2010-01-01

Development of a Land Use Regression Model to Predict Nitrogen Dioxide Concentrations

Teresa Sosa
University of Texas at El Paso, teremont20@yahoo.com

DEVELOPMENT OF A LAND USE REGRESSION MODEL TO PREDICT NITROGEN DIOXIDE CONCENTRATIONS

TERESA SOSA

Department of Civil Engineering

APPROVED:

Wen-Whai Li, Ph.D., Chair

Charles D. Turner, Ph.D.

Raed K. Aldouri, Ph.D.

Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

DEVELOPMENT OF A LAND USE REGRESSION MODEL TO PREDICT NITROGEN DIOXIDE CONCENTRATIONS

by

TERESA SOSA, BS CIVIL ENGINEERING

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN ENVIRONMENTAL ENGINEERING

Department of Civil Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

August 2010

Acknowledgements

I would like to thank Dr. Wen-Whai Li and Dr. Hongling Yang for their support and guidance throughout this thesis. Also, I would like to thank Dr. Charles D. Turner and Dr. Raed K. Aldouri for their participation in my thesis committee. Special thanks to Estrella Molina, Alberto Barud Zubillaga (UTEP), and Gerardo Tarín Torres (SEMARNAT) for assisting with the data acquisition.

To my family, for always being there for me, thank you very much from the bottom of my heart. Special thanks to my parents for their unconditional love and support. Last, but definitely not least, to my husband Gustavo Sosa, thanks for all your patience and encouragement, I love you!

Abstract

Air pollution is a major environmental concern in the El Paso-Juárez region. According to the Ministry of Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales, SEMARNAT) the city of Juárez is one of the city's in México with the highest atmospheric levels of pollution because of its accelerated and unplanned urban growth. One air pollutant of concern is nitrogen dioxide (NO_2) due to its detrimental health effects that have been associated with airway inflammation in healthy people and increased respiratory symptoms in people with asthma. Land use regression modeling is a GIS based approach that seeks to predict pollution concentrations at a given site based on surrounding land use, traffic characteristics, and other geographic variables in a multivariate regression model. This type of model has been a practical and effective method to predict intraurban variation in nitrogen dioxide in several places in North America. It will be useful to create a similar model at the El Paso-Juárez borderland region to assess nitrogen dioxide exposures.

This research evaluates the strength and association of different land regression variables into predicting NO_2 concentrations. Monitoring for NO_2 levels was conducted at 27 locations in the city of Juárez that included 22 schools and 5 homes in a prior study from December 2002-September 2003. Main point sources for NO_2 were identified in the El Paso-Juárez region and include: international ports of entry, cement plants, electric engine factories, and petroleum refineries. Distance to main point sources as well as traffic volume and traffic density on major streets near the monitoring locations were calculated utilizing ArcGIS 9.3.1 and used as predictor variables. Significant pearson correlations with NO_2 concentrations were found with the following predictive variables: distance to a cement plant (DIST_CP), traffic density within the 1000 meter buffer zone (TD_1000), distance to an oil refinery (DIST_OR), distance to electric engine factories (DIST_EE), and distance to a major street with the second highest traffic volume (DIST_2nd). Most of the significant correlations found were consistent with the findings in previous studies conducted in the El Paso-Ciudad Juárez area by Gonzales, et al. (2005) and Smith, et al. (2006). A model built through a stepwise multivariate regression analysis revealed that the three main variables for NO_2 variations include distance to a cement plant (DIST_CP), distance to a major street with the second highest traffic volume (DIST_2nd), and distance to an oil

refinery (DIST_OR), predicting 59 percent of the variation in NO_2 concentration. A bootstrapping analysis of 1,000 iterations evaluated and verified the robustness of the model. Recommendations for future analysis on this pollutant include a closer look into the effect of distance to a cement plant in the variation of NO_2 concentrations, and perhaps the inclusion of other variables that could increase the predictability of the model, such as elevation above sea level.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview of Problem Statement

Air quality is a major problem in the Juárez-El Paso region. Ciudad Juárez has been categorized as one of the places in México with the highest atmospheric levels of pollution because of the accelerated and sometimes unplanned urban growth (SEMARNAT, 2006). Despite revised emission standards and technical improvement in pollution control measures, expanding industrialization and increasing traffic volumes in developing countries will drastically increase total emissions of many air pollutants. Moreover, a study conducted by Li et al. (2001) at the border cities of El Paso and Juárez, where they monitored concentrations for several air pollutants, including nitrogen dioxide (NO_2) demonstrated that levels for these contaminants in Ciudad Juárez were typically more than two times higher than those measured in El Paso. Common sources of pollutants include motor vehicles, power plants, industrial facilities, and dust from unpaved roads. Pollutants coming from these sources and of concern in the border region include particulate matter (PM_{10}, $PM_{2.5}$), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and carbon monoxide (CO). Several studies have shown that air pollutants are associated as having detrimental health effects. Nitrogen dioxide in particular has been associated with increased hospital admissions for cardiovascular diseases (Lau et al., 2009). In addition, short-term exposures to NO_2 may cause lung injury and long-term exposures may reduce immunity and lead to respiratory infections (Han and Naeher, 2006).

Air pollution modeling needs to be able to predict intraurban variation of pollutant concentrations. According to Hoek et al., (2008), the first cohort studies published in the mid-1990s that compared mortality rates between cities, utilized exposure values based on pollutant concentrations measured at a central site within each city. However, in the past decade, various studies have documented significant variation of outdoor air pollution at a small scale within urban areas for pollutants such as nitrogen dioxide (NO_2) and black smoke; furthermore, there is evidence from epidemiological studies that within-city contrasts of air pollutants are associated with larger contrasts than between-city (Hoek et al., 2008). Sahsuvaroglu et al., (2006) in his study recognized the same need: outdoor estimates have shown to be poor predictors of true personal exposures in epidemiological

studies, revealing that while central monitors correlate well with average personal exposures for daily variations in acute studies, they do not account for spatial variations within cities for chronic exposures. Measurements made at one single location are not valid in assessing the chronic exposure of the community to that pollutant. Having a tool that is able to predict intraurban variation of pollutant concentration is of utmost importance. Land use regression modeling is a current option that can aid in predicting intraurban variation of pollutants. Moreover, several studies have been conducted in European cities and in a few cities in the United States, such as San Diego, trying to create tools that will aid in the prediction of air pollution contaminants within cities. Having an understanding of the factors that influence the concentration of pollutants such as NO_2 will help us control those factors and protect the vulnerable population.

1.2 Research Objectives

The main goal of this research is to evaluate the strength and association of different land regression variables into predicting nitrogen dioxide concentrations. The main goal in this research was achieved by conducting the following tasks:

1. Monitoring of nitrogen dioxide levels at several schools and private homes in Ciudad Juárez (conducted in a prior study).
2. Identifying the major contributors for nitrogen oxides in Ciudad Juárez.
3. Establishing predictive variables for nitrogen dioxide.
4. Evaluating the correlations between nitrogen dioxide concentrations and the established predictive variables.
5. Creating a land use regression model that will predict nitrogen dioxide concentrations by running a multiple regression analysis to evaluate the prediction strength of each of the independent variables.
6. Validating the strength of the created model.

1.3 Thesis Organization

This thesis is organized into six chapters. Chapter one provides an introduction into the air pollution problem that is evaluated in this thesis as well as the steps used to address it. Chapter two is

the literature review chapter, providing background information on land use regression models and different locations and studies that have utilized this type of modeling for the prediction of air pollutants focusing on NO_2. Chapter three provides an overview of NO_2 as an air pollutant, its health effects, emission standards in both the United States and México, and sources that influence the concentration of this contaminant in Ciudad Juárez. Chapter four provides a description of the methodology that was followed in this thesis. It includes a description of the monitoring period, study variables identified, and an overview of the statistical analysis that was performed. Chapter five provides the results and discussion of the study analysis that includes a description of the model chosen and its validation. Finally Chapter six presents the main conclusions and recommendations for future studies.

Chapter 2: Literature Review

This chapter provides a review of land use regression models and continues with a review of studies that have used land use regression modeling and its application to predicting nitrogen dioxide concentrations. The second part also gives a background regarding different studies that have focused on predicting NO_2 levels and the different factors that have been found as significant for this particular pollutant.

2.1 Land Use Regression Modeling

Land use regression (LUR) modeling is a GIS-based regression approach that seeks to predict pollution concentrations at a given site based on surrounding land use, traffic characteristics, and other geographic variables in a multivariate regression model (Ryan and LeMasters, 2007; Jerrett et al., 2005; Smith et al., 2006). It uses measured pollution concentrations at a certain location as the dependent, response variable (y) and uses land types variables as the independent variable (x) around the specified location, for a given buffer zone, to predict the measured concentrations. The buffer zone distance varies according to the pollutant being modeled and factors such as decay rate of the pollutant should be taken into account. Levels of pollution can then be predicted for any location using the parameter estimates derived from the regression model.

The incorporation of site-specific variables into this model detects small area variations more effectively than other models. Ryan and LeMasters (2007) in their analysis of land use regression models provided with four classes of geographic variables for this particular type of models that include 1) road type, 2) traffic count, 3) elevation, and 4) land cover. Road type includes variables such as major road, minor road, bus route, highway, and many others. The definition of the variables will differ according to each study. For example, one study defined a major road as one having 50,000 vehicles per day or more; a different study defined a major road as one having an average daily truck count of 10,000 trucks per day or more. Traffic count would usually give the daily traffic volume, or maximum traffic volume in units such as total number of vehicles per day or vehicle/kilometers-hour, again, depending on the study being performed. Elevation encompasses the height above sea level for the monitored

location. The land cover category encompasses a broader type of variables that include: areas for a designated land use (square footage of commercial land use, industrial, etc.), distance from monitored location to point sources, or distance to the coast (if applicable), household density, and population density. This particular study will focus on two classes: traffic count and land cover, since they are the variables that seem to be more relevant for the prediction of NO_2 concentrations.

As far as the number of monitoring sites that should be included for the model, there is no established methodology to determine the required number of monitoring locations. In the reviewed articles, the number of monitoring locations ranged from 20 to 100 sites. Hoek et al. (2008), recommends a total of 40-80 sites as a reasonable number to choose for site-specific monitoring, however, the size of the population and the city can be taken into account to determine the actual number.

One of the main advantages of land use regression modeling recognized by several researchers is its applicability to account for small scale variability in intraurban pollutant concentrations, helping to assess individual's exposure to different pollutants (Ryan and LeMasters, 2007; Jerrett et al., 2005). According to Jerrett et al., (2005), a land use regression model allows adaptation to local areas without additional monitoring or data acquisition, or it can identify areas requiring more intensive monitoring. It also has a relatively low cost of implementation as compared to other modeling options. Land use regression models have been successful for modeling several air pollutants including nitrogen dioxide (NO_2), nitrogen oxides (NO_x), particulate matter with a cutoff diameter of 2.5 μm ($PM_{2.5}$), and volatile organic compounds (VOCs) in both European and North American cities (Hoek et al., 2008).

One disadvantage to land use regression modeling includes its limits in transferability. Although it is recognized that further examination should be conducted in this topic, several authors agree that the methodology and model developed in a specified location can be transferable to a certain extent to another location, but with caution (Jerrett et al., 2005; Hoek et al., 2008; Ryan and LeMasters, 2007). One model developed in a particular city, may not be applicable to a different city if this new city has different topographical characteristics, traffic patterns, or land structure. For example, Jerrett et al. (2005), describes in his review that a model developed for the city of Amsterdam (Netherlands) did not

explain spatial variation in NO_2 concentrations for the city of Hamilton (Canada). It is recommended to transfer the developed model to nearby areas having similar characteristics that will create similar predictor variables. Also, models developed for different cities will have certain variables that describe specific geographic characteristics of that area. For example, in the studies reviewed by Ryan and LeMasters (2007), one conducted in Ohio found that elevation was important for predicting ECAT (Elemental Carbon Attributed to Traffic, a marker of diesel exhaust). Another study led by Ross et al. (2006), in California found that distance to the coast (Pacific Ocean) was a significant predictor for NO_2 levels. These findings emphasize the importance of including local geographic variables in land use regression models.

Jerrett et al. (2005), compared several models that assess intraurban variations, including proximity models, dispersion models, interpolation models, integrated meteorological-emission models, hybrid models, and land use regression models. Compared to the rest of the models, in terms of complexity with respect to suitability requirements and cost of implementation (software, equipment, cost), land use regression modeling had the third place (out of six places, with the sixth place having the most complex, most expensive model). In terms of software expertise, land use regression modeling required knowledge of GIS software, statistical software, and monitor equipment (depending on the type of pollutant to be studied). Its data requirements ranked also in the middle as compared to other models. Table 2.1 below provides more details in the comparison of land use regression modeling and other models. In Ryan and Le Masters (2007) study, where he compares a proximity model to a land use regression model he recognizes that the use of a proximity model may lead to exposure misclassification. The land use regression model resulted in a range of predictor variables that helped in more accurate predictions of ECAT, and in turn reduced exposure misclassification that can arise from a proximity model.

Table 2.1: Comparison of models assessing intraurban variations of pollutants

Model	**Data requirements**	**Implementation Cost**			**Software expertise**	**Transferability**
		Equipment	**Software**	**Personnel**		
Proximity	Traffic volumes, distance from line source, questionnaire	Low	Low	Medium	GIS, Statistics	Low
Geostatistical	Monitoring measurements	Medium	Medium	Low	GIS, Spatial Statistics	Low
Land Use Regression	Traffic volumes, land use, meteorology, monitoring	Medium	Medium	Medium	GIS, Statistics, Monitors	Medium
Dispersion	Traffic volumes, emission from point sources, meteorology, monitoring measurements, topography	High	High	Medium	GIS, Statistics, Monitors, Dispersion	High
Integrated meteorological emission	Traffic volumes, emission from point sources, meteorology, monitoring measurements, topography	High	High	High	GIS, Statistics, Monitors	Medium
Hybrid	Questionnaire, personal monitoring data, other	High	N/A	N/A	Personal Monitors, Survey Design, others	Low

Source: Jerrett et. al., 2005

2.2 Land Use Regression Modeling and NO_2

In a study led by Sahsuvaroglu et al. (2006), about predicting intraurban variations in air pollutants, in Hamilton, Canada, results indicated that a land use regression model is a practical and effective method to predict intraurban variation in nitrogen dioxide within the city. The land use regression model was developed utilizing 107 monitoring sites and it explained 76 percent of the variation in NO_2 concentrations. Significant variables positively correlated with NO_2 included traffic

density, proximity to a highway, and industrial land use. Open land use and distance from the lake were negatively correlated with NO_2 concentrations. The explanatory power of the model was increased with the inclusion of wind patterns, traffic densities, and seasonal validation.

In another study led by Ross et al. (2006), they modeled the intraurban distribution of NO_2 utilizing a land use regression model and data from 39 monitoring locations in the fall of 2003 at San Diego, CA. Using multiple linear regression, they were able to predict 79 percent of the variation in NO_2 levels based on four variables: traffic density within 40-300 m of the sampling location, traffic density within 300-1000 m of the sampling location, length of road within 40 m, and distance to the Pacific coast. Traffic density within 40-300 m alone was the most significant predictor, accounting for 54 percent of the variation. In this specific study, the model was validated by predicting NO_2 levels at 12 validation samples, predicting levels on average within 2.1 ppb, and all of the predicted estimates were within a factor of 1.5 times the range of the observed levels.

In her study of *Characterization of a spatial gradient of nitrogen dioxide across a United States-México border city during winter*, Gonzales et al. (2005), evaluates a set of geographic variables for predicting variation in NO_2 concentrations, utilizing concentration measurements collected at 20 elementary school sites during a 7-day week period in the winter of February, 1999. The results from the study concluded that proximity to vehicle-related sources of NO_2 and site elevation are key predictors for vehicle-related air pollution exposure in the El Paso region. A multivariate regression analysis demonstrated that site elevation and distances to a main highway and to an international port of entry from México explained 79 percent of the variance in passive measurements. Furthermore, her results indicated that the El Paso NO_2 gradient is associated with proximity to areas of localized high traffic density and lower elevation. In addition, her study also indicated that low winds and the predominant meteorological inversions during the winter study period resulted in the confinement of NO_2 near central El Paso and emission sources, while on days of moderate wind, the precursor cloud was dispersed and NO_2 concentrations are lower, but more widely and evenly dispersed across the city. The dispersion pattern also provided an explanation for the significance of site elevation for predicting NO_2 in El Paso.

Smith et al. (2006), conducted another study in the El Paso region to predict NO_2 concentrations as well as benzene, toluene, ethylbenzene, *o*-xylene, and *m,p*-xylene (volatile organic compounds) at a total of 55 schools from predictive equations developed from measurements conducted at 22 schools during the months of November and December, 1999. The predictive equations were developed by regressing the monitored measurements at the 22 schools utilizing a GAM approach on several land-use variables derived from GIS. Their analysis demonstrated that the most important variables for predicting NO_2 levels in El Paso are: elevation above sea-level, traffic intensity within 1000 m of the monitored location, population density, distance to the nearest border crossing, and distance to the nearest oil facility.

In the same study as above (conducted during the months of November and December 1999), but in a different publication, Noble et al. (2003), continuously measured fine and ultrafine particulate matter, gas phase pollutants (that included NO_2), and several meteorological conditions to determine potential surrogates for particulate matter concentration. His results from the NO_2 sampling reveal that this pollutant followed a trend consistent with vehicular traffic: peaking in the mid-morning and early evening. In addition, concentrations were 25-140 percent greater when the wind was coming from the North, as compared to when the wind was coming from the South. These higher concentrations with southerly winds were attributed in part to the fact that average wind speed was lower for southerly than for northerly winds. However, due to the low correlation between wind speed and pollutant concentrations (NO_2 with r = -0.49) other factors may contribute to the high pollutant concentrations, including the 350,000 vehicles registered in Ciudad Juárez, nearly twice the number of those registered in El Paso, and the level of maintenance at which those vehicles are kept. Other Mexican sources for pollutant emissions included brick kilns, municipal waste burning, and maquiladoras- internationally owned production and assembly plants.

Chapter 3: NO_2 and Site Characterization

3.1 Site Characterization

Ciudad Juárez is located geographically in the northern hemisphere, with a northern latitude of 31°07’48” and 31°48’ and a western longitude of 106°06’57” and 106°98’44” and elevation of 1140 meters above sea level. Ciudad Juárez shares its airshed with several cities from the US that include El Paso, Texas and Sunland Park, New Mexico and therefore, air pollutants released in one area can have a great effect in another area, even from a different country.

At the beginning of the 21st century, Ciudad Juárez was witnessing a strong period of growth with a population of 1.2 million in the year 2000. It was expected to continue growing up to 1.5 million by the year 2005 and up to 2.3 million by the year 2020. Economic activities in Ciudad Juárez are focused around the manufacturing industry. In the year of 1995, there were a total of 309 factories; by the year of 2005, that number had increased to 819 (SEMARNAT, 2006). On the other hand, Ciudad Juárez has also experienced growth in average trips per person, which is parallel to the growth of number of vehicles circulating in the city. Public transportation is not widely used or promoted due to the lack of security, cost, comfort, or general organization/coordination. Therefore, the use of private vehicle is the main form of transportation in Ciudad Juárez (SEMARNAT, 2006).

Growth in the number of factories as well as growth in the number of vehicle counts and increased usage has left a mark in the atmosphere of the region. Table 3.1 below provides a breakdown of the different sectors in the city that contribute to nitrogen oxides (NOx) emissions in Ciudad Juárez. In Figure 3.1 we can appreciate that the industry sector contributes 44 percent of the emissions, the transportation sector 51 percent and the commercial sector and other services 5 percent. As can be observed from Table 3.1, the main contributors for nitrogen oxides in Ciudad Juárez come from the transportation sector, especially heavy duty diesel trucks which contribute 31 percent of total NOx emissions, light passenger gasoline vehicles contributing 11 percent of the total NOx emissions, and light gasoline trucks contributing 7 percent of the total NOx emissions (SEMARNAT, 2006). The main contributors under the Industry sector are outlined with two asterisks. For example, power plants under the industry sector come in second place with 21 percent, and electric engine factories 11 percent. Other

sources with smaller, but still significant contributions to nitrogen oxides include cement plants, residential combustion, and others (SEMARNAT, 2006) see Table 3.1.

Table 3.1: NO_x Emission Inventory for Ciudad Juárez (2002)

Sector	**NO_x (tons/yr)**
Industry	**8273**
Petroleum	8
Chemical	26
Metallurgical	3
Automotive	22
Cement**	1300**
Ceramic and glass	3
Power Plants**	4080**
Hospitals	2
Food	10
Products of varied materials	2
Concrete and asphalt	4
Metallic products	420
Plastic products	26
Printing	1
Electronics	194
Medical equipment	3
Light bulbs production	1
Electric engines**	2158**
Textile	8
Commerce and other services	**1000**
Asphalt	10
Brick kilns	29
Outdoor burning	35
Commercial combustion	253
Residential combustion	626
International ports of entry	47
Transportation	**9622**
Light duty vehicles (gasoline)	2071
Light duty buses (gasoline)	1332
Heavy duty vehicles (gasoline)	321
Light duty vehicles (diesel)	15
Light duty buses (diesel)	8
Heavy duty vehicles (diesel)	5848
Motorcycles	28

Source: SEMARNAT, 2006

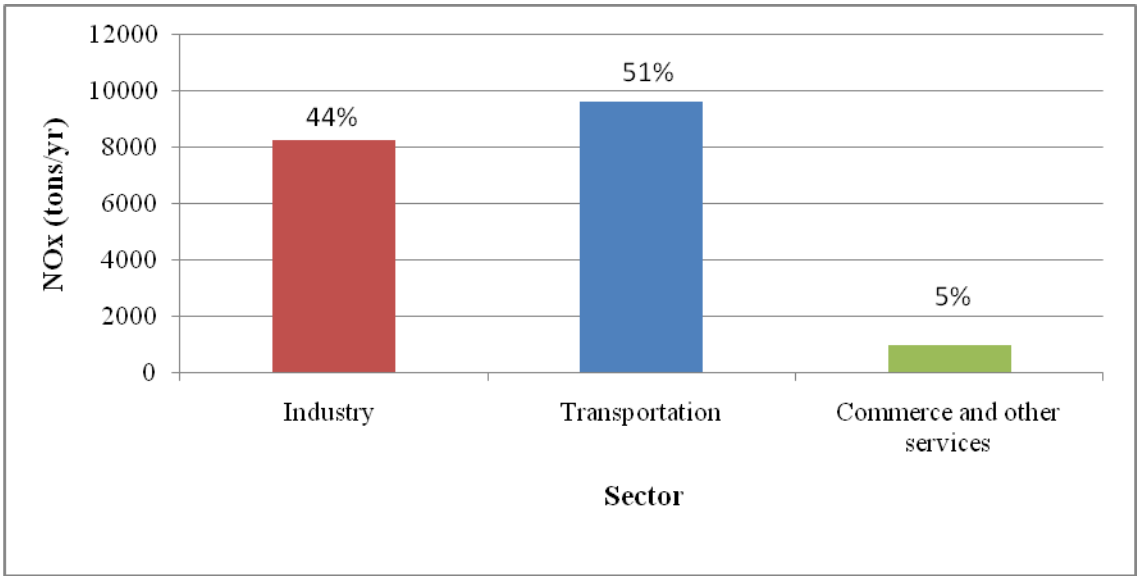

Figure 3.1: NO_x yearly emissions (2002) by sector in Ciudad Juárez

3.2 NO_2 Characterization

Nitrogen dioxide is a reddish brown gas highly reactive in ambient air. Both nitrogen dioxide and nitrogen oxides are formed during the combustion process. In the atmosphere and in industrial devices, NO reacts with O_2 to form NO_2. NO and NO_2 are often treated together as one problem and written as NO_x. Most regulations for NO_x emissions base all numerical values on the assumption that all the NO is converted to NO_2. Nitrogen dioxide can be classified as a primary pollutant when it is directly emitted from its sources, or as a secondary pollutant when produced from chemical reactions in the atmosphere. Nitrogen dioxide is part of the group of nitrogen oxides that are considered as one of the six criteria air pollutants by the Environmental Protection Agency (EPA). Although that list includes other nitrogen compounds, such as nitrous acid and nitric acid, nitrogen dioxide is the component of greatest interest and the indicator for the larger group of nitrogen oxides. Nitrogen dioxide is considered an important pollutant not only for its detrimental health effects, but also for it role in the formation of photochemical smog and ground-level ozone (SEMARNAT 2006, Han and Naeher, 2006). A complex set of simultaneous atmospheric reactions take place in order to form ozone (O_3) from NO_2:

$$NO_2 + h\nu \rightarrow O + NO$$

$$O + O_2 + M \rightarrow O_3 + M$$

$$NO + O_3 \rightarrow NO_2 + O_2$$

where *hv* represents a photon of light of proper wavelength and M represents any other molecule, usually N_2 or O_2, which must carry away some of the energy released in the reaction.

According to the EPA, current scientific evidence links short term NO_2 exposures with adverse respiratory effects including airway inflammation in healthy people and increased respiratory symptoms in people with asthma. According to Table 3.2, long term exposure to concentrations as low as 0.06-0.1 parts per million by volume (ppmv) can lead to respiratory diseases, concentrations above 1.5 ppmv can cause breathing difficulty. The action of nitrogen dioxide involves the reactions of nitrogen oxides with ammonia, moisture, and other compounds to form small particles. These small particles penetrate deeply into the lungs causing or worsening respiratory diseases, aggravating existing heart disease, and leading to increased hospital admissions and premature death. Inside our bodies, nitrogen dioxide reacts with moist tissue to form nitric acid attacking the tissue by acting as a corrosive compound (Turco, 1997).

Another indirect health effect of nitrogen dioxide comes from ozone. Nitrogen dioxide along with volatile organic compounds, heat, and sunlight participate in the formation of ozone. Health effects from exposure to ozone can be as minor as nose and throat irritation or as major as reduction in lung function, increased respiratory symptoms and increased respiratory related emergency visits, hospital admissions, and possibly premature deaths.

