

2010-01-01

Assessing Annual And Seasonal Spatial Variability Of Ambient Pm10 Using Linear Regression Analysis In A United States-Mexico Urban Sprawl

Mario Ivan Garcia

University of Texas at El Paso, migarcia2@miners.utep.edu

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Environmental Engineering Commons](#), and the [Environmental Health and Protection Commons](#)

Recommended Citation

Garcia, Mario Ivan, "Assessing Annual And Seasonal Spatial Variability Of Ambient Pm10 Using Linear Regression Analysis In A United States-Mexico Urban Sprawl" (2010). *Open Access Theses & Dissertations*. 2682.
https://digitalcommons.utep.edu/open_etd/2682

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

ASSESSING ANNUAL AND SEASONAL SPATIAL VARIABILITY OF
AMBIENT PM₁₀ USING LINEAR REGRESSION ANALYSIS IN A UNITED
STATES-MEXICO URBAN SPRAWL

MARIO IVAN GARCIA MARTINEZ

Department of Civil Engineering

APPROVED:

Wen-Whai Li, Ph.D., Chair

Charles D. Turner, Ph.D.

Hector A. Olvera, Ph.D.

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

Copyright ©

by

Mario Garcia

2010

Dedication

I humbly dedicate this thesis to my parents and sister who supported me endlessly in this achievement. Also, I want mentioned my lovely wife who took care of me, when I was caring for everything else, who taught me to appreciate everything that I involved on and for being everything I need and the strongest part of me. I would have never accomplished this with the unconditional help of God and la Virgen de Guadalupe who I deeply thank; to everyone I just mentioned I dedicate the fruits of this effort. Nevertheless, I also dedicate this thesis to my unborn child which has been throughout this journey the greatest source of inspiration, thank you
Mia Fernanda Garcia Alvarez.

ASSESSING ANNUAL AND SEASONAL SPATIAL VARIABILITY OF
AMBIENT PM₁₀ USING LINEAR REGRESSION ANALYSIS IN A UNITED
STATES-MEXICO URBAN SPRAWL

By

Mario Garcia, B.S.

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

May 2010

ACKNOWLEDGEMENTS

I would like to thank each one that helped and inspired me during my thesis. My utmost gratitude goes to my thesis advisor Dr. Wen Whai Li for his supervision, advice and guidance from the very early stage of this research. His involvement with his originality has trigger and nourished my intellectual maturity that exceptionally inspire and enrich my growth as a student, a researcher and a scientist. I am indebted to him more than he knows. I would also like to record my sincere gratitude to Dr. Hector Olvera, for his advice, guidance, and crucial contribution, which made him a backbone of this research and so to this thesis. His truly scientist intuition has made him a constant oasis of ideas and passions in science. He quickly became a role model of a successful researcher in the field for his expertise and research insight. I convey special acknowledgement to Jessica Gamez, who first brought me into the world of research and with whom I began to learn about Air Quality. Not only a great mentor and colleague, she has also been a cornerstone in my professional development.

I will like to dedicate my thesis to my family for their unflagging love and support throughout my life; this thesis is simply impossible without them. I have no suitable word that can fully describe the everlasting love of my mother, Marcel Martinez. I remember many sleepless nights with her accompanying me with a cup of coffee. I remember her constant support when I encountered difficulties and I remember, most of all, her summer math lessons at age five that probably determine the course of my life towards engineering. Mother I will always love you. I am indebted to my father, Mario Garcia, for his care and love. As a typical father in a Mexican family, he worked industriously to support the family and spare no effort to provide the best possible environment for me to grow up and attend school even when doing so meant depriving of certain needs and constant criticisms' from his family. He had never complained in spite of all the hardships in his life. I feel proud of my sister, Marcela Garcia for her courage, free spirit, and determination in life. Finally my wife, Karla Alvarez, who is my inspiration,

because she taught me that only those who have the patience to do simple things perfectly will acquire the skill to do difficult things easily; my muse because she showed me that humbleness is the secret to knowledge and wisdom; and the love of my life because when moments of weakness force me to set aside and mourn she would always sit beside me and shed her tears with me.

I am grateful to have friends like Sergio Saenz, Adrian Muñoz and Juan Sanchez who always found the way to make me laugh even when everything seemed to collapse under my shoulders.

Last but not least I would like to thank everybody who was important to the successful realization of this thesis, as well as expressing my apology that I could not mention personally one by one. I would like to finish by citing a reflection that always encourages me through life. An unknown author tells that two children were skating on a frozen pond, when suddenly the ice broke, and one child fell into the water. The other child seeing that his friend was drowning underneath the ice, took a stone and began to strike the ice with all his strength until he managed to break it and thus save his friend. When firefighters arrived and saw what had happened, they wondered: How did he? The ice is too thick, it is impossible he could have been able to break it with that stone and even so with his hands so small. Just then an old man appeared and said: “I know how he did it.” “how?” asked a firefighter. And the old man answered: - “there was nobody around to tell him that it was impossible to do it.”

ABSTRACT

As air quality issues grow progressively in the consciousness of the global community, stakeholders are pressing for epidemiologic studies of air pollution. The effects of chronic exposure to ambient air pollution remain a difficult challenge due to its substantial small-scale spatial variation. Recent approaches to assess intra-urban exposure have employed the use of proximity-based assessment, interpolation methods, emission-meteorological models, dispersions models, and land-use regression (LUR) models.

This thesis assesses the spatial variability of ambient particulate matter (PM_{10}) obtained by five self-governing models for the area of El Paso. Using multiple linear regression analysis five regression-based equations were developed to predict annual and seasonal intra-urban gradient of PM_{10} . The first four models were limited to seasonal analysis in order to evaluate the sensitivity of PM_{10} mass concentration to conditions attributable specifically to each seasonal period such as: temperature, human activities, diurnal patterns and atmospheric stability; as well as to explore short-term spatial variability for each season. The fifth and final model was developed to estimate annual individual levels of long-term exposure of ambient PM_{10} . Integrated measurements of 7-day periods were collected as part of a three-year monitoring campaign. The samples were measured employing Dichotomous samplers located at 13 sites spread throughout the study region. Three of these sites were co-located with the Texas Commission on Environmental Quality (TCEQ) monitoring stations and thus correlated with the measurements of TCEQ's Tapered Element Oscillating Microbalance (TEOM) samplers for quality-assurance and quality-control (QA/QC). These concentrations were used to fit a multiple linear regression model to assess the association between logarithmic concentrations of PM_{10} and surrogate parameters derived using ESRI ArcGIS 9.2 geographic information system. Integrated local spatial covariates regarding land use, traffic-related, population, property value and physical geography were

derived from circular buffers with radii from 200 m to 5000 m around the monitors. The validated models explained between 74% to 90% of the variation in PM₁₀ concentrations associated with four variables: zoning, traffic-based, nearest distance to the border area, and population. The PM₁₀ surface area was generated with a 3528 grid points distributed orthogonally at a 610 m isotropic resolution capable of characterizing small-scale variations and capturing the relative impact of the bordering urban areas. Overall, high concentrations were perceived in the fall model which continue during the winter season and lessened during the spring where finally moderates abruptly during the summer season. The five models, including the annual model, concede with a southward pattern where all of the cases identified the downtown core area as a hot spot where the air quality is often under relative stress when compared to other areas. In the case of exposure and habitats, the higher burdens seen southward is partly due to the higher pollutant concentrations found in these areas. PM₁₀ concentrations, for example, are greatly increased in the downtown core largely because of the human activity in the area; particulate concentrations are raised not only by local emissions but also by emissions from Ciudad Juarez. In fact, 2.3, 84, and 4% of the population of El Paso were found to be exposed to an increase of 10 µg/m³ of PM₁₀ from spring to summer, summer to fall, and fall to winter season respectively. Developing a health risk of 1 to 10% increase in symptoms related to cough, decline lung function, and asthma attacks. The predictive maps from LUR analysis appear to capture small-area variations in PM₁₀ concentrations assessing short- and long-term exposure. The application of such prediction models showed that a substantial fraction of the variability in PM₁₀ concentrations was explained by land use-zoning variables. This approach results in optimal epidemiologic studies and air-quality management as it reduces the potential for exposure misclassification posed by other modeling methods.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
ABSTRACT	vii
TABLE OF CONTENTS	ix
LIST OF TABLES.....	xiii
LIST OF FIGURES	xv
LIST OF EQUATIONS	xviii
LIST OF ACRONYMS	xx
1 INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM STATEMENT	2
1.3 OBJECTIVE.....	5
1.4 THESIS STRUCTURE	6
2 LITERATURE REVIEW	8
2.1 MODELS.....	8
2.1.1 Proximity-Based Models	8
2.1.2 Statistical-Interpolation Models.....	9
2.1.3 Dispersion Models	11
2.1.4 Land-Use Regression Models.....	13
2.3 LAND-USE REGRESSION MODELS.....	14
2.3.1 Land use Regression Models for NO _x	14
2.3.2 Land use Regression Models for PM _{2.5}	19
2.3.3 Land use Regression Models for PM ₁₀	24
2.4 AMBIENT PARTICULATE MATTER (PM ₁₀)	26
2.4.1 Overview.....	26
2.4.2 Health Effects	28
2.4.3 Sources.....	32
2.2 GEOGRAPHIC INFORMATION SYSTEMS.....	34
2.2.1 Overview.....	34

2.2.2 GIS Applications	35
3 STUDY AREA	38
4 MONITORING PROGRAM.....	41
4.1 SAMPLING LOCATION	41
4.2 SAMPLING EQUIPMENT	44
4.3 FIELD OPERATIONS.....	46
4.4 SAMPLING ANALYSIS.....	49
4.5 DATASET VALIDATION	51
5 GEOGRAPHICAL DATA	66
5.1 BUFFERING	66
5.2 TRAFFIC-BASED VARIABLES	67
5.3 POPULATION	69
5.4 LAND USE-BASED VARIABLES	70
5.5 ADDITIONAL VARIABLES	70
6 CONSTRUCTION OF MODELS	73
6.1 MODEL SELECTION.....	73
6.2 MAP-BASED GENERATION	76
6.3 CROSS-VALIDATION	83
7 ANNUAL MODEL	88
7.1 PM ₁₀ MEASUREMENTS	88
7.2 REGRESSION ANALYSIS	90
7.3 MODEL VALIDATION	95
7.4 CROSS-VALIDATION	105
7.5 DISCUSSION	110
8 SPRING MODEL	120
8.1 PM ₁₀ MEASUREMENTS	120

8.2 REGRESSION ANALYSIS	122
8.3 MODEL VALIDATION	125
8.4 CROSS-VALIDATION	133
8.5 DISCUSSION	136
9 SUMMER MODEL.....	144
9.1 PM ₁₀ MEASUREMENTS	144
9.2 REGRESSION ANALYSIS	146
9.3 MODEL VALIDATION	149
9.4 CROSS-VALIDATION	156
9.4 DISCUSSION	161
10 FALL MODEL	170
10.1 PM ₁₀ MEASUREMENTS	170
10.2 REGRESSION ANALYSIS	171
10.3 MODEL VALIDATION	174
10.4 CROSS-VALIDATION	182
10.5 DISCUSSION	186
11 WINTER MODEL.....	194
11.1 PM ₁₀ MEASUREMENTS	194
11.2 REGRESSION ANALYSIS	195
11.3 MODEL VALIDATION	198
11.4 CROSS-VALIDATION	206
11.5 DISCUSSION	210
12 CONCLUSION	219
REFERENCES	226
APPENDIX A.....	236
APPENDIX B.....	248

APPENDIX C	250
APPENDIX D.....	261
APPENDIX E	266
APPENDIX F	269
APPENDIX G.....	275
APPENDIX H.....	278
CURRICULUM VITA	299

LIST OF TABLES

Table 4-1. Description of the 13 monitoring sites throughout the El Paso region	42
Table 4-2. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the annual model.....	53
Table 4-3. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the spring model	56
Table 4-4. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the summer model	58
Table 4-5. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the fall model	61
Table 4-6. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the winter model	63
Table 7-1. Descriptive statistics of the average annual concentrations for the 13 monitoring stations	88
Table 7-2. Land use regression model summary of the annual iteration.....	91
Table 7-3. Statistics of the model's coefficients for the annual iteration	92
Table 7-4. Collinearity diagnostics of the model's parameters for the annual iteration.....	94
Table 7-5. Descriptive statistics of the annual model's residuals.....	98
Table 7-6. Shapiro-Wilk's W test for normality for the annual model's residuals	101
Table 7-7. Cross-validation results of the five interpolation techniques used to map annual PM ₁₀ mass concentration throughout the El Paso cohort study	107
Table 8-1. Descriptive statistics of the average spring concentrations for the 13 monitoring stations	120
Table 8-2. Land use regression model summary of the spring iteration	123
Table 8-3. Statistics of the model's coefficients for the spring iteration.....	123
Table 8-4. Collinearity diagnostics of the model's parameters for the spring iteration	124
Table 8-5. Descriptive statistics of the spring model's residuals	127
Table 8-6. Shapiro-Wilk's W test for normality of the spring model's residuals	129
Table 8-7. Cross-validation results of the five interpolation techniques used to map spring PM ₁₀ mass concentration throughout the El Paso cohort study	133
Table 9-1. Descriptive statistics of the average summer concentrations for the 13 monitoring stations	144

Table 9-2. Land use regression model summary of the summer iteration.....	147
Table 9-3. Statistics of the model's coefficients for the summer iteration	148
Table 9-4. Collinearity diagnostics of the model's parameters for the summer iteration	148
Table 9-5. Descriptive statistics of the summer model's residuals	151
Table 9-6. Shapiro-Wilk's W test for normality of the summer model's residuals	152
Table 9-7. Cross-validation results of the five interpolation techniques used to map summer PM ₁₀ mass concentration throughout the El Paso cohort study	157
Table 10-1. Descriptive statistics of the average fall concentrations for the 13 monitoring stations	170
Table 10-2. Land use regression model summary of the fall iteration	172
Table 10-3. Statistics of the model's coefficients for the fall iteration	173
Table 10-4. Collinearity diagnostics of the model's parameters for the fall iteration	174
Table 10-5. Descriptive statistics of the fall model's residuals	176
Table 10-6. Shapiro-Wilk's W test for normality of the fall model's residuals	178
Table 10-7. Cross-validation results of the five interpolation techniques used to map fall PM ₁₀ mass concentration throughout the El Paso cohort study	183
Table 11-1. Descriptive statistics of the average winter concentrations for the 13 monitoring stations	194
Table 11-2. Land use regression model summary of the winter iteration	196
Table 11-3. Statistics of the model's coefficients for the winter iteration.....	197
Table 11-4. Collinearity diagnostics of the model's parameters for the winter iteration	198
Table 11-5. Descriptive statistics of the winter model's residuals	200
Table 11-6. Shapiro-Wilk's W test for normality of the winter model's residuals	202
Table 11-7. Cross-validation results of the five interpolation techniques used to map winter PM ₁₀ mass concentration throughout the El Paso cohort study	207
Table 12-1. General overview of the annual and seasonal land use regression models of PM ₁₀	220

LIST OF FIGURES

Figure 3-1. El Paso del Norte delimiting Ciudad Juarez, Mexico; El Paso, Texas; and Sunland Park, New Mexico. Source (Map data © Google, INEGI)	38
Figure 3-2. Temperature average variations for the three-year monitoring campaign. Source (Climate Charts ©2007).....	39
Figure 4-1. Distribution of the thirteen Dichotomous samplers throughout the El Paso region	43
Figure 4-2. Dichotomous sampler allocated at each site. Source (Thermo Scientific Dichotomous manual 1996)	45
Figure 4-3. Dichotomous virtual Impactor. Source (Thermo Scientific Dichotomous manual 1996)	45
Figure 4-4. Field operation route 1 comprising sites A, D, and F. Source (Navteq 2009).....	46
Figure 4-5. Field operation route 2 comprising sites H, I, and K. Source (Navteq 2009).....	47
Figure 4-6. Field operation route 3 comprising sites J and L. Source (Navteq 2009).....	47
Figure 4-7. Correlation of the monitoring site F and TCEQ’s CAMS 12 analyzer for the annual dataset	54
Figure 4-8. Correlation of the monitoring site H and TCEQ’s CAMS 41 analyzer for the annual dataset	54
Figure 4-9. Correlation of the monitoring site J and TCEQ’s CAMS 49 analyzer for the annual dataset	55
Figure 4-10. Correlation of the monitoring site F and TCEQ’s CAMS 12 analyzer for the spring dataset	56
Figure 4-11. Correlation of the monitoring site H and TCEQ’s CAMS 41 analyzer for the spring dataset	57
Figure 4-12. Correlation of the monitoring site J and TCEQ’s CAMS 49 analyzer for the spring dataset	57
Figure 4-13. Correlation of the monitoring site F and TCEQ’s CAMS 12 analyzer for the summer dataset	59
Figure 4-14. Correlation of the monitoring site H and TCEQ’s CAMS 41 analyzer for the summer dataset	59
Figure 4-15. Correlation of the monitoring site J and TCEQ’s CAMS 49 analyzer for the summer dataset	60

Figure 4-16. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the fall dataset	61
Figure 4-17. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the fall dataset	62
Figure 4-18. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the fall dataset	62
Figure 4-19. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the winter dataset	64
Figure 4-20. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the winter dataset	64
Figure 4-21. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the winter dataset	65
Figure 7-1. Linearity test between the sample concentrations and the predicted values for the annual model	96
Figure 7-2. Residual's distribution of the annual model	100
Figure 7-3. Scatterplot of the standardized residuals versus standardized predicted values for the annual model.....	102
Figure 7-4. The funnel shape pattern attributable to the violation of homoscedasticity. Source (Wikipedia ©2009)	103
Figure 7-5. Weighted case of OND for the annual model	104
Figure 7-6. Weighted case of NLR5K for the annual model	105
Figure 7-7. Weighted case of GZ5m for the annual model	105
Figure 7-8. Annual PM ₁₀ spatial gradient using Universal Kriging interpolation method.....	119
Figure 8-1. Linearity test between the sample concentrations and the predicted values for the spring model	126
Figure 8-2. Residual's distribution of the spring model	128
Figure 8-3. Scatterplot of the standardized residuals versus standardized predicted values for the spring model	130
Figure 8-4. Boxplot of the standardized predicted values and standardized residuals for the spring model	130
Figure 8-5. Weighted case of NGZ15k for the spring model	132
Figure 8-6. Weighted case of npop15k for the spring model	132

Figure 8-7. Spring PM ₁₀ spatial gradient using Universal Kriging interpolation method.....	143
Figure 9-1. Linearity test between the sample concentrations and the predicted values for the summer model	150
Figure 9-2. Residual's distribution of the summer model	152
Figure 9-3. Scatterplot of the standardized residuals versus standardized predicted values for the summer model	153
Figure 9-4. Weighted case of OND for the summer model.....	155
Figure 9-5. Weighted case of TR5m for the summer model	155
Figure 9-6. Weighted case of lc2m for the summer model	156
Figure 9-7. Summer PM ₁₀ spatial gradient using Universal Kriging interpolation method.....	169
Figure 10-1. Linearity test between the sample concentrations and the predicted values for the fall model	175
Figure 10-2. Residual's distribution of the fall model.....	178
Figure 10-3. Scatterplot of the standardized residuals versus standardized predicted values for the fall model	179
Figure 10-4. Boxplot of the standardized predicted values and standardized residuals for the fall model	180
Figure 10-5. Weighted case of npop15k for the fall model.....	181
Figure 10-6. Weighted case of ntr1k for the fall model	181
Figure 10-7. Weighted case of NGZ3m for the fall model.....	182
Figure 10-8. Fall PM ₁₀ spatial gradient using Ordinary Kriging interpolation method	193
Figure 11-1. Linearity test between the sample concentrations and the predicted values for the winter model	199
Figure 11-2. Residual's distribution of the winter model.....	202
Figure 11-3. Scatterplot of the standardized residuals versus standardized predicted values for the winter model	203
Figure 11-4. Boxplot of the standardized predicted values and standardized residuals for the winter model	204
Figure 11-5. Weighted case of vf15k for the winter model.....	205
Figure 11-6. Weighted case of pop5m for the winter model	206
Figure 11-7. Weighted case of GZ5m for the winter model.....	206
Figure 11-8. Winter PM ₁₀ spatial gradient using Ordinary Kriging interpolation method	218

LIST OF EQUATIONS

Equation 3-1. PM _{2.5} Fine particulate matter mass concentration	51
Equation 3-2. PM _{10-2.5} Coarse particulate matter mass concentration	51
Equation 3-3. PM ₁₀ Particulate matter mass concentration	51
Equation 6-1. Land use multiple linear regression	73
Equation 6-2. Inverse distance weighted interpolation technique	77
Equation 6-3. Natural neighbor interpolation technique	78
Equation 6-4. SPLINE interpolation technique	79
Equation 6-5. Governing linear equation for estimating the coefficients in the SPLINE equation	79
Equation 6-6. Linear equation for estimating coefficient A_i in the SPLINE equation.....	79
Equation 6-7. Linear equation for estimating coefficient A_i at the x_i position for the SPLINE equation	79
Equation 6-8. Linear equation for estimating coefficient A_i at the y_i position for the SPLINE equation	79
Equation 6-9. Ordinary kriging interpolation technique equation.....	80
Equation 6-10. Simultaneous equation used to determine the semivariance of the point at x_i and y_i	81
Equation 6-11. Simultaneous equation used to determine the semivariance of the point at x_j and y_j	81
Equation 6-12. Simultaneous equation used to determine the semivariance of the point at x_k and y_k	81
Equation 6-13. Governing equation used to determine the weight of $x_{i,j,k}$ and $y_{i,j,k}$	81
Equation 6-14. Universal kriging first order polynomial (plane surface) interpolation technique	82
Equation 6-15. Universal kriging second order quadratic surface polynomial	82
Equation 6-16. Mean of the entire n subsamples for the bootstrap cross-validation equation	84
Equation 6-17. Standard error of the mean of the entire n subsamples for the bootstrap cross- validation equation.....	84

Equation 6-18. Mean estimation for the n -th observations without point i for the jackknife cross-validation equation.....	85
Equation 6-19. Mean estimation for the whole sample for the jackknife cross-validation equation	85
Equation 6-20. Pseudo-value for the jackknife cross-validation equation	86
Equation 6-21. Root mean square error cross-validation technique.....	86
Equation 6-22. Standard root mean square error cross-validation technique	86
Equation 7-1. Multiple linear regression equation for the annual model	95
Equation 7-2. Durbin-Watson test	97
Equation 8-1. Multiple linear equation for the spring model	125
Equation 9-1. Multiple linear equation for the summer model	149
Equation 10-1. Multiple linear equation for the fall model	174
Equation 11-1. Multiple linear equation for the winter model	198

LIST OF ACRONYMS

AATC	Annual average traffic counts
ANOVA	Analysis of variance
AOR	Adjusted odd ratio
APMOSPHERE	Air pollution modeling for support to policy and health environmental risks in Europe
BLUE	Best linear unbiased estimate
BOTA	Bridge of the Americas
CAL	Chronic airflow limitation
CAMS	Continuous monitoring sites
CB	Corrected buffer
CCAAPS	Cincinnati childhood allergy and air pollution
CDJ	Ciudad Juarez
CERM	Center for Environmental Resource Management
CI	Confidence interval
DEM	Digital elevation model
DICHOT	Dichotomous sampler
DRI	Desert Research Institute
ECAT	Element carbon attributable to traffic source
EPA	Environmental Protection Agency
EPWU	El Paso Water Utilities
ER	Emergency room
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GZ5m	Group zoning in a 500 m buffer
H ₀	Null hypothesis
H ₁	Alternative hypothesis
HEI	Health Effects Institute
HHNE	Hispanic Health and Nutrition Examination survey
I-10	Interstate 10
ICE	Immigration and Custom Enforcement
IDW	Inverse distance weighted
IQR	Inter quartile range
lc2m	Land use commercial in a 200 m buffer
LUR	Land use regression
MAE	Mean absolute error
MLR	Multiple linear regression
MPO	Metropolitan Planning Organization
NAAQS	National Ambient Air Quality Standards
NAFTA	North American Free Trade Agreement
NGZ15k	Normalized group zoning in a 1.5 km buffer
NGZ3m	Normalized group zoning in a 300 m buffer
NIEHS	National Institute of Environmental Health Sources

NIH	National Institutes of Health
NLR5K	Normalized length of road in a 5 km buffer
NN	Natural neighbor
NO	Nitric oxide
NO ₂	Nitrogen dioxide
npop15k	Normalized population in a 1.5 km buffer
NRW	North-Rhine Westphalia
ntr1k	Normalized length of truck roads in a 1 km buffer
OK	Ordinary kriging
OND	Opposite nearest distance
OSPM	Operational street pollution model
PdN	Paso del Norte
PM	Particulate matter
PM _{10-2.5}	Coarse particle
ppb	Parts per billion
QA/QC	Quality assure Quality control
RMSE	Root mean square error
RR	Relative risk
SA-LUR	Source-area land use regression
SAVIAH	Small area variation in air quality and health
SEE	Sum of square errors
SCERP	Southwest Consortium for Environmental Research and Policy
SLP	Sunland Park, New Mexico
SO ₂	Sulfur dioxide
SOP	Standard operating procedure
SRMSE	Standardized root mean square error
TCEQ	Texas Commission on Environmental Quality
TEOM	Tapered Oscillating Microbalance
TR5m	Length of truck roads in a 500 m buffer
TRAPCA	Traffic related air pollution and childhood asthma
TSP	Total suspended particulate matter
UK	Universal kriging
UTEP	University of Texas at El Paso
UTM	Universal Transverse Mercator
VIF	Variance inflation factor
VMT	Vehicle miles traveled
WLS	Weighted least squares
WR	Western Refining

1 INTRODUCTION

1.1 BACKGROUND

Over the past decades air quality issues have grown progressively in the consciousness of the global community. Early research programs were devoted in determining a baseline of pollutants that trigger the development of adverse health conditions in human beings. In essence, researchers were dedicated in identifying the relation between the pollutants' chemical composition and the associated adverse human health effects, solely for the purpose to inform decision-makers and thus heightening environmental policy management. Subsequently, policymakers and scientist shifted their approach into evaluating the health repercussions of concentrations of such pollutants by deriving an exposure assessment based on a dense monitoring network throughout a cohort study. However, most urban networks comprise only a few sites, and those which do exist can rarely be taken as representative of the exposure experience by the population as a whole. More recently, despite the improvements in air quality and measurement technology, health exposure evaluation relied on proximity methods and other techniques that debase characterization of pollutants at a relative coarse resolution misclassifying health effects. Former studies in this field were mainly hampered by severe deficiencies in assessing the individual exposure of the study population. Furthermore, between-city comparisons, which relied on citywide averages, were prone to be confounded by intangible differences between the cities. Consequently, nowadays, researchers and decision-makers have identified, as a priority, the need for a more accurate health assessment of chronic exposure to air pollution in a more precise scale. The development of new technologies has prompt more feasible sophisticated methods for exposure assessment. In addition, providing a more stringent

technique for mapping air pollution variability will evaluate more narrowly the health effects of inhaling ambient air pollutants at any given location within the cohort study exerted by long-term average concentrations.

Exposure models predicting the spatial variation of air pollution within cities have been identified as a research priority. The assessment of air pollution exposure using only community average concentrations likely underestimates the health burden attributable to elevated concentrations in the vicinity of source. Health effects may be more perceptible around sources, and these effects are diminished when using average concentrations for the entire community. Exposure to air pollution, however, may vary spatially within a city, and these variations may follow social gradients that influence susceptibility to environmental exposures. The spatial correspondence between high exposure and potentially susceptible populations within cities may further bias estimates that rely on central monitors to proxy exposures over wide areas. Since direct measurement of the total exposure for each study subject is not feasible, indirect methods of exposure assessment are necessary. While surrogate measures, such as distance to roads, have been related to large health effects, they may misclassify exposure because they are not directly estimated from monitored data. Alternatives to surrogate measures arise from models. Methodologies employed to accomplish this task include: proximity-based assessment, emission-meteorological models, statistical interpolation, Land-use regression, and dispersion models.

1.2 PROBLEM STATEMENT

Overall air pollution is a chronic problem for the nearly two million inhabitants in the El Paso-Juarez region. Air pollution in the region has been considered to be the worst along the U.S.-Mexico border, and concentrations of PM_{10} have frequently exceeded the National Ambient

Air Quality Standards (NAAQS) of both countries. The air pollution problem in the region is also unique from the other metropolitan areas in the U.S. for the significant amount of uncontrolled geologic emission sources (such as desert, sand and dust from unpaved roads), as well as mobile sources, residential cooking and heating methods, and other unidentified anthropogenic emission sources. The arid weather, occasional high winds, frequent stagnations, shallow nighttime and morning mixing depth, and complex topography (elevation) preclude simple explanations for the excessive PM_{10} levels in the region. Therefore, high-resolution maps of air pollution for the El Paso area are needed for assessing exposure as part of epidemiologic studies. The development of methods to estimate individual exposures of study populations within a single urban area will allow for the assessment of chronic exposures and health impacts associated with within city variability of air pollution.

Recent surveys proxy the El Paso, Texas as the 6th ‘worst cities’ in the U.S among the nations 25 asthma hot spots. This according to a new study of the country’s 100 largest metropolitan areas; based on measures of the scope and impact of asthma, as well as data for asthma triggers such as air pollution and tobacco smoke. D’Amato et al. (2000) has recently noted that the prevalence of allergy-related respiratory diseases is increasing especially in industrialized countries like the United States at such an alarming rate that it is being called the “Epidemic of the 21st Century”. There is also mounting evidence over the past few years that ultrafine and especially nanoparticulates (or nanoaerosols with mean diameters <100nm) in the atmosphere are far more toxic than expected and pose considerable health risk including mortality, asthma complications, chronic bronchitis, respiratory tract infections, ischemic heart diseases, and stroke. Lung diseases such as asthma, chronic airflow limitation (CAL) and other chronic lung diseases represent a significant burden of illness in the United States and other parts

of the world as they consume a large amount of resources. Central and peripheral airflow obstructions frequently occur in these conditions and may have different pathophysiological mechanisms of obstruction and required different therapeutic interventions. Moreover, asthma is a complex disease state resulting from the interaction of host, environmental and socioeconomic factors. Disparity occurs in asthma prevalence and severity due to differences in these factors between socioeconomic or ethnic groups. Economic, environmental, and cultural factors may play a role in the resulting disparities between white and Mexican-American children. Economic factors likely contribute to the disparity in hospitalization rates. Hispanic children have been found to have fewer physician visits (Butler et al., 1997) and be more likely to be uninsured, even after controlling for income level. Poverty and lack of health insurance are known barriers to routine care and use of controller medications that prevent hospitalization. Physicians caring for Hispanic children with asthma also identify financial barriers to care as the most important reason. Often, families only seek care when the disease state requires emergency care. When inner city children do have access to specialty care for their asthma, there are demonstrable improvements in asthma outcomes. Several studies have subsequently found associations between mortality from cardiovascular and respiratory diseases and long-term average exposure to air pollution. It has been demonstrated that an increased in PM_{10} is typically associated with a 1-10% increase in acute morbidity symptoms, 10-25% increase in chronic morbidity, and 0.5-1.5% in daily mortality (Pope et al., 2006). Yet uncertainties in exposure assessment methodologies continue to raise questions about reliability and accuracy of risk estimates from intra-urban air pollution studies. To date, most assessments of air pollution using community average concentrations or temporal mismatch variations likely underestimates the health burden attributable to elevated concentrations in the study area. Intra-urban air pollution is characterized

by high spatial variability of pollutants with rapid decay from the source and thus health effects may be larger around sources. These effects, however, are diminished when using average concentrations for the entire community. Former American Cancer Society (ACS) studies have relied on between community exposure contrasts at the scale of a metropolitan area giving all residents of the area the same exposure concentrations or even null spatial concentrations (Dockery et al., 1993; Pope et al., 1995; Krewski et al., 2000; the World Health Organization, WHO, 2001; Pope et al., 2002; Jerrett et al., 2003; and Pope et al., 2004). Exposure to air pollution, however, may vary spatially within a city, and these variations may follow social gradients that influence susceptibility to environmental exposures. These exposure misclassifications may prone bias estimates of short- and long-term health effects to ambient air pollution.

1.3 OBJECTIVE

The aim of this thesis is to develop a land based regression (LUR) model to assess and generate accurate, short-term and long-term high-resolution maps of ambient particulate matter within the El Paso region, as a basis for environmental, health policy and epidemiological research. This technique will be used to predict concentrations at any spatiotemporal resolution within the framework of a Geographic Information System (GIS) using sampled PM₁₀ mass concentrations as the dependent variable, exogenous high resolution geographic data and capable software to characterize small-scale variations of air pollution. Consequently, using a 3528 point grid, the predicted values were interpolated using five different methods in order to map the intra-urban gradient of PM₁₀ for a more accurate exposure assessment as well as to determine the parameters that explain the variability of ambient particulate matter and thus their explanatory

influence over the pollutant of interest at every time period modeled. The regression-based models offer the advantages of accounting for sampling levels of the pollutant of interest as well as for individual exposures of the study population allowing the assessment of chronic exposures and health impacts associated with within-city variability in air pollution and thus minimizing exposure misclassification. Furthermore, the resulting maps will identify areas of redundant concentrations that overstress the local airshed. This will help reorient air pollution studies in relation to the population's associated health effects prone by the high concentration areas identified by the map. Also, the project is expected to make important, albeit often indirect, contributions to longer-term policy developments by informing decision-makers and the public about the problems and possible risks of hot spots.

1.4 THESIS STRUCTURE

This report presents the analyses applied to the mass concentrations obtained at different monitoring sites in a period of three years, for characterizing and identifying the intra urban gradient in the PdN region. Chapter 2 is dedicated to a summary review of several studies performing modeling techniques, GIS and its applications, land use regression analysis used in addressing the local air pollution problems, and PM₁₀ health concerns. Chapter 3 describes the main characteristics of the study area. Chapter 4 describes the monitoring program and their main characteristics as well as operation modes. This chapter delineates the instrumentation and sampling design used for the collection of the monitoring sites, dataset validation of the three co-located monitoring stations with the fixed air quality state stations. Chapter 5 describes in detailed the geographical variables employed in the analysis. Chapter 6 corresponds to the model selection, and the cross-validation method analyzed in this research. Many studies have been

performed at a number of locations for the analysis and evaluation of air quality data. Results and discussions from the application of multivariate analyses to annual and seasonal time periods are considered in Chapter 7, 8,9,10, and 11. Main discussions drawn in this chapter are used to identify the location and nature of the PM sources impacting intra-urban gradient. The overall conclusion is listed and described in Chapter 12. This conclusion integrated the findings of the latter chapters in a global and comprehensive frame to yield a better picture of the findings of this study. The Appendixes contains supplementary data that were not presented in the main text of this report. All the employed databases, automatic procedures tables, and analyses applied to the data sets have been recorded on an accompanying Appendix folder.

2 LITERATURE REVIEW

2.1 MODELS

2.1.1 Proximity-based Models

Proximity models represent the most basic approach in differentiating intra-urban air pollution exposures by measuring the proximity of a subject to a pollution source. This method identifies the relationship between air pollution and health outcomes based on the assumption that nearness to emission sources proxies for higher exposure in human populations. A vast majority of research often focused on the association between road proximity and respiratory disease, lung cancer and stroke mortality with most quantified within a buffer of some predefined extent, to differentially classify exposure based on both potential emissions and distance from source.

Venn et al. (2000) findings suggest that traffic counts or emissions near the residence may exacerbate asthma symptoms, yet perceptible evidence supports a link between asthma onset and intra-urban exposure. As case point, after controlling for confounding effects such as age, sex, and race, English et al. (1999) found no evidence of increased risk of asthma in children under 14 with an increase in traffic counts. Yet among children with asthma, the number of medical care visits escalated with higher traffic counts. Moreover, for a study on adult asthma in Hamilton performed by Jerrett et al. (2002), women, aged 20-44 years, within 50 m of a major road were associated with a 50% increased risk of reporting asthma symptoms, however no significant association was found for males. Conversely, Wilkinson et al. (1999) found no association between children, ages 5-14 years, within 150 m of a main road and the number of hospital admissions for treatment of asthma. Although, none of the asthma studies used a

prospective cohort design to assess the question on asthma formation, still these results must be viewed with this limitation in mind.

Notwithstanding the straightforward application of proximity models for the analysis of long-term exposure classification, this method has considerable limitations. Most studies of this type ignore population exposure to other locations rather than the place of interest, potentially leading to misclassification and biased risk estimates. Also given the fact that air pollution represents a continuous spatial process that will decay with distance away from the source, the use of binary buffers will confer exposure mischaracterization for subjects still in the actual influence zone, but outside the buffer. Nevertheless, surrogate measures of pollution such as distance to roadway may endow correlation with markers of traffic pollution, however in complex environments; these basic methods may misclassify exposure when terrain and meteorological conditions modify the exposure experience. Despite the limitations prompt by this method, proximity models may still contribute to environmental epidemiology as a form of exploratory analysis where aetiological suspicion of effects exists, but there is limited prior evidence.

2.1.2 Statistical-Interpolation Models

Interpolation methods reckon on deterministic and stochastic geostatistical techniques. Measurements of the target pollutant are obtained at a set of monitoring stations distributed throughout the cohort study area. Based on this technique, the purpose is to breed estimates of monitoring stations, usually attained at the center of a grid, imposed over the study area, so that a continuous surface of pollution concentration could be determined. Some methods, such as splines, inverse distance weighting (IDW) and Theissen triangulation rely on deterministic or

geometric algorithms that may produce reasonable estimates of the pollutant at unsample sites, but they offer no means of assessing statistically represented errors in the estimates. These methods are simpler to apply and, in this sense, estimates obtained by them may be more appropriate in instances where the sampling network is limited and errors are assumed to be large. Nonetheless, Kriging methods are known as optimal interpolators considering they supply the best linear unbiased estimate (BLUE) of the variable's value at any point in the study area (Burrough and McDonnell, 1998). A major advantage stems from the production of both predicted values and their standard errors at unsample locations. These standard errors quantify the degree of uncertainty in spatial predictions at unsample sites, illustrating where the interpolation is less reliable (Mulholland et al., 1998). In contrast, Kriging models exploit spatial dependence in the data to develop continuous surfaces of pollution. Beyond random error or noise in the data, spatial dependence embodies two types of effects. Foremost are global trends that measure broad tendency in the data over the entire study area as well as local variations that are a function of distance between the points (Burrough and McDonnell, 1998).

The study by Mulholland et al. (1998) examined the relationship between modeled ozone levels and pediatric asthma exacerbation in three-year successive summers with hourly ozone levels obtained from 10 monitoring stations. The study subjects included pediatric emergency room (ER) visits. To capture spatial and temporal information, the spatial average of daily zone values for the day prior to the ER visit was assigned to each patient using zip code of the residence. Results from the study indicated an underestimation between the temporal variation in daily maximum ozone concentrations and the maximum values estimated by the kriging technique. Despite the underprediction, a positive association between ozone and asthma was

found with a 20 parts per billion (ppb) increase in ambient ozone concentration relating to a 4% increase in the ER visit rate.

However, interpolation techniques commit to algorithms which are mechanistic and thus do not take other factor into account such as terrain or localized patterns in other possible predictors. Likewise, algorithms fall apart at the edges due to the lack of data. As a result, informed subjective editing may be required to produce a more realistic statistical surface. Data may be added between measured data point to eliminate the distance weighting or edge effects commonly produced with regular interpolation algorithms. Additionally, geostatistical modeling entails a reasonably dense network of sampling sites. For an urban area the number of sites varies depending on the scale of analysis, scale of variability in the pollutant, local emissions, desired errors in estimates, topography of the cohort study, and prevailing meteorological conditions. Furthermore, the high cost of primary data collection often means researchers must seize data for short temporal periods, which may or may not adequately represent the long-term distribution of the pollutant. Ultimately, researchers may be forced to rely on one or two proxy pollutants that are easily monitored, leaving key aspects of air pollution such as particles out of the mix.

2.1.3 Dispersion Models

Dispersion models generally relied on Gaussian plume equations (Bellander et al., 2001). They use assumptions about deterministic processes making use of data on emissions, meteorological conditions, and topography in estimating spatial exposure estimates of air pollution concentrations. These techniques have the potential advantage of incorporating both spatial and temporal variations of air pollution within a study area without need for dense

monitoring networks. Emission data are classified into two categories, depending on the type of source: First, stationary sources account for air pollution coming from local sources such as home heating and industries. For each emission point a number of release parameters are collected like annual mass emissions, stack height, diameter, temperature, and vertical emissions velocity in addition to information on facility type and location (Hruba et al., 2001). Emissions data are collected on an annual basis, or by pattern of emissions that reflect hourly rates. Secondly, mobile sources include traffic emissions and re-suspended particles. Traffic emissions are usually pre-modeled with an automatic interface in order to translate traffic counts, type of vehicles, speeds, gradients of the road network, and an additional variety of variables in order to translate them into emissions. After the data requirements have been met and the model calibrated, the dispersion model computes the pollution levels for the desired time interval with further data updates being infrequent. These models can therefore provide high-resolution analysis of patterns in health outcomes and environmental factors, and they can be applied with relatively minor alterations for different study areas (Bellander et al., 2001).

In another study, Barna et al. (2002) examined the inter-suburb dispersion of particulate air pollution in Christchurch, New Zealand, during a wintertime particulate pollution episode, employing CALPUFF. The performance of the model was evaluated against observed PM_{10} time series at four monitoring stations. The values for the correlation coefficient between observed and predicted PM_{10} concentrations were found between 0.49 and 0.77. Therefore, these results suggest that the model had a tendency to overpredict PM_{10} concentrations.

In an extensive study, focused on traffic-related pollution, Venkatram et al. (2003) modeled the concentrations near roadways using AERMOD. The dispersion model estimated 10-min average of NO measured at 7m (Unit 1) and 17m (Unit 2) from the edge of a major highway

in Raleigh, North Carolina, aside to observed concentrations during a field study conducted in August. Consequently, the model performance explained 75% of the variation of the observed concentrations with minor bias at Unit 1 and an overestimate at Unit 2 by about 10%. Yet at both sites, 95% of the model estimates were within a factor of 2.5 from the observations.

Pollution concentrations vary substantially in space and time due to source strength. All the features can be accounted for within the dispersion framework by including point and line source models for both mobile and stationary sources. However, these methods under reluctant circumstances may account for some disadvantages such as: relatively costly data input; unrealistic assumptions about dispersion patterns like Gaussians dispersion; a need for extensive cross-validation with monitoring data; and temporal mismatches in data can possibly cause estimate errors. That is, for non-traffic and heavy diesel traffic-related emissions, a dichotomous relationship often exist between the sample interval for the two major inputs to the dispersion model - emissions and meteorology. Emissions data from point or area sources are often reported as annual emission rates, while for heavy diesel traffic, the patterns of flow do not tend to be accurately characterized by hour traffic counts. In synergy with the inherent variability of meteorological data, which are commonly collected at the hourly time-scale, the unmeasured variability in emission sources may include significant exposure error. Additionally, an overriding obstacle in the implementation of these models is the high level of required programming coupled with fairly expensive hardware requirements.

2.1.4 Land use Regression Models

The land use regression methodology seeks to predict pollution concentrations at a given site based on surrogate variables. This method uses measured pollution concentrations y at

location s as the response variable and land use types x within areas around location s as predictors of the measured concentrations. Regression mapping provides a practical approach for the assessment of exposure to parameters related to pollution. The method entails the use of least-square regression modeling to predict pollution surfaces based on pollution monitoring data and existing exogenous independent variables. The main strength of this technique is the empirical structure of the regression mapping, which allows adaptation to local areas without additional monitoring or data acquisition. It also assists in methods that identify areas requiring more intensive monitoring through the installation of additional stations (Kanaroglou et al., 2003). Additionally, this method can be applied to multiple pollutants with an appropriate calibration and thus provide reliable estimations handling independent variables known to significantly affect the concentration of pollutants with a relative shallow cost to implement. Despite the limitation of this method of mischaracterizing and some times over predicting pollutants when moving to cohort studies to areas with much different land use and topography; yet this technique predicts equally or even more accurately than geostatistical interpolation methods and dispersion models (Hoek et al., 2008).

2.2 LAND USE REGRESSION MODELS

2.2.1 Land use Regression Models for NO_x

Former studies have shown that within-community or intraurban exposure gradients may be associated with larger health effects than the between-community exposures used in earlier studies (Pope et al., 2002). Much less, however, is known about the chronic effects of exposure. The acute health effects of short-term exposures to traffic-related pollution have been widely demonstrated. Hoek et al. (2002) reported a near doubling of cardiopulmonary mortality,

RR=1.95 with 95% significance and CI: 1.09-3.52, for subjects living near major roads in a cohort of 5000 people, where control was available for many confounding variables. Traffic-related air pollution is of particular interest from a regulatory and health perspective, as the demand for transportation will outpace improvements in vehicle technologies over the next decades. Subsequently, utmost of LUR studies have modeled nitrogen dioxide (NO₂) as it is the predominant indicator of mobile source influence. Therefore, concentrations of NO₂ are more spatially diverse over a surface area as it has a two- to threefold difference within 50 m or less (Hewitt, 1991).

Briggs et al. (1997) introduced the first LUR in the Small Area Variations In Air quality and Health (SAVIAH) project, scrutinizing traffic-related air pollution in an individual level to assess the risks of respiratory disease of children in four European cities: Amsterdam, Huddersfield, and Prague. Using 80 passive NO₂ monitors, over a 14-day campaign, and variables relating to traffic (density within 0-30m and 30-400m), land use (high-density housing and industrial), and elevation were used to develop stochastic models. Slightly different models were derived for each city due to the differences in characteristics between cities, as well as differences in data availability. Likewise, all models were constrained by requiring that all regression coefficients have the a priori define sign. The final prediction models explained between 61% and 72% of the observed variability in concentrations between sites. Subsequently, the relationship between the response variable and the predictors was tested for reliability from a small sample of 8-10 references sites in SAVIAH prompting correlations values ranging from 0.79 to 0.87.

Moreover, an updated model described by Briggs et al. (2000) investigated the transferability of the SAVIAH project into four United Kingdom urban areas: Hammersmith and

Ealing, undertaken as part of a project to investigate relationships between hospital admissions for respiratory illness and traffic related air pollution; Sheffield, undertaken as part of a study to examine links between self-reported asthma and traffic-related pollution in adolescent children; Northampton, carried out to assess the impacts of traffic management on air pollution and exposure; and Huddersfield, as part of a baseline model transferred elsewhere. The model was initially recalibrated at the local level using the respective cities' network of 10 or more monitoring sites. In this manner, the model was able to provide estimates of mean annual NO₂ concentrations within a factor of 1.5 of the actual mean in 70-90% of cases and within a factor of 2 between 70-100% of the time. Convincingly, the regression models provided an R² bearing between 0.58-0.76 hence explaining the variability between values modeled and observed mean annual concentrations thus proving the possibility to pool effects into a random effects framework within areas of relative homogeneity of land use, meteorology, and vehicle mix. However, the hypothesis of model transferability becomes more dubious as characteristics defining each cohort study diverge amid themselves. The resulting surface variability of pollutants has been hampered by poor correlations and over predicaments in utmost areas. For instance, Gilbert et al. (2005) applied the coefficients from an Amsterdam model (Briggs et al., 2002) to Montreal, Canada. As presented in their example, model transfer resulted in a surface that lacked spatial variability as it only unraveled 54% of the variation of NO₂ mixing ratio. In addition, the performance with 95 monitoring sites spread throughout all Toronto, Ontario was also comparable to applications in European cities. As a consequence, the spatial variability was over predicted in most areas thus implying a low transferability among cities mainly due to the differences in population density, vehicle fleets and city configurations between European and North American cities (Jerrett et al., 2007).

More recently, former studies in North America have expanded variable selection to include distances to major water bodies. Ross et al., (2006) developed land use regression models using traffic, distances to the coast and road length measures that explained nearly 80% of the variation in NO₂ levels in San Diego, California and were able to predict validation locations (locations that were not included in the modeling) to within, on average, 2.1 ppb.

In addition, a study over North-Rhine Westphalia (NRW), Germany; a population consisting of birth cohorts was appraised in whether development of inhalant allergy, asthma and other respiratory conditions in children were influenced by the impact of traffic-related air pollution. The findings of Hochadel et al. (2006) corroborated the hypothesis that traffic-related variables influenced the surface variability of NO₂ in the coordinates of interest as well as the respiratory conditions in children; as the methodology manifests that 90% of the variation of NO₂ was directly correlated to traffic-based parameters despite the variable oxidation time of the mainly NO-emissions to NO₂. As traffic models can be use to assign traffic counts to other roads, and in some cities modeled data are available from transport and highways authorities, several LUR studies have successfully explored the use of the length of specific road types without traffic intensity data (Henderson et al., 2007). In the Vancouver model Henderson et al. (2007) performed two 14-day monitoring periods using 15 routine monitoring sites for a 5-year campaign. The observed two period values resulted in average NO₂ that were within 15% of the actual annual mean for 75 out of 75 cases. Using 55 potential predictors and following the approach of using a priori definition of a required sign of regression slopes for specific variables, NO and NO₂ models gave an R² of 0.62 and 0.57 respectively. Moreover, few studies that evaluated NO found similar R² values to those for NO₂ in the same city. Notwithstanding, Henderson et al. (2007) regression analysis method demonstrated that the NO map developed for

Vancouver showed more pronounced small-scale spatial contrast than the NO₂ map, suggesting the importance of monitoring a primary pollutant.

Albeit, wind direction is an important variable that determines the impact of sources on receptor points, though it has seldom been incorporated into land-use regression models, Arain et al., (2007) assessed the use of wind fields to improve the prediction of air pollution in the Toronto-Hamilton urban airshed. Wind direction fields were constructed from 38 weather stations in the area and the constructed wind direction was found to be a significant predictor of NO₂ concentration. The R² of the prediction increased from 0.65 without wind to 0.69 when wind direction was added to the model. It is not obvious whether the work required to construct the wind fields was worth the effort given this modest increase in prediction. As the author appropriately point out, wind fields in urban areas should be very local in order to be useful. Another approach is taken by a group of Canadian researchers who develop a source-area land use regression (SA-LUR) model, by combining concepts of a box-type dispersion model and land use regression to provide more detailed temporal resolution (Su et al., 2008). Different scenarios with varying degrees of complexity of input data like wind direction and mixing height were evaluated. The SA-LUR model was not superior to LUR in predicting long-term average concentrations, but did provide the flexibility for developing shorter-term exposures (Su et al., 2008). A key problem in the method was the ability to obtain sufficiently spatially detailed meteorological data.

Lastly, the first regression analysis technique performed in the El Paso-Juarez region is inferable to Gonzales et al. (2005). The aim of their study was to characterize the gradient of exposures across the city, particularly in the school district region where children's respiratory health remained overlooked. As the author hypothesized that an intra-urban gradient of NO₂ was

influenced by mobile sources, integrated 7-day measurements of NO₂ were collected at 20 elementary schools bestowing two-sided Ogawa passive diffusion samplers during winter atmospheric conditions. Consequently the NO₂ measurements collected by the passive diffusion tubes at the four monitoring stations co-located with the chemiluminescence analyzers were highly correlated (R^2 of 0.74) with precision of 2.5 ± 2.2 ppb. However, the passive monitors overestimated the analyzer measurements, the +16% bias was not significant in this sample and is similar to the +18% to +35% biases previously reported for passive NO₂ diffusion tube samplers (Hamilton and Heal, 2004). Nevertheless, the regression-based analysis instigated that 81% of the variance in the passive measurements were influenced by surrogate parameters like elevation, distances to a main highway, and distance to an international port of entry from Mexico.

2.2.2 Land use Regression Models for PM_{2.5}

Studies have shown a correlation between air pollution and allergic and respiratory diseases (Venn et al., 2001). In contrast to the combustion of gasoline, the combustion of diesel fuel produces up to 100 times more particulates, and most are fine (0.1-2.5 μm) or ultrafine ($<0.1 \mu\text{m}$) in size [U.S. Environmental Protection Agency (EPA) 2002]. Because of their small size, exposure to PM_{2.5} has been linked to a wide array of health effects such as aggravation of existing heart and lung disease and premature mortality (Pope et al., 2002). Health studies often rely on relative-city central monitor estimates that assign entire metropolitan areas the same level of exposure. Former studies have demonstrated that PM_{2.5} particles can reach the nasal and peripheral airways on inhalation, enhancing the allergic immune response and increasing production of immunoglobulin (Riedl and Diaz-Sanchez 2005). Exposure to these particulates during infancy while the immature immune system is developing may have contributed to the

increase in allergic diseases over the past two decades (Sherrill et al., 1999). Traffic –related variables are a significant factor in predicting air pollution levels, as different traffic parameters yield different levels of pollutant emission or transportation activities.

A modest quantum of models of PM_{2.5} has attempted to predict fine particle concentrations with the land use regression methods to date. Most studies have used informal methods to maximize the contrast in variables hypothesized to be potentially important predictors, by taking account of the distribution of locations to which the model will be applied. One of the only studies was undertaken part of a multicenter study relating traffic-related air pollution with incidence of asthma in three birth cohort studies in three European cities under the Traffic Related Air Pollution and Childhood Asthma study (TRAPCA). Researchers measured PM_{2.5} for representative temporal periods over 1-year in Netherlands, Munich, Germany and Stockholm County, Sweden. They found significant differences in model explanatory power from region to region with R² values ranging from 73% (Netherlands) down to 56% and 50% (Munich and Stockholm, respectively), therefore, those subjects living near major roads were likely to experience substantially higher air pollution, as a substantial fraction of the variability in annual average concentrations for all locations was explained by traffic-related parameters. Although, the limited variability of the monitoring sites in Stockholm County was suggested explanation for this difference (Brauer et al., 2003).

The concentration of PM_{2.5} mass has long been realized to be a large-scale phenomenon (Hoek et al., 2002b). This observation is strongly attested by Hochadel et al. (2006) that show a low level of its spatial variability as well as a poor predictability by traffic-related variables for PM_{2.5} mass (R²=0.9-0.17). This poor predictability of PM_{2.5} mass was an aftermath of a contradictory variable selection procedure, as it involved combinations of two very similar

variables at the same spatial scale with contrary signs of the coefficients, disputing causality. Acceptable models did not even reach a fit (R^2) of 0.2 thus indicating instability in both $PM_{2.5}$ mass models. Conversely, $PM_{2.5}$ absorbance, a marker of diesel exhaust attributable to traffic sources and an appropriate proxy for ambient concentrations of element carbon or soot, turned out to be substantially more predictable by traffic-based variables explaining the surface variability by 65 to 82%. Whilst Hochadel et al. (2006) model predicted positive $PM_{2.5}$ absorbance values and doubtful $PM_{2.5}$ mass concentrations; a strong correlation was found between $PM_{2.5}$ absorbance and NO_2 concentration ($R^2 = 0.93$), whereas $PM_{2.5}$ mass concentration was less strongly correlated with $PM_{2.5}$ absorbance and NO_2 ($R^2 = 0.52$ and $R^2 = 0.41$ respectively), thus delineating real associations between the different pollutants rather than artifacts of the modeling procedure.

Likewise, Ryan et al. (2007) performed a LUR model for the Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS) to compare exposure from trucks and buses with those developed with proximity models (Ryan et al., 2007) and thereupon assess the relationship to infant wheezing. Using CAAPS ambient air sampling network of 24 monitoring sites, a derived marker of diesel combustion-element carbon attributable to traffic sources (ECAT) was seized from $PM_{2.5}$ collected on 37-mm Teflon membrane filters employing an optical reflectance method. Based on elevation, average daily number of trucks within 400 m, and length of bus routes within 400 m the model was able to predict 75% of the variability of the sample levels of ECAT. In contrast, LUR model improved exposure assessment as that of the proximity model by identifying 95% of the infants that were previously designated as unexposed to have ECAT at or below $0.44\mu g/m^3$ thus providing a higher spatial resolution. Nevertheless, the association between ECAT and wheezing without a cold remains significant after adjustment for maternal

smoking, childcare attendance, sex, race, breast-feeding, pet ownership, and viable mold. At $0.5\mu\text{g}/\text{m}^3$ there is almost a 2-fold increase in wheezing with an adjusted odd ratio (AOR) (Ryan et al., 2007) of 1.86, with 95% significance and CI: 1.02-3.39, and at $0.9\mu\text{g}/\text{m}^3$ it increases to greater than 4-fold with an AOR of 4.26, 95% CI: 1.06-17.26.

Diversely, not all models have shown an influence by traffic sources. A recent study by Henderson et al. (2007) sights no traffic-based variable to be significant influential for $\text{PM}_{2.5}$ model, and the effect of automobile density consequently make-believe a weak correlation. Contrarily, the model for $\text{PM}_{2.5}$ absorbance relies entirely on truck density, which suggests that LUR is useful for discriminating between pollutants associated with different sources. Employing 25 monitoring stations in a two-months 7-day campaign, both $\text{PM}_{2.5}$ absorbance and $\text{PM}_{2.5}$ mass models were able to spurred R^2 's of 0.41 and 0.52. Parameters like elevation, commercial, residential and industrial land use were able to predict $\text{PM}_{2.5}$ mass concentration; as for $\text{PM}_{2.5}$ absorbance, truck density was the sole variable to influence the surface variability over Vancouver. These findings confirm that valuable LUR models can be developed in the absence of traffic count data, which are unreliable or nonexistent in many areas.

The representation of pollution levels of such models is heavily dependent on the specific regulatory site locations, many of which are located to represent average urban concentrations. Using such data may fail to capture the actual range in values that exist throughout the area, as the placement of these monitors may exclude “hotspots” or high concentrations. Therefore, some developed LUR models relied solely in exposure assessment rather than evaluating health-related impacts.

As a result, a former study over the five county metropolitan statistical area of Los Angeles found that 69% of the variance in $\text{PM}_{2.5}$ was unfold by traffic density within 300 m,

industrial land area within 5000 m, and government land area within 5000 m of the monitoring sites. Though, slight over-prediction at Long Beach and freeway 710 exceeded both federal ($15\mu\text{m}/\text{m}^3$ annual average) and state standards ($12\mu\text{m}/\text{m}^3$ annual average) predicting somewhat concentrations upward $126\mu\text{g}/\text{m}^3$ (Moore et al., 2007). Westerdahl et al. (2005) found that measured freeway concentration of $\text{PM}_{2.5}$ had a range of 60 to $820\mu\text{g}/\text{m}^3$ on the Interstate-10 (I-10) East Freeway, and that concentrations along major roadways with high traffic density were up to 20 times higher than residential concentrations.

Similarly, in New York City three regression equations were developed to predict fine particulate matter using three-year averages programme in 28 counties as well as a more urbanized 9-county. Region primary $\text{PM}_{2.5}$ emissions available at the county did not enter the final prediction model for $\text{PM}_{2.5}$ the study consisted of 28 counties; hence data were available at a rather crude spatial scale. Ross et al. (2007) found that traffic within a buffer of 300 m or 500 m explains the greatest proportion of variance of 34 to 44% in all three models. Measures of urbanization, specifically population density, explain a significant amount of the residual variation of 7 to 18% after including a traffic variable. Finally, a measure of industrial land use further improves the 28-county and 9-county models based on the 3-yr annual averages, explaining an additional 6%. In total, the final land use regression models explain between 61% and 64% of the variation in $\text{PM}_{2.5}$. On the other hand, it was observed that the coordinates of older routine monitoring sites were not geographically precise. Uncertainties also occur in the traffic data, due to the limited representativeness of traffic counts and difficulties in extrapolating these across the network. As linear regression assumes independence of the residuals, most studies do check whether the residuals of the model exhibit spatial autocorrelation using a variety of techniques including kriging and Moran's I. To calculate Moran's I, however, nearby needs to

be defined—for example, as nearest neighbors. Most studies observed that, while significant spatial autocorrelation is present in the measured concentrations, the residuals of land use regression models are independent, indicating that ordinary least squares regression can be used to obtain valid prediction models. Lastly, land use regression slightly outperformed kriging in predicting $PM_{2.5}$ concentrations at validation sites. The Root Mean Square Error (RMSE) at prediction sites was 1.15 and 1.00 mg/m^3 for LUR in the ‘28 county’ and ‘more urbanized 9 county’ models versus 1.30 and 1.47 mg/m^3 for kriging in these counties. The difference between the two studies is likely due to the much smoother spatial patterns of $PM_{2.5}$, which has a strong regional background component.

2.2.3 Land use Regression Model for PM_{10}

PM_{10} concentrations vary depending on the characteristics of the location of the study. Some cohort areas may report high concentrations due to long-range transport, mainly secondary aerosols and natural sources. Studies performed in more arid areas have reported that the main source of PM_{10} is attributable to soil and street dust, heavy oil burning, traffic exhaust, wood burning, and unidentified sources. Despite the differences in sources of particles smaller than or equal to $10\mu m$, still they remain to somewhat similar, independent of the characteristics of each study area. In particular, chronic exposure to PM_{10} has been associated with increased mortality, morbidity, and risk of harmful effects to the pulmonary and cardiovascular systems (Pope III & Dockery, 2006). However, none of the studies reviewed so far have developed LUR to assess exposure of PM_{10} , instead, all of them relied in solely assessing the performance of regression mapping against other modeling techniques.

The first regression-based model to estimate PM₁₀ mass concentration was undertaken as part of the EU-funded Air Pollution Modeling for Support to Policy on Health and Environmental Risks in Europe (APMOSPHERE) project. The aim of the study was to assess methods to develop accurate, high resolution and updatable maps of emissions and air pollution across the 15 countries in the European Union, as a basis for environmental and health policy and environmental and epidemiological research. Although, Briggs et al. (2005) used ordinary kriging, universal kriging and LUR to assess exposure of PM₁₀, still none of the methods were able to provide satisfactory predictions when cross validated.

In a more recent study, Briggs et al. (in. press) extended their previous APMOSPHERE research conducting a more rigorous assessment of four modeling techniques. In contrast, these modeling approaches (based on nearest monitoring site, kriging, ADMS-Urban dispersion model and land use regression) were compared in their ability to predict mean annual PM₁₀ mass concentrations across 52 monitoring sites in London, UK. Few of the approaches, however, showed good correlations with monitored PM₁₀ concentrations but could not predict values better than simple classification based on site-type. Kriging and the nearest monitor site had the worst performance with a model R^2 of -0.01 and 0.05. Only land use regression, otherwise, achieved acceptable performance levels of correlation ($R^2=0.47$) that might be considered for exposure assessment, though not significantly better than the site type.

Conversely, a European Union study showed an outperformance of universal kriging over regression mapping and ordinary kriging. The two-kriging methods substantially performed better than LUR model giving R^2 of 0.45 for universal kriging and 0.41 and 0.38 for regression mapping and ordinary kriging, respectively (Beleen et al., 2009). However, the low performance of LUR was attributable to mischaracterization of conditions near as each site. The north-south

gradient reflected ingress of dust from sub-Saharan Africa and meteorological variables on secondary PM formation in southern Europe, thus compromising the reliability of exposure assessment for PM₁₀ mass concentration. As indicators comprising metrics on proxies for exposure are usually defined in terms of the source activity or the relationship between the target individual, the population, or the source at a local scale rather than at a universal scale within countries as adopted by Beleen et al. (2009).

2.3 AMBIENT PARTICULATE MATTER (PM₁₀)

2.3.1 Overview

Airborne particulate matter is defined as any substance, except pure water, that exists as a liquid or solid in the atmosphere under normal conditions and is of microscopic or submicroscopic size but larger than molecular dimensions (Seinfeld, 1997, p.50). Airborne particulate matter, which may sometimes be referred as PM₁₀ or PM in this research, can be characterized by its composition, size, origin, and source emission type. PM₁₀ is typically composed of a complex mixture of chemicals such as metals, ions and a large variety of organic compounds, which can undergo exchanges between the gas and condensed phases, producing remarkable compositional changes of PM with time. PM₁₀ embraces three different sizes of particulates often classified depending on their aerodynamic diameter. Fine particles are those which have an aerodynamic diameter equal or smaller than 2.5µm and are characterized by constituents such as sulfate, nitrate, ammonium, metals elemental carbon, and hundreds of different organic carbon compounds. Ultrafine particles are particles defined by their aerodynamic diameter equal or less than 0.1µm. They originate from combustion sources and from homogeneous nucleation of low vapor pressure compounds. In sum, ultrafine and fine

particles are collectively defined as $PM_{2.5}$. Coarse particles ($PM_{10-2.5}$), on the other hand, have an aerodynamic diameter between $2.5\mu m$ and $10\mu m$ and formed primarily by material coming from fugitive dust, construction sources, and reaction of gases in the atmosphere. These processes are composed of street dust, fly ash, nitrates, chlorides, oxides of crustal elements, and organic material. Whilst ultrafine particles, fine particles and coarse particles are defined by the aerodynamic diameter, they collectively are defined as PM_{10} . Particles directly emitted to the atmosphere are termed “primary” particles. Such particles can be fine (e.g., combustion products) or coarse (e.g., dust, plant debris, pollen, sea spray). “Secondary” PM_{10} results from the condensation and deposition of gaseous precursors to the particulate phase. Lastly, EPA is concerned about particles that are 10 micrometers in diameter or smaller because those are the particles that generally pass through the throat and nose and enter the lungs. Once inhaled, these particles can affect the heart and lungs and cause serious health effects.

The El Paso area airborne particulate matter has been identified as the major air pollutant in region. El Paso, Texas is within a tri-state bi-national metropolitan area often referred as Paso del Norte (PdN). PM mass concentrations in the PdN airshed vary significantly on both temporal and spatial scales. The highest PM loadings, excluding dust storms, have been found in the urban centers and industrial areas of the region where local sources strongly influence air quality. A vast part of the PM in the PdN region derives from geological sources, vehicle exhaust, residential cooking and heating, and other identified sources (Li et al., 2001). The arid weather, occasional high winds, frequent stagnations, shallow nighttime and morning mixing depths, and complex topography preclude simple explanations for excessive PM_{10} levels. According to Dattner (1990) who analyzed winter time air quality found that the maximum concentrations of PM_{10} in the PdN occurred during nighttime, from 5 p.m. to 5 a.m., with peaks in the early to

mid-evening and under the occurrence of low mixing heights. Also the PM_{10} concentration gradient across the eastern border of the airshed resulted in a net transport of PM into El Paso from Ciudad Juarez. As a result Dattner (1990) stated that soil-related materials are the dominant material in airborne particulate matter. This was further collaborated in a recent study performed by Li et al. (2001) who indicated that 25% of the PM_{10} concentration is composed of fine particles. Comprising, coarse particles to be approximately 75% of the total PM_{10} concentrations; thus crustal material, a precursor of coarse particles, tends to be higher among the sources of PM_{10} for the PdN region.

2.3.2 Health Effects

Because only small particles can be inhaled into the lungs, U.S. national health standards for the quality of ambient air are based on the mass concentration of “inhalable Particles”, defined to include particles with an aerodynamic diameter of equal than or less than 10 micron. In 1987 the U.S. Environmental Protection Agency (EPA) promulgated a new size-specific air quality standard for environmental particulate matter. This new standard applies only to PM_{10} particles and replaces the original standard for total suspended particulate matter (TSP). PM_{10} standard takes into account particle fraction inhaled into the deeper respiratory tract regions, causing pathologies associated with aerosol pollution. With the exponential growth of vehicular traffic over the past 20 years, larger particles from uncontrolled sources in the urban environment have been replaced by relative higher concentrations of much smaller particles. Several multi-city epidemiological studies have observed associations between adverse health effects and surrogates of traffic exposure (e.g., residence relative to highways) or to exposure estimates based on central monitoring (de Medeiros et al., 2009). In consequence, additional research has

been conducted since the mid1970s documenting the substantial health effects associated with dramatic pollution episodes. Recent epidemiologic studies have provided a more precise quantification of subtle health effects associated with particulate pollution common to contemporary cities.

Numerous studies have assessed acute morbidity effects of particulate pollution by examining short-term temporal associations between lung function measures and respiratory symptoms as well as pollution. Results of most studies suggest that a $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} resulted in less than 1% decline in lung functioning (Pope et al., 1991; Pope and Dockery, 1992; Pope and Kanner, 1993; and Raizenne et al., 1990). Nevertheless, 24-hr PM_{10} concentrations would occasionally exceed $150 \mu\text{g}/\text{m}^3$, in which, during such episodes lung function declines by as much as 7% (Raizenne et al., 1990). In essence, the particulate pollution effect on lung function was generally physiologically small, though statistically significant. In addition to declines in lung function, many of these studies observed increases in respiratory symptoms. A $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} was associated with 1 to 10% increase symptoms such as cough, combine lower respiratory symptoms, and asthma attacks even when observed at PM_{10} levels near or below $150 \mu\text{g}/\text{m}^3$ (Koenig et al., 1993; Pope et al., 1991; Pope and Dockery, 1992; and Schwartz et al., 1994). Moreover, numerous studies have also evaluated morbidity effects of particulate pollution by examining short-term temporal associations between particulate air pollution and hospital admissions, health care visits, or other measures of restricted activity due to illness. Statistically significant associations between hospital and health care visits for respiratory illness and particulate pollution were observed in most, but not all, of these studies. Much of the studies suggested that a $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} on the day of the visit or 1 or 2 days before the visit was typically associated with a 1-4% increase in hospital visits (Vedal et al.,

1991; and Schwartz, 1993).

Measures of lung function and incidence rates of respiratory symptoms have also been compared across communities or neighborhoods with different levels of particulate air pollution. Given the design of these studies and because pollution measures are averages over relative long periods of time (1 year or more); these studies are often interpreted as evaluating chronic or cumulative effects of exposure rather than acute effects. Schwartz et al. (1993) infers that a $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} was typically associated with a decline of less than a 2% in lung function. Likewise, respiratory disease, including emphysema and chronic bronchitis, and the incidence of respiratory symptoms were also associated with particulate pollution. The results suggest that a $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} was typically associated with a 10 to 25% increase in bronchitis or chronic cough (Schwartz et al., 1993; and Chesnut et al., 1991).

Some of the most striking studies of health effects of particulate air pollution are those that observed changes in daily death counts associated with short-term changes in particulate pollution, precise comparisons between studies is difficult, but results of most of the studies suggest that a $10 \mu\text{g}/\text{m}^3$ increase in PM_{10} was associated with an increase in daily mortality of 0.5 to 1.5% (Schwartz, 1991, 1993; and Schwartz and Dockery, 1992). However, the major limitation in assessing chronic mortality may be the lack of true long-term exposure data or what often is referred by the literature as “lag” time. Several studies have demonstrated that air pollution levels, not just for the current day but for several preceding days, affects counts on daily mortality on a given day (Schwartz and Dockery, 1992; and Kelsall et al., 1997). In fact, multiday averages of pollution are found to be better predictors of daily death counts than is a single day’s exposure. A number of studies have assessed this issue using different methodologies (Schwartz, 2000c, 2001; and Zeger et al., 1999). All of them reported increased,

rather than decreased, effects when longer lags were examined. One possible reason for this is that different causes of death are affected by air pollution with different latency periods.

Zanobetti et al. (2003), more recently, provided an insight into the shape of the longer-term responses to particulate air pollution. In their study, they suggested that the adverse response to pollution persist to a month or longer after exposure not only for total mortality but also for respiratory mortality. Using an unrestricted distributed lag, the estimated effects increased to 4.2% (95% significant, and CI: 1.08-7.42) for respiratory deaths and to 1.97% (95% significant, and CI: 1.38-2.55) for cardiovascular deaths. Thereupon, the effect size estimate for PM_{10} doubles when considering long-term effects for all deaths as well as cardiovascular deaths and increases 5-fold for respiratory deaths.

Moreover, and according to the literature, respiratory disease deaths were most strongly associated with particulate pollution levels, but statistical associations were also observed for cardiovascular disease deaths (Pope et al., 1995). The available studies indicate that exposure to airborne particles is associated with hospital admissions for cardiovascular disease. Morris (2001) findings suggest that an increase of $10 \mu g/m^3$ in PM_{10} was associated with increases in admission rates of 0.8% (95% significant and CI: 0.5-1.2%) for congestive heart failure, 0.7% (95% significant and CI: 0.4-1.0%) for ischemic heart disease, and 0.2% (95% significant and CI: -0.2-0.6%) for cerebrovascular accidents. In addition, Dockery (2001) exerted an association with indicators of autonomic function of the heart including increased heart rate, decreased heart rate variability and increased cardiac arrhythmias. Furthermore, the affiliation between pulse rate and PM_{10} had a statistical significance with a positive association between increase pulse rate and PM_{10} . The magnitude of which in terms of increased beats per minute was physiologically small, 0.8 beats per minute increase for $100 \mu g/m^3$ increase in PM_{10} on the previous day.

A large and growing body of literature documents the adverse health effects associated with particulate air pollution at levels common to contemporary cities in the world. Health effects are observed at levels common to many U.S. and Canadian cities, including levels well below current U.S. National Ambient Air Quality Standards. There is no clear evidence of a safe threshold level. Many studies observe that the health effects increase monotonically with pollution levels, often with a near-line dose-response relationship.

2.3.3 Sources

Although there is a trend toward more detailed atmospheric measurements, only a moderate number of studies have been performed in apportionment of ambient particulate matter. Previous studies have identified similar contributors to PM_{10} , though, not all of them agree in the same distribution. In a study carried out in Kuopio, Finland, Hosiokangas et al. (1999) analyzed PM_{10} samples for 38 days. Using receptor modeling and multiple linear regression analysis of sources and their contributions for the measured concentrations of PM_{10} particles, found that the four major sources of PM_{10} were: 46-48% of soil and street dust, 12-18% of heavy fuel oil burning, 10-14% of traffic exhaust, 11% of wood burning, and 15-25% of undefined sources. However, during spring dust episode days, the main source of PM_{10} was soil. In another study Lenschow et al. (2001) carried out a PM_{10} source apportionment study in Berlin, Germany by comparing the chemical composition of particles collected at urban background, regional background and curbside locations. They concluded that 55% of the emissions of PM_{10} were due to vehicle exhaust, combustion (e.g., from heating to power plants) and tire abrasion whereas the rest 45% was attributable to re-suspended soil and unclassified materials. In the same year, another study performed in Athens (Greece) and Birmingham (United Kingdom)

performed a principal component and regression analysis to quantify the contribution of non-combustion sources to the observed PM₁₀ background levels (Vardoulakis and Kassomenos, 2008). This contribution, though, ranged between 45% and 70% in Birmingham and 41-74% in Athens. Even more, it was found that long-range transport of particles from continental Europe had a marked effect on PM₁₀ background levels in Birmingham, while the local weather had a stronger influence on PM₁₀ levels in Athens (Vardoulakis and Kassomenos, 2008). Likewise, using chemical mass balance receptor modeling Mazzera et al. (2001) were able to determined potential important aerosol source type in McMurdo. Soil dust, combustion emissions, and nitrates contributed 57%, 14% and 1%, respectively, to the average estimated PM₁₀ at Hut Point. As a moderate number of studies have attempted to identify source apportionment of PM₁₀ in European cities, the number of research performed in the U.S. remains relative small compared to EU and even more for the Paso del Norte region.

As one of the exiguous studies, Wayne and Hugh (1995) summarized most of the air quality research performed in the PdN region since the 1980's. They concluded that although, industrial sources were not a significant source of PM₁₀, biomass combustions constituted nearly 80% of the total PM₁₀ mass in the winter season; as stagnation events in winter and the complex topography significantly limit the dilution of contaminants. Traffic related sources, on the other hand, may be influence by the vehicles from Juarez as the average emissions from Ciudad Juarez vehicles are three times higher than that from El Paso vehicles, though Vehicle miles traveled for El Paso represent three times more than the VMT's for Ciudad Juarez.

More recently Jeon et al. (2001) studied PM compositions and origins in the PdN region by analyzing 45 time-resolved (2-hr) filter samples. Using principal component analysis and related techniques, four main factors (vehicular emissions, re-suspended urban dust, biomass

combustion, native vegetation detritus and re-suspended agricultural dust, and waste burning) were identified responsible for most of the variance of PM_{10} . Both reports suggested that the PdN regional-specific meteorological, topographic, and emission features contributed to the distinctive pattern of PM levels in the airshed.

2.4 GEOGRAPHICAL INFORMATION SYSTEMS

2.4.1 Overview

A GIS consists of a set of geographical features linked to a database of attributes that are specific to the features, which are integrated through a multi-layer system. The geographical features can have three basic geometric forms: polygons, lines, or points. An attribute is any information that is specific to a geographical feature. For instance, a road can be represented as a geographical feature by using a line and its attributes can be the name of the road, length, average speed, and any other kinds of information specific to this feature. The integration of geographical features to attributes is called coverage. In the past example, the coverage could be a road system for a given region. Geographical analysis of coverages is done with GIS software. Detailed explanation of all the analysis capabilities of GIS systems is beyond the scope of this study, but two relevant examples are discussed below. These two examples are Geocoding and Spatial Intersection.

Geocoding is the process of identifying the coordinates of a geographical location when an address is provided. To geocode, the use of a road coverage that contains addresses as part of its attributes is needed. The roads are divided into segments with address ranges specified for each segments. The program identifies the road by its name, and further locates the segment where the address number falls within. One can make use of the geocoding capabilities to

generate additional coverages. For example, by knowing the addresses of all fire stations within a city, their locations can be geographically located on a separate layer. A fire station coverage can then be generated by saving this layer separately.

Spatial Intersection is the allocation of attributes from one coverage to another where its geographical features (i.e., polygons) overlay one another. Two coverages that can be represented by polygons are census tracts and zip codes. For a given county, these two coverages overlay one another. Let's assume that the census tract coverage contains population information as part of its attributes while zip code coverage does not. Let's say it is necessary to analyze population distribution within a county by zip code. The population attribute of the census tract coverage can be allocated to the zip code polygon by the area proportionality of the intersected section created by the overlay of the census tract polygon. This procedure can also be applied to the feature combination of lines and polygons or point and polygons with insignificance difference.

2.4.2 GIS Applications

Dai and Rocke (2000) applied the data management and spatial analysis capabilities of GIS to spatial allocation of area source emissions. Dai and Rocke's approach consisted of a GIS coverage of actual locations of emitting sources generated through geocoding. The addresses were retrieved from phonebooks. Activity data for each emission source was collected and processed. The study area was then disaggregated into sub county units (e.g. modeling cells) and a grid coverage was created. The data was converted into the modeling unit by spatial intersection of the two coverages. Finally, spatial variations of the activities were mapped using the gridded coverage and emission allocation factors were computed. This procedure allocates

accurately emission estimates to each grid cell. A disadvantage of this procedure, however, is that it requires many resources to generate the coverages. In addition, as mentioned by Lindley et al. (2000), the data is not always available in such format that it can be easily processed.

Modeling mobile source emissions in a GIS framework is another of the many GIS applications (Bachman et al., 2000). Bachman et al. (2000) recognized that mobile emissions are intrinsically spatial and applied a GIS approach to mobile source modeling. The inputs for this type of model were divided into three general categories: fleet activity, fleet characteristics, and operating conditions. In typical mobile source models, vehicles miles traveled (VMT) are calculated for each spatial modeling unit (grid cell). The emission totals are then distributed by using the aggregation of the three inputs mentioned above. By doing so, the individual spatial variability of each of the inputs is not taken into account (i.e., assumptions of uniform distribution of vehicle fleet across a region are made). Bachman et al. (2000) suggested that a GIS framework is ideal for implementation of a modal modeling approach, where emissions rates are a function of specific modes of vehicle operation (engine starts, running exhaust, enrichment, etc.). The method consisted of estimating mobile source emissions for the different operating modes and storing these results on individual layers. An advantage of this method is that the aggregation of layers into grid cells, compatible with photochemical air quality models, can be implemented easily. This type of approach permits individual operating model analysis and yields more accurate estimates.

Bruckman et al. (1992) also identified many benefits of using GIS in the preparation of input data including fleet activity, fleet characteristics, and operating conditions for mobile source emissions. Souleyrette et al. (1992) acknowledged the potential of applying GIS to manage complex spatial data, typically involved in air quality studies. Another powerful

application of GIS was the development of an activity-based model for travel demand forecasting that integrated household activities, land use patterns, traffic flow, and regional demographics. Although not directly developed as an emission model, the model outputs could be used in the evaluation of impacts of mobile emissions on regional air quality (Stopher et al., 1996). Hallmark and O'Neill (1996) developed a model that combines a microscale air quality model (CAL3QHC) with GIS.

GIS has been used as geographic analyzing tools for decision makers. Jensen et al. (2001) generated a GIS-mapping tool which estimated ambient air pollution levels at high spatial and temporal resolutions by the inclusion of an operational street pollution model (OSPM). The system also maps traffic emissions, air quality levels and human exposure at residential addresses, and work place addresses and streets. This was possible with the use of cadastral digital maps, national administrative database of buildings, population information, as well as maps of streets and other geographical features.

Thus, GIS is a tool that provides a wide range of geo-information analyzing capabilities but its contribution to air quality management and modeling remains to be fully explored.

3 STUDY AREA

El Paso city resides within a three-state (New Mexico, Texas, and Chihuahua) bi-national metropolitan areas (The United States and Mexico). Located at 31.79N, -106.42W it lays at the northern part of the Chihuahuan Desert where it is known as the Paso del Norte (PdN) region. The El Paso area wraps around the southern tip of the Franklin Mountains, which run north to south and abruptly end just north of downtown sector of the city, it is separated from Ciudad Juárez, Chihuahua by the Rio Grande River (Figure 3-1). This area, mostly desert with agriculture along the Rio Grande River, includes two valleys; the lower valley which is located in the eastside of the city and the upper valley located in the western part of the city.



Figure 3-1. Paso del Norte area delimiting Ciudad Juárez, Mexico; El Paso, Texas; and Sunland Park, New Mexico. source (Map data ©2009 Google, INEGI).

The elevation for the El Paso area ranges from 2180 m at the Franklin Mountains to 1127 in the valley. These geographic features have a strong influence in the wind pattern. The abrupt deviation of the wind through these topographic features defines the pattern from southeast to northwest and vice versa (see APPENDIX A). With a climate classified as arid (Koppen climate classification), hot summers tend to average 97°F to a low of 68°F, whilst mild and dry winters range from an average high of 55°F to an average low of 28°F (as shown in Figure 3-2 and APPENDIX B).

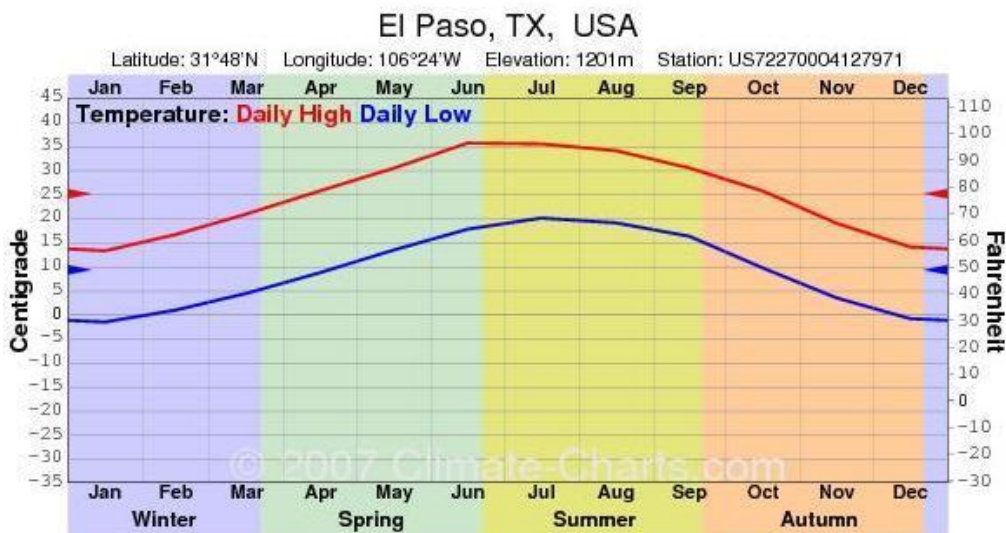


Figure 3-2. Temperature average variations during the three-year monitoring campaign. Source (Climate Charts© 2007).

