument." *Journal of the Air & Waste Management Association*. Vol. 53, pp. 992-1002, August 2003.

55. Frey, H. Christopher; Rouphail, Nagui M.; Unal, A.; and James D. Colyar. "Measurement of On-Road Tailpipe CO, NO, and Hydrocarbon Emissions Using a Portable Instrument." *Proceedings of the Air & Waste Management Association's 96th Annual Conference & Exhibition*, Orlando, Florida, 2001.

BIOGRAPHICAL INFORMATION

Benjamin Afotey hails from the West African country of Ghana. He was born and raised in the city of Accra. He completed his secondary education at St. Peter's Secondary School and Ghana National College. He graduated from the University of Science and Technology, Ghana with a Bachelors of Science degree Chemical Engineering in June, 2000. During his final year, he worked as a student intern at the Cocoa Processing Company, Ghana.

Benjamin completed his Masters in Civil/Environmental Engineering from the University of Texas, Arlington in December, 2003. During this period he was appointed by Dr. Crosby as a graduate teaching assistant.

In January 2005, he began his Doctorial Degree in Civil/Environmental Engineering where he was appointed a graduate research assistant by Dr. Melanie Sattler. The North Central Texas Council of Governments' Aftermarket Technology and Fuel Additive Research Project is one of the research projects he worked on while completing the program.

In December, 2008, he graduated from the university with a Doctorate Degree in Civil/Environmental Engineering. He developed a micro scale model to estimate exhaust emissions of light duty gasoline vehicles.

He is currently working with TCEQ in the Office of Compliance and Enforcement, Houston, Texas. Research interests include emissions measurement, air quality modeling, emissions modeling, and air pollution control technologies.