Table 3.2: Health effects from exposure to different concentrations of NO_2

Concentration (ppmv)	**Health Effect**
0.06-0.1	Long-term exposure promotes disease
1.5-5.0	Breathing difficulty
25-100	Acute bronchitis
150	Death

3.3 NO_2 Emission Standards

The National Ambient Air Quality Standards (NAAQS) for NO_2 are shown in Table 3.3. EPA's NAAQS for NO_2 are designed to protect against exposure to the entire group of nitrogen oxides (NO_x), since, as stated by the EPA, control measures that reduce NO_2 can generally be expected to reduce population exposures to all gaseous NO_x. However, NO_2 is the component of greatest concern and is used as the indicator for the larger group of NO_x. The 1-hour standard of 0.100 ppm was recently set on January 22, 2010 by the EPA. On that same day, EPA also established a new methodology to verify compliance with the 1-hour standard: the 3-year average of the 98^{th} percentile of the annual distribution of daily maximum 1-hour average concentrations. The air basin of the El Paso-Juárez region is monitored by several stations on both sides of the border. El Paso has been under compliance for NO_2 levels in the past years. However, the 8-hour standard for ozone was recently revised to 0.075 ppm (effective May 27, 2008) and according to Ellen Smyth, city's director of Environmental Services, El Paso barely meets this standard.

México also follows a series of pollutant standards entitled *Normas de Calidad del Aire* or Regulations for Air Quality which specify the threshold levels for criteria pollutants generated from anthropogenic as well as natural sources. These regulations are set by the *Secretaría de Salud* or the Health Department as the Official Mexican Regulations (*Normas Oficiales Mexicanas or NOM's)* and nitrogen dioxide is included as part of the criteria pollutants to be regulated. Table 3.4 shows the standards for nitrogen dioxide as set by the NOM's. Looking at the 1-hour regulation of 0.100 ppm set by the EPA in the US, one can contemplate that US regulations are more stringent than Mexican regulations, with a 1-hour regulation of 0.21 ppm.

Table 3.3: National Ambient Air Quality Standards for NO_2

Primary Standards		**Secondary Standards**	
Level	**Averaging Time**	**Level**	**Averaging Time**
0.053 ppm	Annual	Same as Primary	
0.100 ppm	1-Hour	None	

Table 3.4: Regulations set by the Health Department in México

Contaminant	Short Term Exposure		Long Term Exposure	NOM #
	Concentration and average time	**Acceptable maximum frequency**	**(For protection of susceptible population)**	
Nitrogen Dioxide (NO_2)	0.21 ppm (1 hour)	Once a year	-	NOM-023-SSA1-1993

Chapter 4: Methodology

This chapter explains the methodology that was followed throughout this study. It includes a description of the monitoring period for each location, data collection for the dependent variables, a full description of each variable and finally a description of the statistical analysis that was utilized.

4.1 Description of Monitoring Phase

Monitoring for NO_2 levels was conducted at 27 locations in Ciudad Juárez, 22 schools and 5 homes, from December 2002 to September 2003. Table 4.1 shows a list of the schools and homes along with their corresponding identification number assigned for analysis in this study. To distinguish the homes from the schools, the homes are in bold in Table 4.1. The schools and homes were selected based on the fact that they were within 800 meters of a major paved street with high daily traffic counts according to the data provided by the Municipal Institute of Investigation and Planning, IMIP (Molina, 2005 and Holguin et al., 2007).

Ogawa passive samplers, as shown in Figure 4.1, were used to monitor the NO_2 levels and sent to Harvard School of Public Health in Boston, MA for analysis. The NO_2 Ogawa passive samplers are composed of 2 chambers, each one with the assembly stack up as shown in Figure 4.2. After assembly, the loaded samplers are placed into a re-sealable plastic bag and then the plastic bag is placed into a brown airtight container provided by the company distributing the samplers (Ogawa & Co., 1997). It is then taken to the exposure site, and after the exposure period, the same procedure is followed to take the sampler for analysis. Since the samplers were analyzed at an offsite laboratory, they were put in a polypropylene air-tight container and refrigerated at a temperature of 5°C. According to the Ogawa Sampling Protocol (Ogawa & Co., 1997), it is mandatory to protect the monitor from moisture while it is being exposed during the sampling period. To accomplish this task while still retaining proper airflow, it is necessary to use an opaque shelter with a sampler bracket as shown in Figure 4.3. The mounting bracket was then attached to the site by passing a wire through two holes on each side of the bracket and then around the pole or post.

There were a total of 11 monitoring periods from December 2002 to September 2003, and each one consisted of 12 days ±4 hours as shown in Table 4.3. Almost every month had a sampling period, except for January and July 2003. The number of monitoring periods, shown in Table 4.2, varied at each location and ranged from 11 to 3 periods, with an average of 7 monitoring periods per location. One final, average NO_2 concentration was calculated for each of the 27 locations from the concentrations of each monitoring period, at each location. Figures 4.4-4.30 show the variation of the different concentrations that were obtained for each of the schools at each monitoring period. Figure 4.31 shows the location within Ciudad Juárez of the schools and homes that were monitored.

Table 4.1: Elementary schools/homes monitored

Identification number	**Location name**
1	Maclovio Herrera
2	Carmen Serdan
3	Margarita Maza
4	Soledad Herrera Villa
5	Lázaro Cárdenas
6	**Home**
7	Aquiles Serdán
8	Ramona Soto de González
9	José Vasconcelos
10	Gabino Barreda
11	**Familia Jiménez (Home)**
12	Ma. Guadalupe Brena Ponce
13	**Familia Corona (Home)**
14	Francisco Márquez
15	Francisco Villa
16	Francisco Matus Micelli
17	Niñez Mexicana
18	**Home**
19	Teófila Borunda
20	**Home**
21	San Vicente de Paul
22	Hortensia Solís Ontiveros
23	Oscar Flores Sánchez
24	Raramuri
25	Manuel Primo Corral
26	Ricardo Flores Sánchez
27	Ma. Olivia Cárdenas Reyes

Table 4.2: Number of monitoring periods for each location

Identification Number	Total Twelve-Day Monitoring Periods
1	4
2	8
3	5
4	9
5	3
6	5
7	10
8	5
9	9
10	11
11	6
12	7
13	5
14	3
15	5
16	9
17	8
18	5
19	5
20	3
21	11
22	6
23	9
24	6
25	10
26	6
27	7

Table 4.3: Dates for monitoring periods

Monitoring Phase	**Dates**
1	12/02/2002-12/14/2002
2	02/03/2003-02/15/2003
3	02/17/2003-03/01/2003
4	03/05/2003-03/15/2003
5	03/17/2003-03/29/2003
6	03/31/2003-04/12/2003
7	05/05/2003-05/17/2003
8	05/19/2003-05/31/2003
9	06/02/2003-06/14/2003
10	08/18/2003-08/30/2003
11	09/01/2003-09/13/2003

Figure 4.1: Ogawa passive sampler

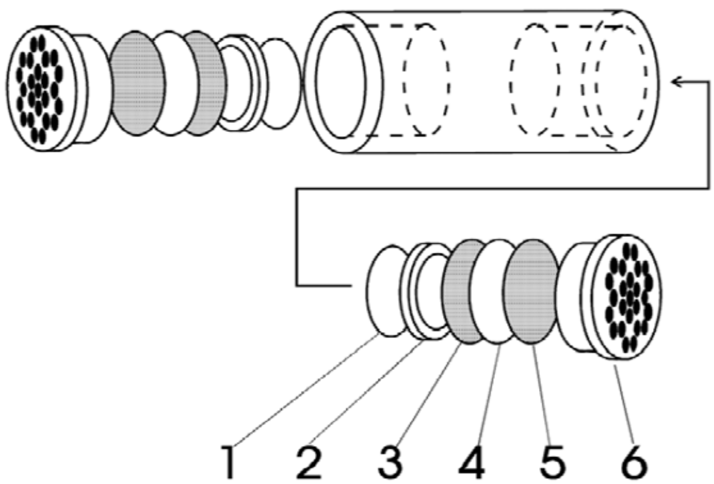

Figure 4.2: Ogawa sampler assembly: 1) solid pad 2) pad retaining ring 3) stainless screen 4) coated collection filter 5) stainless screen 6) diffuser end cap