This bi-national, tri-state region has an urbanized area of approximately 903.8 km² distributed in 72% for El Paso, Texas (ELP) 25% and 3% for Ciudad Juarez, Mexico (CDJ) and Sunland Park, New Mexico (SLP) respectively (Wikipedia, 2009). According to the 2009 U.S Census, the city had a population between 0.6 million and 0.7 million for the entire metropolitan area that covers all of the El Paso County. Ciudad Juarez had a population of nearly 1.5 million inhabitants. Both El Paso and Juarez region add up to a population of nearly two million people (Wikipedia, 2009; U.S. Census Bureau, 2009; and INEGI, 2009). The PdN, is supported by

different economical scenarios. The El Paso, on one hand, has its urban sprawling based around Fort Bliss due to the Federal investments towards this United States Army installation. Ciudad Juarez, on the other hand, has its foundation based on industrial Maquiladoras mainly Automobile industries. Annually, more than 16 million private passenger vehicles and nearly one million freight carriers enter the U.S. via border crossing bridges over the Rio Grande in El Paso. Six international ports of entry are located within the PdN region; three of them experience constant heavy traffic conditions. The international Bridge of the Americas (BOTA) has the heaviest combined traffic of heavy-duty diesel and gasoline fueled vehicles among all U.S. and Mexico ports of entry. Vehicles remain idling in queues an average of 60 to 90 min as they wait for inspection. El Paso area was chosen as the study site. The combination of high population density, industrial influences and meteorological conditions often causes PM_{10} concentrations in the region to exceed the national ambient air quality standards of both countries. The PdN region is unique for the significant amount of uncontrolled geologic emission sources, dust from unpaved roads, less-regulated mobile emissions from aged vehicles, biomass emissions from residential cooking and waste burning, uncharacteristic emissions from brick kilns and house heating, and other unidentified anthropogenic emissions. The arid weather, occasional high winds, frequent stagnations, shallow nighttime and morning mixing layer, and complex topography preclude simple explanations for excessive PM levels (Li et al., 2001). Furthermore, the intra-urban gradient of PM_{10} emissions has not been studied across much of the metropolitan El Paso because the air quality monitoring stations operated by the state and local environmental agencies are located mainly in the central El Paso with fewer stations located in more distant residential areas.

4 MONITORING PROGRAM

4.1 SAMPLING LOCATION

A purpose-designed air monitoring network was deployed from August 2006 to April 2009 in accordance with the EPA guidelines (U.S. EPA 1997). Weekly PM₁₀ samples were collected at 13 sites distributed throughout the El Paso 50 geographic strata. Among the 13 monitoring sites, eight sites were initially deployed in 2006; five sites had to be relocated throughout the campaign. Four of which had to be relocated in August 2007 to increase the spatial coverage and an additional site was relocated in August 2008 due to restricted site accessibility. The total area covered by the network area was 694 km², resulting in a monitoring density of one site per every 50 km². It was paramount to allocate monitors at specific locations where the impact of typical activities in the city, characteristics, and parameters could be assessed more thoroughly. Therefore, nine sites were allocated within residential areas, four sites were positioned near major roadways, and five were located close to the international border to capture the regional impact of PM₁₀ emitted in Ciudad Juarez. In addition, one monitor was located near the BOTA port of entry. Several sites met more than one site selection criteria resulting in double counting of several sites above classifications. All sites were coded for easement purposes. Table 3-1 provides a description of the monitoring sites in terms of name of the site, coded ID, address, duration, and location. Most of the monitors were setup inside local and state agencies own properties. Permission was obtained from El Paso Water Utilities (EPWU) and TCEQ to use their facilities. Both agencies were chosen on the basis of their amenity, electrical power accessibility, existing monitors and most importantly, their fenced installations.

Table 4-1. Description of the 13 monitoring sites throughout the El Paso region.

Name	ID	Address	Longitude (UTM)	Latitude (UTM)	Duration
Lindbergh	A	5155 Charl Ann	31.848	-106.583	2006-2007
Fountain	A2	279 Fountain	31.838	-106.531	2007-Current
Skyline	D	5050 A. Yvette	31.894	-106.426	2006-2007
McCombs	D2	10800 McCombs	31.928	-106.407	2007-Current
UTEP	F	250 Rim Road	31.768	-106.501	2006-Current
Chamizal	H	800 S. San Marcial	31.766	-106.455	2006-2007
Parral	H2	7701 Parral	31.751	-106.357	2007-Current
Cosmos	I	2951 Catnip	31.783	-106.361	2006-Current
Socorro	J	201 South Nevarez	31.662	-106.303	2006-Current
Plant	K	8470 Plant Road	31.704	-106.356	2006-2007
Mimosa	K2	7501 Mimosa	31.736	-106.378	2007-Current
Montana Vista	L	3668 Desert Meadows	31.805	-106.166	2006-2008
Pebble Hills	L2	2775 Joe battle	31.783	-106.270	2008-Current

Monitors A, A2, D2, H2, I, K, K2, L, and L2 were installed at EPWU facilities. Samplers D, F, H, and J were located at TCEQ's stations by virtue of their parallel sampling (Figure 3-3). All samplers were positioned at least 8 feet from any building, with the inlet head standing 5 feet above the ground.

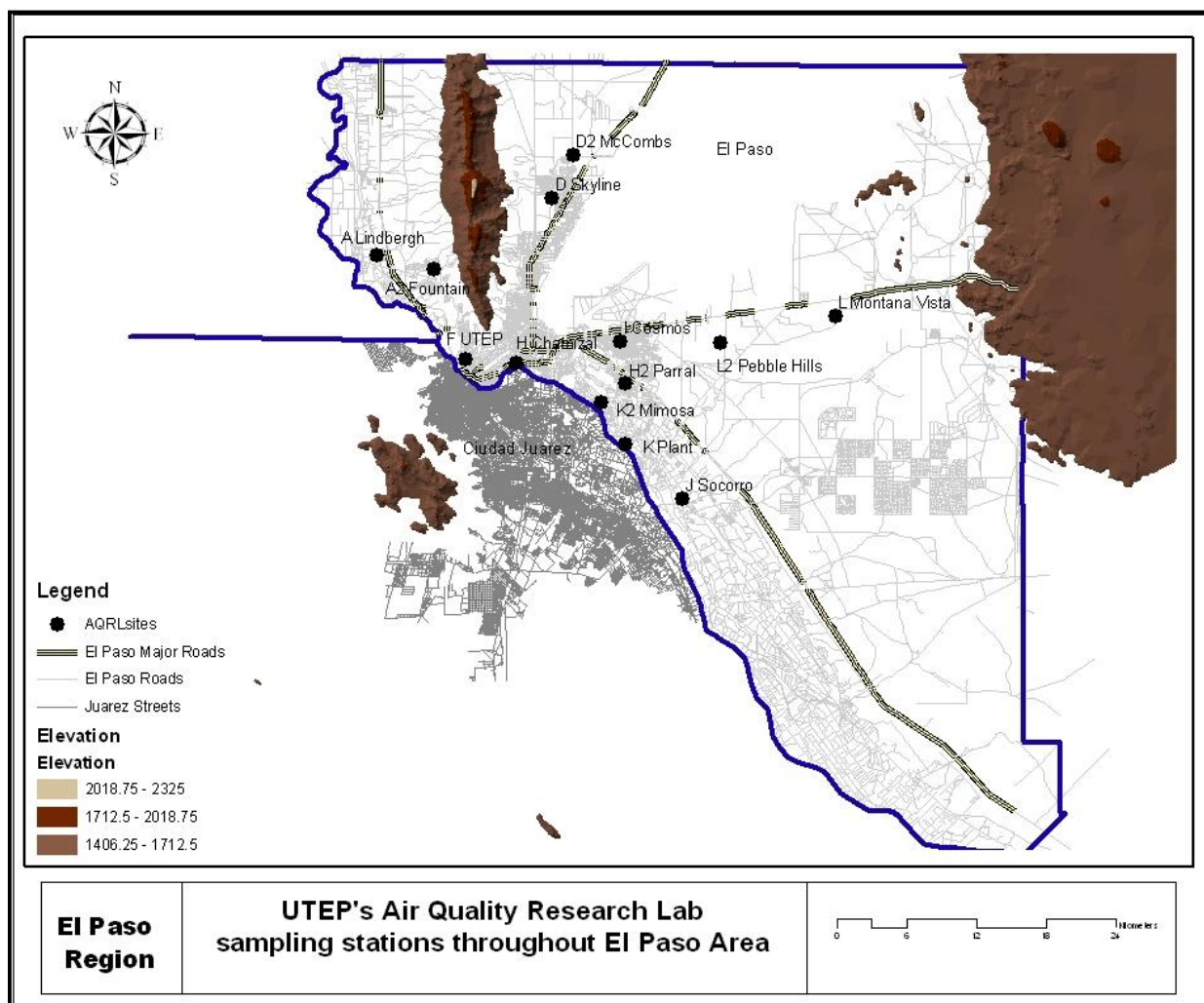


Figure 4-1. Distribution of the thirteen Dichotomous samplers throughout the El Paso region.

PM₁₀ collected at three governmental continuous ambient monitoring sites (CAMS) was employed for cross validation of dichotomous samples and to analyze temporal patterns. Data collected at CAMS sites was not considered for modeling purposes because out of the four sites that measured PM₁₀, two sites were collected with dichotomous samplers, another stopped collection before the beginning of this study, and the remaining site appeared spatially auto correlated with other observations.

4.2 SAMPLING EQUIPMENT

PM₁₀ measurements were sampled using Dichotomous samplers (Anderson Instruments, Inc. ® and Thermo Fischer scientific®). One Dichotomous air sampler was placed at each site and 1 additional air sampler was used for duplicate samples among eight sites. These samplers consisted of two modules: the sampling module and the flow control module as shown in Figure 3-4. The flow control module was set to operate at a low-flow rate of 16.67 L/min, in order to simulate the breathing rate of the human being. A mini-bubble calibrator (Dry Cal® DC-Lite, Butler, NJ), a primary standard calibration device traceable to Bios International Corporation, was used to calibrate the rotameters on the dichotomous samplers (DICHOT). Particle-laden air passing close enough to the inlet head of the DICHOT is forced to take a sharp turn upon entry. Large particles (greater than 10 µm) cannot make this turn into the cap and proceed to move in their original direction due to inertia. After passing through the inlet, the air containing inhalable particles are accelerated at sufficient speed to impact particles larger than 2.5 microns into the virtual impactor (shown in Figure 3-5). Coarse particles are then decelerated at the virtual impactor receiver tube by the miner flow (10% of the total flow) to the coarse particle filter. Fine particles, on the other hand, follow the high velocity flow stream (15 L/min) entering the virtual impactor and are captured on the fine particle filter. Both coarse and fine flows remained relative constant except for small decreases that occur during sampling.

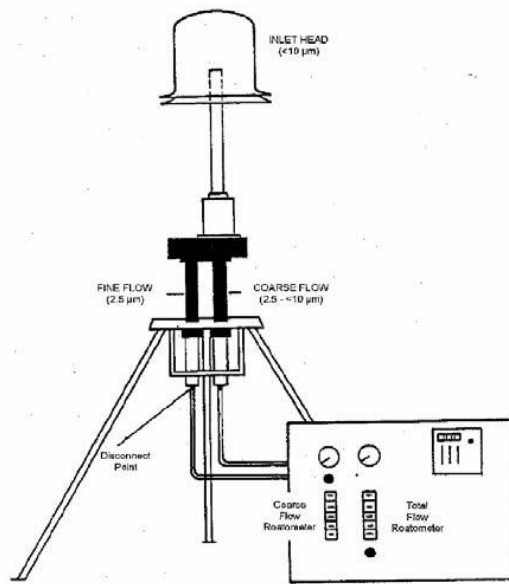


Figure 4-2. Dichotomous sampler allocated at each site. Source (Thermo Scientific Dichotomous manual 1996).

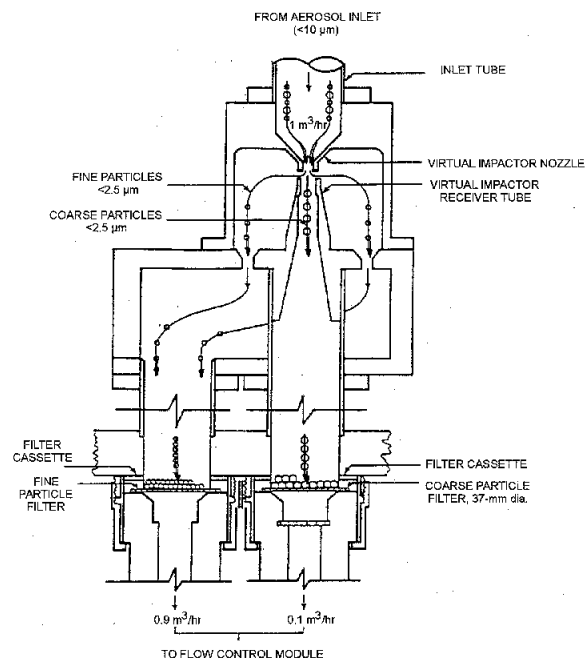


Figure 4-3. Dichotomous virtual Impactor. Source (Thermo Scientific Dichotomous manual 1996).

4.3 FIELD OPERATIONS

Each site was sampled for 7 consecutive days. Sampling could only be done at eight sites simultaneously due to the limited availability of equipment. The logistic of the field operations was carried out by six-researchers. Three-teams of two researchers each were responsible of deployment and retrieval of filters and equipment. The purpose of two persons assigned to each team transpired in order to reduce any unsustainable field uncertainties that solely one researcher, instead could be responsible for. Two teams were assigned with a route comprising of three sites and one with only two. The latter team was responsible for the longest route (Figure 4-4, Figure 4-5, and Figure 4-6).

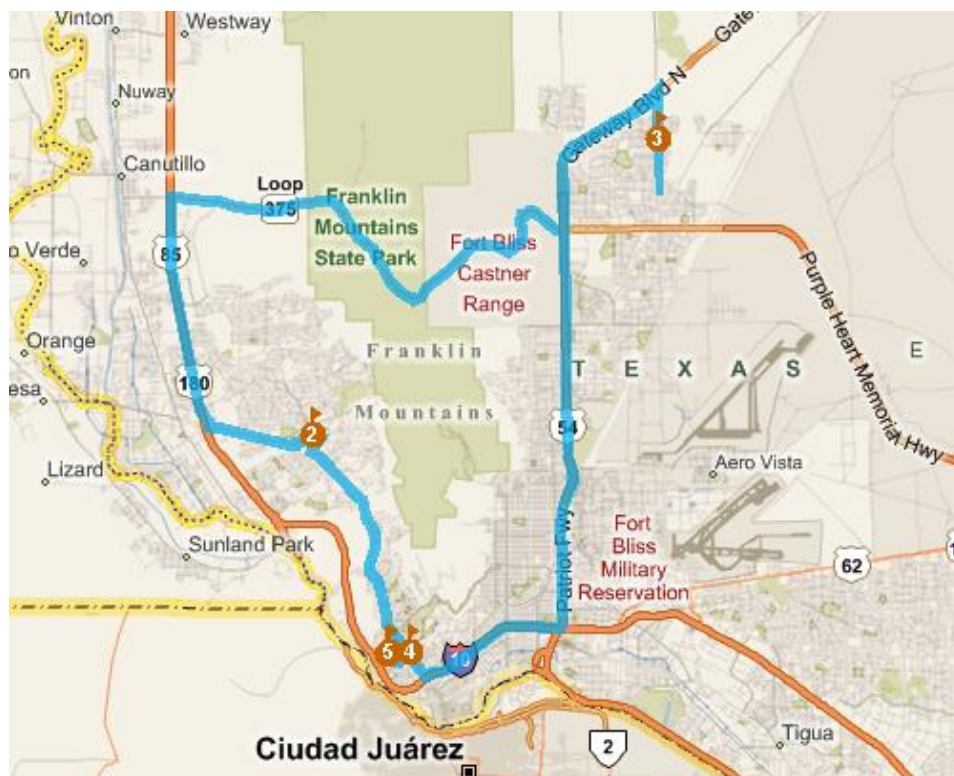


Figure 4-4. Field operation route 1 comprising sites A, D, and F. Source (Navteq ©2009).

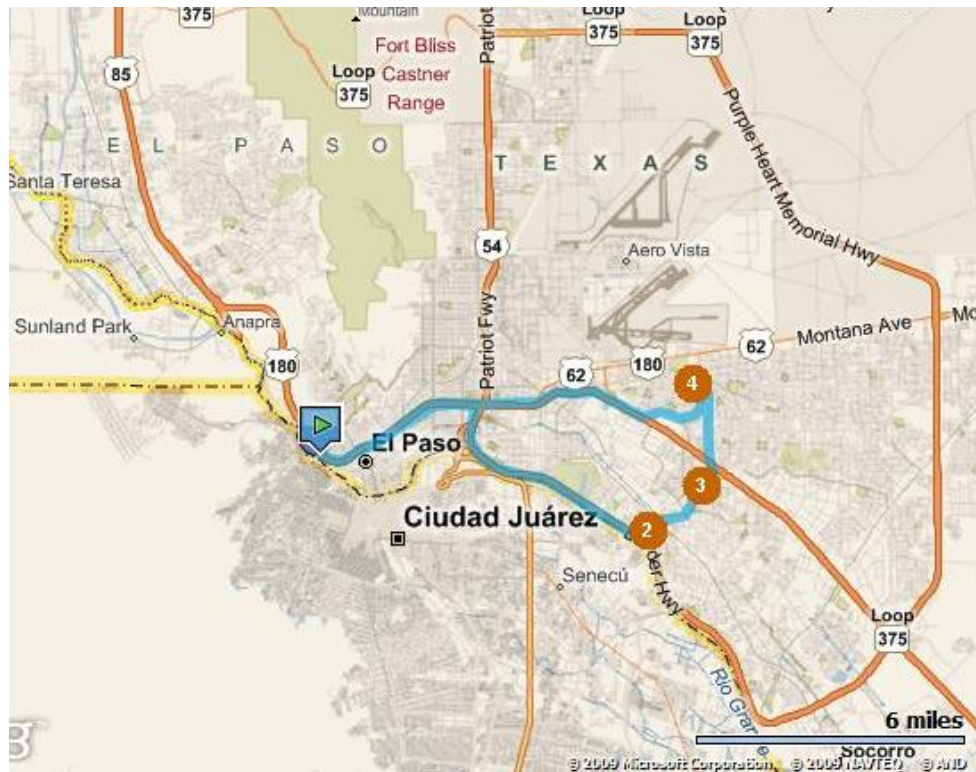


Figure 4-5. Field operation route 2 comprising sites H, I, and K. Source (Navteq ©2009).



Figure 4-6. Field operation route 3 comprising sites J and L. Source (Navteq ©2009).

Initially, the DICHOTS were calibrated at the lab using a primary flow meter (Dry Cal®, Bios International Corporation, Butler, NJ) setting the fine flow to 15 L/min and the coarse flow to 1.67 L/min. Then each team assembled and collocated the samplers at their corresponding site on the assigned field operations day. Every week, each team was sent to their corresponding route to deploy new filters and to retrieve the sampled filters. Each team hauled a chain of custody, petridishes, flow meter and field blanks in some cases, and antistatic gloves. At the monitoring site the field researchers' recorded in the chain of custody the current total fine flow and coarse flow. Then the equipment was turned off and the loaded filters were removed, using the antistatic gloves, and stored in their reciprocal petridishes. Subsequently, the new filters were installed and the time recorded in the chain of custody. Afterwards the samplers were turned on and the total flow and coarse flow affixed accordingly to the chain of custody. Finally the petridishes were capped, and stored with its respective chain of custody in a conditioned-box and thus proceed to the following site. The field blanks, on the other hand, were rotated among each team in a cycle of three-weeks for the first and second team whereas two weeks for the two-site team. During this three-week rotation the field blank was briefly exposed each week at a predefined site, continuing the exposure at the subsequent site the following week. At the evaluation site, the field blank was momentarily installed in the sampler, removed, and stored in its protective container inside the conditioned-box at the evaluation site until it is retrieved back to the weighing laboratory. Moreover, the primary flow meter was also rotated on a weekly basis among the three teams. Each week the flow meter was given to a team to perform a flow check at every DICHOT that the team was responsible for. Likewise, a duplicate dichotomous sampler was deployed at a different site each week in a rotating fashion during the monitoring campaign. Although, when the rotating cycle within the sites of each team concluded, the following team

was responsible to retrieve the sampler from its predecessor site and allocated it to one of their own sites in order to resume the cycling routine.

During the months when temperature exceeded 95°F the DICHOT were brought back to the lab for maintenance. In August 2007 all eight DICHOT were retrieved from their respective sites. At the lab a technician performed a status check and enforced labor to those samplers that required maintenance. The samplers were deployed in September and some relocated at new sites. The succeeding year the DICHOT were not brought back to the lab in an all inclusive campaign, instead they were retrieved one by one and replaced with the duplicate sampler. Maintenance of the DICHOT included replacements of pumps, pressure regulators, tubing, and flow meters.

4.4 SAMPLING ANALYSIS

Standardized operating procedures for the filter media preparation, gravimetric operations, and sampling were followed accordingly to Orquiz (2001). 37-mm-diameter Teflon membrane filters (nominal pore size, 1 μ m; Pall Corp., East Hills, NY) were conditioned for at least 24-hour period on a temperature-humidity control chamber at a relative humidity of 30 \pm 5% with a temperature of 25 \pm 5°C, for temperature and humidity equilibration prior to weighing at the University of Texas at El Paso (UTEP) Air Quality Lab. After pre-equilibration and just before weighing, the lab analyst performed a standardized calibration operating procedure for a microbalance, XP6 (Mettler Toledo®), with a \pm 1 μ g sensitivity. Following the standard operating procedures (SOP), five working weight standards (1 mg, 5 mg, 10 mg, 50 mg and 100 mg) were weighed three times and recorded on a laboratory batch data form before and after a filter lot was hefted. During the sampling campaign the microbalance pre- and post-calibration showed the

exact synonymous correlation; wherein the 1 mg and 5 mg were the single weights to prompt a minimal variability between the actual and measured values giving a residual average of $\pm 2.0\text{E}-3$ and $\pm 1.00\text{E}-3$ mg, respectively. A batch of 20 37-mm diameter Teflon filters were handled each week during the three-year sampling campaign. Two filters (coarse and fine) for each eight sampling sites and one duplicate were hefted three times (within the U.S EPA standard variability of $\pm 15 \mu\text{g}$) and their average weights recorded on the laboratory batch data form. However, for quality assurance and quality control, for every tenth consecutive filter weighed, one filter was pooled from that series and reweighed. In other words, out of the complete filter batch (20 filters) 10% (2 filters) underwent a secondary analysis. Moreover, one lab blank was weighed each week and used as a continuously reference for monitoring possible errors in calibration drift in the microbalance. Besides identifying microbalance errors, lab blanks were used to ascertain any possible laboratory contamination due to improper conditioning of the filters, improper procedure by the lab analyst, or any possible weight changes between pre-sampling and post-sampling, that would compromise the reliability of the measured values and even more the research study. Sources of random error are continuously linked to temperature variances. Filters may experience significant ambient temperature of over 28°C to below 4°C , loss/gain of mass from volatilizing species, or condensation of water vapor in the field reduces measurements precision, hence field blanks served as a gauge to quantified this error. Lastly, the pre-weighed filters were put into petridishes and stored in a control environment (U.S Environmental Protection Agency, 1998) for no more than 30 days prior to sampling. Loaded filters were removed from the field and transported to the Air Quality Laboratory at UTEP for conditioning and gravimetrical analysis. Post-weighing of sample filters followed the same procedure discussed above. Weighing of a lab blank was always performed during the

gravimetric analysis. After the post-weighing protocol, QA/QC procedure showed an average variability of $\pm 12 \mu\text{g}$ for lab blanks just, $\pm 3 \mu\text{g}$ under the EPA's standard (U.S Environmental Protection Agency, 1998), average field blank values for the three year campaign showed an average of $\pm 24 \mu\text{g}$ of mass deposited. Lastly, the calibration weighting for the pre- and post-weighing remained the same for all calibration weights.

All DICHOT measurements were reported in milligrams. After subtracting the post-weigh from the pre-weigh and multiplying it by a factor of 1000 (conversion factor from milligrams to micrograms) the mass variation was then reported in micrograms (μg). Further analysis was required to convert mass to concentration. Using Equation (3-1), and Equation (3-2), fine and coarse concentrations were reported as mass over the volume of air sampled.

$$PM_{2.5} = \frac{\text{Mass}_{2.5}}{(\text{Volume}_{\text{Fine}})} \quad (3-1)$$

$$PM_{10-2.5} = \frac{\text{Mass}_{10-2.5} - \text{Volume}_{\text{Coarse}} * \frac{\text{Mass}_{2.5}}{\text{Volume}_{\text{Fine}}}}{\text{Volume}_{\text{Total}}} \quad (3-2)$$

$$PM_{10} = PM_{2.5} + PM_{10-2.5} \quad (3-3)$$

Subsequently, the concentrations for fine and coarse PM were added up to obtained PM_{10} concentrations, Equation 3-3. Mass concentrations were reported as micrograms of particulate matter per cubic meter of air ($\mu\text{g}/\text{m}^3$).

4.5 DATASET VALIDATION

Ambient PM was measured at 13 sites in the El Paso area using a Dichotomous sampler's network. To evaluate the performance of the monitors, replicate concentrations of co-located referenced TEOMs at three state-operated air pollution monitoring stations were analyzed using

linear regression. The approximately normal distribution of the samples permitted the use of linear regression with the standard Gaussian assumptions. Initially, four monitoring sites were co-located with TCEQ's monitoring stations throughout the sampling campaign. Sites like D (skyline), F (UTEP), H (Chamizal), and J (Socorro) were correlated with state monitoring CAMS stations 72, 12, 41, and 49, respectively. The sparseness of the stations were purposely design to cross-validate one sampler in each stratified quadrant of the city. Site D was located in the northeastern part of the city side by side to CAMS 72 sampler; due to the absence of northwestern TCEQ stations. Site F was established on the west side of the city at the same location as CAMS 12; justified by its proximity to a port of entry and the high commercial activity. Site H was located in Downtown El Paso closed to the border at the same dwelling as CAMS 41. Site J resided at CAMS 49 in the southeastern part of the city. However, not all co-located sites performed a parallel air pollutant sampling. Site F, H, and J started their PM₁₀ monitoring campaign in 1998 and currently continue to sample ambient particulate matter. Site D, on the other hand, had no PM measurement campaign during the time period of the study. Therefore site D as well as CAMS 72 had to be withdrawn from the validation process and only three subsequent samplers were analyzed. Yet, not all of the three monitors sampled during the same time period. For example, site F and J were fixed stations from the onset to the end of the monitoring campaign, whereas site H had to be relocated to Parral in August 2007 to increase the coverage area. The measurements of the three sites were performed in different time periods. Therefore, the sampling period of the three monitors was compared to the sampling period of their corresponding fix station. Weekly concentrations that did not concurred in time were removed from the analysis. The annual data was calculated as the three-year arithmetic mean concentrations for each site. The concentrations for winter, spring, summer, and fall were

calculated as the arithmetic mean of the three terms that each season had within the three-year campaign. Current PM₁₀ measurements were obtained from the TEOM at each CAMS site and averaged into a 1-week average, corresponding to the hours when the co-located DICHOT PM measurements were collected.

Moreover, yearly PM model concentrations were averaged over the weekly concentrations computed from the three-year monitoring campaign. The six PM₁₀ yearly samples (F, H, J, CAMS 12, CAMS 41, and CAMS 41) were statistically analyzed as shown in Table 4-2 with averages ranging from 30.6 µg/m³ in UTEP to 43.1 µg/m³ in CAMS 41, with a standard deviation of 9.9 and 21.4 µg/m³, respectively.

Table 4-2. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the annual model.

Annual Descriptive Statistics					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
UTEP	99	8.4	57.1	30.6	9.9
CAMS 12	133	11.3	69.3	33.3	11.6
Chamizal	42	11.6	73.1	36.4	15.1
CAMS 41	49	12.4	121.9	43.1	21.4
Socorro	112	9.8	81.18	41.4	15.9
CAMS 49	133	10.8	107.1	42.8	16.3

Likewise, linear regression analyses prompt strong correlations between the measured samples under the current research program and the permanent state monitoring stations. Correlation coefficients range from a low 0.75 between site F and CAMS 12 to a high 0.86 among site H and CAMS 41 as shown in Figure 4-7, Figure 4-8, and Figure 4-9. It is possible that the higher concentrations shown by the state monitoring stations are attributable to the quality control and quality assurance methods used to assess their data. Concentrations measured by the TCEQ's TEOM are recorder and send to the state central headquarters in Austin, Texas. Then over a

three-month period the data is evaluated and uploaded into TCEQ's web page. The method used by the state agency to determine outliers or faulty readings is beyond this study. However, the detection limit used by TCEQ agency is responsible for the high average concentrations.

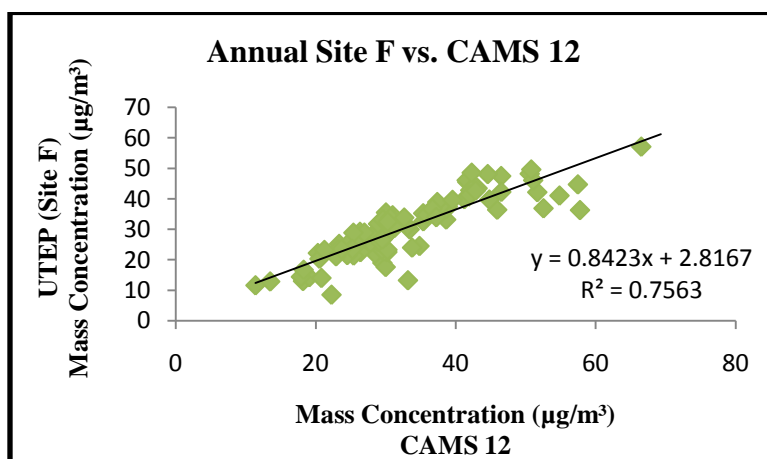


Figure 4-7. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the annual dataset.

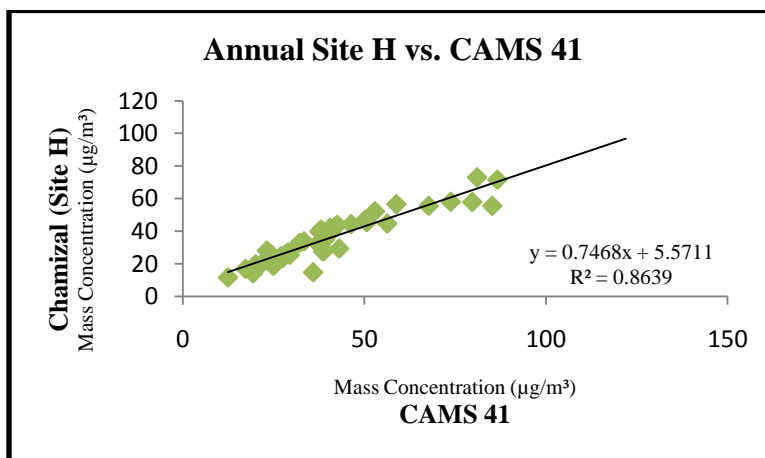


Figure 4-8. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the annual dataset

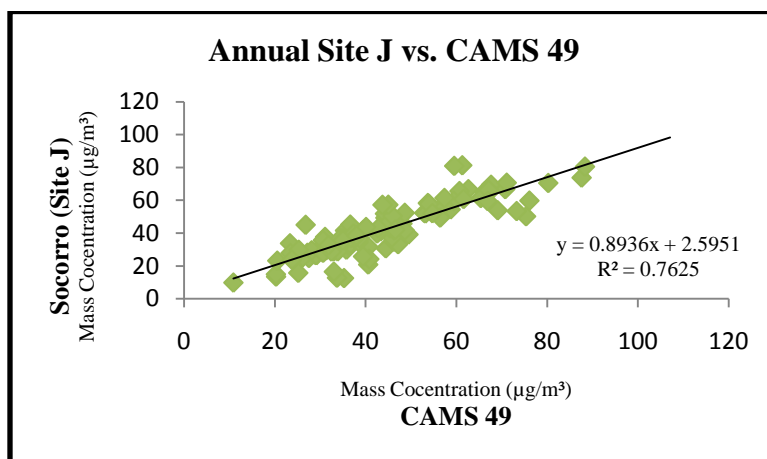


Figure 4-9. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the annual dataset.

For the spring model, concentrations were calculated in a different fashion. Weekly values from March 21st to June 20th were extracted from the three-year measurements programme and averaged in order to obtain a representative value for that particular time period. Remarkably, Chamizal had only 11 sampled weeks in contrast to 21 and 29 for UTEP and Socorro, respectively. Note that the sampled values were performed exclusively for the spring campaign prompting a total of 36 possible measurements. However, modest weekly concentrations observed at Chamizal corresponded to the rotation of the monitor after one year of sampling. During this sampling period (1 year) only 11 weekly concentrations out of the 12 possible samples were reported for the Chamizal site, as one sample had to be dropped from the analysis due to inefficiencies in the field operations.

Table 4-3. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the spring model.

Spring Descriptive Statistics					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
UTEP	21	13.9	49.5	34.1	9.5
Cams12	31	20.7	66.0	38.1	11.2
Chamizal	11	22.0	73.0	39.8	13.8
Cams41	13	24.4	81.0	43.0	16.1
Socorro	29	24.8	81.1	49.2	14.9
Cams49	31	26.7	67.1	47.5	12.5

In regards to the modest amount of samples in a three-year period the Chamizal site was determined to be kept for the analysis as it compromised a complete year of weekly measurements for the spring season as well as weekly concentrations for the remaining seasonal time periods. Average spring concentrations for all six monitors ranged from a low 34.1 to a high 49.4 $\mu\text{g}/\text{m}^3$ with a standard deviation of 9.5 $\mu\text{g}/\text{m}^3$ at UTEP to a relatively high value of 16.1 $\mu\text{g}/\text{m}^3$ shown at CAMS 41 (Table 4-3).

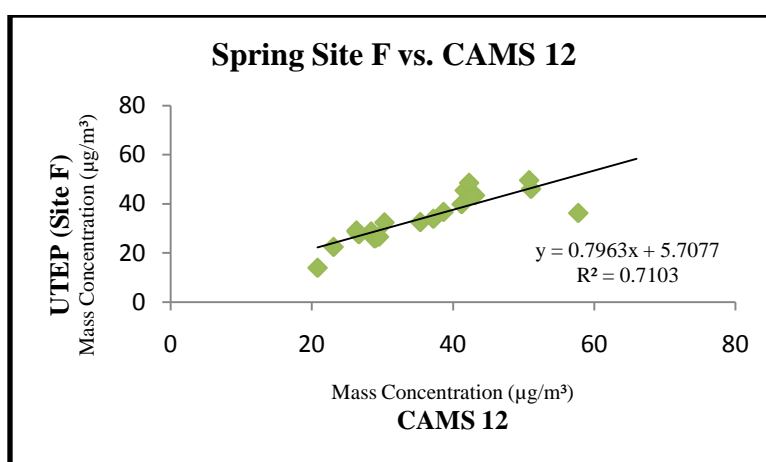


Figure 4-10. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the spring dataset.

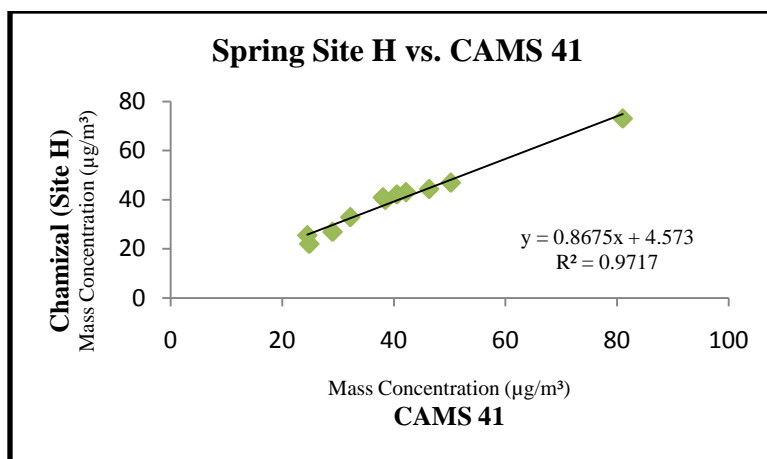


Figure 4-11. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the spring dataset.

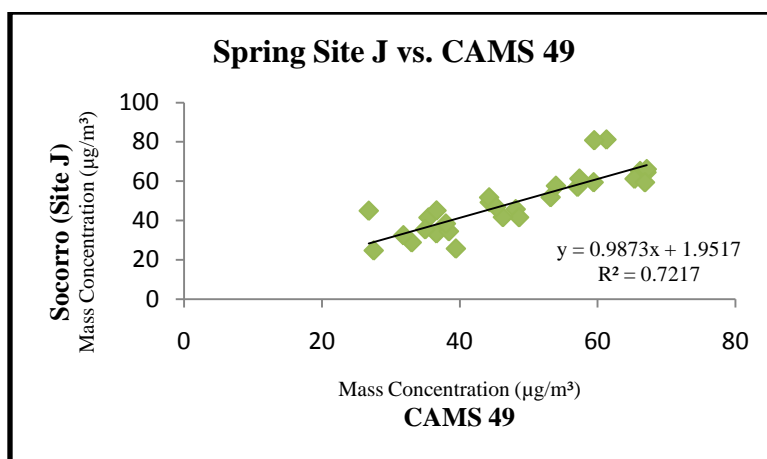


Figure 4-12. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the spring dataset.

Independent of the modest amount of events unveiled by site H, the correlation for co-located samples stayed strong with the R^2 varying from 0.71 between UTEP and CAMS 12 to 0.97 between H and CAMS 41.

However, summer concentrations differ from all other cross-validation models. Summer time period comprised from 21st of June to the 20th of September. During this time frame monitors underwent maintenances for a period of one month in average. As a result, a good amount of weekly concentrations were prone to be lost reducing the number of events to a small number. Weekly concentrations for this particular time of the year were extracted from the

database and averaged for background concentrations for the three-month summer period.

Similar to the spring dataset, Chamizal prompted the smallest number of sampled weeks. Out of the 12 possible measurements one had to be withdrawn as it represented an outlier with a concentration above 450 $\mu\text{g}/\text{m}^3$. The remaining two were dropped from the analysis due to equipment repair and maintenance. Most of the DICHOTS sampled in a continuous mode, thus by the time high temperature reached El Paso a few sensitive parts of the DICHOTS (such as rotameters and pumps) may have malfunctioned under the extreme heat in the summer. As a result, monitors, as mentioned before, were brought back to the lab for repair and maintenance.

Table 4-4 shows the descriptive statistics for the summer where the concentrations seem to be below the previous seasonal concentrations. Values ranged from 23.3 $\mu\text{g}/\text{m}^3$ at UTEP to 38.3 $\mu\text{g}/\text{m}^3$ at CAMS 41. Standard deviation varied from a low 8.8 $\mu\text{g}/\text{m}^3$ at UTEP to a high of 15.6 $\mu\text{g}/\text{m}^3$ at CAMS 41.

Table 4-4. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the summer model.

Summer Descriptive Statistics					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
UTEP	13	8.4	36.3	23.3	8.7
Cams12	24	13.4	45.9	26.2	8.0
Chamizal	9	14.1	58.0	31.9	13.7
Cams41	10	19.3	73.7	38.2	15.5
Socorro	14	12.6	39.1	22.9	8.9
Cams49	24	17.1	45.0	29.6	8.5

The number of weekly samples by TCEQ outnumbers the monitoring sites in the program.

Therefore the correlation between each co-located site was performed under the same corresponding time frame measured in site F, H, and J. Figure 4-13, Figure 4-14, and Figure 4-15

illustrate the correlation between each parallel monitoring site, thus validating the precision of the measured concentrations in each site during the monitoring campaign. This will further help in the development of a stochastic model approach that will derive into a more reliable assessment of the PM₁₀ spatial gradient.

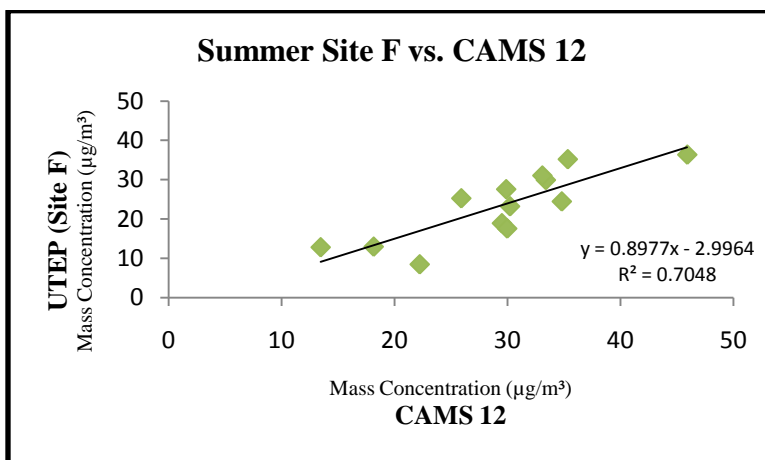


Figure 4-13. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the summer dataset.

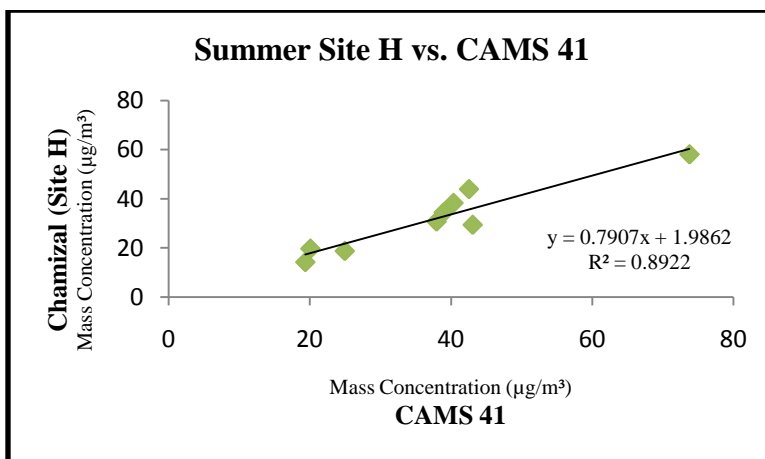


Figure 4-14. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the summer dataset.

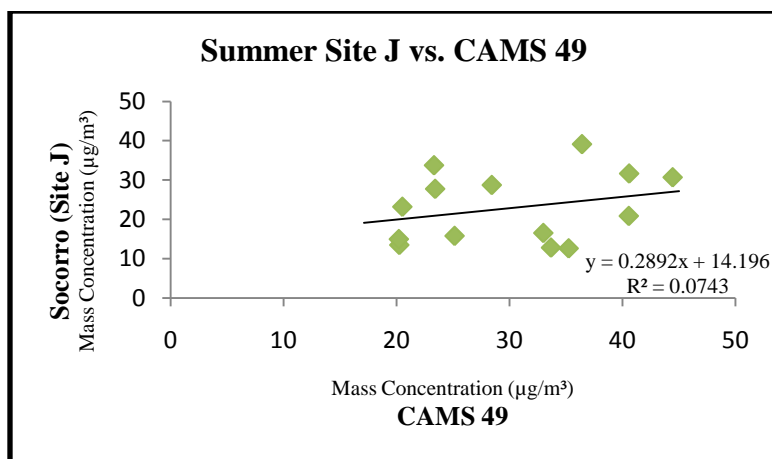


Figure 4-15. Correlation of the monitoring site J and TCEQ’s CAMS 49 analyzer for the summer dataset.

Co-located samples collected at sites F and H were found to be in good agreement with TEOM measurements with a R^2 value of 0.70 and 0.89, respectively. However, poor correlation was found between the values in site J and CAMS 49 (Figure 4-15). The discrepancy between these two sites was attributable to equipment malfunction during the summer season in the first two years of the monitoring campaign.

The fall data was analyzed in order to conduct a preliminary evaluation of the sensitivity to source contributions when the temperature trend begins to descent from an average of 97°F seen in summer. Weekly concentrations from September 21st to December 20th were extracted from the three year database and cross-validated with that recorded at state air quality station. The number of samples was almost similar for sites F and J; however Chamizal site only had a total of 12 samples in one-year.

Table 4-5. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the fall model.

Fall Descriptive Statistics					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
UTEP	34	13.2	72.5	33.9	14.0
Cams12	39	19.0	69.3	36.8	13.2
Chamizal	12	14.8	111.0	47.7	26.6
Cams41	13	23.1	121.9	57.9	29.3
Socorro	35	24.1	95.8	45.2	16.9
Cams49	39	28.0	107.1	50.6	19.8

Three-year weekly averages from September 21st to December 20th were approximately normally distributed (APPENDIX C), with averages ranging from a high value of 57.9 $\mu\text{g}/\text{m}^3$ with a standard deviation of 29.3 $\mu\text{g}/\text{m}^3$ at CAMS 41 to a low value of 33.9 $\mu\text{g}/\text{m}^3$ with a standard deviation of 14 $\mu\text{g}/\text{m}^3$ at site F (Table 4-5).

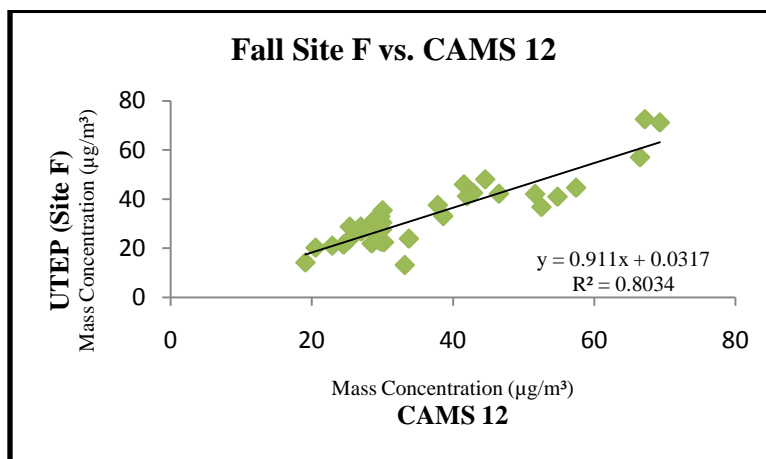


Figure 4-16. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the fall dataset.

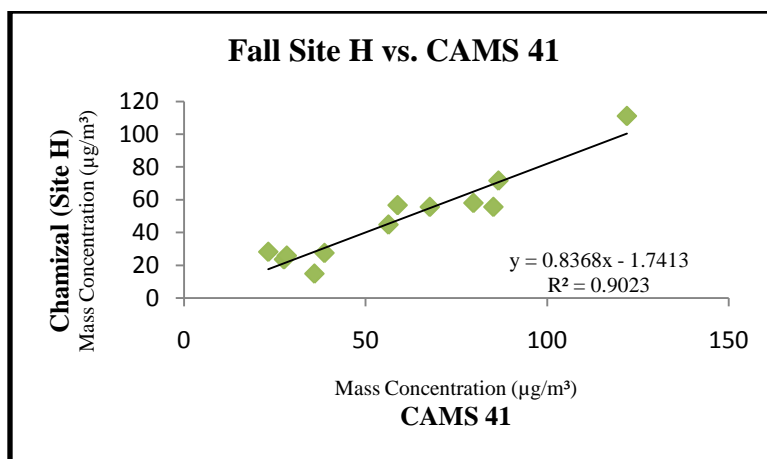


Figure 4-17. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the fall dataset.

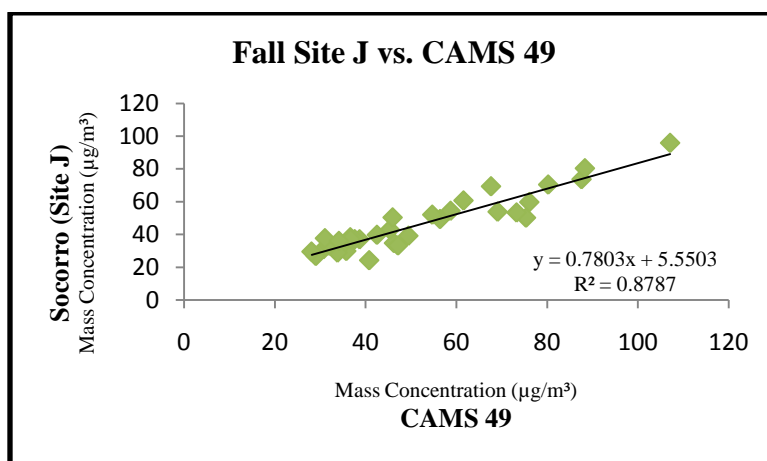


Figure 4-18. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the fall dataset.

Likewise, linear regression analysis shows strong correlations between the measured values and that observed at the state monitoring stations as shown in Figure 4-16, Figure 4-17, and Figure 4-18. Independent of the modest number of events at site H, the correlations between co-located samples at each site were strong, with R^2 values varying from 0.80 between UTEP and CAMS 12 to 0.90 at site H and CAMS 41. Overall the cross-validation measurements performed adequately. The high concentrations were prompted at the Chamizal (Downtown) site due to its surroundings and proximity to the port of entry, whereas much lower concentrations at UTEP (West) were observed as it remains as a residential area for the most part. As the cross-validation

methodology gave high correlation coefficients between the three sites and the co-located state monitoring stations, the reliability of the measurements performed during this research for the most part reckoned to be 75% accurate in detecting the variability of PM₁₀ concentrations among the monitoring stations. Furthermore, this accuracy was prone to be consistent at the remaining monitoring stations as the field operations and logistics were performed under the same standard operation procedure. Moreover, every cross-validated site was comprehended in each of the three routes. Site F was included in route 1, site H comprised route 2 and site J comprised route 3. Therefore the concentrations sampled at the eight purpose-allocated stations, throughout the monitoring program, were analyzed for their corresponding time period considering the reliability evidenced in the cross-validation analysis. Lastly, the winter model concentrations were calculated under the same procedure. Weekly values from December 21st to March 20th were extracted from the three-year measurements programme and averaged in order to obtain a representative value for that particular time period. Concentrations for the winter period range from 30.2 µg/m³ in UTEP to 40.3 µg/m³ in Socorro, with a standard deviation of 9.1 and 14.4 µg/m³ respectively as shown in Table 4-6.

Table 4-6. Descriptive statistics of the 3 co-located monitoring sites with TCEQ's fix stations for the winter model.

Winter Descriptive Statistics					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
UTEP	33	11.5	48.2	30.3	9.1
Cams12	39	11.4	50.6	30.4	9.1
Chamizal	11	11.7	52.2	31.2	12.6
Cams41	13	12.5	52.9	32.2	12.0
Socorro	35	9.9	70.6	40.3	14.4
Cams49	39	10.9	71.1	39.4	12.9

No outliers were required to be removed as concentrations above 150 $\mu\text{g}/\text{m}^3$ did not fall under this particular three-month seasonal three-year average. Thus, when comparing the concentrations from the measured values and the TCEQ samples the correlations range from 0.97 between site H and CAMS 41, and 0.90 between site J and CAMS 49 as shown in Figure 4-19, Figure 4-20, and Figure 4-21.

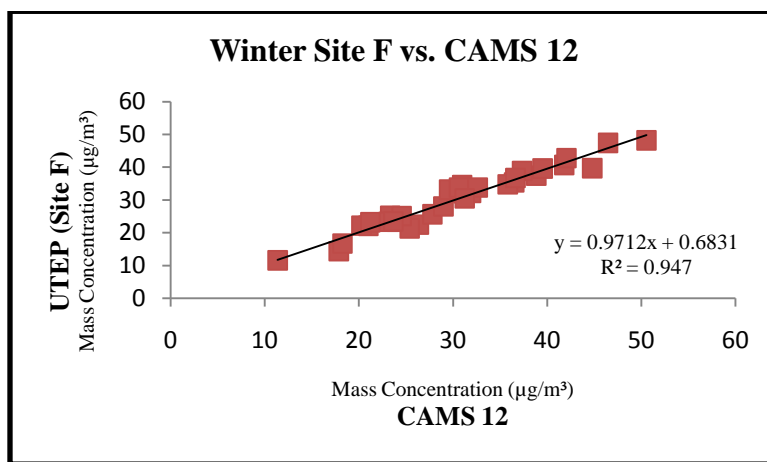


Figure 4-19. Correlation of the monitoring site F and TCEQ's CAMS 12 analyzer for the winter dataset.

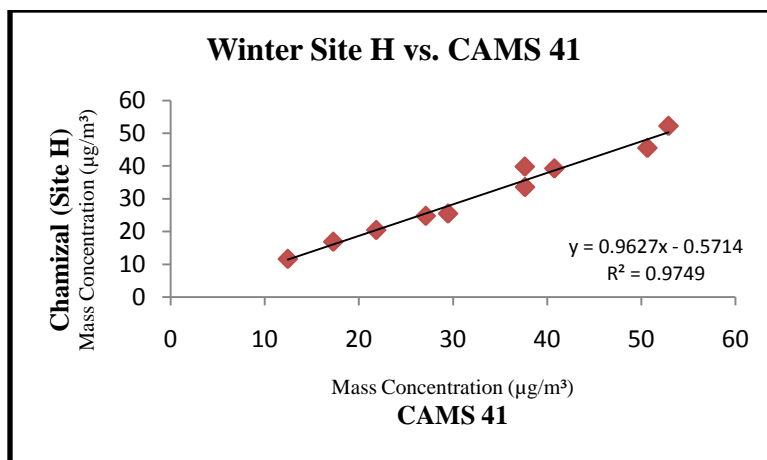


Figure 4-20. Correlation of the monitoring site H and TCEQ's CAMS 41 analyzer for the winter dataset.

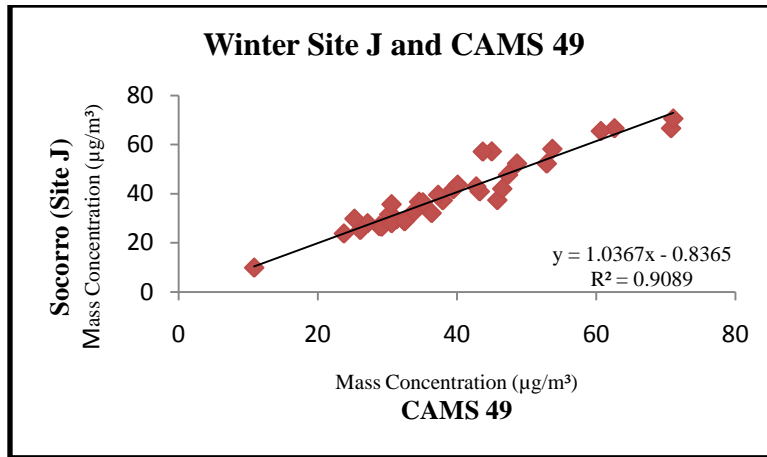


Figure 4-21. Correlation of the monitoring site J and TCEQ's CAMS 49 analyzer for the winter dataset.

5 GEOGRAPHICAL DATA

5.1 BUFFERING

The average concentrations for the different models (yearly, winter, spring, summer, and fall) were related to surrogate variables collected from ESRI ArcGIS 9.2 geographic information system. Data on spatially resolved variables believed to be associated with intra-urban PM₁₀ emissions were collected and classified into the following categories: 1) traffic-related, 2) land use, 3) population, 4) property value, 5) meteorological parameters, and 6) distance to the international border. As with the sampling site selection, a common criterion was followed for the collection of geographic data and for the subsequent modeling. Predictor for variables in LUR models were computed for circular zones around each monitoring site, using buffer functions in GIS. The selection of buffer size was crucial in determining the performance of the model, as the spatial resolution of the estimates was selected to take account of the known dispersion patterns. For every sampled site location, a series of concentric circle buffers with radii of 200 m, 300 m, 500 m, 750 m, 1 km, 1.5 km, 2 km, 3 km, and 5 km were constructed. However, a corrected buffer was used instead of the regular buffer area as portions of wider buffers often laid outside El Paso jurisdictional area, especially those sites along the border. The corrections were developed by intersecting the buffers with the El Paso County political boundary, shearing the area of the buffer that posited outside the jurisdictional area. Thereupon, corrected buffer (CB) will be referenced as the buffer area used in this study. Variables that lay within buffer areas outside the jurisdiction of El Paso were assign a weighted value determined by the radii within El Paso boundary. Each CB size increment determined different variable magnitudes and were defined as new variables. In total 250 variables were generated from which 100 were logarithmically transformed to approximate normal distribution. Further, all geographic

information system layers were then clipped to these buffers. Although an examination of dispersion models suggest that 80-90% of the decay of pollutants occurred within 150 and 200m (Ross et al., 2006), a previous study in El Paso found that a detectable effect could be identified at higher distances necessitating the inclusion of wider buffers (Gonzales et al., 2005). In particular, they found that distance to the U.S Border a statically significant predictor of NO₂ concentration. Finally the NAD 1983 Stateplane Texas Central FIPS 4203 feet coordinate system was used as spatial reference for the geographic features. In fact, some variables required a coordinate transformation.

5.2 TRAFFIC-BASED VARIABLES

Personal automotive and commercial vehicles emit primary particulates and precursors that react with the ambient environment to produce secondary PM air pollution (Brauer et al., 2003). To improve estimates of traffic-generated particulate air pollution five indicator measures were developed as follows: 1) traffic counts, 2) vehicle miles traveled, 3) nearest distance to a major road, 4) truck road length, and 5) road length.

The traffic counts were obtained from El Paso Metropolitan Planning Organization (MPO). These traffic counts were collected from 2007 to 2008 for the city of El Paso. The annual average traffic counts (AATC) is defined as the average number of cars traveling in both directions per year. The AATC point data was spatially assigned to the closest individual road segments in the GIS streetmap dataset. Through Environmental Systems Research Institute (ESRI) the AATC magnitudes were intersected with each circular CB area and centered around the monitoring sites. The traffic counts were computed by averaging all AATC that fell within the sampler CB and then jointed spatially to a mother table.

The vehicle miles traveled (VMT) were also provided by MPO for 2005 and 2007. The miles traveled by a vehicle in El Paso region was computed by multiplying the annual average daily traffic by the roadway segment length, and then multiplying this figure by 365 days to produce an annual VMT segment. Calculating VMTs in GIS was slightly more complicated than traffic counts. Initially, VMTs were spatially represented as segments that in most cases some portion of the segment laid outside the circular zone when masked with the CB layer. Therefore, the spatial VMT dataset was first intersected with their respective CB. Then using a calculator tool in GIS the length of VMT segment within the CB was recalculated and then divided by the original length of the VMT segment in order to obtain a representative factor of the segment. This newly developed factor represented the percentage of the total segment partitioned within a CB. Further, the factor was multiplied by the original VMT values to produce a proportional value of VMTs. Finally, the fractioned VMTs were summed up by site and tabulated in the parent table. An additional variable was generated from 2005 and 2007 VMTs, as they were later averaged.

Likewise, road and truck road length variables were intersected with their corresponding CB area, and then using the same calculator tool from GIS, the new length was calculated. All the new dissected segments within each CB were summed and jointed to the parent table. It is important to mention that no factor (or percentage) was required since the purpose of such variables was to characterize solely the total length of all roads rather than obtaining a proportional value attributable to the segment within the circular area, such the case as traffic counts and VMTs.

Lastly, the nearest distance to a major road was obtained by measuring the shortest distance from the monitoring site to the nearest major road, using the NEAR command in GIS.

The shortest distances were tabulated in the parent matrix.

5.3 POPULATION

Population density is an important factor in determining how much and what type of pollution is produced in a given area. Densely populated areas typically contribute to more traffic-related pollution than sparsely populated areas, and within the city, density may also influence emissions. Higher population increases the traffic pollution from commuting and traveling. The latest U.S. Census Bureau (2009) block group database was used to determine population density within each buffer. Although the population in El Paso County increased since the last census, its spatial distribution within the city is believed to have experienced minimal changes. A GIS coverage map generated by the county of El Paso Appraisal District (updated in June 2008) was employed for calculating the population distribution within each CB. The Population was computed based on the lots within a parcel. Only parcels smaller than 65,000 square feet were considered in the analysis since most of the larger parcels correspond to industrial or recreational land use. Similar to the traffic-related variables, the census block group layer was clipped to the buffers discussed previously. Then the area of the intersected parcels that overlaid within the CB was recalculated. The intersected area was divided by the parcel total area (the pre-intersected parcel) in order to obtain a proportional fraction. Consequently, this proportional fraction was multiplied by the original population attributable to the parcel total area, as a way to represent the population intersected by the CB. All population values were tabulated in the parent matrix. The methodology discussed above assumes an even distribution of the population within a block group, which may be unrealistic but necessary due to the unavailability of the data.

5.4 LAND USE-BASED VARIABLES

Land use is a significant factor in predicting air pollution levels (Briggs et al., 2000), as different land uses yield different levels of pollutant emissions or transportation activities. For example, industrial and heating sources may release gaseous pollutants and PM₁₀, leaving areas surrounding the source to be relatively high in PM₁₀ concentration. Land use-based variables were determined from a geographical database generated by the City of El Paso Planning Department (last updated in December 2006), in which areas of ELP were classified into the following categories such as: low, medium, and high density of Industrial, residential, and commercial. Agricultural, parks, and open space areas were also included in the shape file. Due to the classifications of industrial, commercial, and residential into subcategories, the data was reckoned into unique categories that would have reasonable probability of coverage around an adequate number of pollution monitors. In conducting the convergence, these data were compiled into industrial, residential, commercial, and open space categories found to be associated with ambient pollution concentrations in prior studies (Briggs et al., 2000; Jerrett et al., 2005; and Ross et al., 2006).

For this analysis, land-use categories were used as a mask and overlaid at the CB areas. As the layers were intersected the total area of individual land use categories in each buffer were recalculated. The computed rationed area of each category was further summed and summarized for every monitor site, providing the total amount of area in kilometer square for every particular category at the parent table.

5.5 ADDITIONAL VARIABLES

Several additional variables that showed promise in previous analyses (Gonzalez et al.,

2005; Ross et al., 2006; Hochadel et al., 2006; Ryan et al., 2007; Moore et al., 2007; Jerrett et al., 2007; and Su et al., 2008) were included. These included the nearest distance to the border, elevation and property value. Elevation is considered to influence PM₁₀ concentrations, as the meteorology and complex terrain in ELP significantly limits the mixing and dilution of air pollution within the region resulting in exuberances of air quality. Elevation was considered a predictor variable for PM₁₀ concentrations followed a constant and proportional decline with increasing elevation at any given location. Digital elevation data was taken from USGS 7.5-DEM Series, scale 1: 24,000, and required no further calculations rather than a singular spatial joint between the sites and the elevation layer.

The hypothesis that an environmental justice variable could influence the concentrations of PM₁₀ at a giving location (U.S. Census Bureau 2009) was conceptually assessed in this study. However, defining the variable as “income” is a challenging and difficult to implement. Instead, the value of the property (or lot) was used as the environmental justice factor. As a result, a Parcels’ layer was obtained from El Paso Appraisal District. With the avail of GIS techniques this layer was clipped to the CB database and the area of each lot intersected and recalculated. Using the calculate geometry command in GIS, the new lot area was divided by the pre-dissected lot area in order to represent the proportion of the total area that lay inside the screen areas. Thereupon, the values of the lots were multiplied by the ratio to generate a representing value of the dissected area within each CB. Finally the values were averaged and summarized for every monitoring site. Note that lot areas greater than 65,000 square feet were previously removed from the layer as they may misrepresent the average value of the property within the CB as large parcels correspond to industrial, commercial or recreational lots.

Distance to the border was included in the analysis, as a previous study performed by

Gonzales et al. (2005) found that the nearest distance to a port of entry, altitude, and traffic-related variable explained 81% of the variability of the intra-urban gradient of NO_2 . The proximity to CDJ has prompted the idea of high concentrations of air pollution. Major PM emitting activities in CDJ vary from brick kilns to assembly industry and to numerous unpaved roads (Currey and Pumfre, 2006). This variable was calculated using the NEAR command in GIS in order to estimate the shortest straight-line distance from the monitoring site to the border limits. A preliminary analysis showed a correlation coefficient of $R^2=0.74$ when PM_{10} mass concentrations were correlated to the distance to the border and hence casted some doubts to the validity of the variable. The positive correlation generated higher concentrations at farther distances to the border and vice versa. This contradicted the initial assumption as well as the observed trend by the monitoring network. As a result, a priori approach was developed. Calculating, instead, the shortest straight-line from the El Paso political limit, between the state of Texas and New Mexico, to the monitoring site. This approach generated strong positive regression coefficients between the PM_{10} concentrations and the distance from the state political limits. In other words, using this priori methodology, the nearest distance to the border justified higher concentrations of PM at sites closer to the border between Mexico and the U.S. Furthermore, this layer did not require any further calculations rather than a simple spatial join to the mother table.

6 CONSTRUCTION OF THE MODELS

6.1 MODEL SELECTION

Multiple linear regression (MLR) is a method used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes also called the predicted and the independent variables the predictors. MLR is based on least squares: the model is fit such that the sum-of-squares of differences of observed and predicted values is minimized. The relation between the geographic variables and the annual and seasonal averages air pollution concentrations for the 13 sites was analyzed by multiple linear regression defined by Equation 6-1.

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + e_i \quad (6-1)$$

The regression equation is rendered as a map by multiplying all cells in the contributing variable layers ($x_{1,...,i}$) against their associated coefficients ($\beta_{1,...,i}$) and summing the resulting grids with the constant intercept β_0 and an error term e_i . The distribution of the concentration levels and predictor variables were characterized by percentiles. The association of continuous variables is described by Pearson correlation coefficients. The fit of linear models was assessed by the percentage of explained variation (R^2) and point estimates of the parameters and their given standard errors. Given a reasonable order of the regressors, the models were selected based on their sequential R^2 achieved by the additional inclusion of the respective regressor when the previous variables were already entered into the model. On the contrary, models involving counter-intuitive associations were excluded, even if these gave a strong basis for prediction (as indicated by R^2 and RMSE).

Using SPSS 17.0 (©2009 SPSS Inc.) and PASW statistics 18 (©2009 SPSS Inc.) five separate models were developed, a yearly model that used PM₁₀ average concentrations from 2006 through 2009, and four that included seasonal data (winter, spring, summer, and fall) of three years measurements. Univariate and multivariate mixed-effects regression models were used to predict the PM₁₀ mass concentrations. Variable distributions were evaluated and transformed using natural logarithm in order to approach a normal distribution for those parameter distributions skewed (see APPENDIX D). Candidate predictors were evaluated using primary exploratory data analysis and manual forward selection procedures, based on the sum of squares explained. Initially correlations among peer of predictor variables and association between each predictor variable and PM₁₀ concentrations were reviewed to assist in the model building. The scale with the highest R² was initially included. This procedure was applied first with the most influential predictor variable based on univariate regression. Next, we assessed multiple regression models including the subsequent most influential variable, based on the highest t-score for each variable, and the different spatial scale of the same variable in separate models. The remaining variables were re-evaluated for inclusion, as such, the same variable at different buffer distances was included if it outperformed other variables. Consequently, if the adjusted R² was higher for the multiple regression models, that model was used. Finally, the effect of adding another variable to the core model was evaluated using the adjusted R² until a model with the highest adjusted R² was obtained. Moreover, stepwise and all-subsets procedures were also used to identify strong models that would not otherwise be identified using forward selection. Bivariate correlation coefficients for PM₁₀ mass concentration and all independent variables were estimated. All independent variables were evaluated against the dependent variable using bivariate least-square regression method. The independent variable with the

lowest f probability and smaller than 0.05 was entered into the model equation. Variables already in the regression equation were removed if their probability of f became larger than 0.10 after a new variable was included at a subsequent step. The process continued until no more variables were eligible for inclusion or removal. Generally, the sign of each coefficient was required to be in accordance with the expected causality of the factor, i.e. coefficients for elevation had to be negative and near distance to the border positive. Firstly, models incorporating only elevation and near distance to the board were prioritized to determine the most relevant spatial scales. Afterwards, we checked carefully whether these models could be improved by entering further geographic variables. If a linear coefficient had an incongruent sign, the corresponding independent variable was discarded and the variable selection and model fitting process was repeated. Normal probability plots of the standardized residuals were employed to evaluate the normality of the residuals. Scatter plots of the standardized residuals against the selected predictor variables and the fitted values were employed to test linearity and homogeneity assumptions. Collinearity among the predictor variables was evaluated with the variance inflation factor (VIF). Variables with the highest VIF and variables with the lowest t -scores were removed until a parsimonious prediction model with the highest R^2 and acceptable collinearity was derived. All variables were considered both independently of other variables in the forward selection process and in the combination with other variables in both stepwise and all-subsets procedures. After discarding the models that did not satisfactory meet the underlying assumption inherent to the methods, the model with the best fit was selected for validation.

6.2 MAP-BASED GENERATION

The spatial variability of PM_{10} was mapped for the annual and seasonal models using five different interpolation techniques in ArcGIS (inverse distance weight, thin-plate spline, natural neighbor, and ordinary and universal kriging).

As interpolation techniques are based on the fundamentals of constructing new data points with the range of a discrete set of known data points, and with exclusively 13 particular sampled sites for each model, it was imperative to generate a gridded format in order to sharpen the resolution of the depicted overall PM_{10} trend. By definition, a grid is a spatial data structure that defines space as an array of cells of equal size that are arranged in rows and columns. In the case of a grid that represents a surface, each cell contains an attribute value that represents a change in Z-value. The location of the cell in geographic space is obtained from its position relative to the grid's origin. Thereby, a grid was generated with an origin based on a northeastern reference point determined manually using Google Earth (©2010 Tele Atlas). The x - and y -coordinates, 300 km easting and 3492 km northing in Universal Transverse Mercator (UTM) 13N, were then transferred to a spreadsheet in Microsoft excel where 610 m were uniformly added to the domain's x - and y -coordinates to create precise and evenly distributed spatial points covering the entire ELP, Juarez, and Sunland Park area.

Furthermore, once the point locations were geographically defined on the spread-sheet, the gridded format was imported and projected into ArcGIS. The grid cell, with a domain of 102 km wide and 86 km long or 51 by 43 grid cells, was consequently intersected with the census tract coverage in order to exclude solely the grid points that lay within the cohort study. A final grid with a 610 m by 610 m isotropic resolution covering the entire urban region of the City of El Paso was saved under five different files names each corresponding to every model developed.

After deriving the parsimonious operational models, the coinciding independent parameters were determined for 3528 grid points and incorporated into each attribute table. Thereupon, the variables were multiplied by their reciprocal coefficients and reckoned into a solely field for each grid point. Based upon the measurements at the monitoring sites and with the calculated PM₁₀ concentrations at each grid point five main interpolation approaches were developed.

Accordingly to the literature, there are two main groupings of interpolation techniques: deterministic and geostatistical. Deterministic interpolation techniques (inverse distance weight, SPLINE, and natural neighbor) create surfaces from measured points, based on either the extent of similarity or the degree of smoothing. On the other hand, however, geostatistical techniques (kriging) quantify the spatial autocorrelation among measured points and account for the spatial configuration of the sample points around the prediction location.

Inverse Distance Weight, an exact and local deterministic interpolation technique, is the simplest interpolation method and it explicitly implements the assumption that the interpolating surface is a weighted average of the scatter points and the weight assigned to each scatter point diminishes as the distance from the interpolation point to the scatter point increases. Thus, IDW assumes that each measured point has a local influence that diminishes with distance.

$$Z(x, y) = \frac{\sum_{i=1}^n Z_i \frac{1}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (6-2)$$

Equation (6-2) demonstrates the calculation of the inverse weighted distance, where: $Z(x,y)$ is the inverse distance weighted at point (x,y) ; Z_i is the Z value at known point i ; d_i is the distance between point i and point (x,y) ; n is the number of known points used in estimation; and k is the specified power. Intuitively, this represents the average of the values of the surrounding

points weighted by the inverse of the distance to those points. An important characteristic to remark of IDW interpolation is that all predicted values are within the range of maximum and minimum values of the known points.

Similarly, Natural Neighbor (NN) method, an exact and local deterministic technique, is based on the same basic equation used in IDW. The difference, however, between IDW interpolation and natural neighbor interpolation relies on the method used to compute the weights and the method used to select the subset of scatter points used for interpolation. The natural neighbor interpolation approach is based on a moving weighted average technique that uses geometric relationships in order to choose and weight nearby points. That is, natural neighbor finds the closest subset of input samples to query point and applies weights to them based on proportionate areas in order to interpolate a value.

$$G(x, y) = \sum_{i=1}^n w_i f(x_i, y_i) \quad (6-3)$$

This technique is defined by the Equation (6-3), shown above, where: $G(x,y)$ is the NN estimation at (x,y) ; n is the number of nearest neighbors used for interpolation; $f(x_i,y_i)$ is the observed value at (x_i,y_i) ; and w_i is the weight associated with $f(x_i,y_i)$. The weights (w_i) depend on the area about each of the data points (Voronoi polygons) instead of the distance between data points, as with Inverse Distance Weighting. By comparison, a distance based interpolator such as IDW would assign similar weights to the northern most point and to the north-eastern point based on their similar distance from the interpolation point. Natural Neighbor interpolation, anyhow, assigns weights based on the percentage of overlap.

The third and final deterministic interpolation method used for mapping the spatial variability of PM_{10} was based on the thin-plate SPLINE principal, which is a form of

interpolation where the interpolant is a special type of piecewise polynomial called a spline. This interpolation method, an exact and local technique, creates a surface that passes through the control points and has the least possible change in slope at all points. In other words, thin-plate splines fit the control points with a minimum curvature surface. The approximation of splines is of the form:

$$Q(x, y) = \sum A_i d_i^2 \log d_i + a + bx + cy \quad (6-4)$$

Where: x and y are the x - and y -coordinates of the point to be interpolated; d_i^2 is $(x-x_i)^2 + (y-y_i)^2$ in which x_i and y_i are the x - and y -coordinates of control point i . Thin-plate splines consist of two components: whither $(a + bx + cy)$ represents the local trend function, which has the same form as a linear or first-order trend surface; and $d_i^2 \log d_i$ represents a basis function, which is designed to obtain minimum curvature surfaces. The coefficients A_i , a , b , and c are determined by a linear system of equations.

$$\sum_{i=1}^n A_i d_i^2 \log d_i + a + bx + cy = f_i \quad (6-5)$$

$$\sum_{i=1}^n A_i = 0 \quad (6-6)$$

$$\sum_{i=1}^n A_i x_i = 0 \quad (6-7)$$

$$\sum_{i=1}^n A_i y_i = 0 \quad (6-8)$$

Where: n is the number of control points; and f_i is the known value at control point i . The estimation of the coefficients requires $n + 3$ simultaneous equations. Unlike the IDW method, the predicted values from thin-plate splines are not limited within the range of maximum and minimum values of the known points. Therefore, splines strength lies in its ability to correlate data which doesn't follow any specific pattern without a single polynomial's extreme behavior.

Howbeit, a major problem with thin-plate splines is the steep gradients in data-poor-areas, often referred to as overshoots. But nevertheless, the regularized option modifies the minimization creating a smooth, gradually changing surface with values that may lie outside the sample data range.

Conversely, a geostatistical interpolation method offers assessment of prediction errors with estimated variances. Kriging, an advance, local and exact geostatistical method, differs from the interpolation methods discussed so far because it can assess the quality of prediction with estimated prediction errors. Kriging assumes that the spatial variation of an attribute such as change in grade within an ore body is neither totally random (stochastic) nor deterministic. Instead, the spatial variation may consist of three components: a spatially correlated component, representing the variation of the regionalized variable; a “drift” or structure, representing a trend; and a random error term. The interpretation of these components has led to the development of two types of kriging methods for spatial interpolation, ordinary kriging and universal kriging.

Ordinary kriging, an exact and local method, assumes that the variation in z-values is free of any structural component (drift). Instead of weighting nearby data points by some power of their inverted distance, ordinary kriging relies on the spatial correlation structure of the data to determine the weighting values. This is a more rigorous approach to modeling, as correlation between data points determines the estimated value at an unsampled point. The basic equation used in ordinary kriging is shown below:

$$F(x, y) = \sum_{i=1}^n w_i f(x_i, y_i) \quad (6-9)$$

Where $F(x, y)$ is the estimated value; $f(x_i, y_i)$ is the known value at point x_i, y_i ; w_i is the weight associated with x_i and y_i coordinates; and n is the number of sample points used in

estimation. The weights thus can be derived from solving a set of simultaneous equations. For example, to interpolate at a point P based on the surrounding points P1, P2, and P3 the weights w_1 , w_2 , and w_3 must be found through the solution of equations described below.

$$w_1s(d_{11}) + w_2s(d_{12}) + w_3s(d_{13}) + \lambda = s(d_{1p}) \quad (6-10)$$

$$w_1s(d_{12}) + w_2s(d_{22}) + w_3s(d_{23}) + \lambda = s(d_{2p}) \quad (6-11)$$

$$w_1s(d_{13}) + w_2s(d_{23}) + w_3s(d_{33}) + \lambda = s(d_{3p}) \quad (6-12)$$

$$w_1 + w_2 + w_3 + 0 = 1.0 \quad (6-13)$$

Where: $s(d_{ij})$ is the semivariance between known points i and j ; $s(d_{ip})$ is the semivariance between the i th known point and the point to be estimated; and the λ is a Langrange multiplier which is added to ensure the minimum possible estimation error. Upon that, this method shows that weights used in ordinary kriging involve not only the semivariances between the point to be estimated and the known points; but also those between the known points. This differs from the IDW method, which uses only weights applicable to the point to be estimated and the known points. Moreover, an additional difference between ordinary kriging and other local interpolation methods remarks the ability of kriging to produce a variance measure for each estimated point to indicate the reliability of the estimation.

Dissimilarly to ordinary kriging, universal kriging, an exact and local technique, assumes that the spatial variation in Z values has a drift or a trend in addition to the spatial correlation between the sample points. One of the assumptions made in universal kriging is that data being estimated is stationary. That is, as you move from one region to the next in the scatter point set, the average value of the scatter points is relatively constant. Whenever there is a significant spatial trend in the data values such as a sloping surface or a localized flat region, this

assumption is violated. In such cases, the stationary condition can be temporarily imposed on the data by use of a drift term. By definition the drift is a simple polynomial function that models the average value of the scatter points. The residual is the difference between the drift and the actual values of the scatter points. Since the residuals should be stationary, universal kriging is performed on the residuals and the interpolated residuals are added to the drift to compute the estimated values. Typically, universal kriging incorporates a first-order (plane surface) polynomial in the kriging process,

$$M = b_1x_i + b_2y_i \quad (6-14)$$

Where M is the drift, x_i and y_i are the x -, y -coordinates of the sample point i ; and b_1 and b_2 are the drift coefficients. Nevertheless, universal kriging also incorporates a second order (quadratic surface) polynomial in the process as describe in Equation (6-15).

$$M = b_1x_i + b_2y_i + b_3x_i^2 + b_4x_iy_i + b_5y_i^2 \quad (6-15)$$

Albeit, higher order polynomials are usually not recommended particularly because kriging is performed on the residuals after the trend is removed, leaving little variation in the residuals for assessing uncertainty. Also high order polynomials mean a larger number of the b_1 coefficients, which must be estimated along with weights, and a larger set of simultaneous equations to be solved.

Despite a number of studies comparing interpolation techniques, there is no clear consensus to suggest that any one approach is universally optimal. Instead, the performance of the various methods tends to vary depending upon the character of the underlying spatial

variation being modeled, the specific characteristics of the data concerned (e.g., sampling density, sampling distribution), and the cross-validation results attained.

6.3 CROSS-VALIDATION

Cross-validation is a statistical method for validating a predictive model. This technique assesses how the results of a statistical analysis will generalize to an independent data set. Subsets of the data are held out, to be used as validating sets; a model is fit to the remaining data (a training set) and used to predict for the validation set. Averaging the quality of the predictions across the validation sets yields an overall measure of prediction accuracy. Considering the fact that using the same data but different methods can generate different interpolation results, as well as different predicted values can occur by using the same method but different parameter values. For that reason, the validity of the regression models used for estimating the cohort exposure was assessed using three different methods: cross-validation, Jackknife, and bootstrapping. The three superficially similar statistical techniques involve reusing or re-sampling the data. In each case a single sample of observations was considered as many samples with the same estimation process being applied to each of them. However, the purposes of this reuse of the samples were quite different for each method.

For instance, bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, most often with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient or regression coefficient. In the simplest form of bootstrapping, instead of repeatedly analyzing subsets of the data, you repeatedly analyze subsamples of the data. Each subsample is a random sample with

replacement from the full sample. The estimates of standard errors are valid for many commonly used statistics, generally requiring no major assumptions other than simple random sampling and finite variance. The bootstrap standard error is consistent for some non-smooth statistics such as the median. However, the bootstrap standard error may not be consistent even for very smooth statistics when the population distribution has very heavy tails. Thus, theoretical results on the consistency of the bootstrap standard error are not extensive. After some algebra it can be shown that the bootstrap mean and the bootstrap standard error equations take the form of:

$$\bar{\hat{\theta}} = \frac{1}{n} \sum_{i=1}^n \hat{x}_i \quad (6-16)$$

$$V_n(\bar{\hat{\theta}}) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{x}_i - \bar{\hat{\theta}})^2} \quad (6-17)$$

Where n is the size of the subsamples; \hat{x}_i denotes the mean of the subsamples at the i -th re-sampling stage; $\bar{\hat{\theta}}$ represents the mean of the entire n subsamples; and $V_n(\bar{\hat{\theta}})$ indicates the standard error of the mean of the entire subsample. In essence, the original sample represents the population from which it was drawn. So subsamples from the sample represent what can be expected if many samples were taken from the population. The bootstrap distribution of a statistic, based on many subsamples, represents the sampling distribution of the statistic, based on many samples. Nevertheless, the bootstrap standard error is simply the standard deviation of the bootstrap distribution of that statistic.

The jackknife or “leave one out” procedure is a cross-validation technique first developed by Quenouille to estimate the bias of an estimator. John Tukey then expanded the use of the jackknife to include variance estimation and tailored the name of jackknife because like a jackknife, a pocket knife akin to a Swiss army knife and typically used by boy scouts, this

technique can be used as a “quick and dirty” replacement tool for a lot of more sophisticated and specific tools. Jackknifing, which is a more orderly version of the bootstrap, is used in statistical inference to estimate the bias and standard error in a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife estimator lies in systematically re-computing the statistic estimate leaving out one observation at a time from the sample set. From this new set of observations for the statistic an estimate for the bias can be calculated, as well as an estimate for the variance of the statistic.

The jackknife estimation of a parameter is an iterative process. First the parameter is estimated from the whole sample. Then each element is, in turn, dropped from the sample and the parameter of interest is estimated from this smaller sample. This estimation is called a partial estimate and is governed by Equations 6-18 and 6-19.

$$\bar{t}_{n-1} = \frac{1}{n} \sum_{i=1}^n t_{n-1,i} \quad (6-18)$$

$$t'_n = nt_n - (n-1)\bar{t}_{n-1} \quad (6-19)$$

Where: $t_{n-1,i}$ denotes the mean estimation attained for the n -th observations without point i ; t'_n is defined as the pseudo-value estimation of the n -th observation; t_n is the mean estimation obtained from the whole sample; and \bar{t}_{n-1} represents the parameter mean estimation obtained without the n -th observation. A pseudo-value is then computed as the difference between the whole sample estimate and the partial estimate. These pseudo-values reduce the (linear) bias of the partial estimate (because the bias is eliminated by the subtraction between the two estimates). The pseudo-values are then used in lieu of the original values to estimate the parameter of interest and their standard deviation is used to estimate the parameter standard error, as shown in

Equation (6-20), which can then be used for null hypothesis testing and for computing confidence intervals.

$$\hat{V}_j = \frac{n-1}{n} \sum_{i=1}^{n-1} (t_{n-1,i} - \bar{t}_{n-1})^2 \quad (6-20)$$

It is worth noting that jackknife and bootstrap may in many situations yield similar results. But when used to estimate the standard error of a statistic, bootstrap gives slightly different results when repeated on the same data, whereas the jackknife gives exactly the same result each time (assuming the subsets to be removed are the same). Both methods estimate the variability of a statistic from the variability of that statistic between subsamples, rather than from parametric assumptions. The jackknife is a less general technique than the bootstrap, and explores the sample variation differently. However the jackknife is easier to apply to complex sampling schemes, such as multi-stage sampling with varying sampling weights, than the bootstrap.