Figure 4.3: Ogawa passive sampler with shelter for moisture protection

Source for Fig. 4.1-4.3: Ogawa sampling protocol

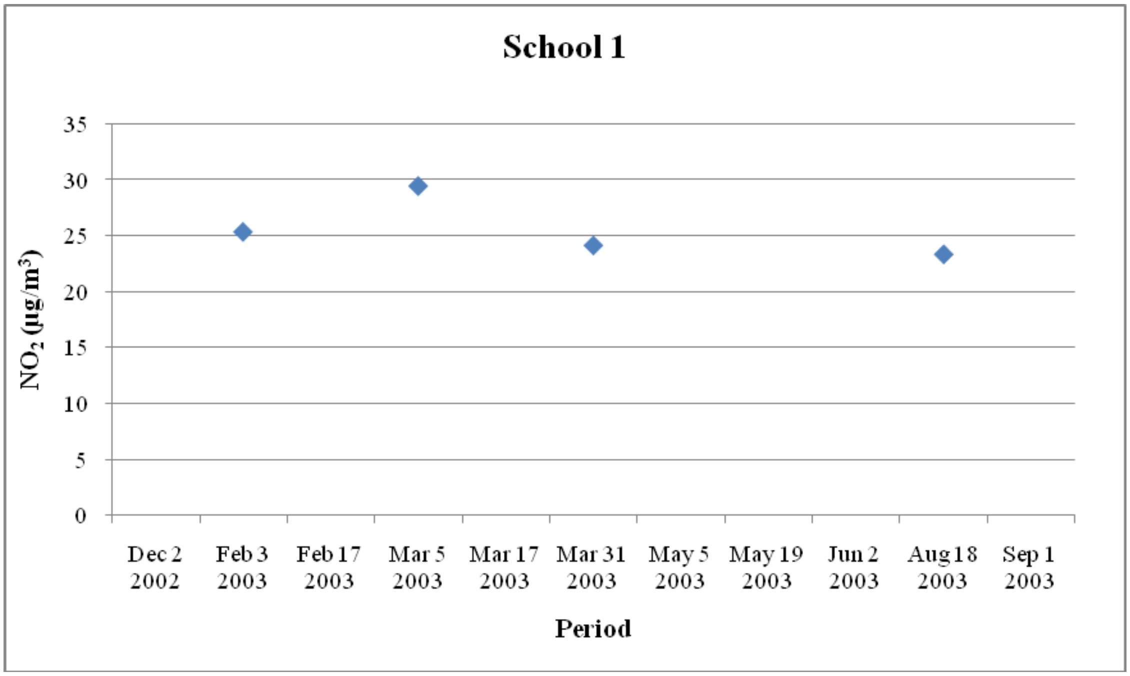

Figure 4.4: NO_2 concentrations from each monitoring period at School 1

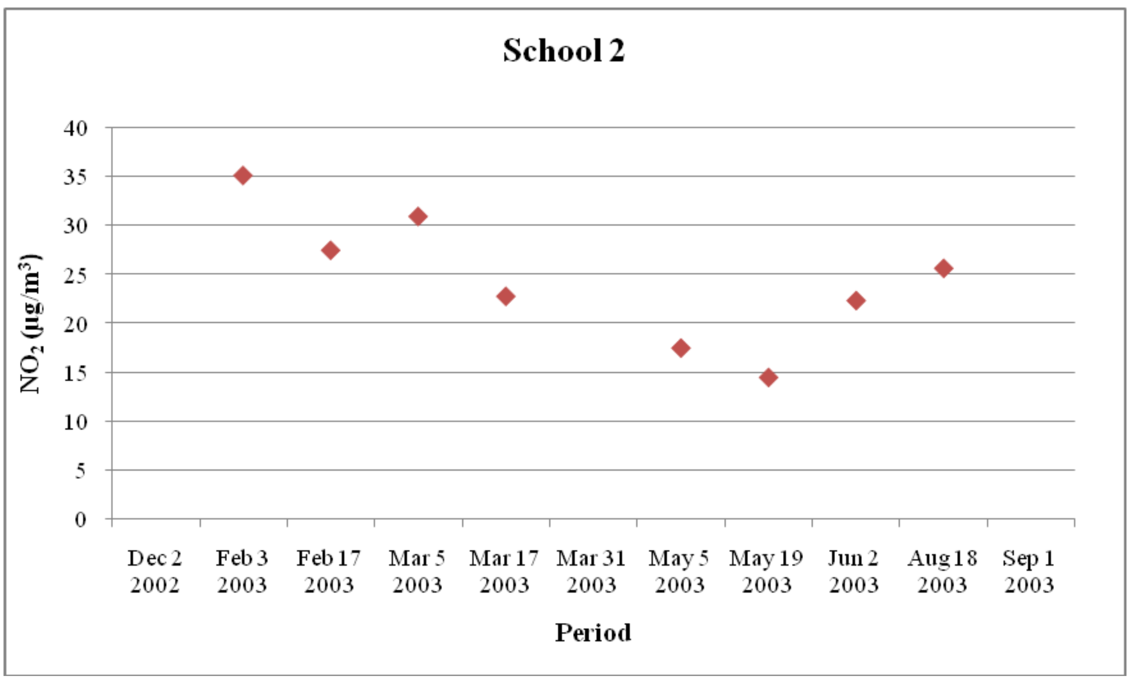

Figure 4.5: NO_2 concentrations from each monitoring period at School 2

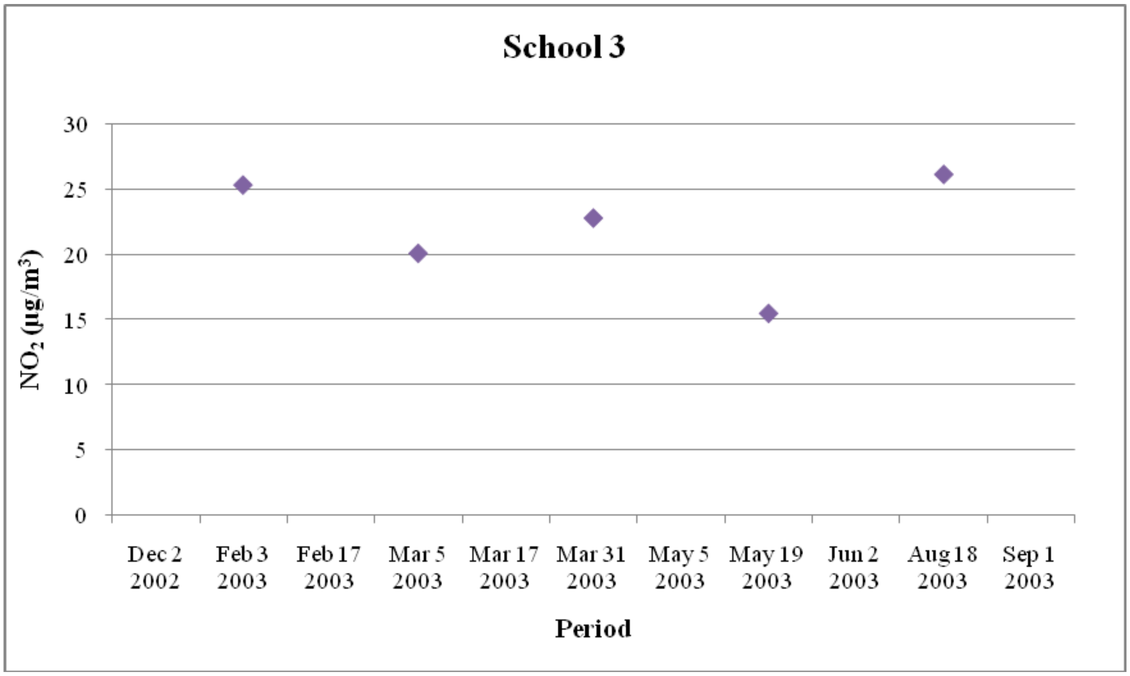

Figure 4.6: NO_2 concentrations from each monitoring period at School 3

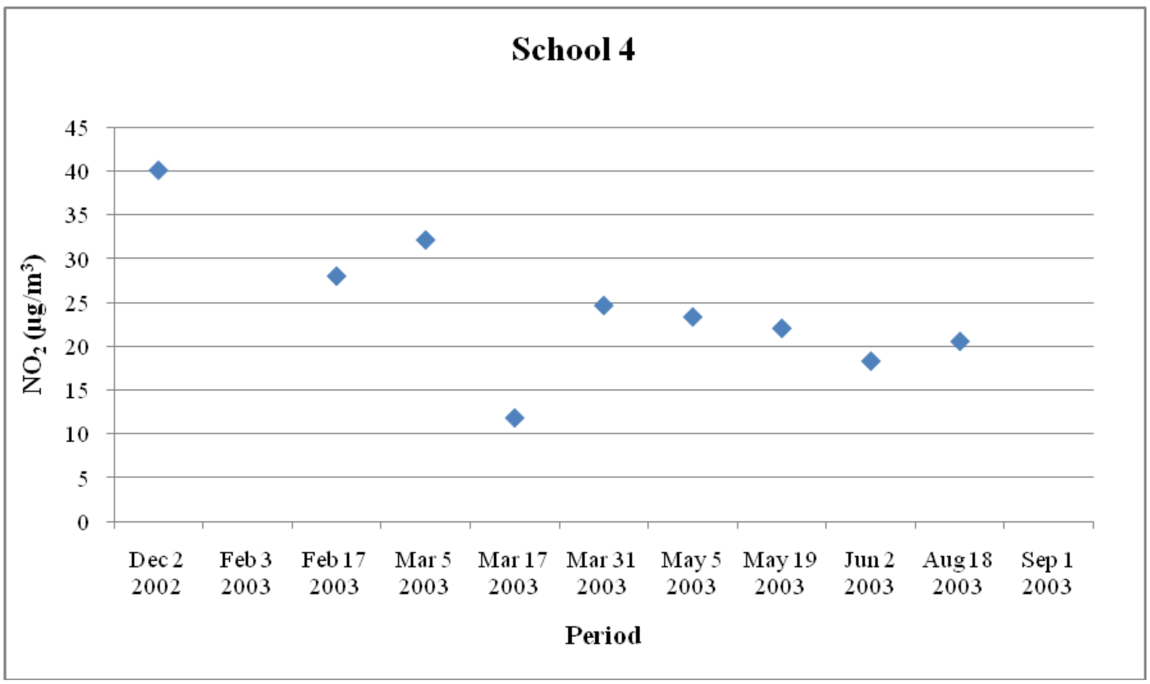

Figure 4.7: NO_2 concentrations from each monitoring period at School 4

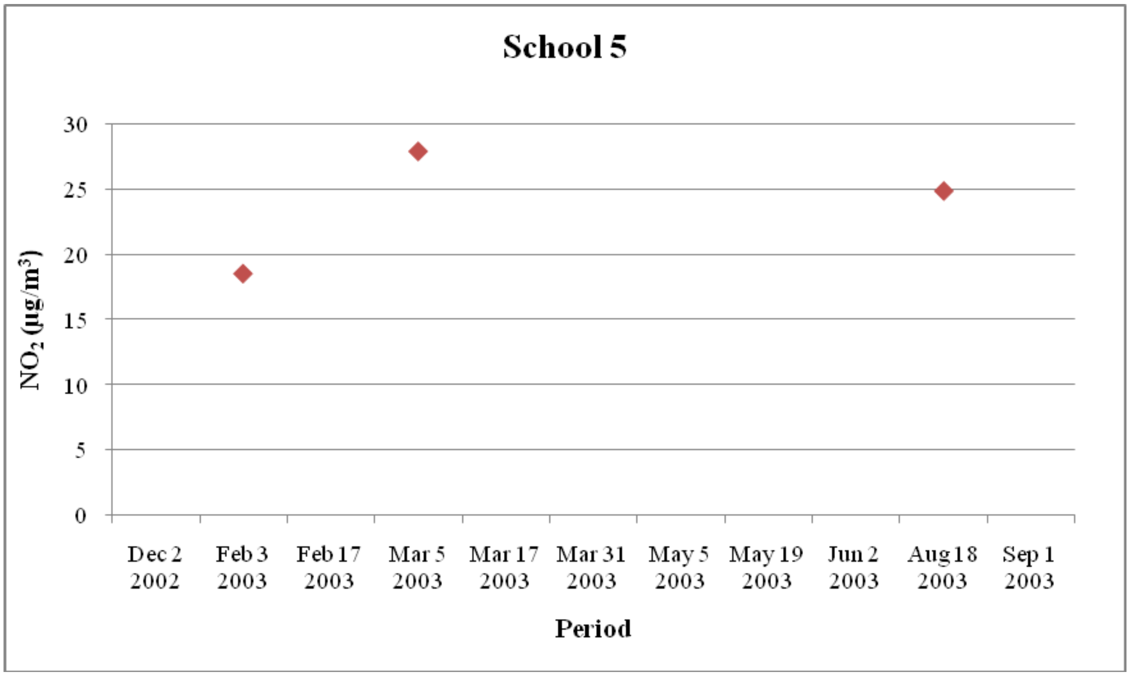

Figure 4.8: NO_2 concentrations from each monitoring period at School 5

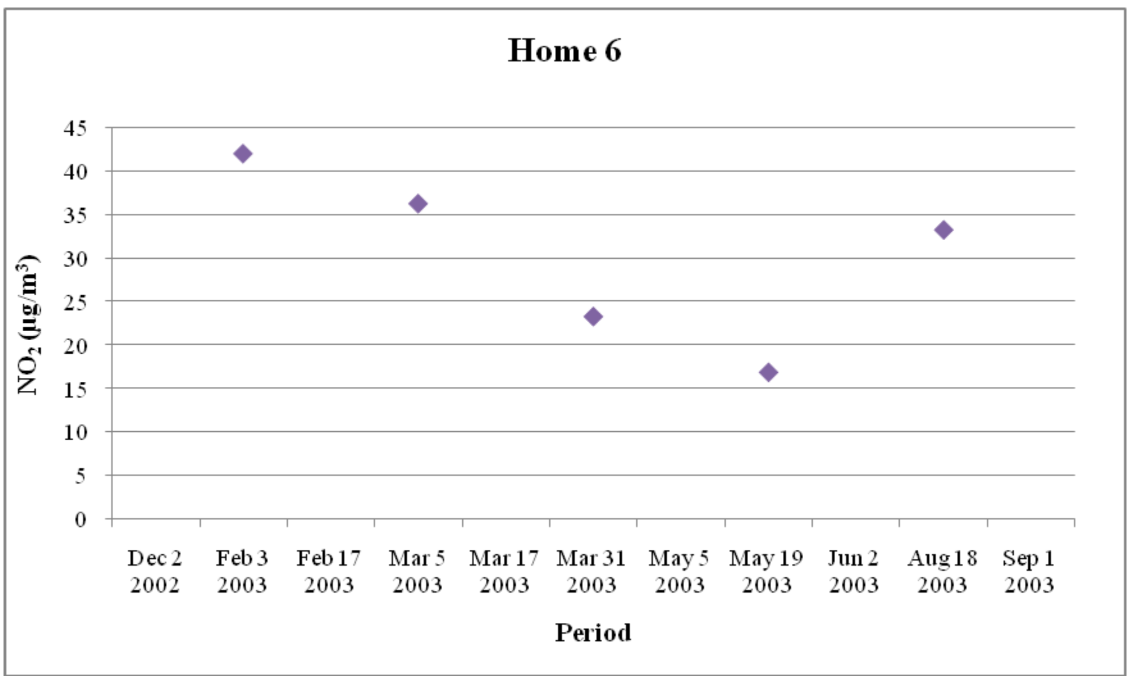

Figure 4.9: NO_2 concentrations from each monitoring period at Home 6

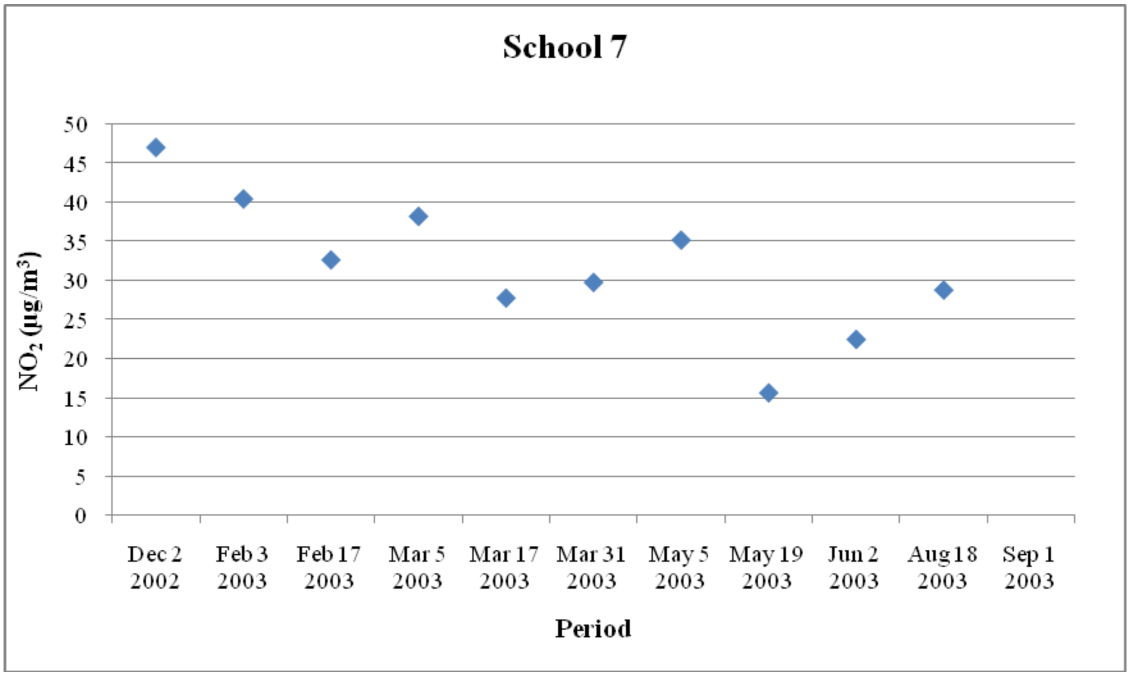

Figure 4.10: NO_2 concentrations from each monitoring period at School 7

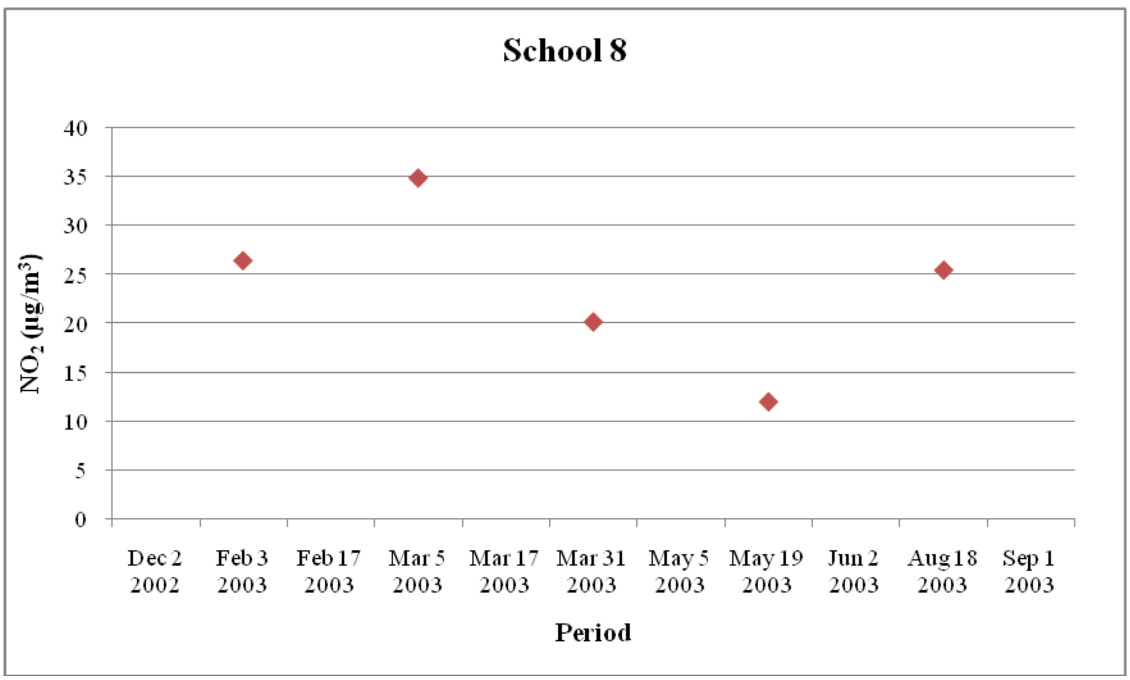

Figure 4.11: NO_2 concentrations from each monitoring period at School 8

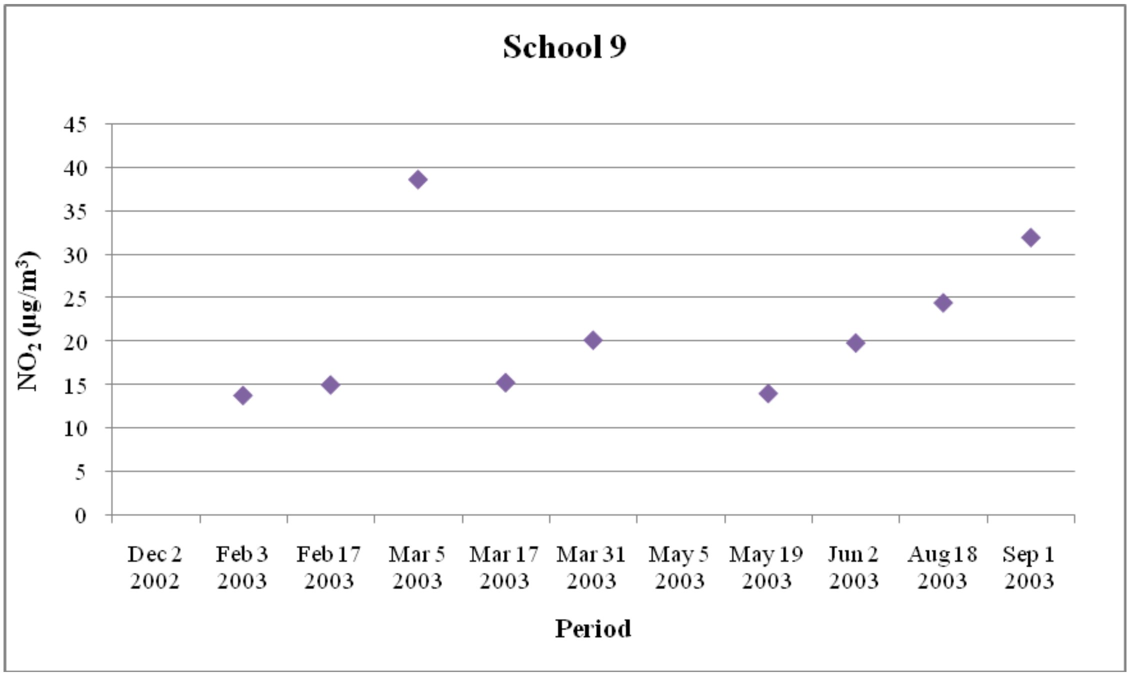

Figure 4.12: NO_2 concentrations from each monitoring period at School 9

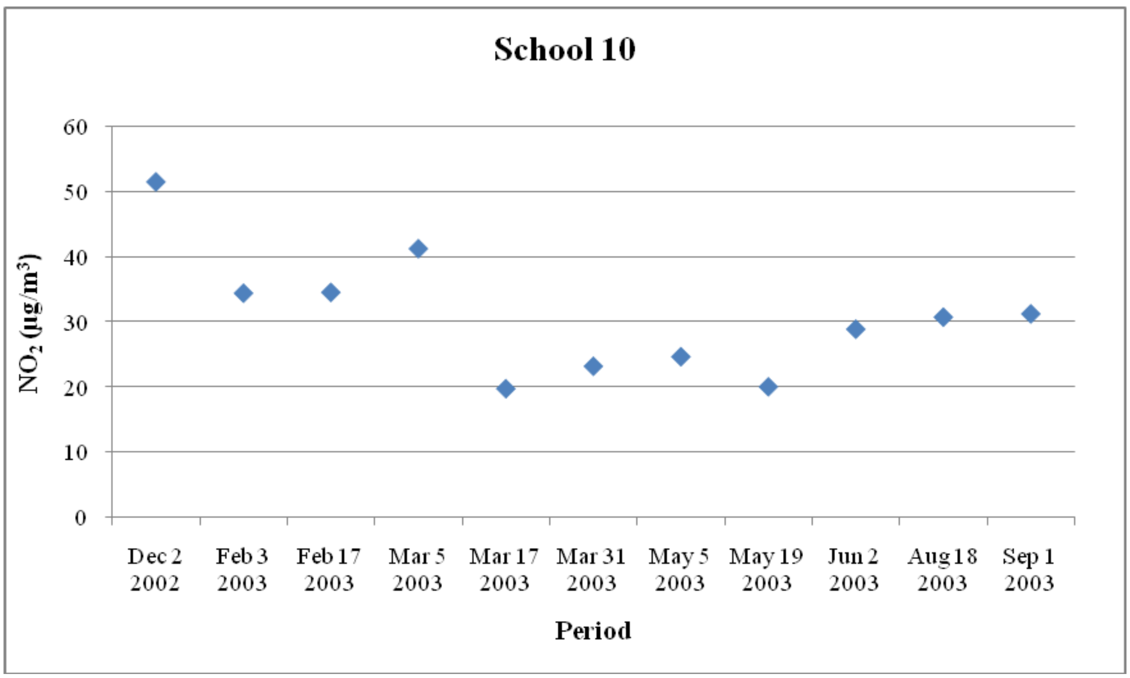

Figure 4.13: NO_2 concentrations from each monitoring period at School 10

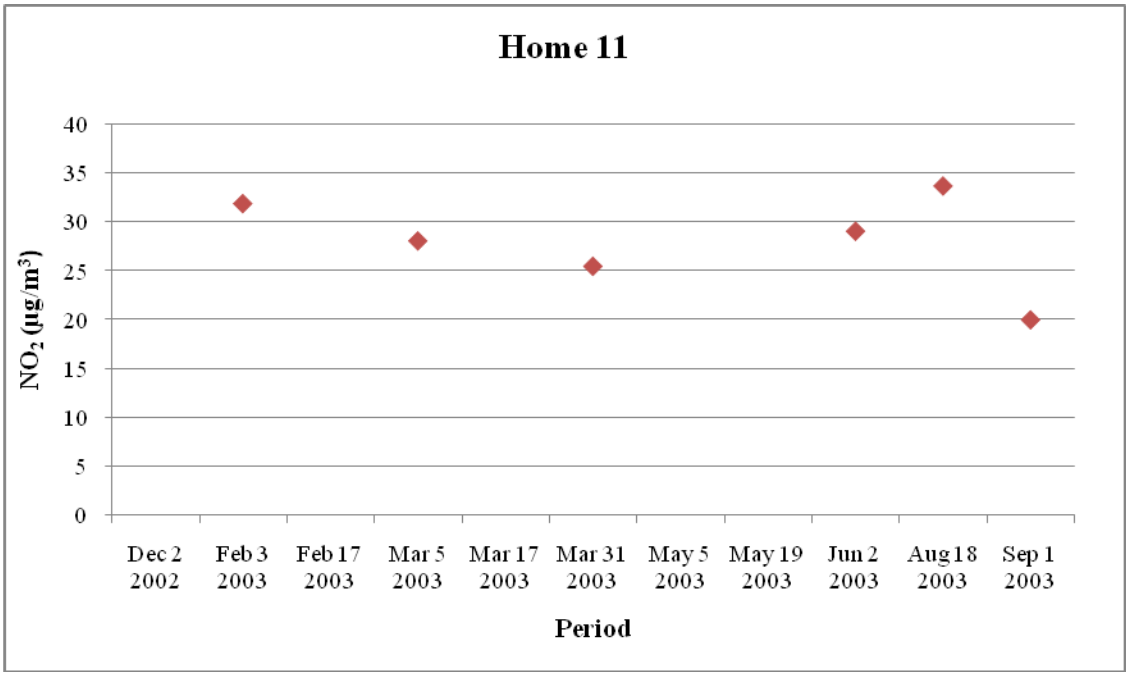

Figure 4.14: NO_2 concentrations from each monitoring period at Home 11

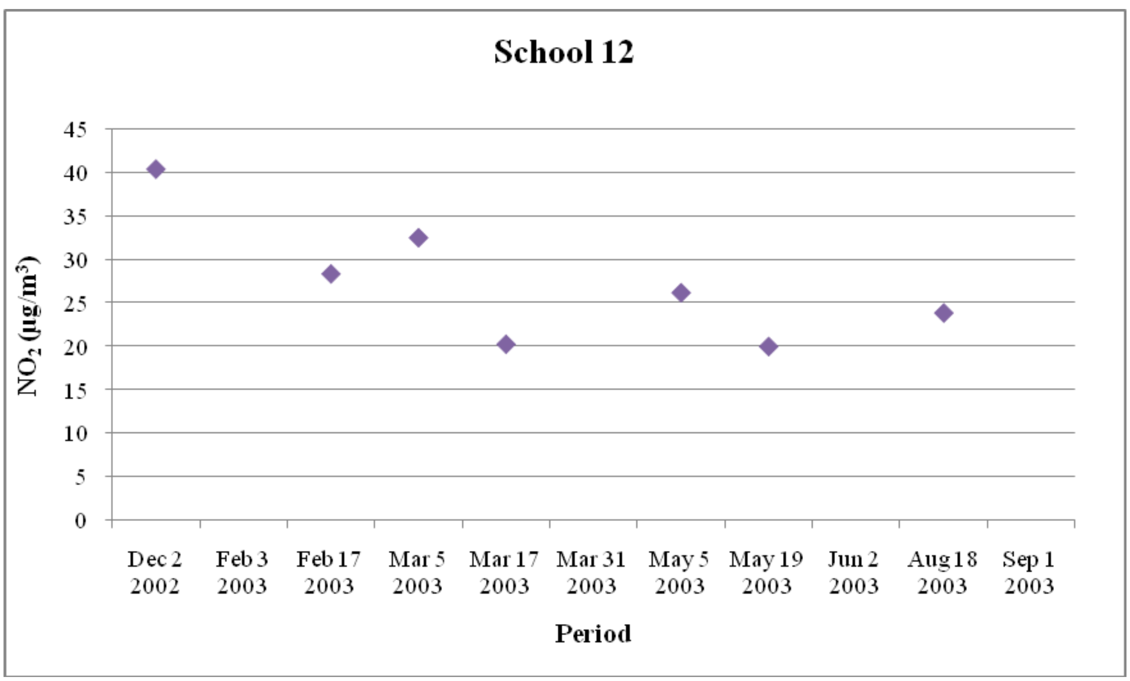

Figure 4.15: NO_2 concentrations from each monitoring period at School 12

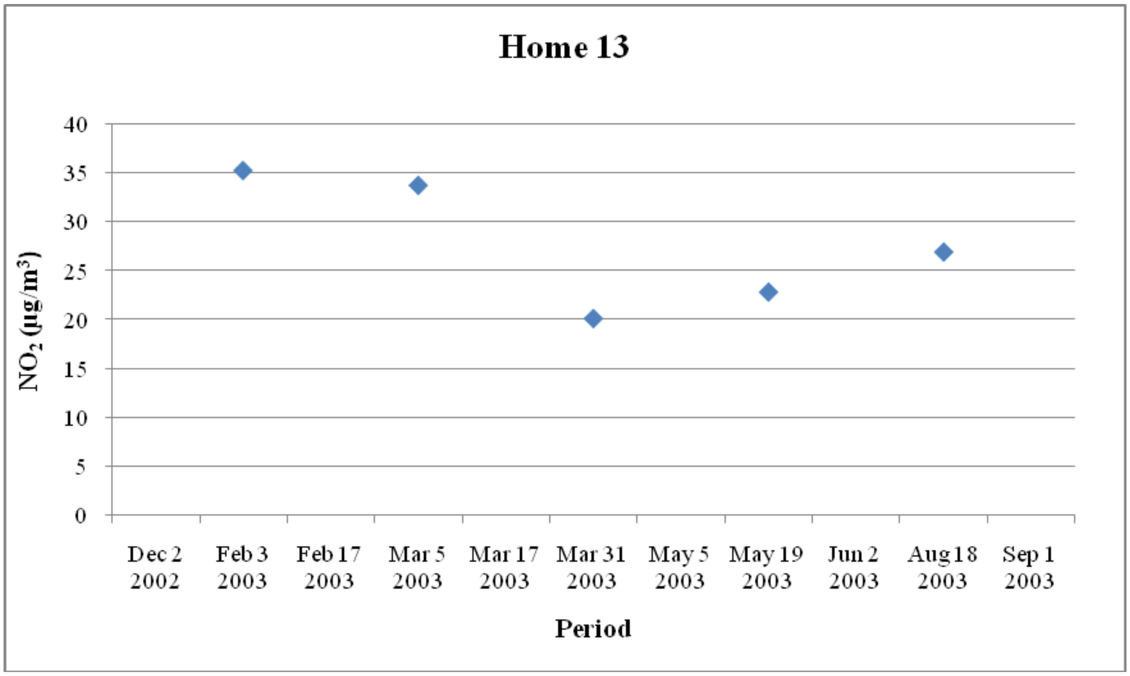

Figure 4.16: NO_2 concentrations from each monitoring period at Home 13

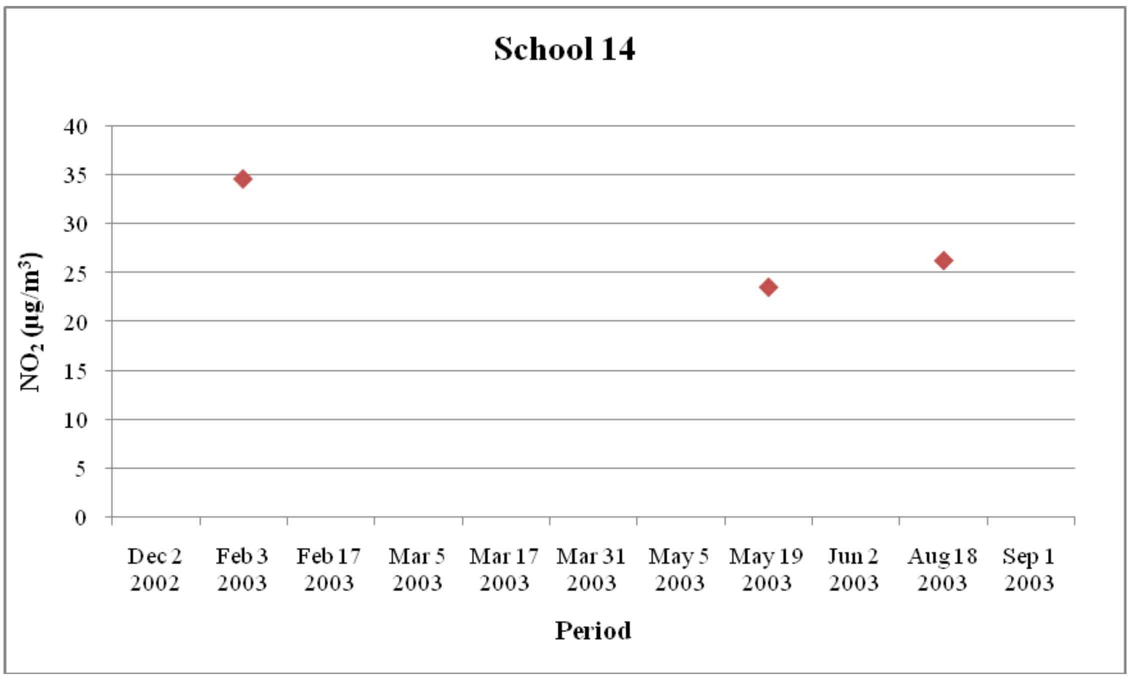

Figure 4.17: NO_2 concentrations from each monitoring period at School 14

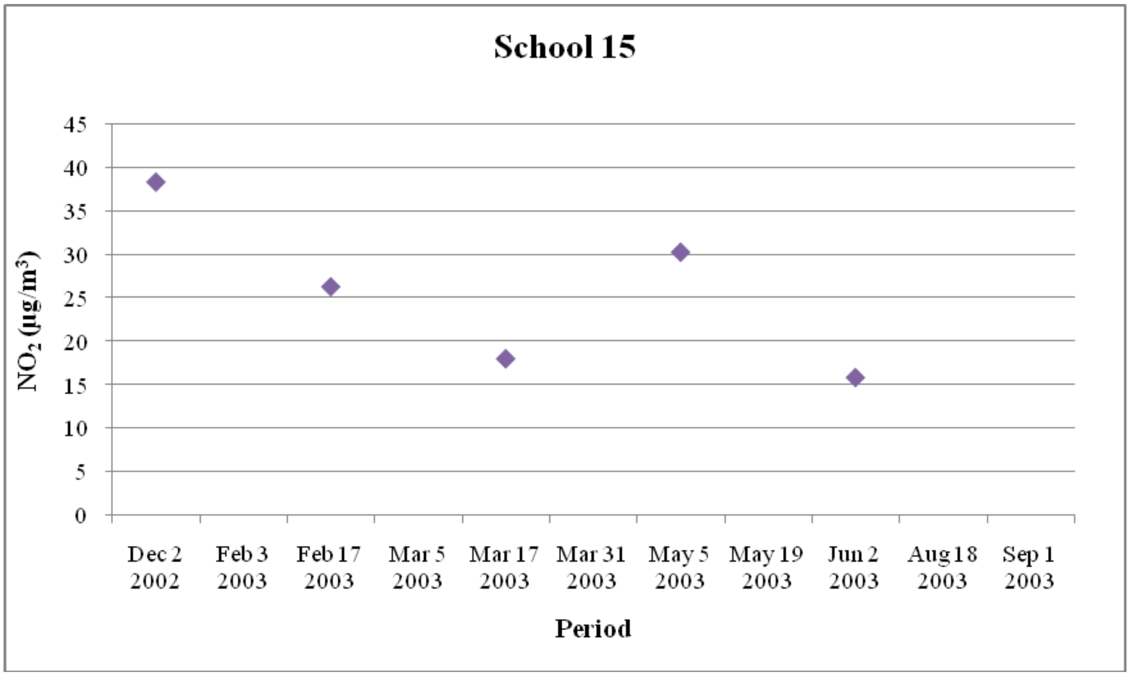

Figure 4.18: NO_2 concentrations from each monitoring period at School 15

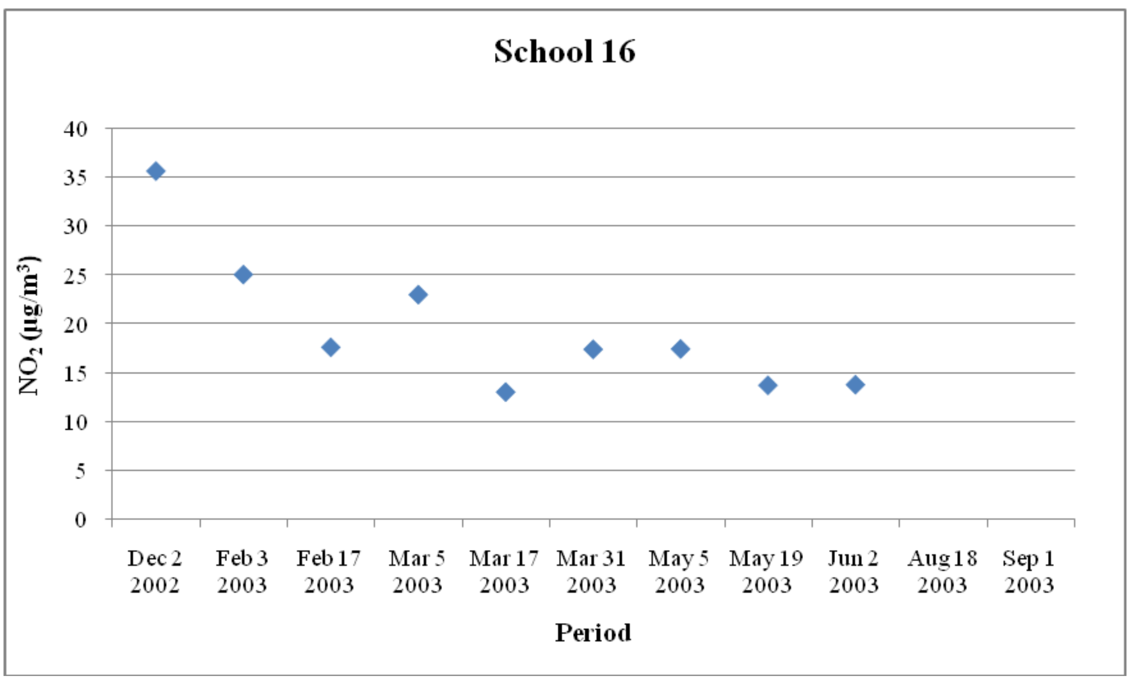

Figure 4.19: NO_2 concentrations from each monitoring period at School 16

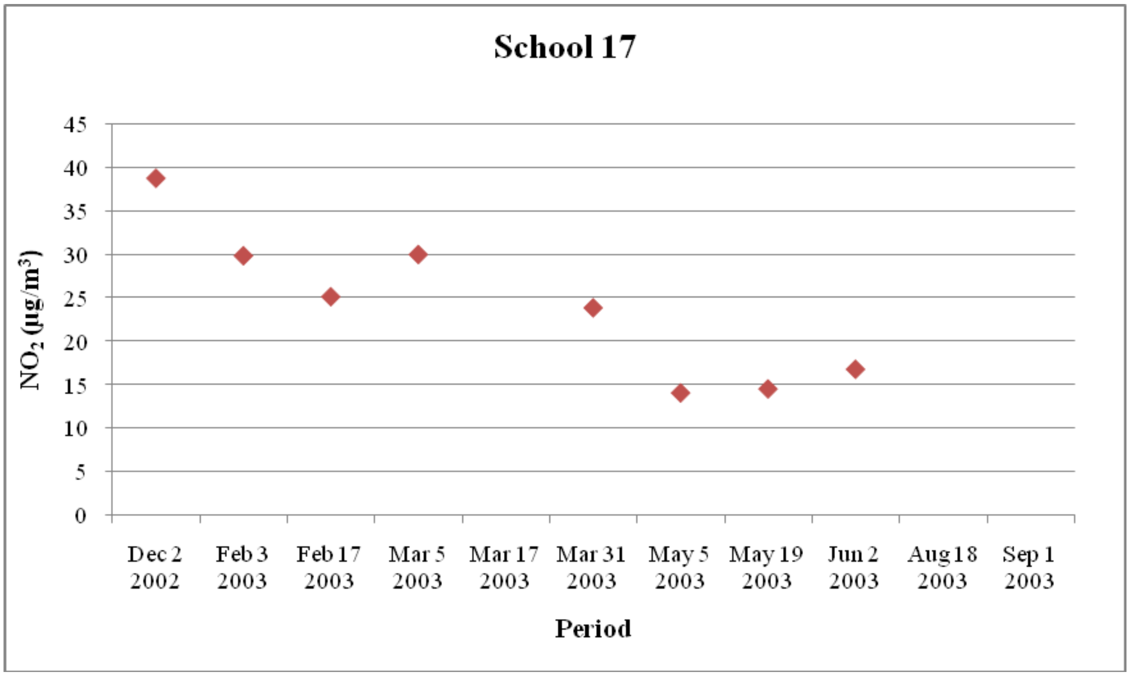

Figure 4.20: NO_2 concentrations from each monitoring period at School 17

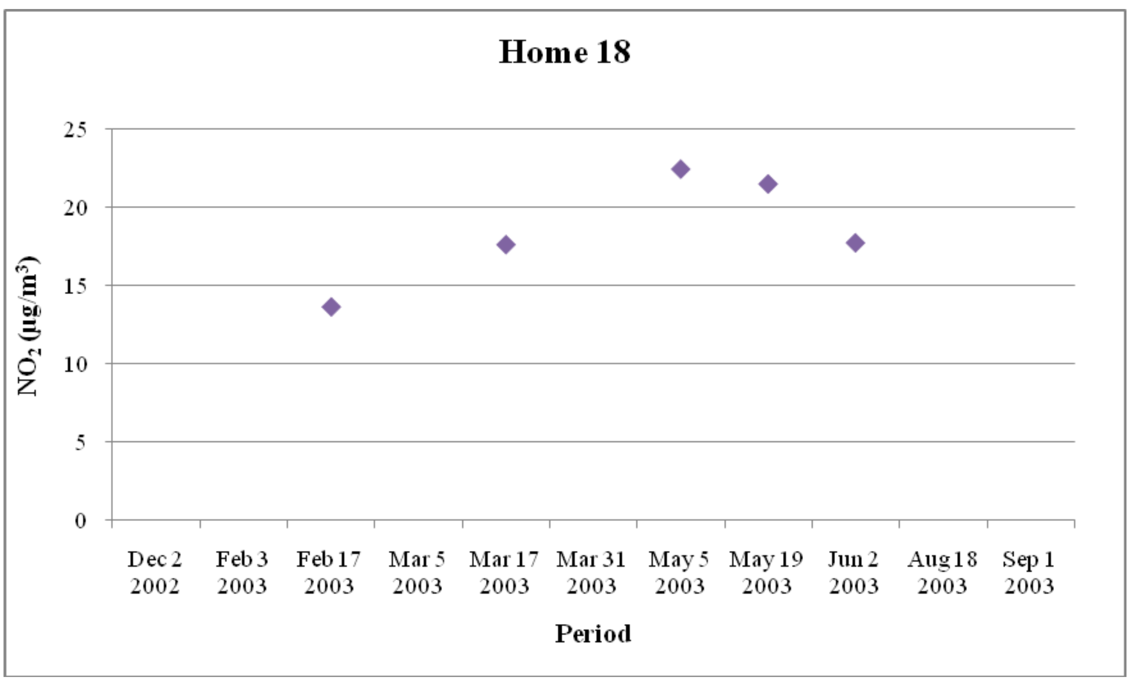

Figure 4.21: NO_2 concentrations from each monitoring period at Home 18

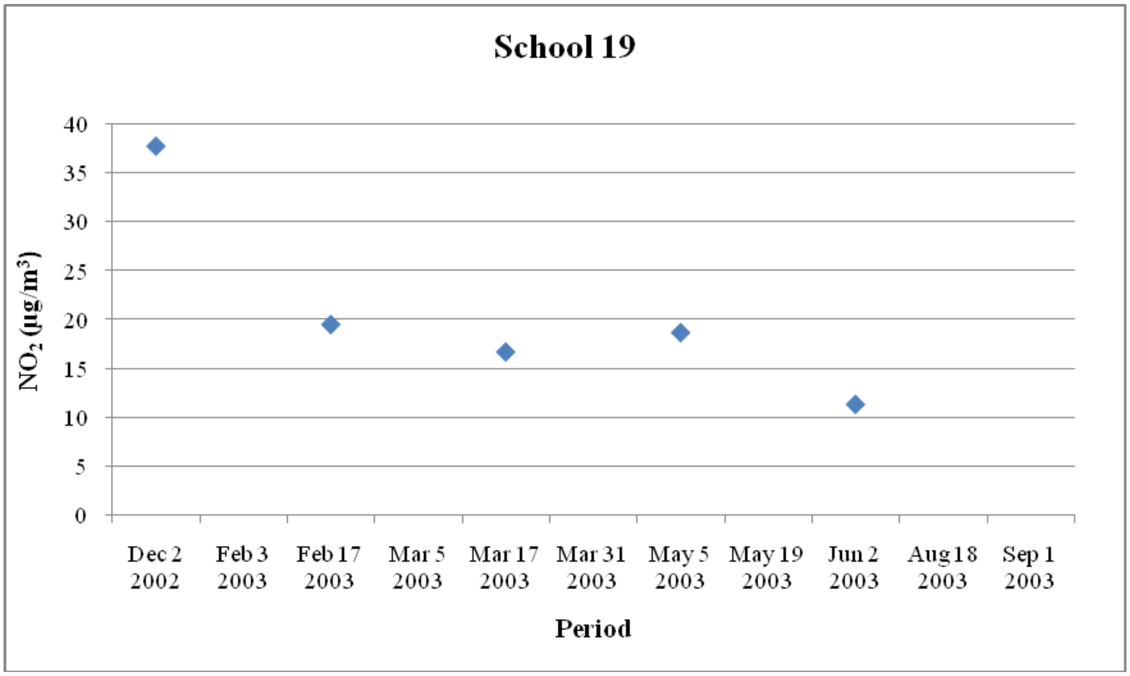

Figure 4.22: NO_2 concentrations from each monitoring period at School 19

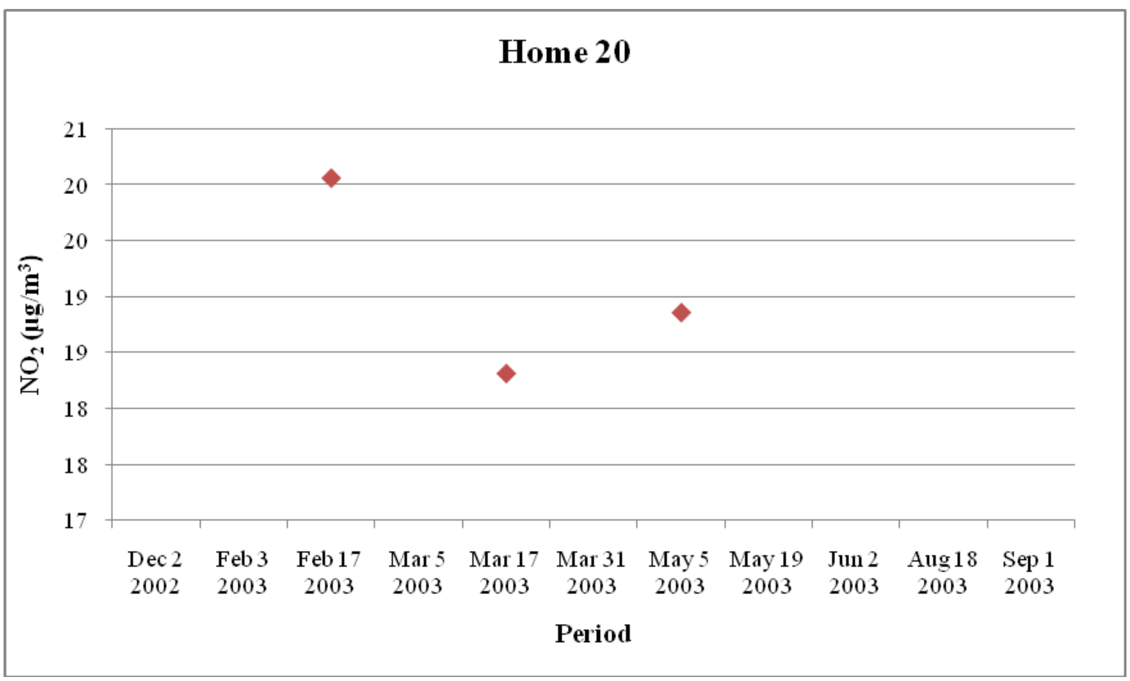

Figure 4.23: NO_2 concentrations from each monitoring period at Home 20

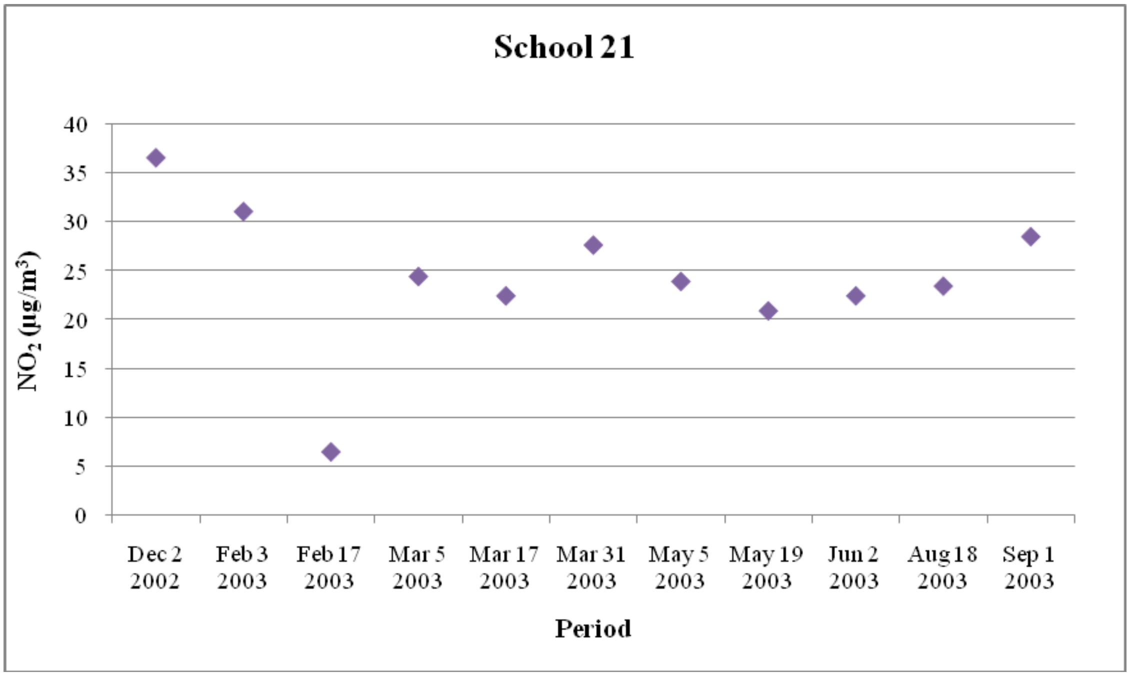

Figure 4.24: NO_2 concentrations from each monitoring period at School 21

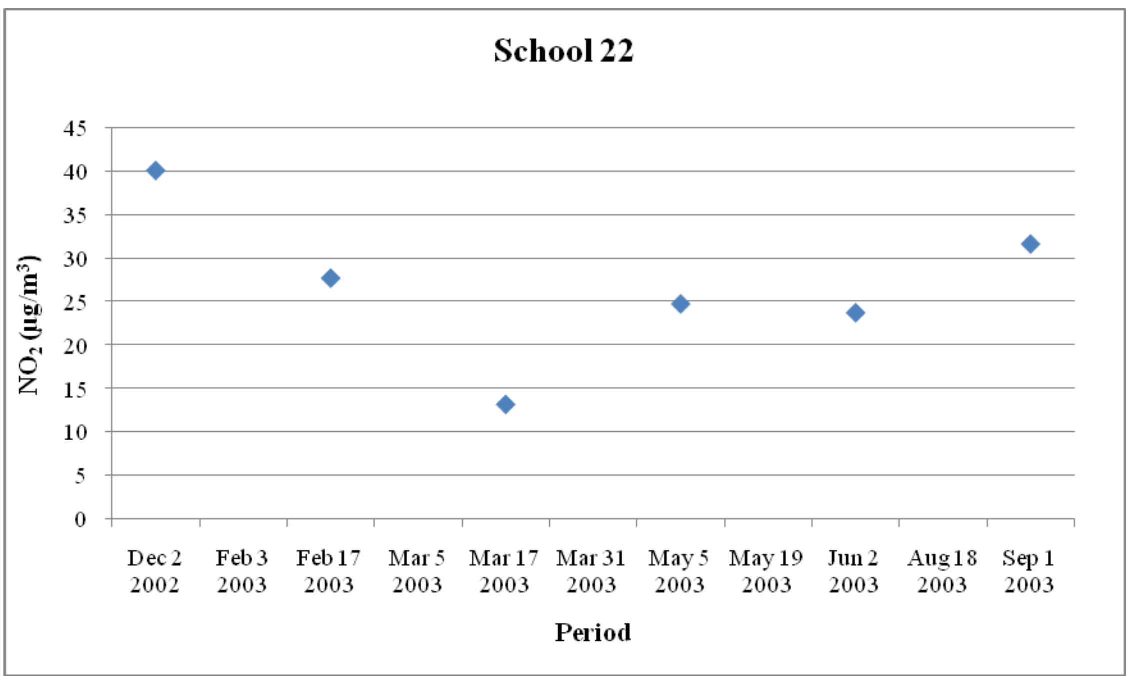

Figure 4.25: NO_2 concentrations from each monitoring period at School 22

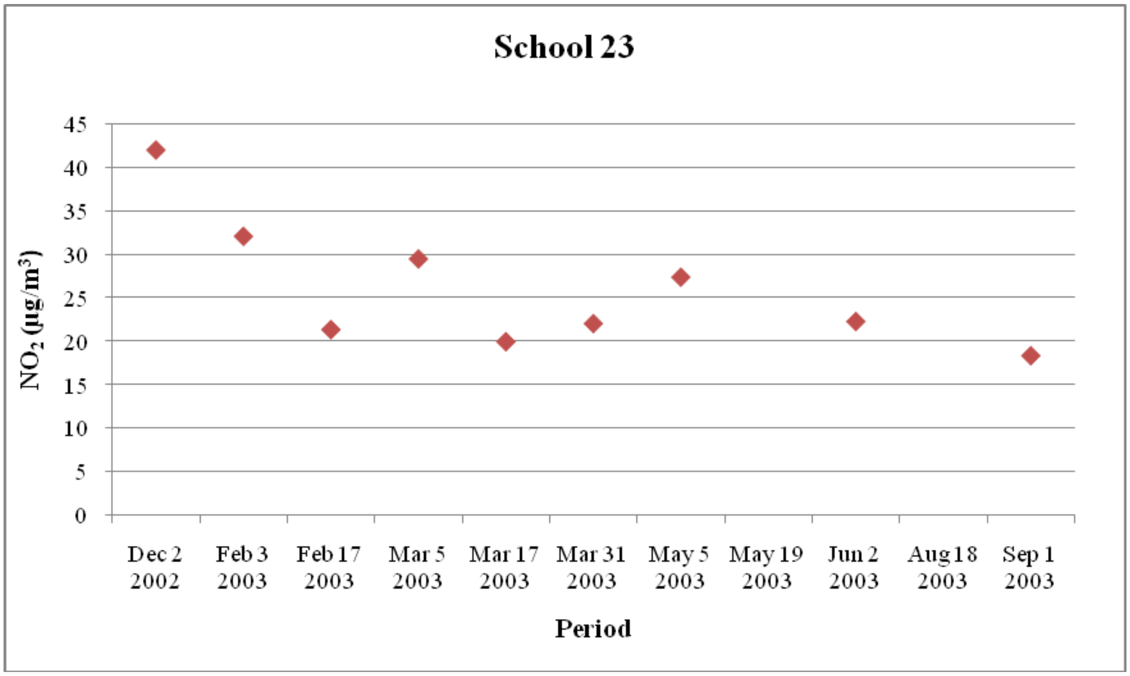

Figure 4.26: NO_2 concentrations from each monitoring period at School 23

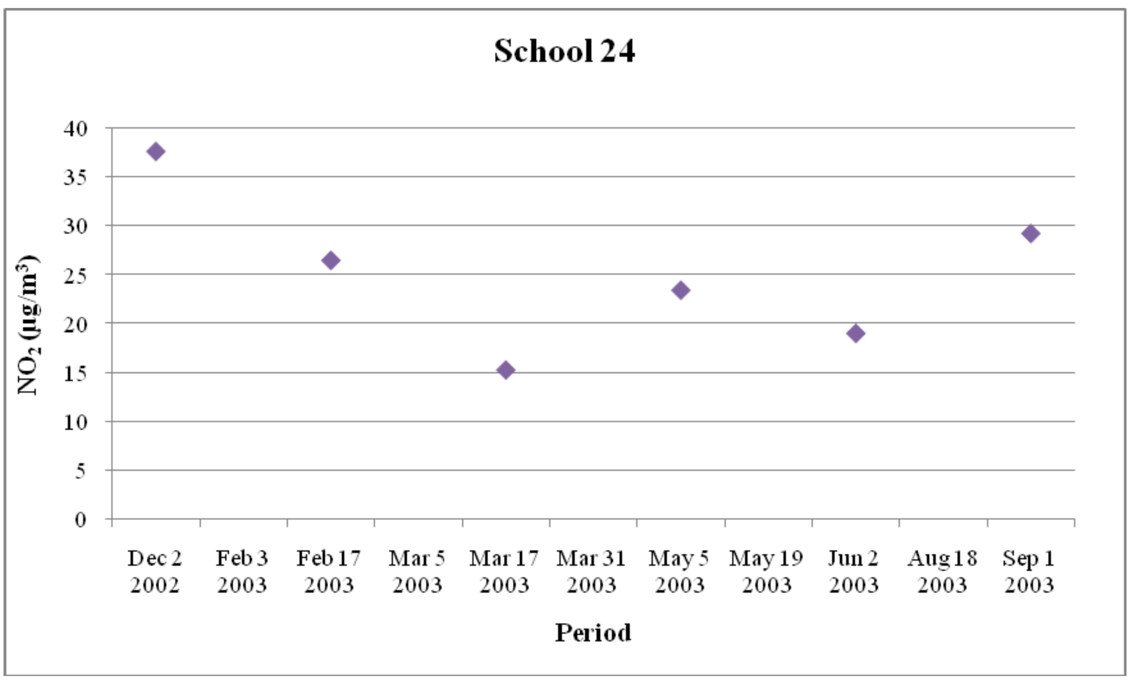

Figure 4.27: NO_2 concentrations from each monitoring period at School 24

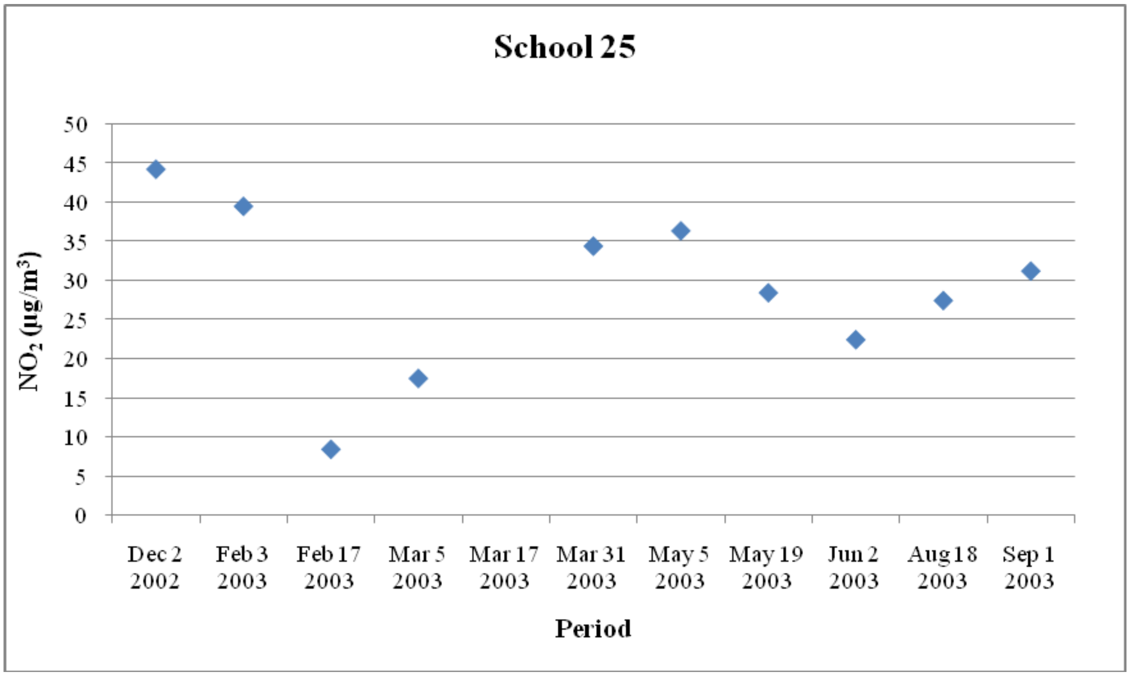

Figure 4.28: NO_2 concentrations from each monitoring period at School 25

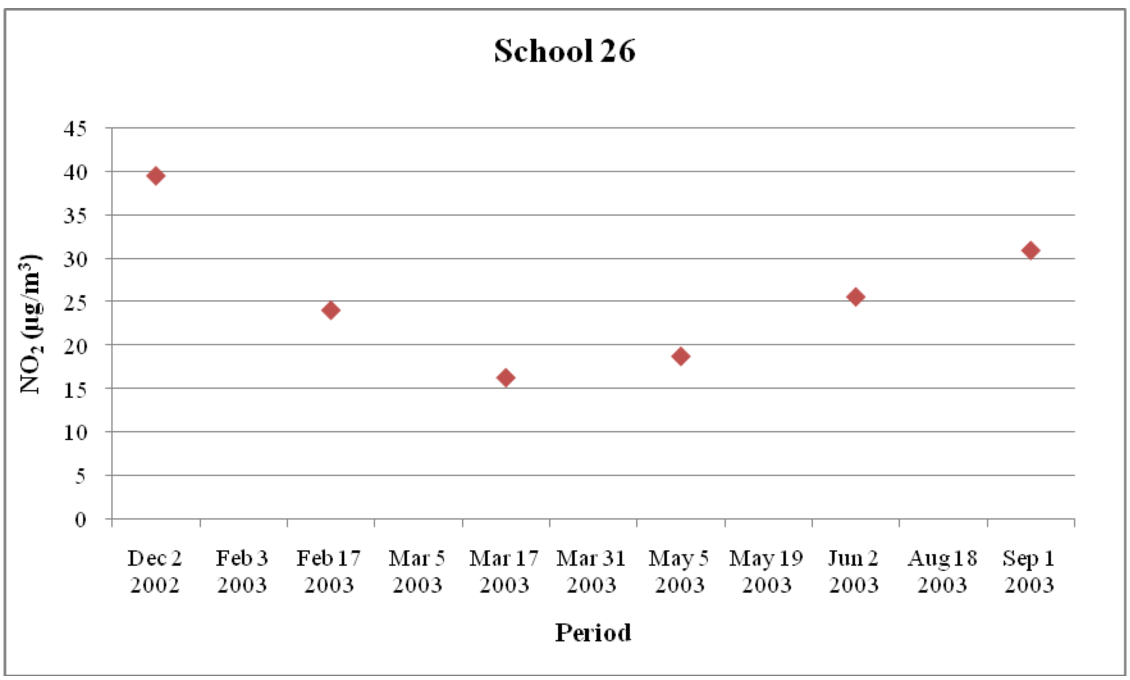

Figure 4.29: NO_2 concentrations from each monitoring period at School 26

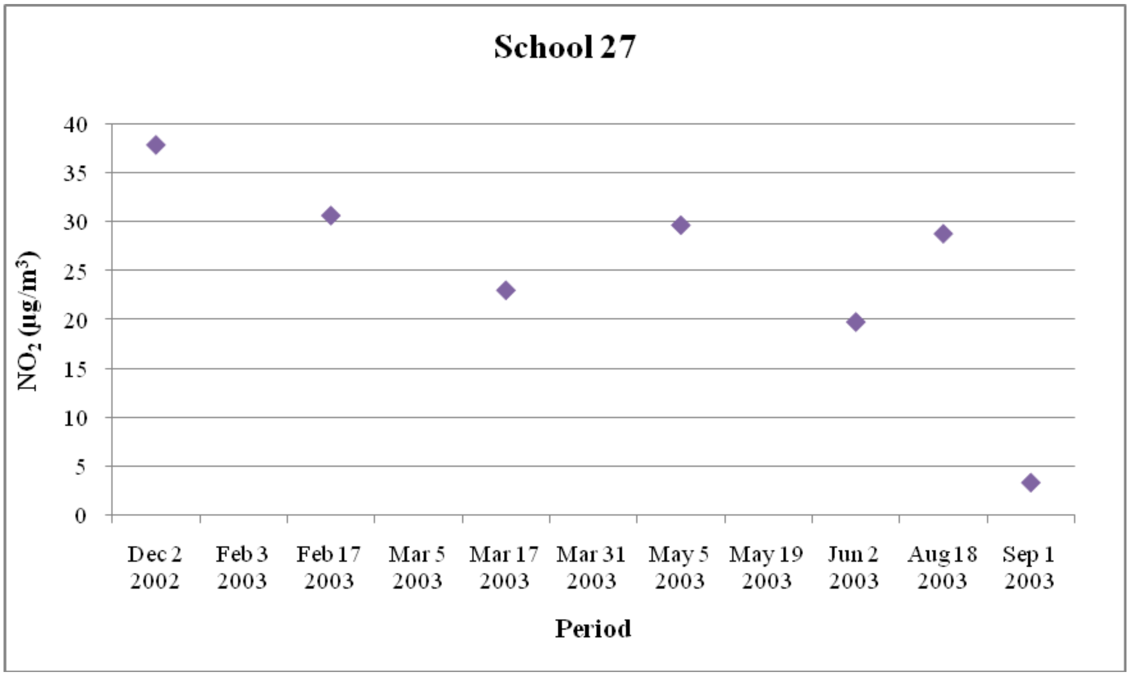

Figure 4.30: NO_2 concentrations from each monitoring period at School 27

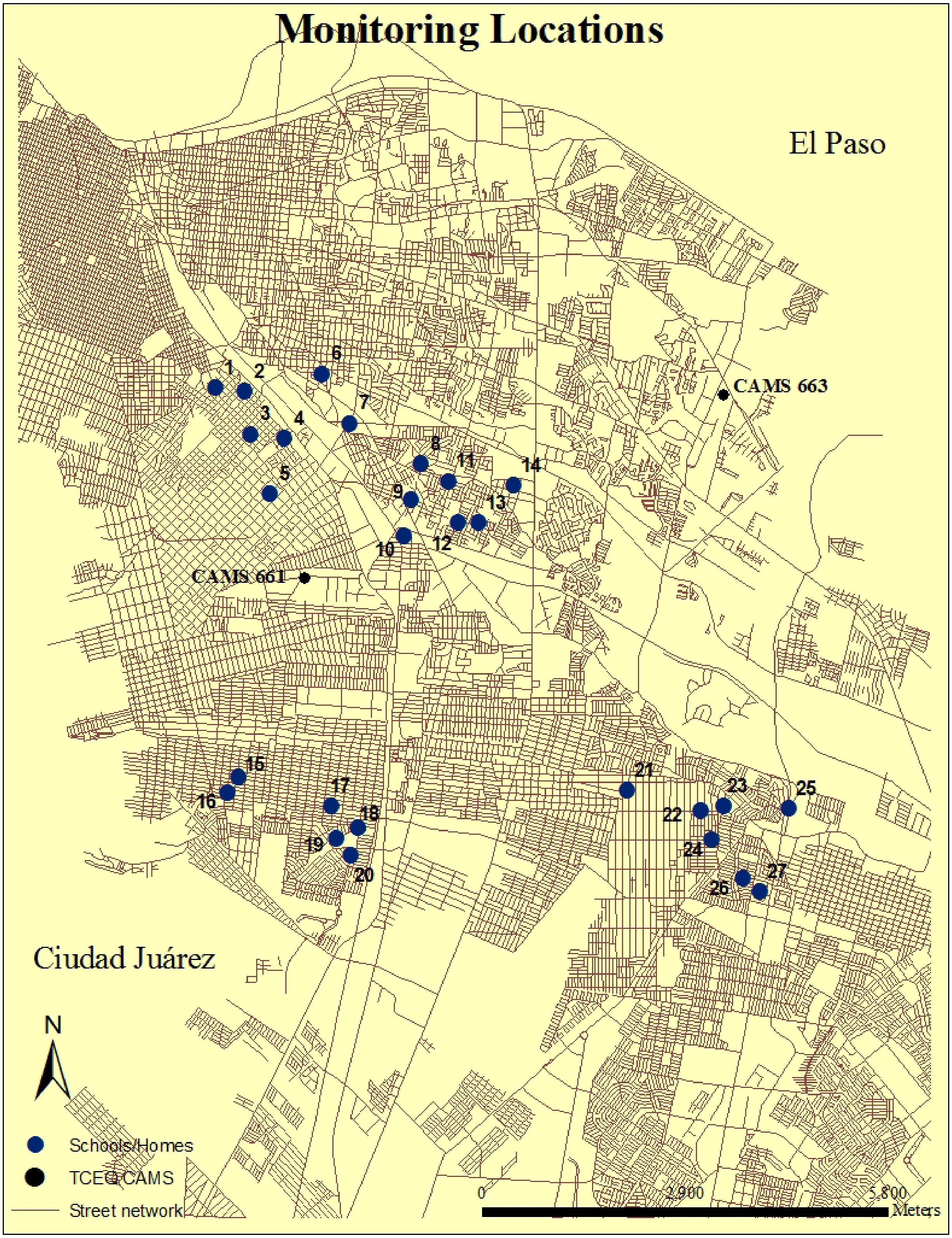

Figure 4.31: Ciudad Juárez area with the location of monitored schools and homes.