Otherwise, after completing the jackknife analysis procedure for each known point, diagnostic statistics were calculated to assess the accuracy of the interpolation method. Two common diagnostic statistics derived from the test samples were used to compare interpolation techniques: the root mean square error (RMSE) and the standardized root mean square error (SRMSE), solely for the kriging methods.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_{i,act} - z_{i,est})^2} \quad (6-21)$$

$$SRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(z_{i,act} - z_{i,est})^2}{s^2}} = \frac{RMSE}{s} \quad (6-22)$$

Where n is the number of points; $z_{i,act}$ is the known value of the i -th point; $z_{i,est}$ is the estimated value of point i -th; s^2 is the variance; and s is the standard error.

In conclusion, the goal of cross-validation was to estimate the expected level of fit of a model to a data set that was independent of the data used to train the model. Thus, it was used to estimate any quantitative measure of fit that were appropriate for the data and model. Using cross-validation, the five interpolation methods used were objectively compared in terms of their respective fractions of misclassified characters. In essence the accuracy of the data to which the five models were fit was examined by bootstrapping, Jackknife, and cross-validation.

7 ANNUAL MODEL

7.1 PM₁₀ MEASUREMENTS

Valid samples were collected at eight concurrent monitoring stations and further re-located to cover a area of 651 km² during the three-year monitoring campaign. During the sampling period, 908 ambient PM₁₀ weekly averaged concentrations were measured. Descriptive statistics of the air pollution measurements are shown in Table 7-1.

Table 7-1. Descriptive statistics of the average annual concentrations for the 13 monitoring stations.

Annual Descriptive Statistics of PM ₁₀ Concentrations (µg/m ³)					
Sites	N	Minimum	Maximum	Mean	Std. Deviation
Lindbergh	46	11.4	66.5	27.3	10.5
Fountain	71	5.4	53.9	21.3	11.9
Skyline	49	6.9	50.3	22.8	11.1
McCombs	69	5.6	51.8	18.9	8.4
UTEP	99	8.5	57.1	30.6	9.9
Chamizal	42	11.7	73.1	36.5	15.2
Parral	76	7.4	74.7	26.5	13.9
Cosmos	115	3.4	69.8	24.5	13.1
Socorro	112	9.9	81.2	41.5	15.9
Plant	46	1.8	91.8	35.1	20.9
Mimosa	67	5.7	123.9	37.1	30.4
Montana	84	4.9	54.7	22.8	10.9
Pebble	32	8.5	50.5	24.1	8.7

Weekly averaged concentrations were depurated by removing values above 150 µg/m³ and greater than three standard deviations as well as concentrations underneath it from their corresponding site. The detection limits for PM₁₀ mass concentrations were defined by three standard deviations of the mass concentration deposit on field blanks. The sample measurements averaged 28.4 µg/m³ with a standard deviation of 7.1 µg/m³. From the 908 sampled weekly

concentrations, 123 and 90 observations were at least 1 standard deviation above or below the mean, respectively. All the measured values exhibited a normal distribution as shown in APPENDIX C. The maximum annual average PM_{10} concentration of $41.4 \mu\text{g}/\text{m}^3$ was observed at the Socorro site near the border and a minimum of $18.9 \mu\text{g}/\text{m}^3$ at the McCombs site, which is the farthest north. The overall highest weekly averaged PM_{10} concentration of $123.9 \mu\text{g}/\text{m}^3$ was measured at Mimosa site, also located near the international border in the eastern part of El Paso whilst the overall lowest concentration of $1.8 \mu\text{g}/\text{m}^3$ was sampled at the Plant site. An evident southward trend with the highest values near the border was observed as well as a less marked northeastern and westward trend. In general, the lowest values were observed within less urbanized areas whereas the highest were near industrialized and commercial areas, major roadways and the international border. The average $PM_{2.5}/PM_{10}$ ratio, calculated from the dichotomous samples, averaged 0.26 with a standard deviation of 0.138. The observed $PM_{2.5}/PM_{10}$ ratio agrees with values previously observed in the same region (Li et al., 2001).

Lastly, a spatial autocorrelation analysis was conducted using Morans's I test, using the nearest neighbor approach, to measure the degree of interdependency among the data arranged for modeling intra-urban PM_{10} mass concentration. After conducting such test, the sites produced a value of 0.06 ($p < 0.0005$) suggesting minimal spatial autocorrelation, since positive spatial autocorrelation ranges from -1 to 1 with 1 representing perfect autocorrelation and -1 perfect dispersion. Whilst the autocorrelation levels remain statistically significant at the 0.05 level, the minor level observed would not be expected to have a significant effect on the model results. In fact, previous studies have observed that spatial autocorrelation has minimal effect on the model fit and performance (Hoek et al., 2008; and Beleen et al., 2007).

7.2 REGRESSION ANALYSIS

Each of the 190 independent variables was tested through the methods previously mentioned using the regression analyst tool in SPSS 17 and PASW Statistics 18 to identify variables that were highly correlated with the PM₁₀ observations. Elevation had the highest bivariate correlation coefficient among all variables and was entered into the model under an enter-forward procedure. Subsequently, the remaining variables were entered in a stepwise procedure to determine the addition of variables that would paired with the elevation. This process was repeated until three variables were defined. However, the results were either hampered by high levels of collinearity between the variables or the coefficients were assigned an opposite sign to what logic suggested. Inconsistencies were shown especially with industrial and commercial zoning and traffic variables as they were given a negative sign. This contradicts the assumption that the higher commercial or industrial areas the denser the traffic and the higher the intra-urban gradient of PM₁₀ concentrations. As a result, a variety of models were developed trying to eliminate variable misinterpretation and to generate a reliable correlation with elevation and PM₁₀. Thereafter conducting nearly 20 different models, no acceptable model fit was obtained nor was a concordance between variables and their corresponding coefficient signs. Thus, elevation was discarded from the analysis and the subsequent highly correlated variable was inputted into the model in the same procedure. It is noteworthy that Montana vista (site L) was excluded from the analysis as it was generating inconsistencies in the model. This site was located in a newly developed area in the outskirts of the city and thus lacked spatial data especially for traffic-based and types of land use parameters. This insufficiency of information generated a leverage effect over the fitted regression plane and thus generated low adjusted R²

coefficients. Considered as an outlier, this site was removed and the analysis continued with the remaining 12 monitoring sites.

Consequently, nearest distance to the boarder (OND) was entered into the model and additional variables were paired with other significant variables in a series of trivariate regression models. This series of analysis identified the pairing with the sum of all the roads' length within the 5 km buffer area (NLR5K) of each monitor. Lastly, industrial and commercial areas were comprised into a unique variable (GZ5m) that was associated within a 500 meter buffer area. Table 7-2 provides a summary of the ability of the three variables (OND, NLR5K, and GZ5m) in predict ambient PM₁₀ mass concentrations model (identified in column 1).

Table 7-2. Land use regression model summary of the annual iteration.

Annual Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.963 ^a	.927	.899	2.28

a. Predictors: (Constant), OND, NLR5K, GZ5m

b. Dependent Variable: PM10_Mean

Where the second column describes the multiple R which is often referred to as the multiple correlation coefficients and it is defined as the correlation between the observed PM₁₀ values and the fitted PM₁₀ concentrations from the model. Thus, if R² is high, multiple R is also high, and a strong positive linear association between the observed PM₁₀ concentrations and the fitted PM₁₀ values is presumable. Secondly, the coefficient of determination, R², is the proportion of variation in PM₁₀ mass concentrations (about its mean) explained by a straight-line relationship. Adjusted R² in column 4 is a modification of R² which adjusts for the number of explanatory terms in the model. Therefore, 89.9% of the variability of ambient PM₁₀

concentration in the El Paso area is explained by OND, NLR5K, and GZ5m with a standard error estimate (s) of 2.28 $\mu\text{g}/\text{m}^3$. This means that the model is expected to predict 89.9% of the ambient PM values to within approximately $\pm 4.56 \mu\text{g}/\text{m}^3$ (or 16%) at a 95% confidence with variables like OND, NLR5K, and GZ5m.

Table 7-3 shows the regression coefficients for the regression model. The regression coefficient (B) for every variable is the amount the average of the dependent variable increases when the independent variable increases one unit and the other independents are held constant. Put in another way, the coefficient is the slope of the regression line: the larger the coefficient, the steeper the slopes, the more the dependent changes for each unit change in the independent.

Table 7-3. Statistics of the model's coefficients for the annual iteration.

Annual Model Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	12.743	2.126		5.993	.000		
GZ5m	22.640	5.202	.630	4.352	.002	.437	2.288
NLR5K	2.552	.589	.524	4.333	.003	.626	1.596
OND	.138	.059	.333	2.326	.048	.446	2.240

a. Dependent Variable: PM10_Mean

The Sig. column illustrates that the nearest distance to the border, group zoning (industrial and commercial), and the sum of the roads' length were statistically significant as the prompted values were under the predefined statistical significance of 0.05. Although OND was just slightly under the predefined significant level when paired, it still represented a reliable predictor for the yearly PM₁₀ concentrations, as the confidence interval did not embrace the value of 0 within its upper and lower confidence bound. The ratio of the beta weights for the independent variables

would determine the relative importance of the independents. Since the three variables are significant, 0 was not within its upper and lower confidence bounds. Tolerance and VIF (column 7 and 8) are tests for multicollinearity (overly high correlation among the independents). The reciprocal of the tolerance is known as the Variance Inflation Factor (VIF). The VIF shows how much the variance of the coefficient estimate is being inflated by multicollinearity. A commonly given rule of thumb is that VIFs of 4 or higher may be reason for concern. This is, however, just a rule of thumb; in fact, some scholars suggest a proper concern when the VIF is over 2.5. Thus, according to the results computed in SPSS shown in Table 7-3, no collinearity problems are presumable for this particular model, since OND, NLR5K nor GZ5m surpasses a VIF value of 2.5. Likewise, another form to evaluate any possible multicollinearity between OND, NLR5K, and GZ5m in the model is defined as tolerance. The higher the intercorrelation of the independent variables in the model, the more likely the tolerance value will approach zero. Computed as the reciprocal product of the VIF, tolerance results for this particular model prompted values above the established rule of thumb. This reaffirms the certainty that none of the three variables are intercorrelated and more importantly that the concentrations estimated by the model may produce lower deviations from the actual fitted regression plane.

Table 7-4 shows a more thoroughly assessment of multicollinearity in the model. To simplify, cross products of the independent variables are factored. There will be as many factors as independents, plus one (for the constant). The first column addresses the dimension representing the variables included plus one constant variable. The ordering of the variables corresponds to the same arrangement shown in the variance proportions. That is, the constant variable corresponds to dimension 1, GZ5m to the 2nd dimension and so forth. The subsequent column identifies the eigenvalue which is a term related in the field of linear algebra and

evaluates linear transformations represented as matrices acting on vectors. The eigenvalues shown in the table below for each variable unfolds little variance in the cross product matrix, as high eigenvalues indicate dimensions (factors) which account for much of the variance in the cross product matrix; eigenvalues close to 0.0, such as in this case, indicate dimensions which explain little variance (i.e. collinearity). Values such as 0.34 for OND, 0.109 for NLR5K, and 0.587 for GZ5m obviously underwent the rule of thumb with dimensions above the 0 value.

Table 7-4. Collinearity diagnostics of the model's parameters for the annual iteration.

Annual Model Collinearity Diagnostics ^a							
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	GZ5m	NLR5K	OND
1	1	3.270	1.000	.01	.01	.01	.00
	2	.587	2.359	.00	.28	.09	.00
	3	.109	5.474	.44	.21	.63	.01
	4	.034	9.827	.55	.49	.26	.99

a. Dependent Variable: PM10_Mean

Condition indices, on the other hand, were used to flag excessive collinearity in the data. The condition indices rule of thumb cut-off for tolerable levels of multicollinearity ranges between 15 and 30. A condition index over 30 suggests serious collinearity problems and an index over 15 indicates possible collinearity problems. If a factor (component) has a high condition index, a subsequent test must be performed by shifting the attention toward the variance proportions column. Criteria selected for "sizable proportion" vary among researchers but the most common criterion is if two or more variables have a variance proportion of 0.50 or higher on a factor with a high condition index. If this is the case, these variables have high linear dependence and multicollinearity is a problem, with the effect that small data changes or arithmetic errors may

translate into very large changes or errors in the regression analysis. Note that it is possible for the rule of thumb for condition indices (above 15), to indicate multicollinearity, even when the rules of thumb for tolerance (greater than 0.25) and VIF (lower than 4) suggest no multicollinearity. Computationally, a "singular value" is the square root of an eigenvalue, and "condition indices" are the ratios of the largest singular value to other individual singular values. This confirms that this model has no collinearity problem since any condition index approaches 15 making it unnecessary to further examine the variance proportions.

Lastly, after determining the absence of collinearity, PM_{10} concentrations were determined to be influenced by the nearest distance to the border, the sum of the roads' length, and the group (industrial, and commercial) zoning as shown in Equation 7-1.

$$PM_{10_{concentration}} = 12.743 + 22.640 \cdot (GZ5m) + 2.55 \cdot (NLR5K) + 0.138 \cdot (OND) \quad (7-1)$$

7.3 MODEL VALIDATION

After determining the appropriate model, standard regression diagnostics were applied to assess the problems such as influential outliers, linearity, independency, normality, and homoscedasticity. Standard multiple regression can only accurately estimate the relationship between dependent and independent variables if the relationships are linear in nature. Therefore if violations of linearity are extremely serious (fitting data to a linear model which are nonlinearly related), the predictions are likely to produce systematic errors, especially when generating usually large or small predictions.

A usual way to detect non-linearity data is by plotting the observed values versus the predicted values. Figure 7-1 unveils a correlation coefficient of 0.92, and a symmetrical point distribution around a diagonal line determining a linearity relationship between PM_{10} mass

concentrations and OND, NLR5K, and GZ5m. Note that, the standardized predicted values were calculated using the model equation for every particular monitoring site previously discussed. In addition, the statistical software suited for this research produces such values in a residual statistics table.

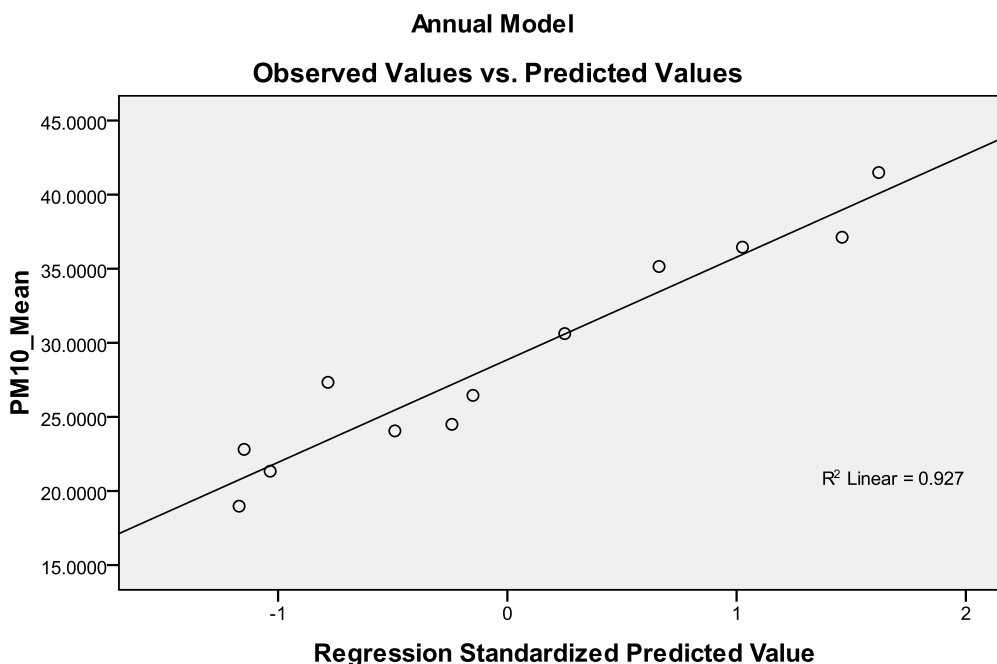


Figure 7-1. Linearity test between the sample concentrations and the predicted values for the annual model.

Residuals are defined as the difference between the observed values and those predicted by the regression equation. Serial autocorrelation is sometimes a byproduct of a violation of the linearity assumption. Residuals thus represent error as in most statistical procedures and were used for three main purposes: (1) to spot heteroscedasticity (increasing error as the observed predicted value increases), (2) to spot outliers (influential cases), and (3) to identify dependency (autocorrealton).

In addition to the Moran's I test, performed previously, a second evaluation was performed under the Durbin-Watson test. This test lays its foundation under the hypothesis of

spatial autocorrelation. The null hypothesis (H_0) states a positive autocorrelation, $d < d_{\text{lower critical value}}$, whereas the alternative hypothesis (H_1) declares a negative correlation among the errors $d > d_{\text{upper critical value}}$. Nevertheless if the d value is greater than $d_{\text{lower critical value}}$ but less than $d_{\text{upper critical value}}$ the test is inconclusive. Thus using the test statistics shown in Equation 7-2, a corresponding d value of 1.802 was achieved.

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (7-2)$$

Where e_i represent the residual between the measured and predicted concentration at point i ; and n is the total number of samples. However, obtaining the $d_{\text{lower critical value}}$ (0.658) and the $d_{\text{upper critical value}}$ (1.864) at 5% significance from APPENDIX E, the test proved to be inconclusive and thus none of the alternatives could be neither accepted nor rejected. Nonetheless, researchers have pointed out that Durbin-Watson values above 1.5 tend to lay in a minimal to non-autocorrelation frame. Additionally the previous Moran's I test proved to have no autocorrelation among the errors.

Table 7-5 contains summary data regarding the residuals (the difference between predicted and actual values). Standard residual (Std. residual), for instance, is the raw residual divided by the standard deviation of residuals. Since the minimum standardized residual is -1.177, at least one prediction is more than one standard deviation below the mean residual. Studentized residuals are very similar to standardized residuals but follow the t distribution. These are used in plots of standardized or studentized predicted values vs. observed values. Deleted residuals, also called "jackknife residuals," compute the standard deviation omitting the given observation prior to standardizing or studentizing the residual. The "deleted residual" row has to do with coefficients when the model is recomputed over and over, dropping one case from the analysis each time.

Table 7-5. Descriptive statistics of the annual model's residuals.

Annual Model Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	20.753918	40.075146	28.857300	6.9254975	12
Std. Predicted Value	-1.170	1.620	.000	1.000	12
Standard Error of Predicted Value	.817	1.975	1.272	.361	12
Adjusted Predicted Value	20.247725	41.281567	28.945593	6.8485819	12
Residual	-2.6868327	3.8951077	.0000000	1.9473947	12
Std. Residual	-1.177	1.706	.000	.853	12
Stud. Residual	-1.292	1.855	-.018	1.019	12
Deleted Residual	-4.1585660	4.6059899	-.0882930	2.8839753	12
Stud. Deleted Residual	-1.359	2.298	.008	1.105	12
Mahal. Distance	.493	7.311	2.750	2.106	12
Cook's Distance	.000	.461	.128	.132	12
Centered Leverage Value	.045	.665	.250	.191	12

a. Dependent Variable: PM10_Mean

The bottom three rows in Table 7-5 are measures of the influence of the minimum, maximum, and mean case of residuals on the model. The Mahalanobis distance takes into account the covariance among the variables in calculating distances. With this measure, the problems of scale and correlation inherent in the Euclidean distance are no longer an issue. To understand how this works, consider that when using Euclidean distance, the set of points equidistant from a given location is a sphere. The Mahalanobis distance stretches this sphere to correct for the respective scales of the different variables and to account for correlation among variables. The rule of thumb suggests that the maximum Mahalanobis distance should not exceed the critical chi-squared value with degrees of freedom equal to number of predictors and a predefined alpha value. Consequently, using three degrees of freedom and an alpha value of 0.001, the threshold

value was determined to be 16.27. The minimum, maximum, and mean Mahalanobis distances displayed in the Table 7-5 denote much lower distances, disregarding any possibility of outliers influencing the development of the linear regression model.

Cook's distance (D) is another measure of the influence of a case. Cook's distance measures the effect on the residuals for all other observations by deleting a given observation. Observations with larger D values than the rest of the data are those which have unusual influence or leverage. Fox (1991: page 34) suggests, as a cut-off for detecting influential cases, values of D greater than $4/(N - k - 1)$, where N is sample size and k is the number of independents. Others suggest using D greater than 1 as the criterion to constitute a strong indication of an outlier problem. In SPSS, the minimum, maximum, and mean Cook's D is displayed in Table 7-5. Using the cut-off suggested by Fox (1991: page 34), a computed threshold value of 0.50 was obtained, thus no apparent problem cases arose since the maximum Cook's D is only 0.46.

The centered leverage value row represents the value of a potential observation to have undue influence on the regression model. That is, a sample observation with low value paired with a high value of sample 2 (or vice versa) may create an exaggerated influence on the regression model. Cases with leverage values less than 0.2 are not a problem, but cases with leverage values of 0.5 or higher may be unduly influential in the model and should be examined. However, for small samples, Vellman and Welsch (1998) suggested that $3*(k/N)$ is the criterion, where k is the number of variables and N is the number of samples. In contrast, the maximum leverage value of 0.665, identified at Chamizal site (H), shown in Table 7-5 was smaller than the computed threshold from the small sample equation of 0.75, thus discarding the possibility of any undue influence from the OND, NLR5K, and GZ5m on the yearly model.

Errors, represented by the residuals, should be normally distributed for each set of values of the independents. Conversely, violations of normality compromise the estimation of coefficients and the calculation of confidence intervals. Sometimes the error distribution is "skewed" by the presence of a few large outliers. Since parameter estimation is based on the minimization of squared error, a few extreme observations can exert a disproportionate influence on parameter estimates. Calculation of confidence intervals and various significance tests for coefficients are all based on the assumptions of normally distributed errors. If the error distribution is significantly non-normal, confidence intervals may be too wide or too narrow. In Figure 7-2, residuals followed a bimodal distribution, based on predicting PM₁₀ mass concentrations from OND, NLR5K and GZ5m. However, the residuals exhibit by the bimodal shape corresponds to the PM₁₀ concentrations themselves.

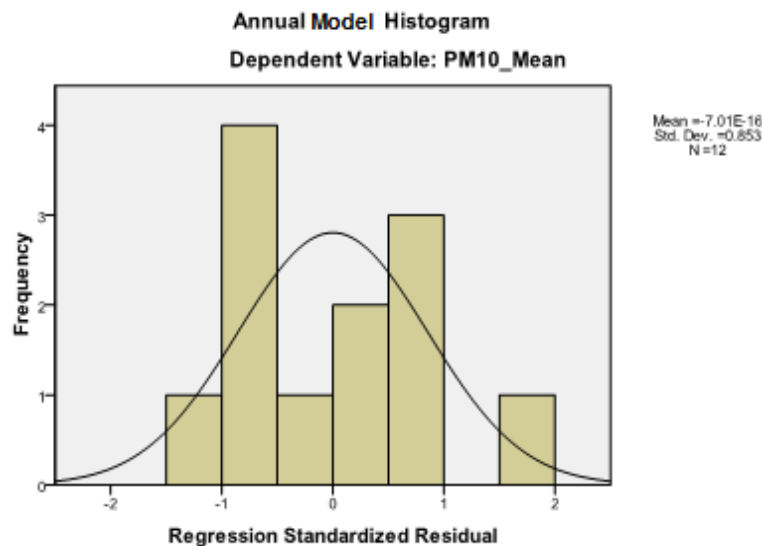


Figure 7-2. Residual's distribution of the annual model.

In addition, the Shapiro-Wilk's (W) test was performed over the model's residuals. This is the standard test for normal distribution and returns a finding of non-significance if residual error is normally distributed. This means that W is the correlation between given data and their

corresponding normal scores, with W equals 1 when the given data are perfectly normal in distribution. If W is significantly smaller than 1, the assumption of normality is not met. In this case, for the given variables, W test was not significant suggesting that the variable's distribution is not significantly different from normal, as shown in Table 7-6.

Table 7-6. Shapiro-Wilk's W test for normality for the annual model's residuals.

Annual Model Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.175	12	.200*	.999	12	.966

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Therefore, Table 7-6 contradicts the residual histogram showing that OND, NLR5K, and GZ5m residuals were normally distributed and that the bimodal shape merely corresponds to the variance in measured concentrations.

The homoscedasticity test, often referred as homoskedasticity, assures that the residuals are dispersed randomly throughout the range of the estimated dependent. Put in another way, the variance of residual error should be constant for all values of the independents. The lack of homoscedasticity may mean that: (1) there is an interaction effect between a measured independent variable and an unmeasured independent variable excluded from the model; or (2) that some independent variables are skewed while others are not. Note that regression analysis is relatively robust regarding homoscedasticity and small to moderate violations of homoscedasticity have only minor impact on regression estimates. Violations, however, make it difficult to gauge the true standard deviation of the forecast errors, usually resulting in confidence intervals that are too wide or too narrow. In particular, if the variance of the errors is

increasing over time, confidence intervals for out-of-sample predictions will tend to be unrealistically narrow.

Figure 7-3, illustrates a disperse pattern of the points with no trend. The distribution of the residuals among the standardized residuals axis demonstrated some ranges of errors below the negative standardized residual axis, while others ranged positive above the 0 value of the y-axis. Moreover, no sign of any outliers was spotted that may be causing this nonconstant error variance, and obviously no error that may be influencing positively or negatively the adjusted R^2 of the model. The scatter plot, however, reflects the fact that the model in this case has low standard errors of estimate explaining a small percentage of variance mainly as a consequence of an adequately fitted model.

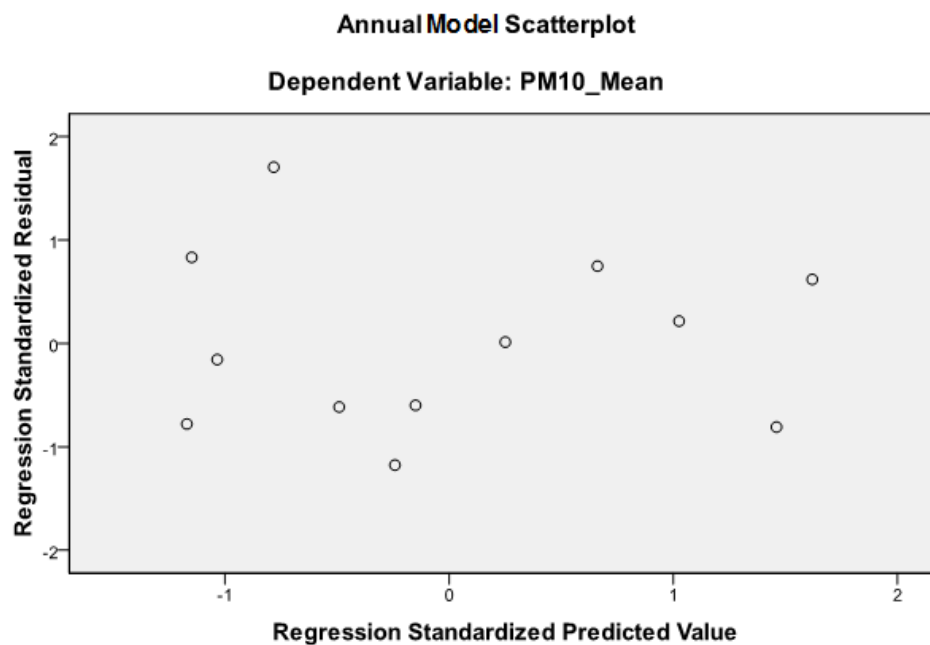


Figure 7-3. Scatter plot of the standardized residuals versus the standardized predicted values for the annual model.

Therefore, the scatterplot of the standardized residuals versus standardized predicted values (Figure 7-3) substantiated the normality (performed previously) and homoscedasticity tests when testing the significance of the model and its parameters. The normality assumption was affirmed as the density of the plotted symbols were denser near the 0 value of the standardized residual axis (y-axis) and quickly abate away from the 0 value, keeping a symmetrical distribution on both sides of the 0 value of the. Likewise, homoscedasticity was attested as the spread of the error values, in the vertical dimension, were the same at any one point on a line as it is at any other point on that line. Nevertheless, the residual plot patterns confirmed the relationship between predictor and criterion as linear.

Heteroscedasticity, on the other hand, may also have the effect of giving too much weight to small subset of the data (namely the subset where the error variance was largest) when estimating coefficients. One possible cause of this might be a skewed rather than normally distributed dependent variable. When homoscedasticity is violated, a simple scatterplot of the independent (x) on the x-axis and the dependent (y) on the y-axis will typically display a "funnel shape," as in the figure below.

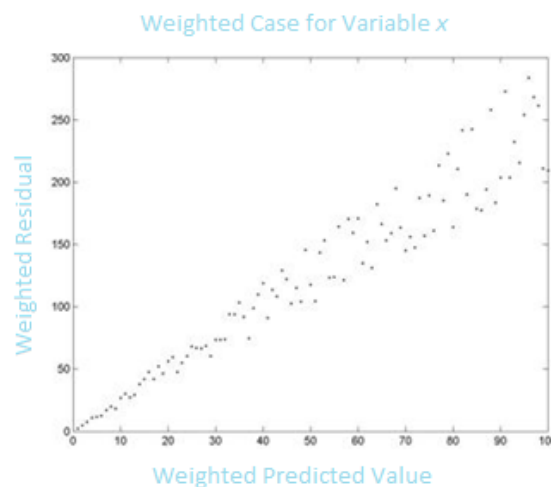


Figure 7-4. The funnel shape pattern attributable to the violation of homoscedasticity. Source (Wikipedia ©2009).

Weighted least squares (WLS) regression causes events with smaller residuals to be weighted more thoroughly in calculating the x coefficients in order to diminish the impact of heteroscedasticity. In the figures below (Figure 7-5, Figure 7-6, and Figure 7-7), cases were weighted by the reciprocal of their estimated point variance, and each assesses the goodness of fit of the weighted model. That is, weighted predicted values are plotted on the x -axis and weighted residual on the y -axis. All three weighted residual plots produced a good fit; the residuals did not form a funnel shape but instead a uniformly distributed pattern around the 0 value of the y axis. Comparing Figure 7-4 with each variable weighted case (Figures 7-5 through 7-7), it can be seen that there is no funnel shape pattern because of the reduction in heteroscedasticity.

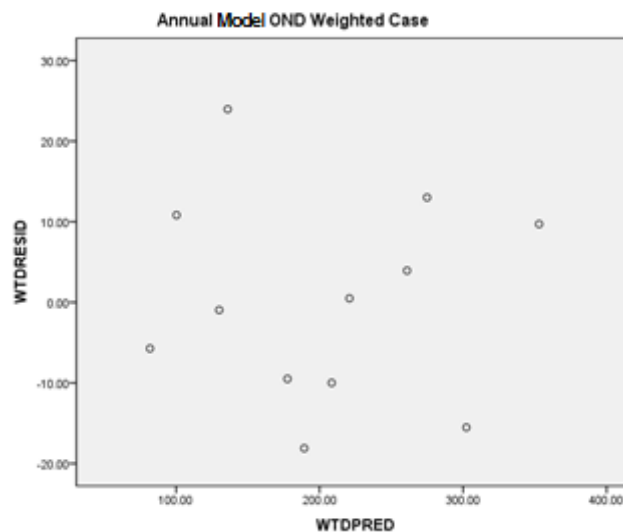


Figure 7-5. Weighted case of OND for the annual model.

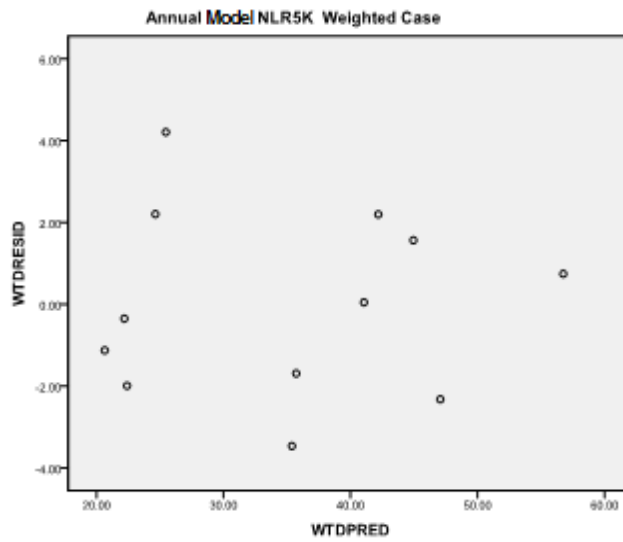


Figure 7-6. Weighted case of NLR5K for the annual model.

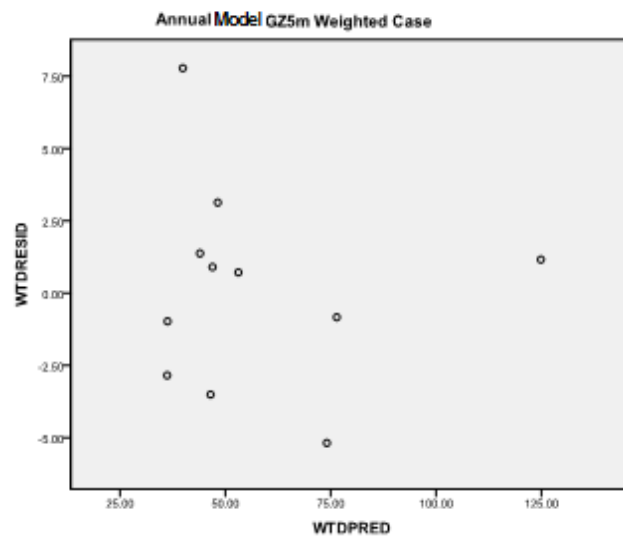


Figure 7-7. Weighted case of GZ5m for the annual model.

7.4 CROSS-VALIDATION

Comparison and assessment of the different methods of air pollution mapping undertaken here was done primarily by comparing their performance at the validation sites using standard measures of model performance. To set these measures in context, they were compared also to the levels commonly quoted in validation studies of formal land use regression models. The

validity of each interpolation model was assessed by comparing the predicted levels of the modeled pollutant to the measured value levels at each grid node included in the LUR model by means of the jackknife (leave-one-out) procedure, a bootstrap to evaluate the sensitivity of the model variance, and the root mean square errors.

First, the bootstrap analysis was undertaken to evaluate the sensitivity of the estimated model variance in the five models. In this interpolation technique, the 3528 predicted values were divided into 147 subsamples of 24 random samples. The variance was calculated as the square root of the sum of the squared differences between the subsample i -th average and the overall mean divided by the number of representatives minus one (3527). Supplementary to this method, histograms of the predicted parameter were validated and derived over 1000 iterations to examine the consistency of the standard variance as presented in APPENDIX F.

Secondly, we validated each model by a comparison of prediction errors obtained by simulating 1,000 jackknife data sets. For each data set, parameter estimates estimated from the interpolated model were obtained on a subset of 3528 locations, where each model was repeatedly parameterized on $N-1$ (3527) data points and then used to predict the excluded locations to predict PM_{10} .

Thirdly, the root mean square errors and the standardized root mean square errors were calculated as the square root of the sum of the squared differences of the observed concentration at site i and the predicted concentration at site i divided by the total grid nodes (3528). In each of the cross-validations, the average difference was not greater than $2.61 \mu\text{g}/\text{m}^3$ for bootstrap, translating into an average relative difference of less than 4.0%. Also, there was a recurrent estimate variance value of $0.01 \mu\text{g}/\text{m}^3$ for the jackknife method, resulting in an error of 15%, and $6.7 \mu\text{g}/\text{m}^3$ for RMSE, which intuitively contrasts the measured values with the predicted values

within an 8% difference. The results of this comparison are presented in Table 7-7.

Table 7-7 Cross-validation results of the five interpolation techniques used to map annual PM₁₀ mass concentrations throughout the El Paso cohort study.

	Annual Model			
	Bootstrap Variance	Jackknife Variance	RMSE	SRMSE
IDW	2.494122	0.00859	6.633	N/A
NN	2.496435	0.00919	6.651	N/A
OK	2.379054	0.00744	6.509	0.865
SPLINE	2.609107	0.01061	6.691	N/A
UK	2.306198	0.00734	6.494	0.953

Amongst the modeling techniques, the predicted concentrations generated by the SPLINE method had the worst performance, with a bootstrap variance of 2.61 $\mu\text{g}/\text{m}^3$, a variation in prediction certainty over 1000 iterations of 0.011 $\mu\text{g}/\text{m}^3$, and a root mean square error of 6.69 $\mu\text{g}/\text{m}^3$. Similarly, the bootstrap histogram demonstrated a presumptuous bimodal shape, as seen in APPENDIX F. In this case, the alteration from normality and the deplorable prediction performance was caused by the fundamentals of the model as it created the surface specifically so that it passes through the control points having the least possible change in slope at all points, fitting the control points with a minimum curvature surface. In other words, the model tends to average to its utmost capacity. Thus, coinciding with the visual pattern (see APPENDIX H) were the incongruous blemishes, of higher concentration at lower polluted areas, broadens more frequently than in the rest of the interpolation techniques. Considering of course, the fact that, the annual PM₁₀ surface tends to have a more moderate disrupted variation pattern, when compared to any other season as the 3-year continuous temporal run of the monitoring data stabilized the monitor readings contributing to the proportion of explained variance.

NN likewise also performed badly, as anticipated from the visual inspection, with a bootstrap variance of 2.496 $\mu\text{g}/\text{m}^3$, a RMSE of 6.65 $\mu\text{g}/\text{m}^3$, and a fractional bias obtained from the jackknife analysis of 0.0091 $\mu\text{g}/\text{m}^3$. Although the bootstrap histogram followed a normal

distribution with no bimodal shape whatsoever, still the model tended to misrepresent the spatial variability of PM_{10} worse than IDW and kriging, mainly because the NN method assigns no data values to the cell center of the perimeter cells of the output raster that fall outside the convex hull (i.e. convex polygon, typically represented by a sequence of its vertices ordered along its boundary), which is defined by the outermost extent of the points. Additionally, influential points were disproportionally weighted over less influential points, especially in the north- and far-east areas, attributable to the Voronoi's intersecting triangles that cause more variance between the measured values and the predicted values and thus the poorly performance of the natural neighbor method. Sequentially, IDW performed better than the latter two methods described above, with a bootstrap difference of $2.49 \mu\text{g}/\text{m}^3$, a fractional bias of $0.008 \mu\text{g}/\text{m}^3$, and a RMSE equal to $6.63 \mu\text{g}/\text{m}^3$ and far less attenuation. It is important to remark that NN and IDW are very similar methods, mainly because they both rely on the weight distance average. However, whereas NN assigns weight based on the percentage of overlap, IDW assigns weights based on their similar distance from the interpolation point. This could be ascertained as both NN and IDW concurred in their visual pattern, shown APPENDIX H, not to mention their similarities in their cross-validation results to which they were put to attest. And although IDW outperformed NN, the model still had some flaws with respect to its ability to predict accurate PM concentrations. A reason for that is attributable to the fact that the IDW average weighted distance cannot be greater than the highest or less than the lowest input values, contradicting or misrepresenting the elevated concentrations measured at the vicinity of the BOTA as well as any intuitively trend. In fact, the bootstrap histogram shows a modest bimodal shape traceable to a single monitoring station located in Chamizal Park, just one mile from the BOTA and extremely close to CDJ and Highway 375. This point had the third highest mean concentration of all

samples. However, there is no reason to believe that the underlying data from this station is inaccurate, as it prompted an 86% correlation when compared to TCEQ's station CAMS40, and therefore the monitor was not excluded.

Lastly, the best performance was seen for both ordinary and universal kriging approach for which the bootstrap variance equals to $2.38 \mu\text{g}/\text{m}^3$ and $2.31 \mu\text{g}/\text{m}^3$ respectively; the jackknife variance for ordinary kriging prompted a value of $0.0074 \mu\text{g}/\text{m}^3$ as well as $0.0073 \mu\text{g}/\text{m}^3$ for universal kriging. For the ELP using 3-yr data the RMSE for ordinary kriging (OK) and universal kriging (UK) resulted in an absolute value of $6.51 \mu\text{g}/\text{m}^3$ and $6.49 \mu\text{g}/\text{m}^3$, respectively; by contrast, the SRMSE on these predictions was $0.865 \mu\text{g}/\text{m}^3$ and $0.953 \mu\text{g}/\text{m}^3$. Nevertheless, the histogram of the predicted estimates derived for each of the 1000 iterations suggested in both studies that the models were robust. However, ordinary kriging showed a slight bimodal shape corresponding to the single predicted location at Bassett Mall and its vicinity, which accounts for a vast amount of commercial area but is somehow less representative in the surface pattern than the universal kriging map; as presented in APPENDIX F. For most metrics, the models derived from the two runs are broadly similar, indicating considerable robustness in the methods in the face of varying input data. There is some tendency, however, for the performance of the ordinary kriging model to be rather worse than the universal kriging, indicating the sensitivity of the methods to the underlying input data. Kriging gives the preferred cross-validation values at the grid nodes, but these generally reflect the way in which the kriged surfaces are fitted to those points. Semivariograms (shown in APPENDIX G) show strong spatial dependency in the data, with low nugget variance compared to the mean concentrations and an indistinct sill (the semivariance value at which the variogram levels off), and thus suggest that the models predicted concentrations well at unsampled locations. Therefore the overall cross-validation diagnostics

concludes that kriging methods demonstrated that the influential points are unlikely to drive the regression model and that universal kriging feebly outperforms ordinary kriging in predicting spatial characterization. All because ordinary kriging assumed that the variation in PM_{10} is free of any structural component. The unusual values in length of road, for example, inflated and deflated the model prediction power when compared to the observations. Ordinary kriging, as a result, oversmooths and hence misses areas with unusually high PM concentrations caused by omitting an external drift. It is also possible that ordinary kriging in this context violates the fundamental assumption of stationarity. Points close to the freeway, for example, exhibited a different relationship between PM_{10} and distance rather than points, for example, in less developed areas far from any major source of PM_{10} . These results indicate that the difference in temporal variation produced relatively moderate errors when comparing measured values to modeled values.

7.5 DISCUSSION

The developed regression model was able to predict PM_{10} mass concentrations for the annual period with high degree of certainty. Geographic variables explained a greater portion of the variability of PM_{10} concentrations. This indicates that the geographic variables used are more closely related to annual PM_{10} measurements. The model was constructed using 908 weekly concentrations from field monitoring sites in the El Paso area. However, the Montanavista site had to be withdrawn from the analysis as it affected the model's predictability with its inclusion or exclusion generating inconsistent associations within the parameters and inverse correlations. The lack of spatial information and its location as an outlier outside the urban sprawl were the main reasons for its withdrawal.

The buffers used in the analysis to measure the parameters were quite large when compared to other studies (Brauer et al., 2003; Hochadel et al., 2006; and Ross et al., 2006). The El Paso study area is a massive and sprawling urban area. This dispersed urban structure, transected by one major highway, commercial areas, and within proximity to Ciudad Juarez leads to a broader regional scale of influence for processes generating PM pollution; the 1500-5000 m buffers surrounding the PM₁₀ monitors reflects this dispersed form of urban development. The calculation of 2, 3, and 5 km-buffers were not quite precise, since they exceeded the boundaries of south El Paso in 5 out of the 13 monitoring sites and had on average 11, 21, and 31%, respectively, of their area in Ciudad Juarez, where no spatial information was procured. Parameters under such buffer areas at those particular sites were given a weighted value depending on their corresponding radii within the ELP boundary, altering utmost the value of the model R² by 0.005%. Thus, the predictability of the annual model for El Paso was not hampered by this problem and did not affect the main findings of the analyses.

On the other hand, predictors based on relatively small buffers also performed superbly when modeling highly concentrated samples. As demonstrated through European studies (Briggs et al., in press.), PM₁₀ varies gradually over space and as such the strength of the small buffers may reflect this small area variation in the pollutant.

Aside from the nearest distance to the border of the sampling location was the most consistent predictor of PM₁₀ levels, accounting for over 84% of the variation. When combining this variable with the sum of the roads' length in a 5 km buffer and the group zoning within 500 meter buffer, the model was able to explain almost 90% of the spatial variation of PM₁₀ in El Paso, Texas. The nearest distance to the border was an influential parameter in the model as the proximity to Juarez has often aroused the speculation of a source for air pollution. The high

industrialized areas, the closeness of brick kilns, unpaved roads, and the traffic activity of antediluvian vehicles along the border in the vicinity city of Juarez has proxy to be a contributor of PM₁₀ concentrations for the El Paso region. Likewise, the sum of the lengths of all roads within a 5 kilometer buffer was included in the model as a traffic-related variable and was able to assist in the significant prediction model. The predictability of the sum of all roads' length is not only due to the actual traffic emissions but also to its association with many other geographic characteristics, for example the population density or the presence of commercial and industrial sources, which notably influence the concentration levels as well. Predictors directly based on traffic counts were of primary interest in the analysis. VMTs per buffer were secondly preferred to maximum traffic intensities, because they seem to reflect the emissions and dispersion process more closely. These variables usually showed lower correlations with the pollutant concentrations than other variables of comparable scale. In general, the sum of the roads' length describing traffic intensity appeared to have greater explanatory power than those describing distance to the nearest road, traffic counts and even VMTs. Furthermore, an additional attempt was made to explain more variability of the pollutant by comparing both traffic metrics in a single model, but only marginal increases in the R² values were obtained. All these results suggested that model built with road length and vehicle density metrics are equally able to explain small-scale variability in pollutant concentrations. This finding confirmed that valuable LUR model can be developed in the absence of traffic metrics, which are unavailable, none existing, or faulty representative in many areas. For example, Briggs et al. (2005) used road type data for their Amsterdam model, but could not speculate on whether better results were achieved with traffic density information. This predictability of traffic related variable corresponds to previous studies which suggested spatial gradients of vehicle emissions in the Paso del Norte

region. Temporal variations in PM_{10} have also been shown to coincide with the morning and evening commute hours at monitoring sites near emission source regions of central El Paso (Jeon et al., 2001; and Noble et al., 2003). Albeit, local traffic-related emissions are only significant closer to the roads. A previous study in the region formulated a LUR model to predict the intra-urban gradient of NO_2 . They cited that the inclusion of the proximity to a port of entry and the addition of distance to a major highway and elevation to the model explained 81% of the variation of the NO_2 gradient for El Paso (Gonzales et al., 2005). Therefore, this model corroborated the findings of Gonzales et al. (2005) that the proximity to Juarez and traffic-based parameters proxy higher concentrations in the cohort study and that the intra-urban gradient increases as the distance to CDJ decreases. Although elevation was not a contributor in the developed model, it was substantiated that elevation has influence over the concentration of PM_{10} as it showed the highest bivariate correlation coefficient. However, this surrogate parameter was not included in the model as it produced inconsistent coefficients when a second or third variable was added to the model as well as spatial autocorrelation among the subsequently included variables. In the third iteration of developing a city-specific LUR model, commercial and industrial areas within 750 m and 1000 m respectively were a strongly predictor in the model. The grouping of the two variables into a single parameter was done in order to contain the model within three variables and thus leave span for other variables to be recognized by the model as surrogates for PM_{10} concentrations rather than concentrating the model into distance to the border and two independent land use variables. Therefore, this newly developed variable (GZ5m) prompted an R^2 of 0.673 when correlated to PM_{10} mass concentrations and showed strong association with background concentrations when included into the model in

conjunction with distance to the border and sum of roads length. Industrial and commercial densities served as a surrogate for the general level of human activity in the vicinity of a monitoring site. These densities have been associated with decreased driving speeds and increase emissions, suggesting that higher density areas may be subject to higher emissions by human activity. Also, the segregation of these land use types in the region creates the need for travel between residential, commercial, and industrial zones. Taken together, these characteristics increase the exposure surface with more spatial variability in ELP urban area.

Additionally, testing for non-linearity, spatial autocorrelation, collinearity, homoscedasticity and normality suggest no violation whatsoever as shown in the previous section. A bootstrap to evaluate the sensitivity of the parameters to the actual samples included in the modeling suggests that the model is robust and relatively stable to the choice of samples. In general, these variables were robust to sample selection, though it was found that group zoning and length of roads variables were affected by the inclusion or exclusion of the Montanavista sample that had modest PM_{10} concentrations. Nevertheless, the fact that group zoning (GZ 500 m) and length of roads (NLR 5 km) in the El Paso city borough, using an entirely different GIS layer and without the influence of the Montanavista sample, was significant lends support to the important role played by group zoning and length of roads for this particular model.

Consequently, after deriving the parsimonious operational regression model a grid covering the cohort study and containing 3528 nodes was developed in order to increase the resolution of the depicted overall PM_{10} trend. At each node the concentration of PM_{10} was calculated applying the three parameters with their corresponding coefficient and joined into a single attribute table. It is important to remark that the measured value located in Fairbanks Avenue and Rushing road at the Northeast area had to be excluded from the analysis as the LUR

model prompted an elevated PM₁₀ concentration much higher than 120% of the second maximum measured value. The truncated LUR predicted value corresponded to a miscalculation by the analysis tool in ArcGIS of the intersected road length's for the 5 km buffer. Further analysis was performed to the surrounding eight nodes in order to detect any more flaws but despite the rigorous authentication, the points gave no further reason to believe the roads length value was confounded and thus the outlier value was excluded from the analysis.

Given the strong predictive power, methods (IDW, NN, OK, UK, AND SPLINE) were combined with LUR to generate a confident PM₁₀ spatial surface. The validity of three models was challenged by three procedures (bootstrap, jackknife, and RMSE) whereas the remaining two models were assessed by an additional method (SRMSE). The bootstrap cross validation procedure for UK yielded a variance over the subsample average of 2.306 $\mu\text{g}/\text{m}^3$ in contrast to the second lowest variance of 2.379 $\mu\text{g}/\text{m}^3$ prompted by the OK method. The fractional bias of 0.00734 $\mu\text{g}/\text{m}^3$ for UK was smaller when compared to the second lowest value of 0.00744 $\mu\text{g}/\text{m}^3$ prompted by the OK method. The predictability of UK was more precise with a RMSE of 6.494 $\mu\text{g}/\text{m}^3$ compared to 6.509 $\mu\text{g}/\text{m}^3$ for the ordinary kriging. Lastly, the SRMSE value for UK was closer to one than the OK meaning that the estimated standard error was more reliable or valid for the uncertainty of the predicted values. Overall, universal kriging based on Gaussian model outperformed IDW, NN, SPLINE, and OK on cross-validation results using the full set of samples. The cross-validation procedure, for example, yielded a bootstrap and jackknife variance, which was small compared to the range of concentrations and not greater than the RMSE, indicating that the influence of the single observations was not too heterogeneous. In addition, 90% of the UK modeled estimates were within a factor of 1.2 times the range of the measured concentration with an inter quartile range (IQR) value of 7.683 $\mu\text{g}/\text{m}^3$, approximately

6% less than the second lowest value cued by OK and 12% by the SPLINE method. The ELP universal kriging model exhibited a bias less evident, but still existing, in the full model that included the validation samples overpredicting 31% of the value locations and underpredicting 29% with a relatively low average absolute value residual or mean absolute error (MAE) of $2.541 \mu\text{g}/\text{m}^3$. The underprediction manifested in the more urbanized area of the El Paso city which is in the proximity of the US-Mexico border, specifically in the area extending from downtown to the southeastern part (Socorro) of the cohort study. Albeit, the overprediction appeared in areas distant from the centralized urbanized sprawl embracing sectors from the Fareast, the Northeast, and to a lesser extent, the upper westerly area of El Paso. All of the observed overpredictions were upwards of $40 \mu\text{g}/\text{m}^3$. Due to unregulated fuel combustion, vehicles are one of the leading sources of particulate air pollution in the ELP basin. Surrounding the ports of entry, especially the BOTA and Downtown Bridge, is a network of roads connected to the freeway use to disperse El Pasoan's and daily commuter's vehicles from residential to industrial and commercial areas. Westerdahl et al. (2005) found that measured freeway concentrations of PM had a range between 60 to $820 \mu\text{g}/\text{m}^3$ on the freeway, and that concentrations along major roadways with high traffic density were up to 20 times higher than residential concentrations. Thus, the overpredicted values in this analysis were at plausible levels. Categorizing the sites by tertiles, 75% of them would be classified identically. Only 25% of the grid nodes near the south boundaries of the El Paso area would change between the extreme tertiles. The divergences might be partially explained by the spatial distribution of PM_{10} which is influence by CDJ. When the predicted exposure was divided by the predicted quartiles it was estimated that 63% of El Paso residents fall into a range of 25 to $35 \mu\text{g}/\text{m}^3$ while 4% of which fall into highest categories for PM_{10} concentrations between 35 to $45 \mu\text{g}/\text{m}^3$.

Finally, in this study LUR in combination with universal kriging successfully predicted small-scale spatial variation of ambient PM_{10} concentrations with no spatial irregularities and adequately reproduced the regional trend identified by the monitoring network. The model predicted lower concentration values within less commercial and industrial development in the Northeast, Far east and West of the city. That is a less marked western and northeastern trend whereas a steeper concentration gradient southward, towards the international border with CDJ. While higher concentrations were mapped near freeways and shallow elevation areas closest in proximity to the border. Although Ciudad Juarez was shown to influence the spatial PM_{10} distribution in El Paso it was unpractical to monitor the entire metropolitan area due to cost and political constraints in order to quantify the impact of CDJ.

As shown in Figure 7-8, the model identifies three noticeable hot spots; first and most pronounced hot spot was located in the surroundings of the downtown area extending to the BOTA port of entry, for which the area accounts for 98% of the buffer zone to group zoning especially commercial land use and 6% lower than the highest sum of length roads seen in a buffer area. This generally spurred concentrations between 40 to 42 $\mu\text{g}/\text{m}^3$ for the downtown area. The Second high concentration spot arose between Interstate 10 and Montana Avenue around Bassett Center Mall which embraces commercial and residential land in its majority; representing 83% of the area of the buffer radii. Likewise the proportionality of the length of roads in this area was 16% lower than the highest value seen in a buffer of 4.81 miles generating concentrations between 39.68 and 41.23 $\mu\text{g}/\text{m}^3$. Thirdly, the Western Refining and circumvent areas prompted a high concentration blemish with values ranging between 36.26 and 40.44 $\mu\text{g}/\text{m}^3$, attributable mainly to the industrialized and commercial areas within the buffer radii representing 96% of the whole circle itself. Nevertheless, the proportional ratio of the sum of

road lengths was the lowest when compared to the latter two hot spots with 30% lower than the highest value computed within buffer radii.

The outcome bearded the long-term spatial variations of averaged PM_{10} concentrations for a localized area within larger metropolitan region and predicted a gradient decreasing with distance from the border. The methodology captured the relative impact of emissions outside the localized monitoring network area and allowed for an optimal utilization of resources.

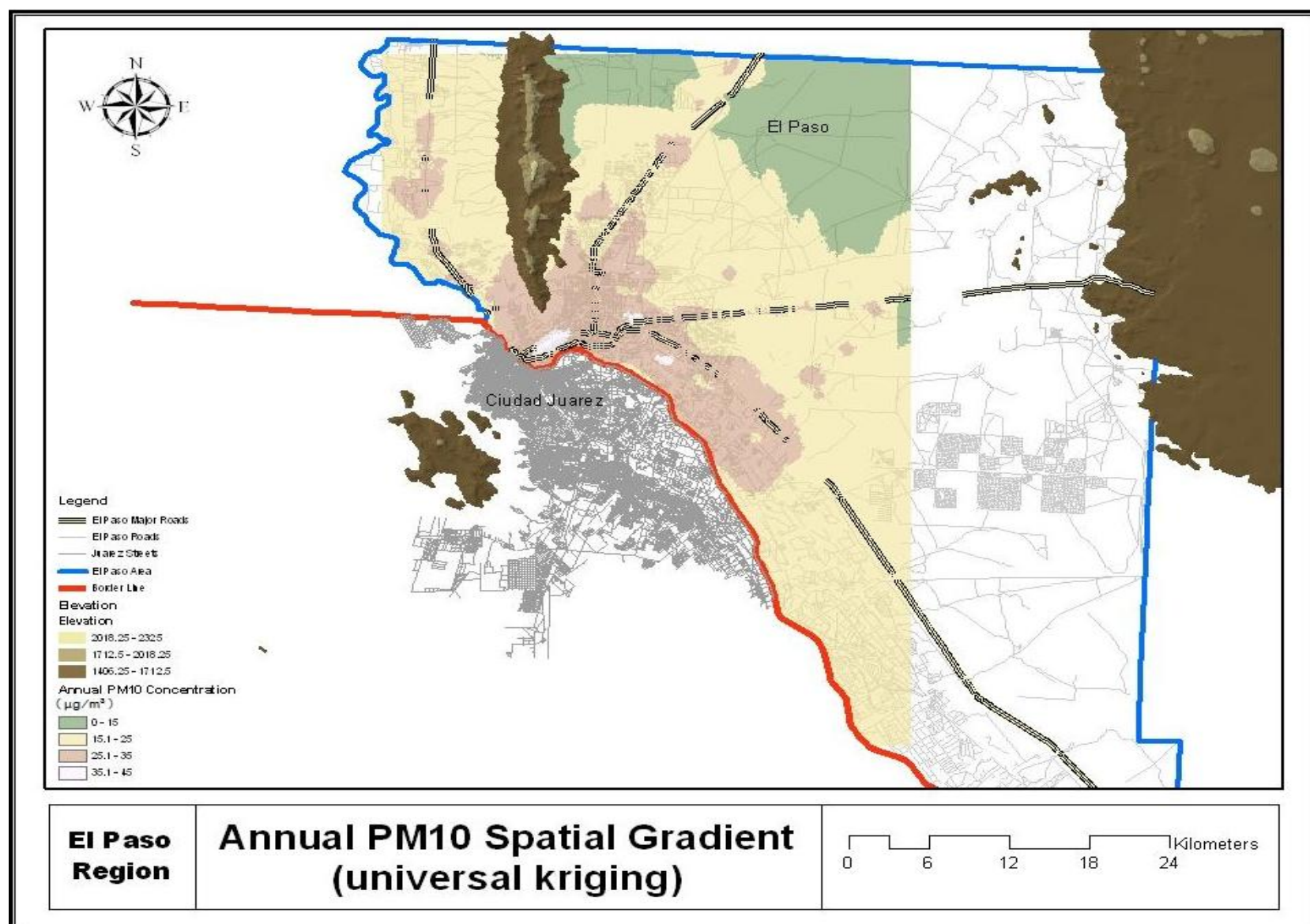


Figure 7-8. Annual PM₁₀ spatial gradient using Universal Kriging interpolation method

8 SPRING MODEL

8.1 PM₁₀ MEASUREMENTS

During this sampling period, 206 ambient PM₁₀ weekly averaged concentrations were measured representing 79.6% of a total of 248 concentrations. The high number of samples measured during this particular period of time corresponded to the performance of the DICHOTS that were deployed after repair and maintenance had been provided and thus coincided with the beginning of the season. Descriptive statistics of the air pollution measurements are shown in Table 8-1.

Table 8-1. Descriptive statistics of the average spring concentrations for the 13 monitoring stations.

Spring Descriptive Statistics for PM ₁₀ Concentrations (µg/m ³)					
	N	Minimum	Maximum	Mean	Std. Deviation
Lindbergh	13	11.4	66.5	24.8	14.5
Fountain	15	14.7	54.0	26.7	11.6
Skyline	11	9.7	50.3	28.1	13.6
McCombs	14	10.9	51.8	22.3	11.9
UTEP	21	14.0	49.6	34.2	9.5
Chamizal	11	22.1	73.1	39.8	13.9
Parral	16	7.4	52.5	27.4	11.2
Cosmos	25	11.3	53.1	26.5	10.7
Socorro	31	26.8	67.1	47.5	12.5
Plant	11	13.3	91.8	41.1	22.3
Mimosa	15	5.7	87.5	27.1	23.9
Montana	23	4.9	54.7	25.1	12.4
Pebble	3	17.5	25.3	22.1	4.1

Weekly averaged concentrations were depurated by removing values above 150 µg/m³ and greater than three standard deviations as well as concentrations underneath it from their

corresponding site. The detection limits for PM_{10} mass concentrations were defined by three standard deviations of the mass concentration deposit on field blanks. The sample measurements averaged $30.13 \mu\text{g}/\text{m}^3$ with a standard deviation of $7.76 \mu\text{g}/\text{m}^3$. From the 206 sampled weekly concentrations, 33 and 25 observations were at least 1 standard deviation above or below the mean, respectively. All the measured values exhibited a normal distribution as shown in APPENDIX C. The maximum spring seasonal average PM_{10} concentration of $46.54 \mu\text{g}/\text{m}^3$ was observed at the Socorro site near the border and a minimum of $22.03 \mu\text{g}/\text{m}^3$ at McCombs site, which is in the farthest north. The overall highest weekly averaged PM_{10} concentration of $91.84 \mu\text{g}/\text{m}^3$ was measured at Plant site, located near the border highway in the eastern part of El Paso. Whilst the overall lowest concentration of $4.91 \mu\text{g}/\text{m}^3$ was sampled at Mimosa site, located at the far eastern part of ELP. An evident southward trend with the highest values near the border was observed as well as a less marked northeastern and westward trend. In general, the lowest values were observed within the western area whereas the highest were near major roadways and the downtown international border. The average $PM_{2.5}/PM_{10}$ ratio, calculated from the dichotomous samples, averaged 0.25 with a standard deviation of 0.116. The observed $PM_{2.5}/PM_{10}$ ratio agrees with values previously observed in the same region (Li et al., 2001).

Lastly, a spatial autocorrelation analysis was conducted using Morans's I test, using the nearest neighbor approach, to measure the degree of interdependency among the data arranged for modeling intra-urban PM_{10} mass concentration. After conducting such test, the sites produced a value of 0.00 ($p < 0.0005$) suggesting minimal spatial autocorrelation.

8.2 REGRESSION ANALYSIS

Each of the independent variables was tested through the methods previously mentioned using the regression analyst tool in SPSS 17 and PASW statistics 18. This identified variables that were highly correlated with the PM_{10} observations. Nearest distance the border (OND) had the highest correlation coefficient among all variables and was entered into the model under an enter-forward procedure. Subsequently, the remaining variables were entered in a stepwise procedure to determine the addition of variables that would paired with the elevation. This process was repeated until three variables were defined. However, the results were either hampered by high levels of collinearity between the variables or the coefficients were assigned an opposite sign to what logic suggested. Inconsistencies were shown especially with commercial land use, population and traffic counts as they were given an opposite sign to what is was expected. This contradicts the assumption that with higher commercial, the higher the human activity, and the denser the traffic-related parameters would be and thus higher the intra-urban gradient of PM_{10} concentrations would be observed. As a result, a variety of models were developed trying to eliminate variable misinterpretation and to generate a reliable correlation with the elevation and PM_{10} . After conducting nearly 40 different models, no acceptable model fit was obtained nor was a concordance between variables and their corresponding coefficient signs. Thus, OND was discarded from the analysis and the subsequent highly correlated variable was inputted into the model in the same procedure. Similarly to the annual LUR, Montanavista (site L) was dropped from the analysis as it was generating inconsistencies in the model. Consequently, log transformed group zoning (NGZ15k) within a 1.5 km buffer was entered into the model. Then, additional variables were paired with other significant variables in a series of trivariate regression models. This series of analysis identified the pairing with the log-

transformed population density within the 1500 m buffer area (npop15k) of each monitor. Table 8-2 provides a model summary and the ability of the two variables (NGZ15k and npop15k) to predict ambient PM₁₀ mass concentrations (indicated by column 1).

Table 8-2. Land use regression model summary of the spring iteration.

Spring Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.939 ^a	.881	.855	3.03

a. Predictors: (Constant), npop15k, NGZ15k

b. Dependent Variable: PM10_Mean

Similar to what observed in the annual model, 85.5% of the variability of ambient PM₁₀ concentration in the El Paso area is explained by NGZ15k and npop15k with a standard error estimate (s) of 3.03 µg/m³. Table 8-3 shows the regression coefficients for the spring model. None of the VIFs values surpasses the rule of thumb of 2.5, thus suggesting no multicollinearity between both variables. Likewise, the tolerance value reaffirms the certainty that none of the variables were intercorrelated and more importantly that the concentrations estimated by the model may produce lower deviations from the actual fitted regression plane.

Table 8-3. Statistics of the model's coefficients for the spring iteration.

Spring Model Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	6.494	3.511		1.850	.097		
NGZ15k	39.499	5.091	.906	7.758	.000	.967	1.034
npop15k	.009	.002	.460	3.938	.003	.967	1.034

a. Dependent Variable: PM10_Mean

Besides tolerance and VIF, the collinearity diagnostics seen in Table 8-4 is an alternative method of assessing if there is too much multicollinearity in the model. To simplify, cross products of the independent variables are factored. There will be as many factors as independents, plus one (for the constant). The first column addresses the dimension representing the variables included plus one constant variable. The ordering of the variables corresponds to the same arrangement shown in the variance proportions. That is, the constant variable corresponds to the dimension 1, NGZ15k to the second dimension and so forth. The subsequent column identifies the eigenvalue which unfolds little variance in the cross product matrix, as high eigenvalues indicate dimensions (factors) which account for much of the variance in the cross product matrix; eigenvalues close to 0.00, such as this case, indicate dimensions which explain modest variance. Values such as 0.039 for npop15k and 0.185 NGZ3m obviously underwent the rule of thumb with dimensions above the 0.00 value.

Table 8-4. Collinearity diagnostics of the model's parameters for the spring iteration.

Spring Model Collinearity Diagnostics ^a						
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	NGZ15k	npop15k
1	1	2.775	1.000	.01	.02	.01
	2	.185	3.872	.01	.63	.20
	3	.039	8.387	.98	.35	.79

a. Dependent Variable: PM10_Mean

With the highest condition index of 8.38 prompt by npop15k, it was clear that the model had no collinearity problems making it unnecessary to examine variance proportions. However, it was required to further assess any problems of multicollinearity as it is possible for condition indices to indicate multicollinearity, even when the rules of thumb for tolerance and VIF were met.

Lastly, after determining the absence of collinearity, PM₁₀ concentrations were determined to be influenced by the normalized population in a 1.5 km buffer area, and normalized group zoning within a 1.5 km buffer area as shown in Equation 8-1.

$$PM_{10} = 6.494 + 39.499 \cdot (NGZ15k) + 0.009 \cdot (npop15k) \quad (8-1)$$

8.3 MODEL VALIDATION

After determining the appropriate model, standard regression diagnostics were applied to assess the problems such as influential outliers, linearity, independency, normality, and homoscedasticity. Standard multiple regression can only accurately estimate the relationship between dependent and independent variables if the relationships are linear in nature. Therefore if violations of linearity are extremely serious (fitting data to a linear model which are nonlinearly related), the predictions are likely to produce systematic errors, especially when generating usually large or small predictions.

Figure 8-1 unveils a correlation coefficient of 0.881, and a symmetrical point distribution around a diagonal determining a linearity relationship between PM₁₀ mass concentrations and NGZ15k and npop15k. Note that the standardized predicted values were calculated using the model equation for every particular monitoring site previously discussed. In addition the statistical software suited for this research produce such values in a residual statistics table.

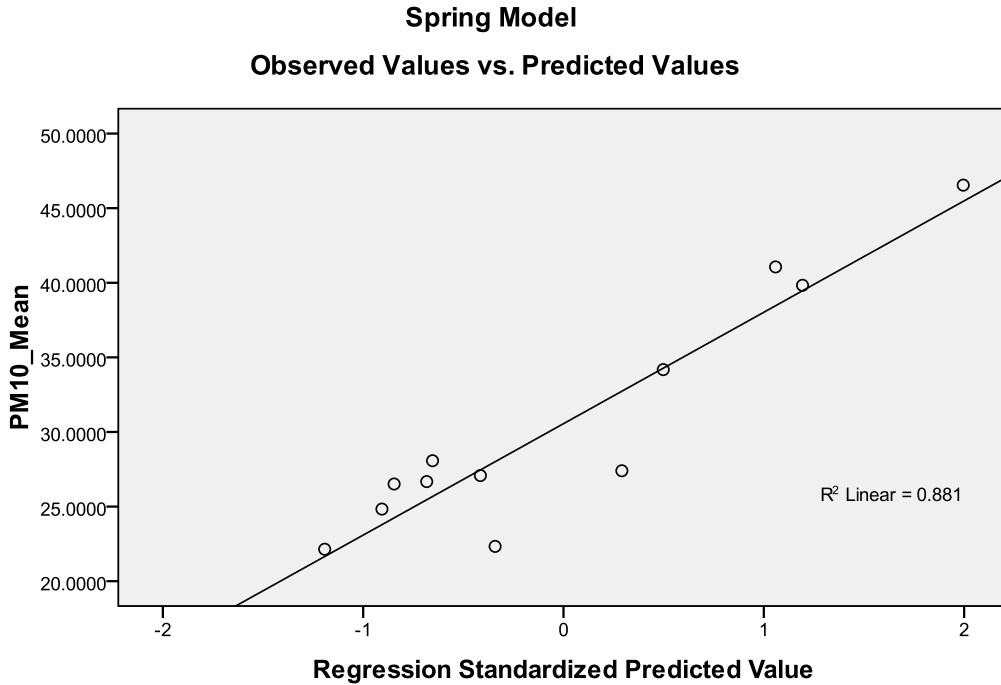


Figure 4-26. Linearity test between the sample concentrations and the predicted values for the spring model.

In addition to the Moran's I test, performed in previous section, a second evaluation was performed under the Durbin-Watson test. Thus using the test statistics shown in Equation 7-1, a corresponding d value of 2.44 was achieved. Although after obtaining the $d_{\text{lower critical}}$ value (0.812) and the $d_{\text{upper critical}}$ value (1.579) at 5% significance from APPENDIX E, the test rejects the null hypothesis of positive autocorrelation and thus accepts the alternative hypothesis demonstrating that the error estimates are not positively autocorrelated.

Table 8-5 contains summary data regarding the residuals (the difference between predicted and actual values). Standard residual (Std. residual) for instance is the raw residual divided by the standard deviation of residuals. Since the minimum standardized residual is -1.803, at least one prediction is more than one standard deviation below the mean residual. Studentized residuals are very similar to standardized residuals but follow the t distribution. Moreover, Mahalanobis distance should not exceed the critical chi-squared value. Using two

degrees of freedom and an alpha value of 0.001, the threshold value was determined to be 13.82. Similarly the minimum, maximum, and mean Mahalanobis distances displayed in the Table 8-5 above denote much lower distances, disregarding any possibility of outliers influencing the development of the linear regression model.

Table 8-5. Descriptive statistics of the spring model's residuals.

Spring Model Residuals Statistics^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	21.653481	45.456493	30.555408	7.4670280	12
Std. Predicted Value	-1.192	1.996	.000	1.000	12
Standard Error of Predicted Value	1.021	2.666	1.449	.462	12
Adjusted Predicted Value	21.318167	41.734745	30.168678	6.9152521	12
Residual	-5.6755395	2.6024718	.0000000	2.7402960	12
Std. Residual	-1.803	.859	.000	.905	12
Stud. Residual	-1.990	.998	.044	1.007	12
Deleted Residual	-6.4872465	4.8075542	.3867303	3.5433196	12
Stud. Deleted Residual	-2.507	.997	-.045	1.181	12
Mahal. Distance	.333	7.599	1.833	2.029	12
Cook's Distance	.000	.650	.113	.189	12
Centered Leverage Value	.030	.691	.167	.184	12

a. Dependent Variable: PM10_Mean

The minimum, maximum, and mean Cook's distance is displayed in Table 8-5. Using the cut-off suggested of D greater than 1, no apparent problem cases arose since the maximum Cook's D is only 0.650. Despite not meeting the Fox (1991) criterion, linear regression is a robust technique that allows the variance in testing procedures or threshold parameters. That is why for this particular model, different thresholds were used to assess the significance of such and power predictability.

Finally the maximum centered leverage value of 0.691 shown in Table 8-5 for the spring model was smaller than the computed threshold from the small sample of 0.75, thus discarding the possibility of any undue influence from the NGZ15K and npop15k in the spring model.

Errors, represented by the residuals, should be normally distributed for each set of values of the independents, as violations of normality compromise the estimation of coefficients and the calculation of confidence intervals. Calculations of confidence intervals for coefficients are based on the assumptions of normally distributed errors. In Figure 8-2, residuals histogram followed a bimodal distribution based on predicting PM_{10} mass concentrations from NGZ3m and npop15k. This suggests that the error distribution is "skewed" by the presence of possible outliers. Since parameter estimation is based on the minimization of squared error, a few extreme observations may have exerted a disproportionate influence on parameter estimates.

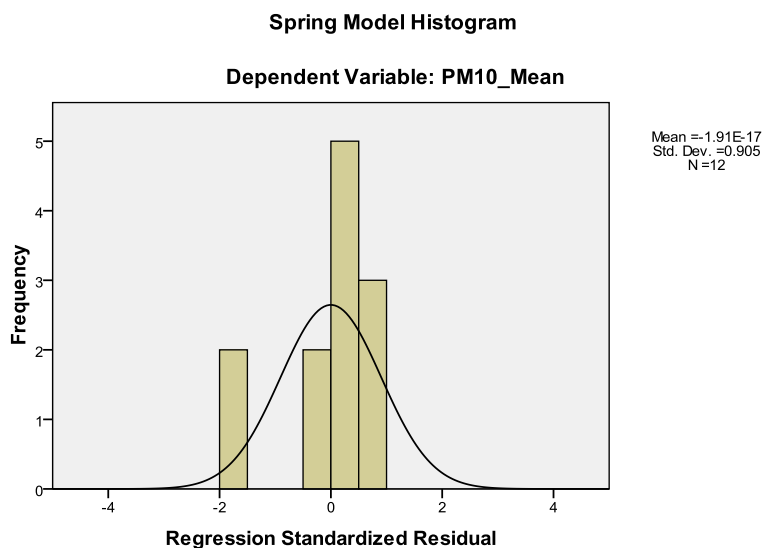


Table 8-2. Residual's distribution of the spring model.

The Shapiro-Wilk's (W) test was also performed for the spring model. In this case, W was significant with a statistic values of 0.78. This suggests that the variable's distribution is significantly different from normal (Table 8-6).

Table 8-6. Shapiro-Wilk's W test for normality of the spring model's residuals.

Spring Model Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.279	12	.011	.778	12	.005

a. Lilliefors Significance Correction

A further investigation revealed that the inclusion of a two monitoring sites, McCombs (D2) and Parral (H), caused the alteration from normality of the error distribution graph which overpredicted PM_{10} concentrations by 20 and 16% respectively. Overall, however, the concentrations shift had little overall effect on the model predictions. Removing both current monitored concentrations the R^2 value (0.85) yielded a R^2 value of 0.86 that was, on average, just 0.6% different, merely insignificant, thus both sites were keep in the analysis.

The homoscedasticity test was performed for the spring model. Figure 8-3, illustrates a disperse pattern of the points with no trend. The distribution of the residuals among the standardized residuals axis demonstrate some ranges of errors below the negative standardized residual axis, while others ranged positive above the 0 value of the y-axis. However, there are two possible outliers in the standardized residuals data. These outliers certainly are demonstrated in Figure 8-4, as the residual boxplot showed the residuals of McCombs and Parral to have values over two standard deviations from the mean.

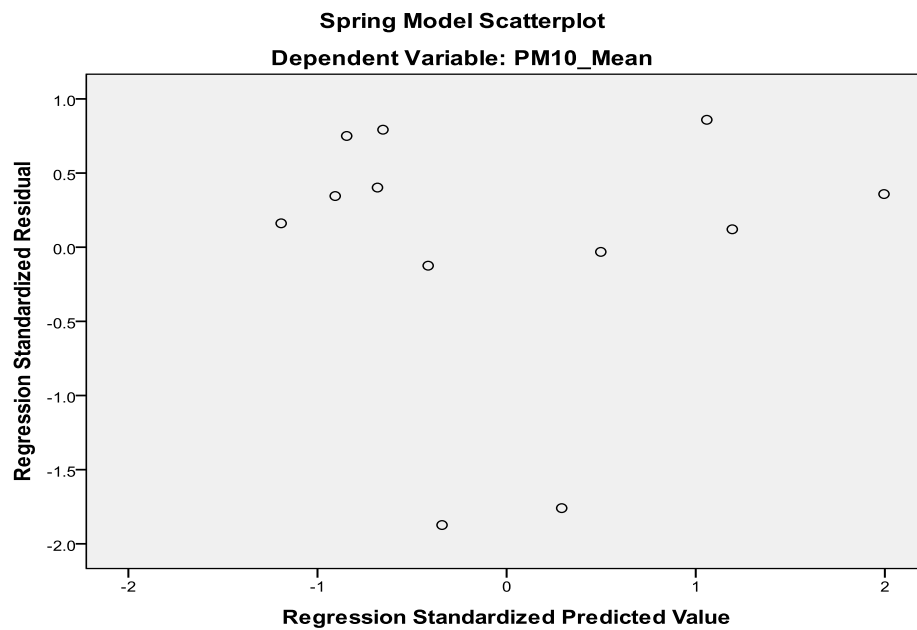


Figure 8-3. Scatterplot of the standardized residuals versus standardized predicted values for the spring model.

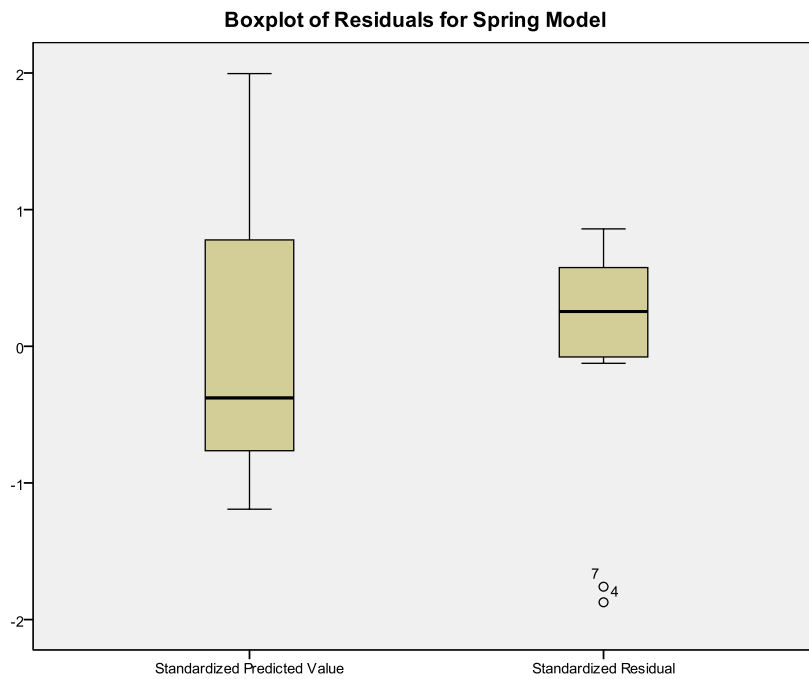


Figure 8-4. Boxplot of the standardized predicted values and standardized residuals for the spring model.

However, Pardoe (2006; page 166) defines that a potential outlier should be considered as an observation with a studentized residual outside the range of -3 to +3. Under this determination the studentized residual remains under the limit area. Yet after removing the observations in question, the boxplot did not detect any additional potential outliers. With no change in the regression line ($\pm 0.6\%$), neither changes in the parameters nor the change in the linear sign; the observations in question were retained in the main analysis as either test were fully met. Therefore, it should be remarked that there was no conclusive evidence to remove the observations from the analysis as they merely represented potential outliers. The scatter plot, nevertheless, reflects the fact that the model in this case has modest standard errors of estimate, explaining the percentage of variance mainly as a consequence of an acceptable fitted model. Furthermore, the scatterplot of the standardized residuals versus standardized predicted values corroborated the influence of McCombs and Parral in the normal distribution. On the other hand, the modest violation generated by both sites in the homoscedasticity test produced a wider range of concentrations between sites, even more disrupted than the annual model. Nevertheless, and regardless of the McCombs and Parral samples, the residual plot patterns endorsed the relationship between predictor and criterion as linear.

Heteroscedasticity was evaluated in Figure 8-5 and Figure 8-6. These cases were weighted by the reciprocal of their estimated point variance, and each was assessed with the goodness of fit of the weighted model. That is, weighted predicted values were plotted on the x -axis and weighted residual on the y -axis. The two weighted residual plots produced a good fit. Although two observations in Figure 8-6 appear to be outliers, the residuals did not form a funnel shape. The two questionable observations corresponded to the residuals of the model measured at McCombs and Parral weighted by the reciprocal of their estimated population density variance.

Nevertheless the two observations remained insignificant for the heteroscedasticity test and merely represented potential outliers.

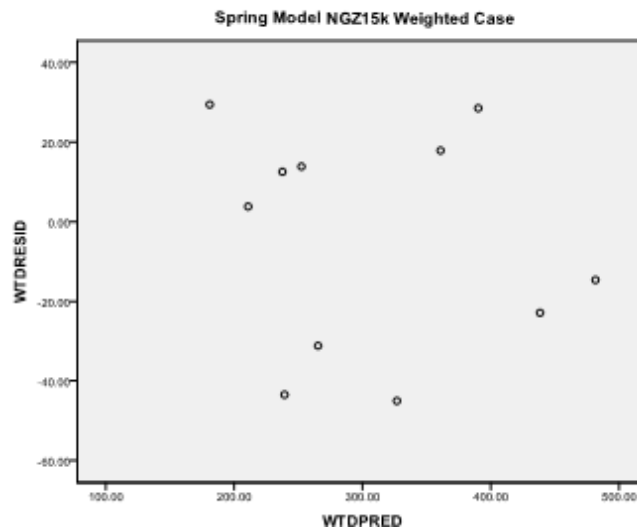


Figure 8-5. Weighted case of NGZ15k for the spring model.

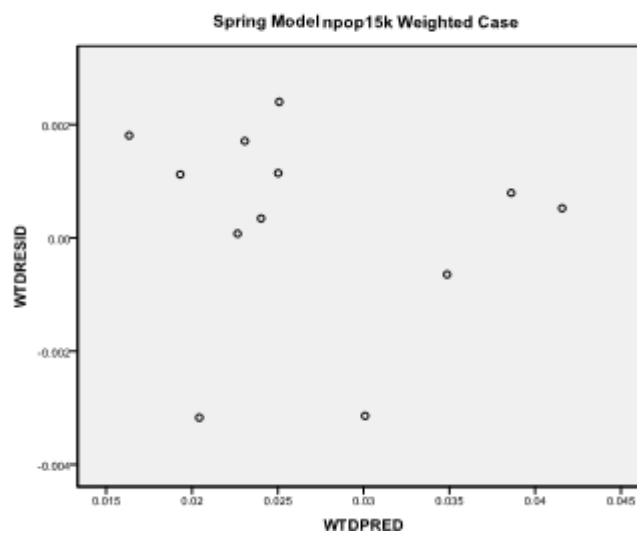


Figure 8-6. Weighted case of npop15k for the spring model.