4.2 Study Variables

The selection of the main sources of nitrogen dioxide was based upon the information provided by the Ministry of Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales, SEMARNAT) and the Texas Comission on Environmental Quality (TCEQ). According to reports such as PROAIRE 2006-2012 compiled by the SEMARNAT and 2009 Point Source Emission Report from the TCEQ, the main sources for nitrogen oxides in Ciudad Juárez and El Paso include traffic, power plants, electric engine factories, cement plants, and oil refineries (El Paso). In order to identify the influence of the emission sources at the monitored areas, distances from these sources to each of the monitored areas were determined. Distances to point sources, such as distance to the cement plant, the three different major ports of entry located in the El Paso-Ciudad Juárez area, and the electric engine factories were estimated utilizing ArcGIS 9.3.1. Even though ports of entry were not explicitly listed in the reports compiled by the TCEQ or SEMARNAT as main sources for nitrogen oxides, SEMARNAT acknowledges traffic as one of the main sources for nitrogen oxides in Ciudad Juárez, therefore, distance to a port of entry was also considered due to the prolonged periods of idling traffic while awaiting border inspection. Three separate variables were used to establish the distance to each of the three major ports of entry: Bridge of the Americas, Paso del Norte Bridge, and the Zaragoza Bridge. Distance to the oil refinery in El Paso, Texas was also considered as another important source for NO_2 emissions due to its close proximity to the border and its place in the list issued by the Texas Commission on Environmental Quality (TCEQ) regarding it as one of the top contributors of NO_x emissions in El Paso. Although power plants were listed as important emitters of nitrogen oxides, they were not considered in this analysis because the closest power plant providing service to the residents in Ciudad Juárez is located at Samalayuca, 40 km (25 miles) South of Ciudad Juárez. Finally, traffic counts were obtained from the IMIP for the major roads in Ciudad Juárez. The definition of a "major road" varied across different articles encountered in this research. For example, Smith et al., (2006) considered in his study those roads with traffic counts above 10,000 vehicles per day; Ross et al., (2006) considered those roads with at least 50,000 vehicles per day. In this study, major roads were defined as those with daily traffic volumes that ranged from 8,250 vehicles/day to 86,471 vehicles/day.

A total of 12 predictor variables were considered in this study and are shown in Table 4.4. The variables listed include distance to the Bridge of the Americas (DIST_BOTA), distance to the Paso del Norte Bridge (DIST_PDN), distance to the Zaragoza Bridge (DIST_ZAR), distance to the cement plant (DIST_CP), distance to the area with the largest agglomeration of electric engine factories (DIST_EE), distance to the oil refinery in El Paso, Texas, (DIST_OR), traffic volume to the major road with the highest daily traffic volume (TV_1st), distance to the major road with the highest daily traffic volume (DIST_1st), traffic volume to the major road with the second highest daily traffic volume (TV_2nd), and distance to the major road with the second highest daily traffic volume (DIST_2nd). The location of the point sources acting as predictor variables are shown in Figure 4.32

As stated by Ross et al. (2006), most dispersion models suggest that 80 -90 percent of the decay of pollutants occurred within 150-200 meters. However, further research revealed that specifically for NO_2, both traffic counts within 500 meters and location within 1500 meters (downwind) of an expressway were statistically significant predictors of NO_2 concentration. Therefore, the last two predictive variables included in this study consisted of two buffer zones of 500 m and 1000 m around each school as shown in Figures 4.33 and 4.34 respectively. Each of these buffer zones were drawn to calculate the traffic density surrounding each school. Traffic density in each buffer zone was calculated by multiplying the street length (major roads) times the traffic volume and then adding each of those values inside each zone, meaning predictor variable (TD_500) consisted of the sum of the daily traffic volume at the major roads times its corresponding street length inside the 500 m buffer zone; (TD_1000) consisted of the sum of the daily traffic volume at the major roads times its corresponding street length inside the 1000 m buffer zone. An overview of these variables and its values are shown in Table 4.4.

Based on the classification provided by Ryan and LeMasters (2007) for land regression variables, the predictors used in this study would fall under two categories: 1) traffic count (TV_1st, TV_2nd, TD_500, and TD_1000) and 2) land cover, which includes in its category distance to various point sources (DIST_EE, DIST_CP, DIST_OR, DIST_BOTA, DIST_PDN, DIST_ZAR, DIST_1st, and DIST_2nd).

In addition to the land regression variables being considered, this study also considered average wind direction throughout the study period for each monitoring location. For this purpose, wind direction data was obtained from TCEQ CAMS 661 (known as Cd Juárez Advance C661 at El Cid Street in the western part of the city) and TCEQ CAMS 663 (known as Cd Juárez Delphi C663 at De la Industria Avenue in the eastern part of the city). Data from each of the monitoring stations was downloaded every hour for each of the monitoring periods as outlined in Table 4.3. The data was then averaged for each school or home according to their individual monitoring periods. Since wind is a vector, the averaging process consisted of calculating the zonal component (U-component) and the meridional component (V-component) utilizing wind speed (v) and direction (d) and the following,

$$V = -v * \cos(d)$$

$$U = -v * \sin(d)$$

Given the U and V components, the following equation was used to calculate the direction,

$$D = \arctan(U/V) + \theta$$

$$\text{Where, } \theta = 180° \text{ if } V >= 0$$

$$\theta = 0° \text{ if } U < 0 \text{ and } V < 0$$

$$\theta = 360° \text{ if } U >= 0 \text{ and } V < 0$$

(Source of equations: http://www.wrh.noaa.gov/sew/fire/olm/transport.htm)

Finally, a North-South component (N-S) and an East-West component (E-W) was obtained for each school or home by calculating the sine and cosine component of each direction.

The data collected was organized in spreadsheets utilizing Microsoft Excel. The averaging process for the wind direction was conducted in Microsoft Excel. Statistical analysis on the rest of the data was conducted utilizing PASW Statistics 18 (SPSS).

Table 4.4: Summary statistics for land-use variables and dependent variable

Predictor Variable	Mean ± SD	Minimum	Median	Maximum
Sampling sites (n=27)				
Average NO_2 ($\mu g/m^3$)	25.16 ± 3.47	18.55	25.17	31.75
DIST_BOTA[A]	9,375 ± 2,994	5,286	7,597	13,886
DIST_PDN[B]	8,685 ± 3,365	3,849	7,515	14,142
DIST_ZAR[C]	9,981 ± 2,771	4,775	10,554	13,690
DIST_CP[D]	5,819 ± 2,207	1,141	6,474	8,987
DIST_EE[E]	6,661 ± 2,361	3,637	5,517	9,868
DIST_OR[F]	11,256 ± 2,231	8,455	10,256	14,249
TV_1st[G]	45,686 ± 11,662	26,400	42,718	86,471
DIST_1st[H]	459 ± 271	54	459	1,050
TV_2nd[I]	29,455 ± 13,630	8,250	27,300	70,966
DIST_2nd[J]	356 ± 241	34	355	1,033
TD_500[K]	38,698 ± 31,705	0	33,456	108,703
TD_1000[L]	169,820 ± 87,846	39,731	149,693	328,899

[A]: Distance to the Bridge of the Americas (m)

[B]: Distance to the Paso Del Norte Bridge (m)

[C]: Distance to the Zaragoza Bridge (m)

[D]: Distance to the cement plant (m)

[E]: Distance to the area with the largest agglomeration of electric engine factories (m)

[F]: Distance to Oil Refinery (m)

[G]: Daily traffic volume to the major road with the highest traffic volume (veh/day)

[H]: Distance to the major road with the highest daily traffic volume (m)

[I]: Daily traffic volume to the major road with the second highest traffic volume (veh/day)

[J]: Distance to the major road with the second highest daily traffic volume (m)

[K]: sum of the daily traffic volume at the major roads times its corresponding street length inside the 500 m buffer zone (veh-km/day)

[L]: sum of the daily traffic volume at the major roads times its corresponding street length inside the 1000 m buffer zone (veh-km/day)

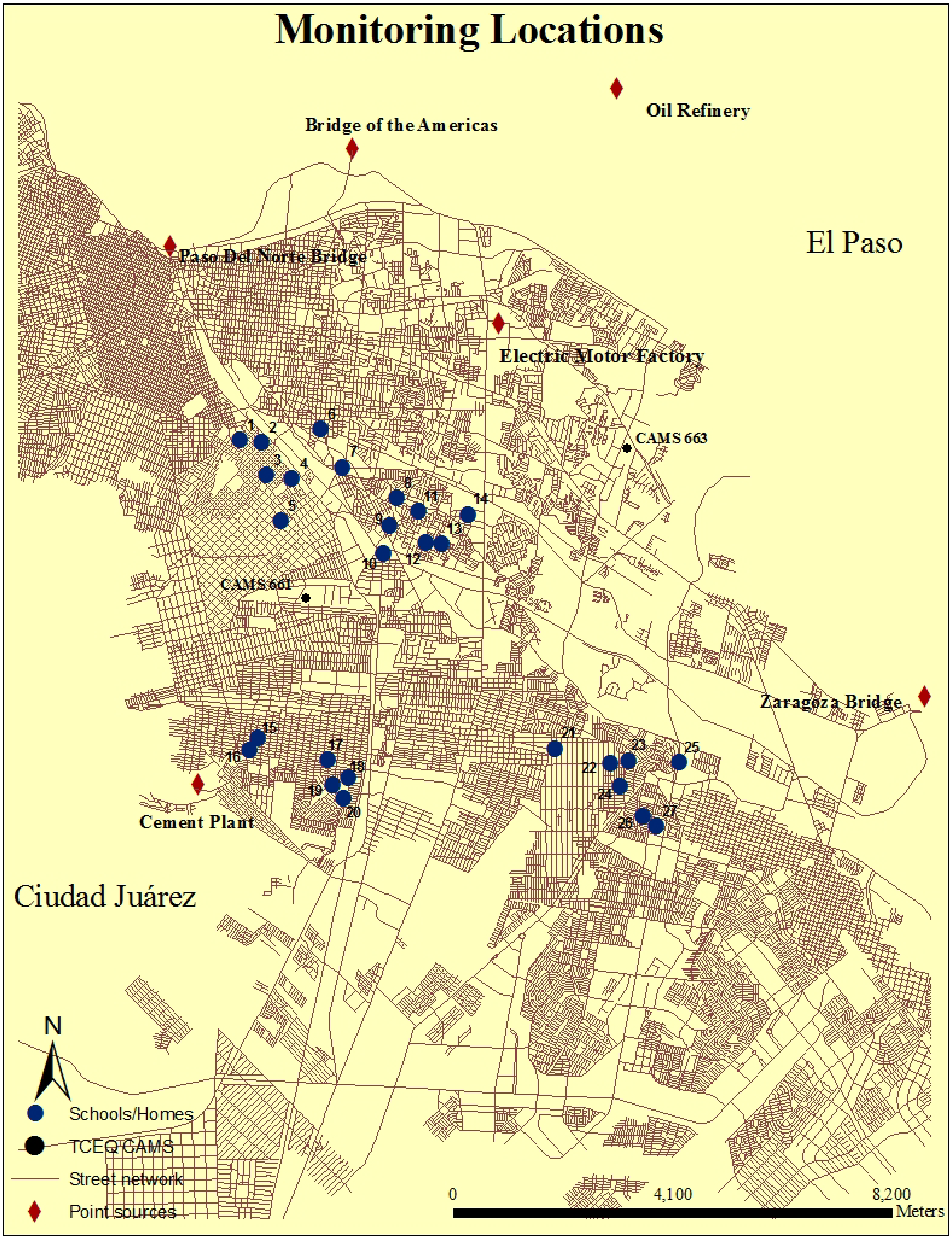

Figure 4.32: Location of point sources for NO_x emissions

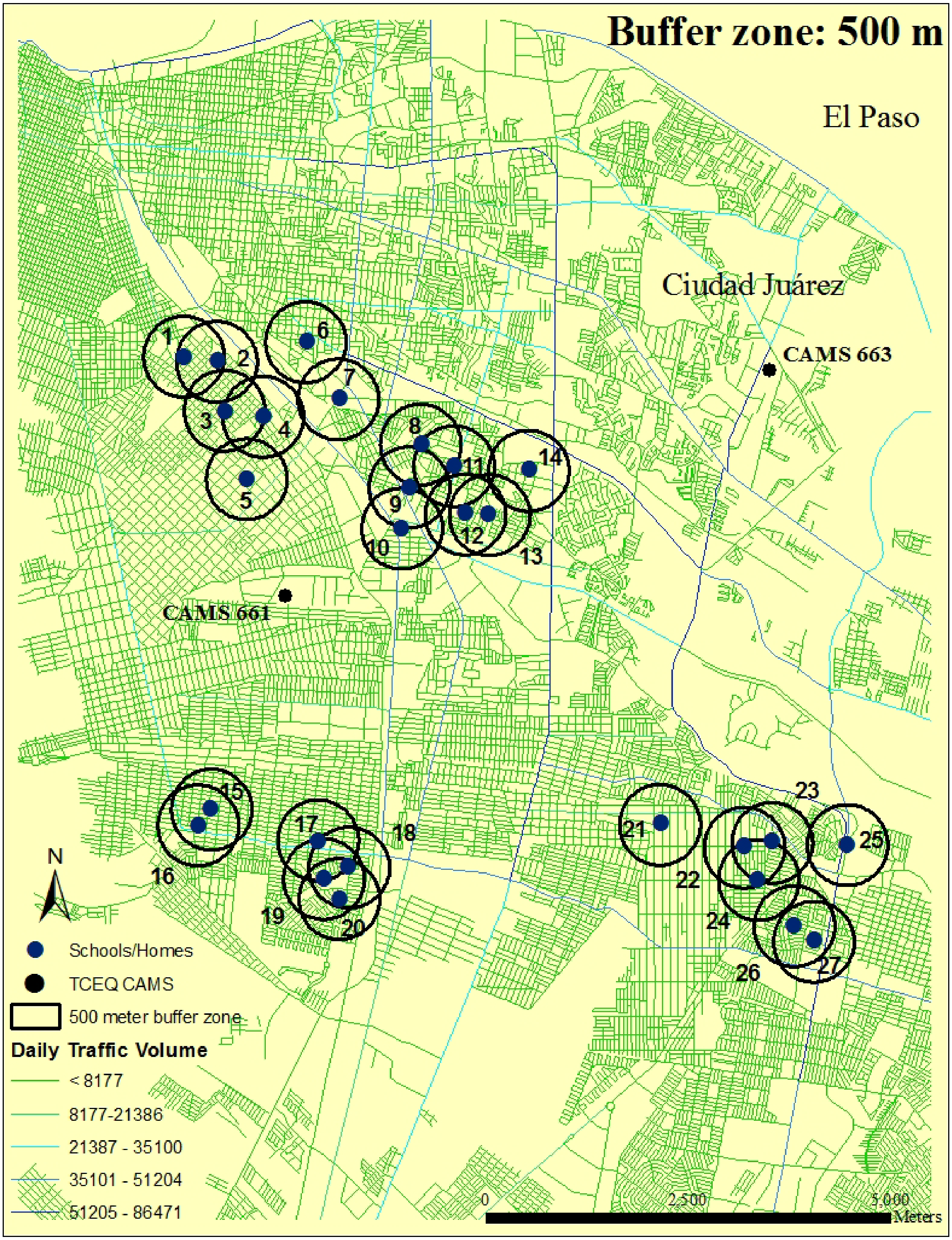

Figure 4.33: 500 meter buffer zones around monitoring locations

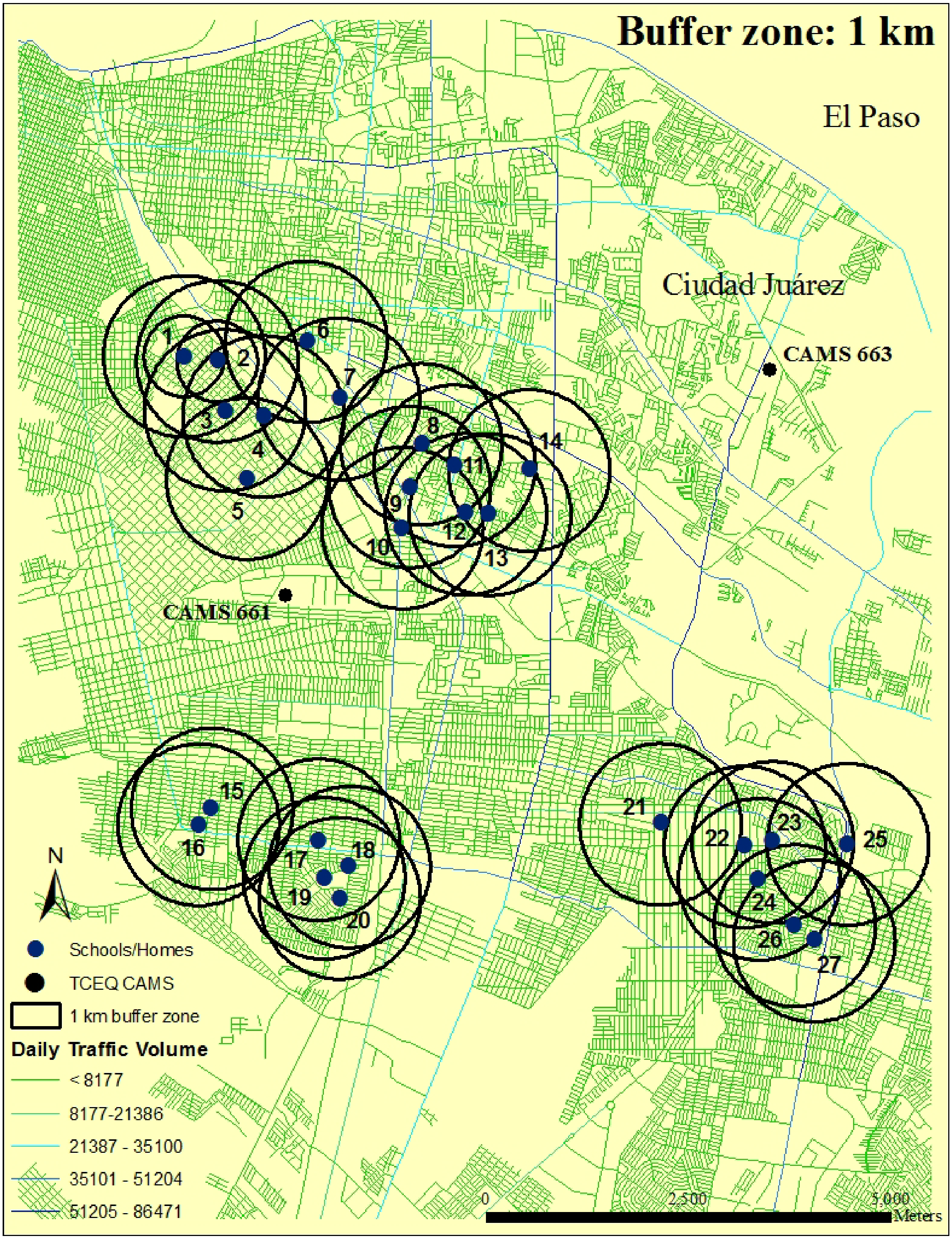

Figure 4.34: One kilometer buffer zones around monitoring locations

4.3 Statistical Analysis

The first step in analyzing any set of data is to understand the nature and characteristics of the different variables. It is important to calculate descriptive statistics to verify the range of the values in the data and to see if there are any outliers. Histograms and scatterplots are important to verify linearity and normality of the data. This type of analysis is known as exploratory data analysis (EDA) and it also serves to verify if the assumptions for the statistical analysis that will be performed are being met. In this particular study, the data was checked for linearity and normality. Linearity, as its name implies, assumes that two variables are related in a linear fashion. The statement means that when points are plotted on a scatterplot, the data will fall in a straight line or in a cluster that is relatively straight. If the data is not linearly related, the plot will look curved. In this case, the data can be transformed to make the variables linearly related (Leech et al., 2005).

The next step in the statistical analysis is to perform Pearson correlations in order to check for the level of association between dependent and predictor variables and to check for multicollinearity among the predictor variables. Multicollinearity occurs when there are high intercorrelations among sets of predictor variables, which can be a sign that two or more of the predictor variables contain overlapping information (Morgan et al., 2007). If variables are highly correlated (with a correlation coefficient of 0.50, 0.60 or above), then one might decide to combine them into a composite variable or eliminate one or more of them if they cannot be combined. In this particular case, the Pearson correlations will help us evaluate the variables to be included in the final model: all of them should have relatively weak correlations within them. The Pearson correlation is a bivariate parametric statistic used when both variables are approximately normally distributed. Parametric tests, such as the Pearson correlations, were designed for data that have approximately normal distributions. Some parametric statistics are said to be “robust” meaning that the assumption can be violated without damaging the validity of the statistic. For example, for parametric tests, the data does not have to show normal distribution at all times, but it can be somewhat skewed, and we can still use statistics designed for parametric tests. To check for normality and skewness, SPSS recommends dividing the skewness by the standard error of skewness, if the result is less than 2.5, then skewness is not significantly different from

normal (Leech et al., 2005). Standard error of skewness (SES) is dependent on the sample size (*n)* of the data and it is obtained by the following formula,

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

Source for formula: Brown, Stan 2010.

Standard error of skewness is dependent upon the sample size, and so a problem with this method is that with large samples, most variables would be found to be non-normal (Leech et al., 2005).

Multiple regression was employed to analyze the correlation between the predictor variables and nitrogen dioxide concentration. Multiple regression analysis attempts to predict a normal dependent variable from a combination of several normally distributed and/or dichotomous independent/predictor variables (Morgan et al., 2007). In this particular study, the multiple regression methodology utilized was the stepwise regression approach. Under the stepwise regression approach variables are either added to or deleted from the regression model at each step of the model building process. The stepwise procedure ends the process with the selection of a best-fitting model, when no variables can be added to or deleted from the last model fitted (Levine et al., 2001).

Model validation was carried out with bootstrapping, a technique for making inferences about a population characteristic based on an estimator derived from a sample drawn from that population. It employs large numbers of repetitive computations to estimate the shape of a statistic's sampling distribution. Bootstrapping involves "resampling" the data with replacement many times, to generate an empirical estimate of the entire sampling distribution (Mooney and Duval, 1993). The resampling of size *n* is drawn from the original sample randomly with replacement. So, although each resample will have the same number of elements as the original sample, through replacement resampling each resample could have some of the original data points represented in it more than once, and some not represented at all. Therefore, each of these resamples will likely be slightly and randomly different from the original sample (Mooney and Duval, 1993)

In this particular case, bootstrapping was conducted on the regression model. The classic regression model holds that the regressors are fixed constants, and that the response is a function of

these fixed constants and a random error term. The only random aspect of the process is the error term, and therefore it is this quantity that should be resampled in bootstrapping.

Chapter 5: Results and Discussion

As outlined in Section 4.3, the first step in analyzing the data was to determine the nature of the variables in order to justify the statistical analysis performed. For this purpose, histograms, skewness, and standard error of skewness were calculated on each of the variables of the database. The histograms for each variable are shown in Appendix B, Figures B1-B13. Table 5.1 shows the calculations that were performed in order to evaluate the normality of the data as outlined in the second paragraph of Section 4.3. As can be observed, Table 5.1 contains a column for skewness, standard error of skewness, and the division of skewness over SES. The fifth column shows whether the test for normality was met; if the division of skewness over SES in the fourth column was less than 2.5, then the cell will display the word "Yes" if the division of skewness over SES was more than 2.5, then the test for normality was not met and the cell will display "No". As can be seen from Table 5.1, this test was met for all of the variables, excluding daily traffic volume to the major road with the highest traffic volume count (TV_1^{st}). A look at the histogram for this variable (Figure B9) shows a relatively normal frequency curve and justified the utilization of parametric tests such as the Pearson correlations.

Table 5.1: Check for normality in the data

	Skewness	**Std. Error of Skewness (SES)**	**Skewness/SES**	**Normality Test**
				Skw/SES <2.5
Average_NO2	-.116	.448	-0.258	Yes
DIST_BOTA	.101	.448	0.225	Yes
DIST_PDN	.192	.448	0.430	Yes
DIST_ZAR	-.644	.448	-1.437	Yes
DIST_CP	-.842	.448	-1.880	Yes
DIST_EE	.032	.448	0.071	Yes
DIST_WR	.111	.448	0.248	Yes
DIST_1ST	.225	.448	0.502	Yes
TV_1ST	1.431	.448	**3.193**	**No**
DIST_2ND	.909	.448	2.029	Yes
TV_2ND	1.006	.448	2.245	Yes
TD_500	1.069	.448	2.387	Yes
TD_1000	.258	.448	0.576	Yes

5.1 Pearson Correlations

The results from the Pearson correlations (*r)* are shown in Table 5.2. The strongest correlations are shown in bold and with either one or two asterisks for emphasis. Significant correlations at a *p-value* of < 0.01 are shown between NO_2 concentrations and distance to the cement plant (DIST_CP), distance to the oil refinery (DIST_OR) and sum of the traffic volume on a major street times its corresponding street length at the 1000 meter buffer (TD_1000). Significant correlations at a *p-value* of 0.05 are shown between NO_2 concentrations and distance to the electric engine factories (DIST_EE) and distance to a major street with the second highest traffic volume (DIST_2ND). Multicollinearity was shown amongst the variables distance to the Bridge of the Americas (DIST_BOTA), distance to the Paso del Norte Bridge (DIST_PDN), distance to the electric engine factories (DIST_EE), and distance to the oil refinery (DIST_OR). The collinearity exists perhaps because those are all point sources located in the same direction, towards the north of Ciudad Juárez (see Figure 4.32). Multicollinearity was also shown amongst highest daily traffic volume in a major street (TV_1ST) and second highest daily traffic volume in a major street (TV_2ND) because their variables reflect similar information. In the final model, none of the variables that exhibited multicollinearity were included except for distance to the oil refinery (DIST_OR).

The Pearson correlations shown from Table 5.2 indicated that DIST_OR was significantly correlated with NO_2 concentrations. The significant negative correlation between NO_2 concentrations and distance to an oil refinery (DIST_OR) indicated that as the distance to the oil refinery increased, the NO_2 concentrations decreased. This finding is consistent with the findings by Smith et al. (2006), where the variable denominated as distance to nearest petroleum facility (OIL_DIST) was a key explanatory variable in the determination of NO_2 concentrations. Surprisingly, distance to the ports of entry was not as significant as we might have expected from the findings reported by Smith et al. (2006), and Gonzales et al. (2005), where both studies conducted in El Paso, TX, found that distance to the nearest border crossing was a significant variable for determining the variation in NO_2 concentrations. Our findings do establish that of all the three ports of entry, distance to the Bridge of the Americas (DIST_BOTA) was the one with the highest negative correlation to NO_2 reflecting the fact that the Bridge of the Americas has a higher traffic volume as compared to the Paso del Norte and Zaragoza

Bridge (see Table 5.5). The distance to the Zaragoza bridge (DIST_ZAR) had the second highest negative correlation out of the three ports of entry and the distance to the Paso del Norte bridge (DIST_PDN) was the lowest. This could be a reflection of the type of traffic crossing each bridge: the Zaragoza bridge has commercial trucks crossing whereas the Paso Del Norte bridge does not include commercial trucks in its daily traffic. As we saw in the NO_x emission inventory for Ciudad Juárez (Table 3.1), heavy duty diesel vehicles contribute to 31 percent of the total yearly emissions of NO_x in the city. Distance to a cement plant (DIST_CP) and distance to electric engine factories (DIST_EE) were variables not included in previous studies conducted in the area such as the ones by Gonzales et al. (2005), or in the same study presented in separate papers by Noble et al. (2003), and Smith et al. (2006). However, the results shown by the Pearson correlations deemed them as significant variables. It was unexpected that DIST_CP was positively correlated with NO_2 concentrations when the opposite would be anticipated. Further investigation would be necessary in why this was the case. It could also be an indication that the emissions from the cement plant are not as significant to the contribution of NO_2 as we might have expected. After all, in the report provided by the SEMARNAT, NO_x emissions from the cement plant were only 7 percent of the total emissions from all sources considered in Ciudad Juárez. Moreover, in an article by Van Oss and Padovani titled *Cement Manufacture and the Environment,* several studies reviewed by them revealed that 90 percent or more of NO_x emissions in cement kilns are nitric oxide (NO) with the rest being NO_2. Another possible explanation could be the location of the cement plant with respect to the monitored locations and the presence of a stack in the cement plant. The height of the stack could serve to disperse the pollutants away from the cement plant, so that as one gets closer to the plant, the pollutant's concentration decreases and as one gets farther from the plant, the pollutant's concentration (in this case, NO_2) increases.

Continuing with Table 5.2, we can see that a significant negative correlation was also found between the distance to a major street with the second highest traffic volume (DIST_2nd) and NO_2 concentrations. As the distance to that street increased, the NO_2 concentrations decreased. This finding was consistent with the findings by Molina (2005) where she used the same set of data for her study. However, in her set of data, the schools were analyzed in groups of 10 clusters, whereas in the present

study, the schools/homes were analyzed individually. Another significant positive correlation was found between the sum of the traffic volume on a major street times its corresponding street length at the 1000 meter radius (TD_1000) and NO_2 concentrations. As the traffic density increased so did the NO_2 concentrations, and vice versa. Again, this finding was consistent with the analysis by Molina (2005) in her study of the schools/homes as clusters, but it was also consistent with the findings by Ross et al. (2006), where he found that traffic density within the 1000 meter buffer zone after subtracting the first 300 meter buffer zone he had taken into account (traffic300-1000) was part of the top 3 strongest predictor variables for NO_2 concentrations in San Diego, CA, and lastly, it was also consistent with the findings by Smith et al. (2006), where he also found that traffic intensity within 1000 m of location, given in vehicles/day-km (INT1000), was an important variable in the determination of NO_2 concentrations. Therefore, a 1000 meter radius has been shown to be a crucial distance between traffic counts and NO_2 concentrations.

Pearson correlations between wind direction and NO_2 concentrations are shown in Table 5.3 for the information obtained from C661 and in Table 5.4 for the information obtained from C663. Wind direction at either of the monitoring stations (C661 and C663) did not show a significant correlation with NO_2. Arain et al. (2007) performed a study evaluating the significance of including wind fields in LUR models for predicting NO_2. His results showed that a small but significant improvement in model performance was made when wind directions effects were included. However, it is also important to consider that the area where he conducted his study (Toronto-Hamilton) had a topography that significantly impacts wind patterns. As Noble et al. (2003) points out in his research, large buildings and urban canyons are known to create localized maxima or minima in pollutant concentration due to channeling and a lack of effective dispersion and dilution. In the case of the El Paso- Ciudad Juárez region, most of its urban area is relatively flat. In the case of El Paso, the exception is the Franklin mountains range beginning within the El Paso city limits in the south and extending northward across the New Mexico border for a distance of about 15 miles. In the case of Ciudad Juárez, most of its population is located in a relatively flat area at the Juárez valley by the Río Grande, although another

part of the population is also established to the west and south of the city where a series of mountain ranges known as Sierra del Presidio, Sierra de Samalayuca, and Sierra de Juárez are located.

Table 5.2: Pearson correlations. Significant correlations at the *p-level* of 0.01 (**) and at the *p-level* of 0.05 (*) are outlined below.

r	NO_2	DIST_BOTA	DIST_PDN	DIST_ZAR	DIST_CP	DIST_EE	DIST_OR	DIST_1st	TV_1st	DIST_2nd	TV_2nd	TD_500	TD_1000
NO_2	1	-0.308	-0.160	-0.230	**0.571****	**-0.447***	**-0.509****	-0.191	0.224	**-0.395***	0.287	0.350	**0.567****
D_BOTA		1	**0.972****	**-0.686****	-0.064	**0.948****	**0.926****	0.073	0.090	-0.150	0.375	0.084	-0.210
DIS_PDN			1	**-0.832****	0.128	**0.851****	**0.813****	0.040	0.194	-0.204	0.473*	0.181	-0.016
DIS_ZAR				1	**-0.622****	-0.453	-0.364	0.017	-0.384*	0.238	**-0.529****	-0.336	-0.358
DIST_CP					1	-0.245	-0.376	-0.052	0.455*	-0.011	0.256	0.251	0.499**
DIST_EE						1	**0.989****	0.110	-0.031	-0.047	0.213	-0.064	-0.461*
DIST_OR							1	0.111	-0.088	-0.055	0.185	-0.082	-0.481*
DIST_1st								1	-0.099	0.142	-0.139	**-0.693****	-0.359
TV_1st									1	-0.100	**0.755****	**0.560****	0.417*
DIST_2nd										1	-0.278	-0.376	-0.325
TV_2nd											1	**0.530****	0.481*
TD_500												1	0.496**
TD_1000													1

Table 5.3: Pearson correlations for wind direction and NO_2 concentrations for C661

r	NO_2	N-S	E-W
NO_2	1.000	.162	-.259
N-S		1.000	-.646
E-W			1.000

Table 5.4: Pearson correlations for wind direction and NO_2 concentrations for C663

r	NO_2	N-S	E-W
NO_2	1.000	.203	-.088
N-S		1.000	-.146
E-W			1.000

Table 5.5: Crossing numbers at the ports of entry for the year of 2003

	2003	
	Trucks	Vehicles
BOTA	13,351	4,679,772
PDN	0	4,173,265
ZAR	313,737	3,370,044

Source: www.elpasompo.org

5.2 Multiple Regression

The variables were further evaluated in a stepwise regression analysis for the creation of a model that will give the main components needed to predict NO_2 concentrations. The nitrogen dioxide model was created using a stepwise regression approach with the inclusion of all 12 predictor variables selected. As shown by the Pearson correlations, distance to the cement plant (DIST_CP) was the strongest predictor of NO_2 levels in the stepwise regression approach, explaining 30 percent of the

variation in the 27 samples. Again, however, the correlation was positive when the opposite would be expected. Along with distance to the cement plant, the second strongest predictor variable was distance to a major road with the second highest traffic volume ($DIST\_2^{ND}$). Together, both variables accounted for 43 percent of the variation in the NO_2 levels in the 27 samples. The last variable to be added to the model was distance to the oil refinery (DIST_OR) and all three variables accounted for 59 percent of the variation in the NO_2 levels in the 27 samples. The final model and its coefficients are shown in Table 5.6. The regression model was validated through the bootstrapping technique by running 1,000 iterations. The histograms obtained are shown in Figures 5.1, 5.2, and 5.3 for each of the significant variables included in the final regression model. The mean of each of the histograms were compared to the standardized beta coefficients obtained from the stepwise regression analysis and are shown in Table 5.7. The closeness of the values shows that the regression model obtained is robust.

Table 5.6: Results of stepwise regression analysis

Variable	**β**	**p-Value**	**Model R^2**
Intercept	29.859	< 0.001	0.594
DIST_CP (m)	0.001	0.007	
DIST_2nd (m)	-0.006	0.005	
DIST_OR (m)	-0.001	0.017	

Table 5.7: Comparison of beta coefficients after a bootstrap with 1,000 iterations

	Standardized beta coefficients from regression model	Mean from standardized beta coefficients after bootstrapping
DIST_CP	0.427	0.430
DIST_2ND	-0.410	-0.420
DIST_OR	-0.371	-0.370

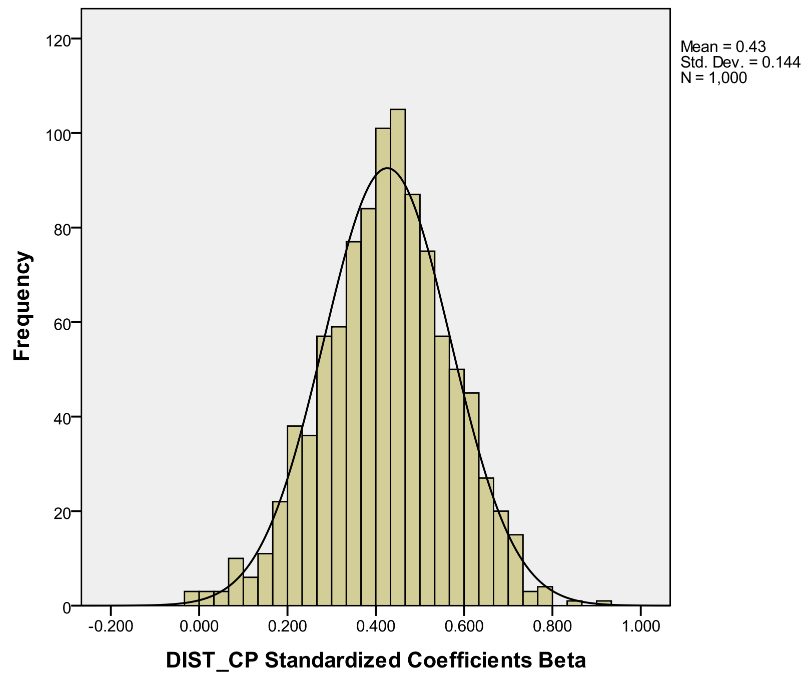

Figure 5.1: Histogram of distance to cement plant variable after a bootstrap with 1,000 iterations

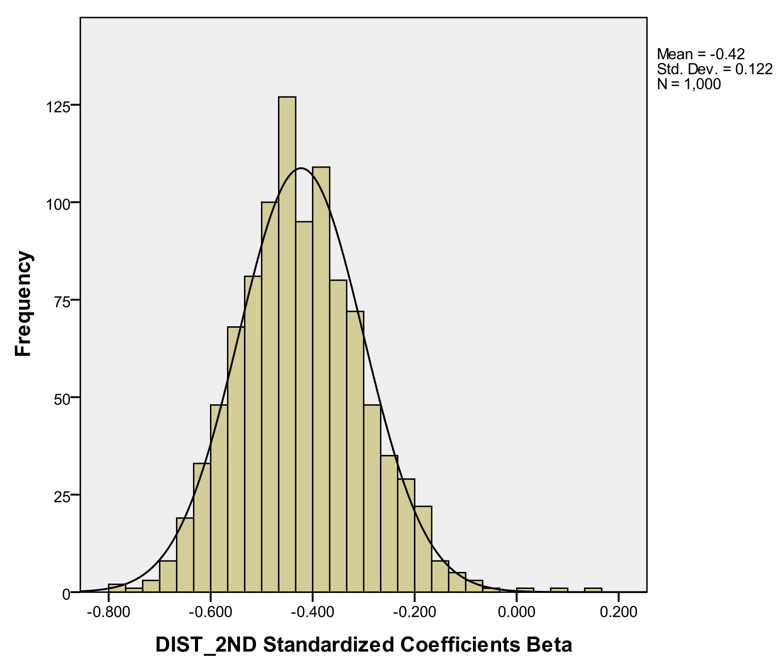

Figure 5.2: Histogram of distance major street with second highest traffic volume variable after a bootstrap with 1,000 iterations

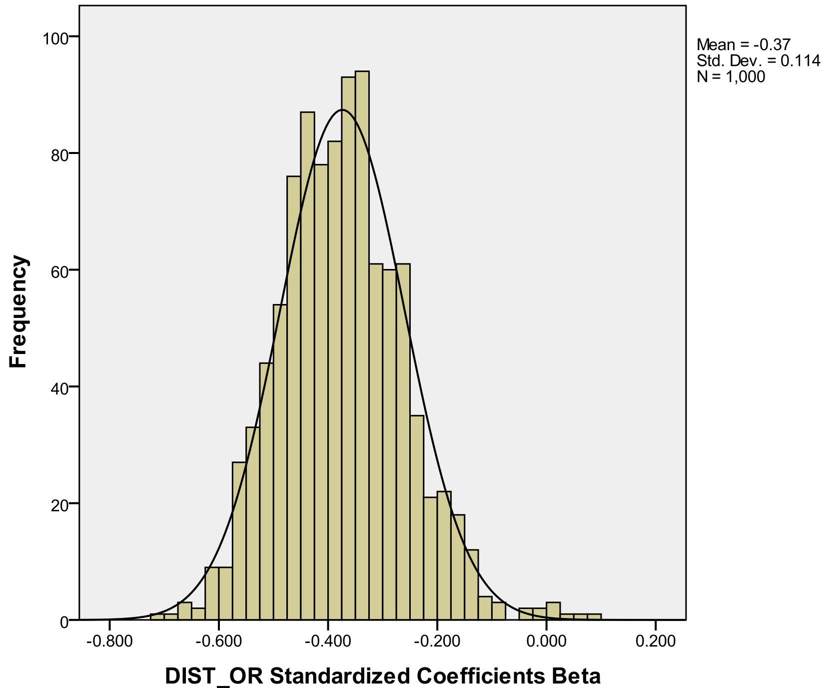

Figure 5.3: Histogram of distance to oil refinery variable after a bootstrap with 1,000 iterations

Chapter 6: Conclusions

This chapter outlines the main conclusions encountered in this research and makes future recommendations.

6.1 Major points

The main objective in this research was to evaluate the strength and association of different land regression variables for predicting NO_2 concentrations in Ciudad Juárez, México. In order to achieve this goal Tasks 1-3 established the database necessary for use in the model. Task 1 monitored NO_2 levels in 27 locations for 12-day periods. Task 2 and 3 consisted of identifying the main contributors to NO_2 emissions in the Ciudad Juárez and El Paso area through the use of reports from the SEMARNAT and TCEQ. A total of 12 variables were identified as significant contributors to NO_2 levels. Statistical analysis performed in the data under Tasks 4 -6, such as pearson correlations, stepwise regression, and bootstrapping to validate the model revealed the following:

- Distance to the cement plant was the most significant variable for predicting NO_2 concentrations as outlined by both the pearson correlation and the stepwise regression approach. However, the correlation was positive meaning that as the distance to the cement plant increased, so did the concentration of NO_2, when the opposite would be expected. The fact that NO_2 levels decrease toward the direct location of the cement plant could be attributed to the presence of a stack driving the emissions away from the cement plant, making the NO_2 concentrations higher as you get further away from the plant and lower as you get closer to the cement plant. No other study in the area had taken this variable into account.
- Traffic density at the 1000 meter buffer zone (TD_1000) was found to have the second strongest positive pearson correlation for variation of NO_2 concentrations. This finding was consistent with a study conducted in the El Paso area by Smith, et al. (2006), in Ciudad Juárez by Molina (2005), and by Ross et al. (2006) in San Diego, CA. Again, this study verified that traffic counts within a 1000 meter radius is an important variable influencing NO_2 concentrations.

- Distance to the oil refinery had the third strongest negative correlation with NO_2 concentrations. This finding was again consistent with the study conducted by Smith, et al. (2006).
- Distance to the electric engine factories had the fourth strongest negative correlation with NO_2 concentrations. This variable was not considered in any of the previous studies reviewed in this thesis, and therefore there was no basis for comparison. However, the NO_x emission inventory for Ciudad Juárez does include electric engine factories as one of the main sources for NO_x emissions in the area.
- Finally the last significant variable encountered in the pearson correlation analysis was distance to a major street with the second highest traffic volume.
- Distance to a port of entry was found as a significant variable for predicting NO_2 concentrations in previous studies conducted in the Ciudad Juárez-El Paso area. The correlation analysis conducted in this study did not find distance to any of the ports of entry as a significant variable at the 0.001 or 0.05 *p-levels.* Nevertheless, pearson correlations in this study found that the distance to the Bridge of the Americas had the strongest negative correlation out of the three ports of entry, demonstrating that the amount of traffic at the bridge and its close proximity to the monitored locations can have an influence in the NO_2 levels.
- Wind direction was not found to be a significant factor for NO_2 concentration variations.
- The number of variables found to be significant for predicting NO_2 levels under the stepwise regression approach where three, predicting 59% of the variation in NO_2 concentrations. These variables are distance to the cement plant (DIST_CP), distance to a major road with the second highest traffic volume (DIST_2ND), and distance to an oil refinery (DIST_OR).
- Finally, the model obtained in the stepwise regression was evaluated with a bootstrap analysis performing 1,000 iterations and the results revealed the robustness of the model.

6.2 Future recommendations

Previous studies did not consider electric engine factories or cement plants as significant contributors to NO_2 levels, specifically in the El Paso- Ciudad Juárez area. However, the present study found that they are indeed significant variables that should be considered, especially for an area such as

Ciudad Juárez, where economic activities are focused around the manufacturing industry. Special attention and further investigation is also needed to look into the influence of cement plants into the variation of NO_2 concentrations. In addition, elevation above sea level was not included in this study. Researchers such as Gonzales, et al. (2005) found it as a significant variable in her own research, therefore the inclusion of this variable in the multiple regression analysis could increase the capability of predicting NO_2 concentrations in our model.

References

Arain A., Blair R., Finkelstein N., Brook J., Sahsuvaroglu T., Beckerman B., Zhang L., and Jerrett M. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment.* 2007; 41: 3453-3464.

Brown S. *Measure of Shape: Skewness and Kurtosis.* Revised 2010. Accessed July 16, 2010. http://www.tc3.edu/instruct/sbrown/stat/shape.htm

DeNevers, N. *Air Pollution Control Engineering.* 2nd edition, Boston: 2000.

Gonzales M., Qualls C., Hudgens E., and Neas L., Characterization of a spatial gradient of nitrogen dioxide across a United States-Mexico border city during winter. *Science of the Total Environment.* 2005; 337: 163-173.

Han X., and Naeher L. P. A review of traffic-related air pollution assessment studies in the developing world. *Environmental International.* 2006; 32: 106-120.

Hoek G., Beelen R., De Hoogh K., et. al. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment.* 2008; 42(33): 7561-7578.

Holguin F., Flores S., Ross Z., Cortez M., Molina M., Molina L., Rincon C., Jerrett M., Berhane K., Granados A., and Romieu I. Traffic-related exposures, airway function, inflammation, and respiratory symptoms in children. *American Journal of Respiratory and Critical Care Medicine.* 2007; 176: 1236-1242.

Jerrett M., Arain A., Kanaroglou P., Beckerman B., Potoglou D., Sahsuvaroglu T., Morrison J., Giovis C. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology.* 2005; 15: 185-204.

Lau J., Hung W.T., Cheung C.S. Interpretation of air quality in relation to monitoring station's surroundings. *Atmospheric Environment.* 2009; 43: 769-777.

Leech N.L., Barrett K.C., and Morgan G.A. *SPSS for Intermediate Statistics: Use and Interpretation.* 2nd edition, New Jersey: 2005.

Levine D., Ramsey P.P., and R.K. Smidt. *Applied Statistics for Engineers and Scientists.* New Jersey: 2001.

Li, W.-W., Orquiz R., Pingitore N. E., Garcia J. H. Jr., Espino T. T., Gardea-Torresdey J., Chow J., and Watson J. G. Analysis of temporal and spatial dichotomous PM air samples in the El Paso-Cd. Juárez air quality basin. *Journal of Air and Waste Management Association.* 2001; 51: 1511-1560.

Molina E. (2005). *Implications of Airborne $PM_{2.5}$ and Nitrogen Dioxide at Cd. Juárez, México.* Unpublished master's thesis, University of Texas at El Paso. 158 p.

Mooney C., Duval R.D. *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Sage Publications: 1993.

Morgan G., Leech N. L., Gloeckner G. W., and Barrett K. C. *SPSS for Introductory Statistics.* 3rd edition, New Jersey: 2007.

National Oceanic and Atmospheric Administrationn (NOAA). Transport Winds Tutorial. Accessed on June 22, 2010. http://www.wrh.noaa.gov/sew/fire/olm/transport.htm

Nitrogen Dioxide. Accessed on March 3, 2010. www.epa.gov/air/nitrogenoxides.

Noble C.A., Mukerjee S., Gonzales M., Rodes C.E., Lawless P .A., Natarajan S., Myers E. A., Norris G. A., Smith L., Ozkaynak H., Neas L. M. Continuous measurement of fine and ultrafine particulate matter, criteria pollutants and meteorological conditions in urban El Paso, Texas. *Atmospheric Environment.* 2003; 37: 827-840.

Ogawa & Co., USA, Inc. NO, NO_2, NO_x, and SO_2 Sampling Protocol Using the Ogawa Sampler (version 6). Accessed on July 12, 2010. http://www.ogawausa.com/passive.html

Paso del Norte Mapa. Accessed on January 2010. http://www.pdnmapa.org

Ryan P., and LeMasters G.K. A review of land use regression models for characterizing intraurban air pollution exposure. PubMed: 2010. Accessed on March 13, 2010.

Ross Z., English P.B., Scalf R., Gunier R., Smorodinsky S., Wall S., and Jerrett M. Nitrogen dioxide prediction in southern California using land use regression modeling: potential for environmental health analysis. *Journal of Exposure Science and Environmental Epidemiology*. 2006; 16: 106-114.

Sahsuvaroglu T., Arain A., Kanaroglou P., Finkelstein N., Newbold B., Jerrett M., Beckerman B., Brook J., Finkelstein M., and Gilbert N.L. A land use regression model for predicting ambient concentrations of nitrogen dioxide in Hamilton, Ontario, Canada. *Journal of Air, Waste, and Management Association.* 2006; 56(10): 1366.

SEMARNAT-INE. Inventario de emisiones de los estados de la frontera norte de México 1999-2005.

SEMARNAT. Programa de gestión de la calidad del aire de Ciudad Juárez (PROAIRE) 2006-2012. 2006.

Smith L., Mukerjee S., Gonzales M., Stallings C., Neas L., Norris G., and Ozkaynak H. Use of GIS and ancillary variables to predict volatile organic compound and nitrogen dioxide levels at unmonitored locations. *Atmospheric Environment.* 2006; 40: 3773-3787.

Turco R. P. *Earth Under Siege: From Air Pollution to Global Change.* New York: 1997.

Van Oss H.G. and Padovani A.C. Cement Manufacture and the Environment (Part II Environmental Challenges and Opportunities). *Journal of Industrial Ecology.* 2003; 7(1).