8.4 CROSS-VALIDATION

Cross-validation was performed using the same 5 methods previously described in Chapter 7. The bootstrap analysis was undertaken using 3528 predicted values that were divided into 147 subsamples of 24 random samples. The variance was calculated as the square root of the sum of the squared differences between the subsample i -th average and the overall mean divided by the number of representatives minus one (3527). Supplementary to this method, histograms of the predicted parameter were validated and derived over 1000 iterations to examine the consistency of the standard variance as presented in APPENDIX F. Also each model was validated a comparison of prediction errors obtained by simulating 1,000 jackknife data sets. Root mean square errors and the standardized root mean square errors were calculated as the square root of the sum of the squared differences of the observed concentration at site i and the predicted concentration at site i divided by the total grid nodes (3528). In each of the cross-validations, the average difference was not greater than $10.25 \mu\text{g}/\text{m}^3$ for bootstrap, translating into an average relative difference of less than 7.0%. Also, there was a recurrent estimation variance value of $0.061 \mu\text{g}/\text{m}^3$ for the jackknife method, resulting in an error of 10%, and $7.521 \mu\text{g}/\text{m}^3$ for RMSE, which intuitively contrasts the measured values with the predicted values within an 12% difference. The results of this comparison are presented in Table 8-7.

Table 8-7. Cross-validation results of the five interpolation techniques used to map spring PM_{10} mass concentration throughout the El Paso cohort study.

	Spring Model			
	Bootstrap Variance	Jackknife Variance	RMSE	SRMSE
IDW	10.08141	0.054159	7.447	N/A
NN	9.647528	0.043323	7.392	N/A
OK	9.603625	0.041428	7.0879	0.7809
SPLINE	10.2501	0.06077	7.521	N/A
UK	9.536393	0.038749	6.7504	0.8107

Amongst the modeling techniques, the predicted concentrations generated by the SPLINE method had the worst performance, with a bootstrap variance of $10.25 \mu\text{g}/\text{m}^3$, a variation in prediction certainty over 1000 iterations of $0.0607 \mu\text{g}/\text{m}^3$, and a root mean square error of $7.521 \mu\text{g}/\text{m}^3$; although the bootstrap histogram demonstrated a presumptuous bell shape, as seen in APPENDIX F. In this case, the deplorable prediction performance was caused by the fundamentals of the model as it created the surface specifically so that it passes through the control points having the least possible change in slope at all points, fitting the control points with a minimum curvature surface. In other words, the model tends to average to its utmost capacity the PM_{10} exposure throughout the cohort study. Thus, coinciding with the visual pattern (see APPENDIX H) were the incongruous blemishes, of higher concentration at lower polluted areas, broadens more frequently than in the rest of the interpolation techniques. Considering the fact that the Spring PM_{10} surface tends to have a more disrupted variation pattern overall, when compared to the annual variation pattern.

IDW likewise also performed badly, as anticipated from the visual inspection, with a bootstrap variance of $10.081 \mu\text{g}/\text{m}^3$, a RMSE of $7.447 \mu\text{g}/\text{m}^3$ and a fractional bias obtained from the jackknife analysis of $0.0541 \mu\text{g}/\text{m}^3$. Although the bootstrap histogram follows a normal distribution with no bimodal shape whatsoever, still the model tended to misrepresent the spatial variability of PM_{10} worse than NN and kriging.

Sequentially, NN performed better than the latter two methods described above, with a bootstrap difference of $9.647 \mu\text{g}/\text{m}^3$, a fractional bias of $0.0433 \mu\text{g}/\text{m}^3$, and a RMSE equal to $7.392 \mu\text{g}/\text{m}^3$ and far less attenuation, though the bootstrap histogram followed a normal distribution with no bimodal shape.

Lastly, the best performance was seen for both ordinary and universal kriging approach for which the bootstrap variance equals to $9.604 \mu\text{g}/\text{m}^3$ and $9.536 \mu\text{g}/\text{m}^3$ respectively; the jackknife variance for ordinary kriging prompted a value of $0.0414 \mu\text{g}/\text{m}^3$ as well as $0.0387 \mu\text{g}/\text{m}^3$ for universal kriging. For the ELP using 3-yr data the RMSE for OK and UK resulted in an absolute value of $7.0879 \mu\text{g}/\text{m}^3$ and $6.7504 \mu\text{g}/\text{m}^3$, respectively; by contrast, the SRMSE on these predictions was $0.7809 \mu\text{g}/\text{m}^3$ and $0.8107 \mu\text{g}/\text{m}^3$. Nevertheless, the histogram of the predicted estimates derived for each of the 1000 iterations suggested in both studies that the models were robust. Albeit ordinary kriging showed a slight bimodal shape traceable to a single grid node, located at UTEP, just 1.5 miles from the Santa Fe international bridge and half a mile from the border with CDJ. This area projected low concentrations considering the fact that the areas surrounding this station are residential and commercial as well as being in close proximity to the border and the downtown port of entry. This point had 1 standard deviation away from the concentration mean of all samples falling into the 95% confidence interval. However, there is no reason to believe that the underlying data from this station is inaccurate, as the value showed a 23% variance from the average value measured at the State fix station CAMS 12 (F). In addition, the correlation between the monitoring site and the fix site prompted a 71% at the validation stage in this analysis and therefore the monitor was not excluded. It is important to mention that universal kriging also denoted the same irregularity at the UTEP location; although the concentration merely showed a 15% variance when compared to average value sampled in CAMS 12 mainly because of the assumption of an external drift.

For most metrics, the models derived from the two runs are broadly similar, indicating considerable robustness in the methods in the face of varying input data. There is some tendency, however, for the performance of the ordinary kriging model to be rather worse than the

universal kriging, indicating the sensitivity of the methods to the underlying input data. Kriging gives the preferred cross-validation values at the grid nodes, but these generally reflect the way in which the kriged surfaces are fitted to those points. Semivariograms (shown in APPENDIX G) show strong spatial dependency in the data, with low nugget variance compared to the mean concentrations and an indistinct sill, and thus suggest that the models predicted concentrations well at unsampled locations. Therefore the overall cross-validation diagnostics concludes that kriging methods demonstrated that the influential points are unlikely to drive the regression model and that universal kriging feebly outperforms ordinary kriging in predicting spatial characterization. All because ordinary kriging assumed that the variation in PM_{10} is free of any structural component. The unusual values in population, for example, inflated and deflated the model prediction power when compared to the observations. Ordinary kriging, as a result, oversmooths and hence misses areas with unusually high and low particulate matter caused by omitting an external drift. It is also possible that ordinary kriging in this context violates the fundamental assumption of stationarity. Points close to residential areas, for example, exhibited a different relationship between PM_{10} and population rather than points, for example, in less developed residential areas far from any major source of PM_{10} . These results indicate that the difference in temporal variation produced relatively moderate errors when comparing measured values to modeled values.

8.5 DISCUSSION

The spring model was constructed using 206 weekly concentrations from field monitoring sites in the El Paso area. The Montanavista site had to be withdrawn for the same reasons discussed in Chapter 7. In the first iteration of developing a city-specific LUR model

commercial and industrial area within 750 m and 3000 m respectively were strongly predictors in the model when paired to elevation. The grouping of all three variables into a single parameter was done in order to contain the model within three variables and leave span for other variables to be recognized by the model as surrogates for PM_{10} concentrations rather than concentrating the model into elevation and two independent land use variables. Therefore, this newly developed log-transformed variable (NGZ15k) was the most consistent predictor of PM_{10} levels, accounting for over 82.3% of the variation in PM_{10} . Industrial and commercial densities served as a surrogate for the general level of human activity in the vicinity of a monitoring site. These densities have been associated with increase emissions, suggesting that with higher density areas may be subject to higher emissions by human activity. Also, the segregation of these land use types creates a need for travel between residential and commercial and industrial areas. Taken together, these characteristics increase the demand for travelling by automobiles, and they create an exposure surface with more spatial variability in the ELP urban area. However, traffic-related variables were not included in the model as they showed a deplorable explanatory power when paired with NGZ15k, regardless of their associations. Predictors directly based on VMTs and traffic counts were of primary interest in the analysis as they explained 63% and 58% respectively of the spatial variability when correlated with PM_{10} , yet none of them could produce a reliable explanatory power when paired with other geographical congruent variable; they were excluded from the analysis based on the preferred explanatory power produced by NGZ15k. Aside from a second stepwise iteration, incorporated into the model normalized Population density within a 1500 meter circular area, serving as a surrogate for the general level of human activity in the vicinity of the monitoring site. Like industrial and commercial areas, population density has been associated with decreased driving speed, as the city's law enforces a lower

speed limit at residential areas. This suggests that areas with higher population density are subject to higher emissions per vehicle (National Resource Council, 2005). A third iteration was pursued, however, no variable seems to correlate significantly enough with population and group zoning to be included in the model. Yielding inverse correlations or variables that were correlated somehow with an untrusted confidence interval over or underestimating ambient PM concentrations. Therefore, the model was contained to only two predictor variables. As mentioned before log-population within a 1500 meter showed a reasonable explanatory power, however when combined with normalized group zoning within a 1.5 km, the model was able to explain 85.5% of the variation of PM_{10} concentrations.

Additionally, the fitted model was tested for non-linearity, spatial autocorrelation, collinearity, homoscedasticity and normality. Two questionable observations draw the attention of possible outliers. The observations corresponded to the average concentrations sampled at McCombs and Parral; thereby after conscious analysis they could not be determined to be outliers. Apart from that, both observations after being excluded and the model reassessed the R^2 had an insignificant shift of merely 0.6%. Thus the observations were determined to have no undue influence on the model fit. Likewise, both residuals were excluded from the standardized residual versus standardized predicted value plot, then the scatterplot re-graphed to the remaining sample observations corroborating that the substantial change in the linear fit did not affect the model fit whatsoever. Moreover, according to Pardoe (2006; page 166), a potential outlier should be considered as an observation with a studentized residual outside the range of -3 to +3. Based on this determination both studentized residual remained under the limit area. Thereby, with no change in the regression line, no changes in the parameters, and no change in the linear sign the observations in question merely represented potential outliers and had insignificant influence in

the homoscedasticity test. In addition, a bootstrap to evaluate the sensitivity of the parameters to the actual samples included in the modeling suggests that the model is relatively stable to the choice of samples. In general, these variables were robust to sample selection, though it was found that normalized group zoning and normalized population variables were affected by the inclusion or exclusion of the Montanavista sample that had modest PM_{10} concentrations. The fact that group zoning (500 m) and length of roads (5 km) in the El Paso city borough, using an entirely different GIS layer and without the influence of the Montanavista sample, was more significant and lends support to the important role played by group zoning and population for this particular model. It is important to mention that the predictions for the El Paso region are not at all compromised by this problem, and on the whole, the named limitations do not affect the main findings of the analysis.

Consequently, after deriving the parsimonious operational regression model a grid covering the cohort study and containing 3528 nodes was developed in order to sharpen the resolution of the depicted overall PM_{10} trend. At each node the concentration of PM_{10} was calculated applying the two parameters with their corresponding coefficient and joined into a single attribute table. It is important to remark that no predicted value had to be withdrawn from the analysis as the highest concentration value ($68.476 \mu\text{g}/\text{m}^3$) was 25% lower than the maximum sample concentration measured at Plant (K). Given the strong predictive power of hybrid models (Jerrett et al., 2005), five interpolations methods (IDW, NN, OK, UK, AND SPLINE) were combined with LUR to generate a confident PM_{10} spatial surface. The validity of three models was challenged by three procedures (bootstrap, jackknife, and RMSE) whereas the remaining two models were assessed by an additional method (SRMSE). The bootstrap cross validation procedure for UK yielded a variance over the subsample average of $9.536 \mu\text{g}/\text{m}^3$ in

contrast to the second lowest variance of $9.603 \mu\text{g}/\text{m}^3$ prompted by the OK method. The fractional bias of $0.0387 \mu\text{g}/\text{m}^3$ for UK was smaller when compared to the second lowest value of $0.0414 \mu\text{g}/\text{m}^3$ prompted by the OK method. The predictability of universal kriging was more precise with a RMSE of $6.750 \mu\text{g}/\text{m}^3$ compared to $7.087 \mu\text{g}/\text{m}^3$ for the ordinary kriging. Lastly, the SRMSE value for UK was closer to one than the OK meaning that the estimated standard error was more reliable or valid for the uncertainty of the predicted values. Overall universal kriging based on Gaussian model outperformed IDW, NN, SPLINE, and OK on cross-validation results using the full set of samples. The cross-validation procedure, for example, yielded a bootstrap and jackknife variance, which was small compared to the range of concentrations and not greater than the RMSE, indicating that the influence of the single observations was not too heterogeneous. In addition, 88% of the UK modeled estimates were within a factor of 1.2 times the range of the measured concentration with an inter quartile range value of $25.271 \mu\text{g}/\text{m}^3$ approximately 12% less than the second lowest value cued by OK and 18% to the SPLINE method. The ELP universal kriging model exhibited a bias less evident, but still existing, in the full model that included the validation samples overpredicting 36% of the value locations and underpredicting 29% with a relatively low average absolute value residual or mean absolute error of $4.399 \mu\text{g}/\text{m}^3$; therefore the overall model tended to overestimate PM_{10} concentrations. The underpredicted values manifested more pronounced in two urbanized areas of the cohort study. These areas have a predominant commercial activity with a moderate residential area. One is located in the proximity of US-Mexico border specifically in the downtown core, where commercial business predominates this area with a modest residential percentage; the second area is located circumvent to Western Refining, which also has a vast amount of industrial, commercial and residential areas. Albeit, the overprediction appeared in areas distant from the

centralized urbanized sprawl embracing sectors from the Fareast along Montana Ave., at the intersection with Highway 375, the uppermost west area above Transmountain Rd., and to a lesser extent, the surrounding upper corner of northeast area. Conversely, categorizing the sites by tertiles, 68% of them would be classified identically. Only 32% of the grid nodes near commercial and industrial areas would change between the extreme tertiles. The divergences might be partially explained by the temporal variance of the sample period which is focused on a three month period for the city PM₁₀ gradient. When the predicted exposure was divided by the predicted quartiles it was estimated that 37% of El Paso residents fall into a range of 40 to 50 $\mu\text{g}/\text{m}^3$ where 1.5% of which fall into the highest categories for PM₁₀ concentrations between 50 to 60 $\mu\text{g}/\text{m}^3$ during the spring time period.

Finally, in this study LUR in combination with universal kriging successfully predicted small-scale spatial variation of ambient PM₁₀ concentrations and adequately reproduced the regional trend identified by the monitoring network. The model predicted the lowest concentration values within the west central part of El Paso. That is a less marked western trend whereas a steeper concentration gradient eastward towards a more populated region. The higher concentrations, though, were mapped at different locations especially at the northeast and east side, although some modest areas in the west also appeared to be affected. Areas prone to have higher particular concentrations in some such way appeared surrounded by residential areas were additional industrial or commercial areas were adjacent in a distance of 1.5 km and vice versa. Nevertheless, several investigators cite wind speed and direction as potential influences (Arain et al., 2007; Su et al., 2007) on PM concentrations. The 1996 Paso del Norte Ozone Study demonstrated how wind speed critically influence NO_x concentrations in El Paso (MacDonald et al., 2001; and Brown et al., 2001). This particular easterly pattern may have been influenced by

the fact that 98% of the CAMS stations registered wind patterns to be from west to east and with 93% of the windspeeds to be above 22 knots. Such high windspeed are attributable to the windstorms that bash the El Paso region precisely in the months of March and April within the analysis time period. On days when the wind is light, PM₁₀ remain confined near central El Paso emission sources, while on days with moderate wind, the precursor cloud is dispersed and PM₁₀ concentrations are more widely and evenly dispersed across the city. Since windstorms are a hallmark of spring in El Paso, measurements of windspeed and direction in addition to estimates of mixed layer growth rate would be needed to expand the current analysis to other seasons. As shown in Figure 8-7 the model was able to identify one noticeable hot spot surrounding the downtown area adjacent to the US-Mexico border with concentrations ranging from 50 to 60 $\mu\text{g}/\text{m}^3$. This hot spot area is predominated by commercial areas, with a lesser, residential percentage. Despite the fact that Ciudad Juarez might have influenced the spatial PM₁₀ distribution in the downtown core, it was impractical to monitor the entire metropolitan area due to cost and political constrains in order to quantify the impact of CDJ. Along with this the model was able to capture the short-term spatial variations of averaged PM₁₀ concentrations for a localized area within larger metropolitan region and predicted a gradient decreasing with distance from the border. The methodology captured the relative impact of emissions outside the localized monitoring network area and allowed for an optimal utilization of resources.

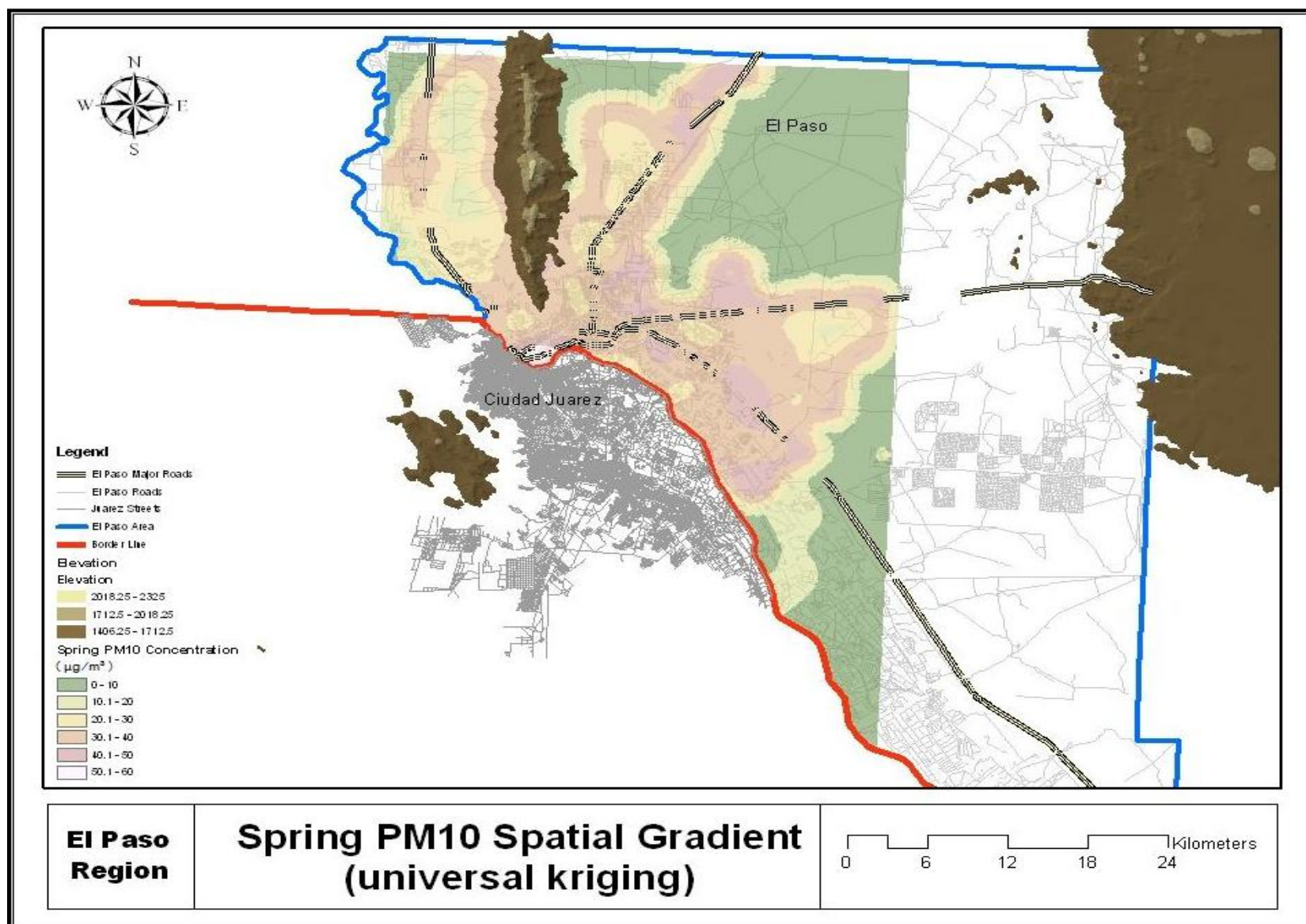


Figure 8-7. Spring PM₁₀ spatial gradient using Universal Kriging interpolation method.

9 SUMMER MODEL

9.1 PM₁₀ MEASUREMENTS

During this sampling period, 148 ambient PM₁₀ weekly averaged concentrations were measured representing 64% of a total of 224 concentrations. The modest number of samples is due to a consistent loss of nearly one month in average of measurements for every site as the samplers were brought to the lab for repair and maintenance during the summer. Descriptive statistics of the air pollution measurements are shown in Table 9-1.

Table 9-1. Descriptive statistics of the average summer concentrations for the 13 monitoring stations.

Summer Descriptive Statistics of PM ₁₀ Concentrations (µg/m ³)					
	N	Minimum	Maximum	Mean	Std. Deviation
Lindbergh	9	17.9	41.9	30.9	7.9
Fountain	11	6.2	25.3	14.1	6.1
Skyline	12	6.9	25.5	12.4	5.4
McCombs	11	5.6	80.6	19.6	20.9
UTEP	13	8.5	36.4	23.4	8.8
Chamizal	9	14.2	58.1	31.9	13.8
Parral	9	9.9	31.2	16.9	6.5
Cosmos	20	4.8	69.8	22.4	17.4
Socorro	14	12.6	39.1	23.0	8.9
Plant	12	16.6	46.5	30.1	9.1
Mimosa	11	10.3	60.2	27.2	14.8
Montana	11	4.9	28.9	14.6	6.8
Pebble	6	8.5	79.1	30.6	25.3

Weekly averaged concentrations were depurated by removing values above 150 µg/m³ and greater than three standard deviations as well as concentrations underneath it from their corresponding site. The detection limits for PM₁₀ mass concentrations were defined by three

standard deviations of the mass concentration deposit on field blanks. The sample measurements averaged $22.84 \mu\text{g}/\text{m}^3$ with a standard deviation of $6.94 \mu\text{g}/\text{m}^3$. In contrast, lower concentrations and less variation than the spring season was observed, as the summer season tends to witness rare stagnations events and thus a better dilution of air pollutant. From the 148 sampled weekly concentrations, 20 and 18 observations were at least 1 standard deviation above or below the mean, respectively. All the measured values exhibited a normal distribution as shown in APPENDIX C. The maximum summer seasonal average PM_{10} concentration of $31.93 \mu\text{g}/\text{m}^3$ was observed at the Chamizal site near the BOTA Bridge and a minimum of $12.35 \mu\text{g}/\text{m}^3$ at Skyline site, which is in the farthest north. The overall highest weekly averaged PM_{10} concentration of $80.55 \mu\text{g}/\text{m}^3$ was measured at McCombs site, also located in farthest north part of El Paso whilst the overall lowest concentration of $4.84 \mu\text{g}/\text{m}^3$ was sampled at Cosmos site, located in a residential area. Is noteworthy, that the highest weekly concentration observed at McCombs site could potentially be the result of construction performed in the proximity of the site during the sampling period. However, the concentration did not surpass the outlier limit of $150 \mu\text{g}/\text{m}^3$ or three standard deviations from the mean and thus was keep in the analysis. An evident southward trend with the highest values near the border was observed as well as a less marked northeastern trend, as seen in Table 9-1. In general, the lowest values were observed within the northeastern and areas whereas the highest were scrutinized near major roadways and the international border.

The average $\text{PM}_{2.5}/\text{PM}_{10}$ ratio, calculated from the dichotomous samples, averaged 0.20 with a standard deviation of 0.127. A spatial autocorrelation analysis was conducted using Morans's I test, using the nearest neighbor approach, to measure the degree of interdependency among the data arranged for modeling intra-urban PM_{10} mass concentration. After conducting

such test, the sites produced a value of -0.01 ($p < 0.0005$) suggesting minimal spatial autocorrelation.

9.2 REGRESSION ANALYSIS

Each of the 190 independent variables was tested through the methods previously mentioned using the regression analyst tool in SPSS 17 and PASW Statistics 18. This identified variables that were highly correlated with the PM_{10} observations. Similar to the annual model Elevation had the highest correlation coefficient among all variables and was entered into the model under an enter-forward procedure. Subsequently, the remaining variables were entered in a stepwise procedure to determine the addition of variables that would paired with the elevation. This process was repeated until three variables were defined. However, the results continued to be hampered by high levels of collinearity when elevation was included in the analysis. The coefficients were assigned an opposite sign to what logic suggested, especially when paired with land-use and traffic-based variables as they were given a negative sign. This contradicted the assumption that with higher land use areas, the denser the traffic-related parameters would be and thus higher the intra-urban gradient of PM_{10} concentrations would be observed. Elevation was discarded from the analysis and the subsequent highly correlated variable was inputted into the model in the same procedure. In the same procedure as the annual and spring models Montanavista (site L) had to be excluded from the analysis as it was generating inconsistencies in the model. The insufficiency of information was generating a leverage effect over the fitted regression plane and thus generated low adjusted R^2 coefficients. After discarding elevation from the analysis, nearest distance to the boarder (OND) was entered into the model and additional variables were paired with other significant variables in a series of trivariate regression models.

This series of analysis identified the pairing with the sum of all the designated truck roads' length within each monitor 500 m buffer area (TR5m) and land use commercial (lc2m) within a 200 meter buffer area. Table 9-2, provides a model summary and the ability of the three variables (OND, TR5m, and lc2m) to predict ambient PM₁₀ mass concentrations (indicated by column 1).

Table 9-2. Land use regression model summary of the summer iteration.

Summer Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.903 ^a	.815	.746	3.41

a. Predictors: (Constant), OND, TR5m, lc2m

b. Dependent Variable: PM10_Mean

Similarly to what discussed in previous Chapters, 74.6% of the variability of ambient PM₁₀ concentration in the El Paso area is explained by OND, TR5m, and lc2m with a standard error estimate (s) of 3.41 µg/m³. This means that the model could expected to predict 74.6% of the ambient PM values to within approximately ±6.82 µg/m³ (or 15%) at a 95% confidence with variables like OND, TR5K, and lc2m. Table 9-3 shows the coefficients statistical results for the spring model. As seen in Table 9-3, none of the VIF's values surpass the rule of thumb of 2.5, thus suggesting no multicollinearity between both variables. Likewise, the tolerance value reaffirms the certainty that none of the variables were intercorrelated and more importantly that the concentrations estimated by the model may produce lower deviations from the actual fitted regression plane.

Table 9-3. Statistics of the model's coefficients for the summer iteration.

Summer Model Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	10.188	3.212		3.172	.013		
lc2m	1.829	.599	.476	3.053	.016	.948	1.055
TR5m	4.984	1.224	.625	4.071	.004	.979	1.021
OND	.172	.061	.443	2.819	.023	.934	1.071

a. Dependent Variable: PM10_Mean

Besides tolerance and VIF, the collinearity diagnostics seen in Table 9-4 is an alternative method of assessing if there is too much multicollinearity in the model. Similar to the previous analysis, the eigenvalues shown in the Table 9-4 shows little variance in the cross product matrix, as high eigenvalues indicate dimensions (factors) which account for much of the variance in the cross product matrix; eigenvalues close to 0.0, such as in this case, indicate dimensions which explain modest variance. Values such as 0.51 for OND, 0.407 for TR5m, and 0.802 for lc2m obviously underwent the rule of thumb with dimensions above the 0 value.

Table 9-4. Collinearity diagnostics of the model's parameters for the summer iteration.

Summer Model Collinearity Diagnostics ^a							
Model	Dimensi on	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	lc2m	TR5m	OND
1	1	2.739	1.000	.01	.03	.05	.01
	2	.802	1.848	.00	.88	.03	.01
	3	.407	2.594	.03	.00	.93	.04
	4	.051	7.318	.96	.09	.00	.94

a. Dependent Variable: PM10_Mean

With the highest condition index of 7.31 prompt by *OND*, it was clear that the model had no collinearity problems making it unnecessary to examine variance proportions. However, it was required to further assess any problems of multicollinearity as it is possible for condition indices to indicate multicollinearity, even when the rules of thumb for tolerance and VIF were met.

Lastly, after determining the absence of collinearity, PM_{10} concentrations were determined to be influenced by the nearest distance to the border, the sum of the truck roads' length within a 500 m buffer area, and the commercial zoning in a 200 m circular area as shown in Equation 9-1.

$$PM_{10} = 10.188 + 1.829 \cdot (lc2m) + 0.172 \cdot (OND) + 4.984 \cdot (TR5m) \quad (9-1)$$

9.3 MODEL VALIDATION

Figure 9-1 unveils a correlation coefficient of 0.815, and a symmetrical point distribution around a diagonal line determining a linearity relationship between PM_{10} mass concentrations and *OND*, *TR5m*, and *lc2m*. Note that the standardized predicted values were calculated using the model equation for every particular monitoring site previously discussed. In addition the statistical software suited for this research produce such values in a residual statistics table.

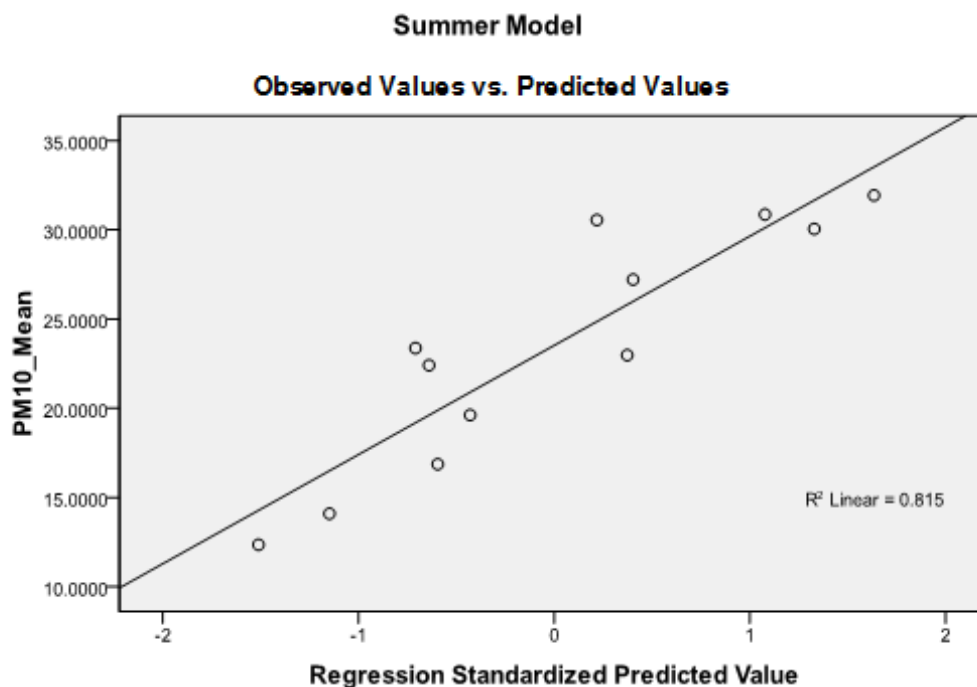


Figure 9-1. Linearity test between the sample concentrations and the predicted values for the summer model.

Table 9-5 contains summary data regarding the residuals (the difference between predicted and actual values). Standard residual (Std. residual) for instance is the standardized residual raw residual divided by the standard deviation of residuals. Since the minimum standardized residual is -0.884, none of the predictions are more than one standard deviation below the mean residual. The bottom three rows are measures of the influence of the minimum, maximum, and mean case on the model. The Mahalanobis should not exceed the critical chi-squared value. Consequently, using three degrees of freedom and an alpha value of 0.001, the threshold value was determined to be 16.27 (see APPENDIX E).

Table 9-5. Descriptive statistics of the summer model's residuals.

Summer Model Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	14.279299	33.526413	23.525950	6.1193868	12
Std. Predicted Value	-1.511	1.634	.000	1.000	12
Standard Error of Predicted Value	1.254	3.231	1.875	.635	12
Adjusted Predicted Value	15.299059	36.749569	23.524791	6.3143938	12
Residual	-3.0175610	5.6950045	.0000000	2.9112249	12
Std. Residual	-.884	1.668	.000	.853	12
Stud. Residual	-1.032	1.794	-.024	.991	12
Deleted Residual	-4.8162675	7.1805453	.0011588	4.4522107	12
Stud. Deleted Residual	-1.037	2.170	.019	1.058	12
Mahal. Distance	.567	8.939	2.750	2.610	12
Cook's Distance	.006	.791	.165	.274	12
Centered Leverage Value	.052	.613	.250	.237	12

a. Dependent Variable: PM10_Mean

The minimum, maximum, and mean Mahalanobis distances displayed in Table 9-5 denote much lower distances, disregarding any possibility of outliers influencing the development of the linear regression model. The minimum, maximum, and mean Cook's distance is displayed in Table 9-5. Using the cut-off suggested of D greater than 1, no apparent problem cases arose since the maximum Cook's D is only 0.791. Despite not meeting the Fox (1991) criterion, linear regression is a robust technique that allows the variance in testing procedures or threshold parameters. That is why for this particular model, different thresholds were used to assess the significance of such and power predictability. The maximum centered leverage value of 0.613 shown in Table 9-5 for the summer model was smaller than the computed threshold from the small sample equation of 0.75, thus discarding the possibility of an undue influence from the

residuals. Figure 9-2 shows that the residuals histogram followed a bimodal distribution based on predicting PM_{10} mass concentrations from OND, TR5M and lc2m. This suggests that the error distribution is "skewed" by the presence of outliers. Since parameter estimation is based on the minimization of squared error, a few extreme observations may have exerted a disproportionate influence on parameter estimates.

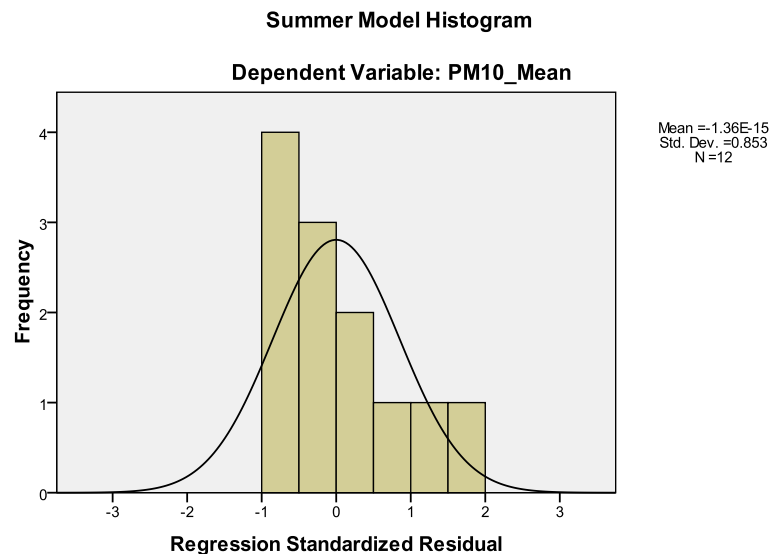


Figure 9-2. Residual's distribution of the summer model.

In addition, the Shapiro-Wilk's (W) test was performed over the model's residuals. Similar to the annual model, for the given variables, W test was not significant suggesting that the variable's distribution is not significantly different from normal, as shown in Table 9-6.

Table 9-6. Shapiro-Wilk's W test for normality of the summer model's residuals.

Summer Model Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.253	12	0.200*	.979	12	.985

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Therefore, Table 9-6 contradicts the residual histogram showing that OND, TR5M and lc2m residuals were normally distributed and that the bimodal shape merely corresponds to the variance in measured concentrations.

Figure 9-3, as obtained from the homoscedasticity test, illustrates a disperse pattern of the points with no trend. The distribution of the residuals among the standardized residuals axis demonstrate some ranges of errors below the negative standardized residual axis, while others ranged positive above the 0 value of the y-axis. Moreover, no sign of any outliers was spotted that may be causing this nonconstant error variance, and obviously no error that may be influencing positively or negatively the adjusted R^2 of the model. The scatter plot, however, reflects the fact that the model in this case has low standard errors of estimate explaining a small percentage of variance mainly as a consequence of an adequately fitted model.

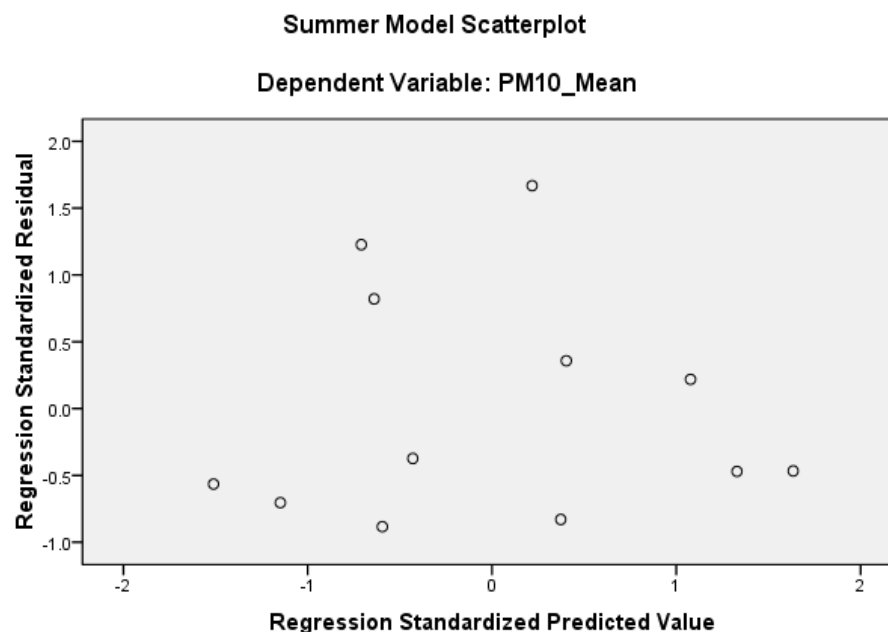


Figure 9-3. Scatterplot of the standardized residuals versus standardized predicted values for the summer model.

Therefore, the scatterplot of the standardized residuals versus standardized predicted values (Figure 9-3) substantiated the normality (performed previously) and homoscedasticity tests when testing the significance of the model and its parameters. The normality assumption was affirmed as the density of the plotted symbols were denser near the 0 value of the standardized residual axis (y-axis) and quickly abate away from the 0 value, keeping a symmetrical distribution on both sides of the 0 value of the. Likewise, homoscedasticity was attested as the spread of the error values, in the vertical dimension, were the same at any one point on a line as it is at any other point on that line. Nevertheless, the residual plot patterns confirmed the relationship between predictor and criterion as linear.

Heteroscedasticity, was also evaluated in Figures 9-4, 9-5, and 9-6. These cases were weighted by the reciprocal of their estimated point variance, and each was assessed with the goodness of fit of the weighted model. That is, weighted predicted is plotted on the on the x -axis and weighted residual on the y -axis. All three weighted residual plots produced a good fit; the residuals did not form a funnel shape but instead a uniformly distribution around the 0 line of the y -axis. Although Figure 9-5 shows the presence of a questionable weighted case for TR5m, after removing and resampling the TR5m weighted case the same distributed pattern was accomplished; determining that this observation had no influence in the model fit. This was substantiated with the standardized residual versus standardized predicted value scatterplot (Figure 9-3) where no outlier was spotted and hence corroborated the normality and homoscedasticity tests previously performed. The observation in question was also responsible for the high value shown in Mahalanobis, Cook's, and centered leverage distances shown in the residual statistics table, however it merely represented a potential outlier.

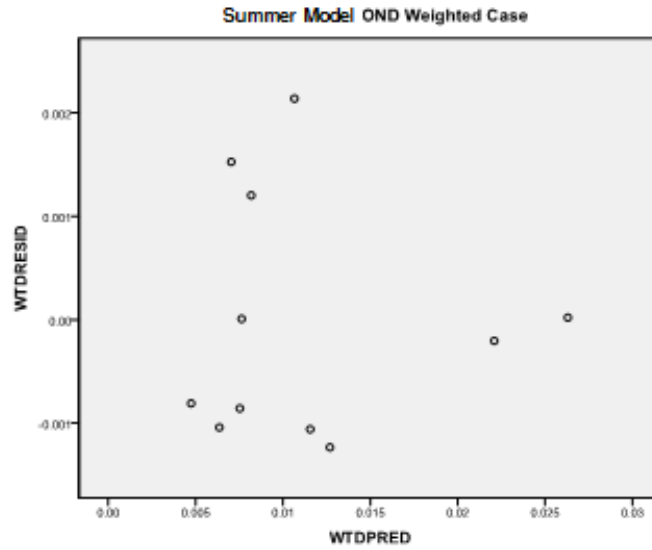


Figure 9-4. Weighted case of OND for the summer model.

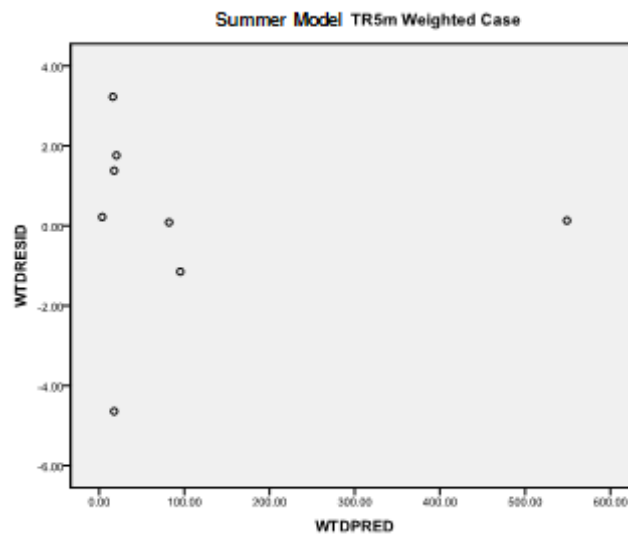


Figure 9-5. Weighted case of TR5m for the summer model.

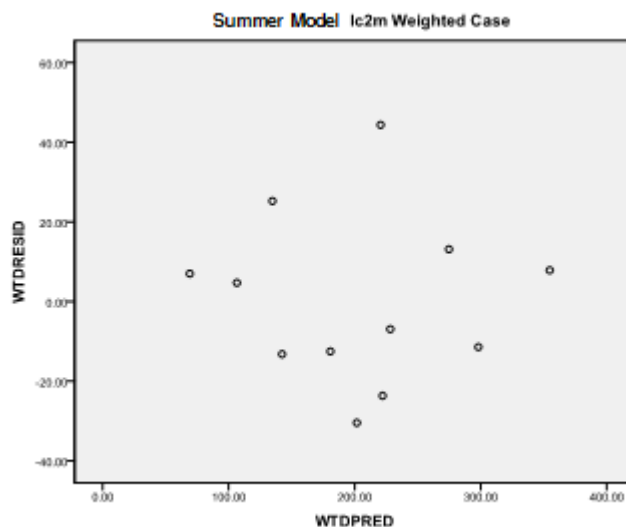


Figure 9-6. Weighted case of lc2m for the summer model.

9.4 CROSS-VALIDATION

Likewise the previous models, the bootstrap analysis was first undertaken to evaluate the sensitivity of the estimated model variance in the five models. In this interpolation technique, the 3528 predicted values were divided into 147 subsamples of 24 random samples. The variance was calculated as the square root of the sum of the squared differences between the subsample i -th average and the overall mean divided by the number of representatives minus one (3527). Supplementary to this method, histograms of the predicted parameter were validated and derived for 1000 iterations to examine the consistency of the standard variance as presented in APPENDIX F.

Secondly, we validated each model by a comparison of prediction errors obtained by simulating 1,000 jackknife data sets. For each data set, parameter estimates estimated from the interpolated model were obtained on a subset of 3528 locations, where each model was repeatedly parameterized on $N-1$ (3527) data points and then used to predict the excluded locations to predict PM_{10} .

Thirdly, the root mean square errors and the standardized root mean square errors were calculated as the square root of the sum of the squared differences of the observed concentration at site i and the predicted concentration at site i divided by 3528, the total grid nodes. In each of the cross-validations, the average difference was not greater than $9.458 \mu\text{g}/\text{m}^3$ for bootstrap, translating into an average relative difference of less than 7.0%. Also there was a recurrent estimate variance value of $0.0569 \mu\text{g}/\text{m}^3$ for the jackknife method, resulting in an error of 10%, and $9.964 \mu\text{g}/\text{m}^3$ for RMSE, which intuitively contrasts the measured values with the predicted values within an 7% difference. The results of this comparison are presented in Table 9-7.

Table 1 Cross-validation results of the five interpolation techniques used to map summer PM_{10} mass concentration throughout the El Paso cohort study.

	Summer Model			
	Bootstrap Variance	Jackknife Variance	RMSE	SRMSE
IDW	9.366721	0.052247	9.604	N/A
NN	9.381451	0.053479	9.629	N/A
OK	9.255762	0.050121	9.463	0.8076
SPLINE	9.458566	0.056977	9.964	N/A
UK	8.793424	0.048024	9.274	0.8197

Amongst the modeling techniques, the predicted concentrations generated by the SPLINE method had the worst performance, with a bootstrap variance of $9.458 \mu\text{g}/\text{m}^3$, a variation in prediction certainty over 1000 iterations of $0.0569 \mu\text{g}/\text{m}^3$, and a root mean square error of $9.964 \mu\text{g}/\text{m}^3$; although the bootstrap histogram preserved normality, only corresponds to a constant variance over 1000 iterations, as seen in APPENDIX F. In this case, the deplorable prediction performance was caused by the fundamentals of the model as it created the surface specifically so that it passes through the control points having the least possible change in slope at all points, fitting the control points with a minimum curvature surface. In other words, the model tends to average to its utmost capacity the PM_{10} exposure throughout the cohort study. Thus, coinciding

with the visual pattern (shown in APPENDIX H) were the incongruous blemishes, of higher concentration at lower polluted areas, broadens more frequently than in the rest of the interpolation techniques. Considering the fact that the summer PM₁₀ surface tends to have a more disrupted variation pattern overall, when compared to the annual and spring models. Although much more conglomerated towards the U.S.-Mexico border.

NN likewise also performed badly, as anticipated from the visual inspection, with a bootstrap variance of 9.381 $\mu\text{g}/\text{m}^3$, a RMSE of 9.629 $\mu\text{g}/\text{m}^3$ and a fractional bias obtained from the jackknife analysis of 0.0534 $\mu\text{g}/\text{m}^3$. In addition, the model average variation showed a deplorable bootstrap histogram displaying an alteration of the normal distribution with a bimodal shape. One reason the model tended to misrepresent the spatial variability of PM₁₀ worse than the IDW and kriging methods was because the NN method assigns no data values to the cell center of the perimeter cells of the output raster that fall outside the convex hull, which is defined by the outermost extent of the points. Additionally, influential points were disproportionately weighted over less influential points, especially in the proximity of Western Refining, accordingly to the Voronoi's intersecting triangles that cause more variance between the measured values and the predicted values and thus the amiss performance of the natural neighbor method.

Sequentially, IDW performed better than the latter two methods described above, with a bootstrap difference of 9.366 $\mu\text{g}/\text{m}^3$, a fractional bias of 0.0522 $\mu\text{g}/\text{m}^3$, and a RMSE equal to 9.604 $\mu\text{g}/\text{m}^3$ and far less attenuation. It is important to remark that NN and IDW are very similar methods, mainly because they both rely on the weight distance average. However, whereas NN assigns weight based on the percentage of overlap, IDW assigns weights based in their similar distance from the interpolation point. This could be ascertained as both NN and IDW concurred

in their visual pattern, shown in APPENDIX H, not to mention their similarities on their cross-validation results to which they were put to attest. And although IDW outperformed NN, the model still had some flaws with respect to its ability to predict accurate PM concentrations. A reason for that is attributable to the fact that the IDW average weighted distance cannot be greater than the highest or less than the lowest input values, contradicting or misrepresenting the elevated concentrations measured at the vicinity of Sunland Park Mall. In fact, the bootstrap histogram shows a modest bimodal shape traceable to the surrounding grid points located in vicinity of WR, just one mile from the US-Mexico border. This point had 1.2 standard deviations away from the concentration mean of all samples falling into the 95% confidence interval. However, there is no reason to believe that the underlying data from this station is inaccurate, as the value showed an 11% increment from the value measured at Mimosa site (K2) which followed the same protocol as the three monitoring sites correlated with the State fix stations, and therefore the monitor was not excluded.

Lastly, the best performance was seen for both universal and ordinary kriging approach for which the bootstrap variance equals to $8.793 \mu\text{g}/\text{m}^3$ and $9.255 \mu\text{g}/\text{m}^3$ respectively, the jackknife variance for universal kriging prompted a value of $0.048 \mu\text{g}/\text{m}^3$ as well as $0.0501 \mu\text{g}/\text{m}^3$ for ordinary kriging. For the ELP using 3-yr data the RMSE for UK and OK resulted in an absolute value of $9.274 \mu\text{g}/\text{m}^3$ and $9.463 \mu\text{g}/\text{m}^3$, respectively; by contrast, the SRMSE on these predictions was $0.8197 \mu\text{g}/\text{m}^3$ and $0.807 \mu\text{g}/\text{m}^3$. Nevertheless, the histogram of the predicted estimates derived for each of the 1000 iterations suggested in both studies that the models were robust. Albeit both kriging methods showed a slight bimodal shape, further investigation revealed that the inclusion of a single sample (located at the area circumvent of WR) resulted in a 13% change in the parameter. This sample location (also disproportionately weighted over less

influential points in the NN and IDW method) had modest commercial area in a 200 m buffer and also the considerably moderate PM_{10} concentration value. Overall, however, the parameter shift had a little but still representative surface map effect on predictions. Replacing the current coefficient value (1.829) with the mean value of the right tail (2.37) yielded predictions that were, on average, just 2% different. For most metrics, the models derived from the two runs are broadly similar, indicating considerable robustness in the methods in the face of varying input data. There is some tendency, however, for the performance of the universal kriging model to be rather better than the ordinary kriging, indicating local stationarity and the sensitivity of the method to the underlying input data. Kriging gives the preferred cross-validation values at the grid nodes, but these generally reflect the way in which the kriged surfaces are fitted to those points. Semivariograms (shown in APPENDIX G) show strong spatial dependency in the data, with modest nugget variance compared to the mean concentrations and an indistinct sill, and thus suggest that the models predicted concentrations well at unsampled locations. Therefore the overall cross-validation diagnostics concludes that kriging methods demonstrated that the influential points are unlikely to drive the regression model and that universal kriging feebly outperforms ordinary kriging in predicting spatial characterization. All because ordinary kriging assumed that the variation in PM_{10} is free of any structural component. The universal kriging variogram (see APPENDIX G) exhibited a more suited model with smaller range and sill than the variogram for ordinary kriging (see APPENDIX G). The unusual value in land use commercial area, for example, inflated and deflated the model prediction power when compared to the observations. Ordinary kriging, as a result, oversmooths and hence misses areas with unusually high and low particulate matter caused by omitting an external drift. It is also possible that ordinary kriging in this context violates the fundamental assumption of stationarity. Points

close to commercial areas, for example, exhibited a different relationship between PM_{10} and commercial areas rather than points, for example, in less developed residential areas far from any major source of PM_{10} . These results indicate that the difference in temporal variation produced relatively moderate errors when comparing measured values to modeled values.

9.5 DISCUSSION

The summer model was constructed using 148 weekly concentrations from field monitoring sites in the El Paso area. However, the Montanavista site had to be withdrawn from the analysis for the same reasons discussed before. In the first iteration of developing a city-specific LUR model the sum of the designated truck road's length of the sampling location was the most consistent predictor of PM_{10} levels, accounting for over 70% of the variation. When combining this variable with the nearest distance to the border and commercial land use type, the model was able to explain almost 74.6% of the variation in the PM_{10} concentrations. Predictors directly based on traffic counts were of primary interest in the analysis. VMTs per buffer were secondly preferred to the sum of road's length, because they seem to reflect the emissions and dispersion process more closely. These variables usually showed lower correlations with the pollutant concentrations than other variables of comparable scale. In general, the sum of the trucks roads' length describing traffic intensity appeared to have greater explanatory power than those variables describing distance to the nearest road, traffic counts and even VMTs. Truck traffic, however, has been proved to be influenced by particles suspected to originate mainly from diesel exhaust engines (Cyrus et al., 2003). The increased demand for freights during August is traceable to some industries depleting their budget's surplus prior to the sealing of their fiscal year and to the highly economic demand. The location of El Paso generates a vast

amount of trafficked freights between states and even between countries. Truck traffic broadens either by freighters going into, going out and even going through the city as a consequence of the international trade between both countries. Expanding trade and truck traffic throughout the region due to the North American Free Trade Agreement has place a strain on the local airshed. With six international ports of entry throughout the PdN region, from which one (BOTA) has the heaviest combined traffic of heavy-duty and gasoline fueled vehicles among the U.S.-Mexico ports of entry (Wilbur and Smith, 2006), it is reckoned that over one million freight carriers enter the U.S. throughout this crossing each year. Therefore, the sum of all designated truck road's length within a 500 meter buffer area collaborated the presumption that truck traffic induces variations in PM_{10} background concentrations. Consistently, the $PM_{2.5}/PM_{10}$ ratio for this particular time period yielded a modest value of 0.20 which derived from an increase in averaged PM traffic-based emissions. The predictability of traffic density is not only due to the actual traffic emissions but also to its association with many other geographic characteristics, for example, the presence of industrial sources and commercial areas; which notably influence the levels of concentrations as well. Since the designated truck road does not only embraces truck traffic but also vehicle traffic; this results correspond to previous studies, which suggested spatial gradients of vehicle emissions in the Paso del Norte region (Jeon et al., 2001; and Noble et al., 2003). In a second iteration, the nearest distance to the border was an influential parameter in the initial model as the proximity to Juarez has often aroused the speculation of a source for air pollution. The closeness of CDJ highly industrialized areas, brick kilns, unpaved roads, and the traffic activity of antediluvian vehicles to the border has proxy to be a contributor of PM_{10} concentrations for the El Paso region. Moreover, a study performed in the ELP region formulated a LUR model to predict the intra-urban gradient of NO_2 . They cited that the inclusion of the

proximity to a port of entry, distance to a major highway, and elevation to the model explained 81% of the variation of the NO₂ gradient for El Paso (Gonzales et al., 2005). Therefore, this model corroborated the findings of Gonzales et al. (2005), that the proximity to Juarez and traffic-based parameters proxy higher concentrations in the cohort study; and that the intra-urban gradient increases as the distance to CDJ decreases. Although elevation was not a contributor in the developed model, it was substantiated to have an explanatory power over the concentrations of PM₁₀. Even after elevation showed the highest bivariate correlation coefficient (75%) in the summer model, however, this parameter was not included in the model as it produced inconsistent coefficients as well as spatial autocorrelation when a second and third variable was added to the model. Subsequently, commercial land use type within a 200 meter buffer area was included in the model in a stepwise iteration as a zoning-based variable and was able to assist in the significance predictability of the model. This variable yielded a bivariate correlation of 0.404 between PM₁₀ concentrations and lc2m. Commercial density served as a surrogate for the general level of human activity in the vicinity of a monitoring site. Such density has been associated with increased traffic activity and decreased driving speeds, hence increased emissions, suggesting that with higher density areas may be subject to higher emissions by human activity. Also, the segregation of commercial land use type creates the need for travel between residential and commercial. Taken together, these characteristics increase the exposure surface with more spatial variability in the ELP urban area.

Additionally, the fitted model was tested for non-linearity, spatial autocorrelation, collinearity, homoscedasticity and normality. Despite the fact that all of the assumptions were met, a questionable observation was detected in the weighted case for the sum of truck road's length under the homoscedasticity test; yet after excluding the observation and the weighted case

resample it was determined that the observation had no influence neither in the sparse trend nor the equality of the observations above and below the weighted residual axis. Therefore, the questionable observation was maintained in the analysis as it represented a merely potential outlier. After all assumptions were met it was determined that this observation was solely responsible for the highly recorder Mahalanobis, Cook's, and centered leverage distances value. Consequently, a bootstrap to evaluate the sensitivity of the parameters to the actual samples included in the modeling suggests that the model is relatively stable to the choice of samples. In general, these variables were robust to sample selection, though it was found that land use commercial and the length of truck roads variables were affected by the inclusion or exclusion of the Montanavista sample that had modest PM_{10} concentrations. Nevertheless, the fact that land use commercial (200 m) and length of truck roads (500 m) in the El Paso city borough, using an entirely different GIS layer and without the influence of the Montanavista sample, was more significant lends support to the important role played by commercial zoning and length of truck roads for this particular model.

Consequently, after deriving the parsimonious operational regression model a grid covering the cohort study and containing 3528 nodes was developed in order to sharpen the resolution of the depicted overall PM_{10} trend. At each node the concentration of PM_{10} was calculated applying the three parameters with their corresponding coefficient and joined into a single attribute table. It is important to remark that seven predicted values had to be withdrawn from the analysis as the highest concentration estimates were 120% higher than the maximum sample concentration measured at McCombs (D2). These cluster values were map along the cohort study and were giving inconsistent characterization of PM gradients. The truncated LUR predicted values corresponded to a miscalculation by the analysis tool in ArcGIS of the

intersected truck road's lengths for the 500 m buffer. Further analysis was performed to the surrounding grid nodes in order to detect any more flaws but despite the rigorous authentication, the points gave no further reason to believe the truck roads length value was confounded and by this problem, and on the whole, the named values do not affect the main findings of the analysis. Given the strong predictive power of hybrid models (Jerrett et al., 2005), five interpolations methods (IDW, NN, OK, UK, AND SPLINE) were combined with LUR to generate a confident PM_{10} spatial surface. The validity of three models was challenged by three procedures (bootstrap, jackknife, and RMSE) whereas the remaining two models were assessed by an additional method (SRMSE). The bootstrap cross validation procedure for UK yielded a variance over the subsample average of $8.793 \mu g/m^3$ in contrast to the second lowest variance of $9.255 \mu g/m^3$ for the OK method. The fractional bias of $0.0480 \mu g/m^3$ for UK was smaller when compared to the second lowest of $0.0501 \mu g/m^3$ prompted by the OK analysis. The predictability of universal kriging was more precise with a RMSE of $9.274 \mu g/m^3$ compared to $9.463 \mu g/m^3$ for the ordinary kriging. Lastly, the SRMSE value for UK was closer to one than the OK meaning that the estimated standard error was more reliable or valid for the uncertainty of the predicted values. Overall universal kriging based on Gaussian model outperformed IDW, NN, SPLINE, and UK on cross-validation results using the full set of samples. The cross-validation procedure, for example, yielded a bootstrap and jackknife variance, which was small compared to the range of concentrations and not greater than the RMSE, indicating that the influence of the single observations was not too heterogeneous. In addition, 84% of the UK modeled estimates were within a factor of 1.2 times the range of the measured concentration with an IQR value of $18.682 \mu g/m^3$ approximately 3% less than the second lowest value cued by OK and 5% to the SPLINE method. The ELP universal kriging model exhibited a bias less evident, but still existing, in the

full model that included the validation samples overpredicting 42% of the value locations and underpredicting 21% with a relatively low average absolute value residual or MAE of 5.544 $\mu\text{g}/\text{m}^3$; therefore the overall model tended to overestimate PM_{10} concentrations. The underpredicted values manifested more pronounced in two urbanized areas of the cohort study: 1) in the vicinity of Cielo Vista Mall where commercial areas predominate as well as being in close proximity to the I-10 where a vast amount of truck traffic traverse, and 2) the area circumvent to Bassett Mall, which also has a thorough amount of commercial areas in addition to its closeness to Montana Ave. and I-10. Intuitively the model misrepresented the PM concentrations in both areas as 3 out of the 21 truncated predicted values that were excluded from the analysis (1 in the circumvent area of Cielo Vista Mall and 2 from the Bassett Mall area) measured concentration levels that were below 120 % of the sample concentrations. For that reason, the model relied on the surrounding measured neighbors to predict concentrations, which were proportional to these values based on the interpolation technique's fundamentals. Conversely, the overprediction appeared at the intersection of Cann Ave. and N. Zaragoza Rd. (in the southeasternly part of El Paso) where there is rather more agricultural land use than commercial type and thereby less trafficked by freighters. This overpredicted area was conspicuous in the projected map as the predicted concentrations ranged from 30.1 to 40 $\mu\text{g}/\text{m}^3$ in an area where the surrounding concentrations ranged between 20.1 to 30 $\mu\text{g}/\text{m}^3$. All of the observed overpredictions were upwards 60 $\mu\text{g}/\text{m}^3$ due to unregulated fuel combustion. Truck traffic is one of the leading sources of particulate air pollution in the El Paso Basin. Surrounding the ports of entry is a network of roads connected to the freeway used to disperse the cargo from suppliers to its distribution and storage centers. There are more transport trucks, which combust diesel fuel, traveling on the I-10 than any other major road in the El Paso basin. Westerdohl et al.

(2005) found that measured freeway concentrations of PM had a range of 60 to 820 $\mu\text{g}/\text{m}^3$ on the 10 east freeway, and that concentrations along major roadways with high traffic density were up to 20 times higher than residential concentrations. Thus, the overpredicted values in this analysis were at plausible levels. Categorizing the sites by tertiles, 73% of them would be classified identically. Only 27% of the grid nodes sites near the I-10 would change between the extreme tertiles. The divergences might be partially explained by the influence by truck traffic diesel emitted particles. When the predicted exposure was divided by the predicted quartiles it was estimated that 7% of El Paso residents fall into categories for PM_{10} concentrations within 30.1 to 50 $\mu\text{g}/\text{m}^3$ during the summer season. Conversely, 2.3% of the residents in ELP exhibit an exposure increase of 10 $\mu\text{g}/\text{m}^3$ contrary to 8% of the residents who generally experience an exposure decrease between the spring and the summer season. Both spring and summer seasons measured concentrations somehow similar in range, however, the characterization in spring was sparser than the summer model where the characterization was more conglomerated towards the border.

Finally, in this study LUR method in combination with universal kriging successfully predicted small-scale spatial variation of ambient PM_{10} concentrations with no spatial irregularities and adequately reproduced the regional trend identified by the monitoring network. The model predicted the lowest values at the upper region of the west side, northeast, and the southeastern area of the cohort study. Such areas represented inferior commercial areas, farther distance to the border and lesser influence by the El Paso major roads. That is, a less marked western and northeastern trend whereas a steeper concentration gradient southward towards the international border with CDJ. Although, Ciudad Juarez were shown to influence the spatial PM_{10} distribution in El Paso it was impractical to monitor the entire metropolitan area due to cost

and political constraints in order to quantify the impact of CDJ. This particular time period implied less variation in the PM spatial gradient. In fact, the concentrations for nearly the entire region (80%) remain at a low range between 10 to 20 $\mu\text{g}/\text{m}^3$. The higher concentrations, though, were mapped at different locations especially at areas close to the freeway with high commercial activity. As shown in Figure 9-7 the model reproduced three noticeable hot spots; first high concentration spot aroused at the surroundings of downtown core spurring concentrations between 55.945 and 82.4939 $\mu\text{g}/\text{m}^3$. Second high concentration area prompted at the junction of I-10 and U.S. Highway 54 (Spaghetti bowl) with concentrations ranging from 59.931 to 66.519 $\mu\text{g}/\text{m}^3$. And finally the area surrounding Yarbrough Plaza adjacent to I-10 which predicted concentrations ranging from 50.276 to 86.915 $\mu\text{g}/\text{m}^3$. The outcome bearded the short-term spatial variations of averaged PM_{10} concentrations for a localized area within larger metropolitan region and predicted a gradient decreasing with distance from the border. Moreover, the methodology captured the relative impact of emissions outside the localized monitoring network area and allowed for an optimal utilization of resources.

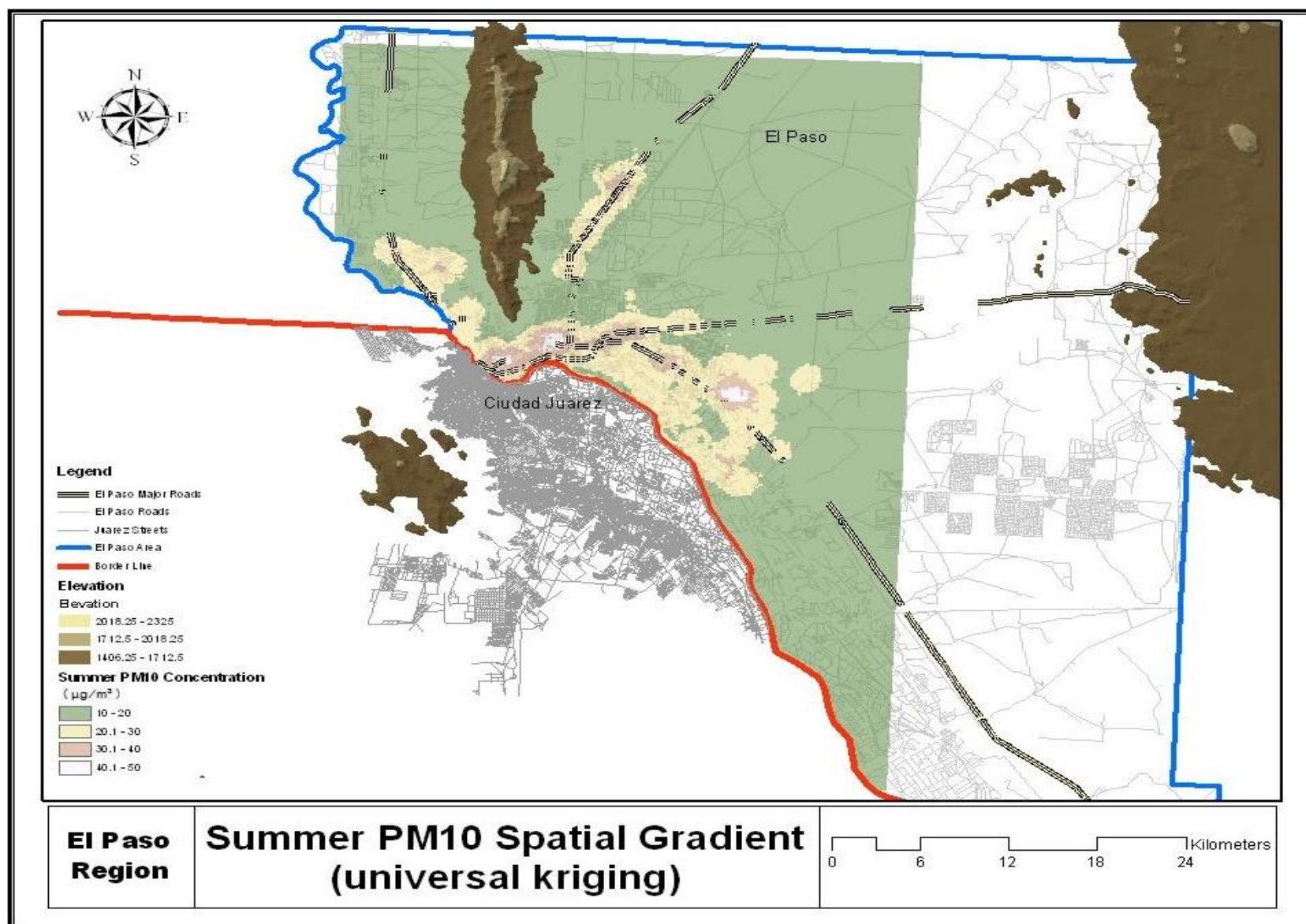


Figure 9-7. Summer PM₁₀ spatial gradient using Universal Kriging interpolation method.

10 FALL MODEL

10.1 PM₁₀ MEASUREMENTS

During this sampling period, 285 ambient PM₁₀ weekly averaged concentrations were measured representing 91% of a total of 312. Descriptive statistics of the air pollution measurements are shown in Table 10-1.

Table 10-1. Descriptive statistics of the average fall concentrations for the 13 monitoring stations.

Fall Descriptive Statistics for PM ₁₀ Concentrations (µg/m ³)					
	N	Minimum	Maximum	Mean	Std. Deviation
Lindbergh	12	11.8	46.4	26.6	9.4
Fountain	24	10.4	46.9	23.2	10.2
Skyline	13	10.1	45.0	26.6	9.7
McCombs	24	8.7	27.5	18.7	4.6
UTEP	34	13.2	72.5	34.0	14.0
Chamizal	12	14.8	111.0	47.7	26.7
Parral	25	10.6	40.6	22.2	7.9
Cosmos	35	12.1	68.9	25.0	13.3
Socorro	35	24.2	95.9	45.2	17.0
Plant	13	1.8	124.6	54.2	32.9
Mimosa	20	17.0	122.2	51.2	27.9
Montana	25	10.1	46.8	25.4	10.6
Pebble	13	9.8	35.0	22.7	6.8

Weekly averaged concentrations were depurated by removing values above 150 µg/m³ and greater than three standard deviations as well as concentrations underneath it from their corresponding site. The detection limits for PM₁₀ mass concentrations were defined by three standard deviations of the mass concentration deposit on field blanks. The sample measurements averaged 35.53 µg/m³ with a standard deviation of 12.49 µg/m³. In contrast, fall measurements

recorded higher concentrations and more variation than the concentrations observed during summer and spring period, as the temperatures begin to decrease, more perpetual stagnations events are perceived and thus limit the dilution of air contaminants. From the 285 sampled weekly concentrations, 42 and 27 observations were at least 1 standard deviation above or below the mean, respectively. All the measured values exhibited a normal distribution as shown in APPENDIX C. The maximum fall seasonal average PM_{10} concentration of $54.24 \mu g/m^3$ was observed at the Plant site near the border and a minimum of $18.73 \mu g/m^3$ at McCombs site, which is in the farthest north. The overall highest weekly averaged PM_{10} concentration of $124.61 \mu g/m^3$ was measured at Plant site, located near the border highway in the eastern part of El Paso. Conversely, it also coincides with the overall lowest concentration of $1.83 \mu g/m^3$. An evident southward trend with the highest values near the border was observed as well as a less marked northeastern and westward trend, as seen in Table 10-1. The average $PM_{2.5}/PM_{10}$ ratio calculated from the dichotomous samples, averaged 0.22 with a standard deviation of 0.131.

Similar to the models, the sites produced a value of 0.00 ($p < 0.0005$) suggesting minimal spatial autocorrelation from the Morans' I test.

10.2 REGRESSION ANALYSIS

Each of the 190 independent variables was tested through the methods previously mentioned using the regression analyst tool in SPSS 17 and PASW Statistics 18. After conducting nearly 33 different models, no acceptable model fit was obtained nor was a concordance between variables and their corresponding coefficient signs. Thus, OND was discarded from the analysis and the subsequent highly correlated variable was inputted into the model in the same procedure. Similar to the latter models Montanavista (site L) was excluded

from the analysis as it was generating inconsistencies in the model. Consequently, log transformed truck road's length (ntr1k) was entered into the model. Then, additional variables were paired with other significant variables in a series of trivariate regression models. This series of analysis identified the pairing with the log-transformed population density within each monitor 1500 m buffer area (npop15k). Lastly, industrial commercial and residential areas were comprised into a unique normalized variable (NGZ3m) that was found to be associated within a 300 meter buffer area. Table 10-2, provides a model summary and the ability of the three variables (NGZ3m, ntr1k, and npop15k) to predict ambient PM₁₀ mass concentrations (indicated by column 1).

Table 10-2. Land use regression model summary of the fall iteration.

Fall Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.937 ^a	.877	.831	5.28

a. Predictors: (Constant), NGZ3m, ntr1k, npop15k

Similarly to the previous models, 83.1% of the variability of ambient PM₁₀ concentration in the El Paso area during fall is explained by NGZ3m, ntr1k, and npop15k with a standard error estimate (s) of 5.28 µg/m³. This means, that the model could expected to predict 83.1% of the ambient PM values to within approximately ±10.56 µg/m³ (or 11%) at a 95% confidence with variables like NGZ3m, ntr1k, and npop15k.

The Table 10-3 shows the statistical coefficients for the fall model. As seen in Table 10-3, none of the VIF values surpass the rule of thumb of 2.5, thus suggesting no multicollinearity among the three variables. Similarly, the tolerance values reaffirmed the certainty that none of

the variables were intercorrelated and more importantly that the concentrations estimated by the model may produce lower deviations from the actual fitted regression plane.

Table 2-3. Statistics of the model's coefficients for the fall iteration.

Fall Model Coefficients^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	1.636	6.192		.264	.798		
npop15k	.015	.005	.494	3.142	.014	.621	1.610
ntr1k	4.875	1.622	.404	3.006	.017	.848	1.179
NGZ3m	41.311	6.732	.904	6.136	.000	.706	1.416

a. Dependent Variable: PM10_Mean

A further assessment for multicollinearity in the model was performed , as multicollinearity may occur even when the rules of thumb for tolerance and VIF were met. Therefore, collinearity diagnostics (Table 10-4) was an alternative method for assessing if there is too much multicollinearity in the model. To simplify, cross products of the independent variables are factored. There will be as many factors as independents, plus one (for the constant). The first column addresses the dimension representing the variables included plus one constant variable. The ordering of the variables corresponds to the same arrangement shown in the variance proportions. That is, the constant variable corresponds to the first dimension, npop15k to the second dimension and so forth. The subsequent column identifies the eigenvalue which is a term related in the field of linear algebra and evaluates linear transformations represented as matrices acting on vectors. The eigenvalues shown in the Table 10-4 unfolds little variance in the cross product matrix, as high eigenvalues indicate dimensions (factors) which account for much of the variance in the cross product matrix; eigenvalues close to 0.0, such as in this case, indicate

dimensions which explain modest variance. Values such as 0.032 for NGZ3m, 0.292 for ntr1k, and 0.671 for npop15k obviously underwent the rule of thumb with dimensions above the 0.00 value.

Table 3. Collinearity diagnostics of the model's parameters for the fall iteration.

Fall Model Collinearity Diagnostics ^a							
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	npop15k	ntr1k	NGZ3m
1	1	3.005	1.000	.01	.01	.03	.02
	2	.671	2.116	.00	.01	.05	.55
	3	.292	3.208	.04	.03	.85	.02
	4	.032	9.706	.96	.95	.07	.41

a. Dependent Variable: PM10_Mean

With the highest condition index of 9.70 prompt by NGZ3m, it was clear that the model had no collinearity problems making it unnecessary to examine variance proportions.

Lastly, after determining the absence of collinearity, PM₁₀ concentrations were determined to be influenced by the normalized population in a 1.5 km buffer area, the normalized sum of the truck roads' length at a 1 km buffer, and normalized group zoning within a 300 buffer area as shown in Equation 10-1.

$$PM_{10} = 1.636 + 41.311 \cdot (NGZ3m) + 4.875 \cdot (ntr1k) + 0.015 \cdot (npop15k) \quad (10-1)$$

10.3 MODEL VALIDATION

After determining the appropriate model, standard regression diagnostics were applied to assess the problems such as influential outliers, linearity, independency, normality, and homoscedasticity. Standard multiple regression can only accurately estimate the relationship

between dependent and independent variables if the relationships are linear in nature. Therefore if violations of linearity are extremely serious (fitting data to a linear model which are nonlinearly related), the predictions are likely to produce systematic errors, especially when generating usually large or small predictions.

Figure 10-1 unveils a correlation coefficient of 0.877, and a symmetrically point distribution around a diagonal determining a linearity relationship between PM₁₀ mass concentrations and NGZ3m, ntr1k, and npop15k. Note that, the standardized predicted values were calculated using the model equation for every particular monitoring site previously discussed. In addition the statistical software suited for this research produce such values in a residual statistics table.

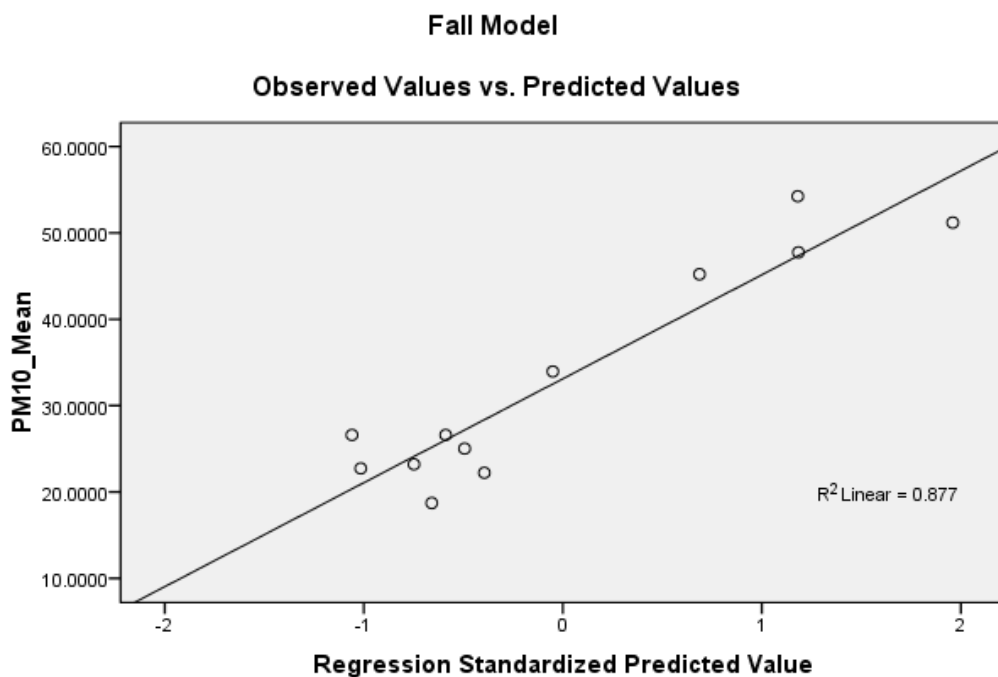


Figure 4 Linearity test between the sample concentrations and the predicted values for the fall model.

In addition to the Moran's I test, performed previous section, a second evaluation was performed under the Durbin-Watson test. Thus using the test statistics shown in Equation 7-3, a

corresponding d value of 2.651 was achieved. However, obtaining the $d_{\text{lower critical}}$ value (0.658) and the $d_{\text{upper critical}}$ value (1.86) at 5% significance from APPENDIX E, the test rejects the null hypothesis of positive autocorrelation and thus accepts the alternative hypothesis demonstrating that the error estimates are not positively autocorrelated. Nonetheless, researchers have point out those Durbin-Watson values above 1.5 proved to have no autocorrelation among the residuals.

Table 10-5 contains summary data regarding the residuals (the difference between predicted and actual values). Standard residual (Std. residual) for instance is the standardized residual raw residual divided by the standard deviation of residuals. Since the minimum standardized residual is -1.227, at least one prediction is more than one standard deviation below the mean residual. Studentized residuals are very similar to standardized residuals but follow the t distribution.

Table 10-5. Descriptive statistics of the fall model's residuals.

Fall Model Residuals Statistics					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	20.368654	56.705864	33.123925	12.0365148	12
Std. Predicted Value	-1.060	1.959	.000	1.000	12
Standard Error of Predicted Value	1.919	5.058	2.897	.983	12
Adjusted Predicted Value	18.867620	64.455894	32.888220	12.8494046	12
Residual	-6.4711018	6.9106584	.0000000	4.4987744	12
Std. Residual	-1.227	1.310	.000	.853	12
Stud. Residual	-1.620	1.691	.009	1.059	12
Deleted Residual	-13.2583961	11.5130644	.2357050	7.2615579	12
Stud. Deleted Residual	-1.849	1.973	.002	1.158	12
Mahal. Distance	.540	9.194	2.750	2.603	12
Cook's Distance	.001	.723	.178	.270	12
Centered Leverage Value	.049	.636	.250	.237	12

a. Dependent Variable: PM10_Mean

The minimum, maximum, and mean Mahalanobis distances displayed in the table above denote much lower distances, disregarding any possibility of outliers influencing the development of the linear regression model. The minimum, maximum, and mean Cook's distance is displayed in Table 10-5. Using the cut-off suggested of D greater than 1, no apparent problem cases arose since the maximum Cook's D is only 0.723. Despite not meeting the Fox (1991) criterion, linear regression is a robust technique that allows the variance in testing procedures or threshold parameters. That is why for this particular model, different thresholds were used to assess the significance of such and power predictability. Finally, the maximum leverage value of 0.636 is shown in Table 10-5 was smaller than the computed threshold from the small sample equation of 0.75, thus discarding the possibility of an undue influence from the residuals.

Error, represented by the residuals, should be normally distributed for each set of values of the independents, as violations of normality compromise the estimation of coefficients and the calculation of confidence intervals. Conversely, violations of normality compromise the estimation of coefficients and the calculation of confidence intervals. Calculations of confidence intervals for coefficients are based on the assumptions of normally distributed errors. In Figure 10-2 shows the residuals histogram followed a bimodal distribution based on predicting PM_{10} mass concentrations from NGZ3m, ntr1k and npop15k.

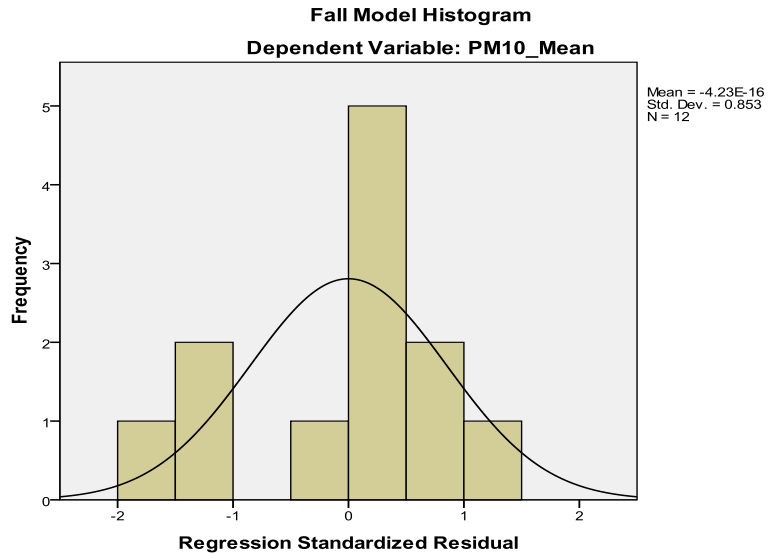


Figure 2 Residual's distribution of the fall model.

In addition, the Shapiro-Wilk's (W) test was performed over the model's residuals. For the given variables, W test was not significant suggesting that the variable's distribution is not significantly different from normal, as shown in Table 10-6.

Table 5 Shapiro-Wilk's W test for normality of the fall model's residuals.

Fall Model Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.140	12	.200 [*]	.993	12	.536

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Therefore, Table 10-6 contradicts the residual histogram showing that OND, TR5M and lc2m residuals were normally distributed and that the bimodal shape merely corresponds to the variance in measured concentrations.

Figure 10-3, illustrates a disperse pattern of the points with no trend, as obtained from the homoscedasticity test. The distribution of the residuals among the standardized residuals axis demonstrate some ranges of errors below the negative standardized residual axis while others ranged positive, above the 0 value of the y-axis. Moreover, no sign of any outliers was spotted that may be causing this nonconstant error variance, and obviously no error that may be influencing positively or negatively the adjusted R^2 of the model. This outlier certainty is demonstrated in Figure 10-4, as the boxplot did not identify any outliers at two standard deviations from the mean. The scatter plot, however, reflects the fact that the model in this case has low standard errors of estimate explaining a small percentage of variance mainly as a consequence of an adequately fitted model.

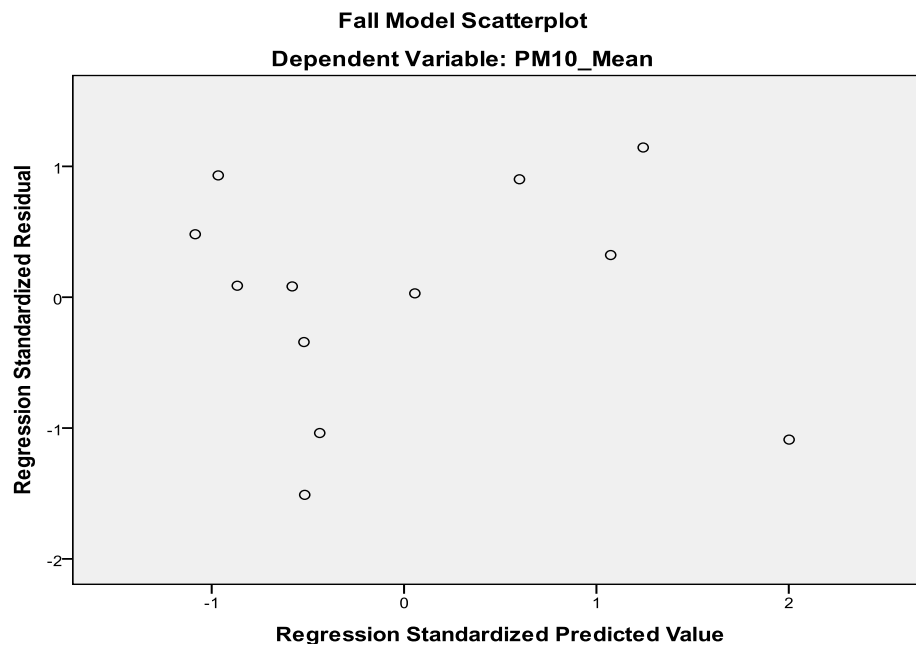


Figure 3 Scatterplot of the standardized residuals versus standardized predicted values for the fall model.

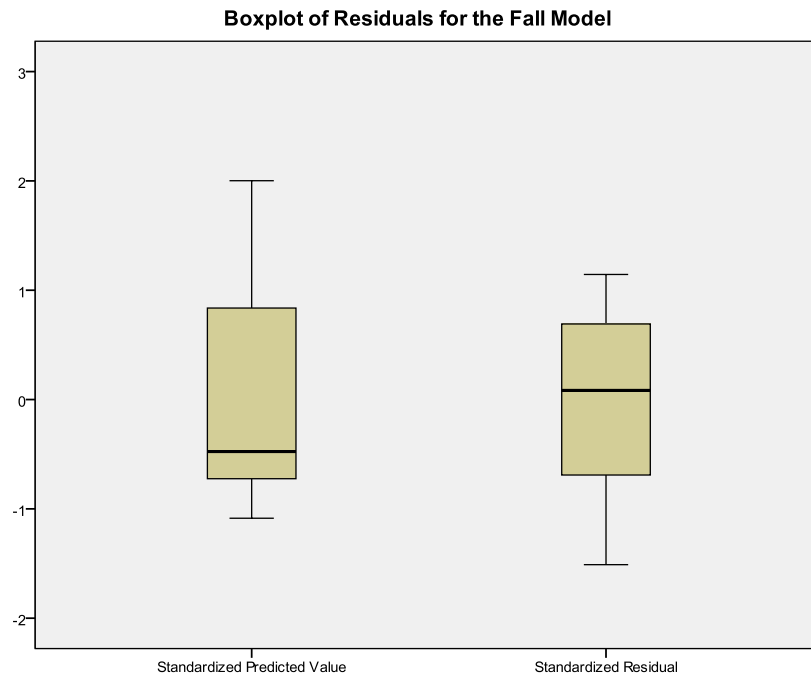


Figure 4 Boxplot of the standardized predicted values and standardized residuals for the fall model.

Therefore, the scatterplot of the standardized residuals versus standardized predicted values (Figure 10-3) substantiated the normality (performed previously) and homoscedasticity tests when testing the significance of the model and its parameters. The normality assumption was affirmed as the density of the plotted symbols were denser near the 0 value of the standardized residual axis (y-axis) and quickly abate away from the 0 value, keeping a symmetrical distribution on both sides of the 0 value of the. Likewise, homoscedasticity was attested as the spread of the error values, in the vertical dimension, were the same at any one point on a line as it is at any other point on that line. Nevertheless, the residual plot patterns confirmed the relationship between predictor and criterion as linear.

WLS regression causes events with smaller residuals to be weighted more thoroughly in calculating the x coefficients in order to diminish the impact of heteroscedasticity. Figures 10-5 through 10-7, cases were weighted by the reciprocal of their estimated point variance, and each

assesses the goodness of fit of the weighted model. That is, weighted predicted is plotted on the x -axis and weighted residual on the y -axis. All three weighted residual plots produced a good fit; the residuals did not form a funnel shape but instead a uniform distribution around the 0.0 line of the y -axis. Therefore, standard errors will be small but estimates will be very similar.

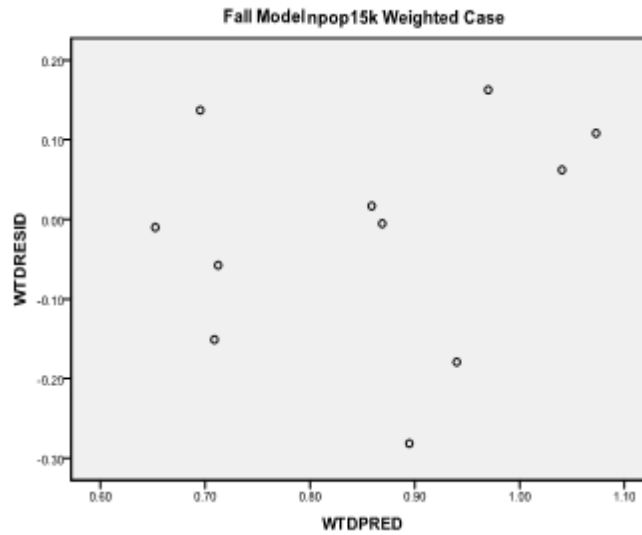


Figure 6 Weighted case of npop15k for the fall model.

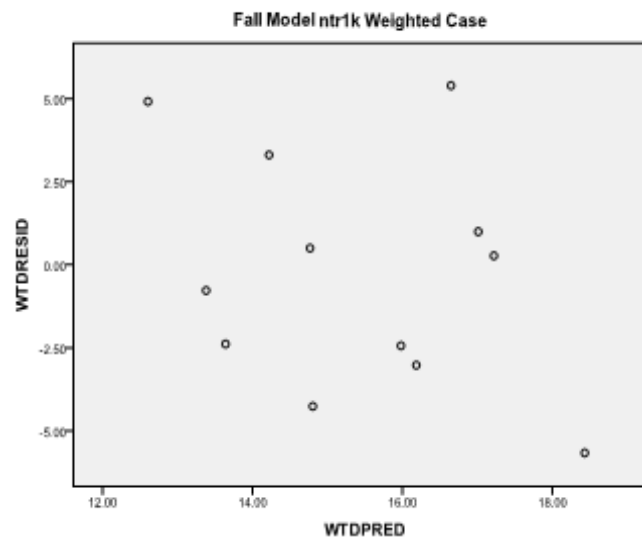


Figure 7 Weighted case of ntr1k for the fall model.

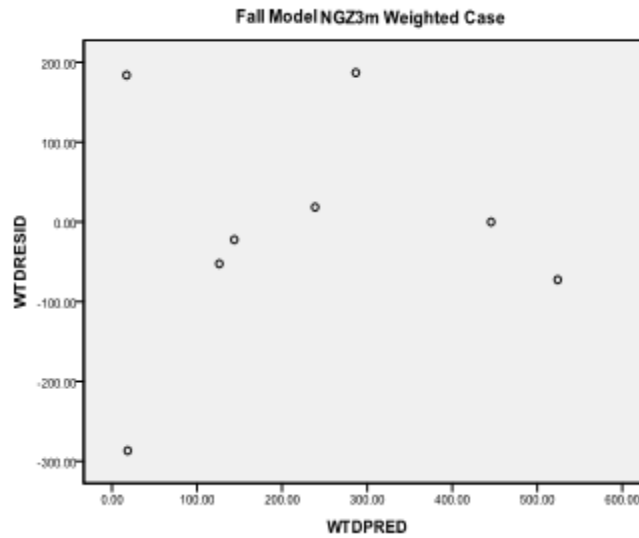


Figure 8 Weighted case of NGZ3m for the fall model.

10.4 CROSS-VALIDATION

The bootstrap analysis was first undertaken to evaluate the sensitivity of the estimated model variance in the five models. Supplementary to this method, histograms of the predicted parameter were validated and derived for 1000 iterations to examine the consistency of the standard variance as presented in APPENDIX F.

Secondly, we validated each model by a comparison of prediction errors obtained by simulating 1,000 jackknife data sets. Thirdly, the root mean square errors and the standardized root mean square errors were calculated as the square root of the sum of the squared differences of the observed concentration at site i and the predicted concentration at site i divided by 3528, the total grid nodes. In each of the cross-validations, the average difference was not greater than $9.660 \mu\text{g}/\text{m}^3$ for bootstrap, translating into an average relative difference of less than 3.0%. Also, there was a recurrent estimation variance value of $0.06 \mu\text{g}/\text{m}^3$ for the jackknife method, resulting in an error of 13%, and $6.691 \mu\text{g}/\text{m}^3$ for RMSE, which intuitively contrasts the measured values

with the predicted values within a 3% difference. The results of this comparison are presented in Table 10-7.

Table 5 Cross-validation results of the five interpolation techniques used to map fall PM₁₀ mass concentration throughout the El Paso cohort study.

	Fall Model			
	Bootstrap Variance	Jackknife Variance	RMSE	SRMSE
IDW	9.40812	0.05346	6.633	N/A
NN	9.458566	0.055363	6.651	N/A
OK	9.368487	0.049374	6.494	0.8076
SPLINE	9.660164	0.059697	6.691	N/A
UK	9.3792	0.051819	6.509	0.7651

Amongst the modeling techniques, the predicted concentrations generated by the SPLINE method had the worst performance, with a bootstrap variance of 9.66 $\mu\text{g}/\text{m}^3$, a variation in prediction certainty over 1000 iterations of 0.05969 $\mu\text{g}/\text{m}^3$, and a root mean square error of 6.691 $\mu\text{g}/\text{m}^3$. Similarly, the bootstrap histogram demonstrated a presumptuous bimodal shape, as seen in APPENDIX F. In this case, the alteration from normality and the deplorable prediction performance was caused by the fundamentals of the model as it creates the surface specifically so that it passes through the control points having the least possible change in slope at all points, fitting the control points with a minimum curvature surface. In other words, the model tends to average to its utmost capacity the PM₁₀ exposure throughout the cohort study. Thus, coinciding with the visual pattern (shown in APPENDIX H) were the incongruous blemishes, of higher concentration at lower polluted areas, broadens more frequently than the rest of the interpolation techniques. Considering the fact that the fall PM₁₀ surface tends to have a more disrupted variation pattern overall, when compared to the annual, spring, and summer models.

NN likewise also performed badly, as anticipated from the visual inspection, with a bootstrap variance of 9.458 $\mu\text{g}/\text{m}^3$, a RMSE of 6.651 $\mu\text{g}/\text{m}^3$, and a fractional bias obtained from the jackknife analysis of 0.0553 $\mu\text{g}/\text{m}^3$. In addition, the bootstrap histogram showed an alteration

of the normal distribution with a bimodal shape. One reason the model tended to misrepresents the spatial variability of PM_{10} worse than IDW and kriging, was because the NN method assigns no data values to the cell center of the perimeter cells of the output raster that fall outside the convex hull, which is defined by the outermost extent of the points. Additionally, influential points were disproportionally weighted over less influential points, especially in the north- and far-east areas, accordingly to the Voronoi's intersecting triangles that cause more variance between the measured values and the predicted values and thus the amiss performance of the natural neighbor method.

Sequentially, IDW performed better than the latter two methods described above, with a bootstrap difference of $9.408 \mu\text{g}/\text{m}^3$, a fractional bias of $0.0534 \mu\text{g}/\text{m}^3$, and a RMSE equal to $6.633 \mu\text{g}/\text{m}^3$ and far less attenuation. It is important to remark that NN and IDW are very similar methods, mainly because they both rely on the weight distance average. However, whereas NN assigns weight based on the percentage of overlap, IDW assigns weights based on their similar distance from the interpolation point. This could be confirmed as both NN and IDW concurred in their visual pattern, shown in APPENDIX H, not to mention their similarities in their cross-validation results to which they were put to attest. And although IDW outperformed NN, the model still had some flaws with respect to its ability to predict accurate PM concentrations. A reason for that is attributable to the fact that IDW average weighted distance cannot be greater than the highest or less than the lowest input values, contradicting and misrepresenting surface concentrations measured at the junction of I-10 and U.S. Highway 54 or as it is commonly known, the "Spaghetti Bowl". Although the bootstrap histogram shows a normal distribution, still the interpolation technique failed to project accurate surface characterization of particulate

matter, thus misrepresenting areas like the spaghetti bowl which the method identified as a hot spot.

Lastly, the best performance was seen for both ordinary and universal kriging approach for which the bootstrap variance equals to $9.368 \mu\text{g}/\text{m}^3$ and $9.379 \mu\text{g}/\text{m}^3$ respectively, the jackknife variance for ordinary kriging prompted a value of $0.0493 \mu\text{g}/\text{m}^3$ as well as $0.0518 \mu\text{g}/\text{m}^3$ for universal kriging. For the ELP using 3-yr data the RMSE for OK and UK resulted in an absolute value of $6.494 \mu\text{g}/\text{m}^3$ and $6.509 \mu\text{g}/\text{m}^3$, respectively; by contrast the SRMSE on these predictions was $0.807 \mu\text{g}/\text{m}^3$ and $0.765 \mu\text{g}/\text{m}^3$. Nevertheless, the histogram of the predicted estimates derived for each of the 1000 iterations suggested in both studies that the models were robust. Albeit universal kriging showed a slight bimodal shape; a conscious investigation revealed that the inclusion of the sample located at the junction of I-10 and Helen of Troy Dr., in the uppermost west area, tended to misrepresented the spatial gradient as the sample location and the circumvent areas are mostly residential use. Thereby this sample, even being in close proximity to the Franklin Mountains, contradicted the relationship between population and PM_{10} concentrations prompt by the fall LUR model (shown in Figure 10-8).

There is some tendency, however, for the performance of the ordinary kriging model to be rather better than for the universal kriging, indicating local stationary and the sensitivity of the method to the underlying input data. Kriging gives the preferred cross-validation values at the grid nodes, but these generally reflect the way in which the kriged surfaces are fitted to those points. Therefore the overall cross-validation diagnostics concludes that kriging methods demonstrated that the influential points are unlikely to drive the regression model and that ordinary kriging feebly outperforms universal kriging in predicting spatial characterization as the data displayed no trend at the scale of the modeling. All because ordinary kriging assumed that

the variation in PM_{10} is free of any structural component. These results indicated that the difference in temporal variation produced relatively moderate errors when comparing measured values to modeled values.

10.5 DISCUSSION

The fall model was constructed using 285 weekly concentrations from field monitoring sites in the El Paso area. In the first iteration of developing a city-specific LUR model, commercial and industrial areas within 750 m and 300 m respectively were strongly predictors in the model. The grouping of all three variables into a single parameter was done in order to contain the model within three variables and thus leaves span for other variables to be recognized by the model as surrogates for PM_{10} concentrations rather than concentrating the model into two land use variables (NGZ3m) and a third independent variable. Therefore, this newly developed log-transformed variable was the most consistent predictor of PM_{10} levels, accounting for over 63.5% of the variation in PM_{10} . Industrial and commercial densities served as a surrogate for the general level of human activity in the vicinity of a monitoring site. These densities have been associated with increase emissions, suggesting that with higher density areas may be subject to higher emissions by human activity. Also, the segregation of these land use types creates the need for travel between residential and commercial or industrial areas. Taken together, these characteristics increase the exposure surface with more spatial variability in the ELP urban area. In a second iteration, predictors directly based on traffic counts were of primary interest in the analysis. VMTs per buffer were secondly preferred to the sum of road's length, because they seem to reflect the emissions and dispersion process more closely. These variables usually showed lower correlations with the pollutant concentrations than other variables of comparable

scale. As a result, the sum of the trucks roads' length describing traffic intensity appeared to have greater explanatory power than those variables describing distance to the nearest road, traffic counts and even VMTs. Truck traffic, however, has been proved to be influenced by particles suspected to originate mainly from diesel exhaust engines (Cyrus et al., 2003). It is possible that the predictability of truck traffic may be depended on: (1) the demand for freights, to seize company's yearly essential appurtenances', increments traceable to the commencement of a new fiscal year; (2) industries depleting their budget's surplus prior to the sealing of their fiscal year based on the calendar year; (3) the increase in sales and restocking by department stores for the seasonal holidays; or (4) that the trade between U.S. cities and states during the fall season may be hampered by low temperatures in northern states that cause temporal closures of some Interstate freeways, making more feasible for companies to diverting truck traffic through I-10 as it preserves a mild temperature average throughout the season. Vice versa to the summer model, hence, the fall time period reckons two months (September and December) were truck traffic broadens either by freighters going into, going out and even going through the city as a consequence of the international trade between both countries. Expanding trade and truck traffic throughout the region due to the North American Free Trade Agreement has place a strain on the local airshed. With six international ports of entry throughout the PdN region, from which one (BOTA) has the heaviest combined traffic of heavy-duty and gasoline fueled vehicles among the U.S.-Mexico ports of entry (Wilbur and Smith, 2006), it is reckoned that over one million freight carriers enter the U.S. throughout this crossing each year. Therefore, the sum of all designated truck road's length, within a 500 meter buffer area collaborated the presumption that truck traffic induces variations in PM_{10} background concentrations. Consistently, the $PM_{2.5}/PM_{10}$ ratio for this particular time period yielded a modest value of 0.22 which derived from an increase in

averaged PM traffic-based emissions. In contrast, both summer and fall models yielded similar $PM_{2.5}/PM_{10}$ ratios, as they both contained truck-traffic-based parameters proxying moderate $PM_{2.5}$ concentrations. The predictability of traffic density was not only due to the actual traffic emissions but also to its association with many other geographic characteristics, for example, the presence of industrial and commercial areas, as mentioned before, which notably influenced the levels of concentration as well. It is remarkable to mention that both variables (NGZ3m and ntr1k) did not show any spatial autocorrelation nor multicollinearity. In a final stepwise iteration, normalized Population density within a 1500 meter was included in the model, serving as a surrogate for the general level of human activity in the vicinity of the monitoring site. Like industrial and commercial areas, population density has been associated with decreased driving speed, as the city's law enforces a lower speed limit at residential areas. This suggests that areas with higher population density are subject to higher emissions per vehicle (National Resource Council, 2005). As mentioned before population within a 1500 meter showed explanatory powers in describing the PM_{10} and when combined with log-transformed group zoning and log-transformed truck road's length, the model was able to explain 83.1% of the variation in the PM_{10} concentrations.

Additionally, the fitted model was tested for non-linearity, spatial autocorrelation, collinearity, homoscedasticity and normality. Although a questionable observation was unfold for the maximum centered leverage distance and cook's distance. However, after excluding the observation and re-evaluating the model regression fit it was determined that the observation had no undue influence on the model fit. In fact, such observation influenced the adjusted R^2 in a ± 0.02 margin, thus the observation was kept in the main analysis as it merely represented a potential outlier; as shown in previous section. In addition, a bootstrap to evaluate the sensitivity

of the parameters to the actual samples included in the modeling suggests that the model is relatively stable to the choice of samples. In general, the predictors were robust to sample selection, though it was found that normalized group zoning and length of truck roads variables were affected by the inclusion or exclusion of the Montanavista sample that had modest PM_{10} concentrations. Nevertheless, the fact that group zoning (300 m) and length of truck roads (1 km) in the El Paso city borough, using entirely different GIS layer and without the influence of the Montanavista sample, was more significant and lends support to the important role played by group zoning and traffic metric for this particular model.

Similarly to the other models previously discussed, the concentration of PM_{10} was calculated applying the three parameters with their corresponding coefficient and joined into a single attribute table. It is important to remark that no predicted value had to be withdrawn from the analysis as the highest concentration value ($70.365 \mu\text{g}/\text{m}^3$) was 20% lower than the maximum sample concentration measured at Plant (K). Five interpolations methods (IDW, NN, OK, UK, AND SPLINE) were combined with LUR to generate a confident PM_{10} spatial surface. The validity of three models was challenged by three procedures (bootstrap, jackknife, and RMSE) whereas the remaining two models were assessed by an additional method (SRMSE). The bootstrap cross validation procedure for OK yielded a variance over the subsample average of $9.368 \mu\text{g}/\text{m}^3$ in contrast to the second lowest variance of $9.379 \mu\text{g}/\text{m}^3$ for the UK method. The fractional bias of $0.0493 \mu\text{g}/\text{m}^3$ for OK was smaller when compared to the second lowest value of $0.05184 \mu\text{g}/\text{m}^3$ prompted by the UK method. The predictability of ordinary kriging was more precise with a RMSE of $6.494 \mu\text{g}/\text{m}^3$ compared to $6.509 \mu\text{g}/\text{m}^3$ for the universal kriging. Lastly, the SRMSE value for OK was closer to one than the UK meaning that the estimated standard error was more reliable or valid for the uncertainty of the predicted values. Overall

ordinary kriging based on Gaussian model outperformed IDW, NN, SPLINE, and UK on cross-validation results using the full set of samples. The cross-validation procedure, for example, yielded a bootstrap and jackknife variance, which was small compared to the range of concentrations and not greater than the RMSE, indicating that the influence of the single observations was not too heterogeneous.

In addition, 85% of the OK modeled estimates were within a factor of 1.2 times the range of the measured concentration with an inter quartile range (IQR) value of $19.316 \mu\text{g}/\text{m}^3$ approximately 6% less than the second lowest value cued by OK and 8% to the SPLINE method. The ELP ordinary kriging model exhibited a bias less evident, but still existing, in the full model that included the validation samples overpredicting 41% of the value locations and underpredicting 22% with a relatively low average absolute value residual or MAE of $5.379 \mu\text{g}/\text{m}^3$; therefore the overall model tended to overestimate PM_{10} concentrations. The underpredicted values manifested more pronounced in two urbanized areas of the cohort study; One underpredicted area with a predominant commercial activity and least residential area was in the proximity of US-Mexico border specifically in the downtown core, where commercial business prevail over this area with a lesser residential percentage; and second underpredicted area was the area circumvent to Western Refining, which also has a vast amount of industrial, commercial and residential areas with an obvious close proximity to I-10. Conversely, the overprediction appeared in areas distant from the centralized urbanized sprawl embracing sectors like the Fareast along Montana Ave., at the intersection of Highway 375 and Montana Ave., the uppermost West area above Transmountain Rd., and to a lesser extent, the topmost Northeast area surrounding Highway 54 of El Paso. All of the observed overpredictions were upwards of $80 \mu\text{g}/\text{m}^3$, due to unregulated fuel combustion. Vehicles are one of the leading sources of

particulate air pollution in the El Paso basin. Surrounding the ports of entry is a network of roads connected to a freeway used to disperse the cargo from suppliers to distribution and storage centers, an even across the border. There is more transport truck, which combust diesel fuel, traveling on the I-10 freeway than any other major road in the ELP basin. Westerdahl et al., (2005) found that measured freeway concentrations of particulate matter had a range of 60 to 820 $\mu\text{g}/\text{m}^3$ on the 10 east freeway, and that concentrations along major roadways with high truck traffic density were up to 20 times higher than residential concentrations. Thus, the overpredicted values in our analysis are at plausible levels. Categorizing the sites by tertiles, 61% of them would be classified identically. Only 39% of the grid nodes near I-10 and commercial and industrial would change between the extreme tertiles. The divergences might be partially explained by the influence of the meteorological conditions and human activity over the spatial distribution of PM_{10} during the fall season. When the predicted exposure was divided by the predicted quartiles it was estimated that 33% of El Paso residents fall into elevated categories for PM_{10} concentrations within 40 to 60 $\mu\text{g}/\text{m}^3$ during the fall season. Conversely, 84% of the residents in El Paso exhibit an exposure increase of 10 $\mu\text{g}/\text{m}^3$ or more from the previous seasonal period. This increment in PM_{10} corresponds to the commencement of stagnation events which in addition to the complex topography significantly limit the dilution of contaminants.

Finally, in this study LUR method successfully predicted small-scale spatial variation of ambient PM_{10} concentrations and adequately reproduced the regional trend identified by the monitoring network. The model predicted the lowest values within less commercial and industrial development in the Northeast of the city. That is a less marked northeastern trend whereas a steeper concentration gradient southward, towards the international border with CDJ. Higher concentrations were mapped near the downtown core and the urbanize area east of the

city. Although, Ciudad Juarez was shown to influence the spatial PM_{10} distribution in downtown El Paso it was unpractical to monitor the entire metropolitan area due to cost and political constraints in order to quantify the impact of CDJ.

As shown in Figure 10-8, the model reproduced one solely noticeable hot spot surrounding the downtown core extending to the BOTA port of entry, where the area comprises mostly group zoning land use, especially commercial use, representing 87% of the area of the buffer radii. Nevertheless, the proportional ratio of the reckoned length of truck roads was 14% lower than the highest value of 3.67 miles seen in a buffer. This generally spurred concentrations for the hot spot between 60 to 80 $\mu g/m^3$; much higher concentrations to what the summer and spring models predicted. The outcome bearded the long-term spatial variations of averaged PM_{10} concentrations for a localized area within larger metropolitan region and predicted a gradient decreasing with distance from the border. Moreover, the methodology captured the relative impact of emissions outside the localized monitoring network area and allowed for an optimal utilization of resources.

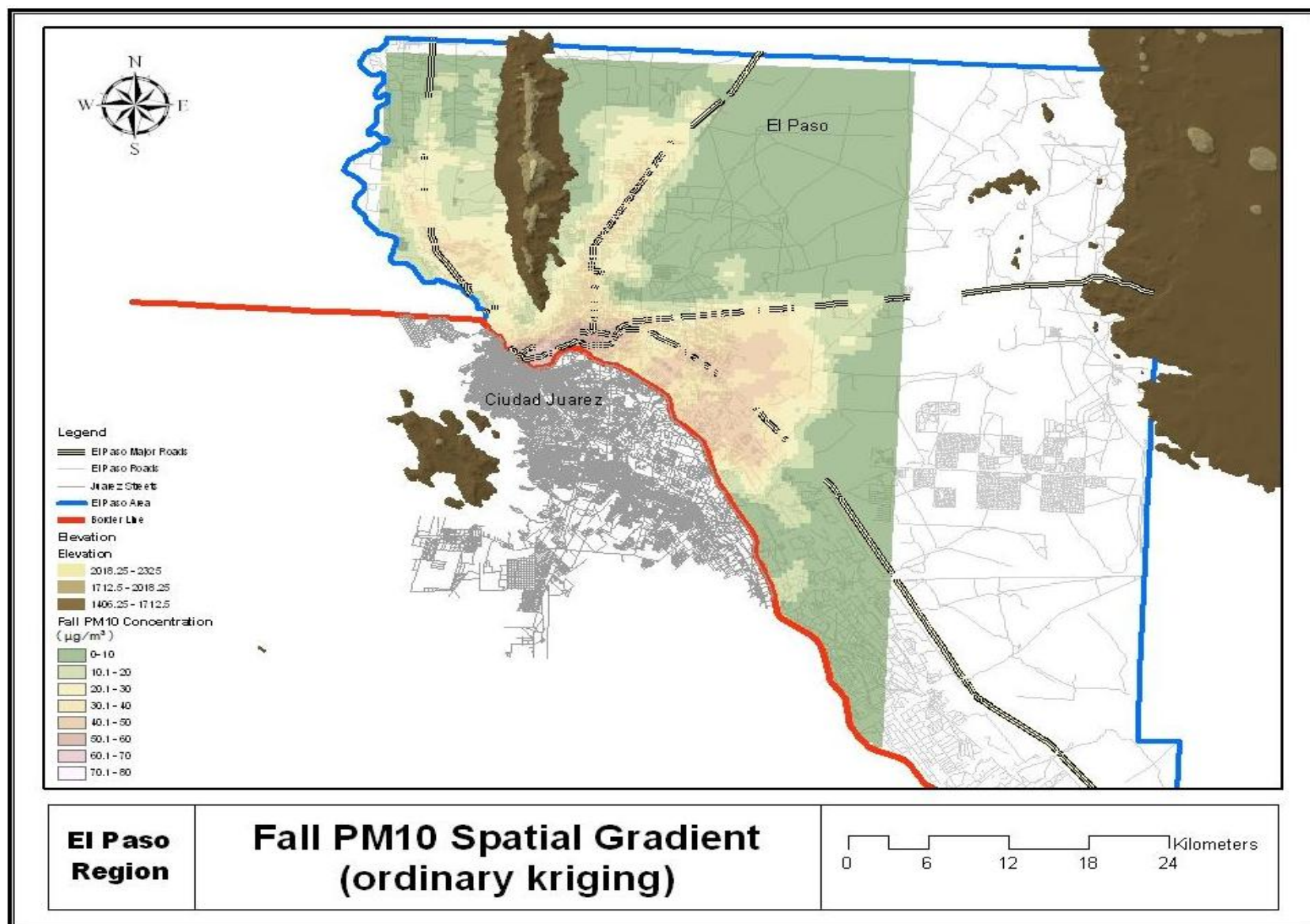


Figure 6 Fall PM₁₀ spatial gradient using Ordinary Kriging interpolation method.

11 WINTER MODEL

11.1 PM₁₀ MEASUREMENTS

During this sampling period, 275 ambient PM₁₀ weekly averaged concentrations were measured representing 95% of a total of 288 concentrations. Descriptive statistics of the air pollution measurements are shown in Table 11-1.

Table 11-1. Descriptive statistics of the average winter concentrations for the 13 monitoring stations.

Winter Descriptive Statistics for PM ₁₀ Concentrations (µg/m ³)					
	N	Minimum	Maximum	Mean	Std. Deviation
Lindbergh	12	16.6	41.6	28.1	8.4
Fountain	21	5.4	49.2	19.2	14.2
Skyline	13	11.8	37.9	24.2	8.0
McCombs	21	7.3	45.4	19.6	9.2
UTEP	33	11.6	48.2	30.3	9.1
Chamizal	11	11.7	52.2	31.2	12.6
Parral	26	8.9	74.7	33.2	18.2
Cosmos	34	3.4	56.0	22.7	10.3
Socorro	35	9.9	70.6	40.3	14.4
Plant	11	5.9	44.2	20.4	10.0
Mimosa	22	5.7	142.1	40.9	43.8
Montana	25	5.9	41.6	21.6	9.5
Pebble	11	13.5	50.5	27.6	10.7

Weekly averaged concentrations were depurated by removing values above 150 µg/m³ and greater than three standard deviations as well as concentrations underneath it from their corresponding site. The detection limits for PM₁₀ mass concentrations were defined by three standard deviations of the mass concentration deposit on field blanks. The sample measurements averaged 27.63 µg/m³ with a standard deviation of 7.35 µg/m³. In contrast, winter measurements

recorded higher concentrations and more variation than the concentrations observed during the summer period, as colder temperatures and stagnation events bash the region thus limiting the dilution of air contaminants. From the 275 sampled weekly concentrations, 44 and 24 observations were at least 1 standard deviation above or below the mean, respectively. All the measured values exhibited a normal distribution as shown in APPENDIX C. The maximum winter average PM_{10} concentration of $40.90 \mu\text{g}/\text{m}^3$ was observed at the Mimosa site near the border and a minimum of $19.17 \mu\text{g}/\text{m}^3$ at the Fountain site, which is in the farthest west. The overall highest weekly averaged PM_{10} concentration of $142.13 \mu\text{g}/\text{m}^3$ was measured at Mimosa site, located in the eastern part of El Paso whilst the overall lowest concentration of $3.37 \mu\text{g}/\text{m}^3$ was sampled at Cosmos site, a residential area. Similarly to the fall model a southward trend with the highest values near the border was observed as well as a less marked northeastern and westward trend.

The average $PM_{2.5}/PM_{10}$ ratio calculated from the dichotomous samples, averaged 0.27 with a standard deviation of 0.124. Similar to the previous models the sites produced a value of -0.05 ($p < 0.0005$) suggesting minimal spatial autocorrelation from the Morans's I test.

11.2 REGRESSION ANALYSIS

Each of the 190 independent variables was tested through the methods previously mentioned using the regression analyst tool in SPSS 17 and PASW Statistics 18. After conducting nearly 20 different models, no acceptable model fit was obtained nor was concordance between variables and their corresponding coefficient signs. Thus, distance to the border was discarded from the analysis and the subsequent highly correlated variable was inputted into the model in the stepwise procedure. Similar to the latter models Montanavista (site L) was excluded from the

analysis as it was generating inconsistencies in the model. Consequently, first volume traffic counts in a 1.5 km buffer area (vf15k) was entered into the model. Then, additional variables were paired with other significant variables in a series of trivariate regression models. This series of analysis identified the pairing with population within each monitor 500 m buffer area (POP5m). Lastly, industrial and commercial areas were comprised into a unique variable (GZ5m) that was found to be associated within a 500 meter buffer area as seen in the previous model. Table 11-2 provides a model summary and the ability of the three variables (vf15k, pop5m, and GZ5m) to predict ambient PM₁₀ mass concentrations (indicated by column 1).

Table 11-2. Land use regression model summary of the winter iteration.

Winter Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.949 ^a	.901	.865	2.74

a. Predictors: (Constant), GZ5m, pop5m, vf15k

As referred in previous chapters, 86.5% of the variability of ambient PM₁₀ concentration in the El Paso area is explained by vf15k, pop5m, and GZ5m with a standard error estimate (s) of 2.74 µg/m³. Meaning, that the model could expect to predict 86.5% of the ambient PM values to within approximately ±5.48 µg/m³ (or 14%) at a 95% confidence with variables like vf15k, pop5m, and GZ5m.

Table 11-3 shows the statistical coefficients for the winter model. As seen in Table 11-3, none of the VIF values surpass the rule of thumb of 2.5, thus suggesting no multicollinearity among the three variables. Similarly, the tolerance values reaffirmed the certainty that none of the variables were intercorrelated and more importantly that the concentrations estimated by the model may produce lower deviations from the actual fitted regression plane.

Table 11-3. Statistics of the model's coefficients for the winter iteration.

Winter Model Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	7.868	3.010		2.614	.031		
pop5m	.008	.002	.460	3.913	.004	.893	1.120
vf15k	.002	.000	.637	5.358	.001	.873	1.146
GZ5m	35.808	4.502	.963	7.954	.000	.841	1.189

a. Dependent Variable: PM10_Mean

A further assessment for multicollinearity in the model was performed, as multicollinearity may occur even when the rules of thumb for tolerance and VIF were met. Therefore, collinearity diagnostics (Table 11-4) was an alternative method of assessing if there is too much multicollinearity in the model. To simplify, cross products of the independent variables are factored. There will be as many factors as independents, plus one (for the constant). The first column addresses the dimension representing the variables included plus one constant variable. The ordering of the variables corresponds to the same arrangement shown in the variance proportions. That is, the constant variable corresponds to the first dimension, pop5m to the second dimension and so forth. The subsequent column identifies the eigenvalue which is a term related in the field of linear algebra and evaluates linear transformations represented as matrices acting on vectors. The eigenvalues shown in the Table 11-4 shows little variance in the cross product matrix, as high eigenvalues indicate dimensions (factors) which account for much of the variance in the cross product matrix; eigenvalues close to 0.0, such as in this case, indicate dimensions which explain moderate variance. Values such as for 0.617 pop5m, 0.261 for vf15k, and 0.47 for GZ5m obviously underwent the rule of thumb with dimensions above the 0 value.

Table 11-4. Collinearity diagnostics of the model's parameters for the winter iteration.

Winter Model Collinearity Diagnostics ^a							
Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	pop5m	vf15k	GZ5m
1	1	3.076	1.000	.01	.01	.02	.03
	2	.617	2.234	.00	.01	.07	.61
	3	.261	3.435	.00	.25	.50	.05
	4	.047	8.128	.99	.72	.40	.31

a. Dependent Variable: PM10_Mean

With the highest condition index of 8.13 prompt by GZ5m, it was clear that the model had no collinearity problems making it unnecessary to examine variance proportions.

Lastly, after determining the absence of collinearity, PM₁₀ concentrations were determined to be influenced by the population in a 500 m buffer area, traffic counts in a 1.5 km circular area, and the group zoning (industrial and commercial) as shown in Equation 11-1.

$$PM_{10} = 7.868 + .008 \cdot (pop5m) + 0.002 \cdot (vf15k) + 35.808 \cdot (GZ5m) \quad (11-1)$$

11.3 MODEL VALIDATION

After determining the appropriate model, standard regression diagnostics were applied to assess the problems such as influential outliers, linearity, independency, normality, and homoscedasticity. Standard multiple regression can only accurately estimate the relationship between dependent and independent variables if the relationships are linear in nature. Therefore if violations of linearity are extremely serious (fitting data to a linear model which are

nonlinearly related), the predictions are likely to produce systematic errors, especially when generating usually large or small predictions.

Figure 11-1 unveils a correlation coefficient of 0.90, and a symmetrically point distribution around a diagonal line determining a linearity relationship between PM₁₀ mass concentrations and vf15k, pop5m, and GZ5m. Note that, the standardized predicted values were calculated using the model equation for every particular monitoring site previously discussed. In addition the statistical software suited for this research produce such values in a residual statistics table.

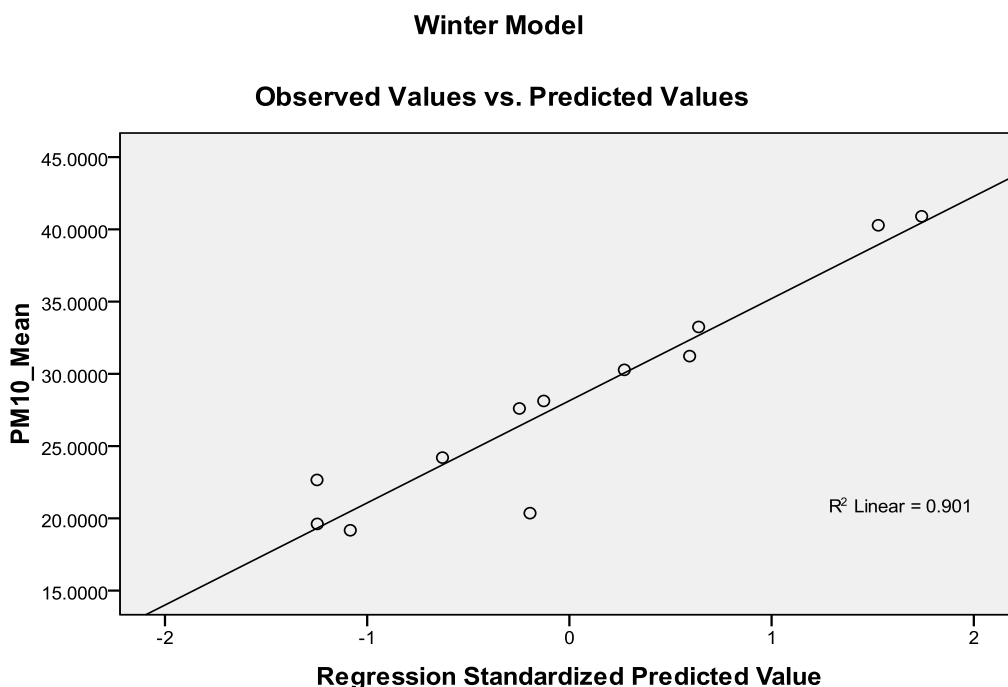


Figure 9 Linearity test between the sample concentrations and the predicted values for the winter model.

In addition to the Moran's I test, performed in previous section, a second evaluation was performed under the Durbin-Watson test. Thus using the test statistics shown in Equation 7-3, a corresponding d value of 2.194 was achieved. However, obtaining the $d_{\text{lower critical}}$ value (0.658) and the $d_{\text{upper critical}}$ value (1.86) at 5% significance from APPENDIX E, the test proven to reject

the null hypothesis and thus accepts the alternative hypothesis demonstrating that the error estimates are not positively autocorrelated.

Table 11-5 contains summary data regarding the residuals (the difference between predicted and actual values). Standard residual (Std. residual) for instance is the standardized residual raw residual divided by the standard deviation of residuals. Since the minimum standardized residual is -2.33, at least one prediction is more than one standard deviation below the mean residual. Studentized residuals are very similar to standardized residuals but follow the t distribution.

Table 11-5. Descriptive statistics of the winter model's residuals.

Winter Model Residuals Statistics^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	19.314922	40.455750	28.139258	7.0698080	12
Std. Predicted Value	-1.248	1.742	.000	1.000	12
Standard Error of Predicted Value	.929	2.182	1.531	.417	12
Adjusted Predicted Value	18.276806	39.838226	27.850603	6.8133154	12
Residual	-6.4031701	3.3429782	.0000000	2.3373696	12
Std. Residual	-2.336	1.220	.000	.853	12
Stud. Residual	-2.491	1.396	.041	.955	12
Deleted Residual	-7.2796092	4.3810945	.2886552	3.0052492	12
Stud. Deleted Residual	-4.919	1.502	-.159	1.592	12
Mahal. Distance	.346	6.057	2.750	1.917	12
Cook's Distance	.001	.284	.069	.095	12
Centered Leverage Value	.031	.551	.250	.174	12

a. Dependent Variable: PM10_Mean

Similarly, the minimum, maximum, and mean Mahalanobis distances displayed in Table 11-5 denote much lower distances than the critical chi square value of 16.27, disregarding any possibility of outliers influencing the development of the linear regression model. The minimum, maximum, and mean Cook's distance is displayed in Table 11-5. Using the cut-off suggested by

Fox (1991: page 34), a computed threshold value of 0.50 was obtained, thus no apparent problem cases aroused since the maximum Cook's D prompted by the model was 0.284. Lastly, the maximum leverage value of 0.551 shown in Table 11-5 was smaller than the computed threshold from the small sample equation of 0.75, thus discarding the possibility of any undue influence from a measurement error by the vf15k, pop5m, and GZ5m for the winter model.

Error, represented by the residuals, should be normally distributed for each set of values of the independents, as violations of normality compromise the estimation of coefficients and the calculation of confidence intervals. Calculations of confidence intervals and various significance tests for coefficients are all based on the assumptions of normally distributed errors. If the error distribution is significantly non-normal, confidence intervals may be too wide or too narrow.

Figure 11-2 shows the residuals followed a bimodal distribution, based on predicting PM₁₀ mass concentrations from vf15k, pop5m and GZ5m. This suggests that the error distribution is "skewed" by the presence of outliers. Since parameter estimation is based on the minimization of squared error, a few extreme observations may have exerted a disproportionate influence on parameter estimates.

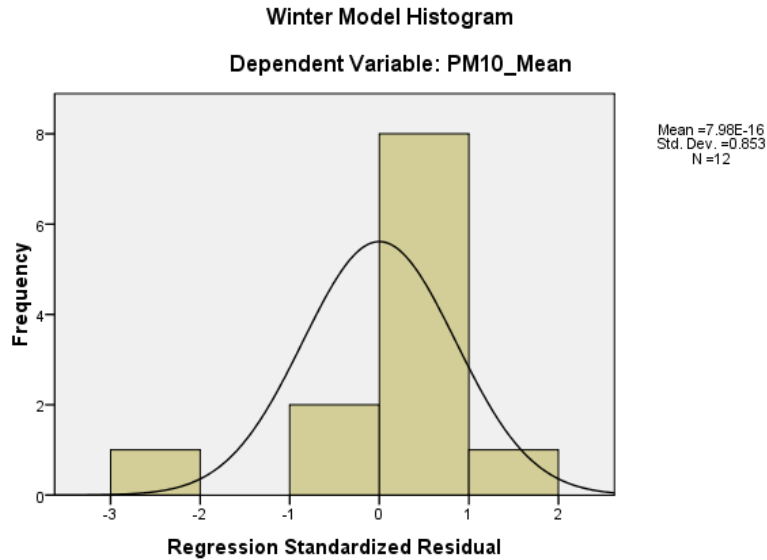


Figure 11-2. Residual's distribution of the winter model.

In addition, the Shapiro-Wilk's (W) test was performed over the model's residuals. In this case, W was significant with a statistic values of 0.79. This suggests that the variable's distribution is significantly different from normal (Table 11-6).

Table 10 Shapiro-Wilk's W test for normality of the winter model's residuals.

Winter Model Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.287	12	.007	.791	12	.007

a. Lilliefors Significance Correction

A further investigation revealed that the inclusion of a particular monitoring site, Plant (K), caused the alteration from normality in the error distribution graph which overpredicted PM₁₀ concentrations by 9 %. Overall, however, the concentrations shift had little overall effect on the model predictions. Removing the current monitored concentrations the R² value (0.865) yielded a R² of 0.868 that was, on average, just 0.4% different, merely insignificant, thus the site was keep in the analysis.

The homoscedasticity test was performed for the winter model. Figure 11-3 illustrates a disperse pattern of the points with no trend. The distribution of the residuals among the standardized residuals axis demonstrate some ranges of errors below the negative standardized residual axis, while others ranged positive above the 0 value of the y-axis. However, there is one possible outlier in the standardized residuals data. This outlier certainly is demonstrated in Figure 11-4, as the residual boxplot showed the residual of Plant to have a value over two standard deviations from the mean.

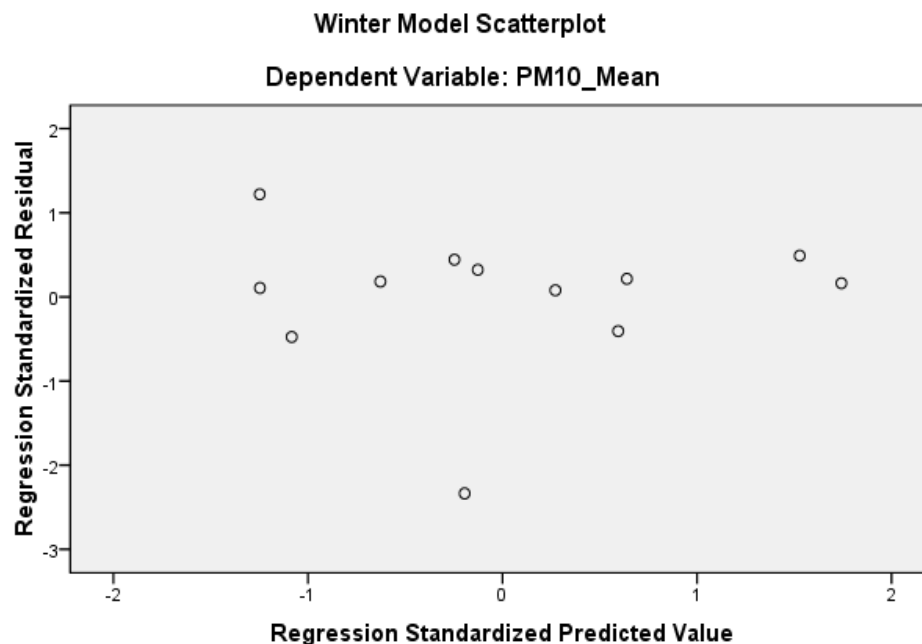


Figure 11-3. Scatterplot of the standardized residuals versus standardized predicted values for the winter model.

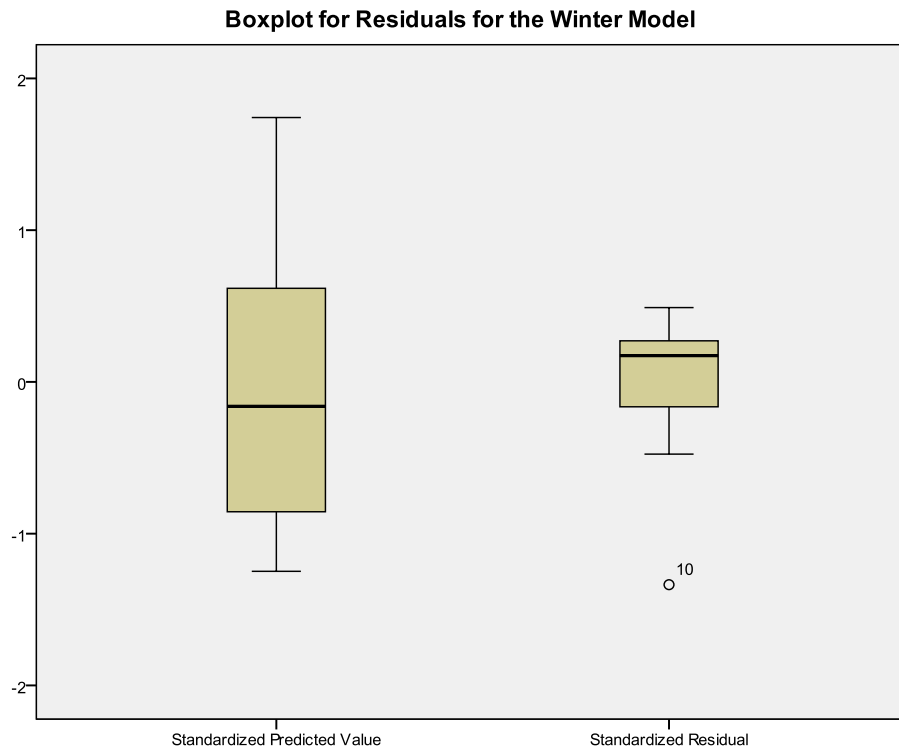


Figure 11-4. Boxplot of the standardized predicted values and standardized residuals for the winter model.

Similarly to the spring model, the observation remained under the limit area defined by Pardoe (2006; page 166). Yet after removing the observations in question, the boxplot did not detect any additional potential outliers. With no change in the regression line ($\pm 0.4\%$), neither changes in the parameters nor the change in the linear sign; the observation in question was retained in the main analysis as either test was fully met. Therefore, it should be remarked that there was no conclusive evidence to remove the observation from the analysis as they merely represented potential outliers. The scatter plot, nevertheless, reflects the fact that the model in this case has modest standard errors of estimate, explaining the percentage of variance mainly as a consequence of an acceptable fitted model. Furthermore, the scatterplot of the standardized residuals versus standardized predicted values corroborated the influenced of Plant in the normal

distribution. On the other hand, the modest violation generated by Plant in the homoscedasticity test produced a wider range of concentrations between sites, even more disrupted than the summer and spring model. Nevertheless, and regardless of Plant samples, the residual plot patterns endorsed the relationship between predictor and criterion as linear.

Heteroscedasticity as also evaluated in Figures 11-5 through 11-7. These cases were weighted by the reciprocal of their estimated point variance, and each assesses the goodness of fit of the weighted model. That is, weighted predicted is plotted on the x -axis and weighted residual on the y -axis. The three weighted residual plots produced a good fit. Although one observation in Figure 4-103 appears questionable, the residuals did not form a funnel shape. Conversely, the observation in question corresponded to the residual of the model measured at Plant weighted by the reciprocal of their estimated zoning parameter variance. The questionable observation remained insignificant for the heteroscedasticity test and merely represented a potential outlier.

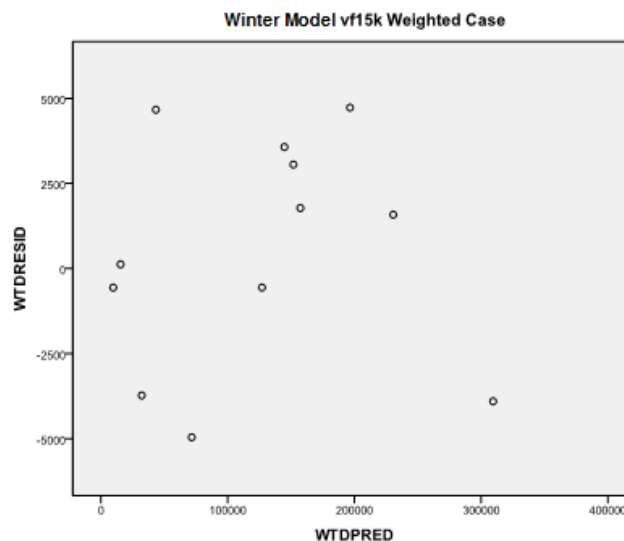


Figure 11-5. Weighted case of vf15k for the winter model.

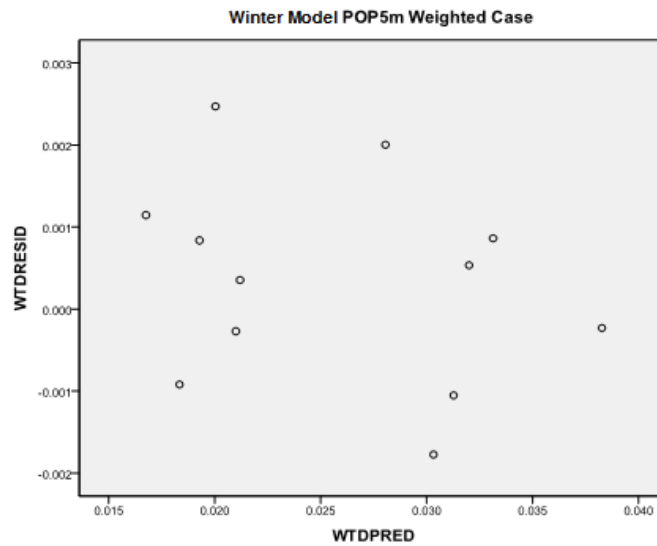


Figure 11-6. Weighted case of pop5m for the winter model.

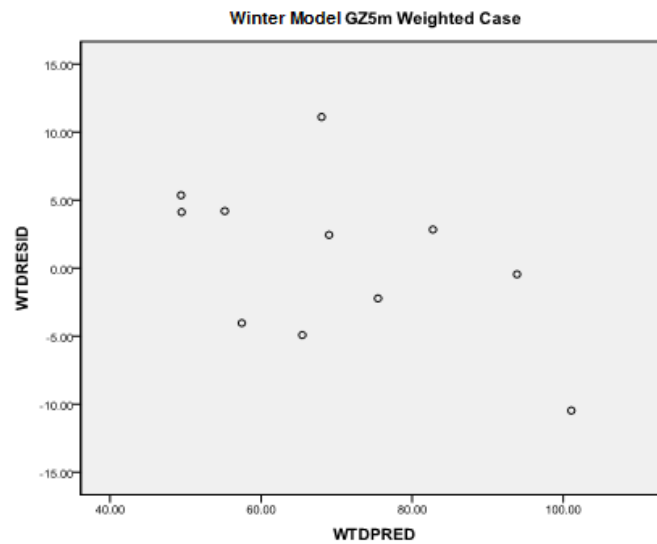


Figure 11-7. Weighted case of GZ5m for the winter model.

11.4 CROSS-VALIDATION

The bootstrap analysis was first undertaken to evaluate the sensitivity of the estimated model variance in the five models. Supplementary to this method, histograms of the predicted parameter were validated and derived for 1000 iterations to examine the consistency of the

standard variance as presented in APPENDIX F. Secondly, we validated each model by a comparison of prediction errors obtained by simulating 1,000 jackknife data sets. Thirdly, the root mean square errors and the standardized root mean square errors were calculated as the square root of the sum of the squared differences of the observed concentration at site i and the predicted concentration at site i divided by 3528, the total grid nodes. In each of the cross-validations, the average difference was not greater than $7.86 \mu\text{g}/\text{m}^3$ for bootstrap, translating into an average relative difference of less than 4.0%. Also, there was a recurrent estimation variance value of $0.041 \mu\text{g}/\text{m}^3$ for the jackknife method, resulting in an error of 4%, and $7.14 \mu\text{g}/\text{m}^3$ for RMSE, which intuitively contrasts the measured values with the predicted values within a 10% difference. The results of this comparison are presented in Table 11-7.

Table 7. Cross-validation results of the five interpolation techniques used to map winter PM_{10} mass concentration throughout the El Paso cohort study.

	Winter Model			
	Bootstrap Variance	Jackknife Variance	RMSE	SRMSE
IDW	7.560209	0.035006	6.299	N/A
NN	7.679548	0.037257	7.025	N/A
OK	7.505061	0.031962	6.287	0.8436
SPLINE	7.854324	0.040947	7.149	N/A
UK	7.530209	0.033006	6.295	0.7436

Amongst the modeling techniques, the predicted concentrations generated by the SPLINE method had the worst performance, with a bootstrap variance of $7.85 \mu\text{g}/\text{m}^3$, a variation in prediction certainty over 1000 iterations of $0.041 \mu\text{g}/\text{m}^3$, and a root mean square error of $7.149 \mu\text{g}/\text{m}^3$. Similarly, the bootstrap histogram demonstrated a presumptuous bimodal shape, as seen APPENDIX F. NN likewise also performed badly, as anticipated from the visual inspection, with a bootstrap variance of $7.679 \mu\text{g}/\text{m}^3$, a RMSE of $7.025 \mu\text{g}/\text{m}^3$ and a fractional bias obtained from the jackknife analysis of $0.0372 \mu\text{g}/\text{m}^3$. Although the bootstrap histogram follows a normal

distribution with no bimodal shape whatsoever, the model still tended to misrepresent the spatial variability of PM_{10} worse than IDW and kriging. Additionally, influential points were disproportionally weighted over less influential points, especially in the far-east area at the intersection of Highway 375 and Montana Ave., accordingly to the Voronoi's intersecting triangles that cause more variance between the measured values and the predicted values and thus the amiss performance of the natural neighbor method. Sequentially, IDW performed better than the latter two methods described above, with a bootstrap difference of $7.560 \mu\text{g}/\text{m}^3$, a fractional bias of $0.035 \mu\text{g}/\text{m}^3$, and a RMSE equal to $6.299 \mu\text{g}/\text{m}^3$ and far less attenuation. And although IDW outperformed NN, the model still had some flaws with respect to its ability to predict accurate PM concentrations. In fact, the bootstrap histogram shows a modest bimodal shape traceable to the surrounding grid points, located in vicinity of WR, just one mile from the US-Mexico border. This point had 2.5 standard deviations away from the concentration mean of all samples falling into the 98.1% confidence interval. However, there was no reason to believe that the underlying data from this station was inaccurate, as the value showed a 25% increment from the value measured at mimosa site (K2) which followed the same protocol as the three monitoring sites correlated with the State fix stations, and therefore the monitor was not excluded. The best performance was seen for both universal and ordinary kriging approach for which the bootstrap variance equals to $7.53 \mu\text{g}/\text{m}^3$ and $7.505 \mu\text{g}/\text{m}^3$ respectively, the jackknife variance for universal kriging prompted a value of $0.033 \mu\text{g}/\text{m}^3$ as well as $0.0319 \mu\text{g}/\text{m}^3$ for ordinary kriging, for the ELP using 3-yr data the RMSE for UK and OK resulted in an absolute value of $6.295 \mu\text{g}/\text{m}^3$ and $6.287 \mu\text{g}/\text{m}^3$, respectively; by contrast the SRMSE on these predictions was $0.743 \mu\text{g}/\text{m}^3$ and $0.843 \mu\text{g}/\text{m}^3$. Nevertheless, the histogram of the predicted estimates derived for each of the 1000 iterations suggested in both studies that the models were robust.

Albeit universal kriging showed a slight bimodal shape; a conscious investigation revealed that the inclusion of a single sample (located at the junction of Montana Ave. and Highway 375) resulted in a 29% change in the parameter. This sample location (also disproportionately weighted over less influential points in the NN method) had the highest traffic volume in the 1.5 km buffer and also the highest PM_{10} concentration value. Overall, however, the parameter shift had a little, but still representative surface map effect on predictions. Replacing the current coefficient value (0.002) with the mean value of the right tail (0.0026) yielded predictions that were, on average, just 0.4% different. Kriging gives the preferred cross-validation values at the grid nodes, but these generally reflect the way in which the kriged surfaces were fitted to those points. Therefore, the overall cross-validation diagnostics concludes that kriging methods demonstrated that the influential points are unlikely to drive the regression model and that ordinary kriging feebly outperforms universal kriging in predicting spatial characterization as the data displayed no trend at the scale of the modeling. The ordinary kriging variogram (see APPENDIX G) exhibited a more suited model with smaller range and sill than the semivariogram for universal kriging (see APPENDIX G). All because, ordinary kriging assumed that the variation in PM_{10} is free of any structural component. The unusual values in traffic counts, for example, inflated and deflated the model prediction power when compared to the observations. Universal kriging, as a result of a drift, misrepresents the smoothness of the pollutant surfaces caused by a drift. Points close to the freeway, for example, exhibited a similar relationship between PM_{10} and traffic counts as well as points, for example, in less trafficked areas far from any major source of PM_{10} . These results indicate that the difference in temporal variation produced relatively minimum errors when comparing measured values to modeled values.

11.5 DISCUSSION

The winter model was constructed using 275 weekly concentrations from field monitoring sites in the El Paso area. Aside from, Group zoning of the sampling location was the most consistent predictor of PM_{10} levels, accounting for over 64% of the variation. When combining this variable with the average first volume of traffic counts in a 1.5 km buffer and the population within 500 meter buffer, the model was able to explain almost 86.5% of the spatial variation PM_{10} in El Paso, Texas. The nearest distance to the border was an influential parameter in the initial model as the proximity to Juarez has often aroused the speculation of a source for air pollution. The highly industrialized areas, the closeness of brick kilns, unpaved roads, and the traffic activity of antediluvian vehicles along the border in the vicinity city of Juarez has proxy to be a contributor of PM_{10} concentrations for the El Paso region. However, OND parameter was dropped from the model analysis as inconsistent correlations were found when paired to other variables. In fact, when a presumable model was achieved, the remaining paired variables were assigned an incongruous adverse correlation sign to what was expected.

In the first iteration of developing a city-specific LUR model commercial and industrial area within 1500 m and 1000 m respectively were strongly predictors in the model when paired to nearest distance to the border. The grouping of all three variables into a single parameter was done in order to contain the model within three variables and leave span for other variables to be recognized by the model as surrogates for PM_{10} concentrations rather than concentrating the model into distance to the border and two independent land use variables. Therefore, this newly developed variable (GZ5m) was the most consistent predictor of PM_{10} levels, accounting for over 64% of the variation in PM_{10} . Industrial and commercial densities served as a surrogate for the general level activity in the vicinity of a monitoring site. These densities have been

associated with increase emissions, suggesting that with higher density areas may be subject to higher emissions by human activity. Also, the segregation of these land use types creates a need for travel between residential and commercial and industrial areas. Taken together these characteristics increase the exposure surface with more spatial variability in the ELP urban area. In a second stepwise iteration predictors directly based on VMTs were of primary interest in the analysis, nearest distance to a road's per buffer were secondly preferred to traffic counts because they seem to reflect the emissions and dispersion process more closely. The first two variables usually showed lower correlations with the pollutant concentrations than other variables of comparable scale. In general, first volume of traffic counts describing traffic intensity appeared to have greater explanatory power than those variables describing distance to the nearest road, sum of road's length, and even VMTs. Furthermore, an additional attempt was persevered to explain more variability of the pollutant by comparing both traffic metrics in a single model, but only marginal increases in the R^2 values were obtained. All these results suggested that model built with road length and vehicle density metrics are equally able to explain small-scale variability in pollutant concentrations. This finding confirmed that valuable LUR model can be developed in the absence of traffic metrics, which are unavailable, none existing, or faulty representative in many areas. For example, Briggs et al. (2005) used road type data for their Amsterdam model, but could not speculate on whether better results were achieved with traffic density information. Wayne and Hugh (1995) showed that combustion sources correspond to 80% of the total PM_{10} particles in the winter season assuring the idea that during cold days the trend of vehicle flows tends to increase more than other seasons. Temporal variations in PM_{10} have also been shown to coincide with the morning and evening commute hours at monitoring sites near emission source regions of central El Paso (Jeon et al., 2001; and Noble et al., 2003).

However, incomplete data provided by the MPO precluded the use of traffic data for all time periods. It is likely that the model based on average daily traffic and average daily morning and afternoon rush hour traffic rather than average yearly traffic might lead to different parameters. It is also possible that the analysis such as those that rely on rush hour traffic could lead to a wider spread of PM_{10} predictions, as it might expected high traffic areas in off-peak times have a significant larger increase in traffic during rush hour. Vice versa, analysis such as those that rely on daily traffic could lead to a more consistent spread of PM_{10} , expecting daily traffic areas to have a significant lower to none increase in traffic during the day. Nevertheless, it is expected that relative rates of traffic remained very similar. Although local traffic related emissions are only significant closer to the roads. Another study perform in the region formulated a LUR model to predict the intra-urban gradient of NO_2 . They cited that the inclusion of the proximity to a port of entry and the addition of distance to a major highway and elevation to the model explained 81% of the variation of the NO_2 gradient for El Paso (Gonzales et al., 2005). Therefore, this model coincides with the findings of Gonzales et al. (2005) that traffic-based parameters proxy higher concentration in the cohort study. Although neither elevation nor distance to the border were contributors in the development of the model, it was substantiated that both parameters had an explanatory power over the winter seasonal PM_{10} concentrations. In fact, even after showing the highest bivariate correlation coefficients for the winter model, 80 and 81%, respectively; both surrogate parameters produced inconsistent coefficients and spatial autocorrelation when a second or third variable was added to the model. The predictability of traffic counts was not only due to the actual traffic emissions but also to its association with many other geographic characteristics, for example, the population density or the presence of commercial and industrial areas; which notably influenced the concentration levels as well.

Moreover, in final stepwise iteration population density within a 500 meter served as a surrogate for the general level of human activity in the vicinity of the monitoring site. Like residential and commercial areas, population density has been associated with decreased driving speed and increased emissions as the city's law enforces lower speed limits at residential areas. This suggests that areas with higher population density are subject to higher emissions per vehicle (National Resource Council, 2005). As mentioned before population within a 500 meter showed explanatory powers in describing the PM_{10} and when combined with traffic counts and group zoning, the model was able to explain 86.5% of the variation in the PM_{10} concentrations.

Additionally, the fitted model was tested for non-linearity, spatial autocorrelation, collinearity, homoscedasticity and normality. One questionable observation drew the attention of possible outlier. The observation corresponded to the average concentration sampled at Plant; thereby after conscious analysis it could not be determined to be an outlier. Apart from that, the observation after being excluded and the model reassessed the R^2 had an insignificant shift of merely 0.4%. Thus the observation was determined to have no undue influence on the model fit. Likewise, the residual was also excluded from the standardized residual versus standardized predicted value plot, then the scatterplot re-graphed to the remaining sample observations corroborating that the substantial change in the linear fit did not affect the model fit whatsoever. Moreover, according to Pardoe (2006; page 166), a potential outlier should be considered as an observation with a studentized residual outside the range of -3 to +3. Based on this determination the studentized residual remained under the limit area. Thereby, with no change in the regression line, no changes in the parameters, and no change in the linear sign the observations in question merely represented potential outliers and had insignificant influence in the homoscedasticity test. In addition, a bootstrap to evaluate the sensitivity of the parameters to the actual samples

included in the modeling suggests that the model was relatively stable to the choice of samples. In general, these variables were robust to sample selection, though it was found that group zoning, population, and traffic counts variables were affected by the inclusion or exclusion of the Montanavista sample that had modest PM₁₀ concentrations. Although, the fact that group zoning (500 m), traffic counts (1.5 km), and population (500 m) in the El Paso city borough, using an entirely different GIS layer and without the influence of the Montanavista sample, was more significant and lends support to the important role played by zoning, traffic related parameters and population for this particular model. It is important to mention that the predictions for the El Paso region are not at all compromised by this problem, and on the whole, the named limitations do not affect the main findings of the analysis. Consequently, after deriving the parsimonious operational regression model a grid covering the cohort study and containing 3528 nodes was developed in order to sharpen the resolution of the depicted overall PM₁₀ trend. At each node the concentration of PM₁₀ was calculated applying the three parameters with their corresponding coefficient and joined into a single attribute table. It is important to remark that no predicted value had to be withdrawn from the analysis as the highest concentration value (94.237 µg/m³) was 50% lower than the maximum sample concentration measured at Mimosa (K2). Given the strong predictive power of hybrid models (Jerrett et al., 2005), five interpolations methods (IDW, NN, OK, UK, AND SPLINE) were combined with LUR to generate a confident PM₁₀ spatial surface. The validity of three models was challenged by three procedures (bootstrap, jackknife, and RMSE) whereas the remaining two models were assessed by an additional method (SRMSE). The bootstrap cross validation procedure for OK yielded a variance over the subsample average of 7.505 µg/m³ in contrast to the second lowest variance of 7.530 µg/m³ for the UK method. The fractional bias of 0.0319 µg/m³ for OK was smaller when compared to the

second lowest value of $0.0330\mu\text{g}/\text{m}^3$ prompted by the UK analysis. The predictability of ordinary kriging was more precise with a RMSE of $6.287\mu\text{g}/\text{m}^3$ compared to $6.295\mu\text{g}/\text{m}^3$ for the universal kriging. Lastly, the SRMSE value for OK was closer to one than the UK meaning that the estimated standard error was more reliable or valid for the uncertainty of the predicted values. Overall ordinary kriging based on Gaussian model outperformed IDW, NN, SPLINE, and UK on cross-validation results using the full set of samples. The cross-validation procedure, for example, yielded a bootstrap and jackknife variance, which was small compared to the range of concentrations and not greater than the RMSE, indicating that the influence of the single observations was not too heterogeneous. In addition, 87% of the OK modeled estimates were within a factor of 1.2 times the range of the measured concentration with an inter quartile range value of $18.272\mu\text{g}/\text{m}^3$ approximately 2% less than the second lowest value cued by UK and 5% to the SPLINE method. The ELP ordinary kriging model exhibited a bias less evident, but still existing, in the full model that included the validation samples overpredicting 35% of the value locations and underpredicting 23% with a relatively low average absolute value residual or mean absolute error of $4.19\mu\text{g}/\text{m}^3$; therefore the overall model tended to overestimate PM_{10} concentrations. The underpredicted values manifested more pronounced in two urbanized areas of the cohort study. One in the proximity of Lomaland Dr. and I-10, where commercial and industrialized areas predominate, not to mentioned residential areas as well; and second was the area circumvent to Western Refining, which also has a vast amount of industrial, commercial and residential areas. Conversely, the overprediction appeared in areas distant from the centralized urbanized sprawl embracing sectors from the Fareast along Montana Ave., at the intersection of Highway 375 and Montana Ave., the uppermost West area above Transmountain Rd., and to a lesser extent, the topmost Northeast area surrounding Highway 54 of El Paso. All of the

observed overpredictions were upwards of $60 \mu\text{g}/\text{m}^3$, due to unregulated fuel combustion. Vehicles are one of the leading sources of particulate air pollution in the ELP basin. Surrounding the ports of entry, especially the BOTA and Downtown Bridge, is a network of roads connected to the freeway use to disperse El Pasoan's and daily commuter's vehicles from residential to industrial and commercial areas. Westerdahl et al. (2005) found that measured freeway concentrations of PM had a range of 60 to $820 \mu\text{g}/\text{m}^3$ on the 10 east freeway, and that concentrations along major roadways with high traffic density were up to 20 times higher than residential concentrations. Thus, the overpredicted values in this analysis were at plausible levels. Categorizing the sites by tertiles, 64% of them would be classified identically. Only 36% of the grid nodes near roads and commercial and industrial would change between the extreme tertiles. Similar to the fall model, the divergences might be partially explained by the influence of the meteorological conditions and human activity over the spatial distribution of PM_{10} during the winter season. When the predicted exposure was divided by the predicted quartiles it was estimated that 12% of El Paso residents fall into highest categories for PM_{10} concentrations during the winter season. Conversely, the range and characterization of the PM_{10} gradient remains homogeneous to what the concentration pattern was perceived for the fall model, although there is a small percentage of resident (4%) that experience an increment in PM_{10} exposure of $10 \mu\text{g}/\text{m}^3$; the similarities of both winter and fall models corresponds to their akin meteorological conditions which critically influence the PM concentrations in El Paso. On days when the mixed layer growth rate is slow, PM_{10} , remain confined near central El Paso emission sources. Low winds, as seen in APPENDIX B, and the predominant meteorological inversions during the current winter study period resulted in a similar dispersion pattern of measured PM_{10} concentrations as the fall model.

Finally, in this study LUR method in combination to ordinary kriging successfully predicted small-scale spatial variation of ambient PM₁₀ concentrations with no spatial irregularities and adequately reproduced the regional trend identified by the monitoring network. The model predicted the lowest values at the outskirts of El Paso, in particular the upper region of the west side, northeast, and the southeastern area of the cohort study that represented an inferior urbanized human activity. That is a less marked western and northeastern trend whereas a steeper concentration gradient southward towards a more urbanized region. The higher concentrations, though, were mapped at different locations especially at areas close to the freeway with high commercial and industrial activity. Although, Ciudad Juarez were shown to influence the spatial PM₁₀ distribution in El Paso it was impractical to monitor the entire metropolitan area due to cost and political constraints in order to quantify the impact of CDJ. As Shown in Figure 11-8, the model reproduced two noticeable hot spots: first high concentration spot aroused at the surroundings of downtown extending to the BOTA port of entry spurring concentrations between 56.866 and 64.975 $\mu\text{g}/\text{m}^3$; Second and more pronounced high concentration spot prompted at the airport and circumvent areas with concentrations ranging from 57.168 to 84.475 $\mu\text{g}/\text{m}^3$. The outcome bearded the short-term spatial variations of averaged PM₁₀ concentrations for a localized area within larger metropolitan region and predicted a gradient decreasing with distance from the border. Moreover, the methodology captured the relative impact of emissions outside the localized monitoring network area and allowed for an optimal utilization of resources.

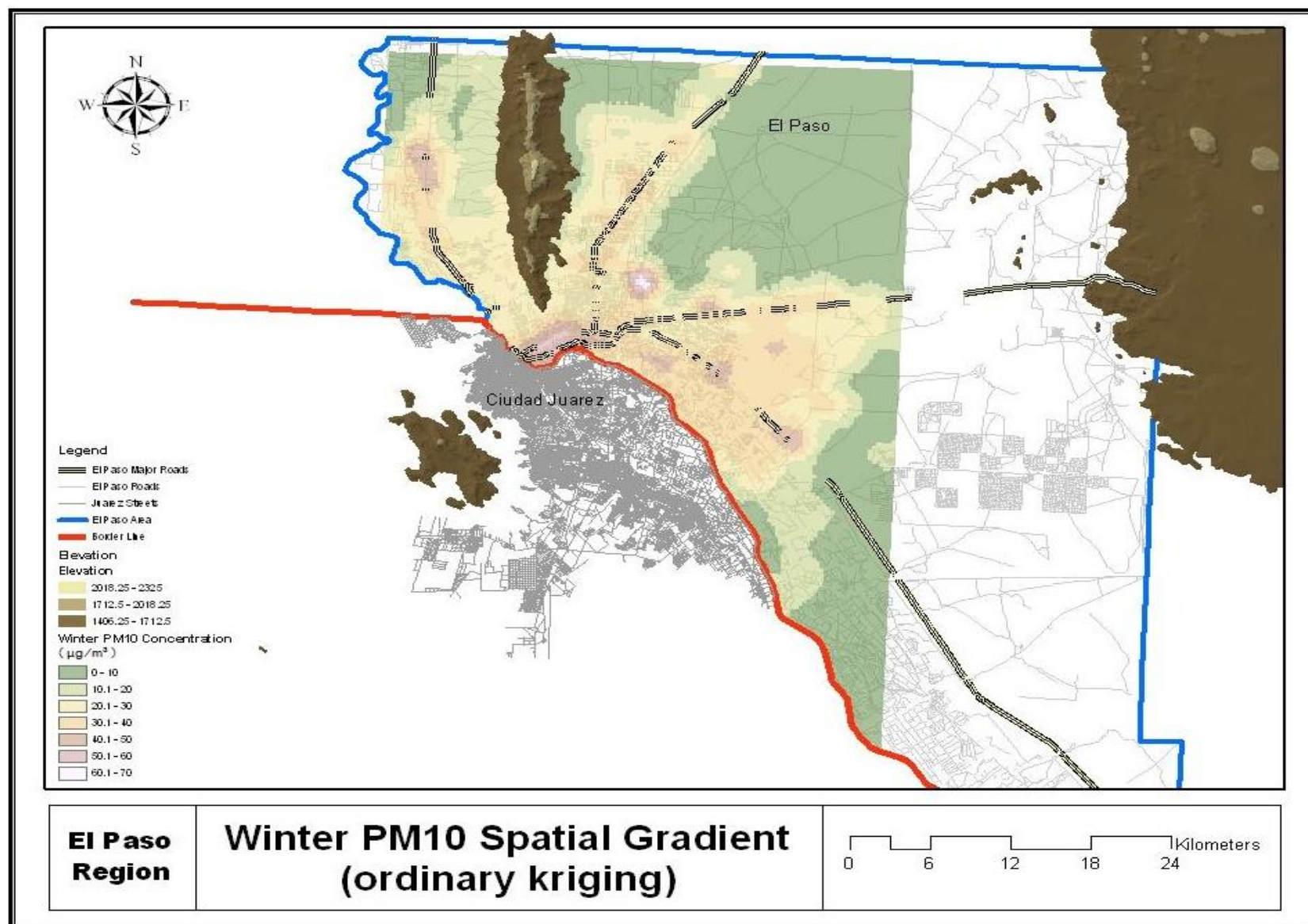


Figure 11-8. Winter PM10 spatial gradient using Ordinary Kriging interpolation method.

12 CONCLUSION

The exposure assessment method used to estimate levels of air pollutants is a critical component of health effect studies. Simple models such as proximity of households to roadways have been used in epidemiologic studies (English et al., 1999; and Hoek et al., 2002). Although PM_{10} levels in households have been found to correlate with distance to roadways and traffic density, the proximity method has been criticized due to a high potential for exposure misclassifications (Jerrett et al., 2005). Assigning exposure based solely on proximity measures ignores other traffic, land use and geographical variables that influence the exposure. On the other hand, the most sophisticated dispersion models that integrated meteorological data and atmospheric chemistry required many data inputs and months of training.

In this study LUR methods successfully predicted small-scale spatial variation of ambient PM_{10} concentrations and adequately reproduced the regional trends identified by the monitoring network. A PM_{10} monitoring network with a limited number of monitors as compared to other LUR studies was employed in this research. The total network coverage area was 651 Km^2 resulting in a monitoring density of one site per every 50 km^2 . Still a high explanatory power was achieved for all models as compared to other PM_{10} LUR modeling efforts in the literature. Moreover, the monitoring network employed in this study was designed based on sitting criteria that targeted both highly traffic and residential areas in addition to the border with the neighboring community of CDJ. The results suggest that the suitability of the land characteristics was captured by the network and less upon the total number of samples. All models (annual, spring, summer, fall, and winter) used the measured levels of a pollutant at sampling locations as the dependent variable in a multiple least-squares regression model. Although the independent

variables in each LUR model were unique, they were broadly categorized into four major classes: (1) land use, (2) traffic, (3) nearest distance to the border, and (4) population. Each LUR model contained definitions of one or more of these class variables depending upon the available data. The final LUR models contained, nearest distance to the border, group zoning, the sum of road's length, the sum of truck road's length, traffic counts and population that accounted for >75% of the variability in the sample PM₁₀ concentrations. Of these, a representative of land use variable was the most significant independent parameter as it explained the variability of PM₁₀ for the annual and all seasonal models. The nearest distance to the border was found to be strongly associated with the level of PM₁₀ for the summer and the annual concentrations, an effect of local geographic variables on land-use regression models. In addition, the contribution of diesel emissions from on-road vehicles to ambient PM particulate exposure gradients was demonstrated in the summer and fall models, as truck-traffic variables were paired with other parameters with explanatory power for such seasonal periods, thus coinciding to the suggestions by Levy et al. (2003) and Zhue et al. (2002b). Population density served as a surrogate for the general level of human activity (including traffic) in the vicinity of the monitoring site and therefore associated with the increased emissions for the fall, winter, and spring time periods.

Table 12-1. General overview of the annual and seasonal land use regression models of PM₁₀.

	MODELS				
	Annual	Spring	Summer	Fall	Winter
Variables	GZ5m, NLR5K, OND	NGZ15K, npop15k	lc2m,TR5m, OND	npop15k, ntr1k, NGZ3m	pop5m, vf15K, GZ5m
Interpolation Method	UK	UK	UK	OK	OK
R ²	89.9%	85.5%	74.6%	83.1%	86.5%
IQR (µg/m ³)	7.7	25.3	18.7	19.3	18.3
MAE (µg/m ³)	2.5	4.4	5.5	5.4	4.2
µ (µg/m ³)	20.5	21.9	12.4	14.3	17.5
σ (µg/m ³)	8.2	15.4	18.5	16.8	13.4
Range (µg/m ³)	13.5-42.7	6.5-68.5	10.4-83.5	7.6-87.7	7.9-94.2

The achievable explanatory power of the four seasonal models was found to vary from 74.6% for the summer model to 86.5% for the winter model; with an intermediate 85.5%, and 83.1% for the spring and fall models, respectively (see Table 12-1). The spring model accounted for the highest IQR and mean concentration of 25.3 $\mu\text{g}/\text{m}^3$ and 21.9 $\mu\text{g}/\text{m}^3$, respectively, although the standard deviation (15.40 $\mu\text{g}/\text{m}^3$) was the second lowest value and the range of concentrations (6.494 to 68.476 $\mu\text{g}/\text{m}^3$) the lowest among the four models. The summer model, on the other hand, had a more stabilized mean concentration of 12.4 $\mu\text{g}/\text{m}^3$ and IQR of 18.7 $\mu\text{g}/\text{m}^3$, but a more relevant concentration range from 10.4 to 85.5 $\mu\text{g}/\text{m}^3$ and consequently a higher standard deviation (18.5 $\mu\text{g}/\text{m}^3$) than the spring model. Both spring and summer models relied on universal kriging to project the spatial gradient of ambient particulate matter which outperformed the remaining interpolation techniques included in the analysis. In both cases universal kriging tended to overpredicted concentrations by 4.33 $\mu\text{g}/\text{m}^3$ for the spring model and 5.544 $\mu\text{g}/\text{m}^3$ for the summer model.

The PM_{10} surface area was generated with a 3528 grid points distributed orthogonally at a 610 m isotropic resolution capable of characterizing small-scale variations and capturing the relative impact of the bordering urban areas. The cohort study, however, had to be redefined with the exclusion of Montanavista site as it was consider outside the urban sprawl. During spring, 37% of the residents in ELP fell on concentrations that range from 40 to 50 $\mu\text{g}/\text{m}^3$ from which 1.5% was exposed to higher concentrations ranging from 50 to 60 $\mu\text{g}/\text{m}^3$. During the summer season, only 7% were exposed to concentrations ranging from 30 to 50 $\mu\text{g}/\text{m}^3$, however, 2.3% of the residents perceived an exposure increase in PM_{10} by 10 $\mu\text{g}/\text{m}^3$ contrary to 8% of the population who discern a decrease in PM_{10} exposure. The spatial range of particulate matter remained analogous for both projected maps, though the spring map presented a more spare

characterization pattern than the summer model, which somehow illustrated a more conglomerated pattern towards the international border. This distribution of the spatial gradient for both models corresponded to the meteorological conditions caused by high windspeed episodes. The fall model prompted the second lowest average concentration of $14.3 \mu\text{g}/\text{m}^3$, although a more disrupted PM_{10} range than the summer model was observed from 7.6 to $87.7 \mu\text{g}/\text{m}^3$ and consequently the second highest standard deviation and IQR of $16.8 \mu\text{g}/\text{m}^3$ and $19.3 \mu\text{g}/\text{m}^3$, respectively. Contrary to the spring and summer model, ordinary kriging outperformed universal kriging and the latter interpolation techniques for this particular model assuming that the concentrations were free from any structure. The fall model showed higher exposure concentrations than the summer model. That is, 33% of the population in ELP was exposed to higher concentrations than the previous season ranging from 40 to $60 \mu\text{g}/\text{m}^3$. In fact, the fall model had a more dispersed pattern of PM_{10} attributable to inversion events that limited the dilution of the pollutant. As a result, 84% of the residents in El Paso were exposed to an increment in PM_{10} concentrations of $10 \mu\text{g}/\text{m}^3$ or more. The winter model, predicted the highest concentration ranging from 7.9 to $94.2 \mu\text{g}/\text{m}^3$ and the second highest mean concentration of $17.501 \mu\text{g}/\text{m}^3$, with the lowest standard deviation of $13.4 \mu\text{g}/\text{m}^3$ and the lowest IQR of $18.3 \mu\text{g}/\text{m}^3$. That is, the variance among the concentration range was more balanced between one grid node and the rest. Nearly 4% of the residents experienced an increment of $10 \mu\text{g}/\text{m}^3$ from the previous season. The annual model showed a more steady set of values than the seasonal models. The annual PM_{10} surface tends to have a modest disrupted variation pattern when compared to any other season as the 3-year continuous temporal run of the monitoring data stabilized the monitor readings contributing to the proportion of explained variance. The annual model had a moderate mean concentration of $20.5 \mu\text{g}/\text{m}^3$ with a low standard deviation of $8.2 \mu\text{g}/\text{m}^3$, a narrow

concentration range of 13.5 to 42.7 $\mu\text{g}/\text{m}^3$, and a minimum IQR of 7.7 $\mu\text{g}/\text{m}^3$. Universal kriging tended to outperform ordinary kriging as this oversmooths and hence misses areas with unusually high particulate matter caused by omitting an external drift. Although universal kriging overpredicted PM_{10} concentrations for the most part by 2.5 $\mu\text{g}/\text{m}^3$, the projected map nevertheless, identified a southward pattern with 4% of the population experiencing concentrations from 35 to 45 $\mu\text{g}/\text{m}^3$.