Appendix A: List of Acronyms

CAMS: Continuous Ambient Monitoring Station

DIST_1st: Distance to a major street close to the monitoring location with the highest daily traffic volume (m)

DIST_2nd: Distance to a major street close to the monitoring location with the second highest daily traffic volume (m)

DIST_BOTA: Distance to the international port of entry of Bridge of the Americas (m)

DIST_CP: Distance to the Cement Plant in Ciudad Juárez (m)

DIST_EE: Distance to an agglomeration of factories fabricating electric engines in Ciudad Juárez (m)

DIST_OR: Distance to the Oil Refinery in El Paso, TX now Western Refinery (m)

DIST_PDN: Distance to the international port of entry of Paso Del Norte (m)

DIST_ZAR: Distance to the international port of entry of Zaragoza (m)

E-W: East-West component of the wind direction

ECAT: Elemental carbon attributed to traffic, a marker of diesel exhaust

EPA: Environmental Protection Agency

GIS: Geographic Information System

IMIP: Instituto Municipal de Investigación y Planeación (Municipal Institute of Investigation and Planning)

LUR: Land use regression

NAAQS: National Ambient Air Quality Standards

NO: Nitric oxide or nitrogen monoxide

NO_2: Nitrogen Dioxide

NOMs: Normas Oficiales Mexicanas (Official Mexican Norms)

NO_x: Nitrogen oxides

N-S: North-South component of the wind direction

SEMARNAT: Secretaría de Medio Ambiente y Recursos Naturales (Ministry of Environment and Natural Resources)

TCEQ: Texas Commission on Environmental Quality

TD_500: Sum of the daily traffic volume at the major roads times its corresponding street length inside the 500 m buffer zone (veh-km/day)

TD_1000: Sum of the daily traffic volume at the major roads times its corresponding street length inside the 1000 m buffer zone (veh-km/day)

TV_1st: Daily traffic volume to the major road with the highest traffic volume (veh/day)

TV_2nd: Daily traffic volume to the major road with the second highest traffic volume (veh/day)

Appendix B: Graphs

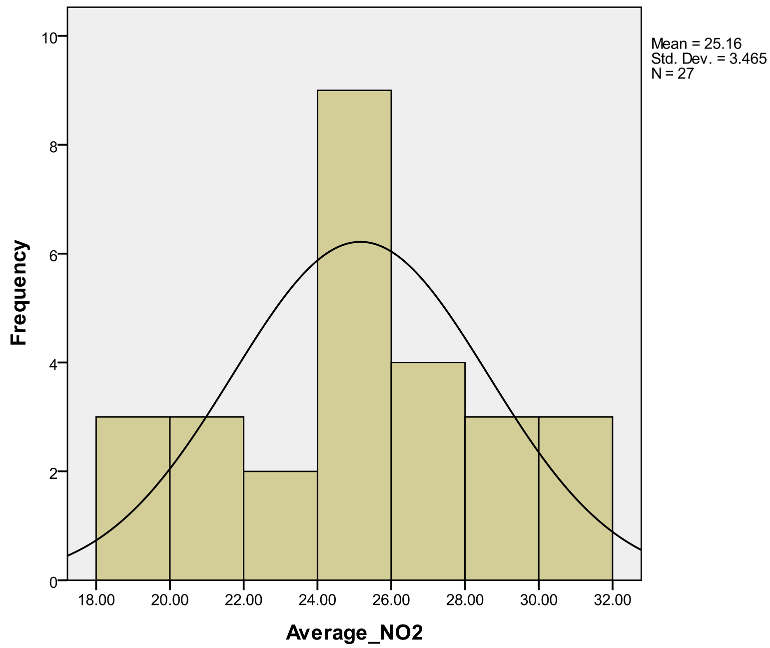

Figure B1: Histogram for NO_2 average concentrations

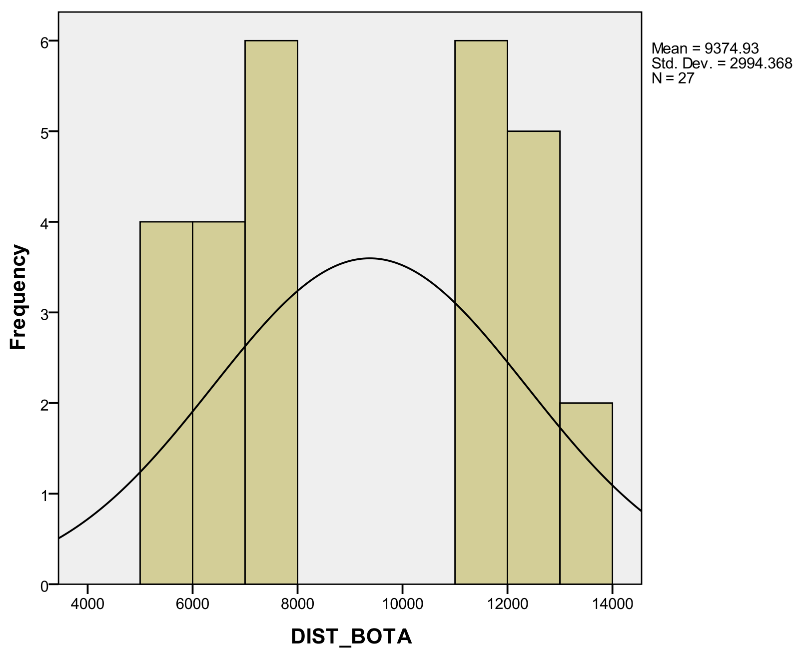

Figure B2: Histogram for distance to Bridge of the Americas port of entry

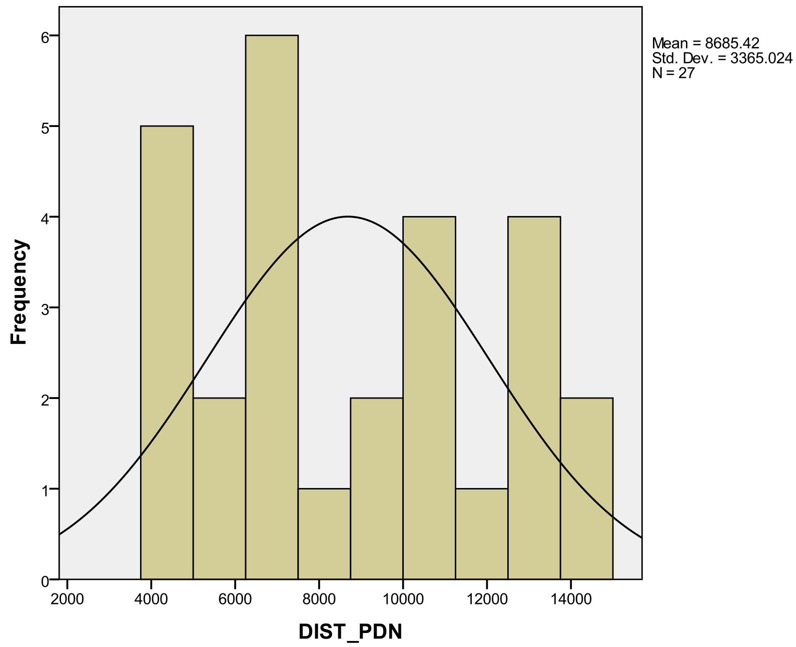

Figure B3: Histogram for distance to the Paso del Norte port of entry

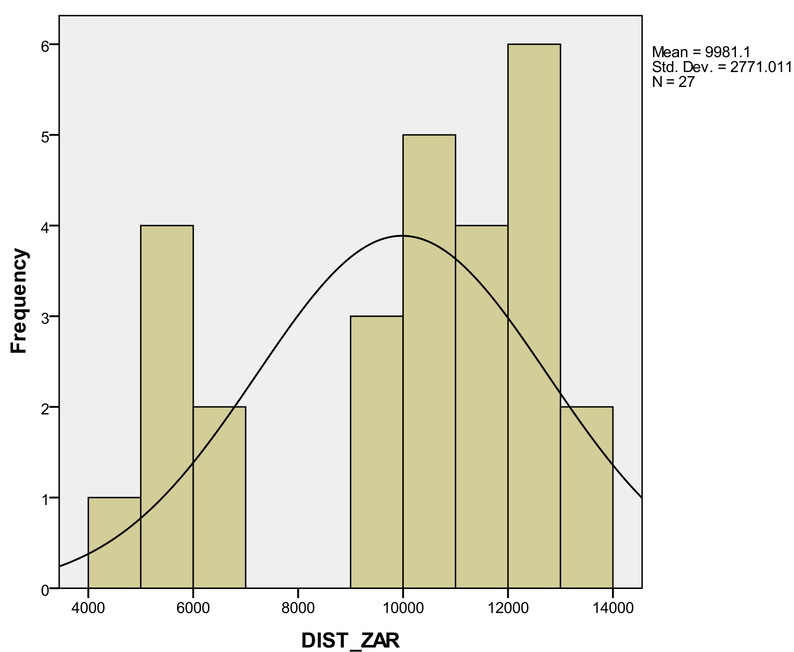

Figure B4: Histogram for distance to the Zaragoza port of entry

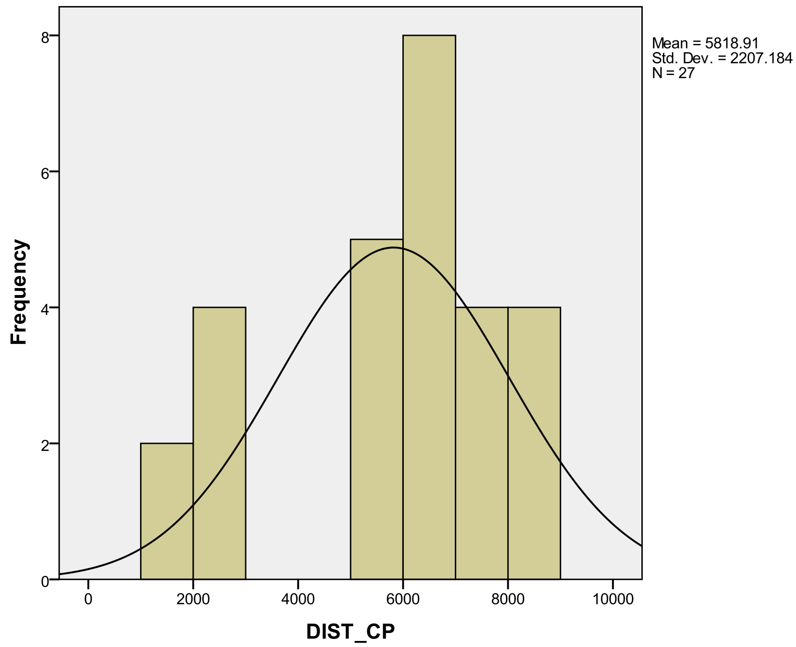

Figure B5: Histogram for distance to the cement plant

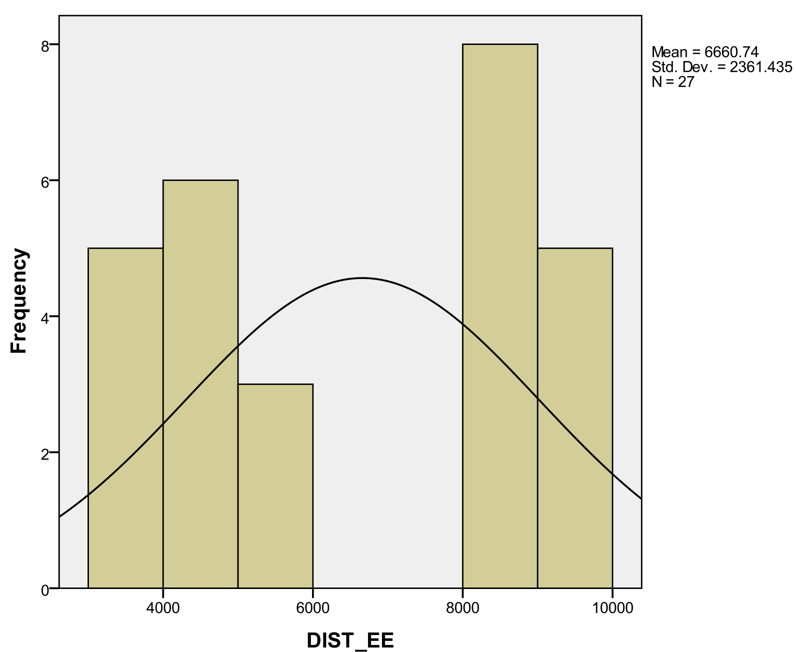

Figure B6: Histogram for distance to the electric engine factories

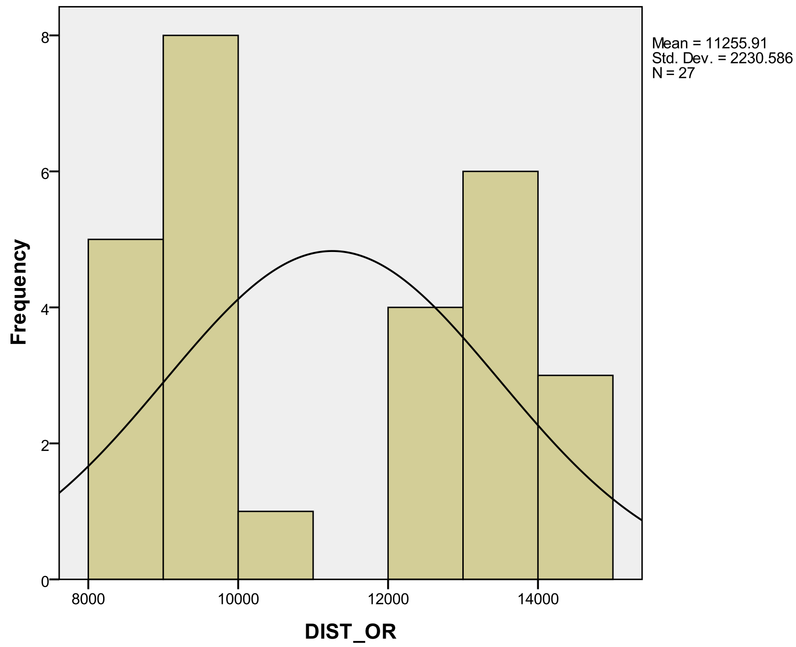

Figure B7: Histogram for distance to the oil refinery

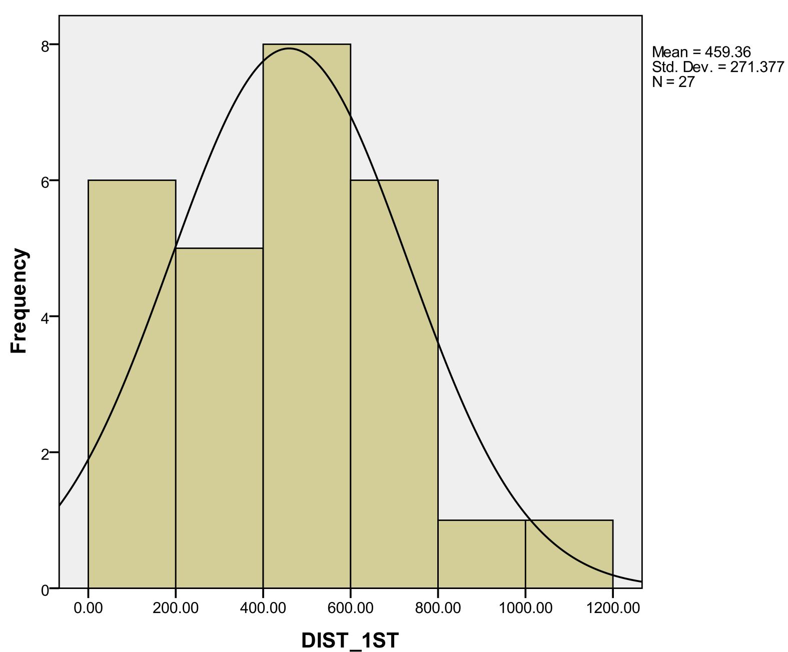

Figure B8: Histogram for distance to a major street with the highest traffic volume

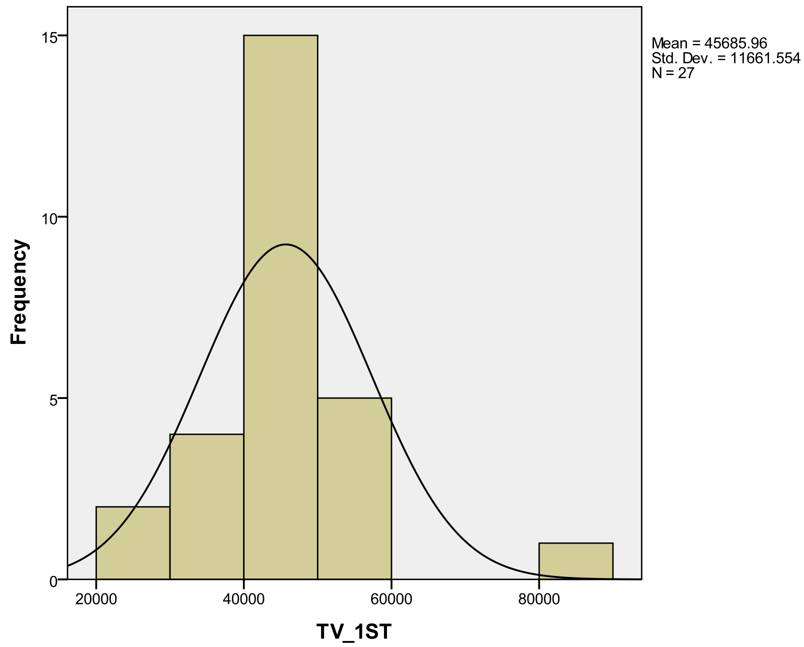

Figure B9: Histogram for the traffic volume on major road with the highest traffic count

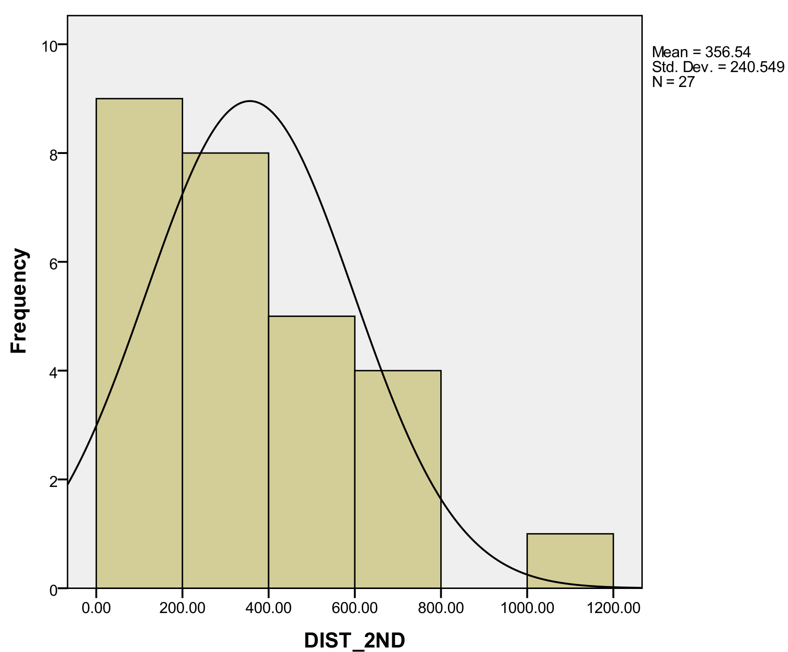

Figure B10: Histogram for the distance to a major road with the second highest traffic volume

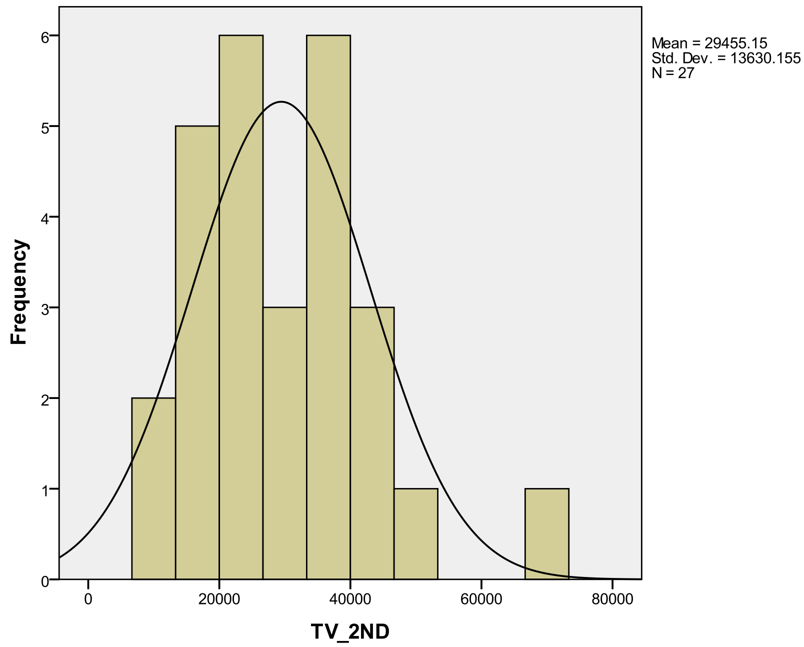

Figure B11: Histogram for the traffic volume on a major road with the second highest traffic count

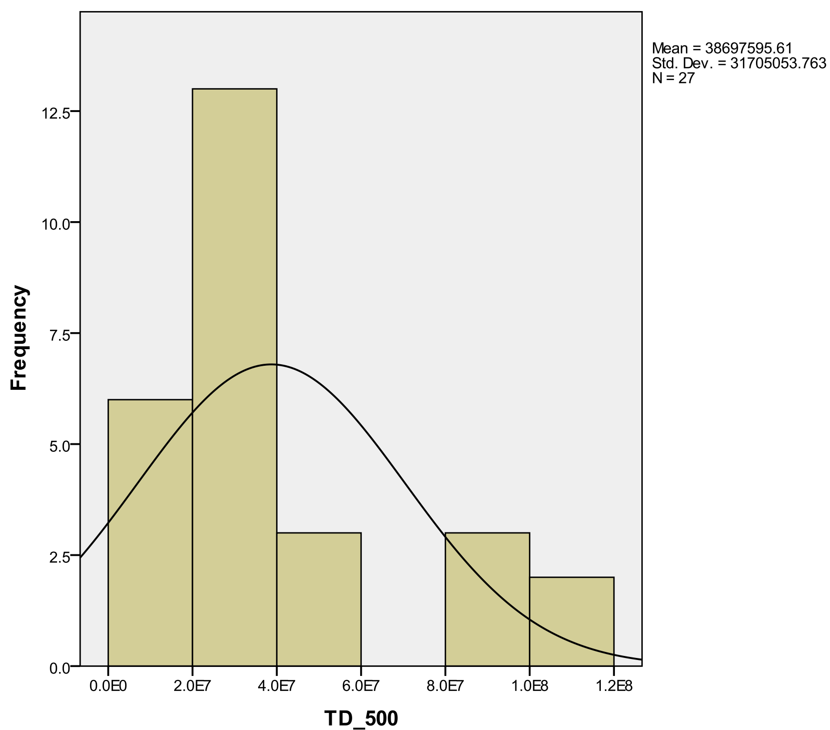

Figure B12: Histogram for the sum of the traffic volume times the street length of major roads in a 500-meter buffer zone

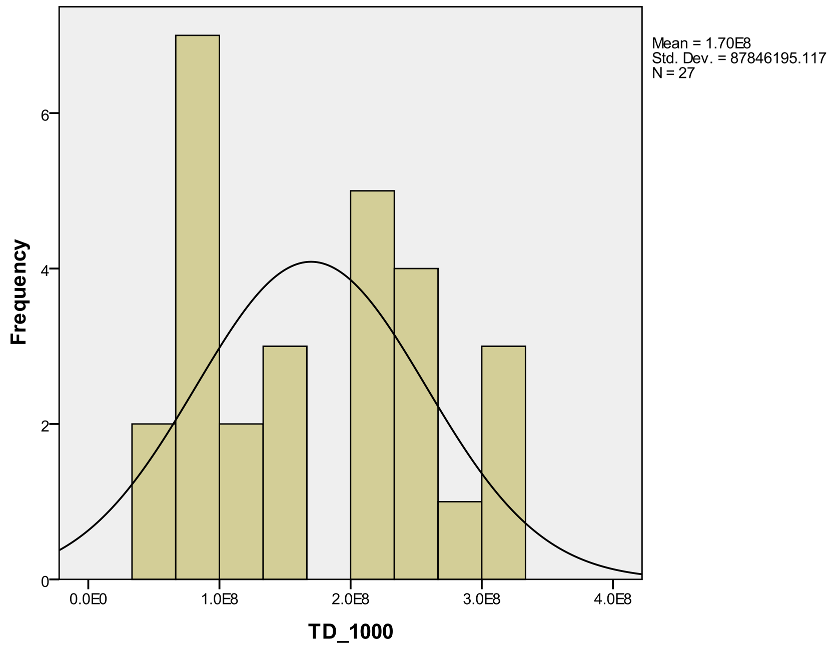

Figure B13: Histogram for the sum of the traffic volume times the street length of major roads in a 1000- meter buffer zone