Overall, high concentrations were perceived in the fall model which continues in the winter, lessens in the spring, and abruptly decreases in the summer. That is, fall and winter models accounted for higher concentration exposure as both seasons are strongly affected by the frequently occurring temperature inversion. The five models, including the annual model, revealed a southward PM pattern where all of the cases identified the downtown core area as a hot spot. It was estimated that 2.3, 84, and 4% of the El Paso residents were exposed to an increment of PM_{10} from spring to summer, summer to fall, and fall to winter season, respectively. According to the literature an increment of 10 $\mu\text{g}/\text{m}^3$ in particulate matter may be associated with; 1) decline of less than 2% in lung function (Schwartz et al., 1993), 2) 1 to 10% increase symptoms such as cough, combine lower respiratory symptoms, and asthma attacks (Koenig et al., 1993; Pope et al., 1991; Pope and Dockery, 1992; and Schwartz et al., 1994), 3) with a 10 to 25% increase in bronchitis or chronic cough (Schwartz et al., 1993; and Chesnut et al., 1991), 4) with an increase in daily mortality of 0.5 to 1.5% (Schwartz, 1991, 1993; and Schwartz and Dockery, 1992), and 5) with increases in admission rates of 0.8% (95% significant and CI: 0.5-1.2%) for congestive heart failure, 0.7% (95% significant and CI: 0.4-1.0%) for ischemic heart disease, and 0.2% (95% significant and CI: -0.2-0.6%) for cerebrovascular accidents (Morris 2001). However, to assess the real health effects of this increase in PM_{10} for

short- and long-term exposure on Hispanic asthmatic children based on residence location remains a further research study.

In general, these models demonstrate that PM_{10} can be predicted by land use regression in with traffic, population, land use, and geographical variables. Although each of the five models discussed in this study are somehow identical in variables or parameters, their similarity reinforces the relationship between PM_{10} and land use variables. Our findings concluded that a combination of smaller and larger scales forms the basis of appropriate prediction models. For El Paso area the buffer radii could be reasonably be restricted to 5000 m in accordance with the extraterritorial jurisdiction and urbanization. PM_{10} concentrations greatly increased in the downtown core due to increased human activity in the area and possibly emissions from Ciudad Juarez. Although our study is considered a relatively simple and brief analysis of the air pollution situations in El Paso, the results nevertheless demonstrate a method for predicting regional PM concentrations, using the high resolution GIS-based data and limited air pollution measurements.

One of the main attributes of the LUR methods showcased in this study was the capacity to capture localized complexities within regions. The results show that a stable predictive surface can be derived for the sub-section of a large metropolitan area. Although, the development of predictive models that could be employed interchangeably in different regions is desired, the ability to adapt local particulates in the models is one of the major advantages of regression mapping. The regression-based mapping technique employed in this study was capable of characterizing small-scale variations of PM_{10} in sections of urban areas and capturing the relative impact of the bordering urban areas. The method successfully accounted for PM_{10} emissions in Ciudad Juarez independently of the complexity and distinct characteristics as compared to El Paso. The use of such techniques in exposure assessment for short- and long-term

epidemiological studies will result in optimal monitoring campaigns and thus better resource managements. Health impact assessments of urban pollution, particularly PM_{10} , on particular sub-populations in localized areas greatly benefited from the applications of these methods. Because of the lack in the number of samples sites available for validation, and the different structures of the regression models, care is needed in comparing the results. Overall, however, it is apparent that the regression method gave extremely good predictions of the pollution levels at the reference sites.

REFERENCES

- Abbey D.E., Nishino N., McDonnell W.F., Burchette R.J., Knutsen S.F., Lawrence B.W., Yang, J.X. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal of Respiratory and Critical Care Medicine* 1999: 159: 373–382.
- Araín M.A., Blair R., Finkelstein N., Brook J.R., Sahsuvaroglu T., Beckerman B., Zhang L., Jerrett M. The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmos Environ* 2007: 41: 3453–3464.
- Bachman W., Sarasua W., Hallmark S., Guensler R. Modeling regional mobile source emission in a Geographic Information System Framework. *Transportation Research Part C*. 2000: 8: 205-229.
- Barna M. G., and Gimson N. R. Dispersion modeling of a wintertime particulate pollution episode in Christchurch, New Zeland. *Atmos Environ* 2002: 36: 3531-3544.
- Beelen R., Hoek G., Fischer P., van den Brandt P.A., Brunekreef B. Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmos Environ* 2007: 41: 1343–1358.
- Beelen R., Hoek G., Pebesma E., Vienneau D., de Hoogh K., Briggs D. J. Mapping of background air pollution at a fine spatial scale across the European Union. *Sci Total Environ* 2009: 407: 1852-1867.
- Bellander T., Berglind N., Gustavsson P., Jonson T., Nyberg F., Pershagen G., and Jarup L. Using geographical information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ Health Perspect* 2001: 109: 633–639.
- Brauer M., Hoek G., van Vliet P., Meliefste K., Fischer P., and Gehring U., et al. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 2003:14(2): 228–239.
- Briggs D., Collins S., Elliott P., Fischer P., Kingham S., and Lebret E., et al. Mapping urban air pollution GIS: a regression-based approach. *Int J Geogr Inf Sci* 1997:11(7): 699–718.
- Briggs D. J., de Hough C., Gulliver J., Wills J., Elliott P., Kingham S., and Smallbone K. A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments. *Sci Total Environ* 2000: 253:151–167.
- Briggs D. J., de Hough C., Gulliver J., Wills J., Elliott P., Kingham S., and Smallbone K. A regression-based method for mapping traffic-related air pollution: Application and testing in four contrasting urban environments. *Sci Total Environ* 2002: 253:151–167.

- Briggs D. The role of GIS: coping with space (and time) in air pollution exposure assessment. *J Toxicol Environ Health A* 2005: 68:1243–1261.
- Briggs D.J., Aaheim A., Dore C., Hoek G., Petrakis M., Shaddick G. Air pollution modelling for support to policy on health and environmental risks in Europe. Final Report Section 6. Imperial College, London. 2005:EVK2-2002-00176: Available from: <http://www.apmosphere.org/>.
- Briggs D., Hoogh C.D. and Gulliver J., Comparative assessment of GIS-based methods and metrics for modeling exposure to air pollution. *J Toxicol Environ Health* in press.
- Brown M.J., Muller C., Wang G., Costigan K. Meteorological simulations of boundary-layer structure during the 1996 Paso del Norte Ozone Study. *Sci Total Environ* 2001: 33: 276:111 – 133.
- Bruckman L., Dickson R.J., Wilkonson J.G. The use of GIS software in the development of emission inventories and emission modeling. *J Air Waste Manage Assoc* 1992:16: 118-132.
- Burrough P., and McDonnell R. Principles of Geographical Information Systems. Oxford University Press, New York, 1998.
- Carter-Pokras O., & Gergen P. J. Reported asthma among Puerto Rican, Mexican-American, and Cuban children, 1982 through 1984. *Am J of Pub Health* 1993: 83: 580-582.
- Chesnut L.G., Schwartz J., Savitz D. A., and Burchfiel C. M. Pulmonary function and ambient particulate matter. Epidemiological evidence from NHANES I. *Arch Environ Health* 1991: 46: 135-144.
- Currey R., and Pumfrey R. Improving air quality in Paso del Norte. In R. Pumfrey (Ed.), the US-Mexican border environment: binational air quality management. 2006: (Vol. Series no. 14, pp 109-122). SCERP Monograph series.
- Cyrys J., Heinrich J., Hoek G., Meliefste K., Lewne M., Gehring U., Bellander T., Fischer P., van Vliet P., Brauer M., Wichmann H.E., Brunekreef B. Comparison between different traffic-related particle indicators: elemental carbon (EC), PM_{2.5} mass, and absorbance. *J Expo Anal Environ Epidemiol* 2003: 13: 134–143.
- Dai J., Rocke D.M. A GIS based approach to spatial allocation of area source solvent emission. *Environmental Modelling & Software* 2000: 15: 293-302.
- D’Amato G. Urban air pollution and plant-derived respiratory allergy. *Clin Exp Allergy* 2000: 30: 628–636.
- Dattner S. El Paso/Juárez 1990 PM₁₀ Receptor Modeling Feasibility Study. Research & Technology Section, Monitoring Operations Division. TNRCC, December 1990,

AS-43.

- De Medeiros A., Gouveia N., Machado R., de Souza M., Alencar G., Novaes H., et al. Traffic-related air pollution and perinatal mortality: a case-control study. *J Environ Health Perspec* 2009; 117(1): 127-132.
- Dockery D. W. Epidemiologic evidence of cardiovascular effects of particulate air pollution. *Environ health perspect* 2001; 109(4): 483-486.
- Einfeld W. and Church H. Short-term winter season PM₁₀ study. Winter season air pollution in El Paso-Ciudad Juarez, EPA-906-R-95-001, 1995.
- English P., Neutra R., Scalf R., Sullivan M., Waller L., and Zhu L. Examining associations between childhood asthma and traffic flow using a geographic information system. *Environ Health Perspect* 1999;107: 761–767.
- Flores G., Abreu M., Olivar M A., Kastner B. Access barriers to health care for Latino children. *Arch Pediatr Adolesc Fam Med* 1998;30:196-205.
- Fox, J. Regression Diagnostics. Quantitative Applications in the Social Sciences Series No. 79. Provides a thorough review of methods of testing the assumptions of regression models. Thousand Oaks, CA: Sage Publications, 1991.
- Fox, J. Nonparametric simple regression. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences Series No.130. Covers local polynomial multiple regression in detail, 2000.
- Fox, J. (2005). Linear models, problems. Kimberly Kempf-Leonard, ed., *Encyclopedia of Social Measurement, Vol. 2*. Amsterdam: Elsevier, 2005, pp. 515-522.
- Gertler A.W. Diesel vs gasoline emissions: does PM from diesel or gasoline vehicles dominate in the U.S. In: Joumard R. (Ed.). Proceedings of the 12th International Symposium: Transport and Air Pollution INRETS, Arcueil, France, 2003, pp. 185–192.
- Gilbert N. L., Goldberg M. S., Beckerman B., Brook J. R., and Jerrett M. Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *J Air Waste Manage Assoc* 2005; 55:1059–1063.
- Gonzales M., Qualls C., Hudgens E.L., Neas L.N. Characterization of a spatial gradient of nitrogen dioxide across a United States –Mexico border city during winter. *Sci Total Environ* 2005; 337: 163–173.
- Hallmark S. and O’Neil W. Integrating Geographic Information Systems for transportation and air quality models for microscale analysis. Transportation Research Record 1551, Transportation Research Board, Washington, 1996, pp. 40-48.

- Han X. & Naeher L. A review of traffic-related air pollution exposure assessment studies in the developing world. *Environmental International*. 2006: 11: 98-117.
- Henderson S., Beckerman B., Jerrett M., Brauer M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ SciTechnol* 2007: 41: 2422–2428.
- Hewitt, C. N. Spatial variations in nitrogen dioxide concentration in an urban area. *Atmospheric Environment* 1991: 25B: 429- 34.
- Hochadel M., Heinrich J., Gehring U., Morgenstern V., Kuhlbusch T., Link E., Wichmann H.E., Kramer U. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. *Atmos Environ* 2006: 40: 542–553.
- Hoek G., Brunekreef B., Goldbohm S., Fischer P., and van den Brandt P.A. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet* 2002a: 360(9341): 1203–1209.
- Hoek G., Meliefste K., Cyrys J., Lewne´ Brauer M., Fischer P., Gehring U., van Vliet P., HeinrichJ., Bellander T., and Brunekreef B. Spatial variability of fine particle concentrations in three European countries. *Atmos Environ* 2002b: 36: 4077–4088.
- Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ* 2008: 42: 7561–78.
- Hosiokangas J., Ruuskanen J., and Pekkanen J. Effects of soil dust episodes and mixed fuel sources on source apportionment of PM₁₀ particles in Kuopio, Finland. *Atmos Environ* 1999: 33:3821-3829.
- Hrubá F., Fabianova E., Koppova K., and Vandenberg J.J. Childhood respiratory symptoms, hospital admissions, and long-term exposure to airborne particulate matter. *J Expos Anal Environ Epidemiol* 2001: 11: 33–40.
- Hu S., McDonald R., Martuzevicius D., Biswas P., Grinshpun S.A., Kelly A., et al. UNMIX modeling of ambient PM_{2.5} near an interstate highway in Cincinnati, OH, USA. *Atmos Environ* 2006: 40(S2): 378–395.
- INEGI. Instituto Nacional de Estadística y Geografía. Retrieved April 22, 2009, from <http://www.inegi.org.mx>.
- Jensen S.S., Berkowicz R., Hansen H.S., Hertel O. A Danish decision-support GIS tool for management of urban air quality and human exposures. *Transportation Research Part D*. 2001: 6: 229-241.

- Jeon S. J., Meuzelaar H.L., Sheya S.A., Lighty J.S., Jarman W.M., Kasteler C., et al. Exploratory studies of PM₁₀ receptor source profiling by GC/MS and principal component analysis of temporally and spatially resolved ambient samples. *J Air Waste Manage Assoc* 2001: 51: 766– 84.
- Jerrett M., Sears M., Giovis C., Burnett R., Kanaroglou P., Elliott S., Cakmak S., Gosselin P., Bedard Y., Maclachlan J., and Cole D. Intraurban air pollution exposure and asthma prevalence in Hamilton, Canada. Presented at the American Association of Geographer's Conference, Los Angeles, USA, 2002.
- Jerrett M., Arain A., Kanaroglou P., Beckerman B., Potoglou D., Sahsuvaroglu T., Morrison J., and Giovis C. A review and evaluation of intraurban air pollution exposure models. *J Expos Anal Environ Epidemiol* 2005: 15:185–204.
- Jerrett M., Arain M.A., Kanaroglou P., Beckerman B., Crouse D., Gilbert D., Brook J.R., Finkelstein N., Finkelstein M.M. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health Part A* 2007: 70: 200–212.
- Kanaroglou P., Jerrett M., Morrison J., Beckerman B., Arain A., Gilbert N., and Brook J. Establishing an air pollution monitoring network for intra-urban population exposure assessment. Proceedings of the transport and air pollution conference, Avignon, France, 2003.
- Kelsall J.E., Samet J.M., Zeger S.L., Xu J. Air pollution and mortality in Philadelphia: 1974–1988. *Am J Epidemiol* 1997: 146: 750–762.
- Koenig J. Q., Larson T. V., Hanley Q. S., et al. Pulmonary function changes in children associated with fine particulate matter. *Environ Res* 1993: 63: 26–38.
- Lenschow P., Abraham H. J., Kutzner K., Lutz M., Preub D. J., and Reichenbacher W. Some ideas about the sources of PM₁₀. *Atmos Environ* 2001: 35: S23-S33.
- Levy J. I., Bennett D.H., Melly S.J., Spengler J.D. Influence of traffic patterns on particulate matter and polycyclic aromatic hydrocarbon concentrations in Roxbury, Massachusetts. *J Expo Anal Environ Epidemiol* 2003: 13: 364– 71.
- Li W., Orquiz R., Garcia J., Espino T., Pingitore N., Gardea-Torresdey J., et al. Analysis of temporal and spatial dichotomous PM air samples in the El Paso-Cd. Juarez air quality basin. *J of the Air & Waste Manage Assoc* 2001: 51(11): 1551-1560.
- Lindley S.J., Conlan D.E., Raper D.W., Watson A.F.R. Uncertainties in the compilation of spatially revolved emission inventories: Evidence from a comparative study. *Atmos Environ* 2000: 34: 375-388.
- Mazzera D. M., Lowenthal D.H., Chow J.C., and Watson J.G. Sources of PM₁₀ and sulfate aerosol at McMurdo stations, Antarctica. *Chemosphere* 2001: 45: 347-356.

- MacDonald C.P., Roberts P.T., Main H.H., Dye T.S., Coe D.L., Yarbrough J. The 1996 Paso del Norte Ozone Study: analysis of meteorological and air quality data that influence local ozone concentrations. *Sci Total Environ* 2001: 276: 93-109.
- Moore D.K., Jerrett M., Mack W.J., Kunzli N. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *J Environ Monitor* 2007: 9: 246–252.
- Morris R. D. Airborne particulates and hospital admissions for cardiovascular disease: a quantitative review of the evidence. *Environ health perspect* 2001: 109(4): 495-500.
- Mulholland J.A., Butler A.J., Wilkinson J.G., Russell A.G., and Tolbert P.E. Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications. *J Air Waste Manage Assoc* 1998: 48: 418–426.
- Nafstad P., Haheim L.L., Oftedal B., Gram F., Holme I., Hijermann I., and Leren P. Lung cancer and air pollution: a 27 year follow up of 16209 Norwegian men. *Thorax* 2003:58: 1071–1076.
- National Resource Council. Modeling Mobile Source Emissions. Washington DC. National Academy Press, 2005:151–152.
- Noble C., Mukerjee S., Gonzales M., Rodes C., Lawless P., Natarajan S., et al. Continuous measurements of and relationship between fine and ultrafine particulate matter, criteria air pollutants and meteorological conditions in El Paso, Texas. *Atmos Environ* 2003: 37: 827–40.
- Pardoe I. Applied regression modeling: A business approach. John Wiley, 2006, pp. 166-168.
- Pope C. A., Dockery D. W., Sengler J. D., and Raizenne M. E. Respiratory health and PM₁₀ pollution: A daily time series analysis. *Am Rev Respir Dis* 1991: 144: 668-674.
- Pope CA III, Dockery DW. Acute health effects of PM₁₀ pollution on symptomatic and asymptomatic children. *Am Rev Respir Dis* 1992:145:1123-1128.
- Pope CA III, Kanner RE. Acute effects of PM₁₀ pollution on pulmonary function of smokers with mild to moderate chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1993: 147:1336-1340.
- Pope CA III, Dockery DW, Schwartz J. Review of epidemiologic evidence of health effects of particulate air pollution. *Inhal Toxicol* 1995a: 7:1–18.

- Pope CA III, Thun MJ, Namboodiri M, Dockery DW, Evans JS, Speizer FE, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med* 1995b; 151: 669–674.
- Pope III CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002; 287: 1132–41.
- Pope, C.A. and Dockery, D.W. Health effects of fine particulate air pollution: lines that connect. *Journal of Air and Waste Management Association*. 2006: 56: 709-742.
- Raizenne M.E., Burnett R.T., Stern B., Franklin C.A., Spengler J.D. Acute lung function responses to ambient acid aerosol exposures in children. *Environ Health Perspect* 1990: 79:179-185.
- Ramos I., May M., Ramos K. Environmental health training of Promotoras in Colonias along the Texas-Mexico border. *Am J of Pub Health* 2001: 91(4): 568-570.
- Riedl M., Diaz-Sanchez D. Biology of diesel exhaust effects on respiratory function. *J Allergy Clin Immunol* 2005: 115: 221–228.
- Ritz B., Yu F., Chapa G., and Fruin S. Effects of air pollution on preterm birth among children born in Southern California between 1989 and 1993. *Epidemiology* 2000: 11: 502–511.
- 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 analyses. *J. Expos Sci Environ Epidemiol* 2006: 16:106–114.
- Ross Z., Jerrett M., Ito K., Tempalski B., Thurston G. D. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmos Environ* 2007: 41: 2255-2269.
- Ryan P. H., LeMasters G. K., Biswas P., Levin L., Hu S., Lindsey M., Bernstein D. I., Lockey J., Villareal M., Hershey G. K. K., and Grinshpun S. A. A comparison of proximity and land use regression traffic exposure models and wheezing in infants. *Environ Health Perspect* 2007: 115:278–284.
- Ryan P.H., LeMasters G.K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal Toxicol* 2007: 19 (Suppl. 1): 127–133.
- Seigneur C. Current understanding of ultrafine particulate matter emitted from mobile sources. Coordinating Research Council, 2008, Report-CRC-A66 (Document CP274-08-01), 1-41.
- Seinfeld J. H., Pandis SN. Atmospheric chemistry and physics. From air pollution to

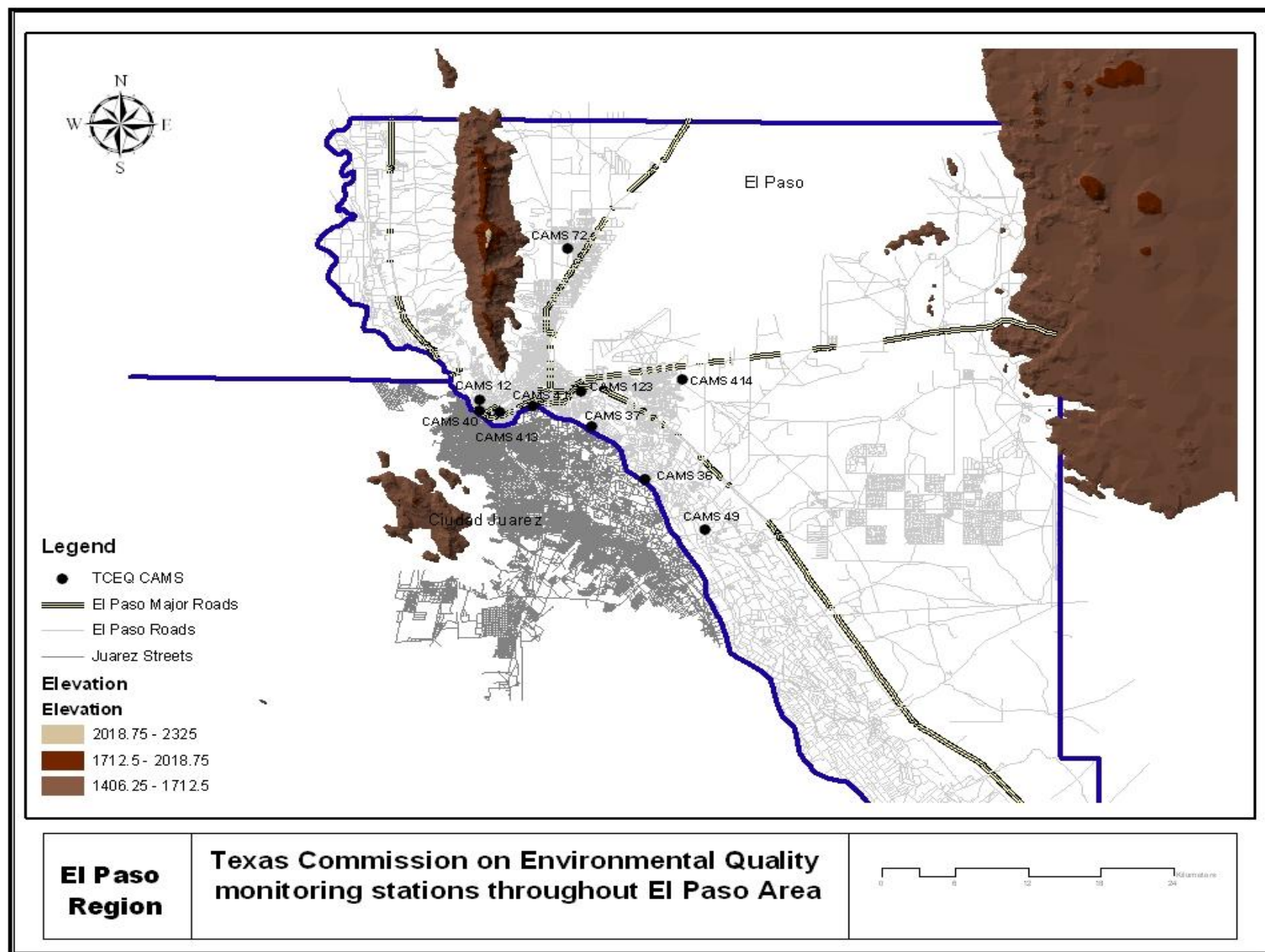
- climate change. John Wiley & Sons, Inc, 1997, pp. 1326.
- Schwartz J., Spix C., Wichmann H.E., Malin E. Air pollution and acute respiratory illness in five German communities. *Environ Res* 1991; 56: 1-14.
- Schwartz J., and Dockery D. W. Particulate air pollution and daily mortality in Steubenville, Ohio. *Am J Epidemiol* 1992a: 135: 12-19.
- Schwartz J., and Dockery D. W. Increased mortality in Philadelphia associated with daily air pollution concentrations. *Am Rev Respir Dis* 1992b: 145: 600-604.
- Schwartz J., Slater D., Larson T.V., Pierson W.E., Koenig J.Q. Particulate air pollution and hospital emergency visits for asthma in Seattle. *Am Rev Resp Dis* 1993: 147:826-831.
- Schwartz J., Dockery D.W., Neas L.M., Wypij D., Ware J.H., Spengler J.D., Koutrakis P., Speizer F.E., Ferris B.G. Jr. Acute effects of summer air pollution on respiratory symptom reporting in children. *Am J Respir Crit Care Med* 1994:150:1234-1242.
- Sherrill D., Stein R., Kurzius-Spencer M., Martinez F.. On early sensitization to allergens and development of respiratory symptoms. *Clin Exp Allergy* 1999: 29: 905–911.
- Souleyrette R.R, Sathisan S.K., James D.E., Lim S. GIS for transportation and air quality analysis. Proceedings of the National Conference on Transportation Planning and Air Quality. 1992, ASCE, pp. 182-194.
- Stopher P.R., Hartgen D.T., Li Y. SMART: Simulation Model for Activities Resource and Travel. Transportation, vol. 23, 1996. Kluwer Academic Publisher, pp. 293-312.
- Su, J.G., Brauer, M., Ainslie, B., Steyn, D., Larsson, T., Buzzelli, M. An innovative land use regression model incorporating meteorology for exposure analysis. *Sci Total Environ* 2008: doi:10.1016/j.scitotenv2007.10.032.
- Texas Natural Resource Conservation Commission. Air Pollution Trends in Texas, Nonattainment Areas. www.tnrcc.state.tx.us/air/monops/eloz.html; Dec 1999.
- Thorpe A. and Harrison R. Sources and properties of non-exhaust particulate matter from road traffic: A review. *Sci Total Environ* 2008: 400: 270-282.
- U.S. Census Burueau. (2009). Retreived April 22, 2009, from El Paso city, Texas-Fact sheet: http://factfinder.census.gov/el_paso-texas.
- U.S. Environmental Protection Agency (U.S. EPA). National Ambient Air Quality Standards (NAAQS) for PM and ozone at <http://www.epa.gov/ttn/oarpg/naaqsfin/>, 2006.

- Vardoulakis S., and Kassomenos P. Sources and factors affecting PM₁₀ levels in two European cities: Implications for local air quality management. *Atmos Environ* 2008; 42(17): 3949-3963.
- Vedal S., Blair J., and Manna B. Adverse respiratory health effects of ambient inhalable particle exposure. *Air Waste Manage Assoc* 1991: 91-810.56.
- Vellman P.F., and Welsch R. E. Efficient computing in regression diagnostics: *American Statistician*, 1998, pp. 234-242.
- Venkatram A. Validation of Concentrations estimated from air dispersion modeling for source-receptor distances of less than 100 meters. Sacramento, California, California Air Resources Board, Research Division. 2003.
- Venn A., Lewis S., Cooper M., Hubbard R., Hill I., Boddy R., et al. Local road traffic activity and the prevalence, severity, and persistence of wheeze in school children: combined cross sectional and longitudinal study. *Occup Environ Med* 2000;57: 152–158.
- Venn A.J., Lewis S.A., Cooper M., Hubbard R., and Britton J. Living near a main road and the risk of wheezing illness in children. *Am J Respir Crit Care Med* 2001;164: 2177–2180.
- Wayne E., Hugh W.C. Winter air pollution in El Paso-Cd. Juárez: A review of air pollution studies in an international airshed. Prepared by Environmental Characterization Department for the US EPA, Region 6. Dallas, Texas. Report No. SAND95-0273, March 1995.
- Westerdahl D., Fruin S., Sax T., Fine P.M., and Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmos Environ* 2005; 39: 3597–3610.
- Wikipedia. (2009). Retrieved April 22, 2009, from El Paso, Texas-Wikipedia: http://en.wikipedia.org/wiki/El_Paso_Texas.
- Wilbur and Smith. (2006). El Paso border improvement plan. El Paso Metropolitan Planning Organization.
- Wilkinson P., Elliott P., Grundy C., et al. Case-control study of hospital admission with asthma in children aged 5–14 years: relation with road traffic in north west London. *Thorax* 1999; 54: 1070–4.
- Wilson W., Chow J., Claiborn C., Fusheng W., Engelbrecht J., Watson J. Monitoring of particulate matter outdoors. *Chemosphere* 2002; 49: 1009-1043.

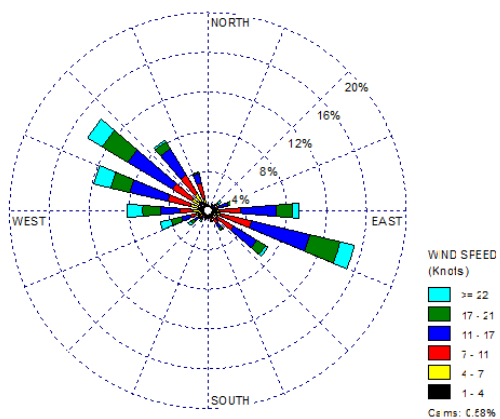
- Zanobetti A., Schwartz J., Samoli E., Gryparis A., Touloumi G., Peacock J., Anderson R. H., Le Tertre A., Bobros J., Celko M., Goren A., Forsberg B., et al. The temporal pattern of respiratory and heart disease mortality in response to air pollution. *Environ health perspect* 2003; 111(9): 1188-1193.
- Zeger S.L., Dominici F., Samet J. Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology*. 1999; 10: 171–175.
- Zhu Y., Hinds W., Kim S., Shen S., and Sioutas C. Study of ultrafine particules near a major highway with heavy-duty diesel traffic. *Atmos Environ* 2002;36: 4323–4335.

Appendix A

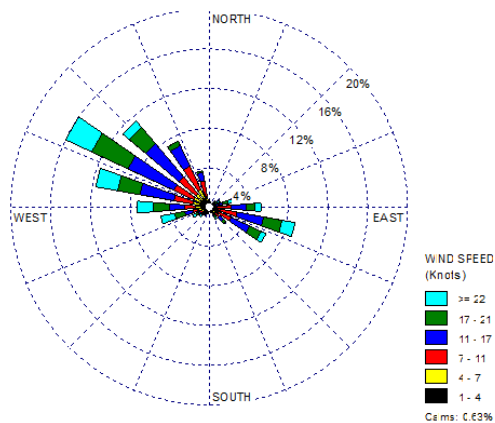
**Annual and Seasonal Wind Roses from the Texas Commission on Environmental Quality
fix stations distributed throughout the El Paso region.**



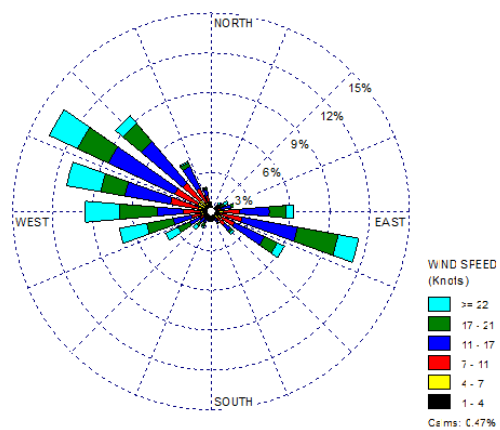
CAMS 12 ANNUAL WINDROSE



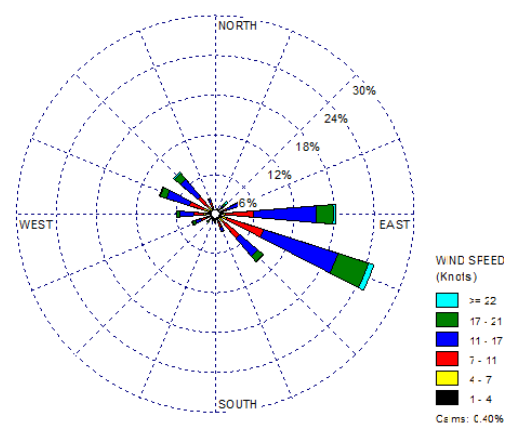
CAMS 12 WINTER SEASON WINDROSE



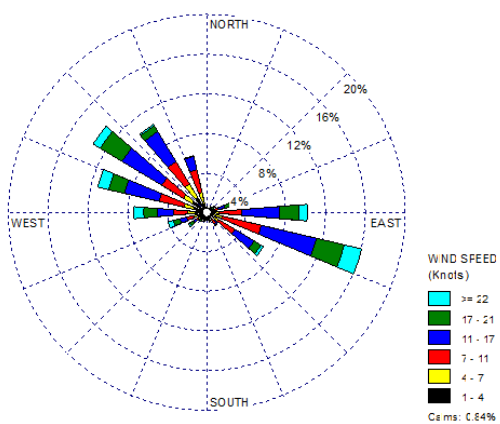
CAMS 12 SPRING SEASON WINDROSE



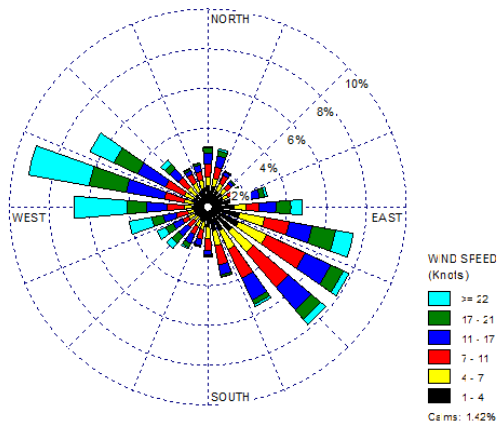
CAMS 12 SUMMER SEASON WINDROSE



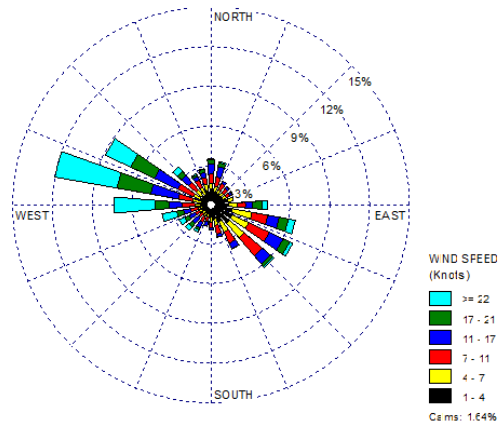
CAMS 12 FALL SEASON WINDROSE



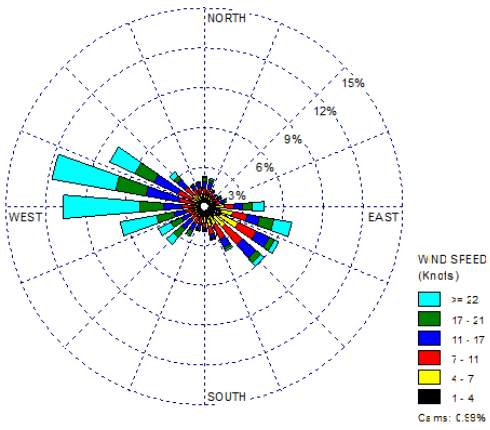
CAMS 36 ANNUAL WINDROSE



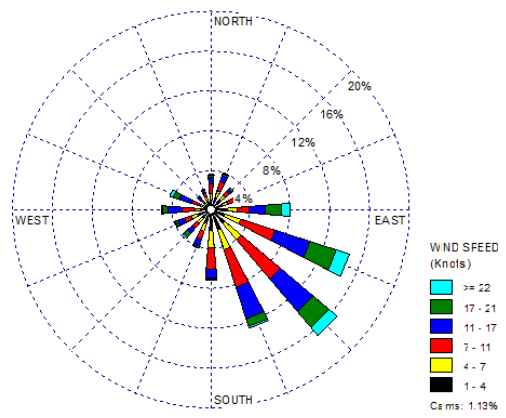
CAMS 36 WINTER SEASON WINDROSE



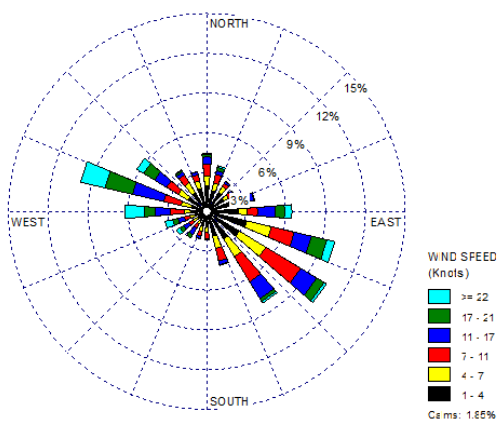
CAMS 36 SPRING SEASON WINDROSE



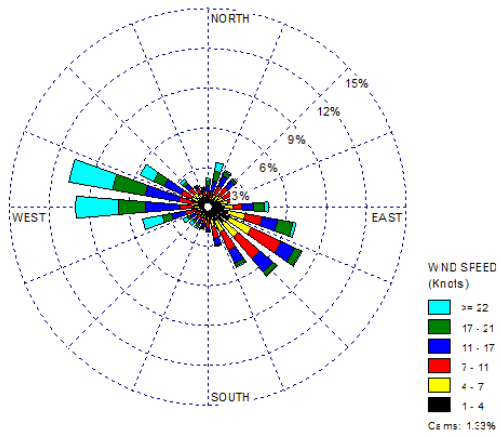
CAMS 36 SUMMER SEASON WINDROSE



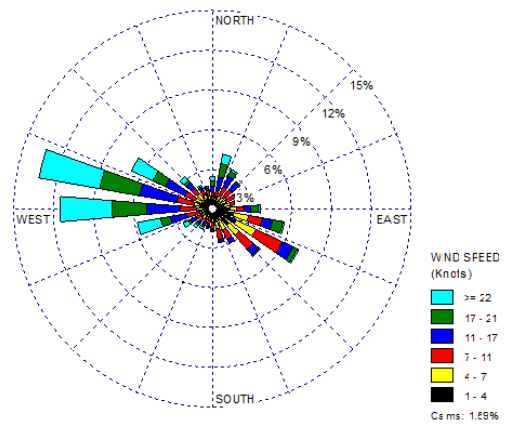
CAMS 36 FALL SEASON WINDROSE



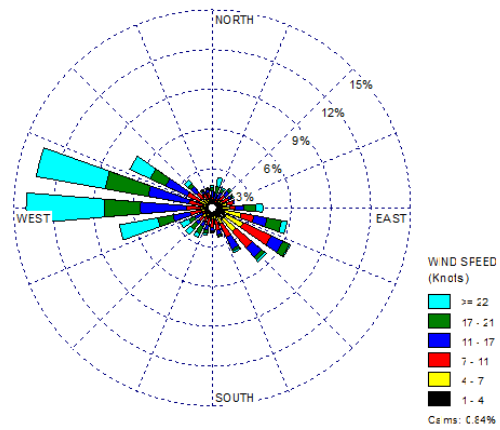
CAMS 37 ANNUAL WINDROSE



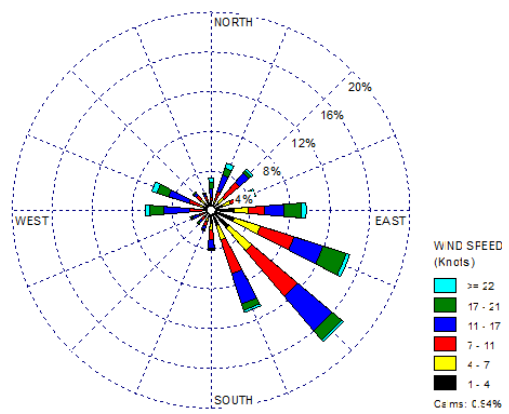
CAMS 37 WINTER SEASON WINDROSE



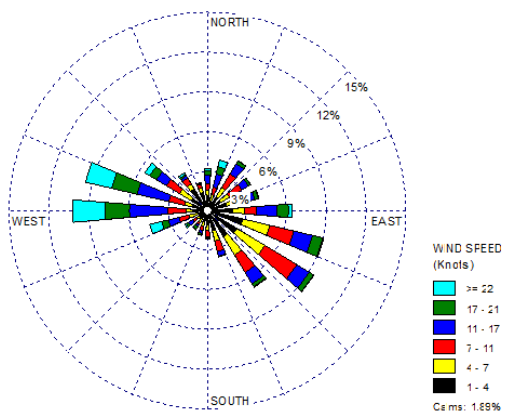
CAMS 37 SPRING SEASON WINDROSE



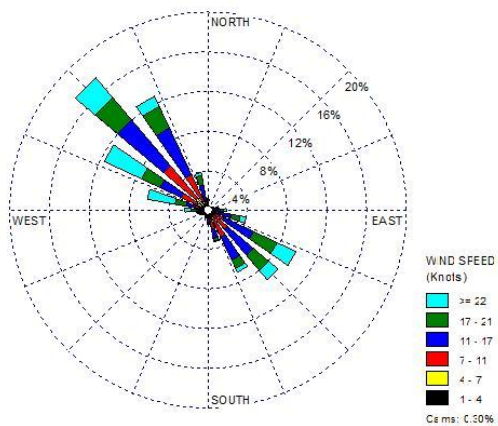
CAMS 37 SUMMER SEASON WINDROSE



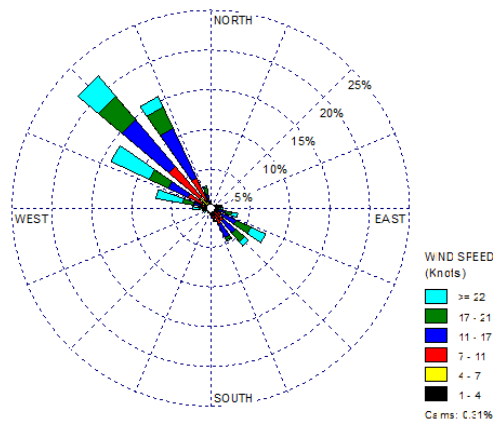
CAMS 37 FALL SEASON WINDROSE



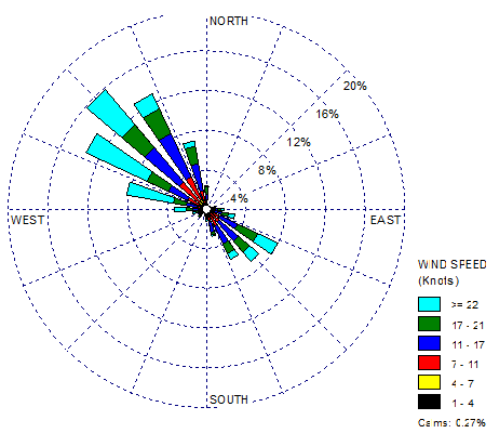
CAMS40 ANNUAL WINDROSE



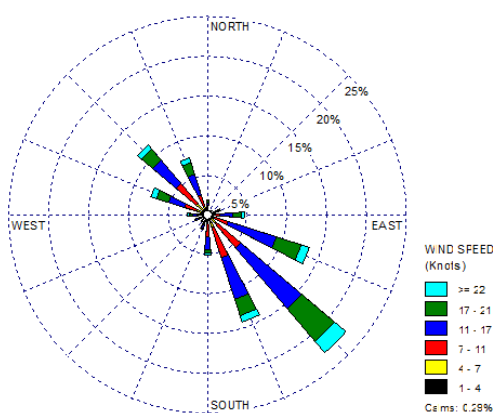
CAMS 40 WINTER SEASON WINDROSE



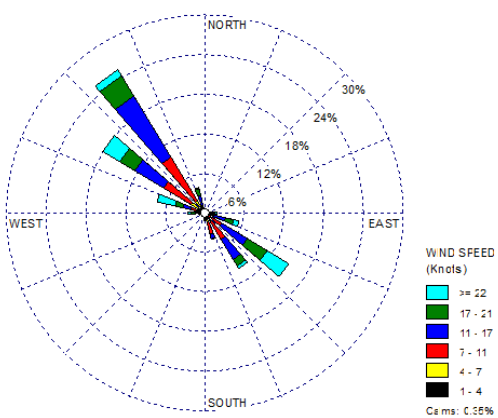
CAMS 40 SPRING SEASON WINDROSE



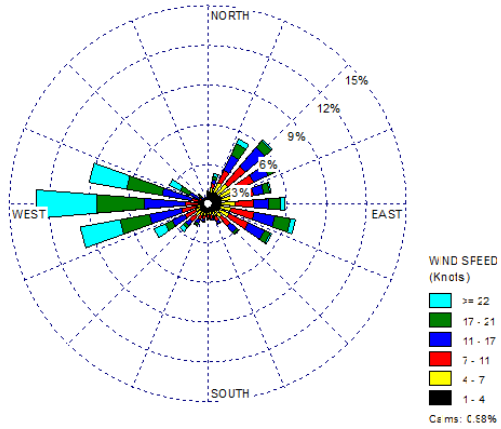
CAMS 40 SUMMER SEASON WINDROSE



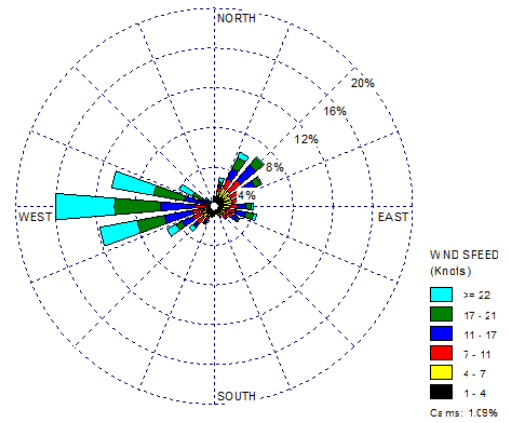
CAMS 40 FALL SEASON WINDROSE



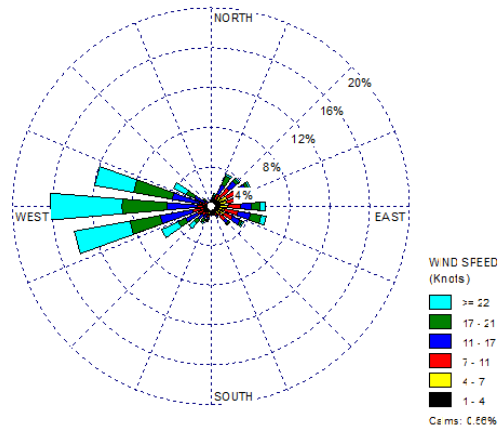
CAMS 41 ANNUAL WINDROSE



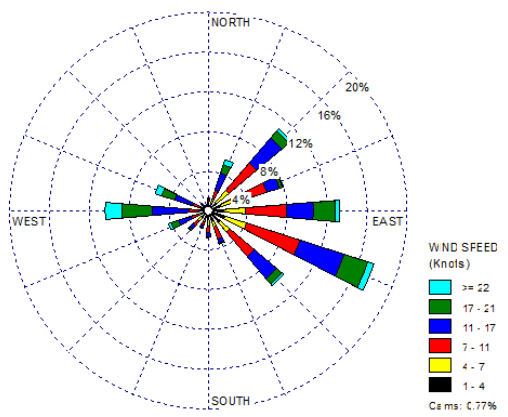
CAMS 41 WINTER SEASON WINDROSE



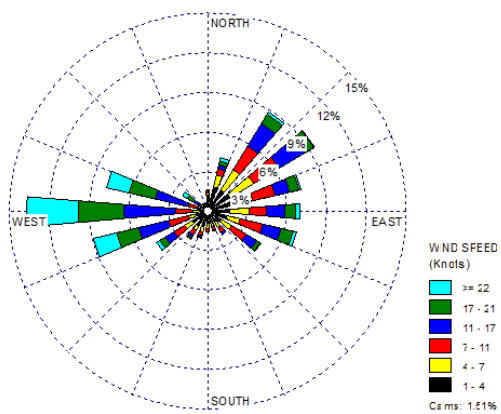
CAMS 41 SPRING SEASON WINDROSE



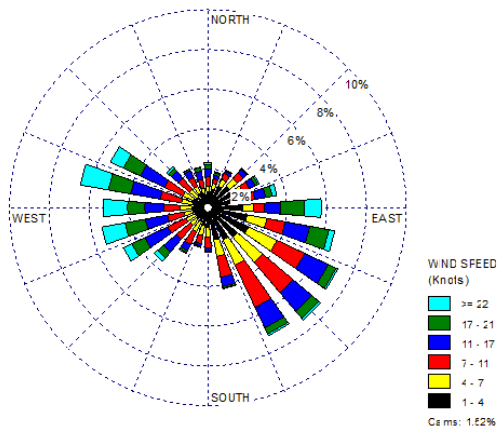
CAMS 41 SUMMER SEASON WINDROSE



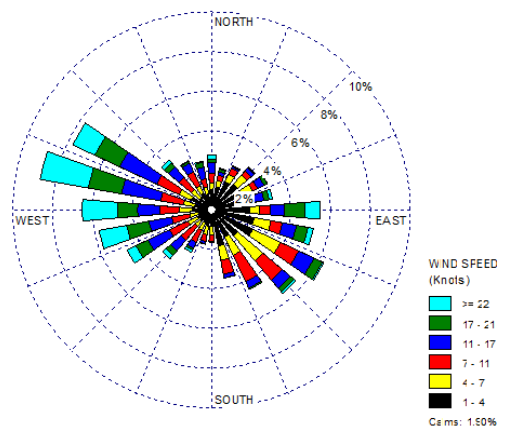
CAMS 41 FALL SEASON WINDROSE



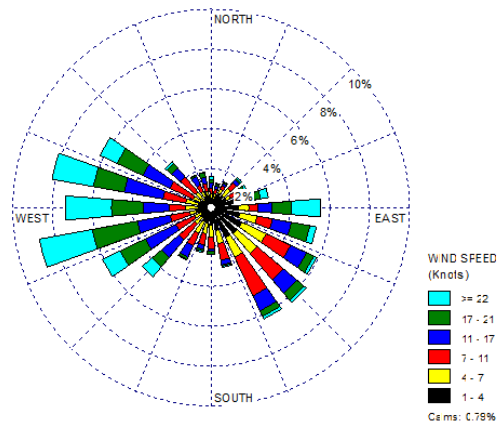
CAMS 49 ANNUAL WINDROSE



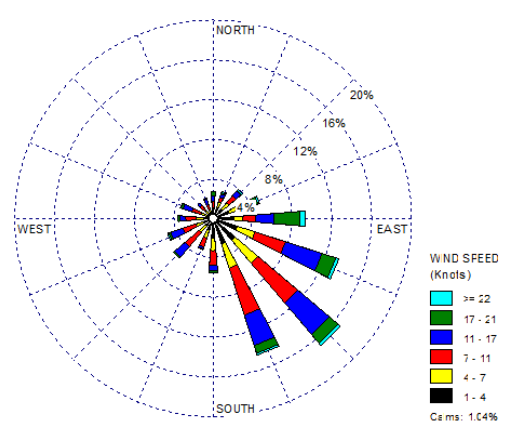
CAMS 49 WINTER SEASON WINDROSE



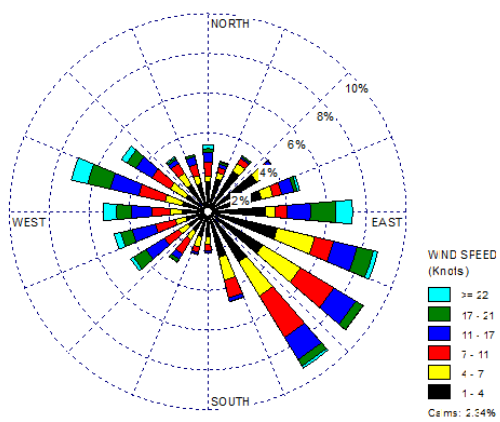
CAMS 49 SPRING SEASON WINDROSE



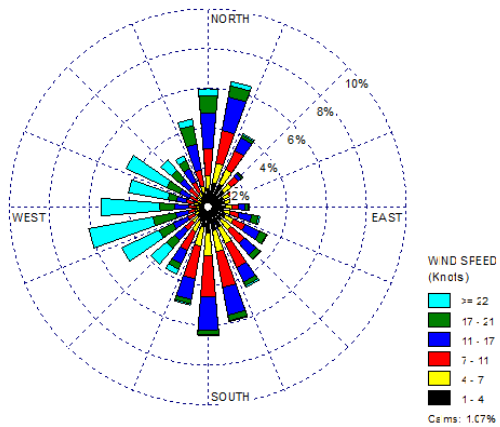
CAMS 49 SUMMER SEASON WINDROSE



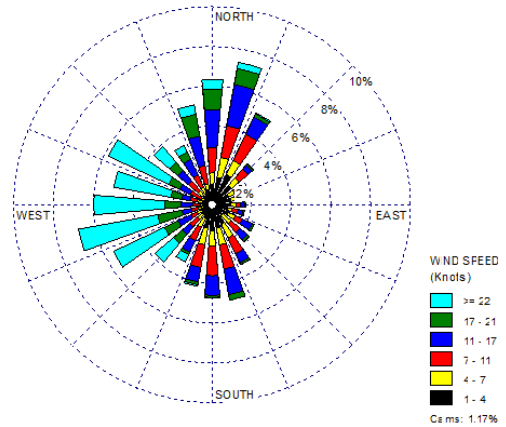
CAMS 49 FALL SEASON WINDROSE



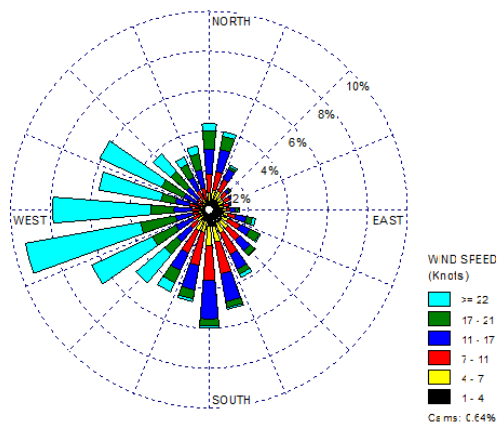
CAMS 72 ANNUAL WINDROSE



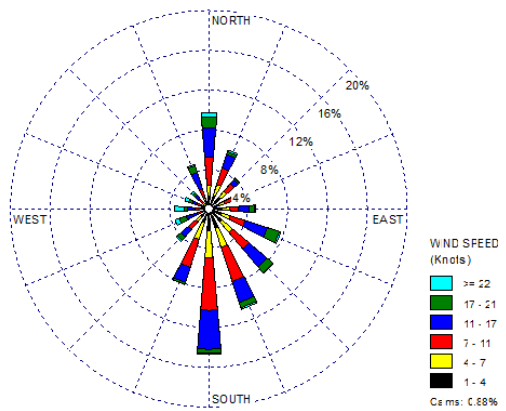
CAMS 72 WINTER SEASON WINDROSE



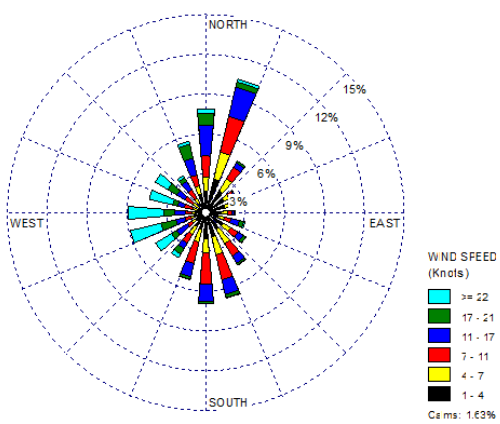
CAMS 72 SPRING SEASON WINDROSE



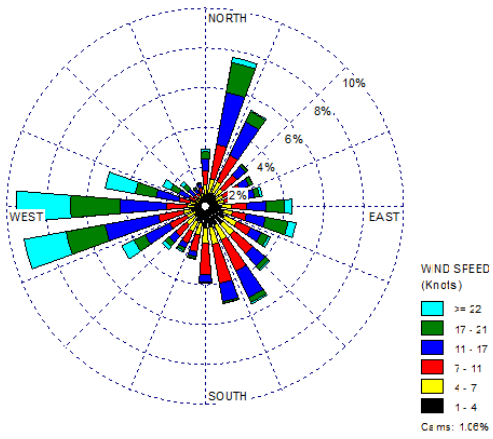
CAMS 72 SUMMER SEASON WINDROSE



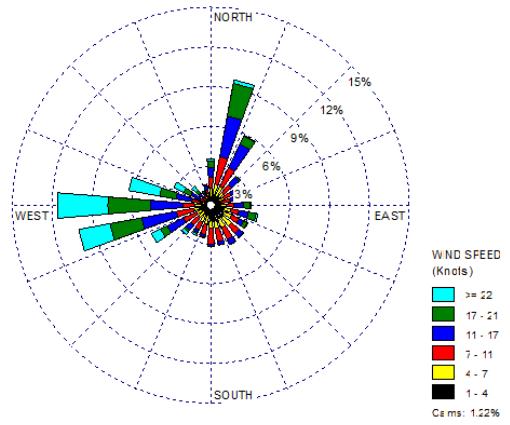
CAMS 72 FALL SEASON WINDROSE



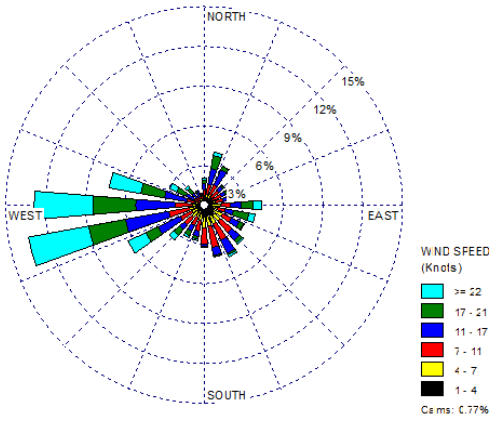
CAMS 123 ANNUAL WINDROSE



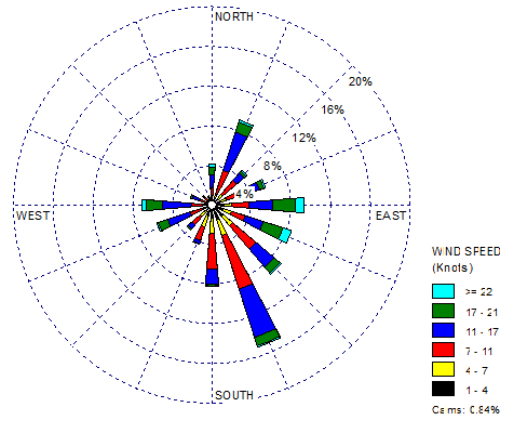
CAMS 123 WINTER SEASON WINDROSE



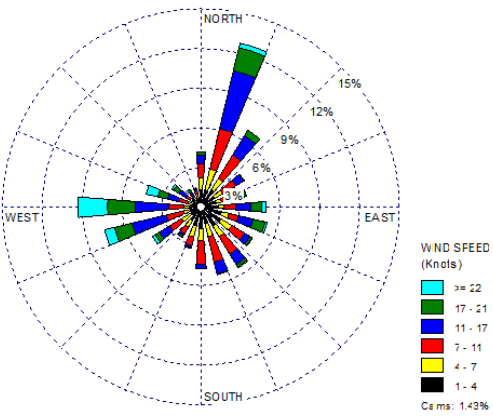
CAMS 123 SPRING SEASON WINDROSE



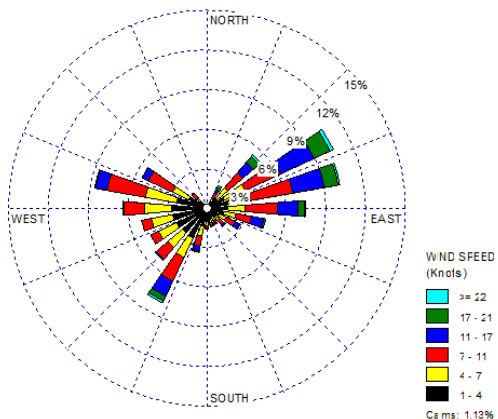
CAMS 123 SUMMER SEASON WINDROSE



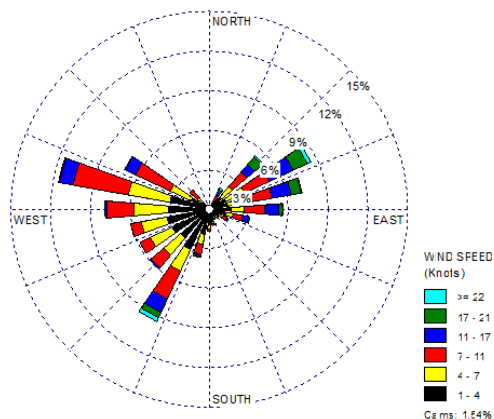
CAMS 123 FALL SEASON WINDROSE



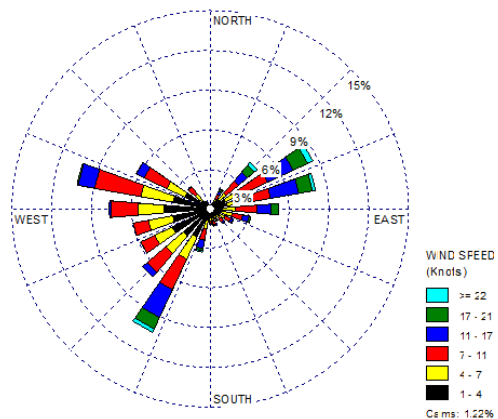
CAMS 413 ANNUAL WINDROSE



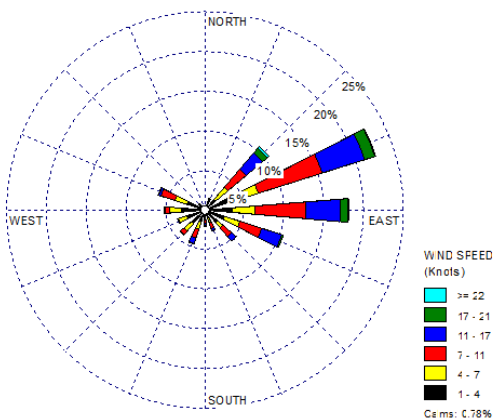
CAMS 413 WINTER SEASON WINDROSE



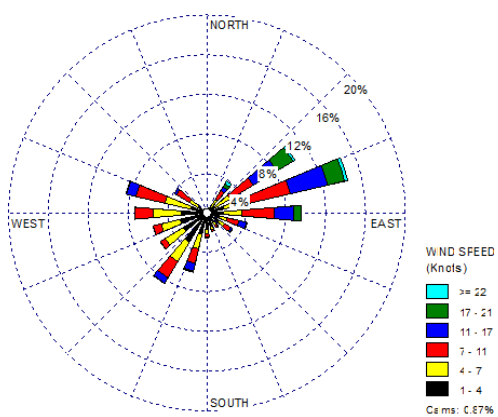
CAMS 413 SPRING SEASON WINDROSE



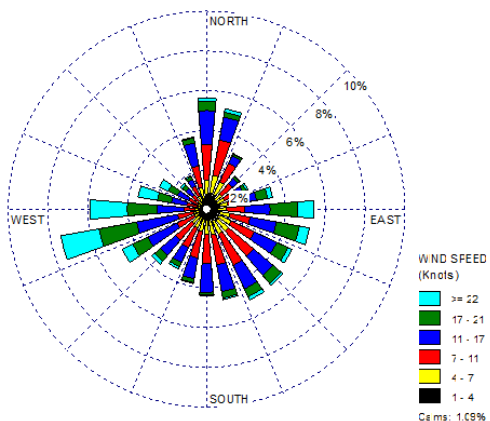
CAMS 413 SUMMER SEASON WINDROSE



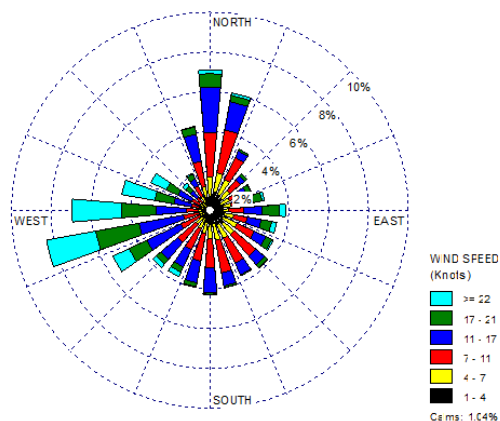
CAMS 413 FALL SEASON WINDROSE



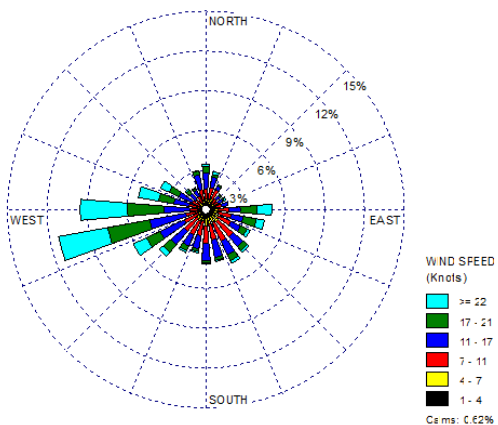
CAMS 414 ANNUAL WINDROSE



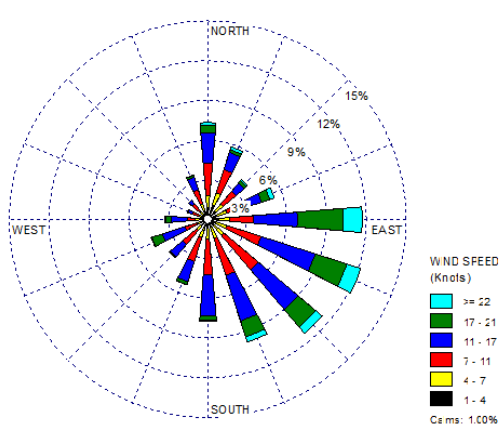
CAMS 414 WINTER SEASON WINDROSE



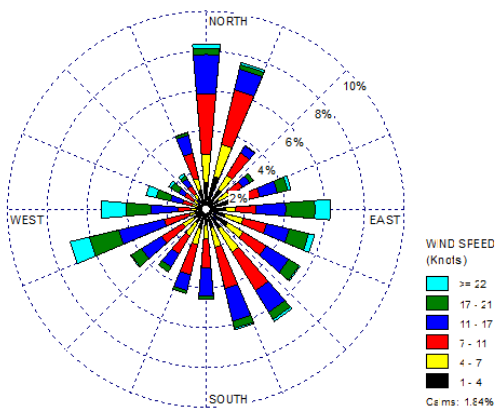
CAMS 414 SPRING SEASON WINDROSE



CAMS 414 SUMMER SEASON WINDROSE

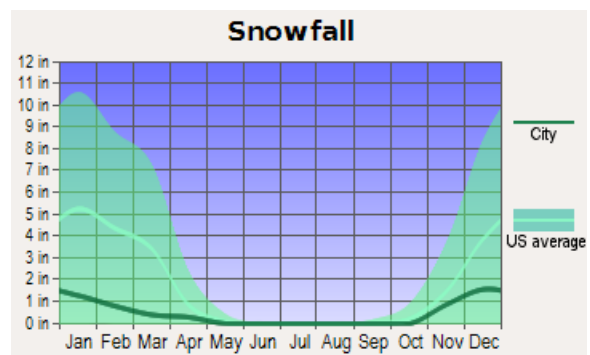
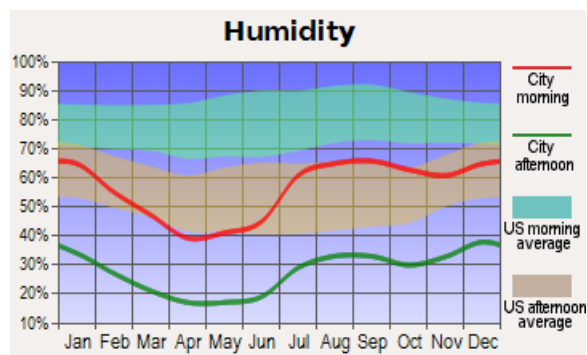
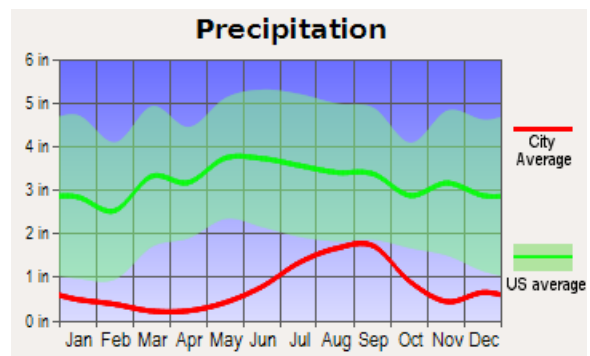
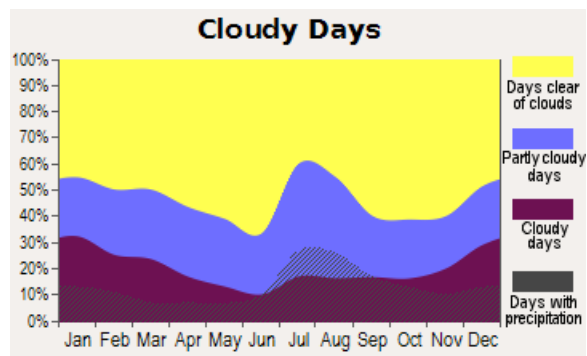
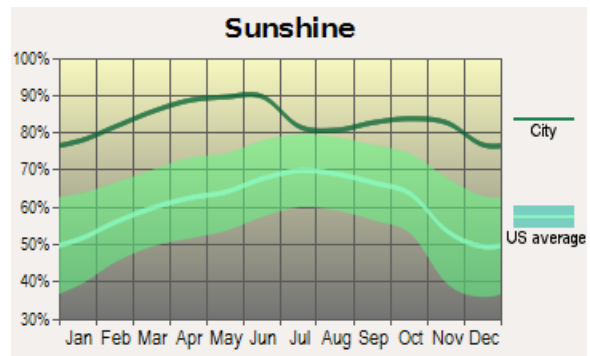
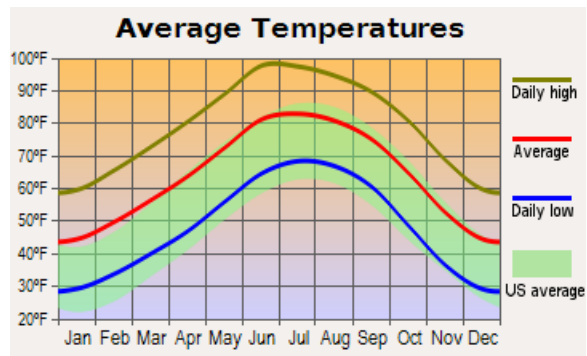


CAMS 414 FALL SEASON WINDROSE



Appendix B

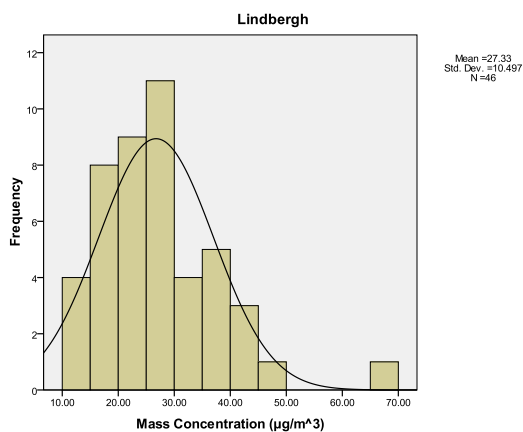
Meteorological characteristics of the three year monitoring campaign for the City of El Paso, Texas.



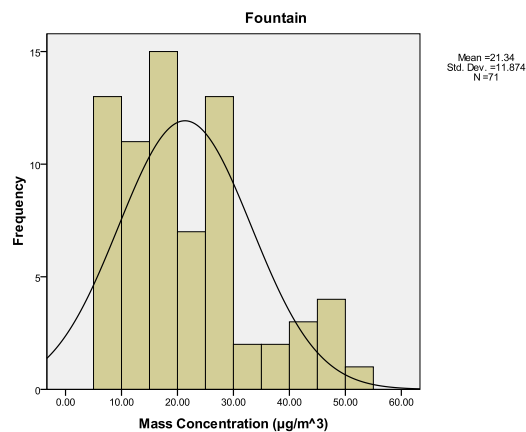
Appendix C

**Annual and Seasonal frequency distribution of the average sampled PM₁₀ concentrations
for every monitoring site during the three year campaign.**

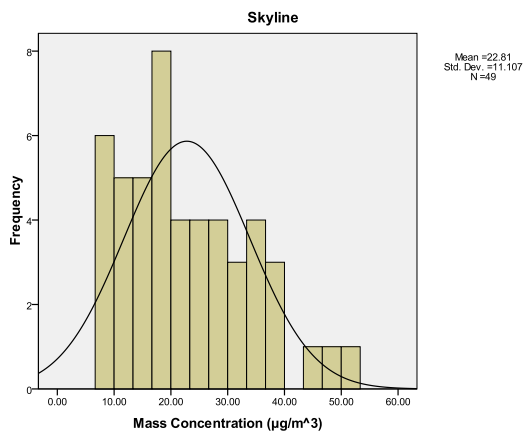
Site A Annual Frequency Distribution



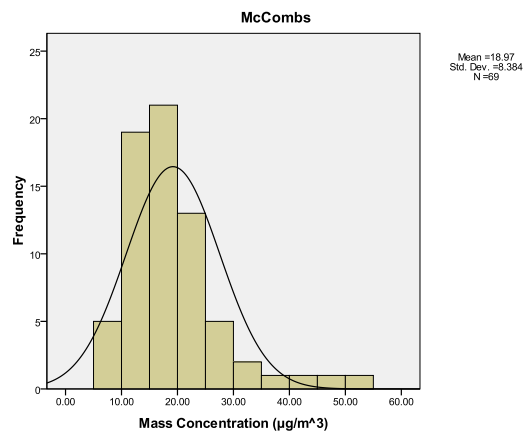
Site A2 Annual Frequency Distribution



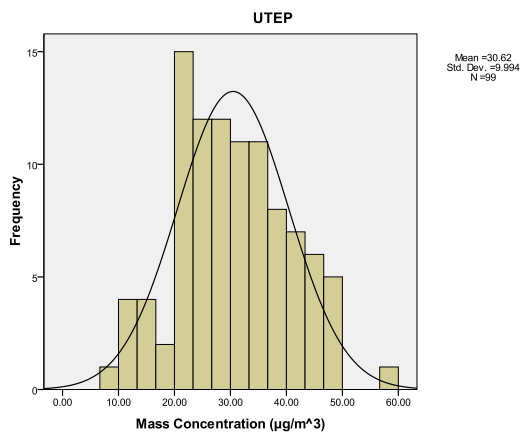
Site D Annual Frequency Distribution



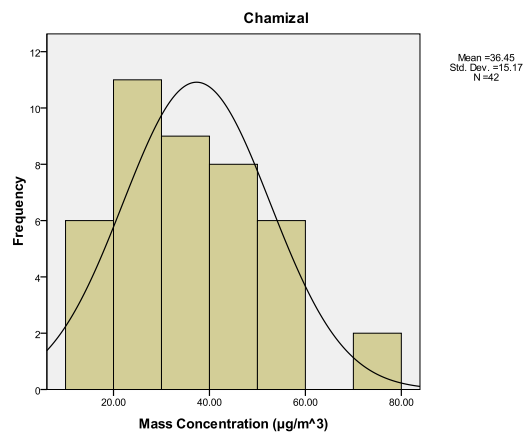
Site D2 Annual Frequency Distribution



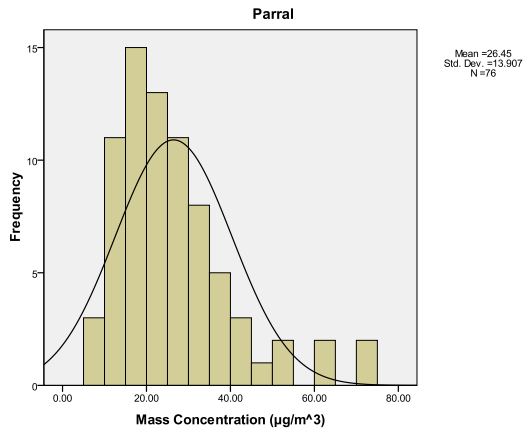
Site F Annual Frequency Distribution



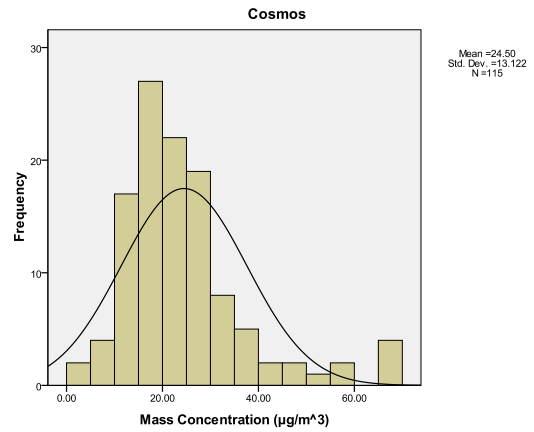
Site H Annual Frequency Distribution



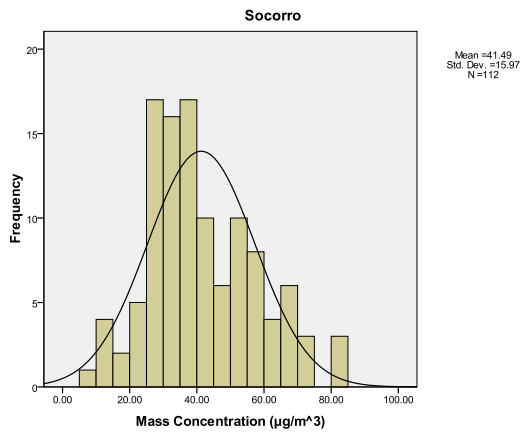
Site H2 Annual Frequency Distribution



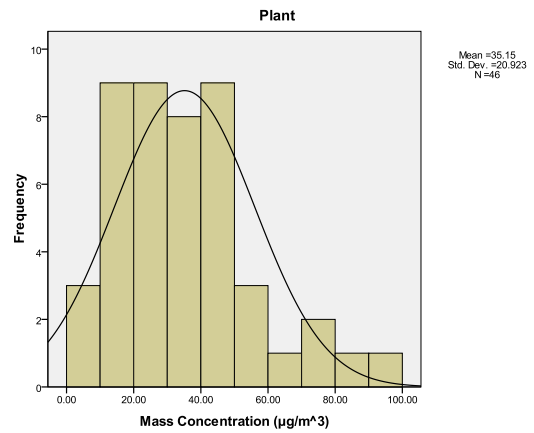
Site I Annual Frequency Distribution



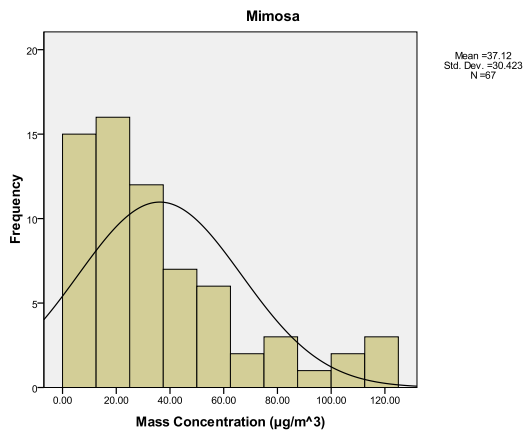
Site J Annual Frequency Distribution



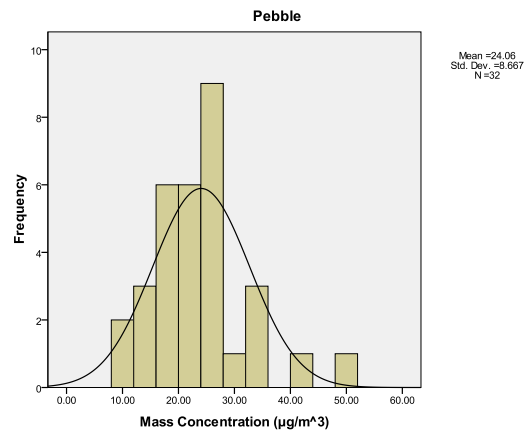
Site K Annual Frequency Distribution

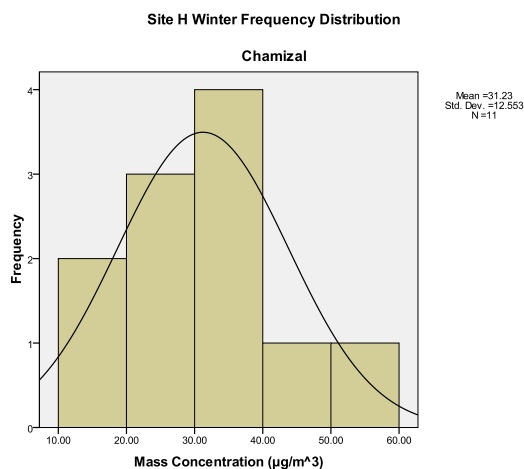
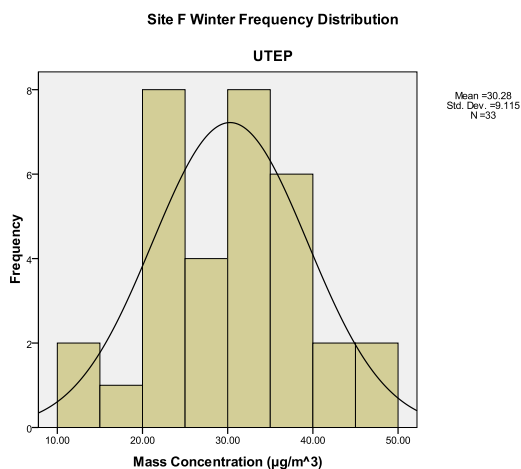
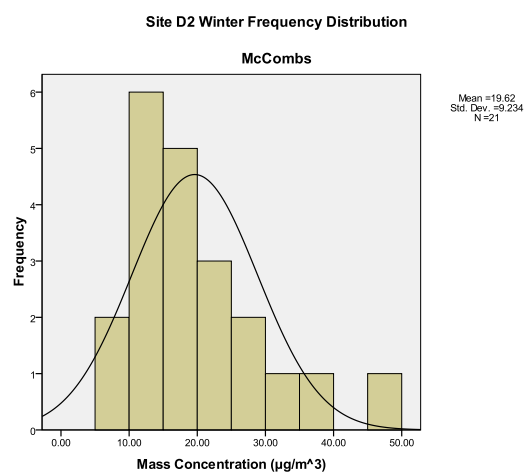
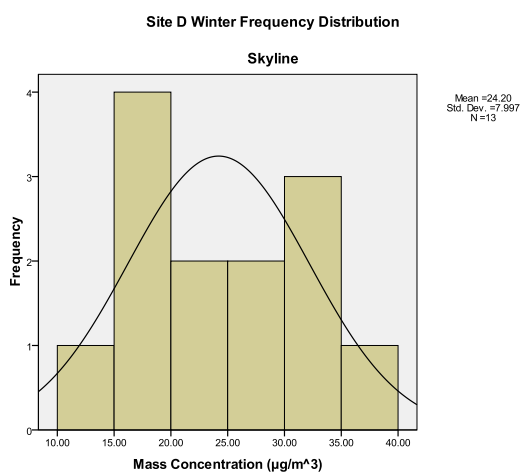
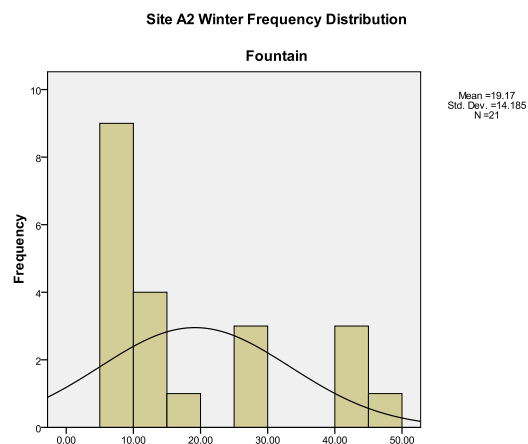
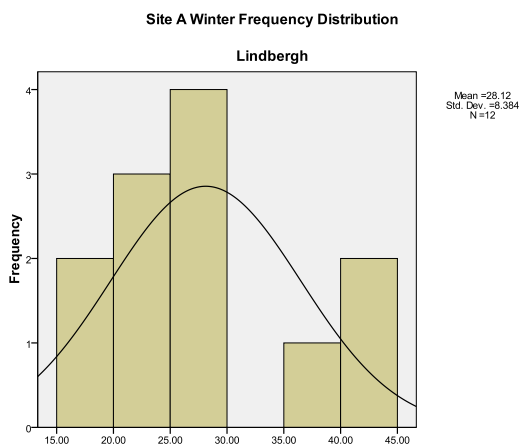


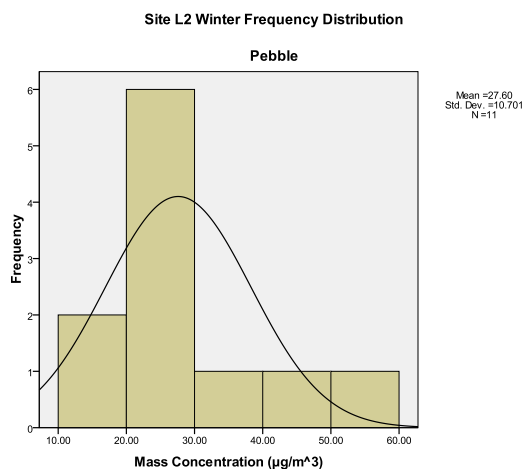
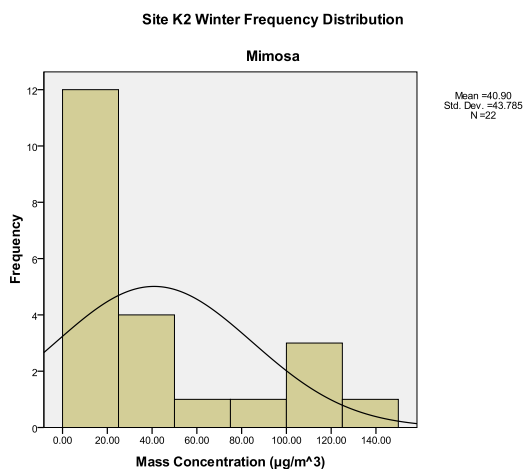
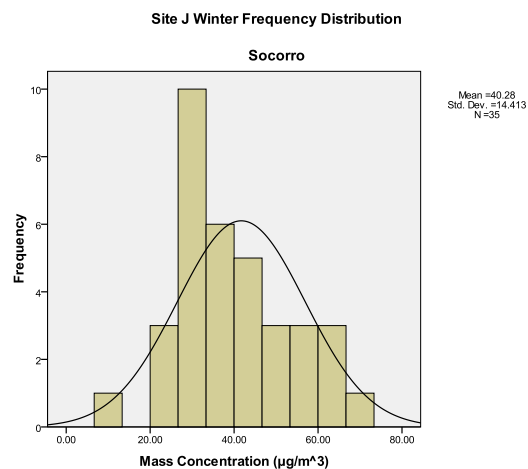
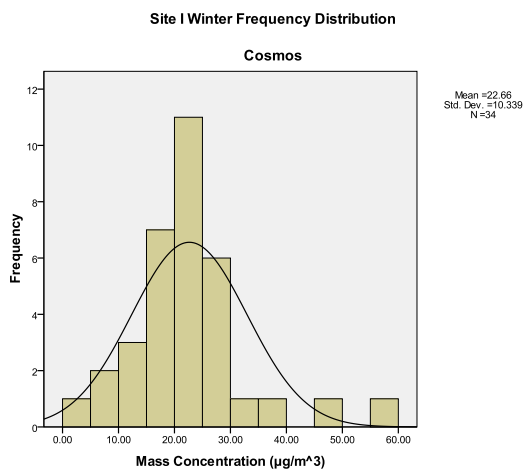
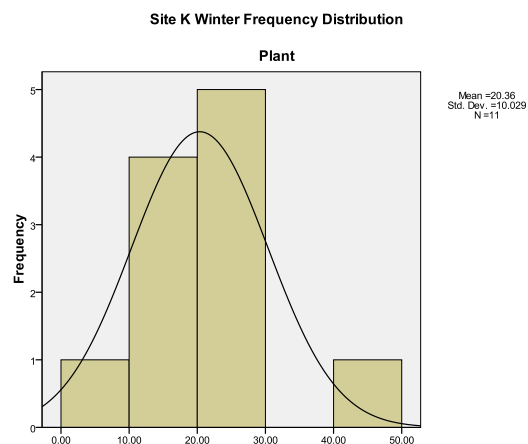
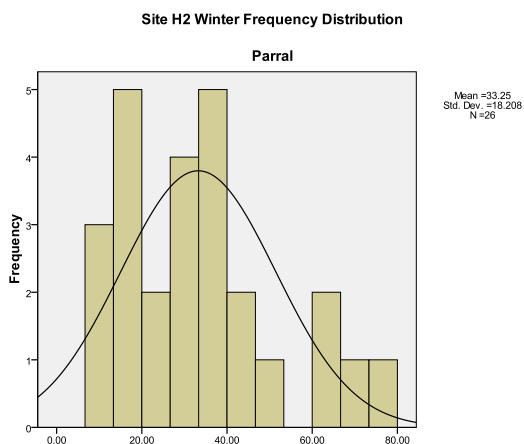
Site K2 Annual Frequency Distribution



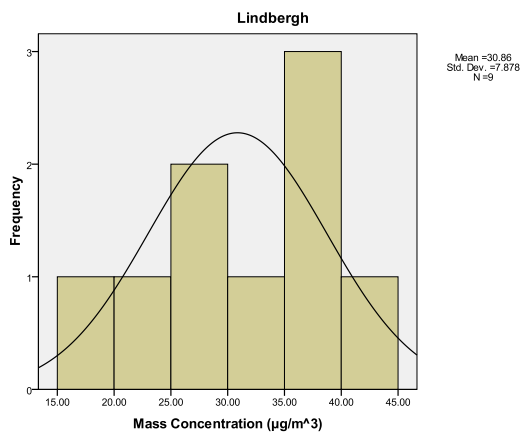
Site L2 Annual Frequency Distribution



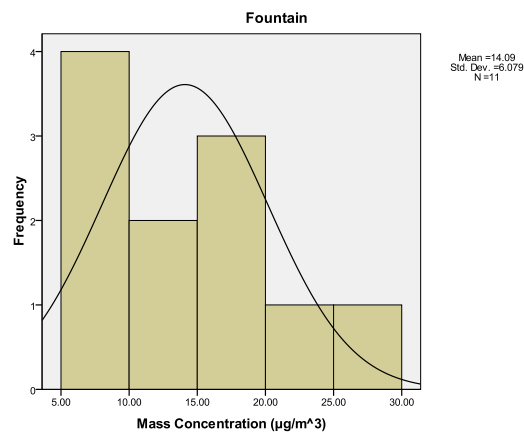




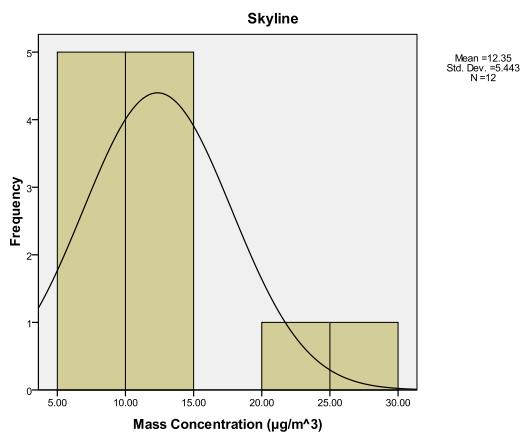
Site A Summer Frequency Distribution



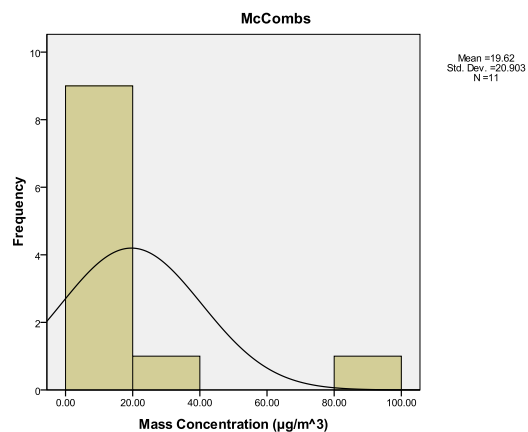
Site A2 Summer Frequency Distribution



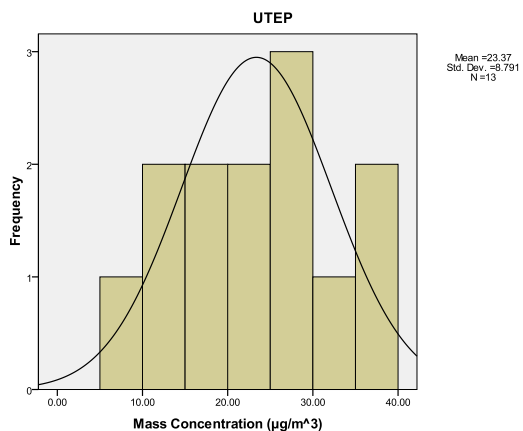
Site D Summer Frequency Distribution



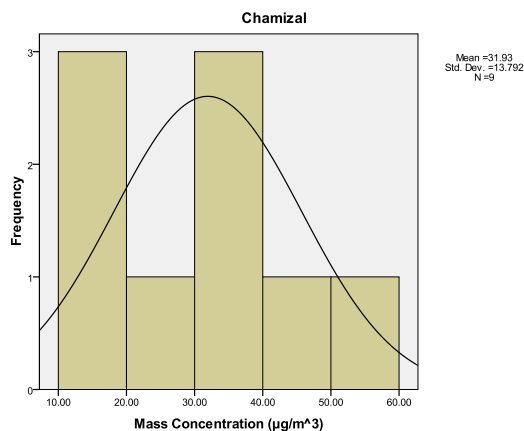
Site D2 Summer Frequency Distribution



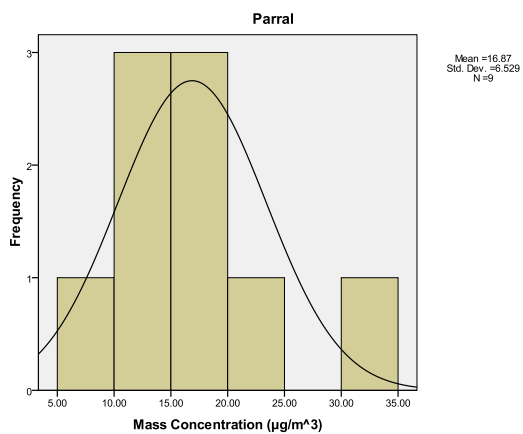
Site F Summer Frequency Distribution



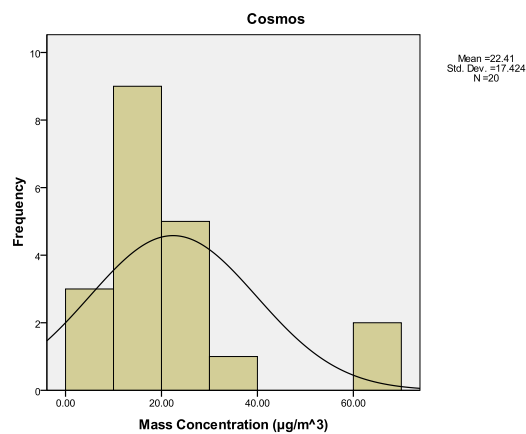
Site H Summer Frequency Distribution



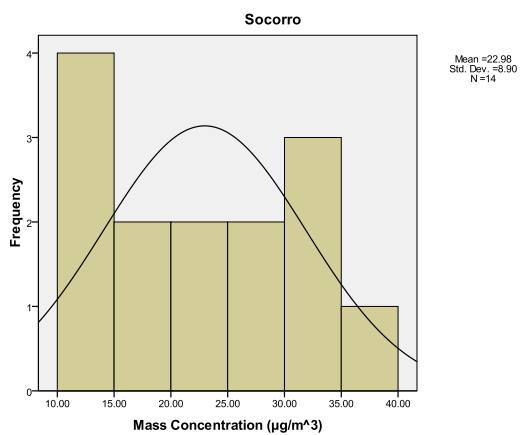
Site H2 Summer Frequency Distribution



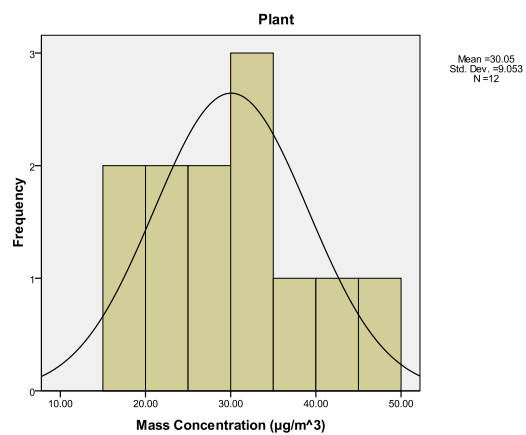
Site I Summer Frequency Distribution



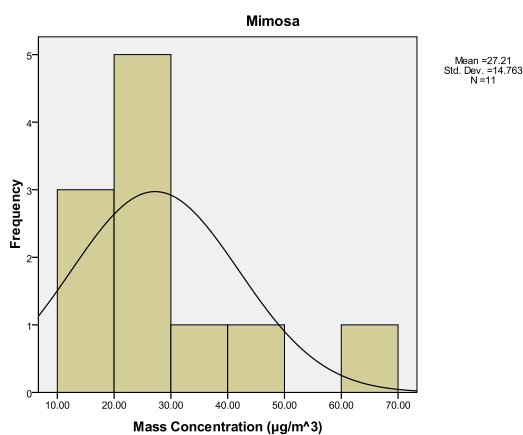
Site J Summer Frequency Distribution



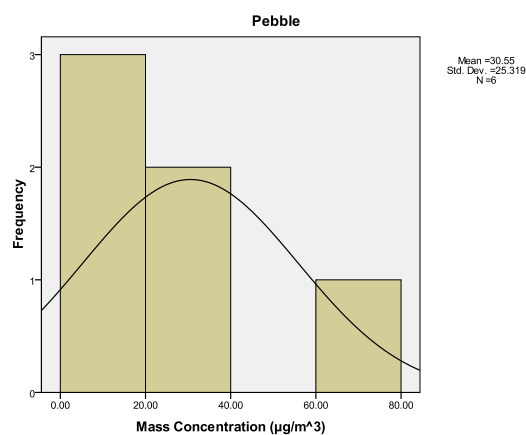
Site K Summer Frequency Distribution



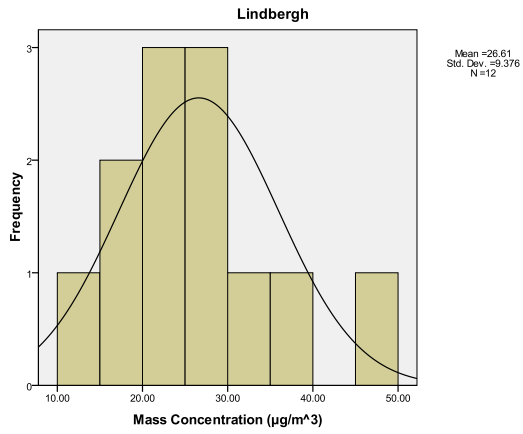
Site K2 Summer Frequency Distribution



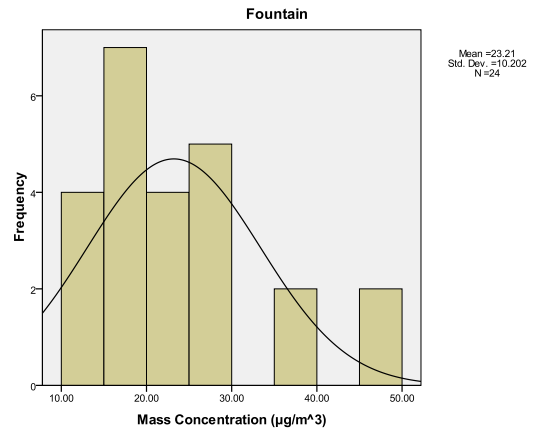
Site L2 Summer Frequency Distribution



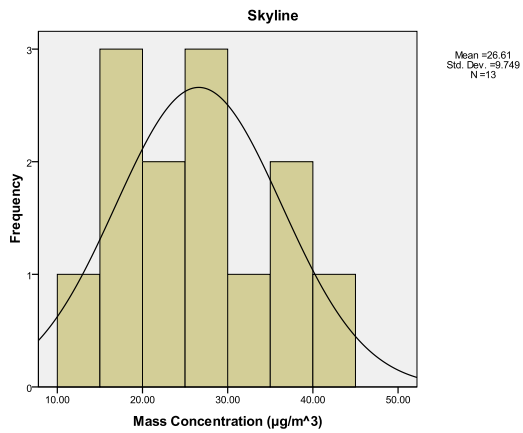
Site A Fall Frequency Distribution



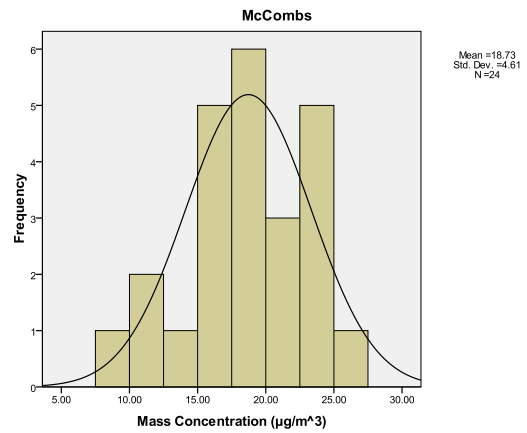
Site A2 Fall Frequency Distribution



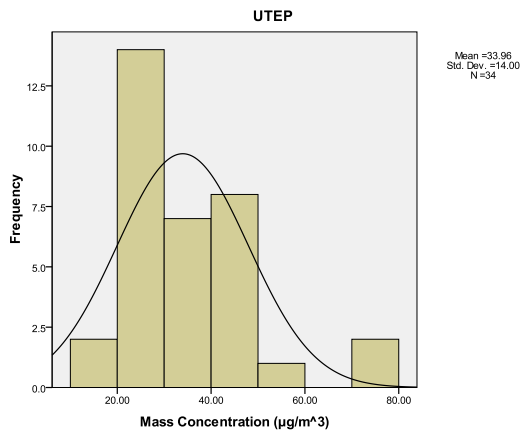
Site D Fall Frequency Distribution



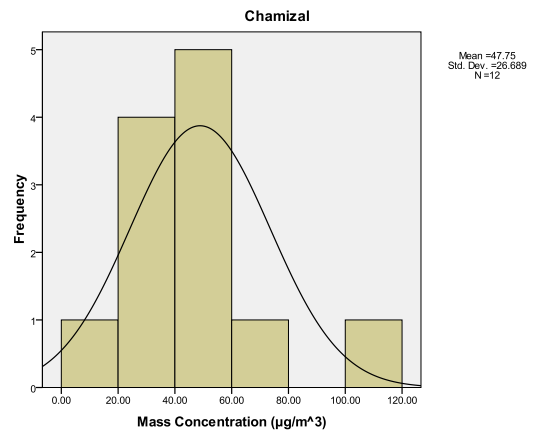
Site D2 Fall Frequency Distribution

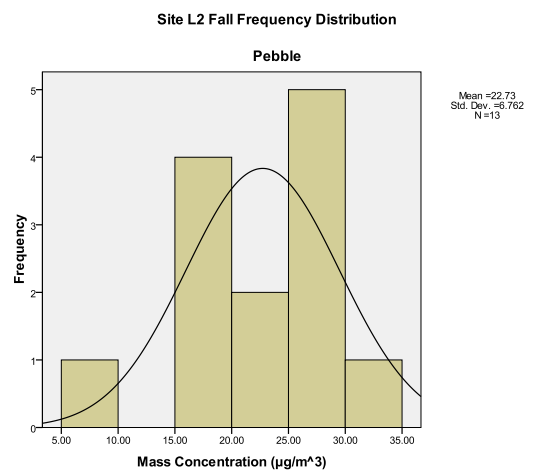
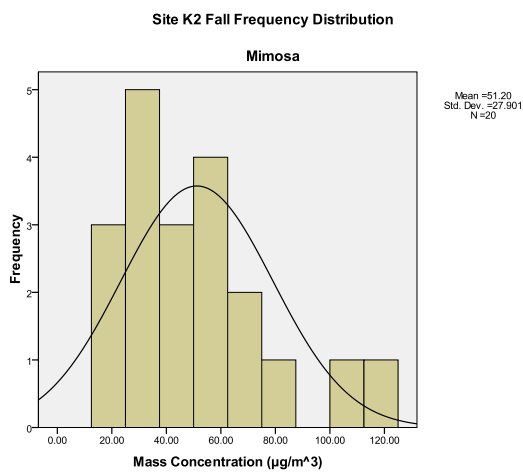
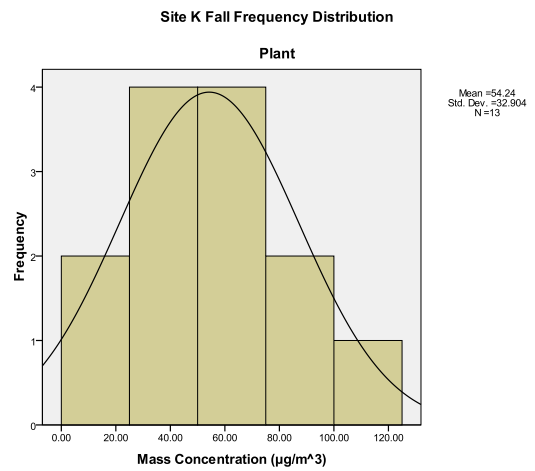
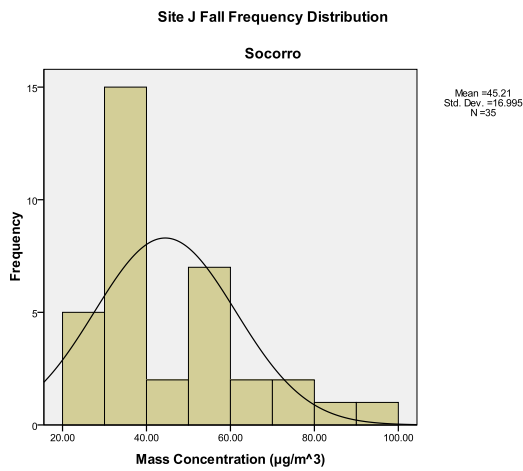
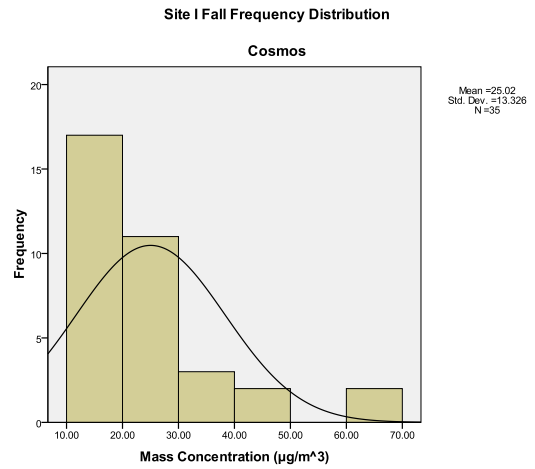
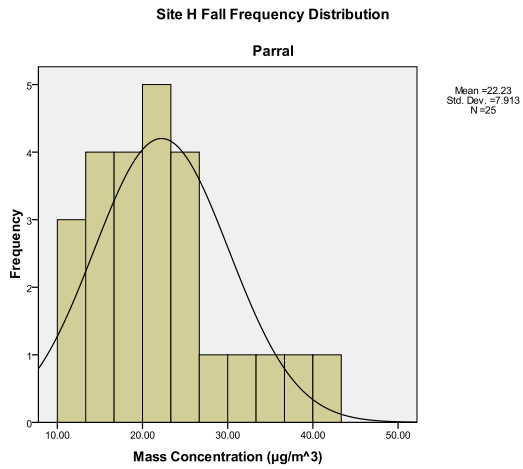


Site F Fall Frequency Distribution

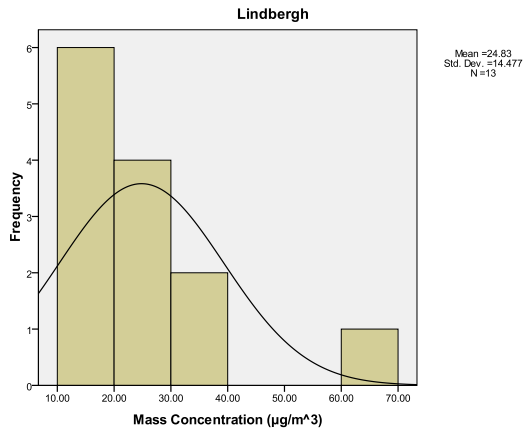


Site H Fall Frequency Distribution

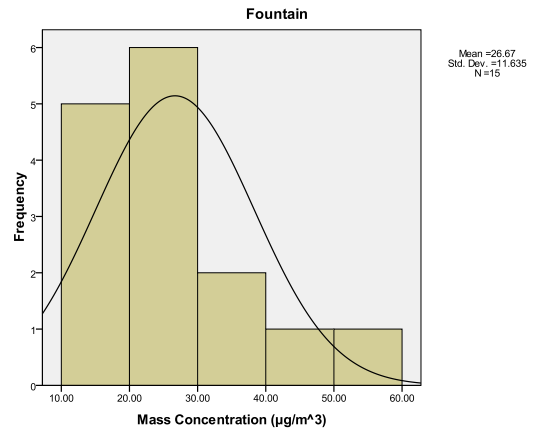




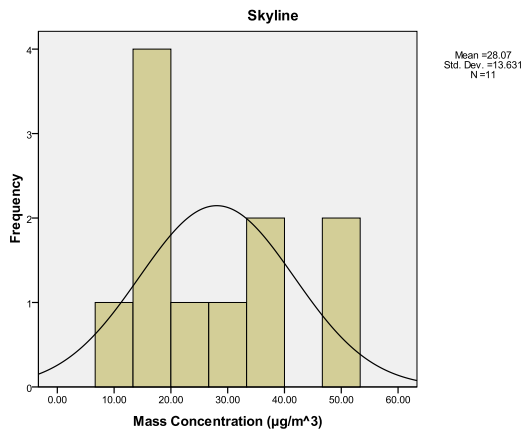
Site A Spring Frequency Distribution



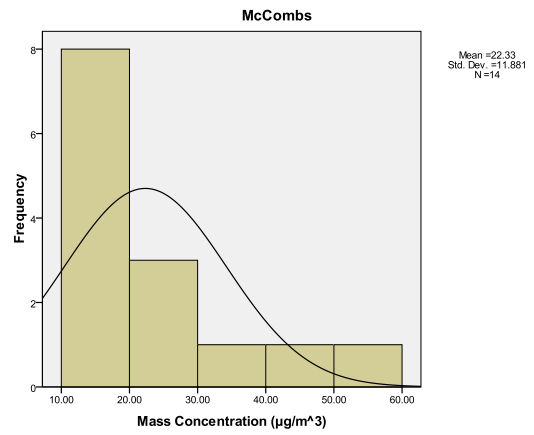
Site A2 Spring Frequency Distribution



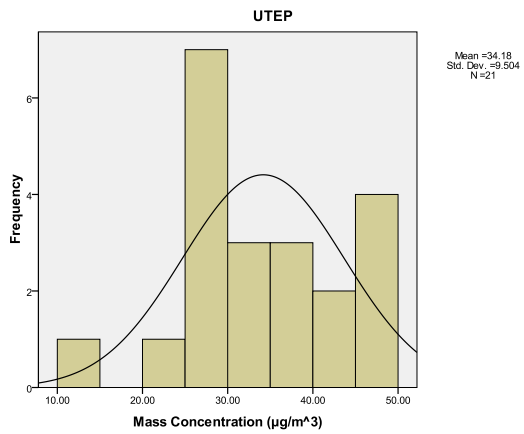
Site D Spring Frequency Distribution



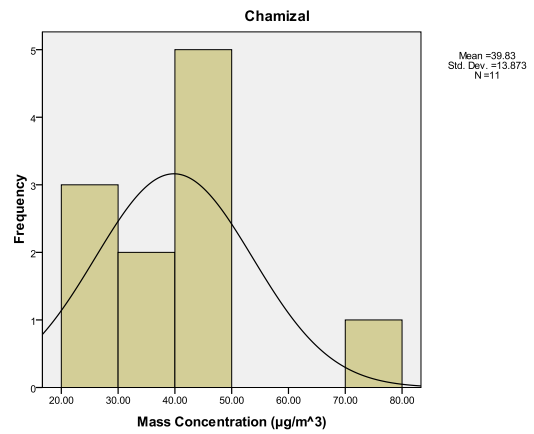
Site D2 Spring Frequency Distribution

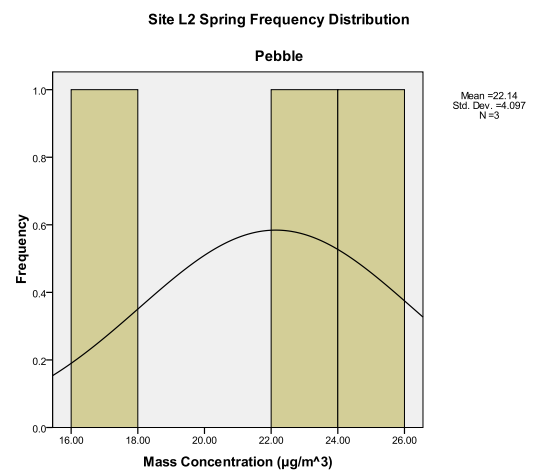
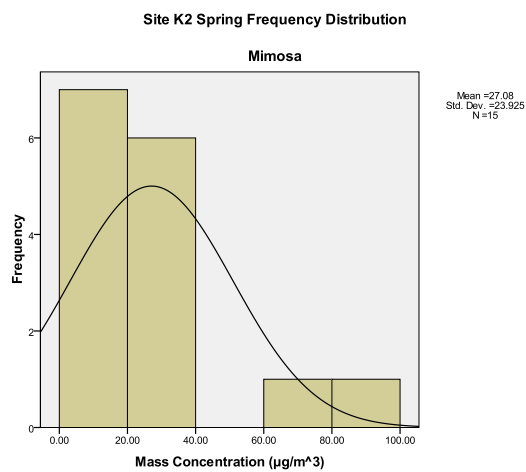
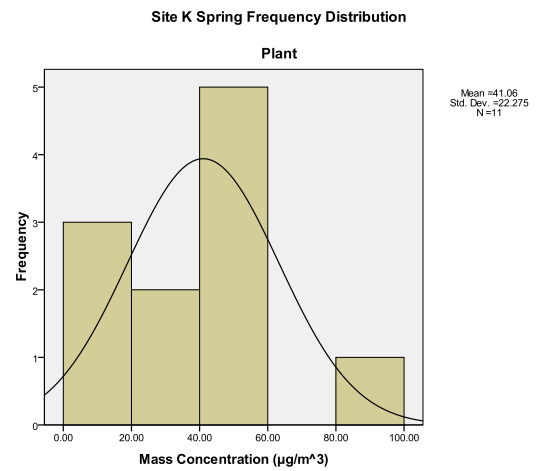
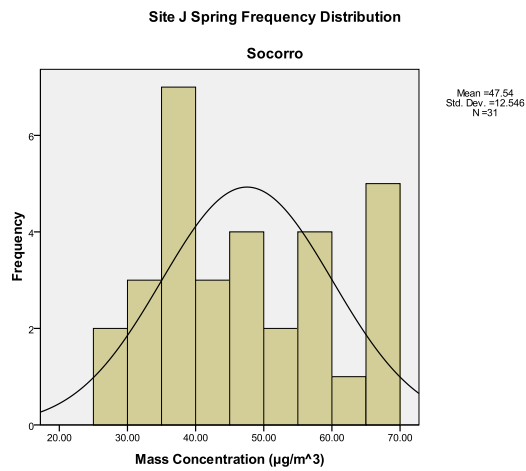
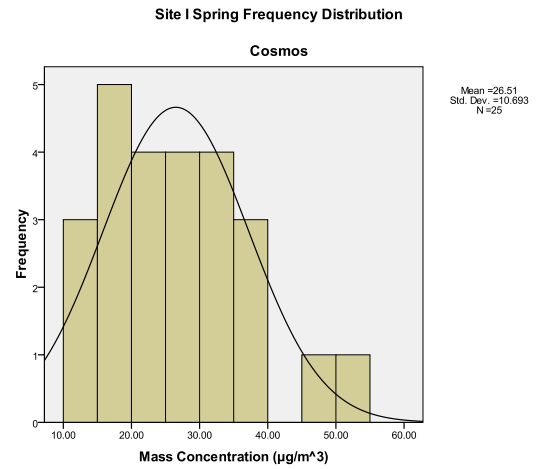
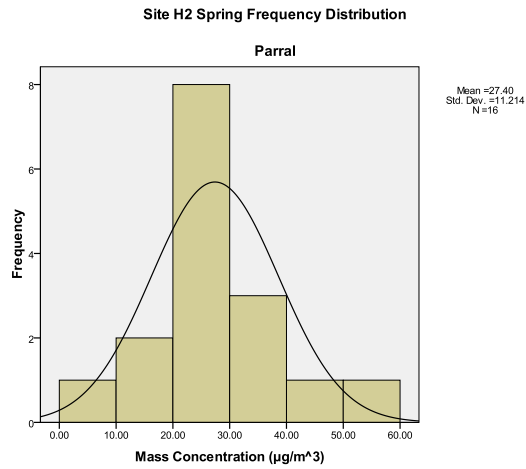


Site F Spring Frequency Distribution



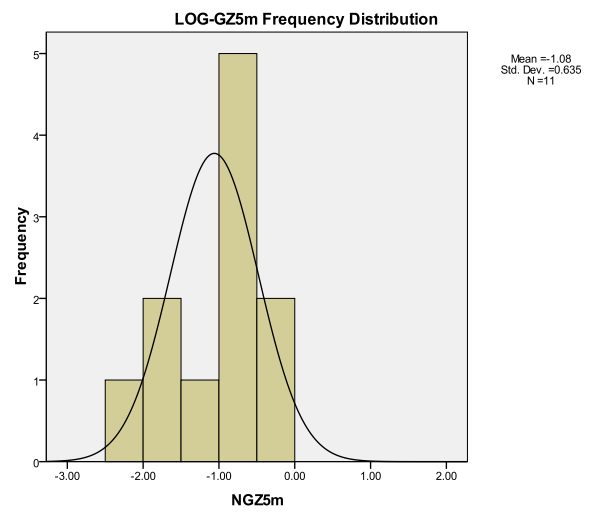
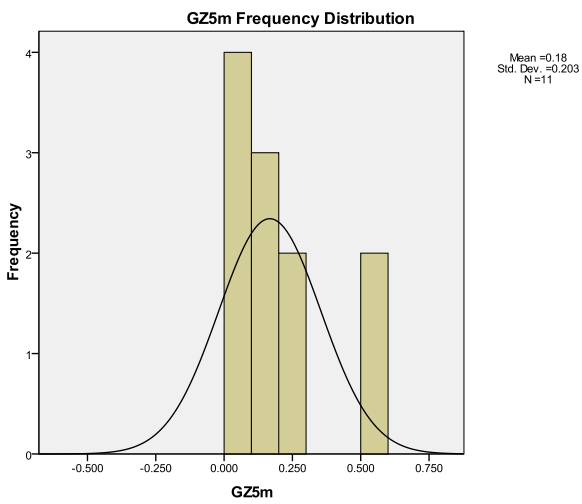
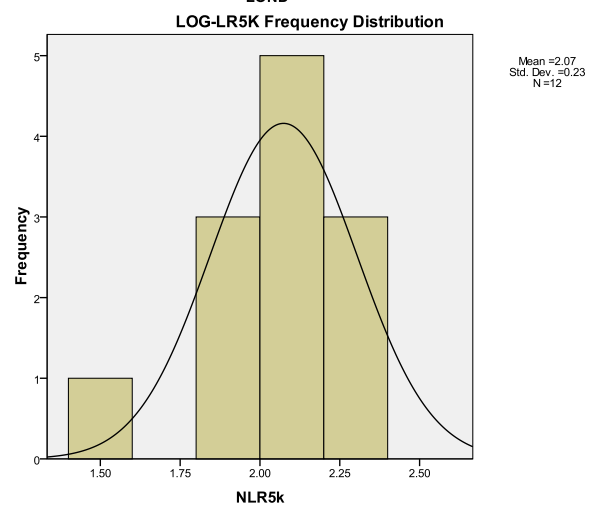
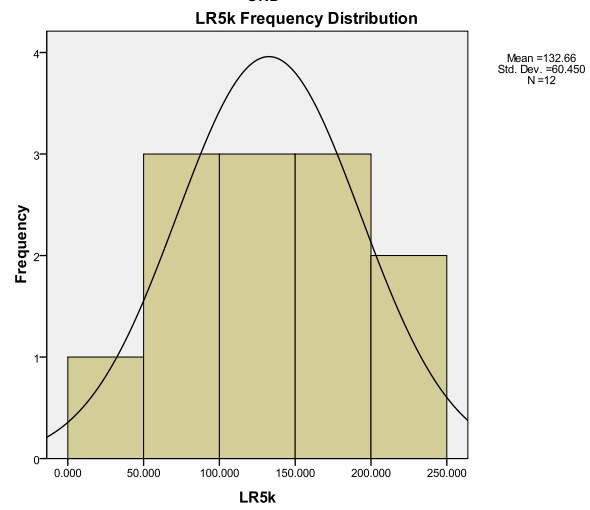
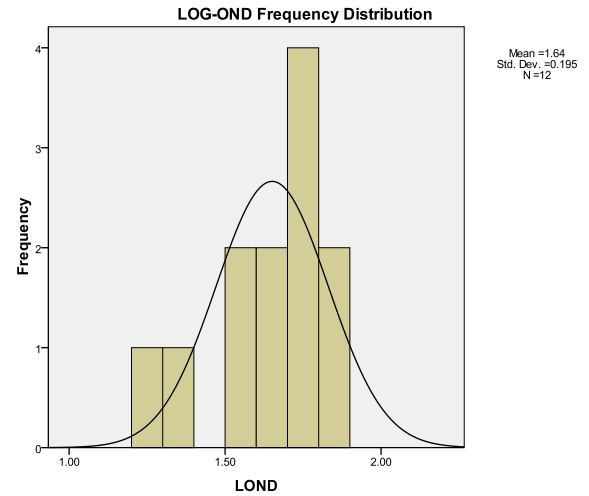
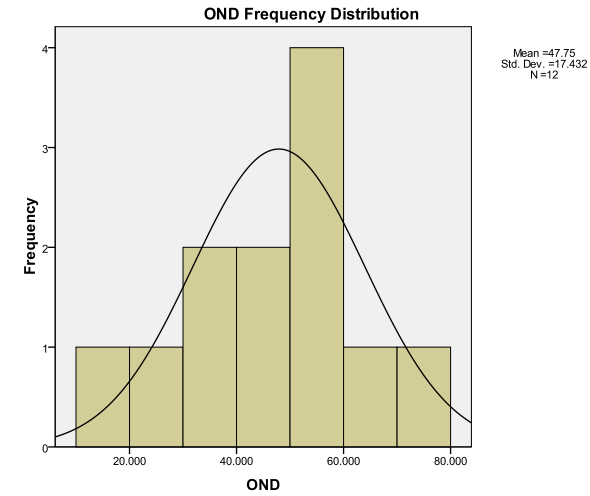
Site H Spring Frequency Distribution

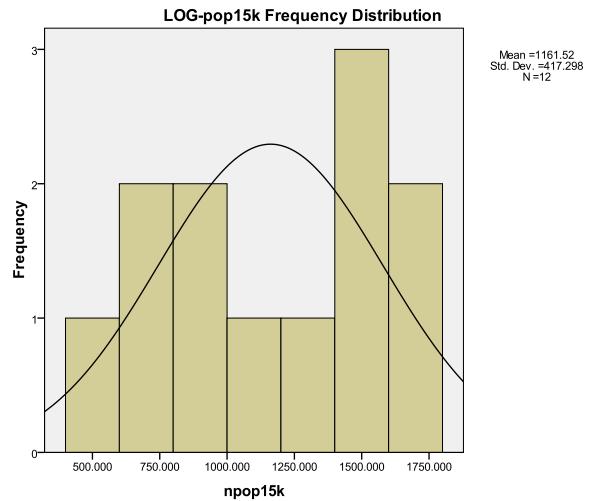
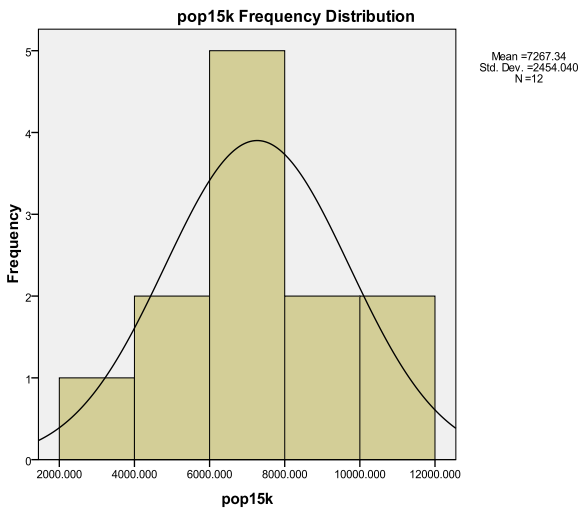
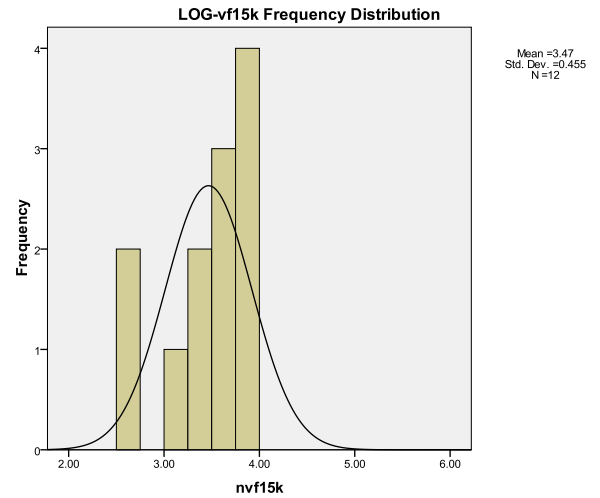
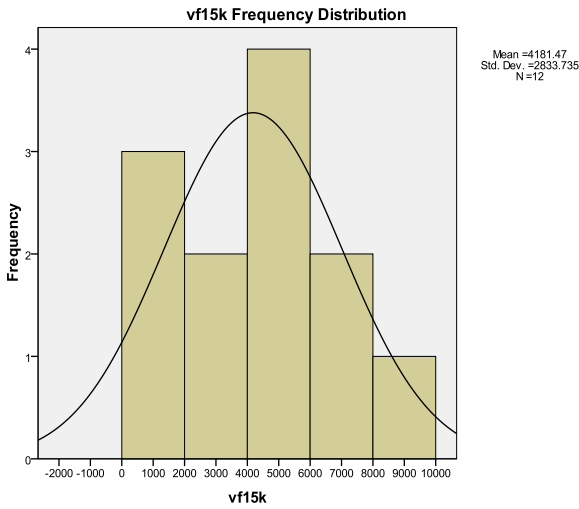
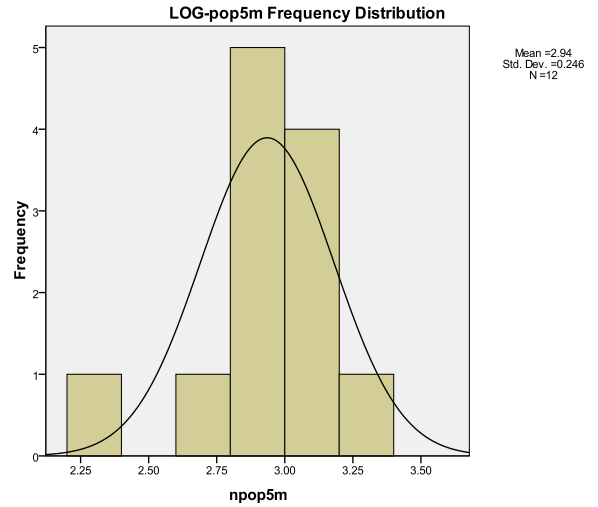
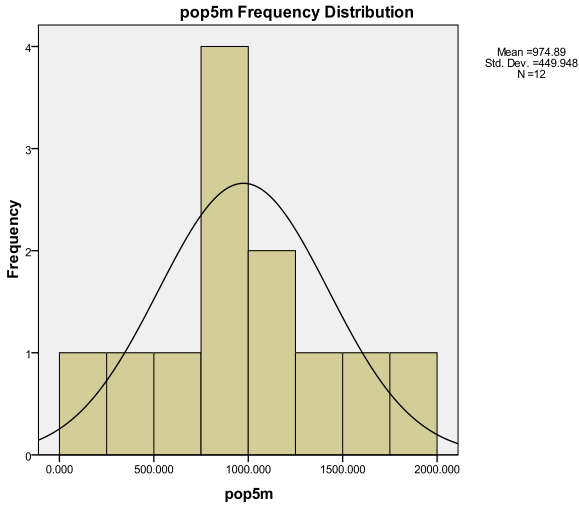


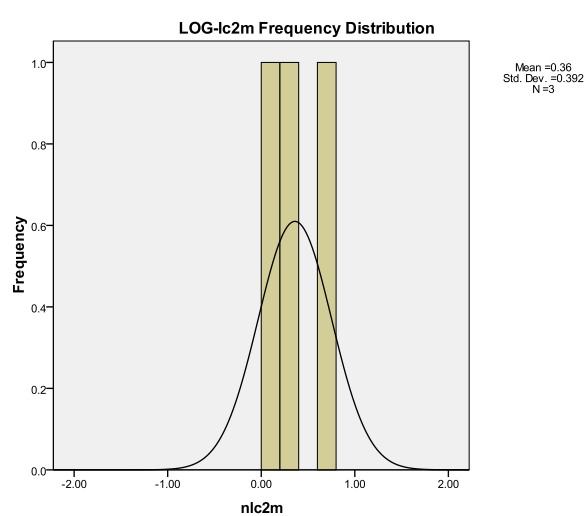
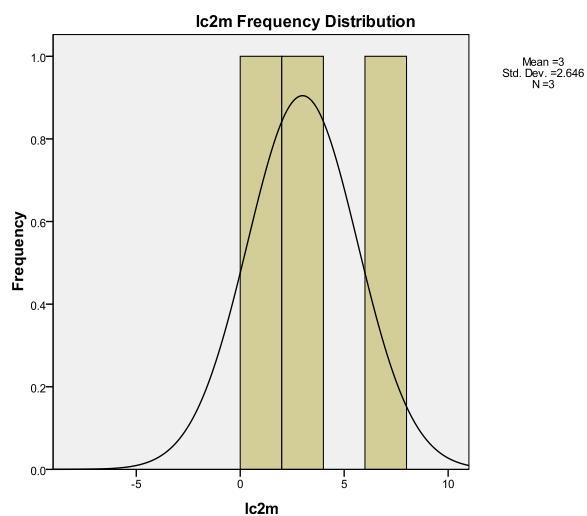
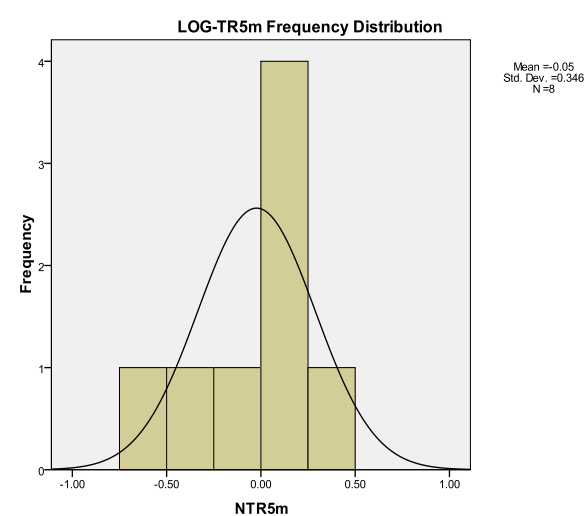
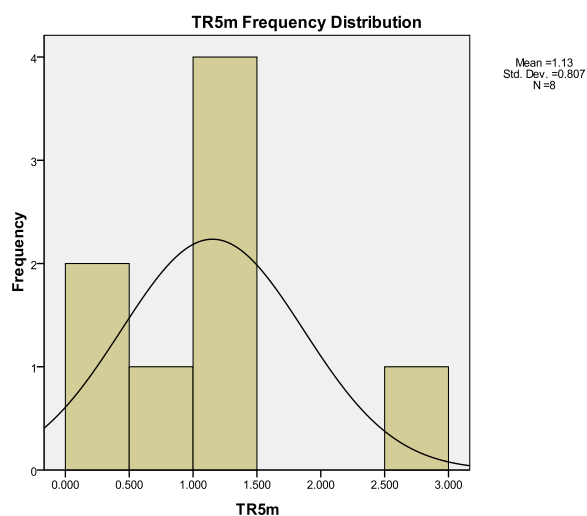
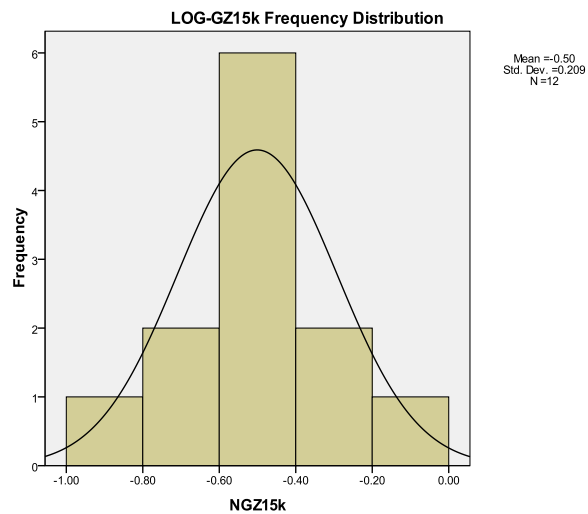
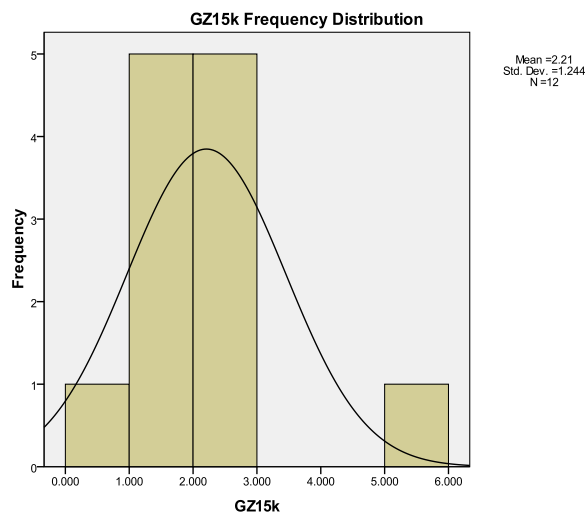


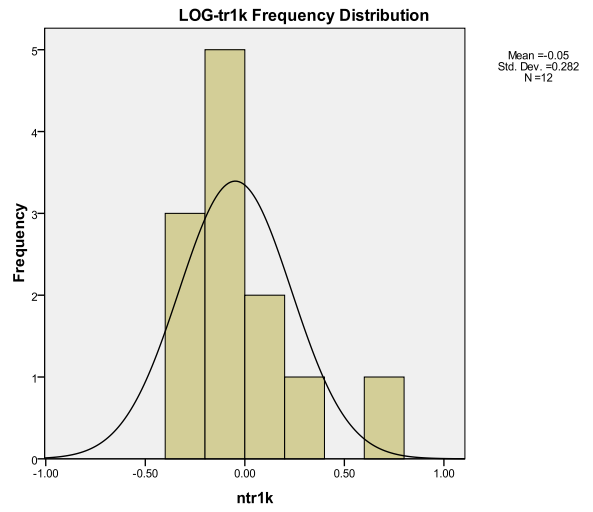
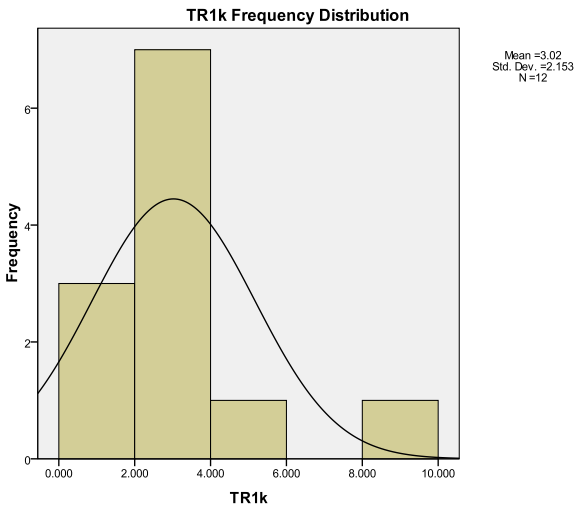
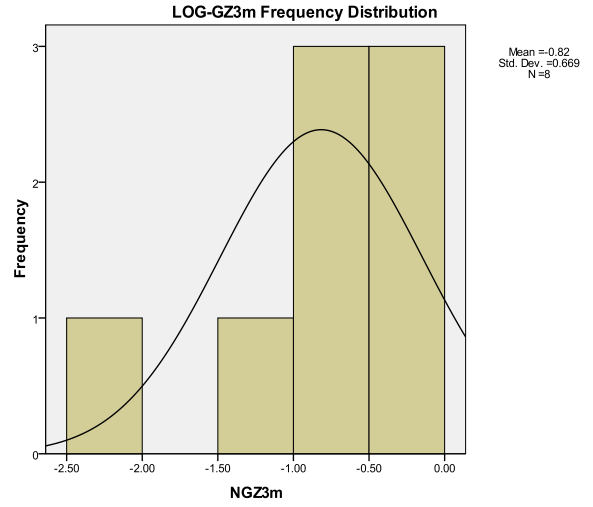
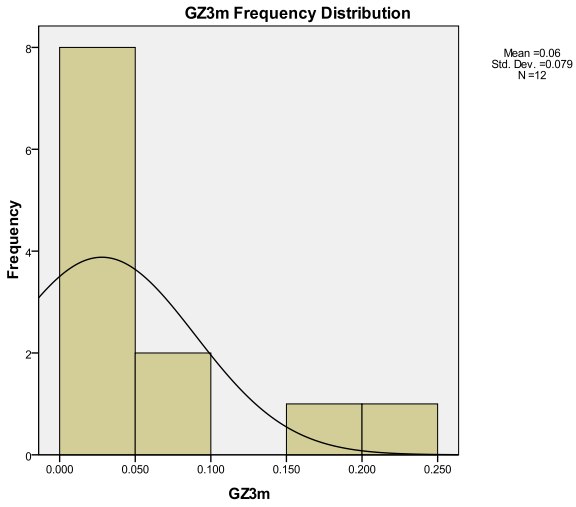
Appendix D

Frequency distribution for every parameter included in the analysis.









Appendix E

Chi-square and Durbin and Watson statistics tables.

Durbin-Watson Statistic: 5 Per Cent Significance Points of dL and dU

	k'=1		k'=2		k'=3		k'=4		k'=5		k'=6		k'=7		k'=8		k'=9		k'=10	
n	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
6	0.610	1.400	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7	0.700	1.356	0.467	1.896	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8	0.763	1.332	0.559	1.777	0.367	2.287	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	----	----	----	----	----	----	----	----	----	----	----	----
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	----	----	----	----	----	----	----	----	----	----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	----	----	----	----	----	----	----	----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	----	----	----	----	----	----
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	----	----	----	----
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	----	----
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.079	1.891	1.015	1.978	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.876	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.197
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877
200	1.758	1.779	1.748	1.789	1.738	1.799	1.728	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

*'k' is the number of regressors excluding the intercept

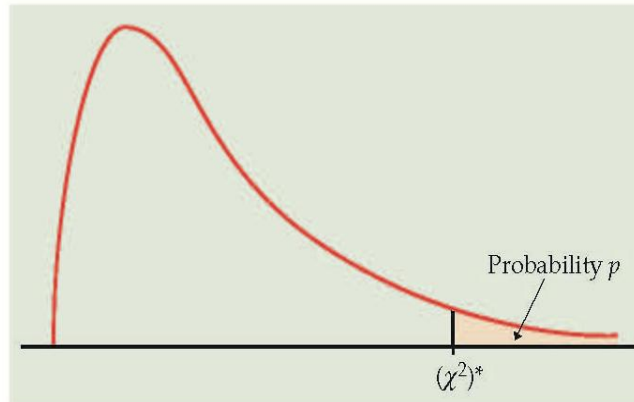


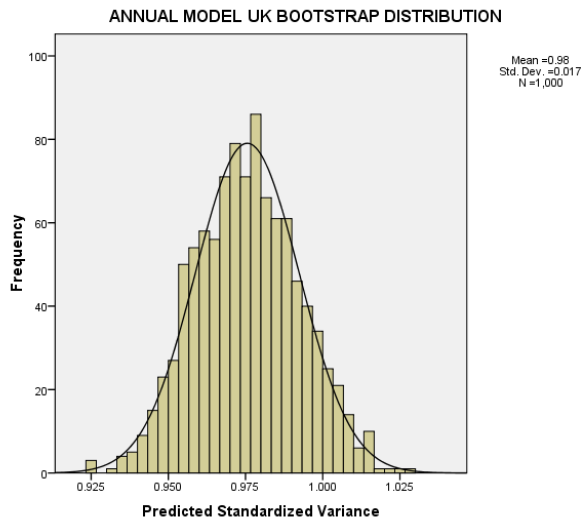
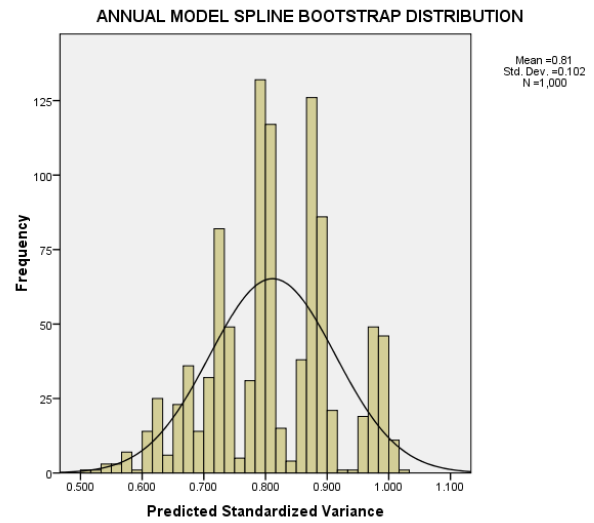
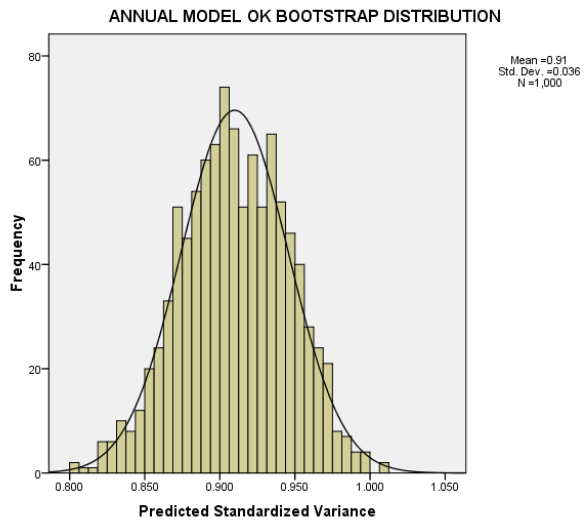
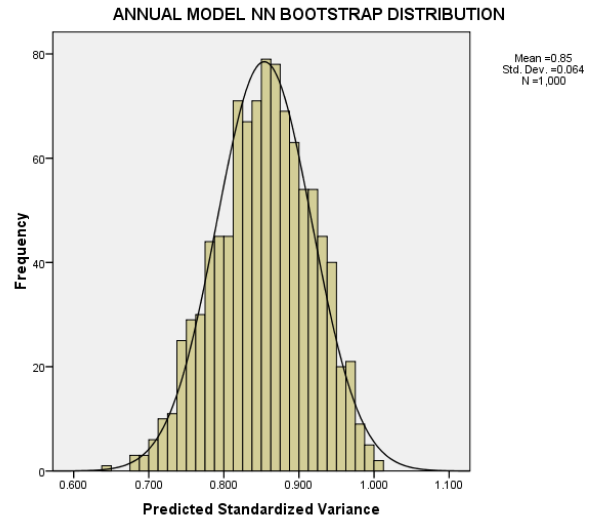
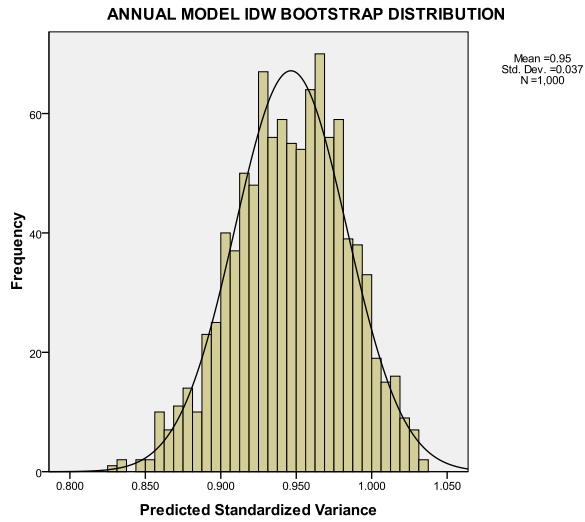
Table entry for p is the critical value $(\chi^2)^*$ with probability p lying to its right.

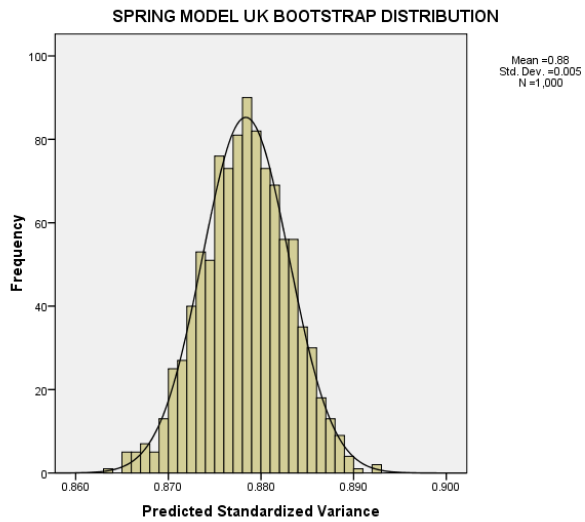
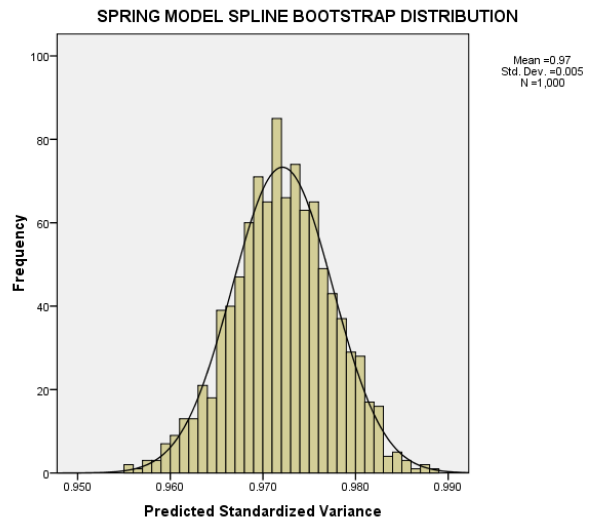
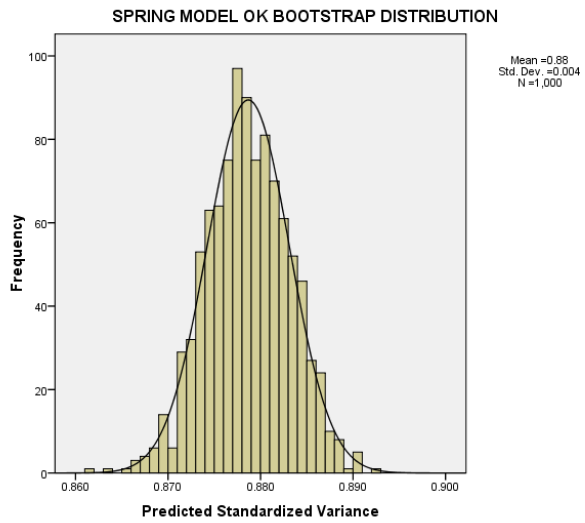
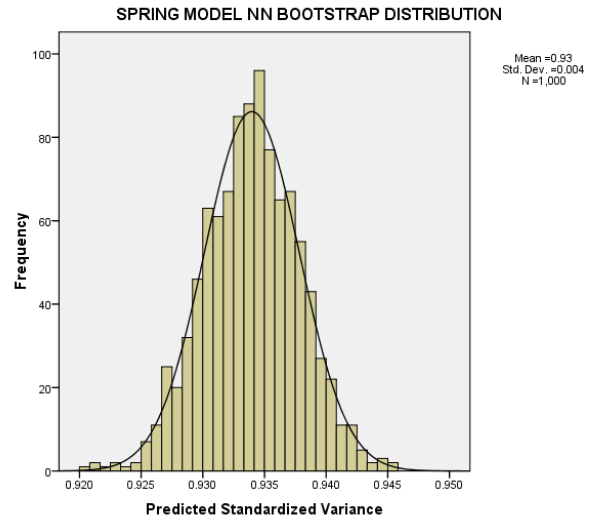
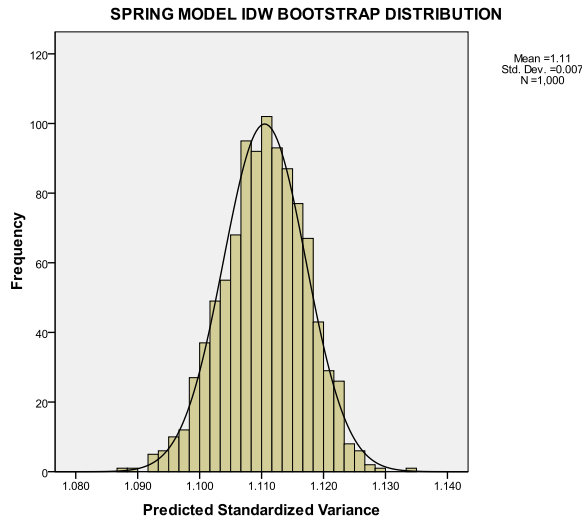
TABLE F χ^2 distribution critical values

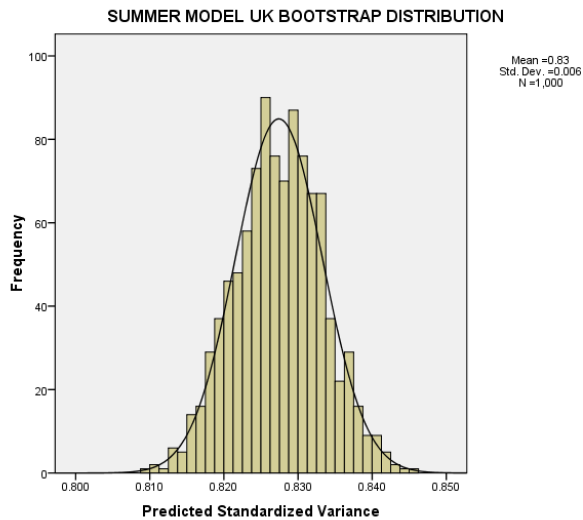
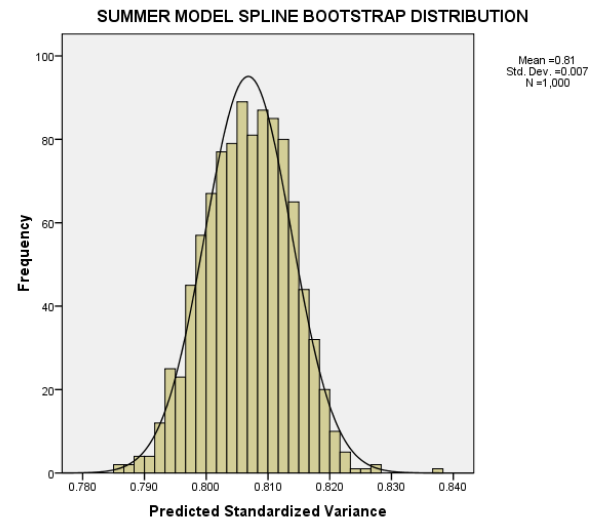
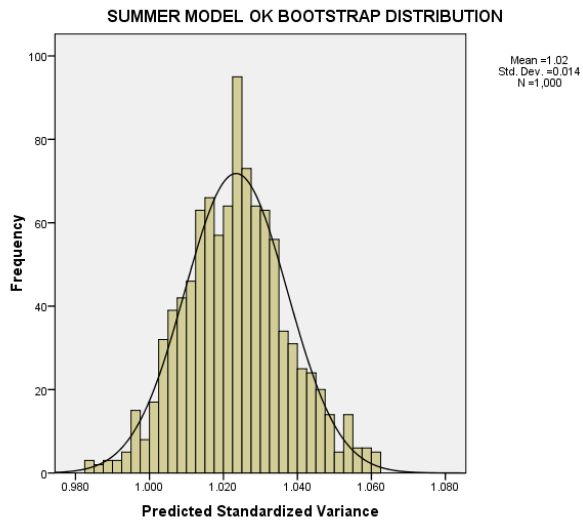
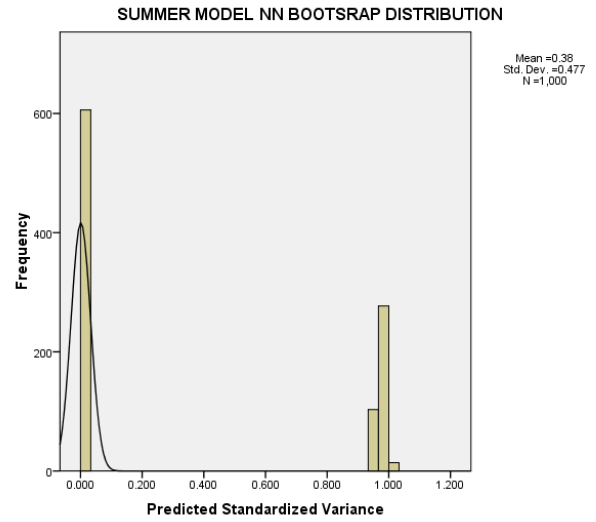
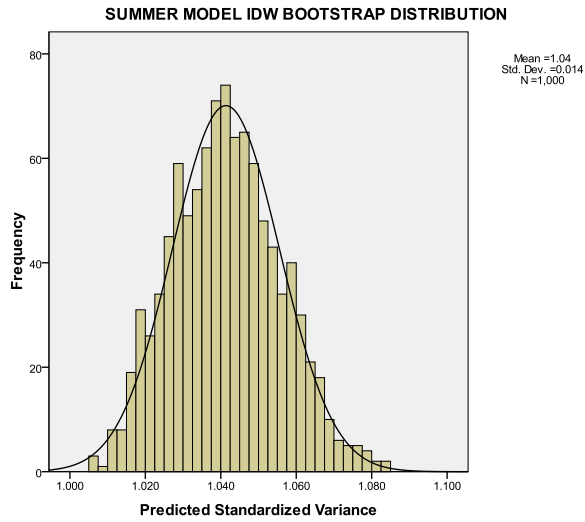
df	Tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

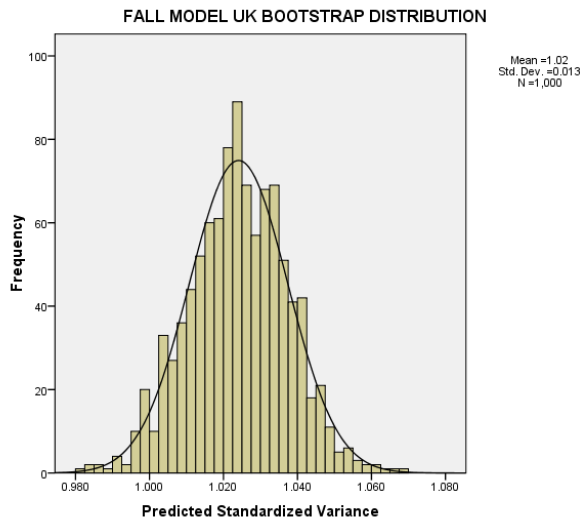
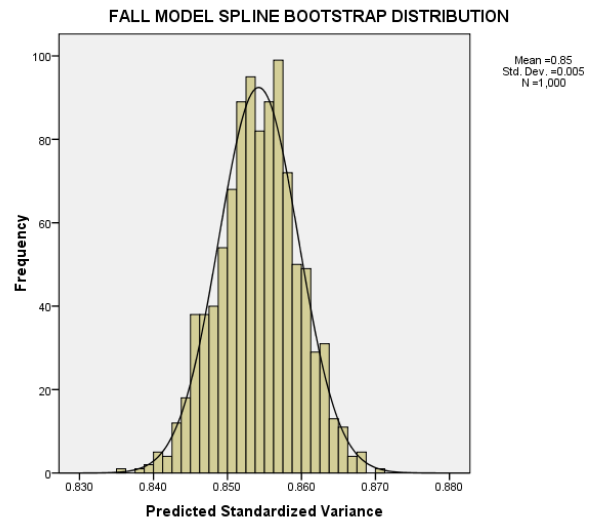
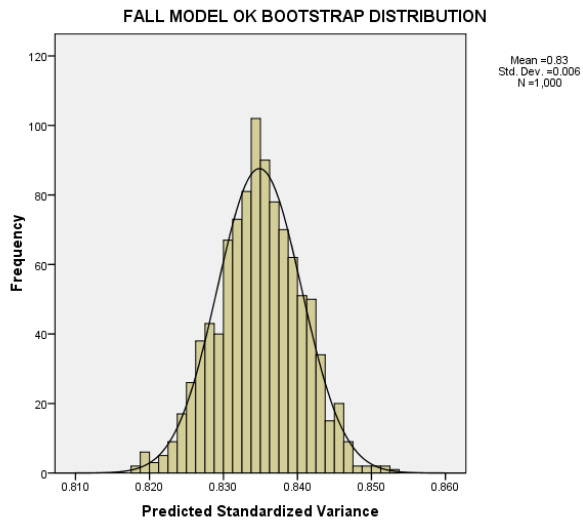
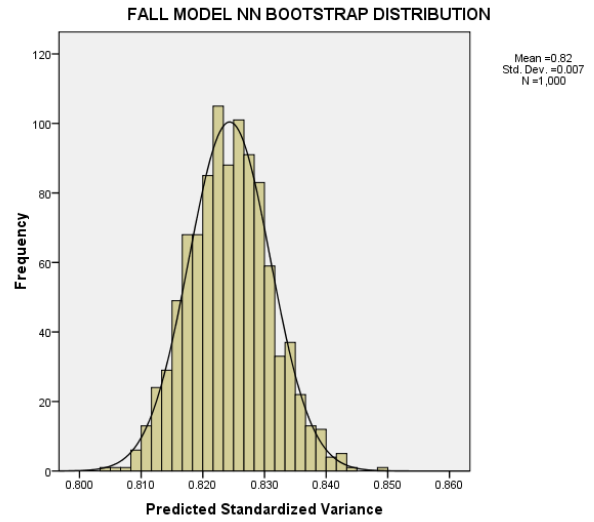
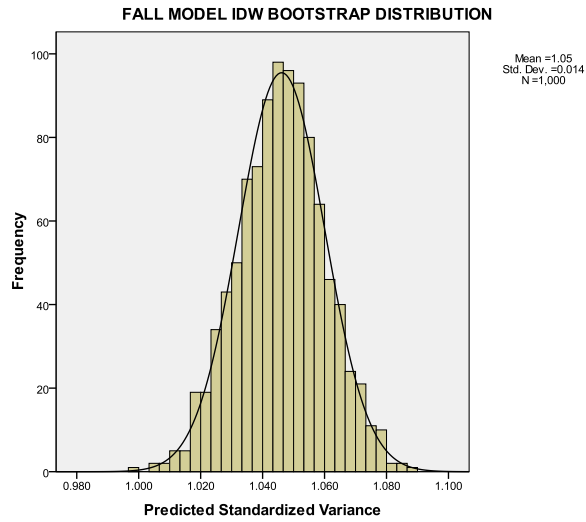
Appendix F

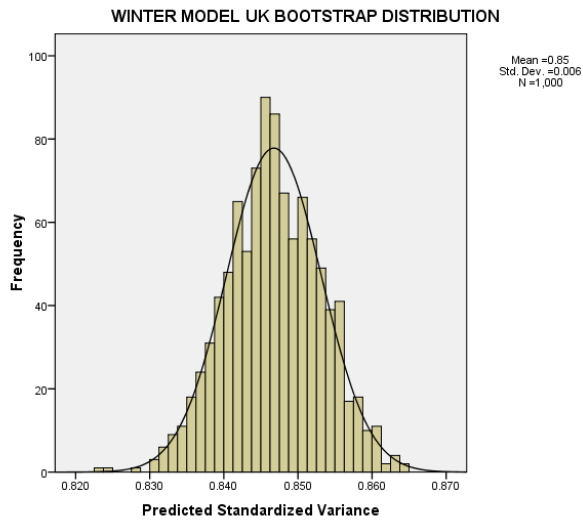
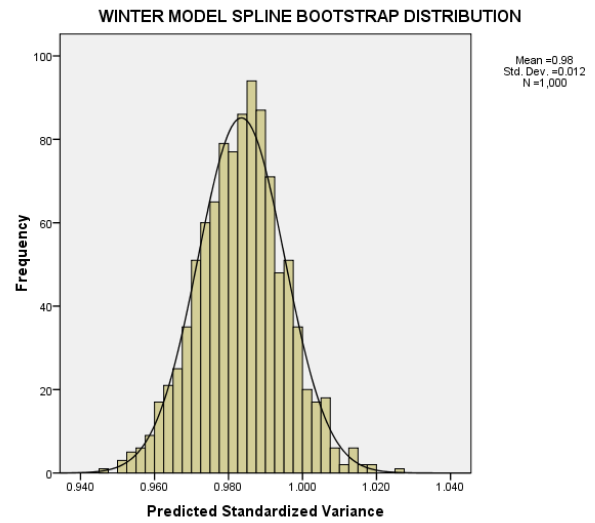
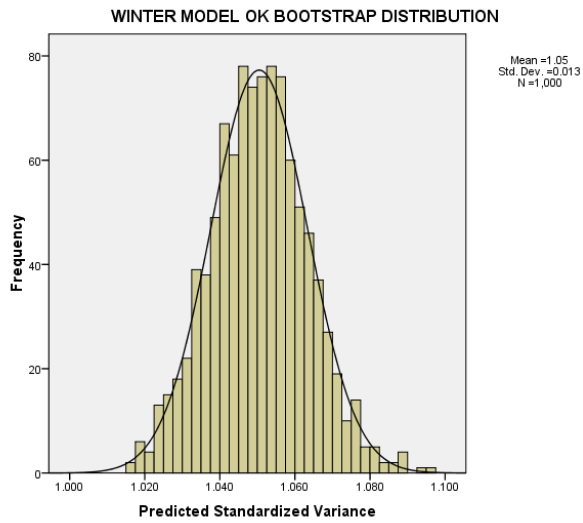
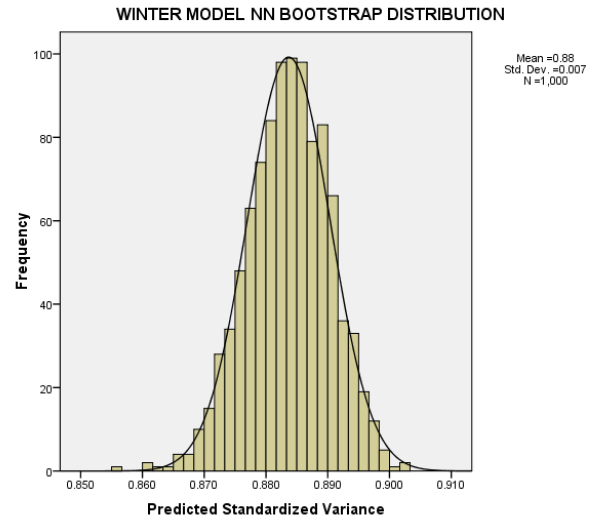
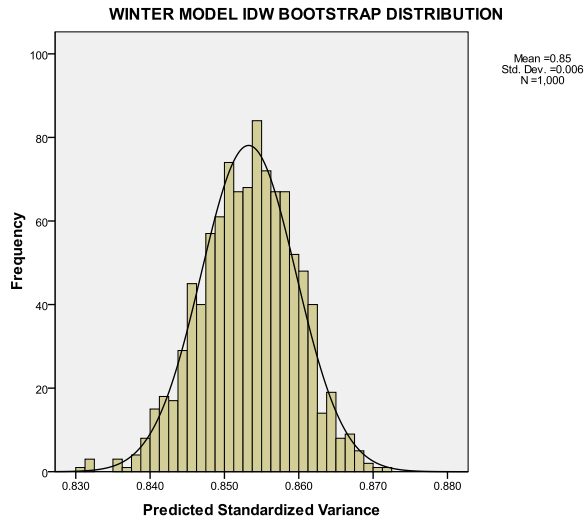
Bootstrap cross-validation for each interpolation technique of the annual and seasonal model.