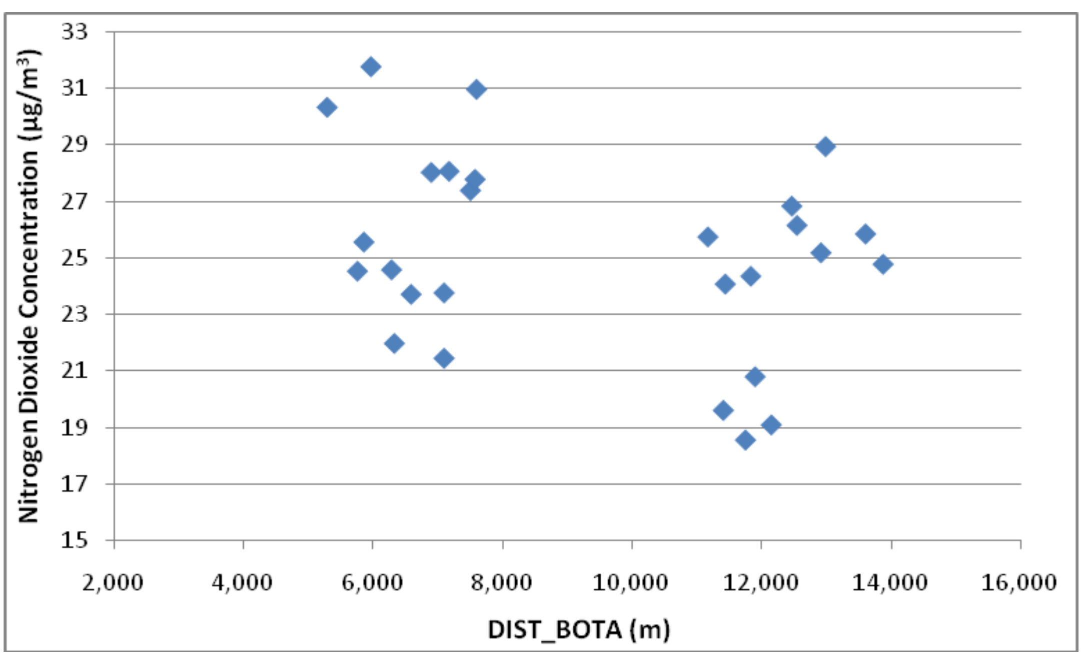

Figure B14: Scatterplot between NO_2 concentration and DIST_BOTA

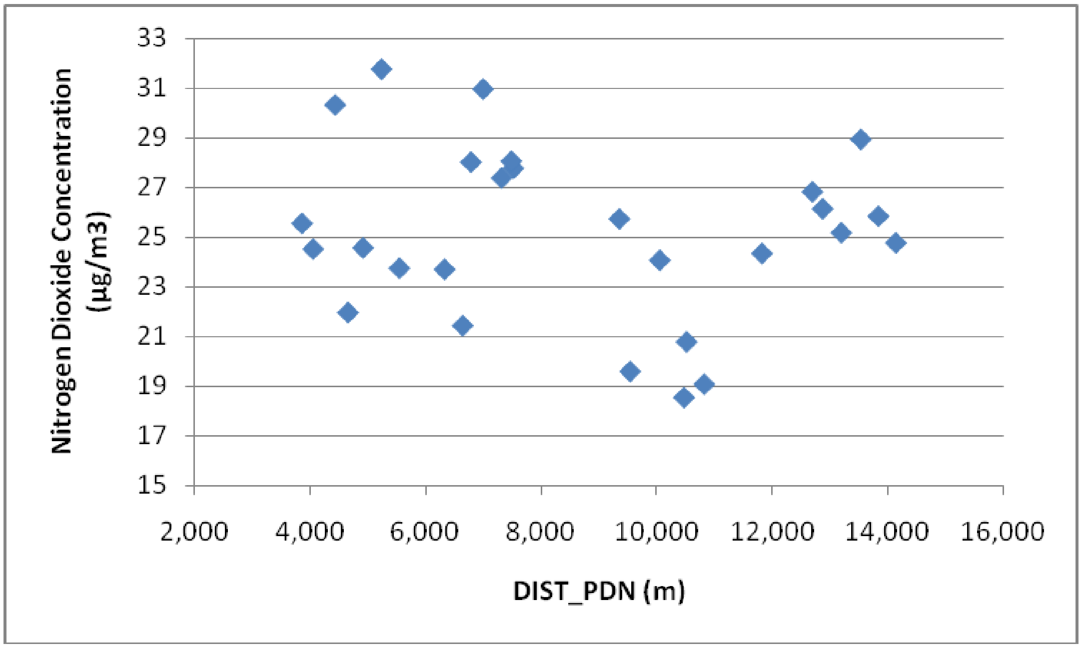

Figure B15: Scatterplot between NO_2 concentration and DIST_PDN

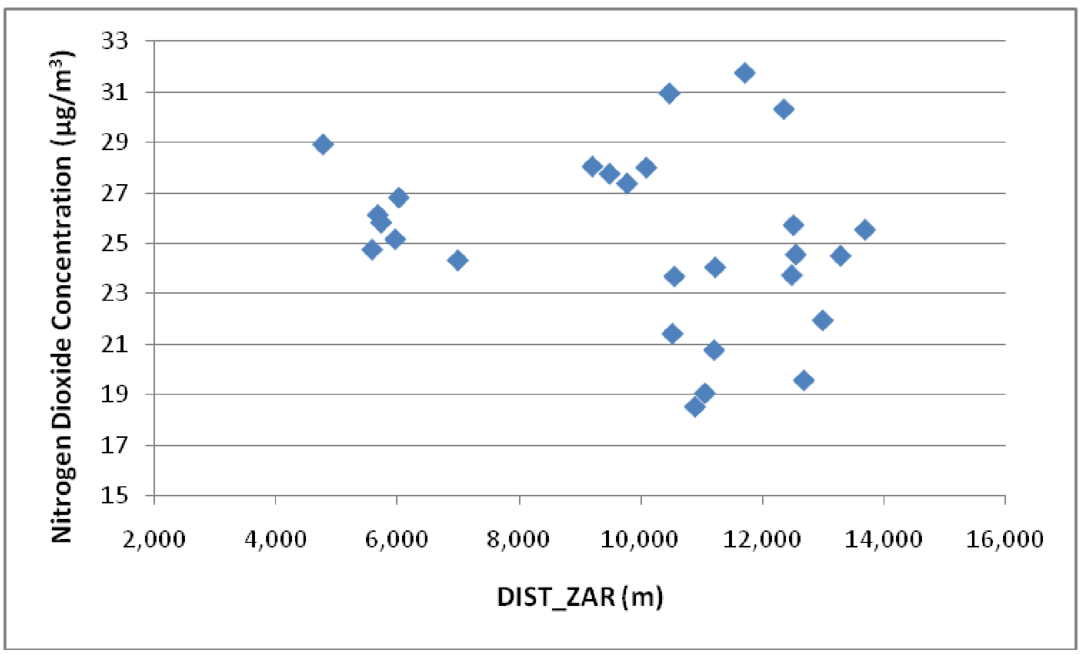

Figure B16: Scatterplot between NO_2 concentration and DIST_ZAR

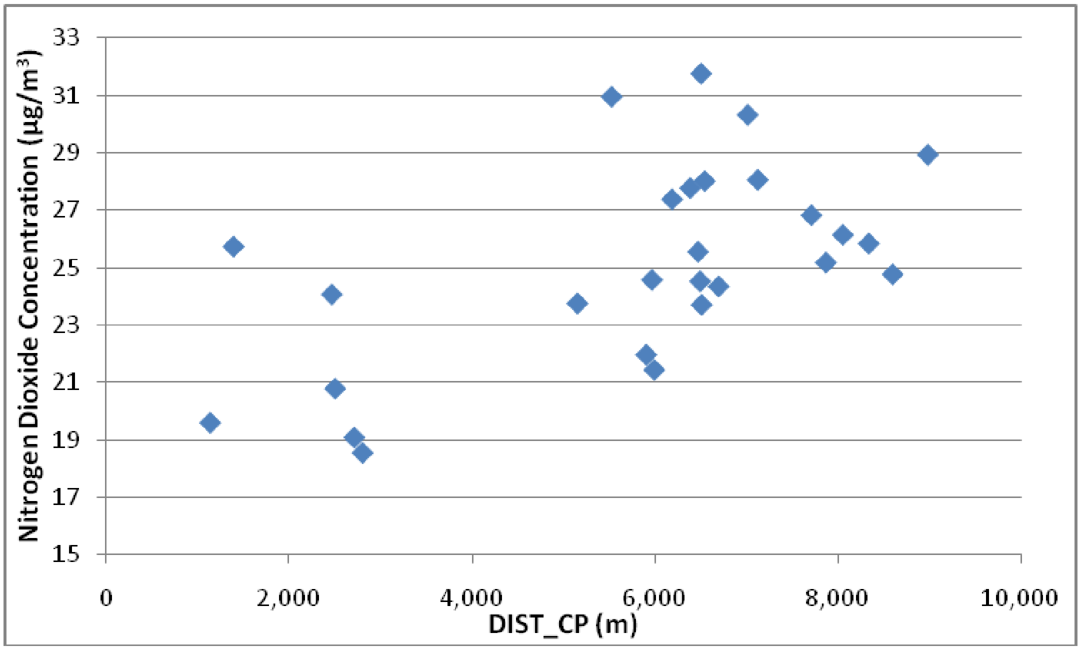

Figure B17: Scatterplot between NO_2 concentration and DIST_CP

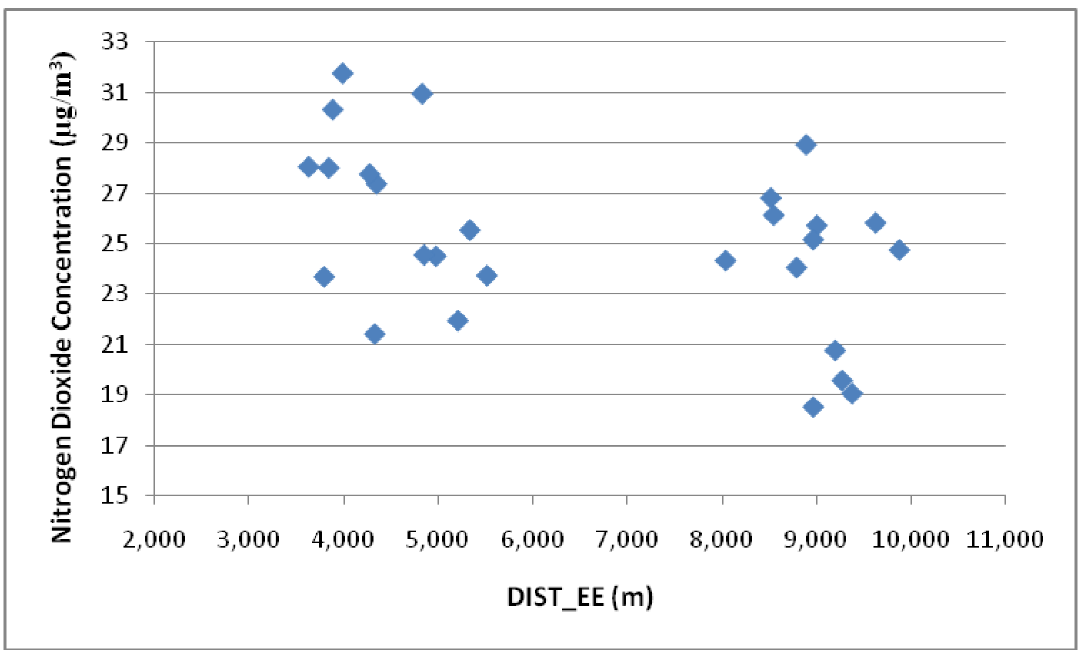

Figure B18: Scatterplot between NO_2 concentration and DIST_EE

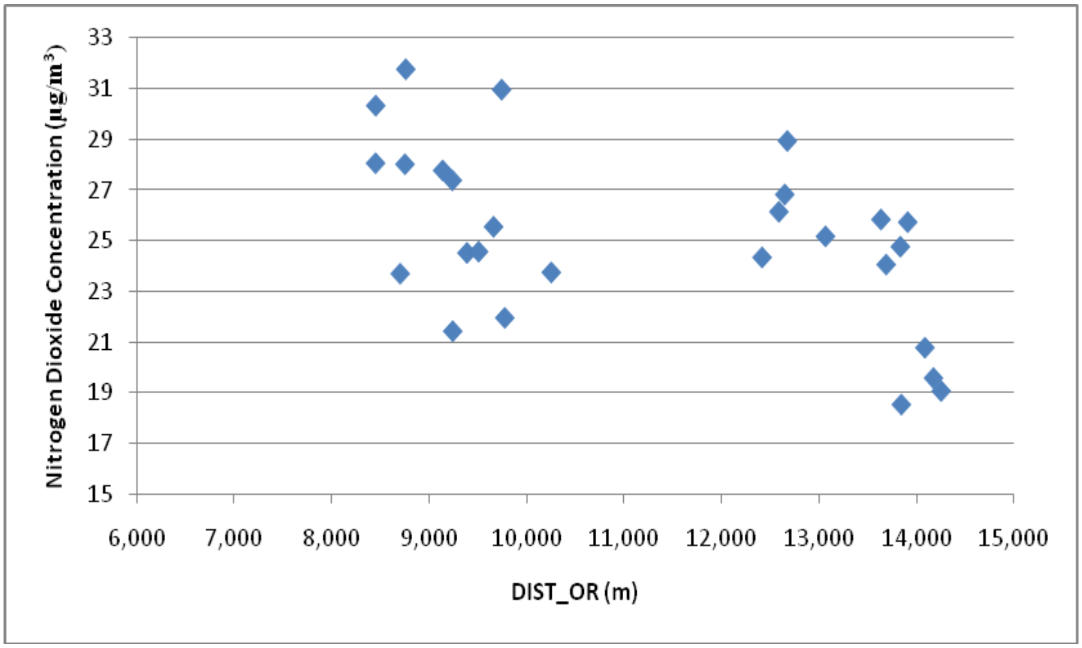

Figure B19: Scatterplot between NO_2 concentration and DIST_OR

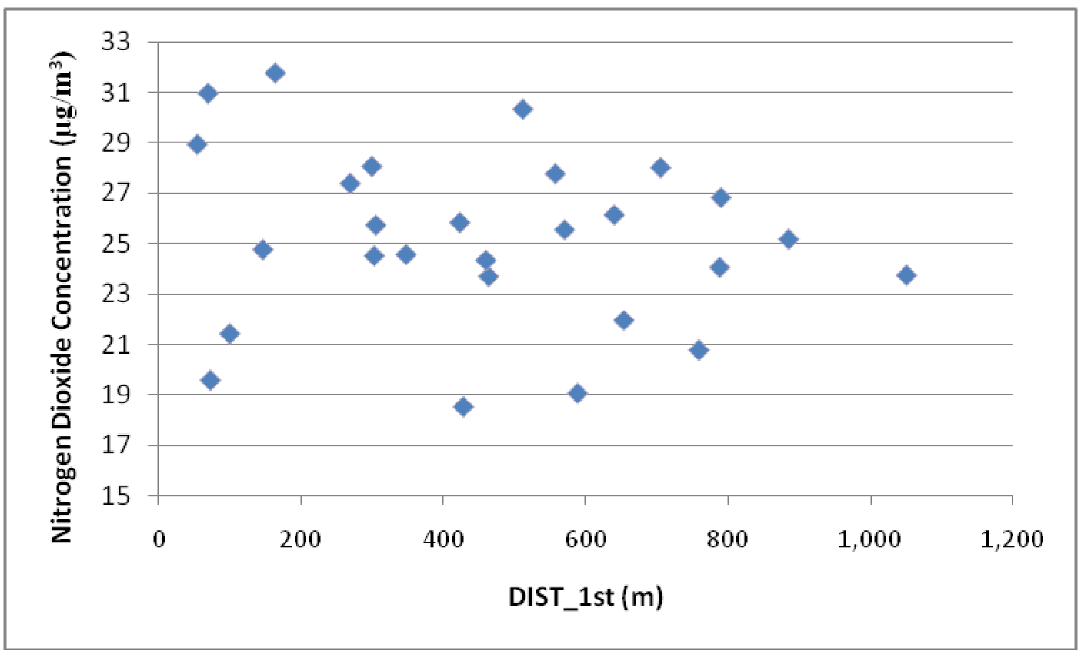

Figure B20: Scatterplot between NO_2 concentration and DIST_1st

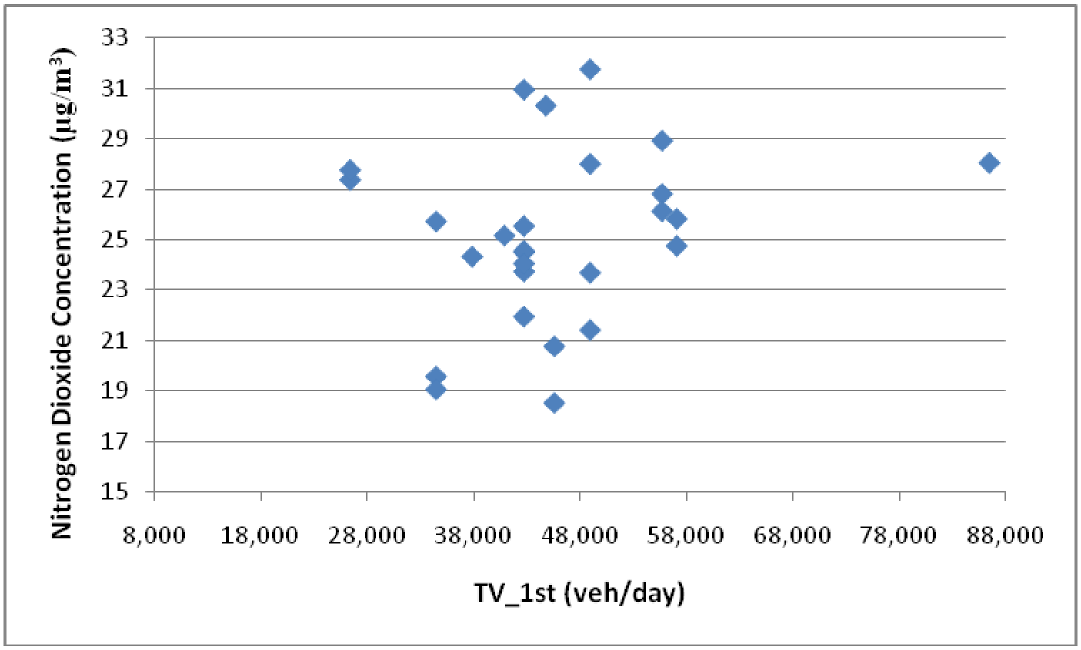

Figure B21: Scatterplot between NO_2 concentration and TV _1st

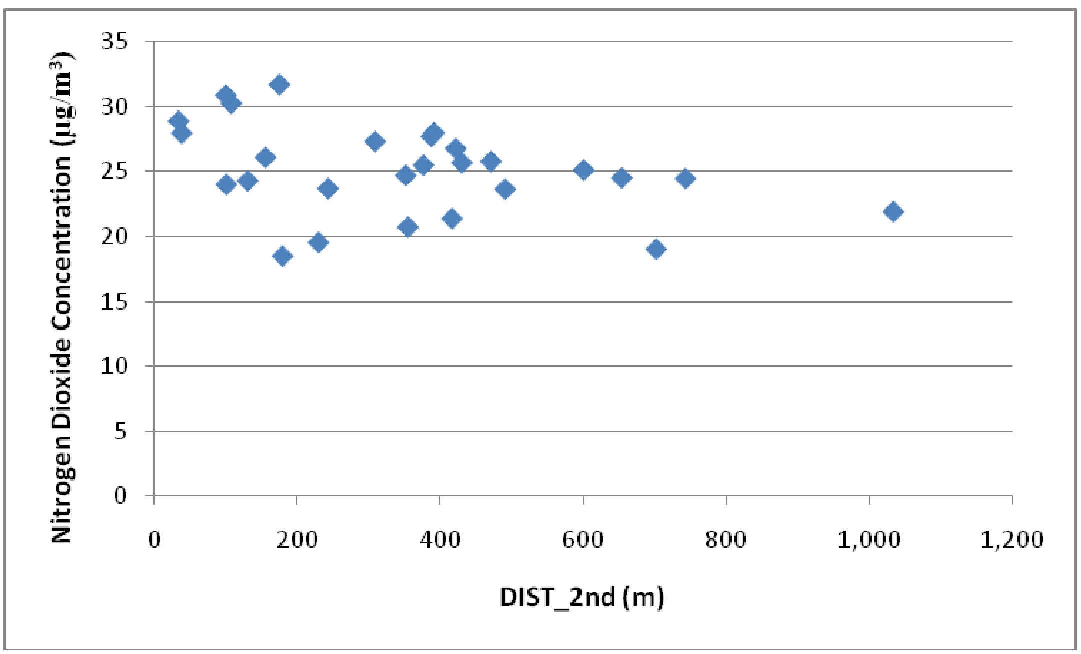

Figure B22: Scatterplot between NO_2 concentration and DIST _2nd

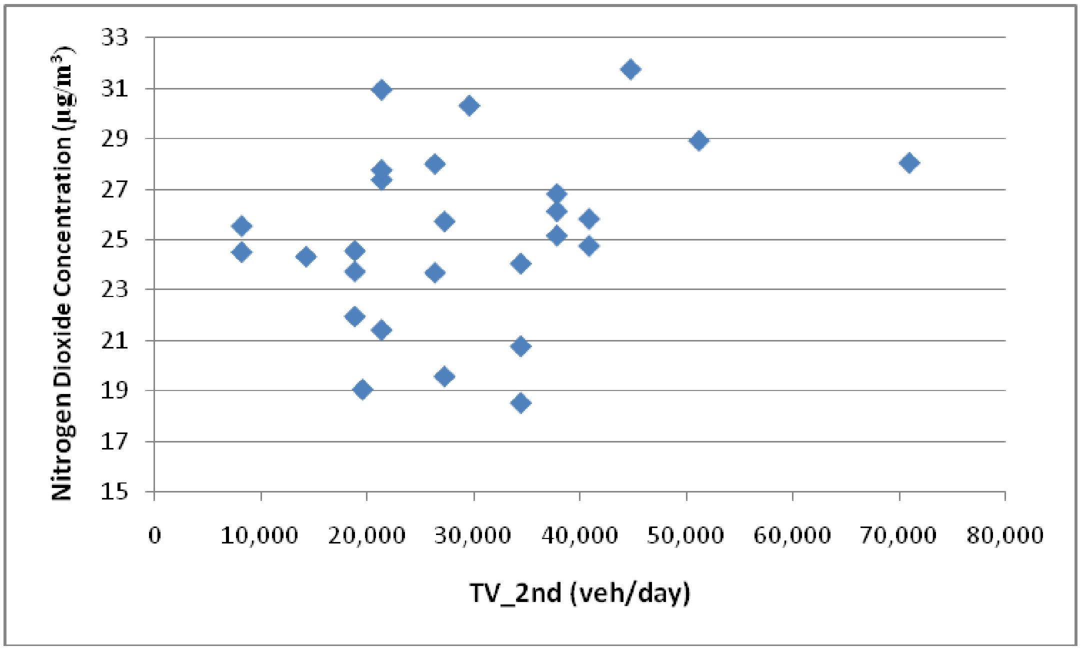

Figure B23: Scatterplot between NO_2 concentration and TV _2^{nd}

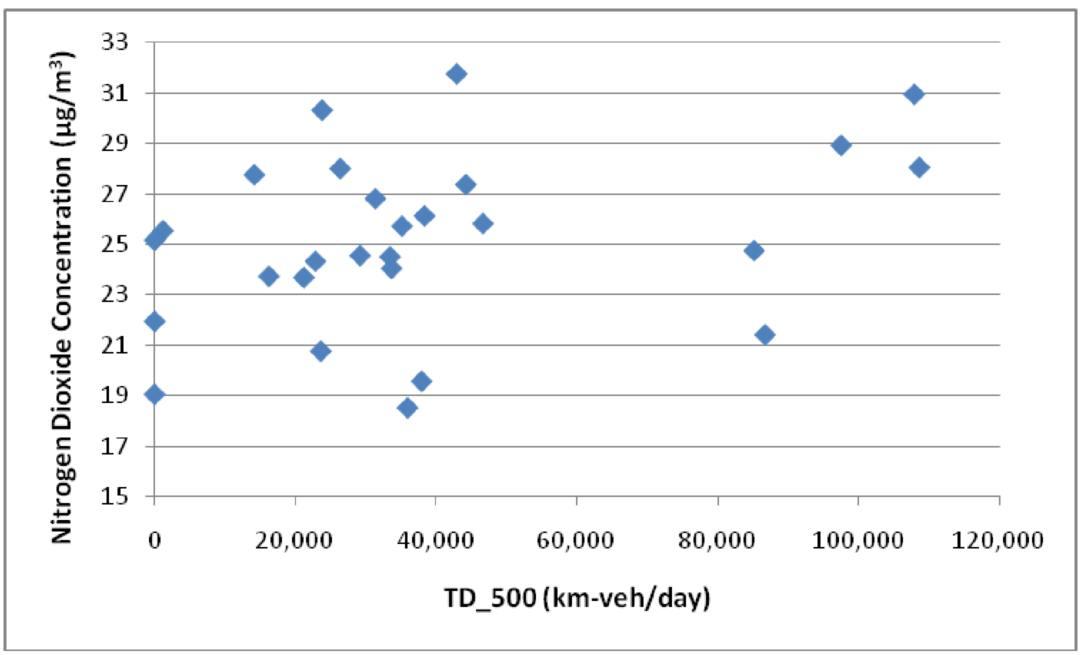

Figure B24: Scatterplot between NO_2 concentration and TD _500

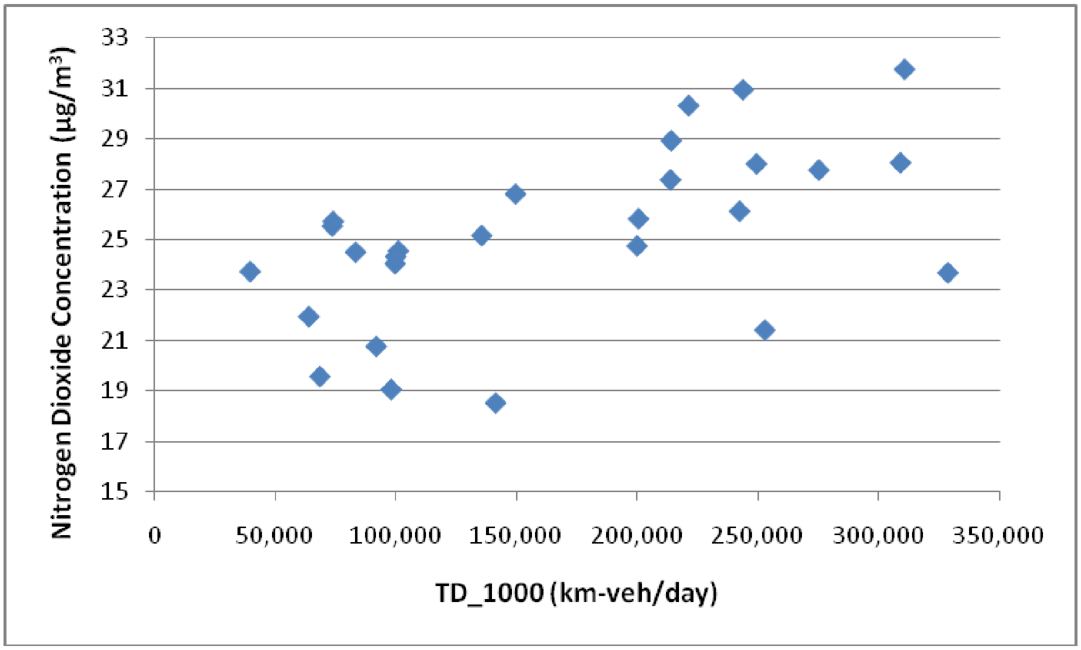

Figure B25: Scatterplot between NO_2 concentration and TD _1000

Vita

Teresa Sosa was born on August 20, 1983 in El Paso, TX. She is the youngest daughter in the family of Teodoro Montoya and Patricia Moreno de Montoya. She graduated from Coronado High School in 2002 and completed her Bachelor's Degree in Civil Engineering from the University of Texas at El Paso in December 2006. In the last year of her master's degree program, Teresa received a fellowship from the National Science Foundation called Bridge to the Doctorate.

Permanent address: 6268 Jerry Turner

El Paso, TX 79932

This thesis/dissertation was typed by Teresa Sosa