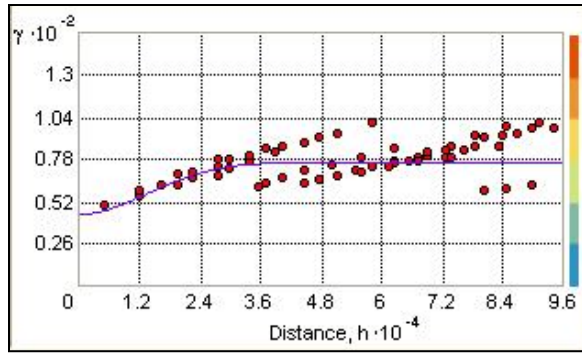




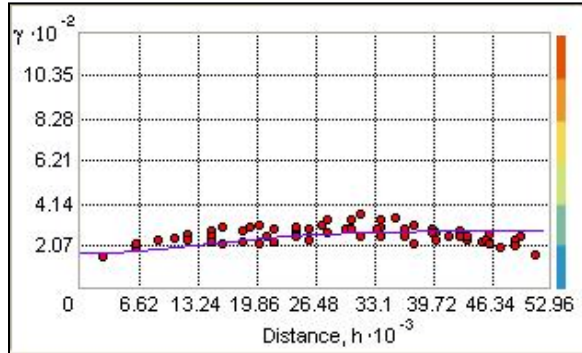


Appendix G

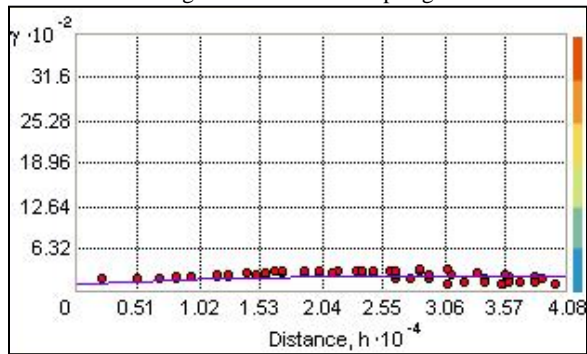
Semivariograms of ordinary and universal kriging methods for each model.



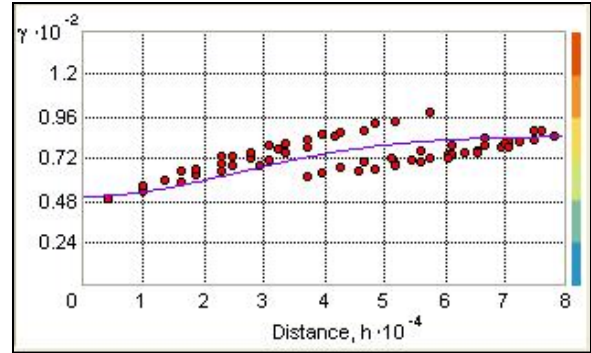
Semivariogram of OK for the annual model.



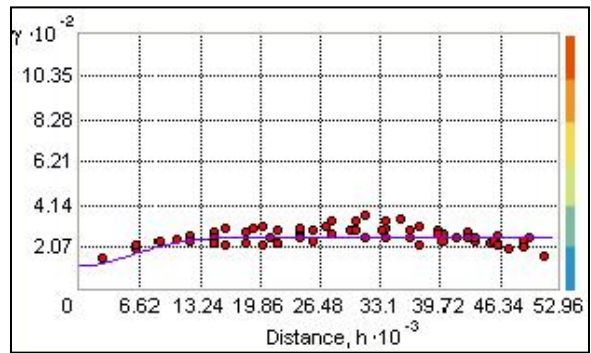
Semivariogram of OK for the spring model.



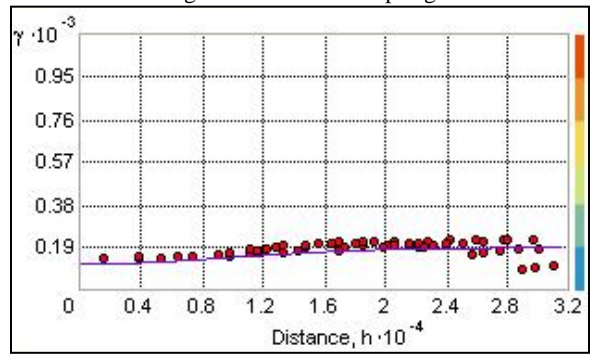
Semivariogram of OK for the summer model



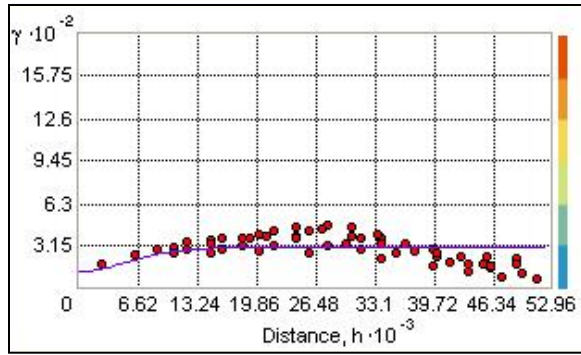
Semivariogram of OK for the annual model.



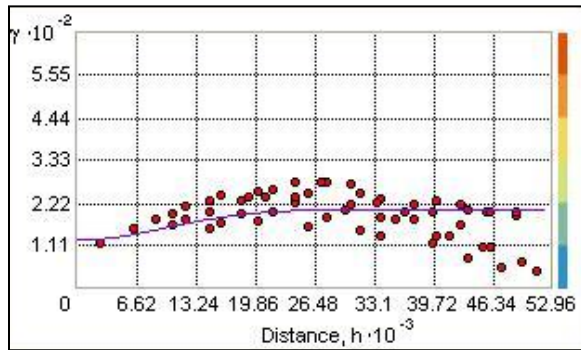
Semivariogram of UK for the spring model.



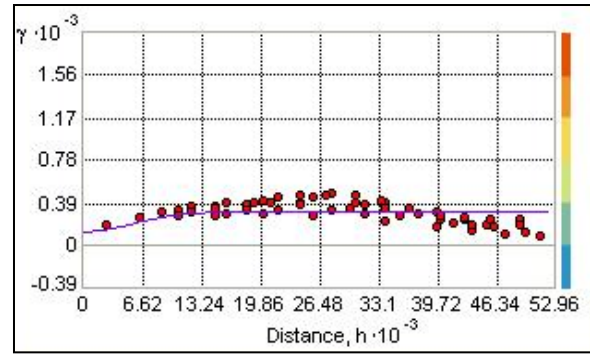
Semivariogram of UK for the summer model.



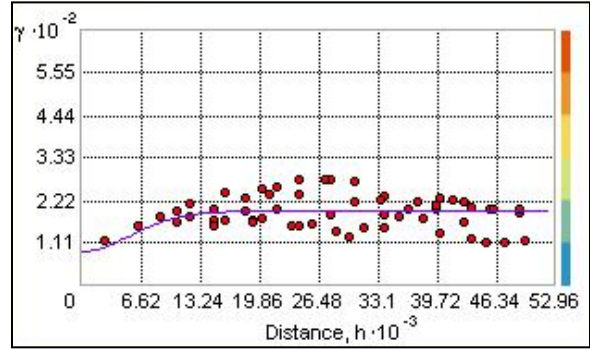
Semivariogram of UK for the fall model.



Semivariogram of UK for the winter model.



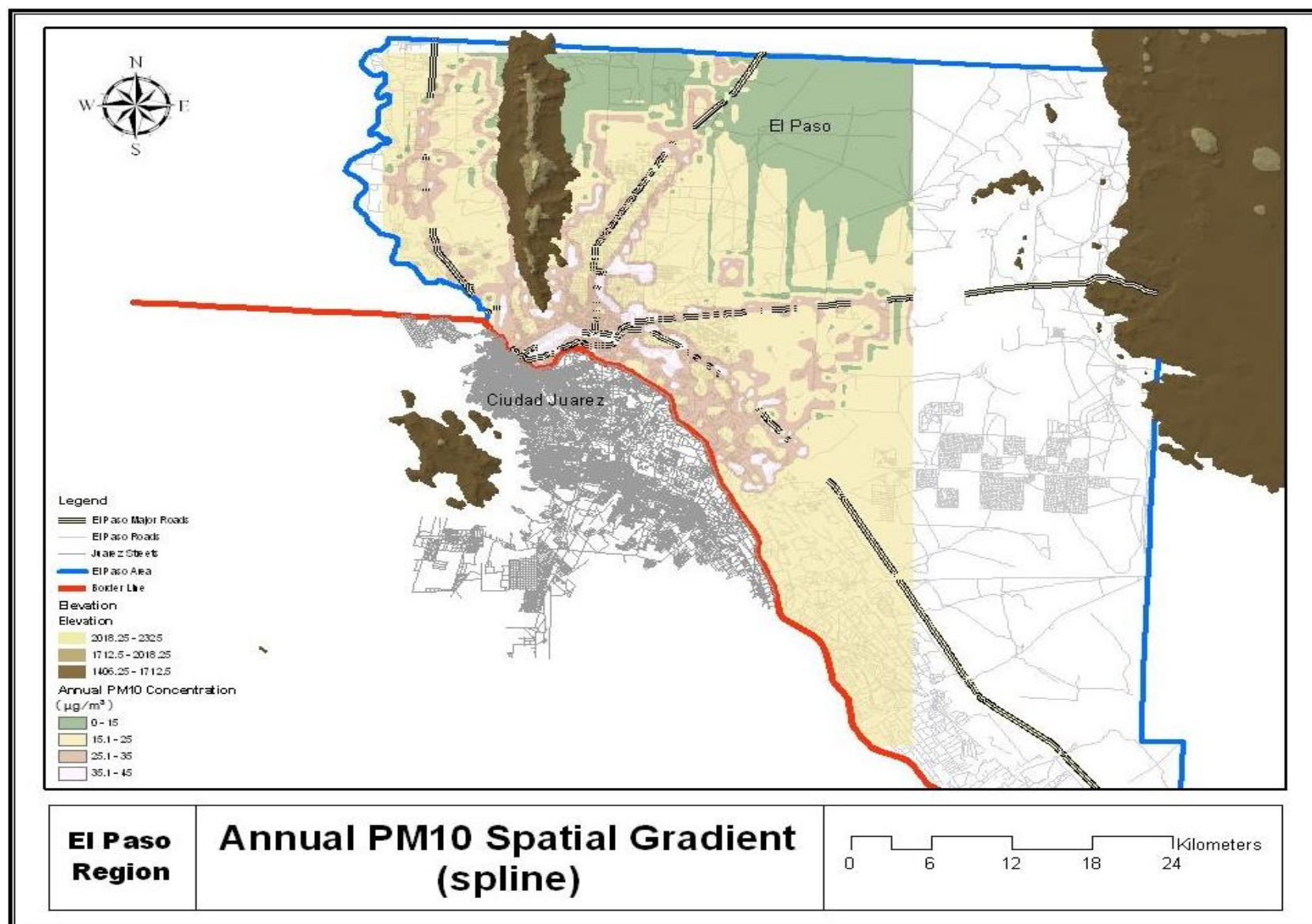
Semivariogram of OK for the fall model.

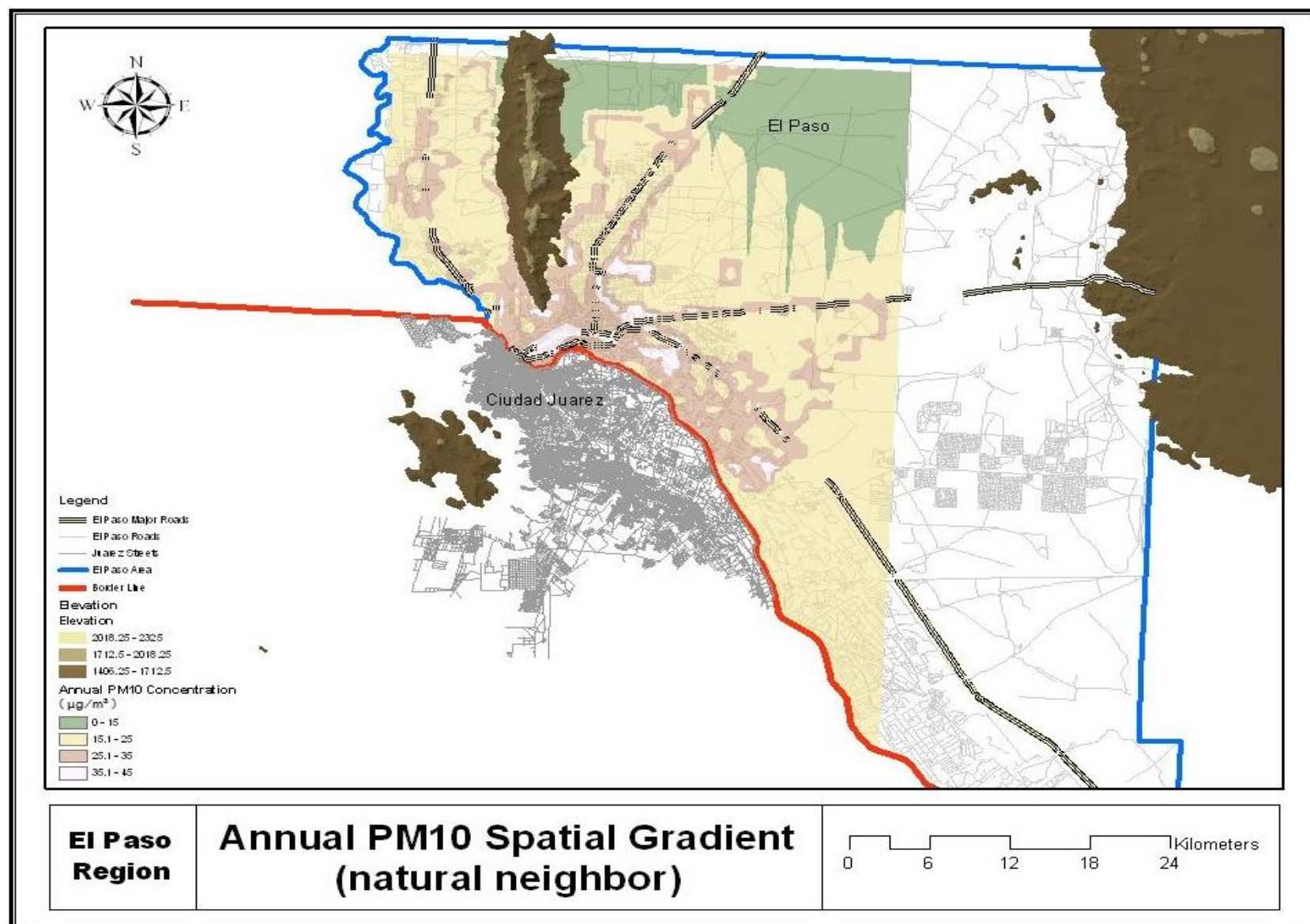


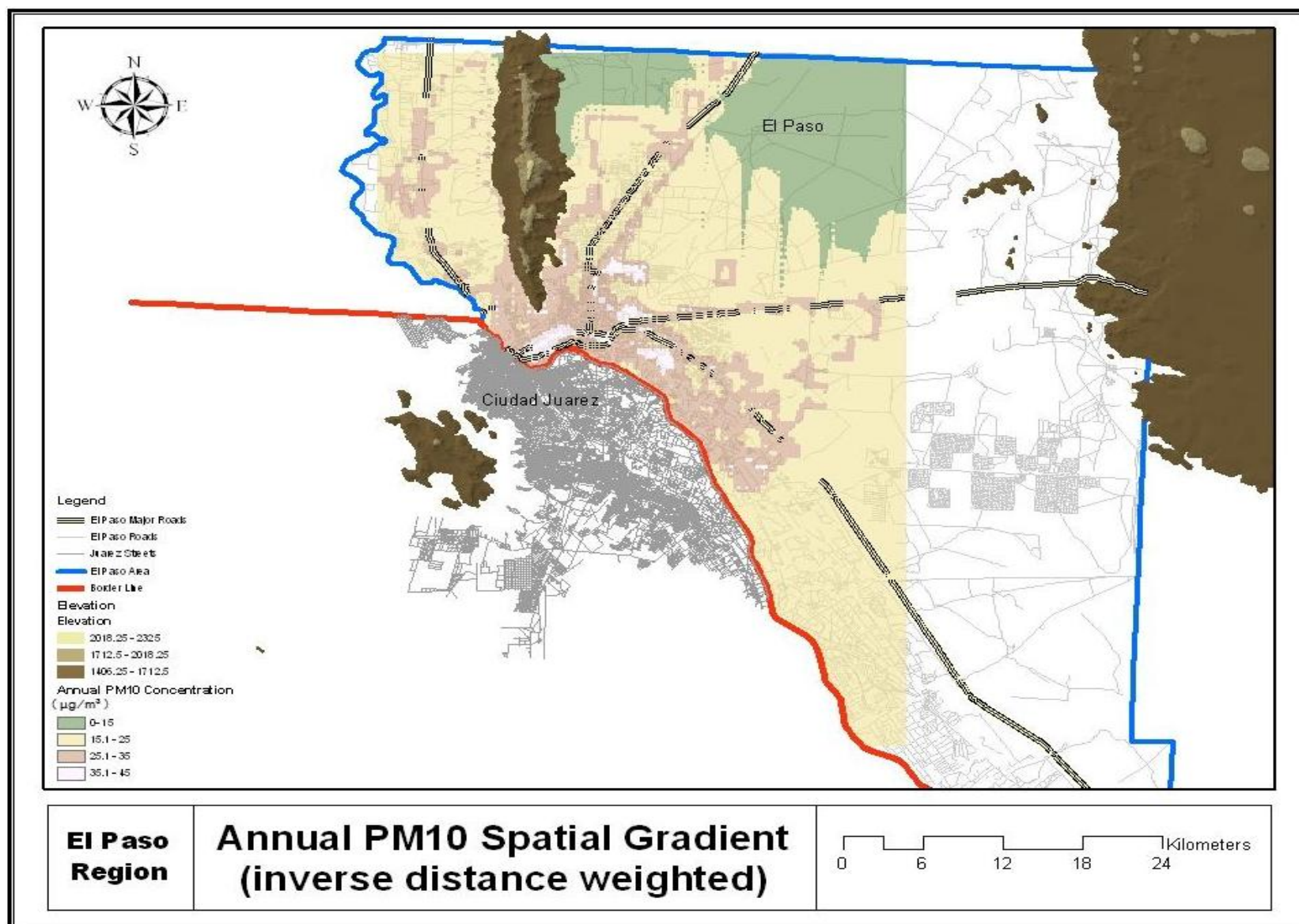
Semivariogram of OK for the winter model.

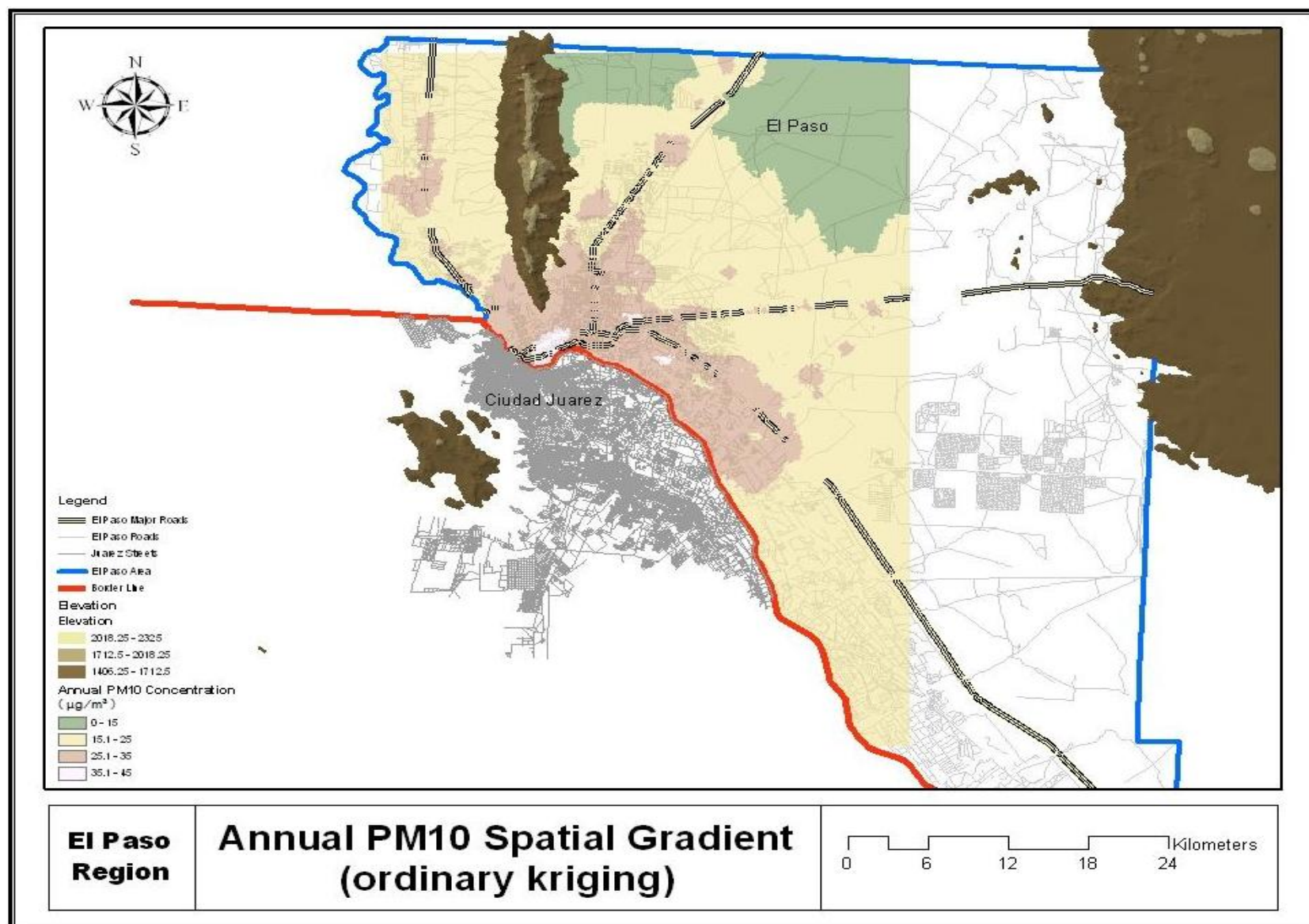
Appendix H

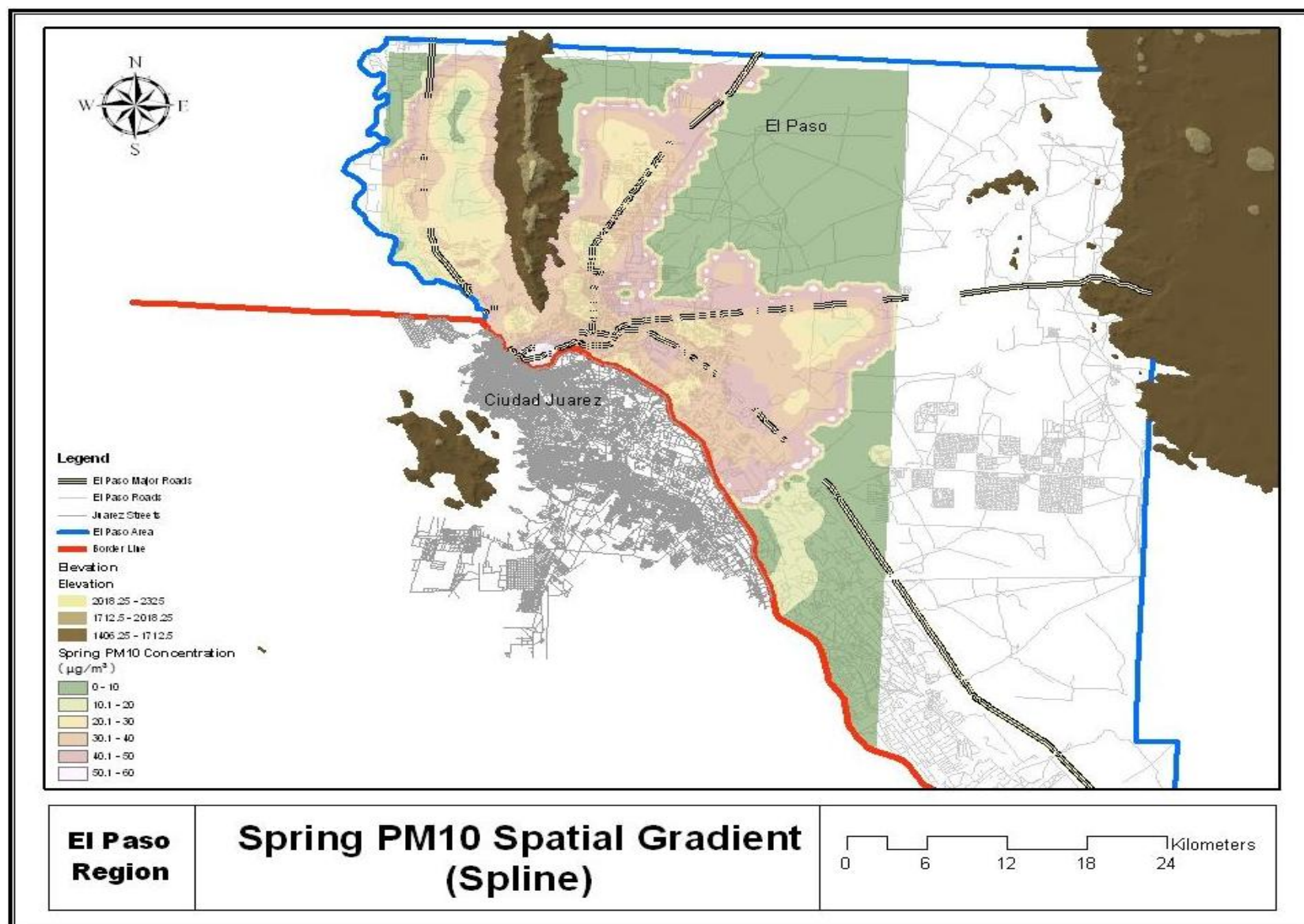
Spatial PM₁₀ gradient for each interpolation technique of the annual and seasonal models.

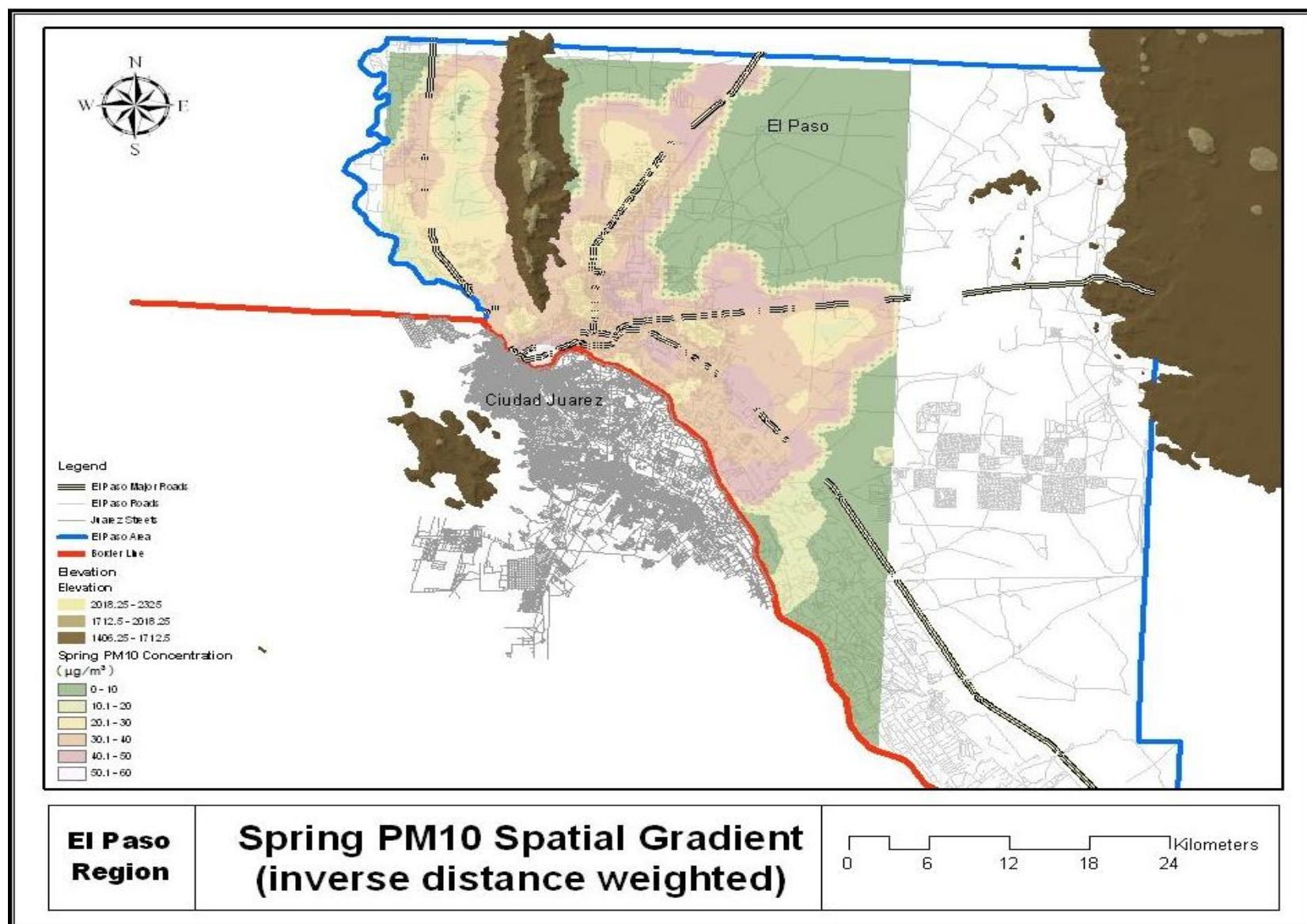


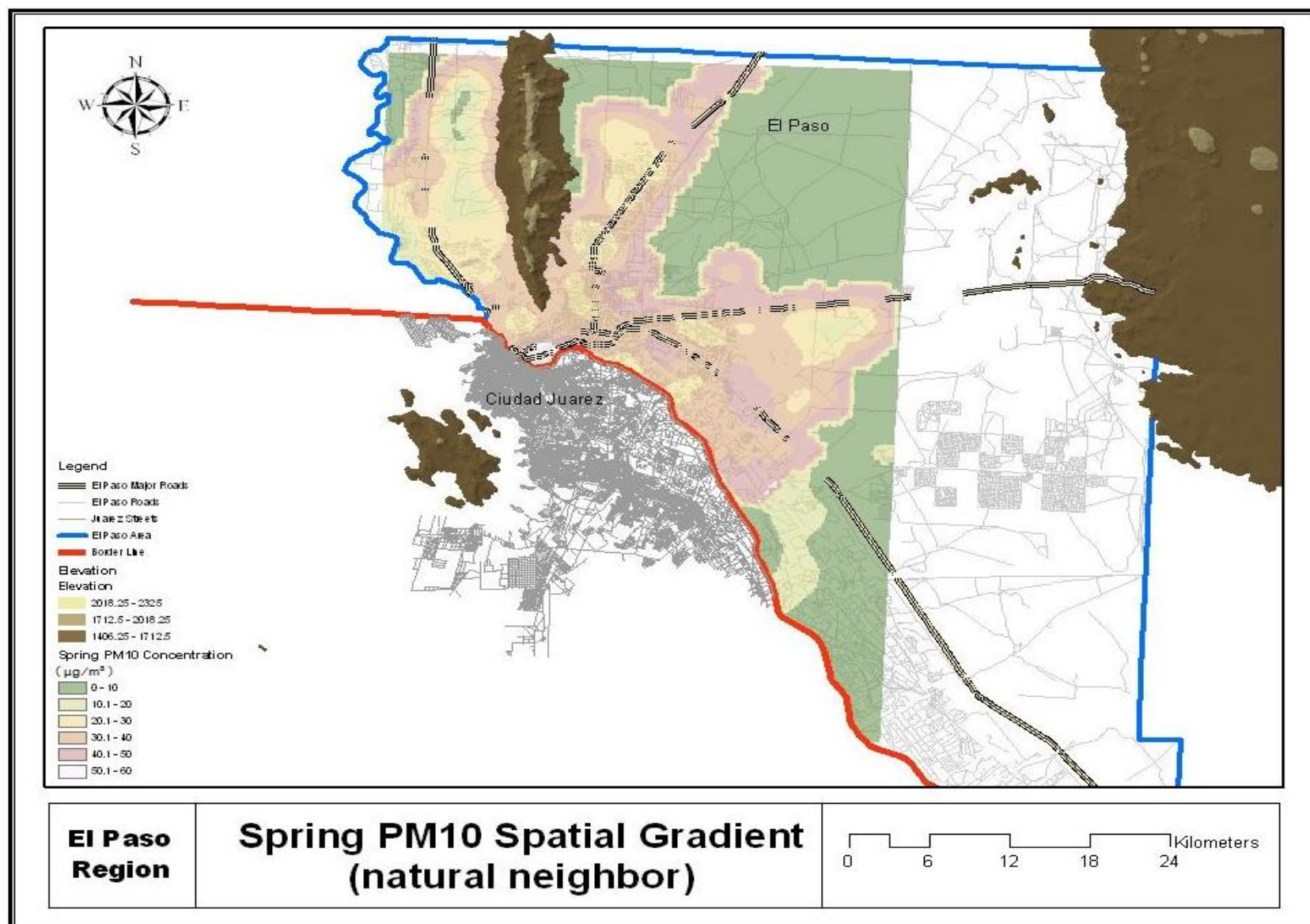


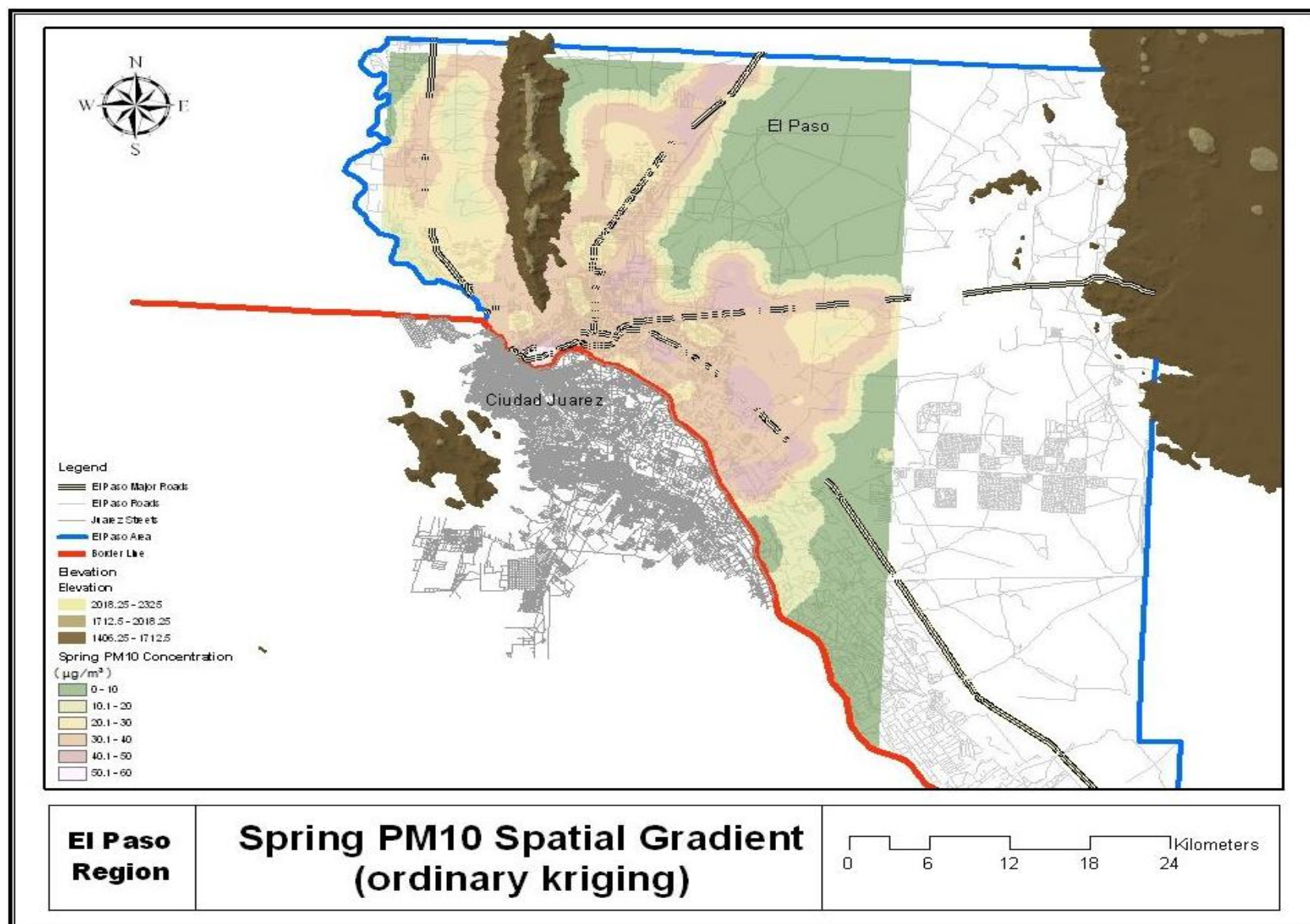


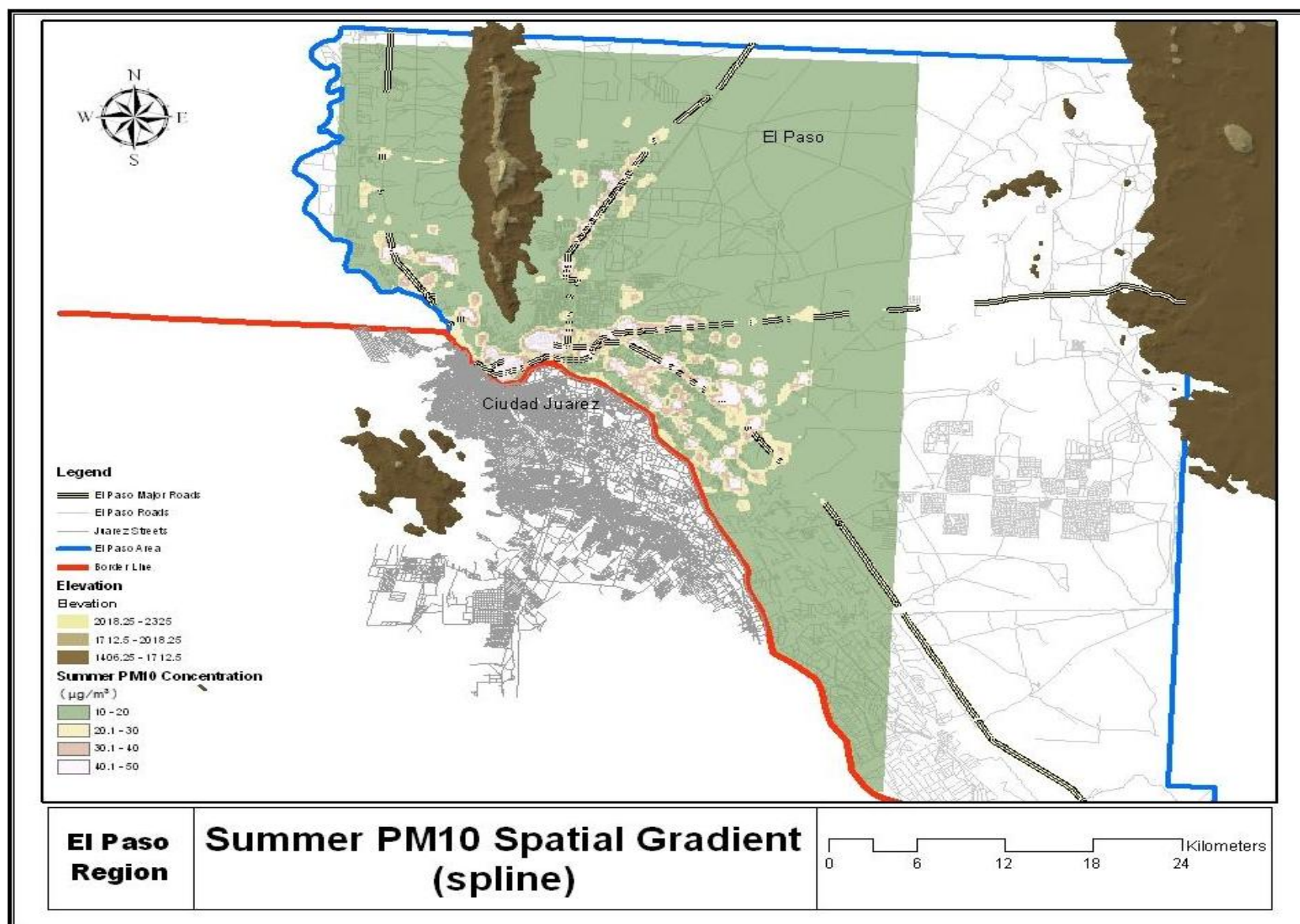


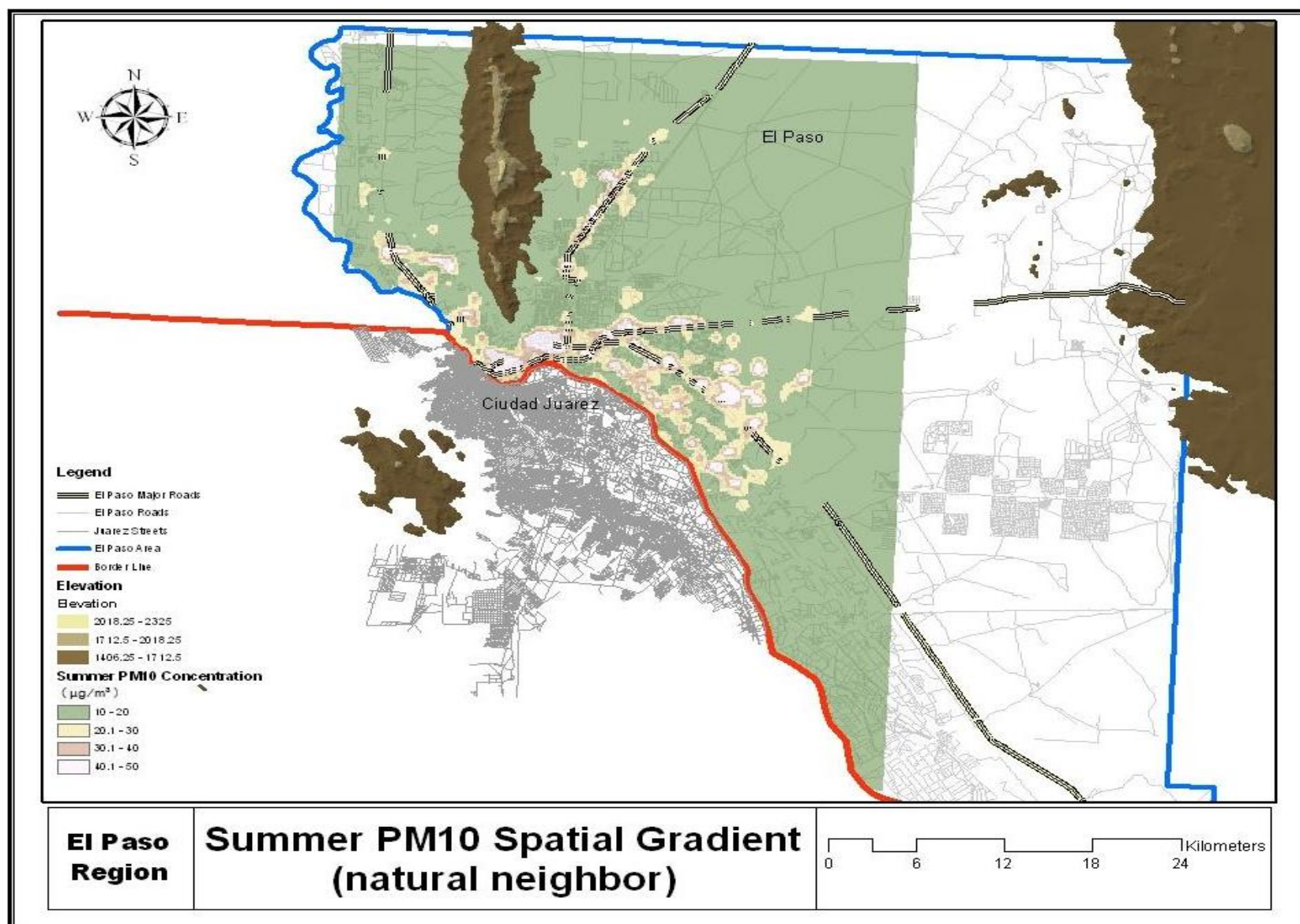


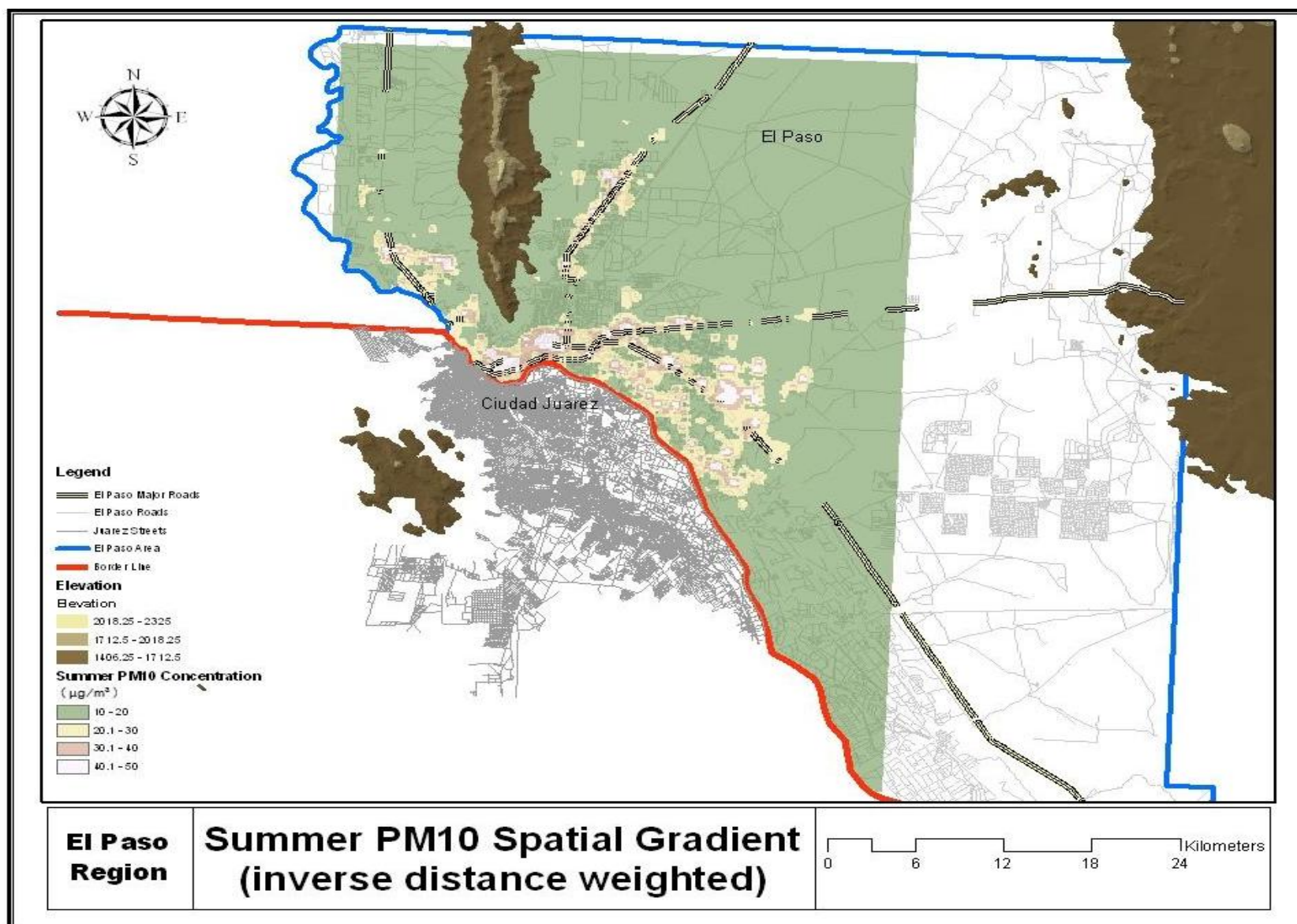


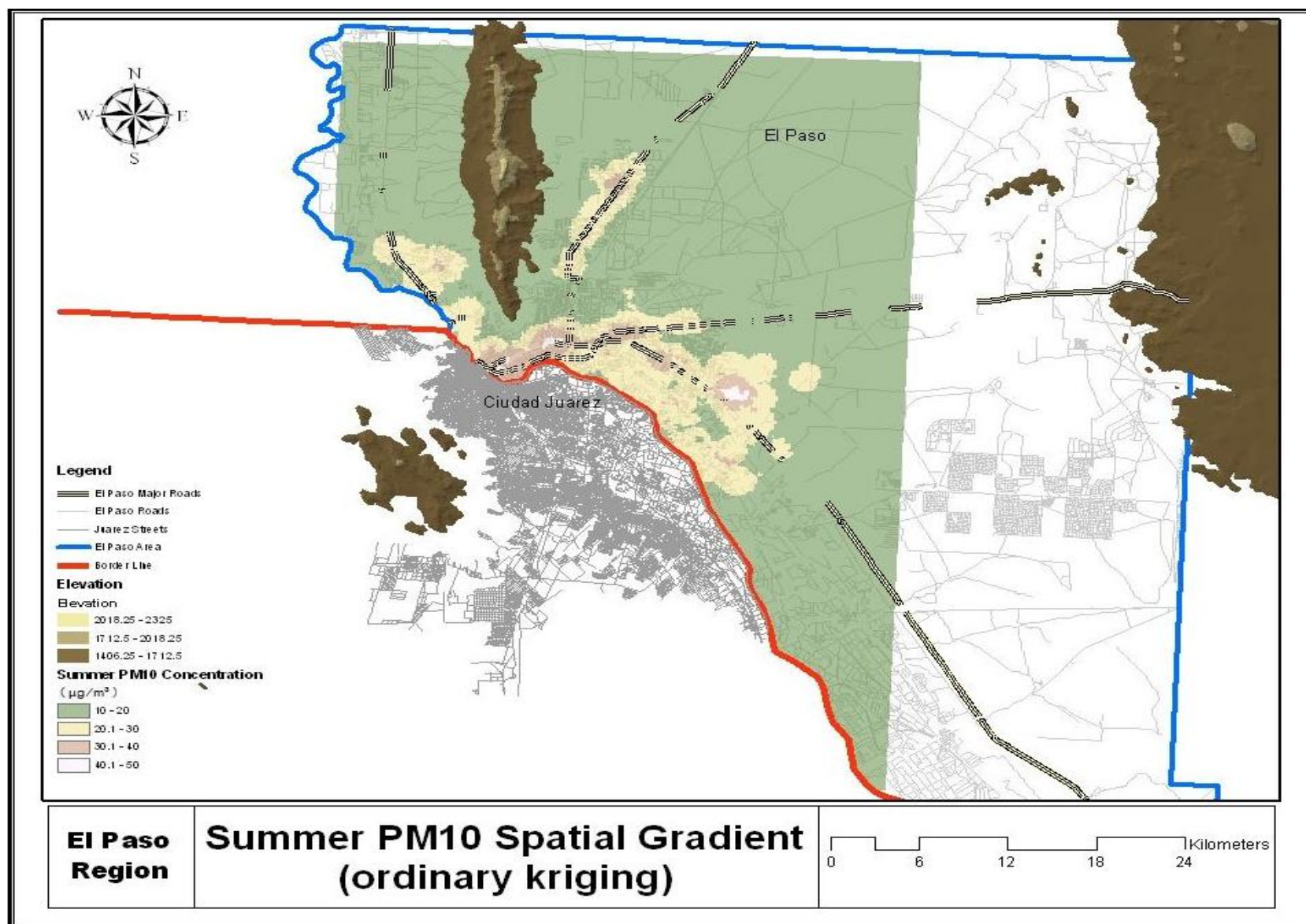


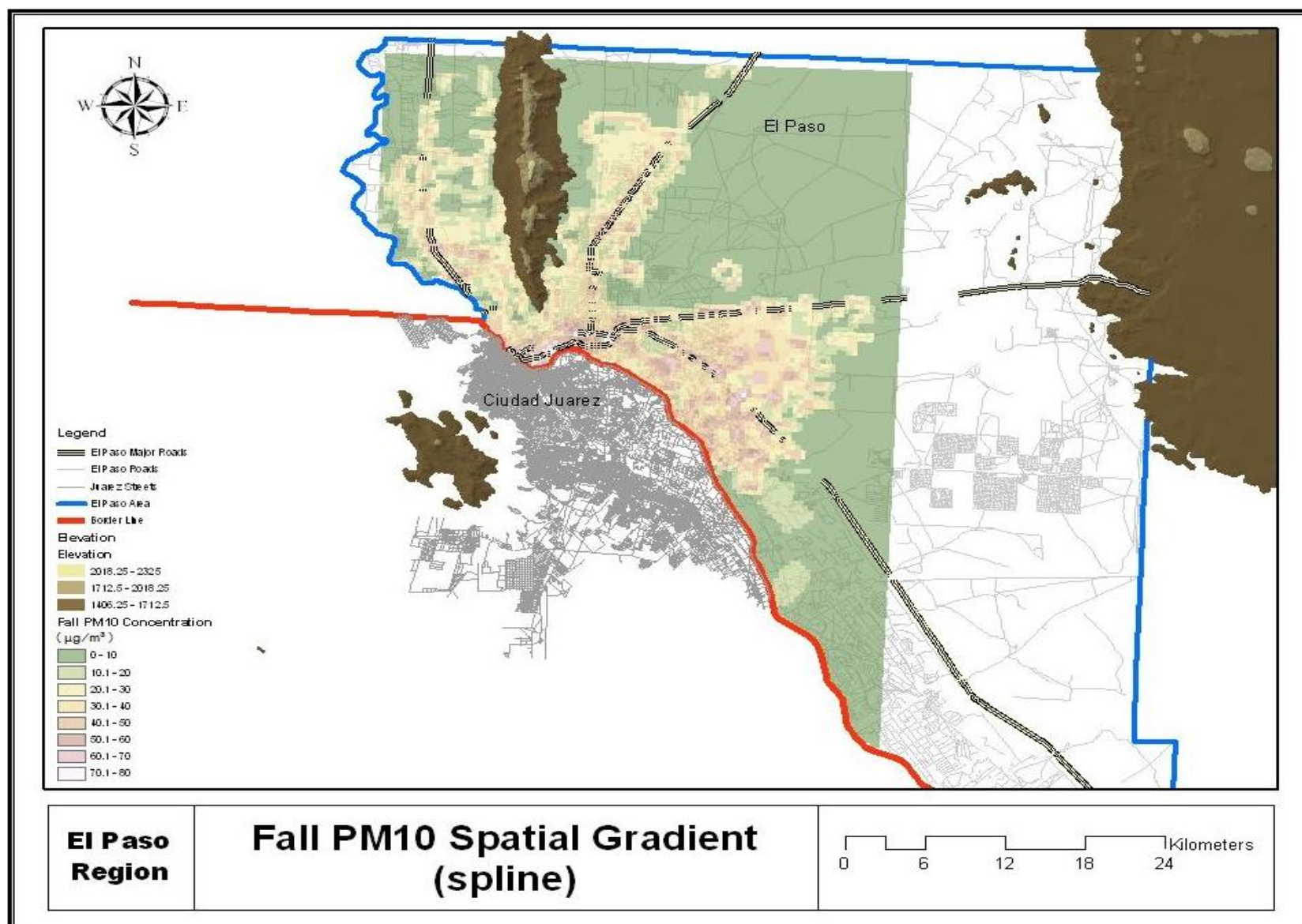


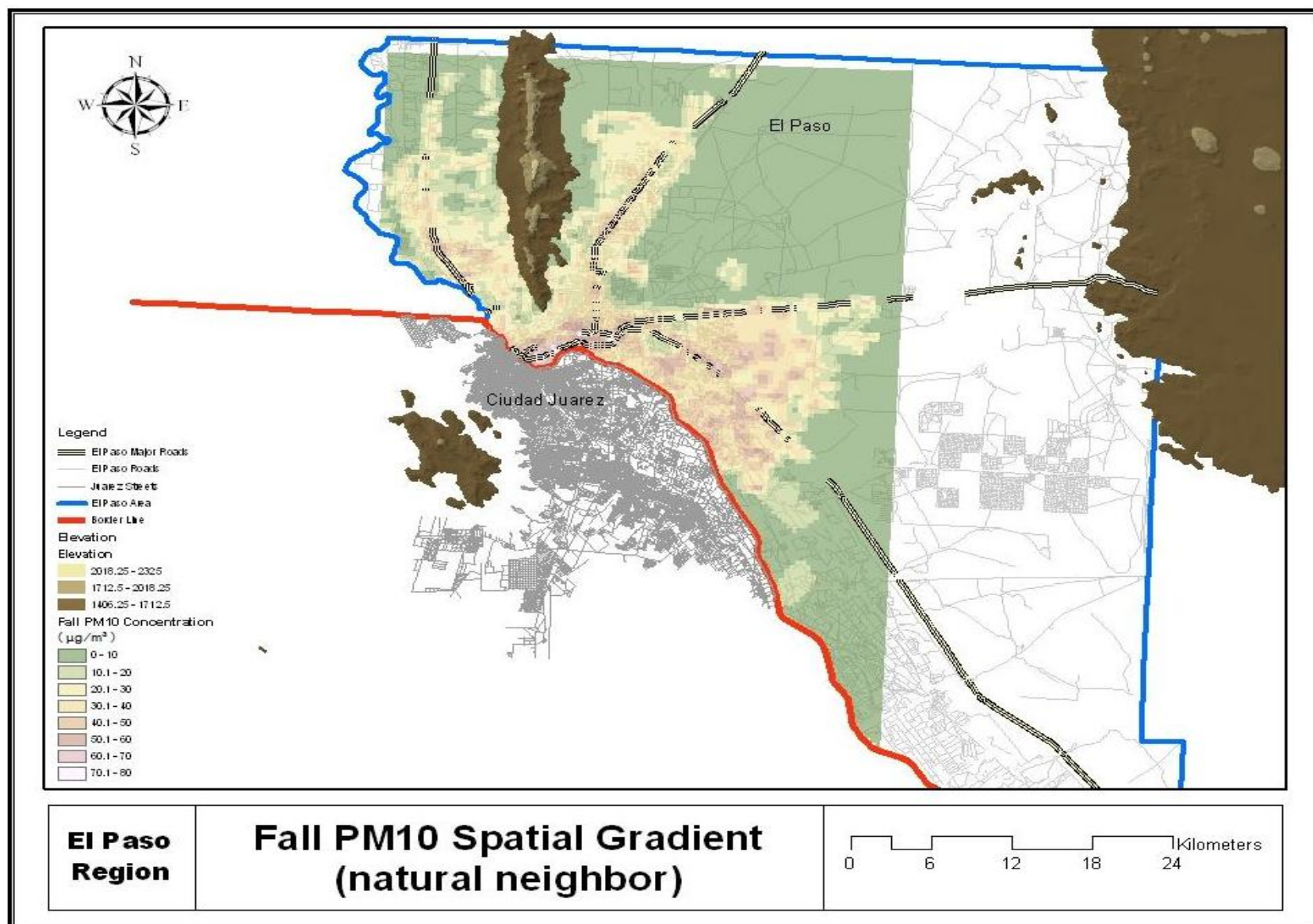


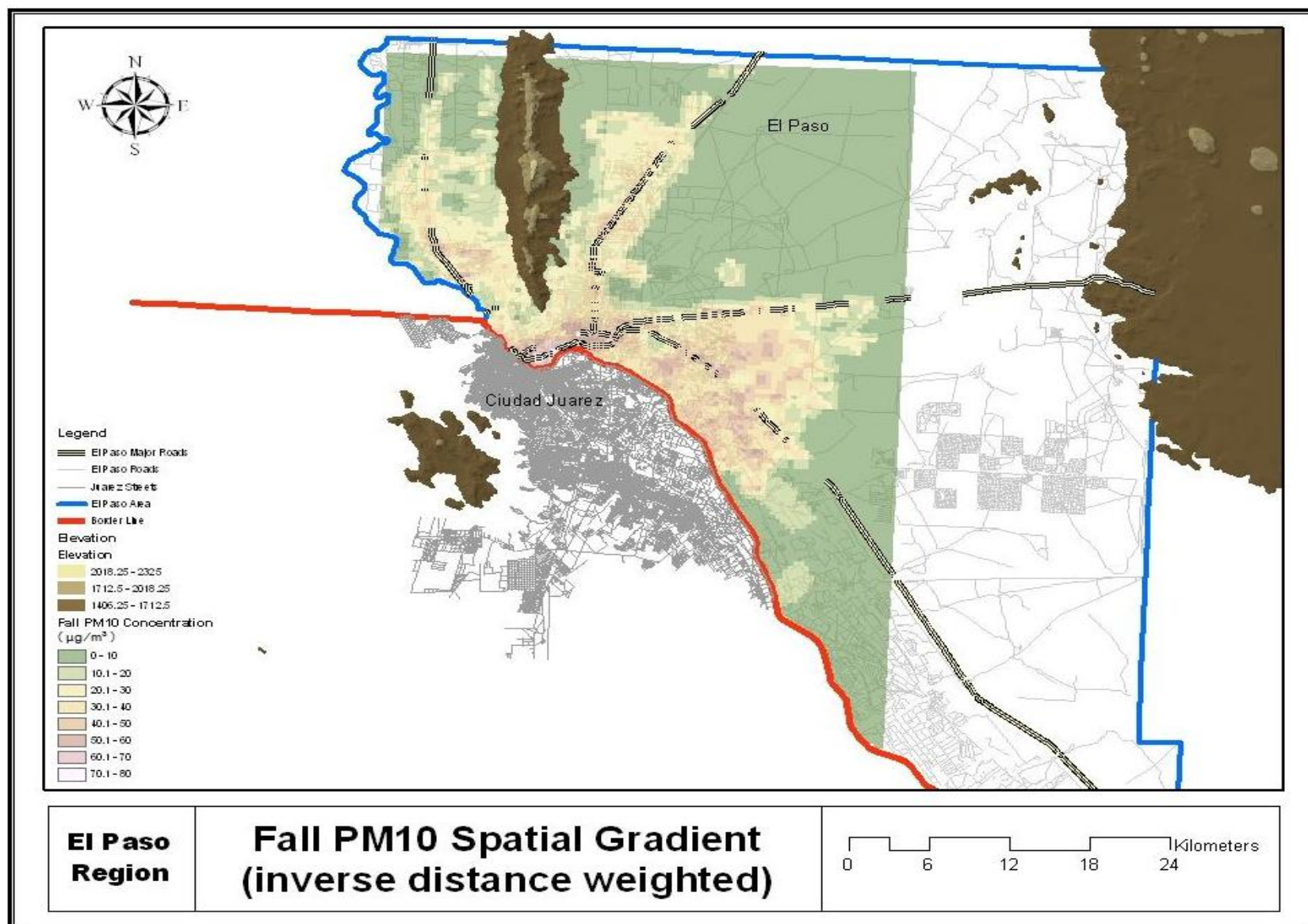


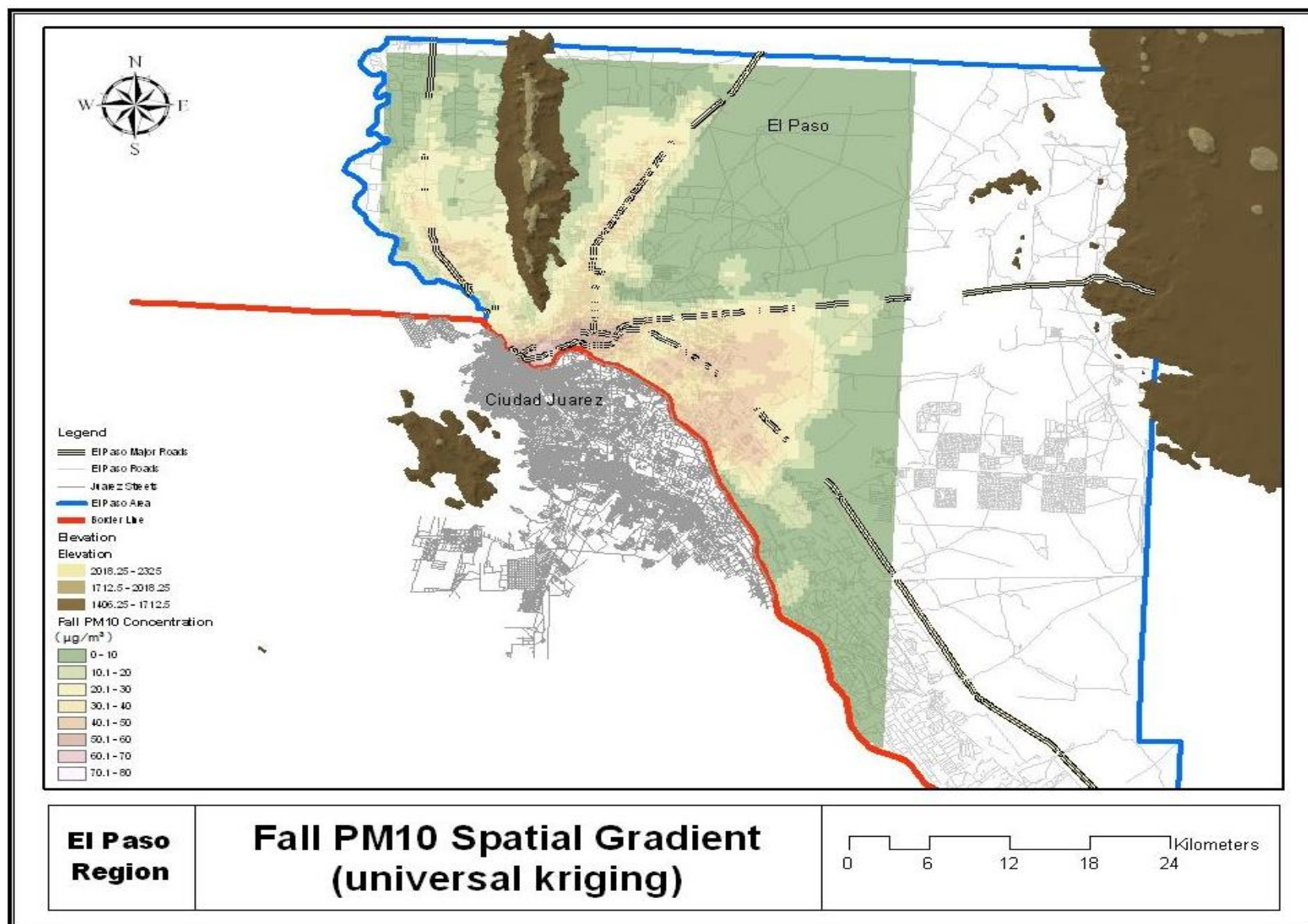


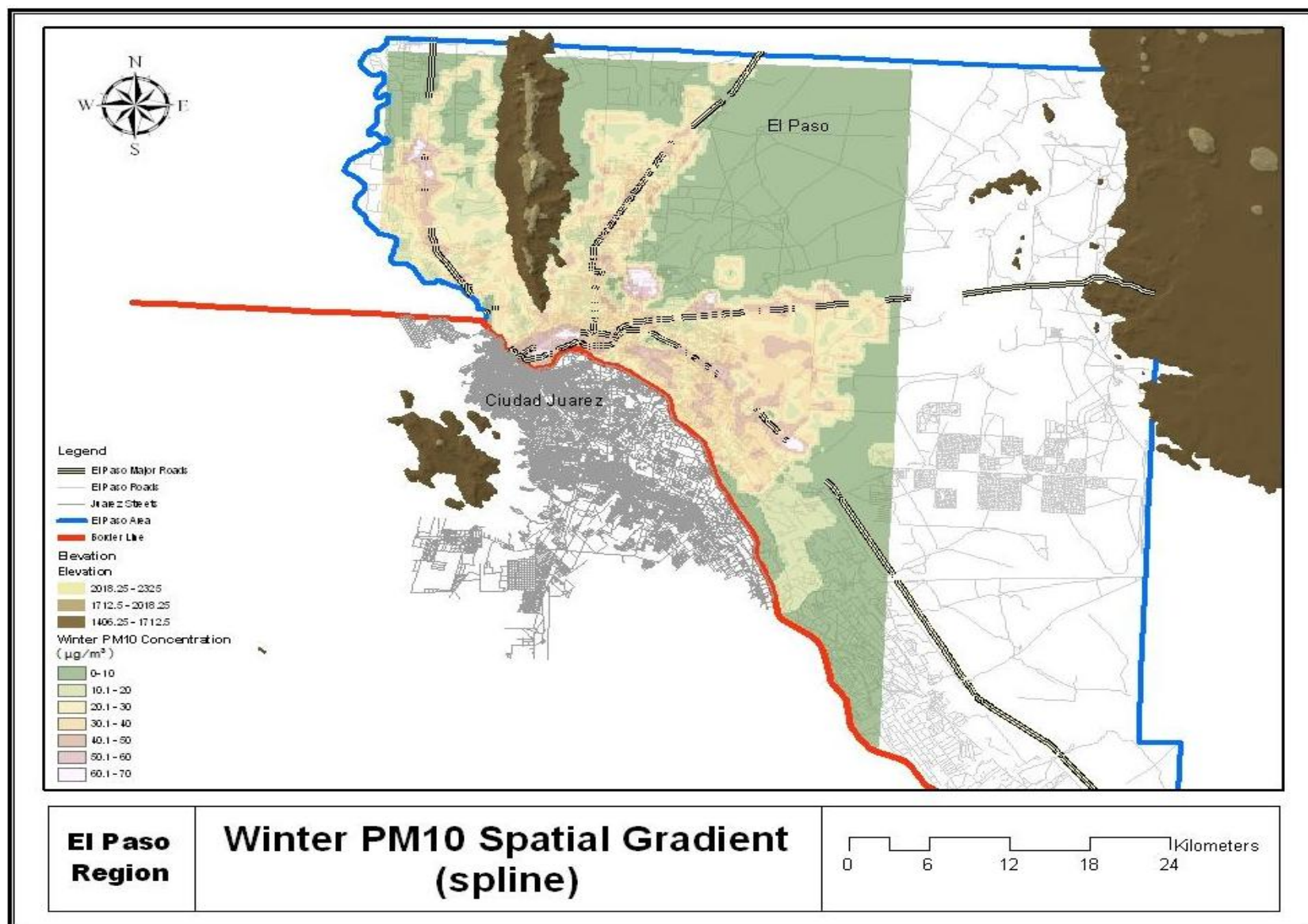


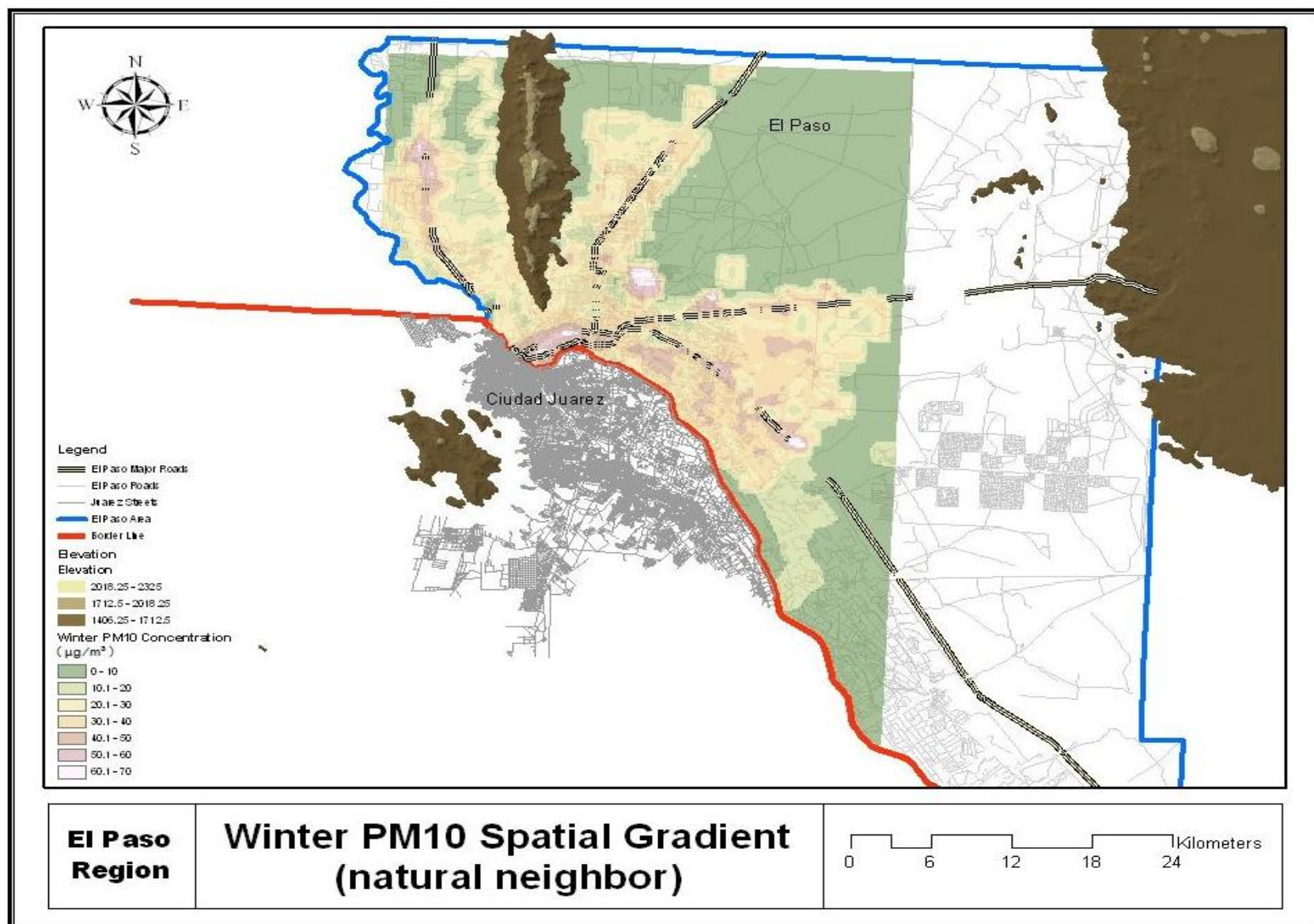


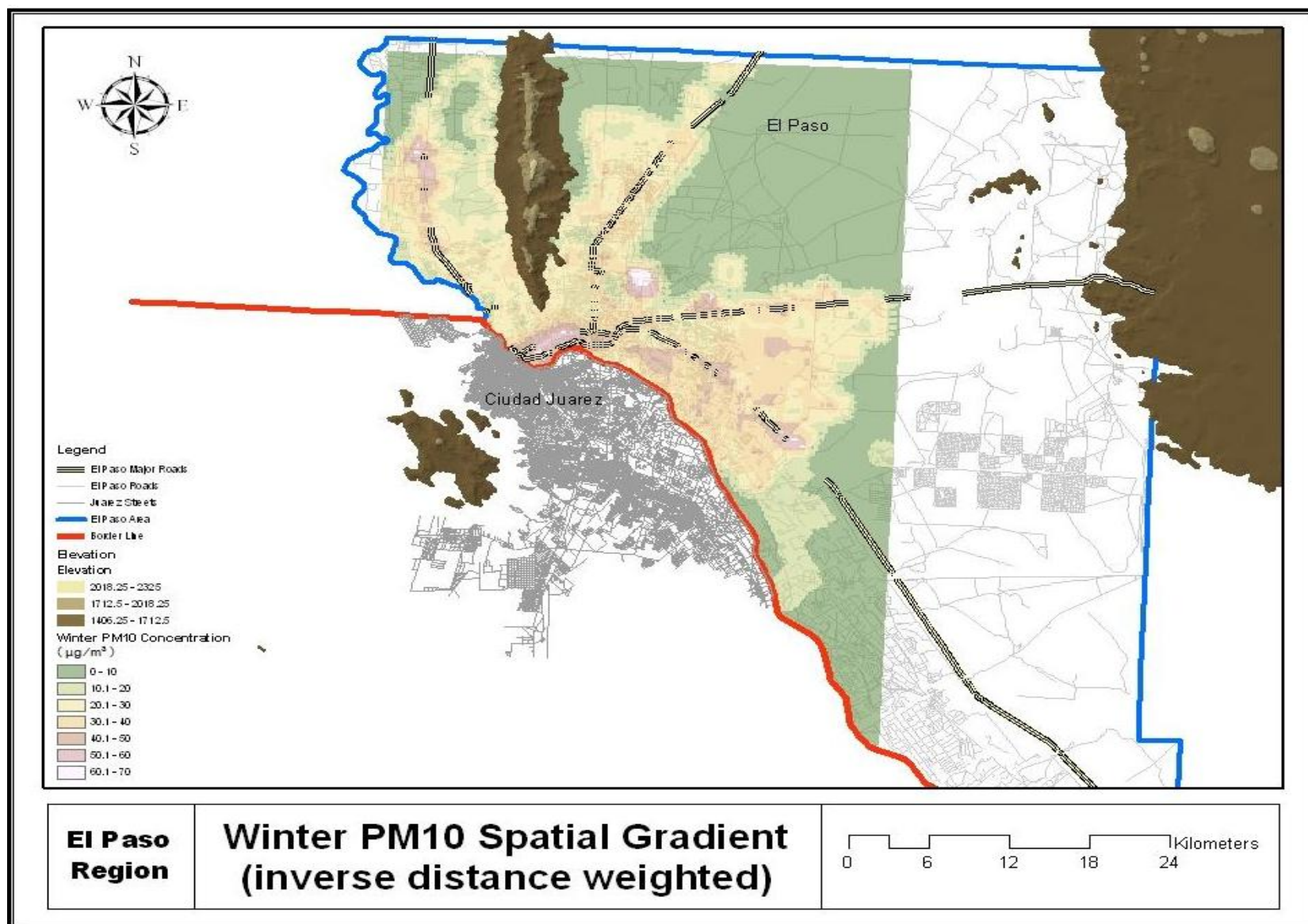


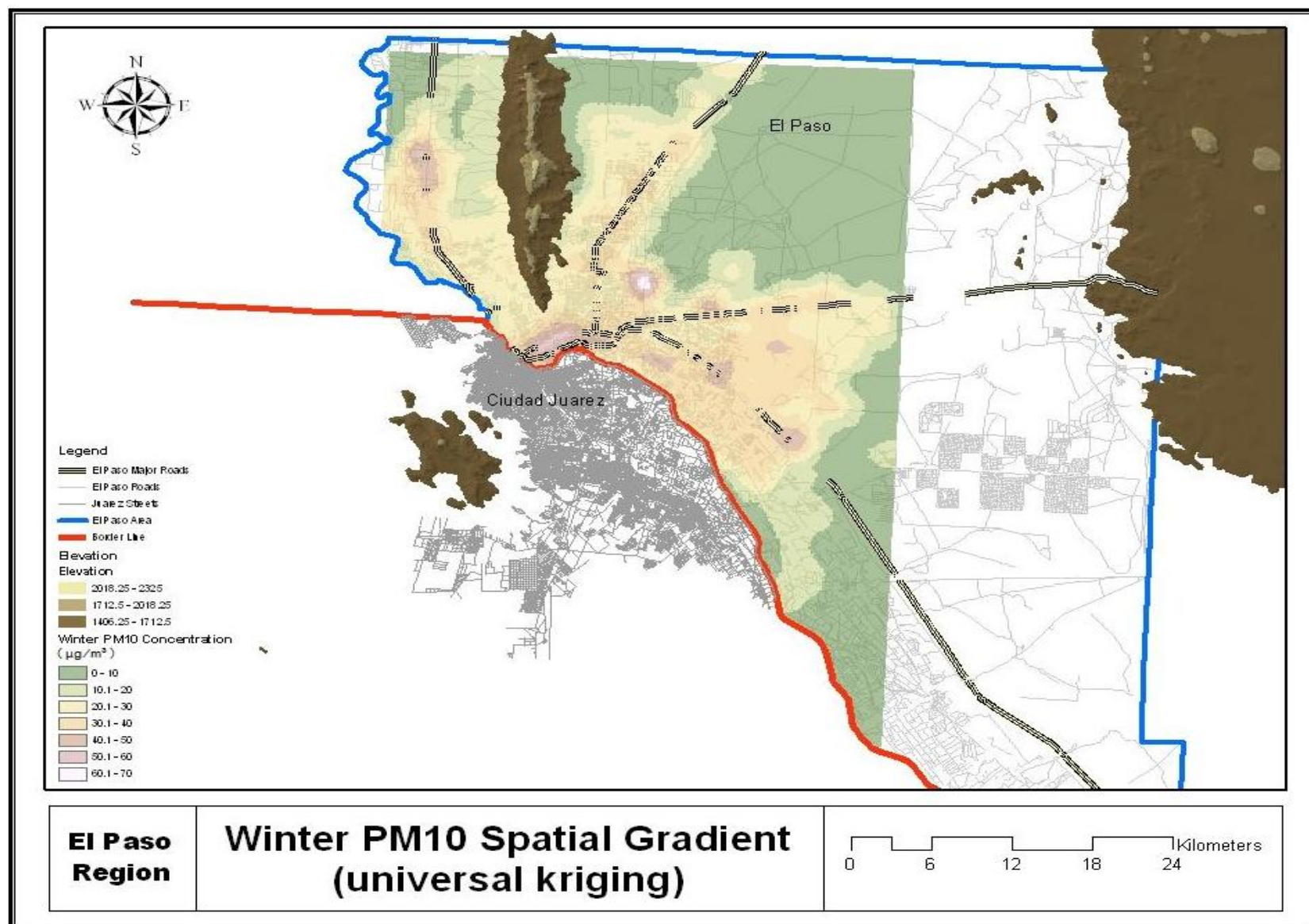












CURRICULUM VITAE

Mario I. Garcia was born and raised in Ciudad Juarez, Chihuahua, Mexico on March 27th, 1983. Born the oldest child of two to Mario Ruben Garcia and Marcela Eugenia Martinez, he was a three year graduate at Preparatoria El Chamizal in 2001. After receiving the High School degree from the Secretaria de Educacion Publica in Ciudad Juarez, he entered The University of Texas at El Paso, graduating with his bachelor's degree in Civil Engineer in May 2007. While pursuing his bachelor's degree he worked as a research assistant to the Civil Engineering Department Chair, Dr. Wen Whai Li at the Air Quality Research Lab. In the fall of 2007, he entered the Graduate School at the University of Texas at El Paso where he continued his research in the field of air quality until March of 2009. Among his accomplishments at this position he conducted the pilot project for NIH "Indoor and Outdoor PM₁₀ mass concentration", he also work in the SCERP "assessing pollutants in schools and asthma events in children", participated with the Desert Research Institute (DRI) "re-evaluating pollution in the southwestern part of the U.S.", and lastly he work and co-published the design of a LUR model to assess the spatial variability of PM_{2.5} in the El Paso area. He left the research air quality lab in March of 2009 for an internship with The Border Environmental Cooperation Commission where he assisted in the NEPA and EPA certifications of the El Paso water runoff project, Los Alcatrazes wastewater collection system, and Yuma water and wastewater collection system. Mario I. Garcia earned his Master's degree of Science in Environmental Engineering in the spring of 2010 from The University of Texas at El Paso where his thesis defense, "Assessing Annual and Seasonal Spatial Variability of Ambient PM₁₀ Using Linear Regression Analysis in a United States-Mexico Urban Sprawl" was supervised by Dr. Wen Whai Li and Dr. Charles D.

Turner from the Civil Engineer department as well as Dr. Hector A. Olvera from the Center for Environmental Resource Management.

Permanent address: Mision de Santa Rita #8033
Fraccionamiento Las Misiones
Ciudad Juarez, Chihuahua, Mexico
C.P. 32530

This thesis was typed by Mario Ivan Garcia Martinez.